Differential Bumble Bee Gene Expression Associated With Pathogen Infection And Pollen Diet

Jonathan J. Giacomini (jjgiacom@ncsu.edu)
North Carolina State University
Lynn S. Adler
University of Massachusetts Amherst https://orcid.org/0000-0003-2125-5582
Benjamin J. Reading
North Carolina State University https://orcid.org/0000-0002-0778-4069
Rebecca E. Irwin
North Carolina State University https://orcid.org/0000-0002-1394-4946

Research Article

Keywords: parasitism, medicinal, infection, immune

Posted Date: November 29th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-912647/v1

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at BMC Genomics on March 29th, 2023. See the published version at https://doi.org/10.1186/s12864-023-09143-5.

Abstract

Background: Diet and parasitism can have powerful effects on host gene expression. However, how specific dietary components affect host gene expression that could feed back to affect parasitism is relatively unexplored in many wild species. Recently, it was discovered that consumption of sunflower (*Helianthus annuus*) pollen reduced severity of gut protozoan pathogen *Crithidia bombi* infection in *Bombus impatiens* bumble bees. Despite the dramatic and consistent medicinal effect of sunflower pollen, very little is known about the mechanism(s) underlying this effect. However, sunflower pollen extract increases rather than suppresses *C. bombi* growth *in vitro*, suggesting that sunflower pollen reduces *C. bombi* infection indirectly via changes in the host. Here, we analyzed whole transcriptomes of *B. impatiens* workers to characterize the physiological response to sunflower pollen consumption and *C. bombi* infection to isolate the mechanisms underlying the medicinal effect. *B. impatiens* workers were inoculated with either *C. bombi* cells (infected) or a sham control (un-infected) and fed either sunflower or wildflower pollen *ad libitum*. Whole abdominal gene expression profiles were then sequenced with Illumina NextSeq 500 technology.

Results: Among infected bees, sunflower pollen upregulated immune transcripts, including the anti-microbial peptide hymenoptaecin, Toll receptors and serine proteases. In both infected and un-infected bees, sunflower pollen upregulated putative detoxification transcripts and transcripts associated with the repair and maintenance of gut epithelial cells. Among wildflower-fed bees, infected bees downregulated immune transcripts associated with phagocytosis and the phenoloxidase cascade.

Conclusions: Taken together, these results indicate dissimilar immune responses between sunflower- and wildflower-fed bumble bees infected with *C. bombi*, a response to physical damage to gut epithelial cells caused by sunflower pollen, and a strong detoxification response to sunflower pollen consumption. Identifying host responses that drive the medicinal effect of sunflower pollen in infected bumble bees may broaden our understanding of plant-pollinator interactions and provide opportunities for effective management of bee pathogens.

Background

Organisms are exposed to a wide range of environmental challenges, such as fluctuations in nutrient availability, ingestion of toxins, and exposure to pathogens. As a result, organisms modulate gene expression patterns at the transcriptional level to cope with these environmental challenges. Interactions between diet and pathogen infection can create feedbacks in gene expression that impact organism health. For example, in phytophagous insects, nutrient availability (1–4) or phytotoxins (5, 6) in the diet can reduce host immune gene expression, thus making a consumer more vulnerable to infection, or may enhance immune gene expression (7, 8), thus conveying health benefits to the consumer. Despite the large body of literature on multitrophic interactions (9–11), studies on whole genome transcriptomic responses to different diets and the interplay between diet and parasite infection remain rare. Here we focus on the relationship between bumble bees, pollen diet, and a protozoan pathogen to shed light on

how specific dietary components affect host bee gene expression that could feed back to affect pathogen infection.

Consumption of sunflower (*Helianthus annuus*) pollen was recently discovered to reduce severity of the gut protozoan pathogen *Crithidia bombi* in *Bombus impatiens* worker and queen bumble bees by at least 80% relative to control pollen (12–14). The medicinal effect of sunflower pollen may extend beyond bumble bees and *C. bombi*. For example, sunflower pollen reduced infection by the microsporidian *Nosema ceranae* in honey bees *Apis mellifera* (12, 15). Further, Asteraceae pollen protected mason bees (*Osmia*) from brood parasitism (15). Despite the dramatic and consistent medicinal effect of sunflower pollen on *C. bombi* infection in *B. impatiens*, we have yet to identify the mechanism(s) underlying this effect. However, since sunflower pollen extract increased rather than suppressed *C. bombi* growth *in vitro* (16), it is likely that sunflower pollen reduces *C. bombi* infection via changes in host physiological functions, such as immune and detoxification systems, or physical changes in the gut environment that prevent parasite growth and reproduction. There is an increasing interest in dietary ingredients that are appropriate to support insect pollinator health, including digestive and immune functions. Unfortunately, population declines have been observed for a number of bee species worldwide (17, 18) due to multiple stressors, including poor nutrition, habitat loss and pathogens (19, 20). Thus, identifying mechanisms that underly the medicinal effect of sunflower pollen in infected bumble bees may broaden our understanding of pollinator disease ecology and provide opportunities to effectively manage bee pathogens.

Chemical or physical properties of sunflower pollen may have indirect negative effects on *C. bombi* mediated through changes in host bumble bee physiology. Sunflower pollen has relatively low protein content and lacks the essential amino acids methionine and tryptophan (21). Many microbial gut parasites rely on their host for nutrition (22), and thus poor host nutrition can limit parasite growth and reproduction. However, the consumption of buckwheat pollen (*Fagopyrum escueluetum*), which matched sunflower pollen in crude protein and amino acid content, did not reduce *C. bombi* infection (12), and the consumption of a presumably nutritionally balanced sunflower pollen diet diluted with a diverse wildflower pollen blend (1:1 ratio by weight) significantly reduced *C. bombi* infection in bumble bees (23), ruling out poor host nutrition as the mechanism. Sunflower pollen also contains plant defensive compounds, including neochlorogenic acid, quercetin glycosides (24), and flavonoids (25, 26), as well as a variety of saturated fatty acids and sterols that may have antimicrobial properties (27, 28). Adler et al. (29) demonstrated that triscoumaroyl spermidine and rutin (a proxy for quercetin glycosides), as well as nine fatty acids found in sunflower pollen, failed to reduce *C. bombi* infection *in vivo* when mixed into non-medicinal control pollen diets. However, consumption of chlorogenic acid, which is similar in structure to neochlorogenic acid found in sunflower pollen (24), reduced *Crithidia* sp. infection in bumble bees (30). Similar to what was found for sunflower pollen extracts, Palmer-Young et al. (31) demonstrated that chlorogenic acid did not have a direct toxic effect on *Crithidia* sp. cells, suggesting an indirect effect mediated through changes in host bumble bee physiology.

Analyzing changes in the bumble bee transcriptome in response to different diets and infection may shed light on the molecular pathways involved in the medicinal effect of sunflower pollen consumption. If sunflower pollen reduces infection via changes in the host bumble bee immune system, we would expect transcripts associated with canonical immune signaling pathways, including the Melanization and Encapsulation, Toll, Jak/STAT, IMD/JNK, or RNAi pathways (32), to be differentially expressed in sunflower pollen-fed bees compared to bees fed wildflower pollen control diet. In addition, a variety of detoxification genes are found in the genome of bumble bees, although to a lesser extent than other phytophagous insects (33). Several studies have found that such genes play a major role in bee metabolism of phytotoxins and xenobiotics found in honey and pollen (34–37) and may also elicit an immune response in bees (38). Congruent expression of putative immune and detoxification genes may indicate that plant defensive chemicals play an important role in the medicinal effect. Alternatively, if echinate sunflower pollen decreases digestibility, or causes physical damage to the bee gut, then we may expect changes in gene expression associated with plasma membrane repair that mediate active resealing of membrane disruptions to maintain homeostasis.

The objective of this study was to use a RNAseq-based whole transcriptome approach to identify key molecular pathways involved in the medicinal effect of sunflower pollen consumption in bumble bees infected with *C. bombi*. We analyzed differences in gene expression profiles of adult *B. impatiens* workers inoculated with live *C. bombi* cells or a sham control, and then either fed a sunflower or wildflower pollen diet. Using a combination of traditional frequentist statistics and machine learning techniques, we found that consumption of sunflower pollen enhances bumble bee immune response to *C. bombi*, stimulates detoxification processes and upregulates genes associated with physical damage to or remodeling of gut epithelial cells. The data generated in this study provide a strong foundation to further explore the chemical and structural properties of sunflower pollen that drive the medicinal effect in bumble bees.

