Multi-omics approaches define novel aphid effector candidates associated with virulence and avirulence phenotypes

Peter Thorpe[1], Simone Altmann[2], Rosa Lopez-Cobollo[3], Nadine Douglas[4,5], Javaid Iqbal[3], Sadia Kanvil[3], Jean-Christophe Simon[6], James C. Carolan[4], Jorunn Bos[2,7*], Colin Turnbull[3*]

[1]Division of Computational Biology, School of Life Sciences, University of Dundee, Dundee, DD5 4EH, UK; [2]Division of Plant Sciences, School of Life Sciences, University of Dundee, Dundee, DD5 4EH, UK; [3]Department of Life Sciences, Imperial College London, London SW7 2AZ, UK; [4]Department of Biology, Maynooth University, Republic of Ireland; [5]School of Biology and Environmental Science, University College Dublin, Dublin 2, Republic of Ireland; [6]Institut de Génétique, Environnement et Protection des Plantes (IGEPP), INRAE, 35653 Le Rheu, France; [7]The James Hutton Institute, Invergowrie, Dundee DD2 5DA, UK. *Authors for correspondence: j.bos@dundee.ac.uk, c.turnbull@imperial.ac.uk.

ABSTRACT

Background. Compatibility between plant parasites and their hosts is genetically determined by both interacting organisms. For example, plants may carry resistance (R) genes or deploy chemical defences. Aphid saliva contains many proteins that are secreted into host tissues. Subsets of these proteins are predicted to act as effectors, either subverting or triggering host immunity. However, associating particular effectors with virulence or avirulence outcomes presents challenges due to the combinatorial complexity. Here we use defined aphid and host genetics to test for co-segregation of expressed aphid transcripts and proteins with virulent or avirulent phenotypes.

Results. We compared virulent and avirulent pea aphid parental genotypes, and their bulk segregant F1 progeny on *Medicago truncatula* genotypes carrying or lacking the *RAP1* resistance quantitative trait locus. Differential gene expression analysis of whole body and head samples, in combination with proteomics of saliva and salivary glands, enabled us to pinpoint proteins associated with virulence/avirulence phenotypes. There was relatively little impact of host genotype, whereas large numbers of transcripts and proteins were differentially expressed between parental aphids, likely a reflection of their classification as divergent biotypes within the pea aphid species complex. Many fewer transcripts intersected with the equivalent differential expression patterns in the bulked F1 progeny, providing an effective filter for removing genomic background effects. Overall, there were more upregulated genes detected in the F1 avirulent dataset compared with the virulent one. Some genes were differentially expressed both in the transcriptome and in the proteome datasets, with aminopeptidase N proteins being the most frequent differentially expressed family. In addition, a substantial proportion (27%) of salivary proteins lack annotations, suggesting that many novel functions remain to be discovered.

Conclusions. Especially when combined with tightly controlled genetics of both insect and host, multi-omics approaches are powerful tools for revealing and filtering candidate lists down to plausible genes for further functional analysis as putative aphid effectors.

KEYWORDS: Aphid, transcriptomics, proteomics, saliva, effector, virulence, avirulence

Background

Crop losses due to insect pests represent an enduring challenge for agriculture and global food security. Aphids are a major problematic group, due both to the direct damage they cause by phloem sap feeding and to indirect effects through acting as vectors for transmission of many viruses. Impacts of pests are further exacerbated by the breakdown of genetically based crop resistance mechanisms due to selection pressures driving pest evolution, as well as evolved insecticide resistance.

In contrast to related fields such as plant-pathogen interactions, the molecular relationships that determine (in)compatibility of plant-aphid interactions are relatively poorly understood. Specific resistance to plant pathogens frequently involves recognition of pathogen effectors, often by resistance proteins (R) characterised by nucleotide-binding and leucine rich repeat (NLR) domains. Several coiled coil domain NLR proteins have been implicated in resistance to aphids and their close relatives. For example, Mi-1, Vat and Bph14 confer resistance to certain biotypes of *Macrosiphum euphorbiae* (potato aphid) [1], *Aphis gossypii* (melon-cotton aphid) [2] and *Nilaparvata lugens* (brown planthopper) [3], respectively. These NLR receptors are predicted to be involved in direct or indirect recognition of molecular signatures that insects, like plant pathogens, release inside their hosts. Indeed, aphids secrete multiple effector proteins into their saliva, that are then predicted to be delivered into plant tissues to modulate host cell processes and to suppress or trigger host defences [4–7]. Although there is one recent report of the BISP effector from brown planthopper, an aphid relative, interacting with the BPH14 NLR in rice [8], there are currently no examples where cognate aphid effector and NLR pairs have been fully defined. Improved molecular insights into virulence and resistance mechanisms taking place during both compatible and incompatible plant-aphid interactions are therefore a priority, and can provide essential knowledge for future development of durable aphid control strategies.

The availability of extensive genome, transcriptome and resequencing resources for the model aphid species *Acyrthosiphon pisum* (pea aphid) [9, 10] have enabled comprehensive genome-wide explorations. There are also genomic sequences now available at NCBI and AphidBase (https://bipaa.genouest.org/is/aphidbase/) for more than 25 species of aphids and close relatives, often associated with gene predictions and transcriptomes [11]. In addition, several papers have

attempted to define the aphid effectorome, either by direct analysis of salivary proteins, or by transcriptomics of salivary glands, coupled with filters for predicted secreted, non-trans-membrane proteins [12–17]. Beyond the true aphids (superfamily Aphidoidea), there are now genomic resources for sister groups within the Hemiptera such as planthoppers, leafhoppers, psyllids, whitefly and scale insects (https://www.ncbi.nlm.nih.gov/assembly/?term=hemiptera) that likewise are major crop pests, alongside genomes for triatomines and bed bugs, hemipterans that feed on animal rather than plant hosts. Outside the Hemiptera, genomic data have been published for sucking pests such as thrips and spider mites that feed on plant tissues other than phloem [18–20]. Genome, transcriptome and proteome comparisons across clades may enable definition of putative effector subsets that are necessary for different feeding modes, and may provide insights into conserved and divergent modes of action in terms of how the plant immune system is targeted to enable successful parasitism.

Despite the wide range of functional genomics studies published to date, one common limitation is the lack of understanding of the differences in effector complements between virulent (host-compatible) and avirulent (host-incompatible) genotypes. Genetic differences operate at several taxonomic levels. First, there are major differences across aphid species in their host preferences and host compatibilities. Some species, such as peach potato aphid (*Myzus persicae*) are generalists that can feed on at least 400 known plant species, making them widespread crop pests [21]. Others are specialists, such as pea aphid (*A. pisum*) that exclusively feeds on legumes (Fabaceae). Second, there is substantial diversity within species such as *A. pisum* that has led to its description as a species complex comprising several host races that each have a strong preference for particular legume species, supported by robust molecular marker fingerprints for each host race [22, 23]. There is evidence of divergence and differential expression of chemosensory gene families such as odorant receptors across different pea aphid biotypes [24, 25], but causative relationships have yet to be established for genes and proteins that govern the range of compatible and incompatible interactions seen. There is also clear evidence that some host races can survive and sometimes thrive as migrants on hosts outside their preferred species range [22]. Finally, at the intra-specific level for both aphids and hosts, there can be a wide range of compatibilities. For example, from testing eight genotypes of *A. pisum* in combination with 23 different *Medicago truncatula* (Mt) accessions, we discovered high diversity in both species that did not correspond particularly strongly to host races or to geographic origins of the host lines [26]. Parallel to this, crossing two divergent pea aphid biotypes to generate F1 recombinant populations uncovered Mendelian segregation of virulence/avirulence on Mt genotypes carrying the *RAP1* aphid resistance QTL [27, 28].

Here, we report global exploration of the molecular basis for aphid virulence and avirulence on defined host genotypes. Specifically, we aimed to link phenotypes to candidate effectors and related

genes by multiple comparisons of the transcriptomes and proteomes of two divergent parental pea aphid clones, along with the transcriptomes of segregating avirulent and virulent pooled individuals from within F1 cross populations (Fig. 1). We also critically analysed the effectiveness of combined omics approaches as a means to robustly uncover proteins with pivotal biological roles, such as effectors that determine the difference between virulent and avirulent outcomes.

Results and Discussion

Generation and analysis of aphid populations for RNA-Seq analyses

In our previous work [27], we had demonstrated Mendelian segregation of inheritance of virulent and avirulent phenotypes in F1 pea aphid populations derived from a cross between N116 and PS01 (virulent and avirulent parental clones, respectively) when infested on *M. truncatula* hosts carrying the *RAP1* resistance QTL [28]. On this basis, we reasoned that the molecular basis of the difference between virulent and avirulent aphids could be revealed by transcriptomic and proteomic analysis. However, there were likely to be thousands of genetic and gene expression differences between the parental genotypes, that are representatives of phenotypically contrasting biotypes within the highly diverse pea aphid species complex [22, 26]. This makes it difficult to discern unrelated genomic background differences from causative genes responsible for suppressing host immunity or for triggering R-gene dependent defences. To address this challenge, we employed a bulk segregant analysis (BSA-) RNA-Seq approach that would both reduce the genetic background effects and allow us to test for heritability of differentially expressed (DE) genes across parental and F1 generations. Enabling this strategy first required us to re-create the segregating F1 populations previously reported [27].

We induced sexual forms of PS01 and N116 and conducted reciprocal crosses, leading to screening of a total of 78 F1 clones on two host plant genotypes carrying *RAP1*: Jemalong A17 (hereafter A17), the original source of the identified *RAP1* QTL, and a resistant near-isogenic line (RNIL) derived from a mapping population [29] using A17 as one of the parents. The *RAP1* aphid resistance QTL is highly effective against PS01 aphids, typically resulting in high mortality, whereas N116 aphids are unaffected. Progeny were verified as true F1 hybrids by a panel of seven SSR markers [22] and by screening for maternal inheritance of secondary symbionts reported in the pea aphid [30]. Using a virulence index based on a combination of aphid survival and reproduction, F1 clones were first ranked according to performance on A17. Phenotypes ranged from fully virulent to fully avirulent (Supplementary Material 1A), similar to previous findings [27], although in the present experiment the population as a whole did not display complete segregation into discrete virulent and avirulent categories. As also previously shown, resistance in the RNIL was slightly weaker than in A17, with F1 clones ranging from virulent to avirulent, and importantly performance on the two host genotypes

was significantly correlated (Pearson r 0.72, P $1.82e^{-13}$). All F1 clones were virulent on hosts lacking *RAP1* (Supplementary Material 1B). We then selected 22 sibling F1 clones from each end of the distribution to provide two bulk sample sets with the strongest virulent (VIR) and avirulent (AVR) phenotypes for subsequent transcriptomic analysis. Fig. 2 shows the complete separation of the selected clones into virulent and avirulent classifications. As a final check prior to RNA-Seq experiments, we re-confirmed separation of survival rates of these two subsets of clones on both resistant host genotypes (Supplementary Material 1C).

Transcriptomic analyses

We first ran an RNA-Seq experiment using the parental clones N116 and PS01 infested onto either A17 or the susceptible DZA315.16 host (hereafter DZA) for 24 h prior to collection of heads for RNA extraction. The multiple aims were to enrich for transcripts from salivary glands that express candidate effectors, to uncover the transcriptome differences between the parental aphid genotypes, and to reveal the impact of host plant genotype. Each aphid x host combination was replicated three times, giving a total of 12 libraries, ranging from 6.8 to 10.6 million reads uniquely mapped to the reference genome (Supplementary Material 2A).

Hierarchical clustering and principal components analysis (PCA) of the transcriptomic expression profiles both indicated that the replicates of each treatment were closely correlated in all cases, so no datasets needed to be discarded (Fig. 3A,B). These analyses additionally revealed that samples were separated largely by aphid genotype rather than host plant treatment. Overall, the transcriptomes of the two aphid genotypes on A17 plants were clearly differentiated, with a total of 483 genes significantly upregulated in N116 and 452 in PS01 (log2 fold change >2.0, FDR <0.05; Supplementary Material 3; Fig. 3C). Similarly, on DZA host plants, 395 and 363 genes were upregulated in N116 and PS01, respectively. In contrast, expression of relatively few genes, between three and 27, across all the pairwise comparisons, was significantly affected by the host plant (Supplementary Material 3; Fig. 3C). Functions of the DE genes are considered below, in conjunction with the other transcriptomic and proteomic experiments.