Results

Inoculation efficacy and pollen consumption. Only a subset of the initial group of bees were chosen for RNA sequencing (see Methods: Inoculation treatment). The remainder (hereafter termed non-RNAseq bees) were reserved to determine infection prevalence and intensity under the treatment conditions to assess inoculation efficacy. All but one (out of 18) of the non-RNAseq bees fed wildflower pollen were infected with *C. bombi*, suggesting successful inoculation for the RNAseq bees. Infected wildflower-fed bees had an average *C. bombi* intensity of 33.44 ± 12.3 cells/0.02 µL (mean ± SE), which is an approximately 42-fold change increase compared to the initial inoculum. Sunflower pollen significantly reduced the prevalence of *C. bombi* infection by 3507 % ($\chi^2 = 10.678$, $p = 0.001$) and the intensity of infection by 8.77 % ($\chi^2 = 5.866$, $p = 0.015$) relative to wildflower pollen. The interaction between pollen diet and the average daily rate of pollen consumption, as well as the main effect of average daily rate of pollen consumption, did not have a significant effect on *Crithidia* infection intensity or prevalence ($\chi^2 < 0.388$, $p > 0.226$, for both) in the non-RNAseq bees. Of the RNAseq bees, sunflower-fed bees had 24%

lower average rate of pollen consumption compared to wildflower-fed bees ($F_{1,16}$ = 5.374, p = 0.034; Figure S2, *Supporting Information*).

Assembly and BLAST. In total, between 13,702,311 and 60,011,808 cleaned reads were obtained after sequencing and trimming (Table S1, *Supporting Information*). The average mapping rate of clean reads to the *Bombus impatiens* genome was 91.38 ± 1.2 %, resulting in 22,726 unique transcripts. Reannotating the transcripts using OmicsBox blastx using the nr database against all arthropods yielded a total of 17,077 hits. The greatest number of top BLAST hits were found in *B. impatiens*, with the top five from *Bombus* (Figure S3, *Supporting Information*), giving us confidence in our read quality.

Differential gene expression - Infected bees: sunflower vs. wildflower. Among infected bees, 40 transcripts were differentially expressed between sunflower- and wildflower-fed bees based on the DESeq2 model (FDR < 0.05; Table S2, *Supporting Information*). Notably, four proteins associated with the innate immune system were significantly upregulated in infected sunflower-fed bees (Figure 1), including the anti-microbial peptide *hymenoptaecin* (XP_003494933), a *serine protease inhibitor dipetalogastin* (XP_012236217), a plasma membrane-bound glycoprotein *alkaline phosphatase 4-like* (XP_012241779), and *digestive cysteine proteinase 1* (XP_003494144). Additionally, four proteins associated with detoxification and oxidative stress were upregulated in infected sunflower-fed bees (Figure 2), including *glucose dehydrogenase [FAD, quinone]-like* (XP_012248181), *cytochrome P450 9e2-like* (XP_033174299) *oxidation resistance protein 1 isoform X6* (XP_024222768) and *beta−1,4−glucuronyltransferase 1* (XP_003491810). Two proteins associated with gut morphology were upregulated in sunflower-fed bees (Figure 3), including the glycoside hydrolase *endochitinase* (XP_012241960) and the gamma secretase *nicastrin* (XP_012247186).

Based on machine learning, we were able to correctly classify sunflower-fed from wildflower-fed infected bumble bees in 100% (SD = 0%) of the instances when using the top 160 through 105 ranked transcripts (Figure S4, *Supporting Information*). Re-training the SMO model with the randomized data sets, the overall mean percent correct classification was 47.31% (SD = 24.09%), the average kappa statistic was -0.04 (SD = 0.46) and the AUROC was 0.48 (SD = 0.24), indicating that true learning occurred in the optimized SMO model with 160 top ranked transcripts. The IDs, gene functions and expression levels of the top 160 ranked transcripts are presented in Table S3 (*Supporting Information*).

Notably, machine learning identified importance of a number of proteins associated with the innate immune system that were upregulated in infected sunflower-fed bees compared to infected wildflower-fed bees (Figure 1), including Toll pathway receptor *spaetzle 4* (XP_033361233), several serine proteases, including *transmembrane protease serine 9-like* (XP_033180356), *serine protease inhibitor dipetalogastin* (XP_012236217) and *probable serine/threonine-protein kinase samkC* (XP_012245395), *chymotrypsin-1* (XP_003485243), *trypsin-3* (XP_012240481) and *trypsin-1* (XP_012240481). Machine learning also identified several proteins associated with detoxification that were upregulated in infected sunflower-fed bees (Figure 2), including *UDP-glucuronosyltransferase 2B17-like* (XP_033176691), three transcripts for *cytochrome P450 9e2-like* (XP_033174299 and XP_003484581), two transcripts for *oxidation resistance*

protein 1 (XP_024222768), *thioredoxin reductase 1* (XP_012247756) and *E3 ubiquitin-protein ligase MARCH5* (XP_003492433). A transcript for *probable cytochrome P450 305a1* (XP_003484727) was downregulated in infected sunflower-fed bees. Six proteins associated with gut morphology, epithelium repair and maintenance were also identified by machine learning, which were upregulated in infected sunflower-fed bees (Figure 3), including two transcripts for *chitinase-3-like protein 1* (XP_003488774), *actin 5c* (XP_014484761), *catenin alpha* (XP_017011441), *Partitioning defective 3 homolog* (XP_012236781), and two transcripts for *lysosomal aspartic protease* (XP_003489428).

Functional enrichment - Infected bees: sunflower vs. wildflower. We did not find significant enrichment of any Gene Ontology (GO) biological process or molecular function terms for DEGs from the DESeq2 model. However, GO enrichment analysis based on the combination of the top 160 ranked transcripts identified by machine learning indicated significant enrichment of proteolysis, glucosidase, hydrolase, carboxypeptidase, exopeptidase and peptidase activities, as well as several carbohydrate metabolic processes in infected sunflower-fed bees compared to infected wildflower-fed bees. (Figure S5, *Supporting Information*), suggesting a metabolic response to xenobiotics and pollen nutrients (i.e., proteins, lipids and starches).

IPA: Canonical Pathways

We used Qiagen Ingenuity Pathway Analysis (IPA) software to further interpret functions of differentially expressed transcripts in each treatment pairwise comparison. A total of 104 out of the 160 top ranked optimal transcripts identified by machine learning were successfully mapped into IPA; 45 transcripts were uncharacterized, and we were unable to find human, rat or mouse orthologs for 11 transcripts. The top enriched canonical pathway was NRF2-mediated oxidative stress response (Table S4; *Supporting Information*), which was predicted to be activated in infected sunflower-fed bees compared to infected wildflower-fed bees (z-score = 2.00, p-value < 0.0001). This pathway elicits a cellular defense response to oxidative stress, including induction of detoxifying enzymes and antioxidant enzymes. Several other canonical pathways that overlap with the NRF2-mediated oxidative stress response pathway and play a role in response to oxidative stress were also enriched, including the thioredoxin pathway, acetone degradation, nicotine degradation II & III pathways, several melatonin degradation pathways, the epithelial adherens junction signaling pathway, and the aryl hydrocarbon receptor signaling pathway. The epithelial adherens junction signaling pathway also plays an important role in the maintenance of epithelial cell layers.

Differential gene expression - Uninfected bees: sunflower vs. wildflower. Among uninfected bees, 10 transcripts were differentially expressed between sunflower- and wildflower-fed bees based on the DESeq2 model (FDR < 0.05; Table S5, *Supporting Information*). Seven transcripts were upregulated in uninfected sunflower bees, four of which were characterized: *dehydrogenase [FAD, quinone]-like* (XP_033180074), *jerky protein homolog-like* (XP_012248162), *RNA-directed DNA polymerase from mobile element jockey-like* (XP_012244110) and the *putative odorant receptor 92a* (XP_024226497). Three

proteins were down-regulated in uninfected sunflower bees, one of which was characterized: *dynein beta chain, ciliary-like* (XP_033357315).