We next undertook a larger RNA-Seq experiment, sampling whole aphid bodies in order to capture transcripts from all tissues. Using aphids infested onto A17 host plants for 24 h, we again compared N116 and PS01 parental clones, but this time alongside the bulked segregant pools of VIR and AVR F1 clones described above. Five biological replicates for each gave a total of 20 RNA libraries each containing 14 to 22 million reads that uniquely map to the reference genome (Supplementary Material 2B).

Similar to the heads experiment, multivariate analysis by hierarchical clustering and PCA both indicated that all replicates within each sample type grouped together, and that each sample type was clearly differentiated. As expected, the genetically divergent parents were again highly separated, whereas the two pooled F1 datasets were much closer to each other, as they contain 50% of each parental genome, with each pool representing the average transcriptome of multiple independent F1 clones (Fig. 4A,B).

Differentially expressed genes were identified for all pairwise comparisons between samples (Fig. 4C). The number of up and down-regulated genes between the parental pairs and the pair of F1 pools are shown in Fig. 5A, with the gene lists provided in Supplementary Material 3. Several hundred genes were differentially expressed in both the whole-body and head comparisons of the parents. Some of these DE genes likely reflect genomic differences between the parental clones that are representatives of divergent pea aphid biotypes. However, relatively few DE genes were detected in the F1 samples, with only 24 genes up-regulated in the VIR pool and 64 in the AVR pool. These numbers can also be interpreted as a higher number of genes being down-regulated in the VIR F1 aphids. Fig. 5B,C show the overlaps across head and whole-body datasets for N116/VIR and PS01/AVR, respectively. Unexpectedly, the intersections of DE genes revealed subsets where the direction of expression was opposite between the parental pair and the F1 pooled pairs, with three genes upregulated in N116 and AVR F1, and 13 genes upregulated in PS01 and VIR F1 (Fig. 5D, Fig. 7G,H). Moreover, very few genes were upregulated in both parental N116 and VIR F1 pool datasets. A plausible explanation is that the genes governing virulence in N116 are not the same as those that result in virulent phenotypes in the F1 population. Each individual in the F1 population carries a random 50% of the genome of each parent, creating a high degree of combinatorial complexity. Nonetheless, the DE genes in the F1 data derive from the average across the 22 individuals used to create each bulk RNA pool, and are therefore likely to be biologically relevant to virulence or avirulence functions rather than background genomic noise. Such genes merit further exploration in both parental and F1 genotypes.

Quantitative proteomic analysis of saliva and salivary glands.

To determine whether differences exist between the salivary protein profiles of the two parental aphid clones, a comparative analysis of salivary gland and salivary proteomes was conducted. A total of 2343 and 2276 high confidence proteins were detected from salivary glands of N116 and PS01, respectively (Supplementary Material 4), with 2038 proteins (80%) common to both (Fig. 6A). Each biotype had similar proportions of non-annotated proteins (PS01: 5.4 % and N116: 6.2%) and proteins predicted to have secretion signals (PS01: 16.6% and N116: 17.3%). These proportions of secreted and non-annotated proteins are typical for pea aphid biotypes [12, 31]. Two major clusters were revealed

by PCA (Fig. 6C), corresponding to the two aphid genotypes. Principal Components 1 and 2 account for 64% of the variation, indicating distinct protein profiles in the salivary glands of each genotype. This distinction was further supported by quantitative analysis that identified 235 statistically significant differentially abundant (SSDA) proteins ($p<0.05$), with 136 and 99 proteins having higher abundances in N116 and PS01 salivary glands, respectively (Fig. 6E; Supplementary Material 4). Relative fold changes (RFC) ranged from −48.5 to +140.0 indicating that even when both genotypes engage in compatible interactions with the same plant type (*V. faba* in this case) the salivary gland profiles are divergent both qualitatively and quantitatively.

Of the 136 SSDA salivary gland proteins with increased abundance in N116, 60 (44%) were predicted to be secreted and 27 (20%) had no annotations. Similar proportions were observed within the 99 SSDA proteins with increased abundance in PS01, with 33 (33%) and 18 (18%) proteins having a secretion signal or no annotations, respectively. These proportions of secreted and non-annotated proteins within the differentially abundant sets are substantially higher than the corresponding proportions in the background salivary gland proteomes described above. Of the top ten proteins with the highest relative abundance in N116, seven had no annotation: ACPISUM_000319 (ACYPI007553; RFC 140.0) and ACPISUM_029783 (LOC100573424; RFC 64), ACPISUM_008675 (LOC100162547; RFC 32), ACPISUM_016335 (Not annotated; RFC 26), ACPISUM_017388 (LOC103309964; RFC 21.1), ACPISUM_003551 (LOC100534636; RFC, 21.1) and ACPISUM_009099 (LOC112598674, 18.4). The other proteins in the top ten were a kinase ACPISUM_015393 (developmentally-regulated protein kinase 1; RFC 64) and two aminopeptidases (ACPISUM_009259; RFC 36.8 and ACPISUM_005699; RFC 22.6). Of the top ten proteins with highest abundances in PS01 in comparison to N116, two were uncharacterised: ACPISUM_007394 (LOC100572241; RFC 48.5) and ACPISUM_007714 (LOC100534636; RFC 11.3); and two were glutathione S-transferases (ACPISUM_019160 and ACPISUM_001883, both RFCs of 8.6). Other proteins included a different developmentally-regulated protein kinase (ACPISUM_005630; RFC 17.1), a peroxidase (ACPISUM_020816; RFC 9.8), a prostatic spermine-binding protein (ACPISUM_004331; RFC 8), peroxidasin (ACPISUM_019870; RFC 6.5), an ATPase subunit (ACPISUM_009308; RFC 5.7) and glyoxylate reductase (ACPISUM_021751, RFC 4.9).

We next examined aphid saliva proteins. Although the samples are collected from artificial diets, these salivary secretomes are likely to be highly similar to the proteins delivered into plant tissues during interactions with the host, and therefore are predicted to include the entire set of effectors. We focussed on categorisation of the total salivary protein lists, and of the DE proteins. Although the analysis of saliva revealed far fewer proteins than from the salivary gland samples, there is again a clear distinction between the two genotypes. A total of 69 and 97 high confidence proteins were found in N116 and PS01 saliva, respectively (Fig. 6B; Supplementary Material 4) with 22 (32% for N116) and

50 (52% for PS01) proteins, being deemed unique to each. A large proportion (30% for PS01 and 25% for N116) of the salivary proteomes had no annotations, indicating their potential phylogenetic restriction to aphids. In addition, 39% and 32% of the proteins had predicted canonical secretion signals for PS01 and N116 saliva, respectively. Notably, although saliva proteins detected in diet samples have, by definition, been secreted, the majority appear not to have canonical secretion signals. Explanations range from incomplete/incorrect gene models to non-canonical or alternative secretion mechanisms. Our results highlight the importance of combining several approaches when attempting to identify potential effectors and molecular determinants of virulence/avirulence. Omitting proteins without secretion signals from bioinformatic pipelines may result in many effector candidates being overlooked.

As with the salivary glands, PCA of the salivary proteins completely resolved two groups, with PC1 and PC2 accounting for 94% of the total variation (Fig. 6D). Label free quantitative analysis using MaxQuant identified 47 SSDA proteins with 12 and 35 proteins having higher abundance in N116 and PS01 saliva, respectively (Fig. 6F; Supplementary Material 4). Notably, N116 saliva comprises fewer detected proteins and fewer SSDA proteins than PS01, possibly pointing to a strategy that enables evasion of host defences. If, for example, one or more of the proteins uniquely detected in PS01 saliva act as avirulence factors due to cognate receptors in the host plant, their absence or low abundance in N116 may result in a compatible interaction. However, it remains to be experimentally determined whether these genotypic differences in type or number of saliva proteins are causatively associated with virulence or avirulence.

Most of the salivary proteins identified here have previously been associated with pea aphid saliva including multiple members of M1 and M2 metalloprotease families, along with peroxidases, glutathione-S-transferases, glucose dehydrogenase and regucalcin [12, 32]. Apart from the Aminopeptidase N (APN) category discussed in detail below, the most frequent annotation was for unknown proteins: 20-26% of the total saliva list for each clone, and 21% of the DE saliva proteins. Four out of the ten DE unknown proteins also featured within the top 20 proteins by MS intensity or protein coverage. High proportions of unknown proteins have been noted in earlier studies of aphid saliva and the salivary gland predicted secretome [31]. In addition, a homologue of a salivary effector previously characterised for *Myzus persicae* (Mp1) [33] had a higher abundance in PS01 saliva (ACPISUM_000421; RFC 14). The relative fold changes of salivary proteins ranged from -2352 for regucalcin to 724 for members of the APN (M1 metalloprotease) family, which represented the most differentially abundant proteins in PS01 and N116 saliva, respectively. Although these RFC values can be considered arbitrary due to imputation of low abundant values in samples where the proteins are in fact absent, there is very clear divergence of salivary proteomes both in the proteins uniquely

detected in one or other genotype, and in the large differences in apparent abundance of several proteins present in both genotypes. The full lists of proteins exclusively found in the saliva or salivary gland proteomes of both genotypes are provided in Supplementary Material 4, with 25 and five proteins exclusive to the salivary glands and saliva of N116, respectively. For PS01, the corresponding numbers were 10 and 13 proteins exclusive to the salivary glands and saliva, respectively. These proteins were present in all replicates of one genotype while being absent in all replicates of the other, strongly supporting their status as candidate effectors, that may individually or collectively determine the VIR and AVR phenotypes observed for each genotype on different host plants.

Comparison of the quantitative differences in protein abundance across both the saliva and salivary gland datasets revealed clear similarities in the two proteomes analysed for each genotype. Five proteins that were of higher abundance in N116 saliva were also more abundant in N116 salivary glands in comparison to their PS01 counterparts. A similar trend was observed for nine PS01 salivary and salivary gland proteins (Supplementary Material 4), with the RFCs for these proteins positively correlated across both biological sample types. The fact that the abundances of these salivary gland proteins are mirrored at the level of externally delivered oral secretions highlights the robustness of both analyses, and points to likely roles as virulence or avirulence determinants in two genotypes with distinct host preferences. Such proteins represent excellent candidates for future characterisation to determine their effector status, especially those that are also supported by DE transcript profiles (Table 1).

Overlap between transcriptomics and proteomics datasets

Across the transcriptomics and proteomics experiments, we analysed all the intersections then extracted the proteins and DE gene subsets that showed the greatest overlaps (Table 1; Supplementary Material 3 and 4), partitioning into genes/proteins associated with virulence, in N116 or the VIR F1 pool, or with avirulence, in PS01 or the AVR F1 pool. The number of DE genes or proteins in the head transcriptome, whole body transcriptome and salivary gland proteome datasets were broadly similar between VIR and AVR samples. However, the PS01 saliva protein and the AVR F1 pool transcript lists were longer than those for N116 saliva and VIR F1 pool transcripts, reflected by larger intersections in the former. Over half (33/64) of genes upregulated in the AVR F1 pool were also in at least one other list, whereas only three out of 24 intersected from the VIR F1 pool data. Whole body RNA-Seq data for a selection of these intersected genes are plotted in Fig. 7. Several of the AVR-upregulated genes shown are annotated as enzymes with hydrolase, glycosidase or peroxidase functions. Other annotations include a transcription factor and proteins of unknown function. Genes on the VIR side included ACPISUM_013796 (myrosinase 1-like) and ACPISUM_019971 (glutathione hydrolase 1 proenzyme-like), although these were not found in saliva. Across the multiple

experiments, the two most frequently found genes in the AVR data were ACPISUM_021997 (regucalcin-like) previously reported as a Ca-binding protein [32], present in all lists except heads RNA, and ACPISUM_029930 (uncharacterized protein LOC100575698), present in all five lists. These AVR-related salivary proteins represent strong candidates for functional effectors, based on the multiple strands of evidence for their differential expression and importantly for co-segregation of their expression with the avirulence phenotype in the F1 population. We have therefore uncovered heritable differences in salivary proteins that associate with avirulence, in this case an incompatible phenotype on Mt hosts carrying the *RAP1* QTL [27, 28]. Intriguingly, however, we found no equivalent strong candidates for salivary proteins that might represent the dominant virulence factor predicted by previous genetic studies [27]. Alternative explanations for the Mendelian segregation found in that study could be that the proposed "virulence" gene is not an effector per se, but instead could be an upstream positive regulator, or a negative regulator of one or more effectors that act as avirulence factors detected by a *RAP1* dependent pathway.