Based on machine learning, we were able to correctly classify uninfected sunflower-fed from wildflower-fed bumble bees in 100% of the instances with the top 141 through 114 ranked transcripts (Figure S6, *Supporting Information*). Re-training the SMO model with the randomized data sets, the overall mean percent correct classification was 51.28% (SD = 21.54%), the average kappa statistic was 0.03 (SD = 0.44) and the AUROC was 0.51 (SD = 0.22), indicating that true learning occurred in the optimized SMO model with 141 top ranked transcripts. The IDs, gene functions and expression levels of the top 141 ranked transcripts are presented in Table S6 (*Supporting Information*).

Similar to gene expression patterns for infected bees, machine learning identified proteins associated with the immune system (Figure 1), detoxification (Figure 2) and gut morphology (Figure 3) as important for distinguishing between sunflower and wildflower bees. Notably, sunflower pollen upregulated the pro-inflammatory regulator *tyrosine-protein phosphatase* (XP_012236351) and *major royal jelly protein 1* (XP_012247599), the latter of which has been shown to have antimicrobial effects in bees (39) and upregulated in response to *Crithidia* sp. infection (40). In addition, sunflower bees upregulated *WD repeat domain-containing protein 83* (XP_012239808), a scaffold protein that regulates the Extracellular signal Related Kinase (ERK) cascade associated with an inflammatory response to wounding (41) and enterocyte gut epithelial cell proliferation (42).

Functional enrichment - Uninfected bees: sunflower vs. wildflower. We did not find significant enrichment of any GO biological process or molecular function terms for DEGs identified in the DESeq2 model. However, GO enrichment analysis based on the combination of the top 141 top ranked transcripts identified by machine learning indicated that a number of biological processes involving pigmentation and oxidation-reduction (redox) reactions were significantly enriched (Figure S7, *Supporting Information*), indicating a detoxification response in uninfected sunflower-fed bees.

IPA: Canonical Pathways

A total of 68 out of the 141 top ranked optimal transcripts identified by machine learning were successfully mapped into IPA; 58 transcripts were uncharacterized, and we were unable to identify human, rat or mouse orthologs for 15 transcripts. Consistent with the IPA analysis for infected bees, a number of significant canonical pathways associated with xenobiotic metabolism, and gastrointestinal physiology were enriched in uninfected bees fed sunflower pollen. The top enriched canonical pathway was epithelial adherens junction signaling pathway (Table S4, *Supporting Information*), followed by sorbitol degradation, stearate biosynthesis, calcium signaling, and protein kinase A signaling. The NRF2-mediated oxidative stress response and the LPS/IL-1 Mediated Inhibition of RXR pathways were also enriched, both of which were predicted to be activated in infected sunflower-fed bees compared to infected wildflower-fed bees. IPA was unable to predict activation of any of the enriched canonical pathways in uninfected sunflower-fed bees.

Differential gene expression - Sunflower-fed bees: infected vs. uninfected. Among sunflower-fed bees, 12 transcripts were differentially expressed between infected and uninfected bees based on the DESeq2 model (FDR < 0.05; Table S7, *Supporting Information*). Of immune transcripts, the antimicrobial peptide *hymenoptaecin* (XP_003494933), the plasma membrane-bound glycoprotein *alkaline phosphatase 4* (XP_012241779), and the proteolytic enzyme *trypsin alpha-3-like* (XP_003491285) were upregulated in infected bees (Figure 1). The detoxification protein *glucose dehydrogenase [FAD, quinone]-like* (XP_033180074) was downregulated in infected bees (Figure 2). The glycoside hydrolase *endochitinase* (XP_012241960) was significantly upregulated in infected sunflower-fed bees (Figure 3).

Machine learning fairly predicted infection treatment among sunflower-fed bees (Figure S8, *Supporting Information*), in most instances only reaching 80% correct classification with considerably large standard deviation (> 30%). The best machine learning classification was obtained from the top 80 through 78 ranked transcripts (% CC: 90.00 + /- 30.02; mean +/- SD); IDs and gene functions are presented in Table S6 (*Supporting Information*). Re-training the SMO model with the randomized data sets, the overall mean percent correct classification was 47.65% (SD = 20.06%), the average kappa statistic was -0.04 (SD = 0.36) and the AUROC was 0.48 (SD = 0.20). Substantial overlap between variation (SD) around the average percent correct classification between the negative control and optimized model, indicates that true learning failed in the SMO model with 80 top ranked transcripts.

Functional enrichment - Sunflower-fed bees: infected vs. uninfected. We did not find significant enrichment of any GO biological process or molecular function terms for DEGs identified in either the DESeq2 model or the top 80 ranked transcripts identified by machine learning. Since machine learning poorly classified infection status among sunflower-fed bees, and so few transcripts were differentially expressed in the DESeq2 model, we did not perform IPA analysis to avoid misleading results.

Differential gene expression - Wildflower-fed bees: infected vs. uninfected. Among wildflower-fed bees, 17 proteins were differentially expressed between infected and uninfected bees based on the DESeq2 model (FDR < 0.05; Table S9, *Supporting Information*). Notably, the proteolytic enzyme *digestive cysteine proteinase 1* (XP_003494144) and the detoxification enzyme *glucose dehydrogenase [FAD, quinone]-like* (XP_012248181) were downregulated in infected bees (Figure 1). No proteins associated with gut morphology were differentially expressed between infected and uninfected wildflower-fed bees (Figure 3).

The best machine learning classification was obtained from the top 98 through 41 ranked transcripts (% CC: 100.00 ± 00.00; mean ± SD; Figure S9, *Supporting Information*); the IDs, gene functions and expression levels are presented in Table S10 (*Supporting Information*). Re-training the SMO model with the randomized data sets, the overall mean percent correct classification was 47.0 % (SD = 20.3 %), the average kappa statistic was -0.03 (SD = 0.41) and the AUROC was 0.48 (SD = 0.22), indicating that true learning occurred in the optimized SMO model with 98 top ranked transcripts.

Notably, machine learning identified importance of several proteins associated with an immune response that were downregulated in infected wildflower bees (Figure 1), including a serine/threonine kinase *inhibitor of nuclear factor kappa-B kinase subunit epsilon* (XP_003486634) and *serine/threonine-protein*

phosphatase 1 regulatory subunit GAC1-like (XP_033179143). Machine learning also identified *luciferin 4-monooxygenase-like isoform X6* (XP_003491563), which was downregulated in infected bees (Figure 3) and in previous work is associated with detoxification in honey bees (36). No proteins associated with gut morphology were identified by machine learning as important for distinguishing infected from uninfected wildflower bees.

Functional enrichment - Wildflower-fed bees: infected vs. uninfected. We did not find significant enrichment of any GO biological process or molecular function terms for either the DEGs identified in the DESeq2 model or the top 98 ranked transcripts identified by machine learning.

IPA: Canonical Pathways

A total of 39 out of the 98 top ranked optimal transcripts identified by machine learning were successfully mapped into IPA; 56 transcripts were uncharacterized, and we were unable to identify human, rat or mouse orthologs for 3 transcripts. Six canonical pathways were enriched in infected compared to uninfected wildflower-fed bees (Table S4, *Supporting Information*). The top enriched canonical pathway was Choline Degradation I, driven solely by large downregulation of choline dehydrogenase (*Chdh*) in infected bees. Several nucleotide metabolism pathways were enriched; upregulation of phosphoribosyl pyrophosphate synthetase 1 (*Prps1*) was associated with enrichment of the PRPP Biosynthesis I pathway, and downregulation of acid phosphatase 3 (*Acp3*) was associated with enrichment of the NAD Phosphorylation and Dephosphorylation, Urate Biosynthesis/Inosine 5-phosphate degradation, Guanosine Nucleotides Degradation III and Adenosine Nucleotides Degradation II pathways.