Gene Ontology analysis

We undertook Gene Ontology (GO) analysis to reveal functional categories and genes that were enriched in the differentially expressed gene and protein data sets. Using a FDR of <0.05, many gene sets contained few or no significantly enriched terms (Table 2; Supplementary Material 5). For the whole-body transcriptome data, aminopeptidase N (APN) proteins were strongly enriched, with different genes within this family upregulated in each of the parental aphids (discussed further below). These trends were reinforced by comparison of parental transcriptomes in the heads RNA-Seq analyses where APN proteins were similarly enriched in both parents. The DE gene sets between the pooled VIR and AVR F1 samples indicated no enriched terms in the VIR data, and only a single term among the AVR upregulated genes: glucosidase II complex, that localises to the ER. These two gene sets are both relatively small (64 and 24 genes), reducing the likelihood of finding significant trends.

Because very few significantly enriched terms were revealed by the initial GO analyses, we applied a lower stringency to inform wider trends in each of the DE gene sets. Here, we examined all terms for which at least two genes and a significant P value (<0.05) were returned. For the DE gene sets from RNA-Seq of heads, the majority of enriched terms were associated with the virulent N116 parent on both host genotypes. Although there was obvious redundancy of many terms, a substantial proportion (30-40%) for N116 relate to energy metabolism including mitochondria, TCA cycle, oxidative phosphorylation and lipid metabolism. In contrast, the PS01 enriched terms included several for protein processing including peptidases, proteolysis and protein glycosylation; and several for ATP-related transport (Supplementary Material 5). When each parental aphid genotype was compared separately for its differential responses to the two host genotypes (A17 and DZA), no significant terms

were found for PS01, and only one weakly significant term for N116: polytene chromosome puffing. The equivalent GO analysis of whole body RNA-Seq data returned significantly enriched terms for both aphid genotypes, including several for protein modification (Supplementary Material 5).

For the DE datasets from salivary gland proteomes, the lower stringency analysis revealed enrichment of distinct functional categories for each parental genotype. For N116, protein modification terms were prevalent including peptidase activity, serine-type endopeptidase inhibitor activity, negative regulation of protein metabolic process, aminopeptidase activity, protein kinase binding and regulation of protein phosphorylation. In contrast, for PS01, ATPase terms were predominant including several related to membrane transport, as also found in the PS01 heads RNA-Seq data (Supplementary Material 5).

Exopeptidases are abundant in saliva, and the majority are DE between aphid genotypes

The saliva protein total and DE lists were much shorter, precluding formal GO analysis, but manual inspection indicated high proportions of exopeptidases: a total of 29 different proteins (Table 3), representing 22-34% of the protein list for each genotype. These were mainly APN proteins but also four members of the Angiotensin Converting Enzyme (ACE) family that are M2 metalloproteases with carboxypeptidase activity. The abundance of APNs in the saliva protein list broadly corroborates the major enriched GO categories detected in the transcriptome analyses.

Most of the exopeptidases detected from aphid saliva (23/29; 79%) were differentially abundant between the parental aphid genotypes. Twenty-two of the 29 saliva exopeptidases were also found in the salivary gland proteomes, with many showing the same direction of differential expression (9 APN, 2 ACE). Moreover, 15 (60%) of the APN proteins were DE in heads and/or whole body RNA-Seq samples (Table 3). Previous reports on pea aphid saliva and salivary gland components have also reported multiple APN and ACE proteins [12, 13, 32, 34]. Similar to our findings, one of these studies reported 11 APN genes that were differentially expressed in a biotype-specific manner, with five of these detected as proteins in saliva [13]. Taking all the evidence together, it is clear that the APN family is highly diversified in pea aphids and represents a major component of the salivary proteome by several measures: the high total number of proteins detected, many of these proteins are high abundance (13 of 20 top scoring in both N116 and PS01 saliva), and most are differentially expressed between aphid genotypes.

Aphid and mammalian ACE proteins have similar sequences and may have broadly similar functions as dipeptidases or by cleaving a single amino acid from the C terminus. However, mammalian ACE proteins are membrane anchored whereas aphid ACEs carry secretion signals, consistent with their detection in saliva. The exact catalytic functions and biological roles of aphid ACE and APN proteins remain to be determined. Cleavage of proteins and peptides could relate to targeting host proteins

such as those involved in defensive sieve-tube blocking as shown at least for the atypical extrafascicular phloem exudate of cucurbits [35]. Alternatively, although there is currently no direct evidence, exopeptidases may act on other salivary protein components, for example to process effectors into active forms. Another non-mutually exclusive possibility is a role in aphid nutrition, with many insects using extra-organismal (extra-oral) digestion typical of arthropods including Hemiptera, enabling nutrition capture from large hosts/prey [32, 36]. Exopeptidases typically release N or C terminal single amino acids and dipeptides, potentially enabling supply of essential amino acids, some of which cannot be biosynthesised directly from the enzyme repertoires of hemimetabolous aphids.

Multi-omic approaches to detecting candidate effectors

We compared the efficiencies of the four different experiments in terms of detecting aphid candidate effectors and related genes: RNA-Seq of heads and whole bodies, and proteomics of saliva and salivary glands. For all datasets, we focussed mainly on differential expression between the highly divergent parental clones N116 and PS01. Because saliva represents the "ground truth" of proteins predicted to be delivered into plant host tissues, we additionally considered saliva proteins that were detected but not DE. Although the proteomics methods are highly sensitive, there are likely to be some further low abundance salivary proteins that were not detected here. In addition, there may be some salivary proteins that are only expressed in response to aphids interacting with their host plants, and hence would not be found in artificial diet samples. Likewise, some proteins may not be stable under the artificial diet conditions. As a case study, we selected the significantly enriched exopeptidases, that comprised the large APN family and the smaller group of ACE proteins. We compared success of detecting genes from the saliva data in the other three experiments, and noted whether the same DE patterns were found (Table 3). The overall trends were broadly correlated, with 18/24 (75%) DE saliva proteins also found to be DE in at least one of the other approaches. Only two genes showed a mismatch in DE direction: ACPISUM_009259 between salivary gland and whole body; and ACPISUM_020790 between saliva and salivary gland. Individually, RNA-Seq of heads was the most effective experiment (14/24) at corroborating the DE saliva protein evidence, followed by RNA-Seq of whole bodies (10) and proteomics of salivary glands (8).

There are several reports where effectors are predicted from aphid salivary gland transcriptomes or proteomes, or other transcriptome datasets, typically filtering for presence of a signal peptide or other secretion motif, and absence of transmembrane domains [12–17]. For our exopeptidase data (Table 3), we detected an additional seven APNs in salivary gland proteomes or the transcriptome data, that were not found in saliva, of which five were DE in at least one dataset. Their absence from saliva indicates these proteins may be considered false positives for candidate effectors, although some low expressed proteins may go undetected. We considered which of the approaches was the most

effective at detecting candidate effectors, and whether multiple omics approaches are advantageous, noting that all require substantial resource investment. Although saliva collection is an exacting and time-intensive procedure, saliva proteomics provided the greatest coverage of candidate effectors here, and quantitative analysis of mass spectrometry data enables robust assignment of differential expression. Of the other approaches, RNA-Seq of heads may be the most effective means to complement the saliva analyses by reinforcing evidence of differential expression, but in the work here did not greatly extend the effector lists per se.

Conclusion

In this study, we demonstrated that transcriptomics and proteomics are both highly effective tools for discovering differentially expressed aphid genes and proteins. The protein subsets present in saliva are likely candidates for effectors with virulence and/or avirulence functions in host plants, and represent priorities for further study especially to determine if differential protein abundance is inherited into the segregating F1 aphid populations. Precise biochemical functions and host targets of most of these effectors are also currently unknown even in cases, such as the exopeptidases, where there are confident gene annotations. Exopeptidases are dominant in saliva by number of different proteins, by frequency of differential abundance, and by quantity. Likewise, there are many proteins of unknown function, with a substantial proportion found at high levels in saliva. Some of these unknown proteins may prove to be pivotal in explaining aphids' unique and highly successful lifestyle as phloem feeders.

Methods

Aphids and crossing

Pea aphid (*Acyrthosiphon pisum*) clones were maintained on tic bean (*Vicia faba minor*) as described in [26]. Parental genotypes were PS01 and N116. PS01 is a biotype adapted to *Pisum sativum* whereas N116 is adapted to *Medicago sativa* [26]. Reciprocal crosses were made between PS01 and N116 to generate F1 hybrid populations, following the protocol of [27]. In brief, parthenogenetic females were induced to generate sexual forms by transfer to short days and lower temperatures to simulate autumn. Eggs resulting from controlled matings were collected onto moist filter paper in petri dishes, and subjected to 90 to 105 days at 4°C to induce exit from diapause. Individual hatchlings were subsequently used to generate multiple parallel clonal F1 lineages. Parents and progeny were genotyped with a set of seven microsatellite markers [22] to verify correctness of crosses. All new F1 progeny were maintained for at least three generations before testing performance on different host plants.

Plants and assessment of virulence

Based on previous findings [27], PS01 aphids are avirulent on *Medicago truncatula* J A17 that carries the resistance QTL, *RAP1* [28]. Near isogenic lines (NILs) derived from a cross (LR4 [29]) between A17 and *M. truncatula* DZA315.16 were also used. PS01 is likewise incompatible with the resistant NIL (RNIL), but is compatible with the susceptible NIL (SNIL) and with DZA315.16. N116 aphids are compatible with all these genotypes. F1 progeny were tested for virulence on both A17 and RNIL, based on [26]. Briefly, five nymphs of each clone were infested onto ten A17 or RNIL plants, then scored for survival and production of new nymphs 10 d later. At least 40 F1 clones each of PS01 x N116 and N116 x PS01 were screened. An overall virulence index was adapted from a calculation proposed in [37]:

Virulence index = log2 (mean number surviving out of 5 x number of nymphs produced + 1)

Virulent (VIR) clones were defined as index >4 and >5 on A17 and RNIL, respectively, and avirulent (AVR) clones were correspondingly defined as index <2 and <4. The different category thresholds on A17 and RNIL reflect the latter's slightly lower resistance. Clones falling into the same phenotype category (VIR or AVR) on both A17 and RNIL were then subject to a further confirmation screen where survival on A17 and RNIL was counted 5 d after infestation. In the confirmation experiment, four plants were used for each aphid x host combination, with five aphids infested onto each plant. Cutoffs were >80% survival for virulence on both hosts, and <20% and <70% for avirulence on A17 and RNIL, respectively. A few F1 clones showed relatively high survival at 5 days but had very weak growth, and therefore were categorised as AVR. Only F1 clones displaying the same phenotype category on all screening experiments were used subsequently in molecular experiments.

Sampling for RNA-Seq

Heads experiment: Young adult aphids of clones N116 and PS01, cultured on *Vicia faba minor*, were infested onto either A17 or DZA315.16 *M. truncatula* plants for 24 h, then heads (40 per sample) were dissected and frozen immediately on dry ice then stored at -80°C. Three replicates were done for each aphid x plant combination.

Whole body experiment: Samples were parental aphid clones (N116 and PS01) and pools of VIR and AVR F1 progeny. Aphids of each individual genotype, age 2 to 3 d, were placed on independent A17 plants for 24 h then frozen in liquid nitrogen and stored at -80°C until processing. A total of 22 VIR and 22 AVR F1 aphid clones were collected individually, before pooling five aphids of each genotype to comprise one sample. Five biological replicates were analysed for both parental and pooled F1 genotypes.

RNA extraction, library preparation and sequencing

Heads were dissected and processed as described in [16]. Total RNA was extracted using a plant RNA extraction kit (Sigma-Aldrich). Illumina TruSeq stranded mRNA-Seq libraries were sequenced at the Genome Sequencing Unit at the University of Dundee on an Illumina HiSeq 2000.

RNA for the BSA-RNA-Seq analysis was isolated from three two to three day old nymphs of parental lines (N116, PS01), 22 VIR F1 lines and 22 AVR F1 lines, using the Norgen Plant and Fungal RNA kit (Sigma E4913). The RNA isolation followed the instructions of the company supplementing Lysis buffer C with ß-mercaptoethanol. An on-column DNase digest was performed (RNase-Free DNase Set, Qiagen) and the concentration of each sample determined via a Qubit fluorometer with the QubitTM RNA Broadrange (BR) assay kit (Thermo Fisher Scientific). Samples corresponding to five replicates of each of the parental lines and the VIR and AVR F1 pools were used to generate a total of 20 Illumina TruSeq stranded mRNA-Seq libraries which were sequenced in 150 bp paired-end mode on an Illumina HiSeq4000 at Edinburgh Genomics.

RNA-Seq data processing and visualisation.

Illumina RNA sequence reads were subjected to quality control using FastQC. The reads were the trimmed using Trimmomatic (version 0.32) Q15, min length 55. The trimmed fastq files were then quasi mapped to the nucleotide gene sequences for the pea aphid using salmon version 1.1. For the pilot study, STAR (2.4.1b) [38] was used to map the reads to the pea aphid genome and HTseq counts was used to quantify the gene expression using AphidBase_OGS2.1b gene annotations.

Clone-specific *de novo* RNA-Seq assemblies (from both the heads and whole-body studies) were individually and collectively generated using Trinity version 2.9.1. All the data were pooled into one for the “collective” assembly, which was used for transcript differential expression analysis. The individual assemblies were used for gene prediction at a later stage. All RNA-Seq assemblies were quality filtered using Transrate to reduce the probability of mis-assembled transcripts. Predicted coding sequences were generated using TransDecoder (with PFAM and BLAST guides). Diamond was used to search against GenbankNR database. Differential expression analysis was performed using EdgeR. Heatmaps and expression profile clustering were generated using the ptr script from within the Trinity package.

During early analysis, following visual assessment of RNA-seq read mapping and initial differential expression results, we found that the original pea aphid gene predictions (AphidBase_OGS2.1b) and the gene predictions from [39] did not fully match those generated by the de novo transcriptome assemblies. Therefore, gene annotation was re-predicted on the published pea aphid genome (OGS2.1b) to improve the accuracy of the gene models. Funannotate, in Other Eukaryotic mode, was used to predict the genes using the *de novo* RNA-Seq assembly generated above, with RNA-Seq data

mapped using STAR (see above). A total of 29,930 genes were assigned codes in the format ACPISUM_0xxxxx, with the annotations provided at doi.org/10.5281/zenodo.11103500 [40].

To assign the various gene calls from the original genome assembly, bedtools intercept was used to identify genes with overlapping coordinates. If the genes overlapped, then they were considered the same gene. A simple BLAST approach could not be used here due to the duplicated nature of aphid assemblies. A combination of reciprocal best BLAST hit, Orthofinder and MCL clustering were used to assign genes between the clones as orthologues.

Saliva Collection

For proteomics samples, N116 and PS01 were maintained separately on *Vicia faba* c.v. The Sutton, grown in standard potting compost and kept at 20°C and a photoperiod of 16-h light/8-h dark. Approximately 3,000 mixed aged aphids were positioned on 30 perspex rings (radius 4.5 cm, height 5 cm), each containing 4.5 ml of a chemically-defined diet, formulation A from [41], held between two stretched sheets of Parafilm™. The aphids were reared on the diets at 20°C with 18h light and 6h dark for 24 h after which the diets were pooled and collected and stored at -80°C until required. Four independent replicates were produced by pooling the collected diet from two daily collections (approximately 150 ml). Pooled diets were concentrated using a Vivacell 250 Pressure Concentrator (Sartorius Mechatronics, UK) using a 5000 Da molecular weight cut-off (MWCO) polyethersulfone (PES) membrane. When the final volume had reached 5 ml it was removed and 1 ml of filtered sterilised PBS (phosphate-buffered saline) supplemented with Roche cOmplete™ protease inhibitor cocktail (PIC) was added. The resulting mixture was further concentrated to approximately 250 µl using a Vivaspin 6 centrifuge concentrator (Sartorius Mechatronics, UK) with a 5000 Da MWCO PES membrane, purified using a 2D Clean-up Kit (GE HealthCare) following the manufacturer's instructions. The resulting protein pellet was suspended in 25 µl 6 M urea, 2 M thiourea, 0.1 M Tris-HCl, pH 8.0 and re-quantified using the Qubit Fluorometer. Four independent biological replicates per genotype were subjected to mass spectrometry.

Salivary glands

The salivary glands from 14-16 day old adult aphids of N116 and PS01 were dissected in ice-cold PBS and transferred to 60 µl PBS supplemented with PIC. Forty pairs of salivary glands were pooled per replicate and homogenized with a motorised, disposable pestle. Sixty microliters of 12 M urea, 4 M thiourea, and PIC was added and the samples were homogenised further and centrifuged at 9,000 × *g* for 5 min to pellet cellular debris. The supernatant was removed and quantified, and 100 µg of protein was purified using a 2D Clean-up Kit (GE HealthCare) following the manufacturer's instructions with the exception that 400 µl of precipitant and co-precipitant were used in the first step. The resulting

protein pellet was re-suspended in 30 µl 6 M urea, 2 M thiourea, 0.1 M Tris-HCl, pH 8.0 and re-quantified using the Qubit Fluorometer. Four biological replicates per genotype were subjected to mass spectrometry.

Protein sample digestion for mass spectrometry

The digestion protocol was the same for both saliva and salivary gland samples and involved the addition of 50 µl ammonium bicarbonate, reduction with 0.5 M dithiothreitol at 56°C for 20 min and alkylation with 0.55 M iodoacetamide at room temperature for 15 min, in the dark. One µl of a 1% w/v solution of ProteaseMax Surfactant Trypsin Enhancer (Promega) and 1 µg of Sequence Grade Trypsin (Promega) were added, then samples were incubated at 37°C for 18 h. Digestion was terminated by adding 1 µl of 100% trichloroacetic acid (Sigma Aldrich) and incubating at room temperature for 5 min. Samples were centrifuged for 10 min at 13,000 *x g* and the supernatant was removed to new microcentrifuge tubes.

Mass spectrometry and proteomic data analysis

One µg of digested peptide was loaded onto a Dionex Ultimate 3000 (RSLCnano) chromatography system connected to a QExactive (ThermoFisher Scientific) high-resolution accurate mass spectrometer. Peptides were separated by an increasing acetonitrile gradient on a Biobasic C18 PicofritTM column (100 mm length, 75 µm ID), using 120 and 50 min reverse phase gradients for salivary glands and saliva, respectively, at a flow rate of 250 nl min^{-1}. All data were acquired with the mass spectrometer operating in automatic data dependent switching mode. A high-resolution MS scan (300-2000 Da) was performed using the Orbitrap to select the 15 most intense ions prior to MS/MS.

Protein identification and normalisation was conducted using the Andromeda search engine in MaxQuant (version 1.6.17.0; http://maxquant.org/) to correlate the data against the predicted protein set generated in this study (ACPISUM_Proteins; 30891 entries) using default search parameters for Orbitrap data. False Discovery Rates were set to 1% for both peptides and proteins and the FDR was estimated following searches against a target-decoy database. Two searches were conducted for both N116 and PS01 saliva and salivary glands. The first involved a combined search of the raw files for each genotype separately to generate comprehensive proteomes for the saliva or salivary gland (hereafter All Identified Proteins). The second involved a quantitative search of the raw files for all biological replicates (n=4) for the saliva or salivary glands. Quantitative and statistical analyses were conducted in Perseus (Version 1.6.1.1 http://maxquant.org/) using the normalized label-free quantitation (LFQ) intensity values from each sample. The data were filtered to remove contaminants, and peptides identified by site. LFQ intensity values were log_2 transformed and samples

were allocated to their corresponding groups. A data imputation step was conducted to replace missing values with values that simulate signals of low abundant proteins chosen randomly from a distribution specified by a downshift of 2.1 times the mean standard deviation (SD) of all measured values and a width of 0.1 times this SD. Normalized intensity values were used for principal components analysis. A two-sample t-test was performed using a cut-off value of $p \leq 0.05$ to identify statistically significant differentially abundant (SSDA) proteins. Volcano plots were produced by plotting –Log p-values on the y-axis and Log_2 fold-change values on the x-axis to visualize differences in protein abundance between the two genotypes.

Gene annotations and Gene Ontology analysis

Secretion signal properties were predicted using SignalP4.1 [42]. Non-annotated genes were defined as those with the following descriptors: hypothetical protein, uncharacterized protein, NA or ACYPIxxxxx without any other assigned function. GO enrichment analyses were performed using GOseq [43].

Data availability

Genome annotations: zenodo.org/records/11103500 [40]

RNA-Seq: Pea aphid clones N116 and PS01 reared on *Medicago truncatula* A17 and DZA315.16, dissected heads: BioProject PRJNA757589, ncbi.nlm.nih.gov/bioproject/PRJNA757589/

RNA-Seq: Pea aphid clones N116, PS01 and bulk F1 hybrid progeny reared on *Medicago truncatula* A17, whole body samples: BioProject PRJNA757896, ncbi.nlm.nih.gov/bioproject/PRJNA757896

Scripts: github.com/peterthorpe5/Pea_aphid_on_medicago_DZA_A17

Proteomics: mass spectrometry data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository [44], dataset identifiers PXD053355 and PXD053620.

References

1. Rossi M, Goggin FL, Milligan SB, Kaloshian I, Ullman DE, Williamson VM. The nematode resistance gene *Mi* of tomato confers resistance against the potato aphid. Proc Natl Acad Sci. 1998;95:9750–4.

2. Dogimont C, Chovelon V, Pauquet J, Boualem A, Bendahmane A. The *V at* locus encodes for a CC - NBS - LRR protein that confers resistance to *A phis gossypii* infestation and *A . gossypii* -mediated virus resistance. Plant J. 2014;80:993–1004.

3. Du B, Zhang W, Liu B, Hu J, Wei Z, Shi Z, et al. Identification and characterization of *Bph14* , a gene conferring resistance to brown planthopper in rice. Proc Natl Acad Sci. 2009;106:22163–8.

4. Rodriguez PA, Bos JIB. Toward Understanding the Role of Aphid Effectors in Plant Infestation. Mol Plant-Microbe Interactions®. 2013;26:25–30.

5. Elzinga DA, De Vos M, Jander G. Suppression of Plant Defenses by a *Myzus persicae* (Green Peach Aphid) Salivary Effector Protein. Mol Plant-Microbe Interactions®. 2014;27:747–56.

6. Wang W, Dai H, Zhang Y, Chandrasekar R, Luo L, Hiromasa Y, et al. Armet is an effector protein mediating aphid-plant interactions. FASEB J. 2015;29:2032–45.

7. Mugford ST, Barclay E, Drurey C, Findlay KC, Hogenhout SA. An Immuno-Suppressive Aphid Saliva Protein Is Delivered into the Cytosol of Plant Mesophyll Cells During Feeding. Mol Plant-Microbe Interactions®. 2016;29:854–61.

8. Guo J, Wang H, Guan W, Guo Q, Wang J, Yang J, et al. A tripartite rheostat controls self-regulated host plant resistance to insects. Nature. 2023;618:799–807.

9. The International Aphid Genomics Consortium. Genome Sequence of the Pea Aphid Acyrthosiphon pisum. PLoS Biol. 2010;8:e1000313.

10. Li Y, Park H, Smith TE, Moran NA. Gene Family Evolution in the Pea Aphid Based on Chromosome-Level Genome Assembly. Mol Biol Evol. 2019;36:2143–56.

11. Shigenobu S, Yorimoto S. Aphid hologenomics: current status and future challenges. Curr Opin Insect Sci. 2022;50:100882.

12. Boulain H, Legeai F, Guy E, Morlière S, Douglas NE, Oh J, et al. Fast Evolution and Lineage-Specific Gene Family Expansions of Aphid Salivary Effectors Driven by Interactions with Host-Plants. Genome Biol Evol. 2018;10:1554–72.

13. Boulain H, Legeai F, Jaquiéry J, Guy E, Morlière S, Simon J-C, et al. Differential Expression of Candidate Salivary Effector Genes in Pea Aphid Biotypes With Distinct Host Plant Specificity. Front Plant Sci. 2019;10:1301.