Discussion

Consuming sunflower pollen resulted in the upregulation of transcripts associated with multiple physiological processes. Among infected bees, sunflower pollen upregulated transcripts associated with the Toll-mediated innate immune system, putative detoxification transcripts and transcripts associated with the repair and maintenance of gut epithelial cells. Among uninfected bees, sunflower pollen upregulated similar detoxification transcripts and transcripts associated with repair and maintenance of gut epithelial cells, but not a Toll-mediated immune response. In uninfected wildflower-fed bees, we did not detect upregulation of the same Toll-mediated immune response or detoxification transcripts in infected bees, but instead found signs of immune deactivation. Taken together, these results suggest that consuming sunflower pollen causes a different immune response than consuming wildflower pollen as well as a detoxification response. Moreover, gene expression patterns suggest that sunflower pollen may cause remodeling of or damage to the gut lining; such changes to the gut lining have the potential to feed back and enhance an effective immune response to *C. bombi* infection. Further research exploring the effects of plant defensive compounds from sunflower pollen on bee physiology, as well as chemical or physical damage caused by sunflower pollen consumption, may provide functional insight into the mechanisms underlying the medicinal effects of sunflower pollen, as well as broaden our understanding of how diet influences disease ecology.

One major hypothesis for the mechanism underlying the medicinal effect of sunflower pollen is that detoxification enhances the immune system. We found concurrent upregulation of both a Toll-mediated immune response and several putative detoxification enzymes in infected sunflower-fed bees, but not in uninfected sunflower-fed bees or wildflower-fed bees regardless of infection treatment. In insects, P450 cytochrome enzymes (CYPs) are known to regulate both pathogen infection, by producing reactive oxygen radicals (e.g., nitric oxide), and detoxification of xenobiotics (43). CYP variants play a role in resistance to fungal infection in silkworms (44), varroa mite resistance in honey bees (45), metabolism of potent insecticides in bumble bees (46), and *Crithidia* infection in bumble bees (40). In honey bees, the secondary metabolite*s* found in pollen and pesticides upregulate both detoxification and immunity transcripts, including several CYPs and the AMPs *abaecin* (36) and *hymenoptaecin* (47, 48), which enhanced the immune response of honeybees against both a microsporidian pathogen and viral infections (47). Oxidoreductases, including FAD-GLD that was upregulated in sunflower-fed bees, also play a major role in insect detoxification and act as a messenger to induce immune-related transcripts (49). Interestingly, Cox-Foster and Stehr (50) suggested that FAD-GLDs interact with phenoloxidase, and play an important role in the killing mechanism of pathogens by reducing quinone, which leads to the production of superoxide radicals that create a toxic environment for pathogens. Further research is needed to determine if detoxification of phytochemicals in sunflower pollen is independent of the immune response to *C. bombi* infection in bumble bees.

A major challenge for living organisms is to maintain homeostasis in the face of multiple internal and external stressors, such as pathogen infection, variation in nutrient supply and exposure to toxins. In response, complex immune systems have evolved to eliminate the potential threat and re-establish homeostasis without causing excessive damage to healthy cells and tissues. We found signs of deactivation of an immune response in infected wildflower-fed bees 72 hours post-inoculation, including downregulation of functional communication along the gut-brain axis, serine kinases, serine protein phosphatases, acid phosphatase and a cysteine proteinase. This pattern agrees with another study that found temporal expression of immune transcripts associated with Toll and melanization immune pathways in insects exposed to pathogens, reducing expression at 72 hours post-inoculation (51). Downregulation of an immune response may indicate that infection has bypassed the host's first line of defense and reflect diverting energy to other components of host physiology. This may prevent the toxic accumulation of reactive oxygen species and the production of energetically costly immune effectors, such as AMPs. Consequently, temporal variation in immune responses may obscure the underlying mechanism of medicinal sunflower pollen. We thus propose that follow-up studies track changes in host bumble bee physiology over time for an improved understanding of how different pollen or nectar diets mediate pathogen infection.

In addition to digestion and nutrient absorption, the digestive tract plays an important role in protecting an organism from absorption of ingested xenobiotics that cause oxidative stress, such as plant defense compounds or pesticides. Sunflower pollen consumption strongly upregulated multiple enzymes that are involved in the primary metabolism of xenobiotics, including a P450 cytochrome (CYP), a quinone oxidoreductase, two glucuronosyltransferases, an ATP-binding cassette (ABC) transporter, thioredoxin

reductase and E3 ubiquitinating proteins. A number of phenolic compounds that may cause oxidative stress in bees have been reported in sunflower pollen, including neochlorogenic acid (*5-O-Caffeoylquinic* acid) (24). The oxidation of phenolic compounds, such as neochlorogenic acid, can form quinones that inhibit protein digestion in herbivores (52) or have direct toxic effects on insects (53). Two P450 enzymes (3a4 and 2e1) and a quinone intermediate play a key role in the metabolism of neochlorogenic acid (54). Our BLAST results found that human CYP 3a4 is homologous to CYP 9e2 (e-value = 3.60E-75), the latter of which was strongly upregulated in infected and uninfected sunflower-fed bees. This, along with the simultaneous upregulation of *glucose dehydrogenase [FAD, quinone]*-like (FAD GLD) in sunflower-fed bees, which catalyzes the reaction of D-glucose and a quinone to yield D-glucono-1,5-lactone and a quinol, suggests detoxification of neochlorogenic acid found in sunflower pollen. Quinones, as either a by-product of neochlorogenic acid metabolism or as a natural component of sunflower pollen, may thus play a major role in the mechanism underlying the medicinal effect of sunflower pollen.

Pesticides are commonly used on sunflower crops to suppress weeds, herbivorous insects and plant pathogens (55) and can pose a substantial risk for bees (56). The upregulation of multiple detoxification enzymes in sunflower-fed bees could thus be an indication of pesticide contamination in sunflower pollen. However, while the pesticide residues in pollen used in this study were not measured, both the sunflower and wildflower pollen were sourced from the same suppliers as in Giacomini et al. (2018), which did measure pollen pesticide levels. In that study, a greater diversity of pesticide residues was found in wildflower compared to sunflower pollen, all but two of which were at trace levels. The two that were above trace levels were both miticides used to treat varroa mites in honey bee colonies. Sunflower pollen also contained a different miticide used to treat varroa in honey bees. Given that pesticide levels were low overall and greater in wildflower than sunflower pollen, it seems unlikely that pesticides are responsible for upregulation of detoxification transcripts in sunflower-fed bees.

In addition to detoxification and an immune response, gene expression patterns indicated wound healing activity in the abdominal gut tissues of *B. impatiens* in response to consuming sunflower pollen, including the enrichment of the Epithelial Adherens Junction Signaling pathway (EAJS), as well as transcripts associated with the formation of cellular surface protrusions, proliferation of fibroblasts and activation of signaling pathways in the brush border membrane. Wound healing in the digestive tract is a dynamic process that requires coordination between the proliferation of new cells, the reorganization of intracellular matrices and both inter- and intra-cellular signaling pathways that facilitate cell-to-cell adhesion (57). The cells of the epithelial layer are joined together by tight junctions composed of a branching network of transmembrane proteins, thus forming a contiguous and relatively impermeable membrane. Adherens junctions are specialized intercellular junctions, in which actin filaments are linked to cadherin molecules of adjacent cells via catenin molecules. These junctions perform multiple functions, including initiation and stabilization of cell-to-cell adhesion (58, 59). Disruption of epithelial cells by sunflower pollen may thus trigger a wound healing response or remodeling of the gut in bumble bees that involves the reformation of intercellular junctions. Since *Crithidia* sp. require tight adhesion to the gut lining in bumble bees to establish infection (60), it is plausible that phytosterols in sunflower

pollen or echinate spines, which are a particularly notable trait of sunflower pollen morphology (61), cause damage to the gut lining, which in turn prevents adhesion and reduces proliferation of *Crithidia* sp.