14. Wang D, Yang Q, Hu X, Liu B, Wang Y. A Method for Identification of Biotype-Specific Salivary Effector Candidates of Aphid. Insects. 2023;14:760.

15. Bos JIB, Prince D, Pitino M, Maffei ME, Win J, Hogenhout SA. A Functional Genomics Approach Identifies Candidate Effectors from the Aphid Species Myzus persicae (Green Peach Aphid). PLOS Genet. 2010;6:e1001216.

16. Thorpe P, Cock PJA, Bos J. Comparative transcriptomics and proteomics of three different aphid species identifies core and diverse effector sets. BMC Genomics. 2016;17:172.

17. Nicolis VF, Burger NFV, Botha A-M. Whole-body transcriptome mining for candidate effectors from Diuraphis noxia. BMC Genomics. 2022;23:493.

18. Rotenberg D, Baumann AA, Ben-Mahmoud S, Christiaens O, Dermauw W, Ioannidis P, et al. Genome-enabled insights into the biology of thrips as crop pests. BMC Biol. 2020;18:142.

19. Grbić M, Van Leeuwen T, Clark RM, Rombauts S, Rouzé P, Grbić V, et al. The genome of Tetranychus urticae reveals herbivorous pest adaptations. Nature. 2011;479:487–92.

20. Huo S-M, Yan Z-C, Zhang F, Chen L, Sun J-T, Hoffmann AA, et al. Comparative genome and transcriptome analyses reveal innate differences in response to host plants by two color forms of the two-spotted spider mite Tetranychus urticae. BMC Genomics. 2021;22:569.

21. Mathers TC, Chen Y, Kaithakottil G, Legeai F, Mugford ST, Baa-Puyoulet P, et al. Rapid transcriptional plasticity of duplicated gene clusters enables a clonally reproducing aphid to colonise diverse plant species. Genome Biol. 2017;18:27.

22. Peccoud J, Ollivier A, Plantegenest M, Simon J-C. A continuum of genetic divergence from sympatric host races to species in the pea aphid complex. Proc Natl Acad Sci. 2009;106:7495–500.

23. Caillaud MC, Mondor-Genson G, Levine-Wilkinson S, Mieuzet L, Frantz A, Simon JC, et al. Microsatellite DNA markers for the pea aphid Acyrthosiphon pisum. Mol Ecol Notes. 2004;4:446–8.

24. Eyres I, Jaquiéry J, Sugio A, Duvaux L, Gharbi K, Zhou J, et al. Differential gene expression according to race and host plant in the pea aphid. Mol Ecol. 2016;25:4197–215.

25. Eyres I, Duvaux L, Gharbi K, Tucker R, Hopkins D, Simon J, et al. Targeted re-sequencing confirms the importance of chemosensory genes in aphid host race differentiation. Mol Ecol. 2017;26:43–58.

26. Kanvil S, Powell G, Turnbull C. Pea aphid biotype performance on diverse *Medicago* host genotypes indicates highly specific virulence and resistance functions. Bull Entomol Res. 2014;104:689–701.

27. Kanvil S, Collins CM, Powell G, Turnbull CGN. Cryptic Virulence and Avirulence Alleles Revealed by Controlled Sexual Recombination in Pea Aphids. Genetics. 2015;199:581–93.

28. Stewart SA, Hodge S, Ismail N, Mansfield JW, Feys BJ, Prospéri J-M, et al. The *RAP1* Gene Confers Effective, Race-Specific Resistance to the Pea Aphid in *Medicago truncatula* Independent of the Hypersensitive Reaction. Mol Plant-Microbe Interactions®. 2009;22:1645–55.

29. Julier B, Huguet T, Chardon F, Ayadi R, Pierre J-B, Prosperi J-M, et al. Identification of quantitative trait loci influencing aerial morphogenesis in the model legume Medicago truncatula. Theor Appl Genet. 2007;114:1391–406.

30. Peccoud J, Mahéo F, de la Huerta M, Laurence C, Simon J-C. Genetic characterisation of new host-specialised biotypes and novel associations with bacterial symbionts in the pea aphid complex. Insect Conserv Divers. 2015;8:484–92.

664 31. Carolan JC, Caragea D, Reardon KT, Mutti NS, Dittmer N, Pappan K, et al. Predicted Effector
665 Molecules in the Salivary Secretome of the Pea Aphid (*Acyrthosiphon pisum*): A Dual
666 Transcriptomic/Proteomic Approach. J Proteome Res. 2011;10:1505–18.

667 32. Carolan JC, Fitzroy CIJ, Ashton PD, Douglas AE, Wilkinson TL. The secreted salivary proteome of
668 the pea aphid ***Acyrthosiphon pisum*** characterised by mass spectrometry. PROTEOMICS.
669 2009;9:2457–67.

670 33. Pitino M, Hogenhout SA. Aphid Protein Effectors Promote Aphid Colonization in a Plant Species-
671 Specific Manner. Mol Plant-Microbe Interactions®. 2013;26:130–9.

672 34. Kaloshian I, Walling LL. Hemipterans as Plant Pathogens. Annu Rev Phytopathol. 2005;43:491–
673 521.

674 35. Furch ACU, Van Bel AJE, Will T. Aphid salivary proteases are capable of degrading sieve-tube
675 proteins. J Exp Bot. 2015;66:533–9.

676 36. Cantón PE, Bonning BC. Extraoral digestion: outsourcing the role of the hemipteran midgut. Curr
677 Opin Insect Sci. 2020;41:86–91.

678 37. Ferrari J, Via S, Godfray HCJ. POPULATION DIFFERENTIATION AND GENETIC VARIATION IN
679 PERFORMANCE ON EIGHT HOSTS IN THE PEA APHID COMPLEX. Evolution. 2008;62:2508–24.

680 38. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-
681 seq aligner. Bioinformatics. 2013;29:15–21.

682 39. Thorpe P, Escudero-Martinez CM, Cock PJA, Eves-van Den Akker S, Bos JIB. Shared
683 Transcriptional Control and Disparate Gain and Loss of Aphid Parasitism Genes. Genome Biol Evol.
684 2018;10:2716–33.

685 40. Thorpe P. Genome annotations for: Multi-omics approaches define novel aphid effector
686 candidates associated with virulence and avirulence phenotypes. 2024.

687 41. Prosser WA, Douglas AE. The aposymbiotic aphid: An analysis of chlortetracycline-treated pea
688 aphid, Acyrthosiphon pisum. J Insect Physiol. 1991;37:713–9.

689 42. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from
690 transmembrane regions. Nat Methods. 2011;8:785–6.

691 43. Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: accounting
692 for selection bias. Genome Biol. 2010;11:R14.

693 44. Perez-Riverol Y, Bai J, Bandla C, García-Seisdedos D, Hewapathirana S, Kamatchinathan S, et al.
694 The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences.
695 Nucleic Acids Res. 2022;50:D543–52.

696

Tables

Table 1 Genes and proteins overlapping in multiple experiments. All genes shown that are represented in at least three datasets, plus all genes intersected between F1 transcriptome and at least one other dataset. Saliva and salivary gland data are proteins, head and body data are transcripts. A. Proteins and upregulated genes in virulent aphids (N116, VIR F1 pool); B. Proteins and upregulated genes in avirulent aphids (PS01, AVR F1 pool). Y = protein present and/or RNA differentially expressed. Full gene and protein lists are in Supplementary Material 3 and 4.

Gene	Annotation	Saliva	Salivary gland	Parent head	Parent body	F1 body
A) N116 & VIR F1						
ACPISUM_000319	ACYPI007553	Y	Y		Y	
ACPISUM_006458	aldo-keto reductase family 1 member B10-like		Y	Y	Y	
ACPISUM_025240	aminopeptidase N	Y	Y	Y	Y	
ACPISUM_005699	aminopeptidase N	Y	Y	Y	Y	
ACPISUM_025168	aminopeptidase N	Y	Y	Y	Y	
ACPISUM_009258	aminopeptidase N	Y		Y	Y	
ACPISUM_024778	aminopeptidase N	Y		Y	Y	
ACPISUM_026844	aminopeptidase N	Y		Y	Y	
ACPISUM_025015	aminopeptidase N	Y	Y	Y	Y	
ACPISUM_023906	Apoptosis inducing protein		Y	Y	Y	
ACPISUM_020864	F-actin-capping protein subunit alpha		Y	Y	Y	
ACPISUM_023535	glutamate-gated chloride channel-like			Y		Y
ACPISUM_019971	glutathione hydrolase 1 proenzyme-like		Y	Y	Y	
ACPISUM_010531	hypothetical protein X975_16721				Y	Y
ACPISUM_013751	LYR motif-containing protein 4		Y	Y	Y	
ACPISUM_013796	myrosinase 1-like				Y	Y
ACPISUM_006164	---NA---		Y	Y	Y	
ACPISUM_023321	papain inhibitor-like		Y	Y	Y	
ACPISUM_009624	proline-rich extensin-like protein EPR1		Y	Y	Y	
ACPISUM_028519	single-stranded DNA-binding replication protein A		Y	Y	Y	
ACPISUM_025560	ubiquinone biosynthesis monooxygenase COQ6, mitochondrial		Y	Y	Y	
ACPISUM_008675	uncharacterized protein LOC100162547		Y	Y	Y	
ACPISUM_007320	uncharacterized protein LOC100167449	Y	Y	Y	Y	
ACPISUM_001031	uncharacterized protein LOC100571631		Y	Y	Y	
ACPISUM_016519	uncharacterized protein LOC100573156		Y	Y	Y	
ACPISUM_010687	uncharacterized protein LOC103309122		Y	Y	Y	
ACPISUM_017388	uncharacterized protein LOC103309964	Y	Y	Y	Y	
ACPISUM_009099	uncharacterized protein LOC112598674		Y	Y	Y	
ACPISUM_027918	vacuolar protein sorting-associated protein 29		Y	Y	Y	
B) PS01 & AVR F1						
ACPISUM_000957	AGAP002382-PA-like protein			Y	Y	Y
ACPISUM_015173	AGAP011571-PA-like protein			Y		Y
ACPISUM_002223	aminopeptidase N		Y	Y	Y	
ACPISUM_003737	aminopeptidase N	Y	Y	Y	Y	
ACPISUM_023448	aminopeptidase N	Y	Y	Y	Y	
ACPISUM_028967	aminopeptidase N	Y	Y	Y	Y	
ACPISUM_021545	aminopeptidase N	Y	Y	Y		
ACPISUM_009259	aminopeptidase N		Y	Y	Y	
ACPISUM_009580	anoctamin-1-like			Y	Y	Y
ACPISUM_012705	CD63 antigen			Y	Y	Y
ACPISUM_006933	cuticular protein			Y	Y	Y
ACPISUM_019160	glutathione S-transferase 1-1-like	Y	Y	Y	Y	
ACPISUM_019168	glutathione S-transferase 1-1-like	Y	Y	Y	Y	
ACPISUM_001883	glutathione S-transferase D7-like	Y	Y	Y		
ACPISUM_016389	histone acetyltransferase KAT6B isoform X1				Y	Y
ACPISUM_009097	multidrug resistance-associated protein 1		Y	Y	Y	
ACPISUM_011553	---NA---		Y	Y	Y	
ACPISUM_011754	---NA---	Y	Y	Y	Y	
ACPISUM_004702	---NA---				Y	Y
ACPISUM_021569	---NA---				Y	Y