While this study is unable to differentiate between chemically- or mechanically-induced damage caused by sunflower pollen consumption, upregulation of both a detoxification response and a response to wound healing is consistent with recent evidence that closely related *Taraxacum* pollen damaged the gut lining of *B. terrestris* bumble bees (62). Similarly, that study was unable to differentiate between a chemical or mechanical cause of damage since a non-*Taraxacum* pollen diet spiked with phytosterols found in *Taraxacum* pollen and crushed *Taraxacum* pollen both induced damage to the gut lining of the digestive tract. On one hand, crushed pollen could increase abrasiveness and cause mechanical damage to the gut lining. Alternatively, crushed pollen could release phytochemicals that would be otherwise trapped in undigested pollen grains, and thus increase exposure to toxins that damage the gut lining. In our study, both sunflower pollen and the control wildflower pollen diets were provided to bees in the form of a paste, which required mechanical breakdown of honey-bee collected pollen pellets before adding water. During that process a small proportion of pollen grains are indeed fractured (JJG, *personal observation*), but since both diets were treated the same, we can rule out mechanical damage to the gut lining caused by fragmented pollen grains. However, *Helianthus sp*. pollen grains are much more echinate than *Taraxacum sp.* pollen, so we cannot rule out mechanical damage caused by intact pollen grains. If abrasiveness of echinate pollen causes damage to the gut lining, then pollen diets that contain a high proportion of pollen species with echinacious spines, regardless of plant family, will cause damage.

In insects, the peritrophic membrane (PM) is regularly shed and replaced via a well-regulated synthesis and turnover of chitin (63) to facilitate both growth and morphogenesis. The PM effectively functions to filter small molecules and aid in nutrient absorption, as well as protect the gut epithelium from damage from abrasive foods or pathogen invasion (40, 62, 64, 65). Infected bumble bees fed sunflower pollen increased expression of *chitinase-3-like protein 1* compared to wildflower-fed bees and increased expression of *endochitinase* compared to uninfected sunflower-fed bees. Differential expression of host bumble bee chitinases in response to *Crithidia* infection has been demonstrated in previous work (40), but an effect on *Crithidia* infection has not been detected. One hypothesis is that host bumble bees respond to *C. bombi* infection by increasing the turnover of the peritrophic membrane (PM) that lines the insect midgut, which physically removes *C. bombi* cells from the digestive tract. However, we did not see upregulation of PM-associated transcripts in infected wildflower-fed bees, indicating a synergistic interaction between sunflower pollen consumption and *C. bombi* infection, possibly mediated by abrasive damage to the PM by sunflower pollen. If the combination of sunflower pollen consumption and *C. bombi* infection increases the turnover of PM, then we may expect differences in the amount of chitin in the bumble bee peritrophic membrane, which can be quantified by image processing (66).

Conclusions

The data generated from this study provide an important foundation to disentangle the mechanism(s) underlying the medicinal effect of sunflower pollen in bumble bees. Chemical or mechanical properties of

sunflower pollen may enhance the bumble bee immune system, facilitating targeted destruction of *C. bombi* cells. Detoxification of phytotoxins found in sunflower pollen, such as neochlorogenic acid, may generate a toxic environment for *C. bombi* cells, or may stimulate and enhance a host immune response. Similarly, echinate sunflower pollen or phytosterols may cause damage to the gut lining, directly preventing growth and reproduction of *C. bombi*, or stimulating an effective host immune response. Future research should focus on disentangling effects of chemical and physical properties of sunflower pollen on host bumble bee physiology, and the implications for *C. bombi* infection. Identifying plant traits and host physiological responses that drive the medicinal effect of sunflower pollen in infected bumble bees may broaden our understanding of pollinator disease ecology and provide opportunities for effective management of bee pathogens.

Methods

Study System. *Bombus impatiens* is a native eusocial bee species in North America, ranging from Maine to Ontario to the eastern Rocky Mountains and south through Florida (67). They are generalists that visit a range of agricultural and native plants. *B. impatiens* have also been domesticated for crop pollination services throughout much of North America (68–70), subsequently making them a widely utilized study species. An annotated reference genome for *B. impatiens* (71) is available from the National Center for Biotechnology Information (NCBI). At the time of this study, NCBI BIMP_2.2 (GenBank assembly accession: GCA_000188095.4) contained 13,161 transcripts that code for 24,471 proteins.

Crithidia bombi (Zoomastigophora: Trypanosomatidae) is an infectious protozoan gut pathogen that can be contracted at flowers via fecal transmission and can also be horizontally transmitted within colonies (72, 73). *Crithidia* sp. reduce learning and foraging efficiency in worker bumble bees (74, 75), slow colony growth rates, especially early in the colony life cycle (76), reduce the likelihood of successful reproduction in wild colonies (77), and reduce infected queen fitness (78). *Crithidia* sp. infection is common; for example, *Crithidia* sp. infected over 60% of wild-caught *B. impatiens* in western MA (79) and commercial colonies can have high levels of infection (74).

Sunflowers (*Helianthus* sp.) belong to a large and diverse family (Asteraceae) with over 32,000 described species (80). *Helianthus annuus* is a major domesticated oilseed crop cultivated worldwide and a native US wildflower (81). With nearly two million acres of sunflowers planted in the US (82) and ten million acres planted in Europe annually (83), the high abundance of cultivated sunflowers combined with large nectar and pollen yields make it an important resource for bees.

Preparing inoculation treatments. Live *Crithidia bombi* cells were harvested from three wild *B. impatiens* workers collected near Stone Soup Farm, Hadley, MA, USA in 2014 (42.363911 N, -72.567747 W) and housed in commercial colonies of *B. impatiens* thereafter. The *Crithidia* species was identified in a previous study and confirmed to be *C. bombi* (84). Both the *C. bombi* source colony and experimental colony used in this experiment were purchased from Koppert Biological Systems (Howell, MI, USA). Colonies were fed with 30% sucrose solution and mixed wildflower pollen throughout their lifetimes and

housed in a dark room at 21 – 24ºC and ~50% rh. We made *C. bombi* inoculum using an established protocol (12, 85, 86). Briefly, bee digestive tracts of 15 workers, excluding the honey crop, were removed with forceps, placed into 1.5 mL microcentrifuge tubes with 300 µL of distilled water, and ground with a pestle. We allowed each sample to rest at room temperature for 4-5 hours so that gut material settled and *C. bombi* cells could ascend into the supernatant. *Crithidia bombi* cells were counted from a 0.02 µL sample of supernatant per bee with a Neubauer hemacytometer under a compound light microscope at 400X magnification. We then mixed 150 µL of the supernatant with distilled water to achieve a concentration of 2400 cells μL^{-1}. The sample was then mixed with an equal volume of 50% sucrose solution to yield inoculum with 1200 cells μL^{-1} in 25% sucrose. We made the sham inoculum following the same procedure as the *C. bombi* inoculum, but instead used the digestive tracts of five bees from the un-infected experimental colony.

Preparing pollen diets. We prepared two pollen diet treatments – sunflower and wildflower. Honey bee-collected sunflower pollen pellets were obtained from Changge Hauding Wax Industry (China) and sorted by color to remove impurities. We verified a pure batch of sunflower pollen by staining five samples with basic fuschin dye (87) and visually confirming only sunflower pollen was present with a compound microscope at 400X magnification. Honey bee-collected mixed wildflower pollen pellets were obtained from Koppert Biological Systems (Howell, MI, USA) and microscopically confirmed to contain < 5% Asteraceae pollen, identified by having spines on the exine (61). Experimental pollen diets were provided to bees as a paste produced by mixing ground pollen pellets with distilled water to achieve a uniform consistency.

Inoculation treatment. Experimental adult worker bumble bees were obtained from a single commercial *B. impatiens* colony that was determined to be uninfected by screening five workers using the methods described in Preparing inoculation treatments. Workers were removed from the colony and placed into individual plastic containers (7.5 cm x 10 cm x 5 cm) with mesh screen flooring. We starved the bees for 3-5 hours and then fed each a 10 µL drop of either the *C. bombi* inoculum or sham control; bees were assigned at random to inoculation treatment. The dose of *C. bombi* inoculum contained 12,000 *C. bombi* cells, which is within the concentration range bees are exposed to when foraging on flowers in the wild (88). Only bees that consumed the entire droplet (n = 120; 60 with *C. bombi* and 60 with sham inoculum) were used in the experiment.

All bees were then randomly assigned within inoculation treatment to either the sunflower (n = 60) or wildflower pollen diet (n = 60). Each day we fed bees fresh pollen paste of their assigned treatment, packed into an inverted lid of a 1.5 mL microcentrifuge tube and 1 mL of 30% sucrose via a filled and inverted plastic 1.5 mL microcentrifuge tube plugged with cotton (Richmond Dental & Medicine, Charlotte, NC, USA). We harvested tissue samples for RNA extraction 72 hours post-inoculation. We selected this time period because *C. bombi* counts start to diverge between bees fed sunflower vs. wildflower pollen between 48- and 72-hours post-inoculation (Figure S1; *Supporting Information*). Moreover, a recent study demonstrated that consuming sunflower pollen for approximately the first 72 hours or for 7 days after inoculation both reduced *C. bombi* intensity in bumble bees compared with control pollen (89). We

harvested tissue samples for RNA extraction from five workers per treatment that had the greatest average daily rate of pollen consumption (see Pollen consumption below). In total, we sequenced 20 samples: 5 replicates for each inoculation treatment and pollen diet. The remaining 100 bees (referred to as non-RNAseq-bees) were reserved to indirectly determine inoculation efficacy by measuring *C. bombi* infection.

Pollen consumption. Previous work showed that consuming higher concentrations of sunflower pollen had a stronger medicinal effect (23). Because it was not feasible to control how much pollen an individual bee consumed in our study, estimating pollen consumption was important to effectively model the relationship between diet and gene expression. To estimate consumption of pollen over the 72-hour period, we recorded the weight of each pollen feeder each day before placing it into the container with the bee and also 24 hours later. We accounted for feeder weight change caused by evaporation by placing an additional 30 pollen and nectar feeders (15 per pollen type) into containers that lacked a bee. Each day bees were provided fresh sucrose and pollen, yielding three days (post inoculation) of pollen consumption and evaporation measurements. We were not able to estimate nectar consumption because nectar feeders often leaked.

All statistical analyses using linear models were conducted with R version 4.0.2 (90). To estimate pollen consumption, we calculated evaporation-adjusted net consumption based on change in weight of the pollen feeder for each bee per day. Using the evaporation controls, we fit separate linear regressions for each day and pollen type, with initial weight regressed against weight 24-hr later. We then used the *predict* function in R to calculate an evaporation-adjusted feeder weight, yielding a net consumption estimate for each bee each day. Consumption variables (day 1, day 2, day 3, average daily rate (mg/day) and total) were strongly correlated based on Pearson's product moment correlations ($t > 4.538$, $df = 34$, $p < 0.001$ for all combinations). We thus focused solely on average daily pollen consumption rate for all gene expression analyses (see Differential gene expression analysis), as this was the metric used to select bees for RNA sequencing. We used ANOVA to test for differences in average daily pollen consumption rate between pollen diets and inoculation treatments for RNAseq bees. Model estimated means and Tukey-adjusted pairwise comparisons were obtained using the “emmeans” package (91).

Efficacy of inoculation. To verify that bees inoculated with *C. bombi* were infected and that sunflower pollen reduced *C. bombi* infection relative to wildflower pollen, we measured *C. bombi* prevalence and infection intensity of a random subset of the remaining bees that were not selected for RNA extraction, but also consistently consumed their pollen treatments over the first 72 hours [n(sunflower pollen) = 15 bees, n(wildflower pollen) = 18 bees]. Each bee was dissected and *C. bombi* cells were counted as in Preparing inoculation treatments, with the addition that all tools were washed with 70% ethanol and thoroughly dried between bees to prevent cross-contamination. We measured prevalence as the presence (1 or more *C. bombi* cells) or the absence of *C. bombi* cells per 0.02 µL sample, and *C. bombi* infection intensity as the number of flagellate *C. bombi* cells per 0.02 µL. We also removed the right forewing of each bee to measure marginal cell length, a proxy for bee size (92).

We used generalized linear models to analyze how pollen diets affected *C. bombi* infection prevalence and intensity. *Crithidia bombi* prevalence models were fit with a binomial distribution and infection intensity models were fit with a negative binomial distribution using the “MASS” package (93).

RNA extractions and sequencing. For bees selected for RNA sequencing, at 72 hours post-inoculation, the bees were anesthetized in a container of dry ice for 2 min. Using flame-sterilized forceps, we removed the abdomen of each anesthetized bee and placed it into a sterile 2 mL microcentrifuge tube with 2 mL of RNA-stabilizing reagent (RNAlater; ThermoFisher, Waltham, MA, USA; cat. No. AM7021). Each abdomen was slightly torn open with forceps for the RNA-stabilizing reagent to fully saturate the tissue sample and stored at 4°C for 24 hours. All samples were then kept in a -80°C freezer until RNA extraction. *Crithidia bombi* is a gut pathogen in bumble bees, and since our interests were in the effects of diet, we focus the sequencing on abdominal tissues. We did not use whole-bees to avoid potential tissue-specific gene expression patterns (e.g., differences between brain and gut gene expression), which has been shown in other insects (94) and may make it difficult to disentangle gut-specific responses.

Total RNA samples were submitted to the NCSU Genomic Sciences Laboratory for Illumina RNA library construction and sequencing. Purification of messenger RNA (mRNA) was performed using oligo-dT beads in the NEBNExt Poly(A) mRNA Magnetic Isolation Module. Complementary DNA (cDNA) libraries for Illumina sequencing were constructed using the NEBNext Ultra Directional RNA Library Prep Kit (NEB) and NEBNext Mulitplex Oligos for Illumina (NEB) using the manufacturer-specified protocol. Double-stranded cDNA was purified, end repaired, and “a-tailed” for adaptor ligation. Following ligation, samples were processed for a final fragment size (adapters included) of 400-550 bp using sequential AMPure XP bead isolation (Beckman Coulter, USA). Prior to library construction, RNA integrity, purity, and concentration was assessed using an Agilent 2100 Bioanalyzer with an RNA 6000 Nano Chip. Library enrichment was performed and specific indexes for each sample were added during the protocol-specified PCR amplification. The amplified library fragments were purified and checked for quality and final concentration using an Agilent 2100 Bioanalyzer with a High Sensitivity DNA chip. The final quantified libraries were pooled in equimolar amounts for clustering and sequencing on an Illumina NextSeq 500 DNA sequencer, using a 75 bp x 2 single end sequencing reagent kit. The software package Real Time Analysis was used to generate raw bcl (base call files), which were then de-multiplexed by sample into fastq files. Low-quality bases and adapter sequences were removed from raw sequence data for each sample using the Trimmomatic software package. Clean reads were mapped to the *B. impatiens* genome (BIMP 2.2) with HiSat2 version 2.1.0 (95) using default parameters. Gene expression was quantified using StringTie version 2.0 (96) to determine the number of reads uniquely mapping to exons and summed at the transcript level using gene features annotated in the NCBI *B. impatiens* annotation file (BIMP 2.2; GCA_000188095.4). Here after, for each transcript product mentioned throughout we report NCBI RefSeq protein accession IDs for *B. impatiens* or blast top hit taxa if *B. impatiens* was unavailable.

Differential gene expression analysis. We used a negative binomial generalized linear model using DESeq2 version 1.28.1 (97) in R to test for differences in gene expression between treatments. We tested for effects of treatment (pairwise comparisons of pollen diet and inoculation treatment) on gene

expression and included pollen consumption rate as a continuous covariate to control for variation caused by differences in average daily pollen consumption among bees. We used the Wald test to assess the significance of differentially expressed transcripts (DETs) and corrected for multiple testing using the Benjamini-Hochberg method with a cutoff of FDR < 0.05. Shrunken log2 fold changes for normalized transcript counts were obtained using lfcShrink function in DESeq2 with a shrinkage estimator based on a normal prior (97).

Machine Learning analysis. Transcriptomic data often suffers from the ‘curse of dimensionality’ due to having many more features than samples (98). Small sample sizes and the rapid loss of degrees of freedom thus make it a poor fit for traditional linear statistics, like regression and ANOVA (99). Standard analyses typically use multiple testing corrections to control false discovery rate (FDR). This method does not consider the highly interactive system of the transcriptome and often fails to detect small changes in gene expression (100). Machine learning, or artificial intelligence tools, can be used to address these challenges by building models from the data rather than fitting the data to rigid models.