ACPISUM_025236	---NA---				Y	Y
ACPISUM_014327	---NA---			Y	Y	Y
ACPISUM_016390	---NA---			Y	Y	Y
ACPISUM_017200	---NA---			Y	Y	Y
ACPISUM_027631	---NA---			Y	Y	Y
ACPISUM_028853	---NA---			Y	Y	Y
ACPISUM_019381	neural cell adhesion molecule L1 isoform X1			Y	Y	Y
ACPISUM_020816	peroxidase-like	Y	Y	Y	Y	
ACPISUM_019857	peroxidase-like	Y	Y	Y		
ACPISUM_019870	peroxidasin homolog		Y	Y	Y	
ACPISUM_000958	phospholipase DDHD2-like			Y		Y
ACPISUM_006758	piggyBac transposable element-derived protein 4-like			Y	Y	Y
ACPISUM_022113	piwi-like protein Siwi				Y	Y
ACPISUM_010778	predicted protein			Y	Y	Y
ACPISUM_019013	protein ABHD18			Y	Y	Y
ACPISUM_021997	regucalcin-like	Y	Y		Y	Y
ACPISUM_021999	regucalcin-like	Y	Y	Y		
ACPISUM_001383	replication protein A 70 kDa DNA-binding subunit-like		Y	Y	Y	
ACPISUM_015166	TBC1 domain family member 19			Y	Y	Y
ACPISUM_014232	tubulin glycylase 3A-like				Y	Y
ACPISUM_008377	uncharacterized family 31 glucosidase KIAA1161-like			Y	Y	Y
ACPISUM_008379	uncharacterized family 31 glucosidase KIAA1161-like			Y	Y	Y
ACPISUM_008380	uncharacterized family 31 glucosidase KIAA1161-like			Y	Y	Y
ACPISUM_012348	uncharacterized protein LOC100158692	Y	Y	Y		
ACPISUM_018433	uncharacterized protein LOC100158721				Y	Y
ACPISUM_007487	uncharacterized protein LOC100160601	Y	Y	Y		
ACPISUM_016065	uncharacterized protein LOC100161530		Y	Y	Y	
ACPISUM_007076	uncharacterized protein LOC100163035			Y	Y	Y
ACPISUM_016064	uncharacterized protein LOC100570074		Y	Y	Y	
ACPISUM_029311	uncharacterized protein LOC100570454	Y	Y	Y	Y	
ACPISUM_008664	uncharacterized protein LOC100570454	Y	Y	Y		
ACPISUM_007394	uncharacterized protein LOC100572241	Y	Y	Y	Y	
ACPISUM_021703	uncharacterized protein LOC100575642	Y		Y	Y	
ACPISUM_029930	uncharacterized protein LOC100575698	Y	Y	Y	Y	Y
ACPISUM_006906	uncharacterized protein LOC100575848			Y	Y	Y
ACPISUM_003989	uncharacterized protein LOC103307823		Y	Y	Y	
ACPISUM_024374	uncharacterized protein LOC107882950				Y	Y
ACPISUM_015285	uncharacterized protein LOC107883982			Y	Y	Y
ACPISUM_000491	uncharacterized protein LOC111028731		Y	Y	Y	
ACPISUM_027814	uncharacterized SDCCAG3 family protein-like			Y	Y	Y

704

705

706 **Table 2 Significantly enriched GO terms within differentially expressed transcript and**
707 **protein data.** Terms enriched at FDR<0.05, after manual curation to remove redundancies,
708 retaining the terms with lowest FDR. Full lists of enriched terms are in Supplementary
709 Material 5.

GO Category	Term	Ontology group	No. genes in DE set	P value	FDR
Whole body RNA					
N116 up					
0004177	aminopeptidase activity	MF	8	2.72E-06	0.0270
PS01 up					
0004177	aminopeptidase activity	MF	9	1.43E-07	0.0014
0017177	glucosidase II complex	CC	4	2.09E-06	0.0104
AVR F1 up					
0017177	glucosidase II complex	CC	3	2.42E-06	0.0240
VIR F1 up	*No terms*				
Heads RNA					
N116 up on A17					
0045271	respiratory chain complex I	CC	11	1.55E-07	9.63E-05
0005743	mitochondrial inner membrane	CC	22	2.01E-07	0.0001
0004177	aminopeptidase activity	MF	10	2.13E-06	0.0009
0016491	oxidoreductase activity	MF	31	3.24E-06	0.0012
0005875	microtubule associated complex	CC	25	1.72E-05	0.0052
0019395	fatty acid oxidation	BP	5	2.44E-05	0.0064
0042826	histone deacetylase binding	MF	4	0.00011	0.0275
0045239	tricarboxylic acid cycle enzyme complex	CC	3	0.00014	0.0327
0004448	isocitrate dehydrogenase activity	MF	3	0.00015	0.0338
N116 up on DZA					
0006635	fatty acid beta-oxidation	BP	5	1.05E-06	0.0082
0004177	aminopeptidase activity	MF	9	2.66E-06	0.0082
0004449	isocitrate dehydrogenase (NAD+) activity	MF	3	1.40E-05	0.0198
0006099	tricarboxylic acid cycle	BP	6	3.54E-05	0.0389
PS01 up on A17					
0004177	aminopeptidase activity	MF	10	7.77E-07	0.00771
PS01 up on DZA					
0004177	aminopeptidase activity	MF	11	5.59E-09	5.53E-05
Salivary gland proteins					
N116 up					
0003983	UTP:glucose-1-phosphate uridylyltransferase activity	MF	3	4.97E-06	0.0213
PS01 up	*No terms*				

710

711

712 **Table 3. Comparison of expression patterns of exopeptidases detected in saliva and salivary**
713 **glands.** All detected proteins are listed, along with whether they were differentially expressed, and
714 whether the patterns were also reflected in the transcriptomes. Sal = saliva; SG = salivary gland; Y =
715 protein present.

ACPISUM v3 gene	Nearest ACYPI gene(s)	Presence/absence				Differential expression			
		Sal N116	Sal PS01	SG N116	SG PS01	Sal	SG	heads	whole
Aminopeptidase N									
ACPISUM_005699-T1	ACYPI080623 ACYPI070600 ACYPI005810	Y	Y	Y	Y				
ACPISUM_025168-T1	ACYPI068031	Y		Y					
ACPISUM_027632-T1	ACYPI073645	Y		Y					
ACPISUM_009259-T1	ACYPI007868	Y		Y	Y				
ACPISUM_009258-T1	ACYPI007868	Y	Y						
ACPISUM_024778-T1	ACYPI072916	Y							
ACPISUM_025015-T1	ACYPI061522 ACYPI21510	Y		Y					
ACPISUM_023174-T1	ACYPI006366	Y	Y						
ACPISUM_025240-T1	ACYPI070333	Y	Y	Y					
ACPISUM_000115-T1	ACYPI000001			Y					
ACPISUM_003737-T1	ACYPI067691	Y	Y		Y				
ACPISUM_021545-T1	ACYPI085147 ACYPI002583	Y	Y	Y	Y				
ACPISUM_029674-T1	ACYPI072988	Y	Y	Y	Y				
ACPISUM_023448-T1	ACYPI010198	Y	Y	Y	Y				
ACPISUM_028967-T1	ACYPI083965	Y	Y	Y	Y				
ACPISUM_000246-T1	ACYPI086097 ACYPI43770 ACYPI068046	Y	Y						
ACPISUM_010796-T1	ACYPI071232 ACYPI33244	Y	Y	Y	Y				
ACPISUM_012062-T1	ACYPI22813	Y	Y	Y	Y				
ACPISUM_018507-T1	ACYPI21711 ACYPI084528 ACYPI003165	Y	Y	Y	Y				
ACPISUM_006298-T1	ACYPI41708 ACYPI22605	Y	Y	Y	Y				
ACPISUM_019635-T1	ACYPI060722		Y						
ACPISUM_019937-T1	ACYPI49161	Y	Y	Y	Y				
ACPISUM_026119-T1	ACYPI083984	Y	Y	Y	Y				
ACPISUM_017858-T1	ACYPI54528 ACYPI001911			Y	Y				
ACPISUM_002219-T1	ACYPI44040	Y	Y	Y	Y				
ACPISUM_014203-T1	ACYPI067721			Y	Y				
ACPISUM_018506-T1	ACYPI21557	Y	Y						
ACPISUM_019609-T1	ACYPI001203	Y	Y						
ACPISUM_019610-T1	ACYPI071951			Y	Y				
	Aminopeptidase N total proteins detected	24	20	21	17				
Angiotensin converting enzyme									
ACPISUM_008374-T1	ACYPI000733	Y	Y	Y	Y				
ACPISUM_024301-T1	ACYPI084554	Y	Y	Y	Y				
ACPISUM_024303-T1	ACYPI071320	Y	Y	Y	Y				
ACPISUM_020790-T1	ACYPI008911		Y	Y					

Key
not detected
detected, not DE
up in N116
up in PS01

716

717

718 **Figure legends**

719 **Figure 1. Summary of transcriptome and proteome analysis pipeline.** Resistant and susceptible host
720 plants carried or lacked the *RAP1* aphid resistance QTL, respectively. For all experiments, virulent N116
721 and avirulent PS01 aphids were compared. In addition, BSA-RNA-Seq was done on whole body pooled
722 samples of F1 virulent and avirulent aphids.

723 **Figure 2. Virulence phenotypes of parental aphid clones and selections from F1 populations used**
724 **for BSA-RNA-Seq.** Tested on two *M. truncatula* genotypes carrying the *RAP1* locus: Jemalong A17 and
725 a resistant near isogenic line (RNIL) derived from a cross between A17 and DZA315.16. The parental
726 genotypes and selections from the F1 populations shown here were all used for the BSA-RNA-Seq
727 experiment. Data are expressed as virulence index, assessed 10 d after infestation. Phenotypes of F1
728 clones were classified using the following virulence index cut-offs: A17 VIR >4, AVR <2; RNIL VIR >4,
729 AVR <4.5. Orange circles are NP (N116 female x PS01 male); blue triangles are PN (PS01 female x N116
730 male); red is N116, and green is PS01, with each of three parental data points from a separate batch
731 of F1 tests. The full population phenotype data are provided in Supplementary Material 1.

732 **Figure 3. Transcriptome analysis of aphid heads.** Samples were dissected heads from PS01 and N116
733 genotypes infested on *Medicago truncatula* A17 or DZA315.16 for 24 h, with n=3 biological replicates.
734 Aphid genotype PS01 is avirulent on *M. truncatula* A17, and all other combinations represent
735 compatible interactions. A. Clustering of transcriptional responses of pea aphid, showing samples
736 clustered more strongly based on aphid genotype than on host interaction; B. Principal components
737 analysis. The top two principal components explain >68% of the variation among transcriptional
738 responses. Samples group largely by aphid genotype rather than host interaction; C. Numbers of genes
739 differentially expressed between the different aphid genotypes on different *M. truncatula* genotypes.
740 Of the 935 DE genes between PS01 and N116 on A17, 483 were up in N116 and 452 were up in PS01.
741 Of the 758 DE genes on DZA hosts, 395 were up in N116 and 363 were up in PS01. Accompanying gene
742 lists and annotations are provided in Supplementary Material 3.

743 **Figure 4. Transcriptome analysis of whole aphids.** Aphids were infested on *Medicago truncatula* A17
744 for 24 h, with n=5 biological replicates. A. Clustering of transcriptional responses of pea aphid PS01,
745 N116, bulked F1 VIR and AVR progeny. Responses within biological replicates are more strongly
746 correlated than responses among different aphid genotypes; B. Principal components analysis. The
747 top 2 principal components explain >45% of the variation among transcriptional responses of PS01,
748 N116, AVR and VIR F1 progeny replicates, and separate the responses of the different aphid genotypes
749 and F1 pools. C. Numbers of genes differentially expressed between the different aphid genotypes
750 and pools. Accompanying gene lists and annotations are provided in Supplementary Material 3.

751 **Figure 5. Differential gene expression in pea aphid genotypes N116, PS01 and bulked F1 pools of**
752 **virulent and avirulent progeny.** A**.** Numbers of genes up- versus down-regulated in comparisons of
753 parent genotypes N116 (virulent) and PS01 (avirulent), and their F1 progeny pools, all on A17 host
754 plants. Orange bars represent the numbers of genes up-regulated in genotype N116 or the VIR F1 pool
755 compared with the genotype PS01 and the AVR pool, respectively. Blue bars represent the numbers
756 of genes up-regulated in genotype PS01 or the AVR F1 pool compared with the genotype N116 and
757 the VIR pool, respectively; B. Overlaps in genes up-regulated in genotype N116 whole body and head
758 tissues compared to genotype PS01, and up-regulated in the VIR F1 pool compared to the AVR F1 pool;
759 C. Overlaps in genes up-regulated in genotype PS01 whole body and head tissues compared to
760 genotype N116, and up-regulated in the AVR F1 pool compared to the VIR F1 pool; D. Overlaps in up-
761 regulated genes among whole body transcriptomes of N116, PS01, VIR F1 pool and AVR F1 pool.