We applied support vector machines (SVM) for classification using Weka 3.8.4 (101) to the gene expression profiles of each pairwise treatment comparison. Classification in machine learning is the task of learning to distinguish data points that belong to two or more categories in a dataset. Feature selection techniques can then be used to select a reduced number of variables that can maintain accurate classification. For example, comparing infected bumble bees fed either sunflower or wildflower pollen, support vector machines can be used to determine how well pollen diet can be classified from gene expression profiles. Feature selection can then be used to determine a reduced set of transcripts that accurately classify pollen diet, and are thus important (i.e., akin to statistical significance). To optimize data dimensionality for feature selection, we first selected a subset of transcripts from each pairwise treatment from the DESeq2 models that were differentially expressed based on an uncorrected p-value < 0.05. Transcripts (attributes) were then ranked using the InfoGain attribute evaluator and Ranker search method. This process evaluates the worth of an attribute by measuring the information gain with respect to the treatment (102). Specifically, InfoGain measures the difference in the Shannon's entropy of the system H(S) before a new attribute X is introduced, and H(S|X) is the entropy of the system after the attribute X has been introduced. We then created a series of ranked datasets, each including a subset of the top-ranked transcripts in a serial manner.

Preliminary classifier runs demonstrated that a support vector machine SMO, using the 10-fold (stratified hold-out) cross-validation method, correctly classified an average of 82.90% (SD = 27.68%) of bee treatment effects based on gene expression profiles, and thus was used in our analyses. SMO implements John Platt’s sequential minimal optimization algorithm for training a support vector classifier by globally replacing all missing values and transforming nominal attributes, in our case transcripts, into binary attributes (103). The machine-learning algorithm SMO has been successfully used to analyze gene expression profiles (104). We used the ranked data sets to train the SMO algorithm using both the 10-fold (stratified hold-out) and 66% split cross-validation method. For each data set, we repeated model training

100 times and used the classification performance metrics percent correct classification (%CC) and kappa statistic (k) to evaluate model performance.

We tested the efficacy of the optimized SMO model using a negative control method for machine learning. We first created 10 randomized data sets using the DESeq2 normalized counts with treatment randomly assigned. We then re-ran the SMO model training using the number of attributes that provided the best classification. Since there are always two class types, the predicted correct classification rate from random assignment should be approximately 50%, based on the Law of Probability, and the kappa statistic should be close to zero for a randomized negative control to demonstrate that true learning occurred in the optimized SMO models. This approach is detailed in previous studies (105, 106).

Gene Ontology enrichment analysis. Protein descriptions and gene ontology (GO) annotations for transcript sequences were obtained using OmicsBox version 1.4.11 software (https://www.biobam.com/omicsbox/). First, a BED formatted file of transcript coordinates was parsed from the NCBI BIMP 2.2 Annotation release. We then used bedtools version 2.29.2 (107) to extract nucleotide sequences based on the BED file coordinates. A BLASTX search was then performed with an E-value of 10^{-25} against all arthropod sequences in the NCBI non-redundant database, with the number of hits restricted to 20, followed by GO mapping and annotation for the resulting hits. We then ran InterProScan annotation for the sequences using the default settings and merged InterProScan GO annotations with BLASTX annotations. GO enrichment analysis was performed for all treatment comparisons to find significantly (FDR < 0.05) enriched GO biological process and molecular function terms in the test set of DEGs with respect to the reference set. We used the publicly available databases GeneCards (108) and UniProtKB/Swiss-Prot (109) as additional primary sources of information about DEGs.

IPA canonical pathway analysis. We used Qiagen Ingenuity Pathway Analysis (IPA) software to further interpret the differential expression of transcripts in each treatment pairwise comparison. IPA Knowledge Base maintains a large set of databases that consist of curated metabolic and signaling pathways. Transcripts were manually mapped to human, mouse, or rat ortholog gene IDs or those of other species based on UniProtKB accession numbers for use in IPA. We also performed an additional blastx for all transcripts using the same methods described in Gene Ontology enrichment analysis, but restricted to the *Homo sapiens*, *Mus* and *Rattus* taxonomies. Using these two different methods allowed us to double-check ambiguous orthologous gene symbols. We then performed a Core Analysis in IPA to determine enrichment of relevant canonical metabolic and signaling pathways based on gene expression patterns. We repeated Core Analysis for each pairwise treatment comparison based on the list of important transcripts identified using machine learning. The significance of the association between each gene set and a Canonical Pathway was determined from a p-value of overlap calculated using a right-tailed Fisher's Exact Test. In addition, IPA calculates a z-score based on the gene expression fold change values of each gene to estimate the state of activation or inhibition of each pathway. We report gene symbols (*Homo sapiens*, *Mus* or *Rattus*) for each gene product mentioned hereafter in the IPA canonical pathway results.

List of abbreviations

DET: Differentially expressed transcripts

FDR: False Discovery Rate

SD: Standard Deviation

SMO: Sequential minimal optimization

IPA: Ingenuity Pathway Analysis

SVM: Support Vector Machines

ABC: ATP-binding cassette

CYP: Cytochrome P450

LFC: Log Fold Change

AUROC: Area Under the Receiver Operating Characteristics

EAJS: Epithelial Adherens Junction Signaling pathway

Declarations

Ethics approval and consent to participate. Ethics approval is not required for experimental studies with *Bombus impatiens* bumble bees and the protozoan *Crithidia bombi*. Consent to participate is not relevant.

Consent for publication. Not applicable.

Competing interests. The authors declare that they have no conflict of interest.

Availability of data and materials. The datasets generated and/or analyzed during the current study are available in the NCBI Sequence Read Archive repository (BioProject ID: PRJNA780223), [https://www.ncbi.nlm.nih.gov/bioproject/ PRJNA780223].

Funding. This work was supported by a National Science Foundation Graduate Research Fellowship (NSF DGE-1746939) and USDA-AFRI-2018-08591. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding agencies.

Author contributions. JJG, LSA, and REI designed research; JJG performed research; JJG and BJR designed the analysis pipeline and JJG performed the analyses; JJG and REI wrote the paper with feedback from all authors.

Acknowledgements. We thank Sara June Connon for assistance with bumble bee colony maintenance, Andy Baltzeger for insight about experimental design and Betsy Scholl for assistance with bioinformatics.

References

1. Cotter SC, Simpson SJ, Raubenheimer D, Wilson K. Macronutrient balance mediates trade-offs between immune function and life history traits. Funct Ecol. 2011;25(1):186–98.
2. Di Pasquale G, Salignon M, Le Conte Y, Belzunces LP, Decourtye A, Kretzschmar A, et al. Influence of pollen nutrition on honey bee health: do pollen quality and diversity matter? PLOS ONE. 2013 Aug 5;8(8):e72016.
3. Brunner FS, Schmid-Hempel P, Barribeau SM. Protein-poor diet reduces host-specific immune gene expression in *Bombus terrestris*. Proc R Soc B. 2014;281:20140128.
4. Roger N, Michez D, Wattiez R, Sheridan C, Vanderplanck M. Diet effects on bumblebee health. J Insect Physiol. 2017 Jan 1;96:128–33.
5. Smilanich AM, Dyer LA, Chambers JQ, Bowers MD. Immunological cost of chemical defence and the evolution of herbivore diet breadth. Ecol Lett. 2009;12(7):612–21.
6. Tan W-H, Acevedo T, Harris EV, Alcaide TY, Walters JR, Hunter MD, et al. Transcriptomics of monarch butterflies (*Danaus plexippus*) reveals that toxic host plants alter expression of detoxification genes and down-regulate a small number of immune genes. Mol Ecol. 2019;28(22):4845–63.
7. Laurentz M, Reudler JH, Mappes J, Friman V, Ikonen S, Lindstedt C. Diet quality can play a critical role in defense efficacy against parasitoids and pathogens in the Glanville fritillary (*Melitaea cinxia*). J Chem Ecol. 2012;38(1):116–25.
8. Smilanich AM, Langus TC, Doan L, Dyer LA, Harrison JG, Hsueh J, et al. Host plant associated enhancement of immunity and survival in virus infected caterpillars. J Invertebr Pathol. 2018;151:102–12.
9. McCann K, Hastings A, Huxel GR. Weak trophic interactions and the balance of nature. Nat Lond. 1998;395:794–798.
10. Agrawal AA. Mechanisms, ecological consequences and agricultural implications of tri-trophic interactions. Curr Opin Plant Biol. 2000;3(4):329–35.
11. Abdala-Roberts L, Puentes A, Finke DL, Marquis RJ, Montserrat M, Poelman EH, et al. Tri-trophic interactions: bridging species, communities and ecosystems. Ecol Lett. 2019;22(12):2151–67.
12. Giacomini JJ, Leslie J, Tarpy DR, Palmer-Young EC, Irwin RE, Adler LS. Medicinal value of sunflower pollen against bee pathogens. Sci Rep. 2018;8(1):14394.
13. LoCascio GM, Aguirre L, Irwin RE, Adler LS. Pollen from multiple sunflower cultivars and species reduces a common bumblebee gut pathogen. R Soc Open Sci. 2019;6(4):190279.
14. Fowler AE, Stone EC, Irwin RE, Adler LS. Sunflower pollen reduces a gut pathogen in worker and queen but not male bumble bees. Ecol Entomol. 2020;45(6):1318–26.