762 **Figure 6. Comparative proteomic analysis of salivary glands and saliva for pea aphid genotypes N116**
763 **and PS01.** Venn diagrams of the number of proteins shared and found exclusively in A) salivary glands
764 and B) saliva identified for both genotypes. Principal Components Analysis (PCA) of C) salivary glands
765 and D) saliva distinguishes both genotypes clearly. Volcano plots based on $-\log_{10}$ p values and $\log_2$ fold
766 differences highlighting the statistically significant differentially abundant (SSDA) proteins ($p \leq 0.05$) for
767 E) salivary glands and F) saliva. Annotations are shown for the top 12 proteins of increased and
768 decreased abundances.

769 **Figure 7. Selected differentially expressed genes from whole body transcriptomes**. A,B
770 representative genes upregulated both in virulent parent and in virulent F1 pool; C-F representative
771 genes upregulated both in avirulent parent and in avirulent F1 pool. G,H representative genes with
772 opposite regulation between parent and F1 pairs. Each point represents an individual RNA-Seq library
773 (n=5). *** indicates FDR<0.001.

774

Acknowledgements

Funding

We thank the Biotechnology and Biological Sciences Research Council for funding to CT (BB/N002830/1) and JB (BB/N002660/1). We thank Umer Rashid and Martin Selby for expert technical assistance.

Author contributions

JB, CT, JC, PT, SA and RLC designed the experiments. PT, SA, RLC, ND, JCS, JI and SK conducted the experiments. PT, SA, RLC, JB, CT, JC, ND and JCS analysed the data. CT, JB, JC and PT wrote the paper. All authors approved the submitted manuscript.

Conflicts

The authors declare that they have no competing interests.

Supplementary Materials

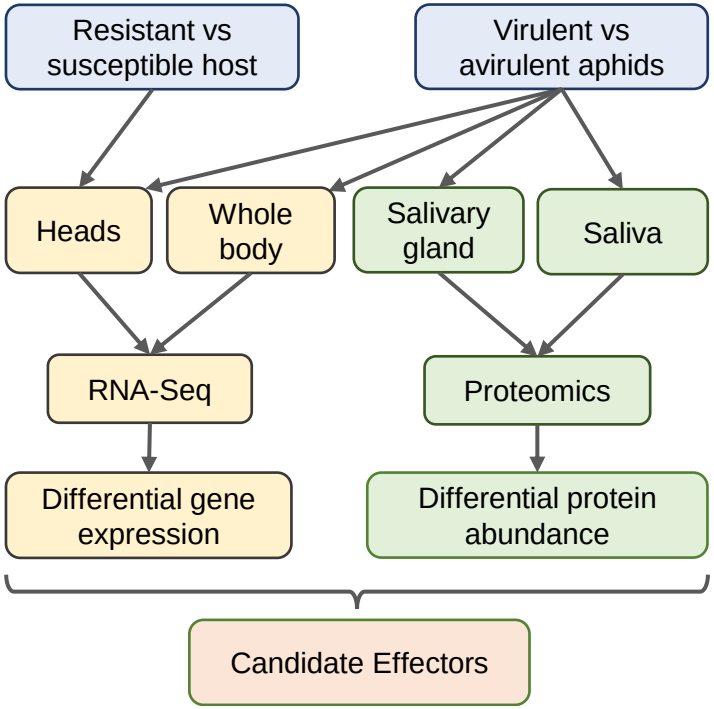
Supplementary Material 1. F1 aphid phenotyping.

Supplementary Material 2. Read mapping summary.

Supplementary Material 3. RNA-Seq Head and whole body differentially expressed genes.

Supplementary Material 4. Salivary gland and saliva proteomics.

Supplementary Material 5. GO enrichment

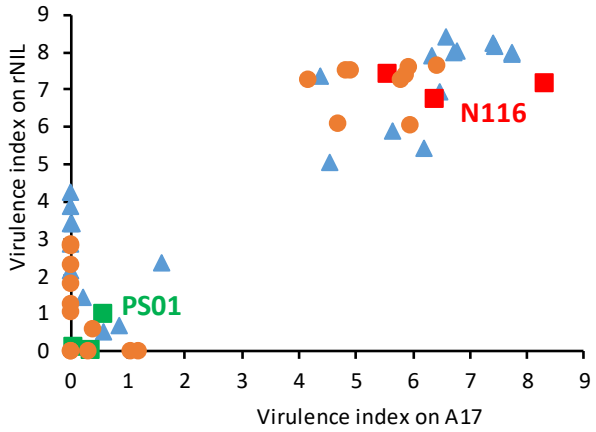


792

793

794 **Figure 1. Summary of transcriptome and proteome analysis pipeline.** Resistant and susceptible host plants carried or lacked the *RAP1* aphid resistance QTL,
795 respectively. For all experiments, virulent N116 and avirulent PS01 aphids were compared. In addition, BSA-RNA-Seq was done on whole body pooled samples
796 of F1 virulent and avirulent aphids.

797

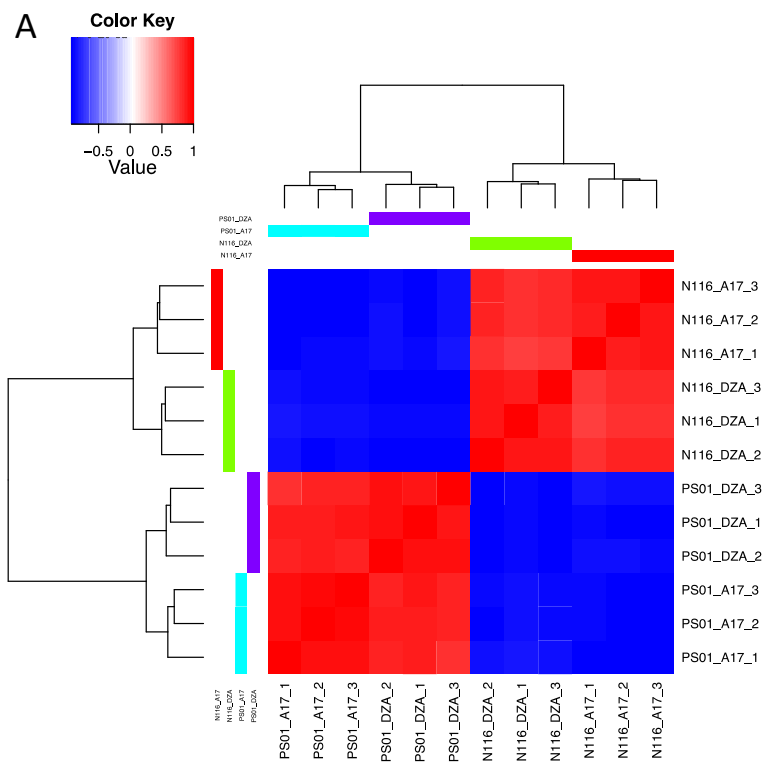


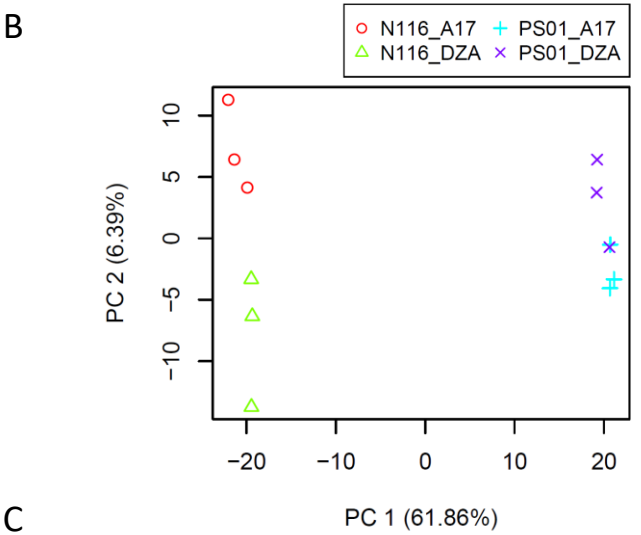
798

799

800 **Figure 2. Virulence phenotypes of parental aphid clones and selections from F1 populations used for BSA-RNA-Seq.** Tested on two *M. truncatula* genotypes
801 carrying the *RAP1* locus: Jemalong A17 and a resistant near isogenic line (RNIL) derived from a cross between A17 and DZA315.16. The parental genotypes
802 and selections from the F1 populations shown here were all used for the BSA-RNA-Seq experiment. Data are expressed as virulence index, assessed 10 d after
803 infestation. Phenotypes of F1 clones were classified using the following virulence index cut-offs: A17 VIR >4, AVR <2; RNIL VIR >4, AVR <4.5. Orange circles
804 are NP (N116 female x PS01 male); blue triangles are PN (PS01 female x N116 male); red is N116, and green is PS01, with each of three parental data points
805 from a separate batch of F1 tests. The full population phenotype data are provided in Supplementary Material 1.

806

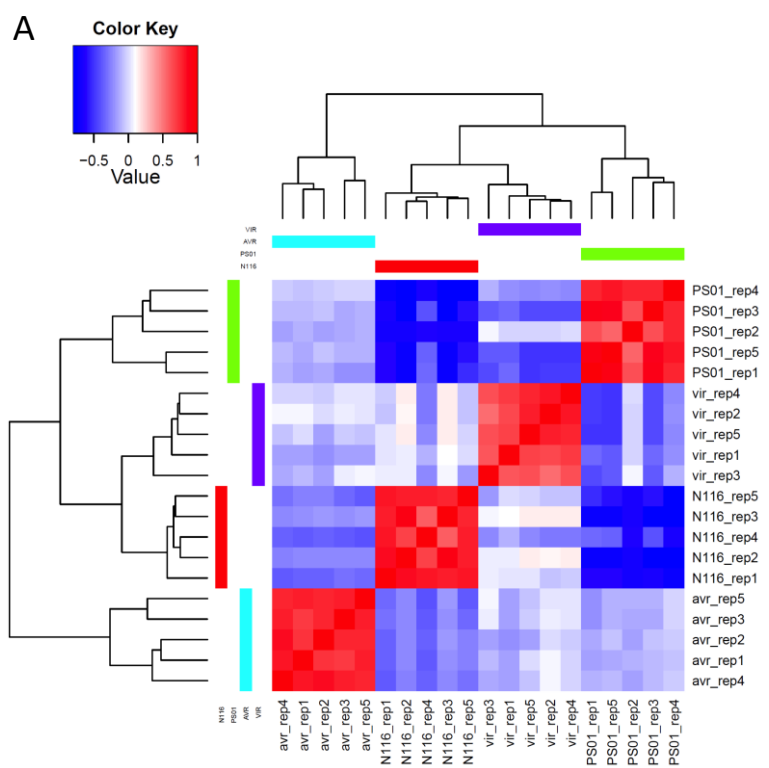


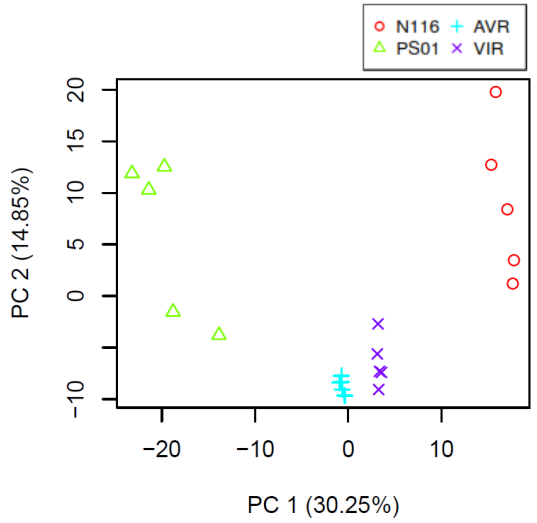


C

	N116/A17	N116/DZA	PS01/A17	PS01/DZA
N116/A17	0	33	935	845
N116/DZA	33	0	782	758
PS01/A17	935	782	0	29
PS01/DZA	845	758	29	0

Figure 3. Transcriptome analysis of aphid heads. Samples were dissected heads from PS01 and N116 genotypes infested on *Medicago truncatula* A17 or DZA315.16 for 24 h, with n=3 biological replicates. Aphid genotype PS01 is avirulent on *M. truncatula* A17, and all other combinations represent compatible interactions. A. Clustering of transcriptional responses of pea aphid, showing samples clustered more strongly based on aphid genotype than on host interaction; B. Principal components analysis. The top two principal components explain >68% of the variation among transcriptional responses. Samples group largely by aphid genotype rather than host interaction; C. Numbers of genes differentially expressed between the different aphid genotypes on different *M. truncatula* genotypes. Of the 935 DE genes between PS01 and N116 on A17, 483 were up in N116 and 452 were up in PS01. Of the 758 DE genes on DZA hosts, 395 were up in N116 and 363 were up in PS01. Accompanying gene lists and annotations are provided in Supplementary Material 3.





C

	N116	Vir F1 pool	PS01	Avr F1 pool
N116	0	148	577	269
Vir F1 pool	148	0	241	88
PS01	577	241	0	246
Avr F1 pool	269	88	246	0

Figure 4. Transcriptome analysis of whole aphids. Aphids were infested on *Medicago truncatula* A17 for 24 h, with n=5 biological replicates. A. Clustering of transcriptional responses of pea aphid PS01, N116, bulked F1 VIR and AVR progeny. Responses within biological replicates are more strongly correlated than responses among different aphid genotypes; B. Principal components analysis. The top 2 principal components explain >45% of the variation among transcriptional responses of PS01, N116, AVR and VIR F1 progeny replicates, and separate the responses of the different aphid genotypes and F1 pools. C. Numbers of genes differentially expressed between the different aphid genotypes and pools. Accompanying gene lists and annotations are provided in Supplementary Material 3.

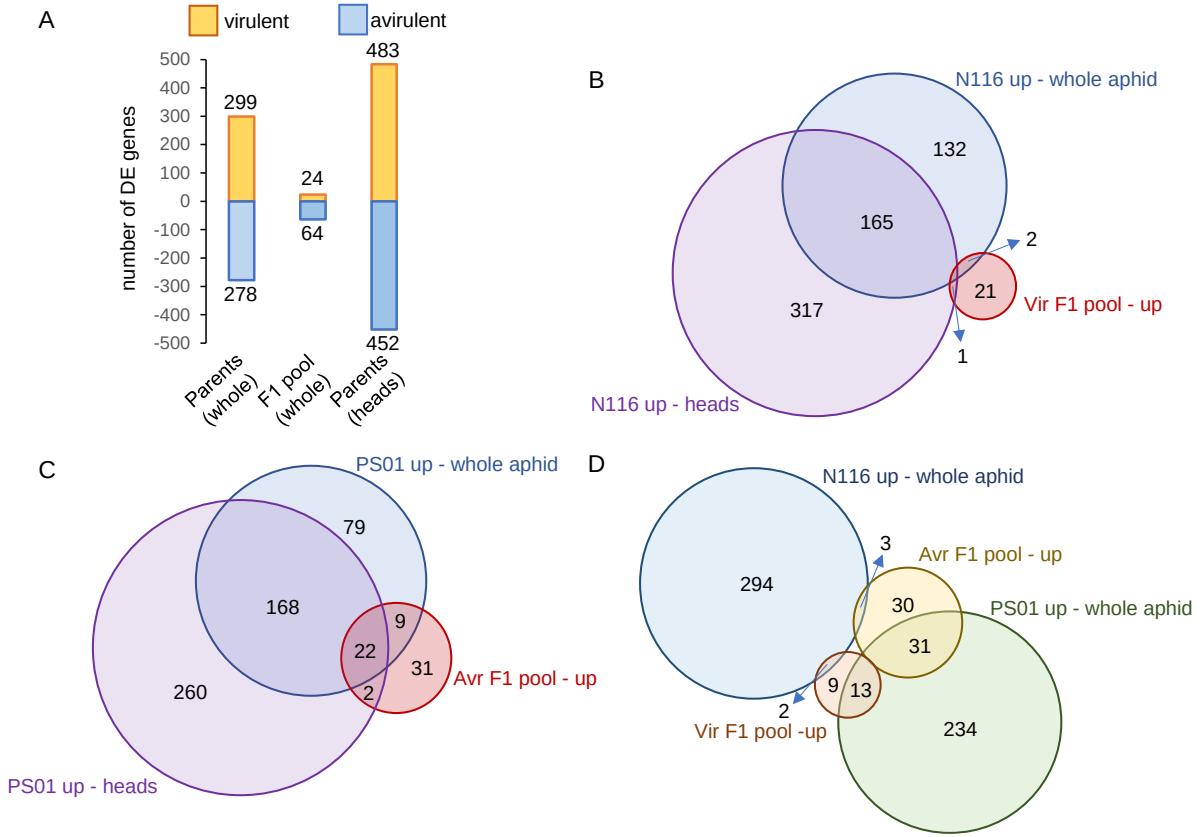


Figure 5. Differential gene expression in pea aphid genotypes N116, PS01 and bulked F1 pools of virulent and avirulent progeny. A. Numbers of genes up- versus down-regulated in comparisons of parent genotypes N116 (virulent) and PS01 (avirulent), and their F1 progeny pools, all on A17 host plants. Orange bars represent the numbers of genes up-regulated in genotype N116 or the VIR F1 pool compared with the genotype PS01 and the AVR pool, respectively. Blue bars represent the numbers of genes up-regulated in genotype PS01 or the AVR F1 pool compared with the genotype N116 and the VIR pool, respectively; B. Overlaps in genes up-regulated in genotype N116 whole body and head tissues compared to genotype PS01, and up-regulated in the VIR F1 pool compared to the AVR F1 pool; C. Overlaps in genes up-regulated in genotype PS01 whole body and head tissues compared to genotype N116, and up-regulated in the AVR F1 pool compared to the VIR F1 pool; D. Overlaps in up-regulated genes among whole body transcriptomes of N116, PS01, VIR F1 pool and AVR F1 pool.

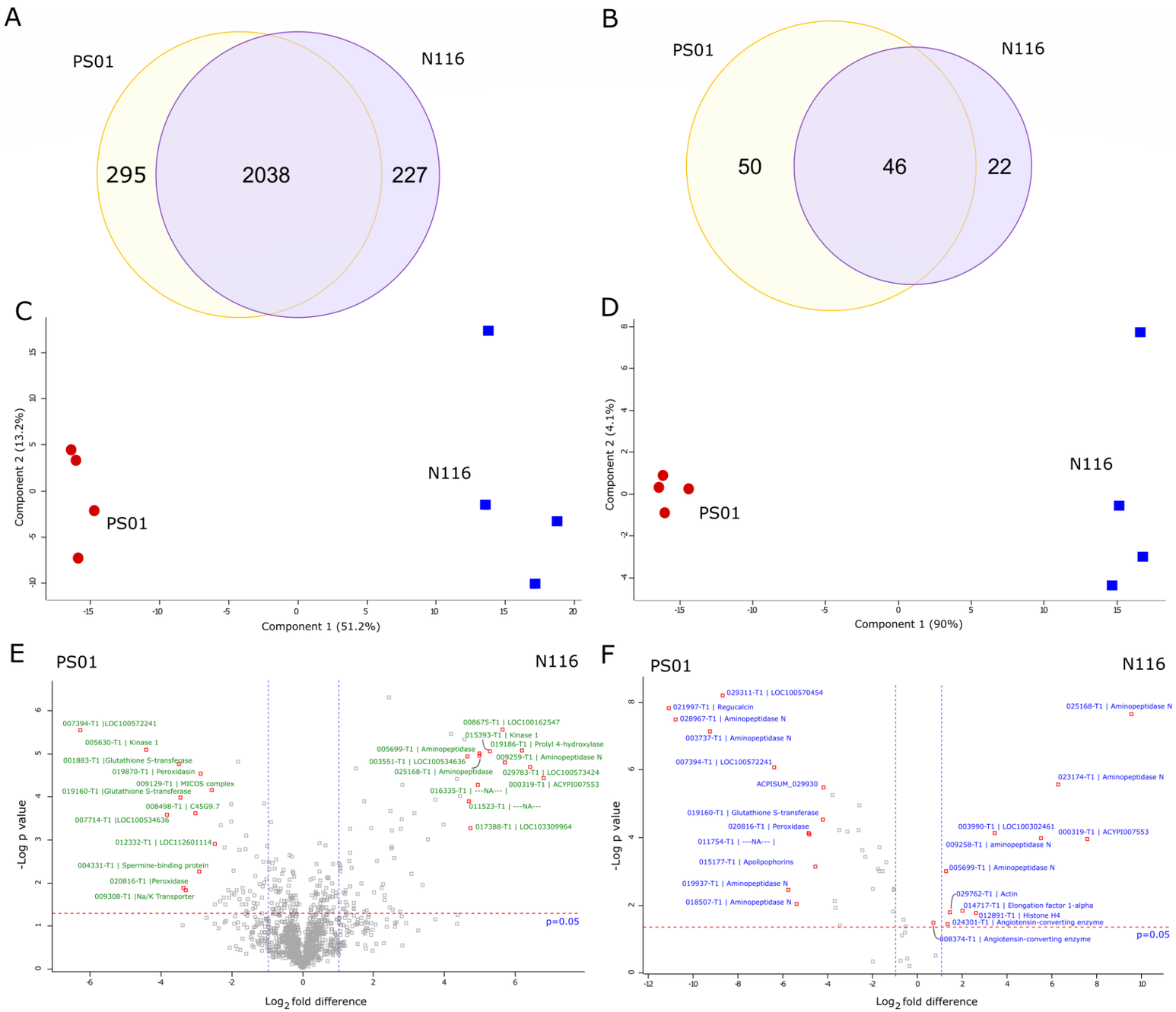
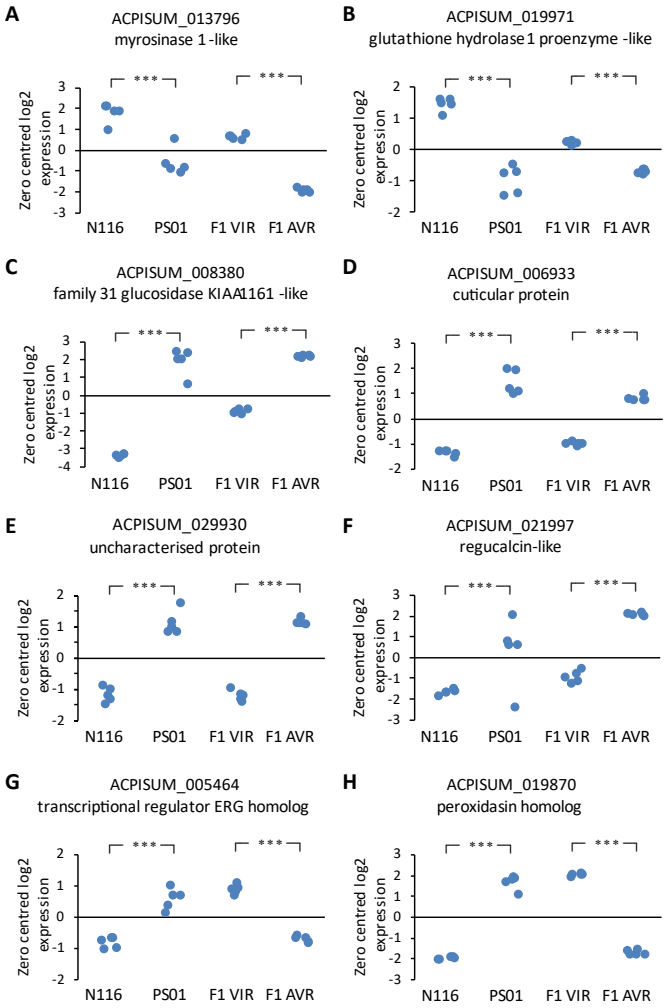


Figure 6. Comparative proteomic analysis of salivary glands and saliva for pea aphid genotypes N116 and PS01. Venn diagrams of the number of proteins shared and found exclusively in A) salivary glands and B) saliva identified for both genotypes. Principal Components Analysis (PCA) of C) salivary glands and D) saliva distinguishes both genotypes clearly. Volcano plots based on -$\log_{10}$ p values and $\log_2$ fold differences highlighting the statistically significant differentially abundant (SSDA) proteins ($p \leq 0.05$) for E) salivary glands and F) saliva. Annotations are shown for the top 12 proteins of increased and decreased abundances.



835

836 **Figure 7. Selected differentially expressed genes from whole body transcriptomes**. A,B representative genes upregulated both in virulent parent and in
837 virulent F1 pool; C-F representative genes upregulated both in avirulent parent and in avirulent F1 pool. G,H representative genes with opposite regulation
838 between parent and F1 pairs. Each point represents an individual RNA-Seq library (n=5). *** indicates FDR<0.001.