15. Spear DM, Silverman S, Forrest JR, McPeek MA. Asteraceae pollen provisions protect *Osmia* mason bees (Hymenoptera: Megachilidae) from brood parasitism. Am Nat. 2016;187(6):797–803.
16. Palmer-Young EC. Pollen extracts increase growth of a trypanosome parasite of bumble bees. PeerJ. 2017;5:e3297.
17. Goulson D, Hanley ME, Darvill B, Ellis J, Knight ME. Causes of rarity in bumblebees. Biol Conserv. 2005;122(1):1–8.
18. Colla SR, Packer L. Evidence for decline in eastern North American bumblebees (Hymenoptera: Apidae), with special focus on *Bombus affinis* Cresson. Biodivers Conserv. 2008;17(6):1379–91.
19. Goulson D, Lye GC, Darvill B. Decline and Conservation of Bumble Bees. Annu Rev Entomol. 2008;53(1):191–208.
20. Cameron SA, Lozier JD, Strange JP, Koch JB, Cordes N, Solter LF, et al. Patterns of widespread decline in North American bumble bees. Proc Natl Acad Sci. 2011 Jan 11;108(2):662–7.
21. Nicolson SW, Human H. Chemical composition of the 'low quality' pollen of sunflower (*Helianthus annuus*, Asteraceae). Apidologie. 2012 Oct 3;44(2):144–52.
22. Coop RL, Holmes PH. Nutrition and parasite interaction. Int J Parasitol. 1996;26(8):951–62.
23. Giacomini JJ, Connon SJ, Marulanda D, Adler LS, Irwin RE. The costs and benefits of sunflower pollen diet on bumble bee colony disease and health. Ecosphere. 2021;00(00):e03663.
24. Kostić AŽ, Milinčić DD, Gašić UM, Nedić N, Stanojević SP, Tešić ŽL, et al. Polyphenolic profile and antioxidant properties of bee-collected pollen from sunflower (*Helianthus annuus* L.) plant. LWT. 2019;112:108244.
25. Kyselka J, Bleha R, Dragoun M, Bialasová K, Horáčková Š, Schätz M, et al. Antifungal polyamides of Hydroxycinnamic acids from sunflower bee pollen. J Agric Food Chem. 2018;66(42):11018–26.
26. Palmer-Young EC, Farrell IW, Adler LS, Milano NJ, Egan PA, Junker RR, et al. Chemistry of floral rewards: intra-and interspecific variability of nectar and pollen secondary metabolites across taxa. Ecol Monogr. 2018;89(1):e01335.
27. Lin S, Mullin CA. Lipid, polyamide, and flavonol phagostimulants for adult western corn rootworm from sunflower (*Helianthus annuus* L.) pollen. J Agric Food Chem. 1999;47(3):1223–9.
28. Kostić AŽ, Pešić MB, Trbović D, Petronijević R, Dramićanin AM, Milojković-Opsenica DM, et al. The fatty acid profile of Serbian bee-collected pollen–a chemotaxonomic and nutritional approach. J Apic Res. 2017;56(5):533–42.
29. Adler LS, Fowler AE, Malfi RL, Anderson PR, Coppinger LM, Deneen PM, et al. Assessing chemical mechanisms underlying the effects of sunflower pollen on a gut pathogen in bumble bees. J Chem Ecol. 2020;46:649–58.
30. Egan PA, Adler LS, Irwin RE, Farrell IW, Palmer-Young EC, Stevenson PC. Crop domestication alters floral reward chemistry with potential consequences for pollinator health. Front Plant Sci. 2018;9:1357.

31. Palmer-Young EC, Sadd BM, Adler LS. Evolution of resistance to single and combined floral phytochemicals by a bumble bee parasite. J Evol Biol [Internet]. 2016 [cited 2017 Feb 27]; Available from: http://onlinelibrary.wiley.com/doi/10.1111/jeb.13002/full
32. Barribeau SM, Sadd BM, Du Plessis L, Brown MJ, Buechel SD, Cappelle K, et al. A depauperate immune repertoire precedes evolution of sociality in bees. Genome Biol. 2015;16(1):83.
33. Claudianos C, Ranson H, Johnson RM, Biswas S, Schuler MA, Berenbaum MR, et al. A deficit of detoxification enzymes: Pesticide sensitivity and environmental response in the honeybee. Insect Mol Biol. 2006;15:615–36.
34. Mao W, Rupasinghe SG, Johnson RM, Zangerl AR, Schuler MA, Berenbaum MR. Quercetin-metabolizing CYP6AS enzymes of the pollinator *Apis mellifera* (Hymenoptera: Apidae). Comp Biochem Physiol B. 2009;154:427–34.
35. Mao W, Schuler MA, Berenbaum MR. CYP9Q-mediated detoxification of acaricides in the honey bee (*Apis mellifera*). Proc Natl Acad Sci USA. 2011;108:12657–62.
36. Mao W, Schuler MA, Berenbaum MR. Honey constituents up-regulate detoxification and immunity genes in the western honey bee *Apis mellifera*. Proc Natl Acad Sci USA. 2013;110:8842–6.
37. Johnson RM, Mao W, Pollock HS, Niu G, Schuler MA, Berenbaum MR. Ecologically appropriate xenobiotics induce cytochrome P50s in *Apis mellifera*. Plos One. 2012;7:e31051.
38. Schmehl DR, Teal PE, Frazier JL, Grozinger CM. Genomic analysis of the interaction between pesticide exposure and nutrition in honey bees (*Apis mellifera*). J Insect Physiol. 2014;71:177–90.
39. Fratini F, Cilia G, Mancini S, Felicioli A. Royal Jelly: An ancient remedy with remarkable antibacterial properties. Microbiol Res. 2016;192:130–41.
40. Riddell CE, Garces JDL, Adams S, Barribeau SM, Twell D, Mallon EB. Differential gene expression and alternative splicing in insect immune specificity. BMC Genomics. 2014;15(1):1–15.
41. Vomastek T, Schaeffer H-J, Tarcsafalvi A, Smolkin ME, Bissonette EA, Weber MJ. Modular construction of a signaling scaffold: MORG1 interacts with components of the ERK cascade and links ERK signaling to specific agonists. Proc Natl Acad Sci. 2004;101(18):6981–6.
42. Goessling W, North TE, Loewer S, Lord AM, Lee S, Stoick-Cooper CL, et al. Genetic interaction of PGE2 and Wnt signaling regulates developmental specification of stem cells and regeneration. Cell. 2009;136(6):1136–47.
43. Feyereisen R. Insect P450 enzymes. Annu Rev Entomol. 1999;44(1):507–33.
44. Xing D, Yang Q, Jiang L, Li Q, Xiao Y, Ye M, et al. RNA-seq analyses for two silkworm strains reveals insight into their susceptibility and resistance to *Beauveria bassiana* infection. Int J Mol Sci. 2017;18(2):234.
45. Jiang S, Robertson T, Mostajeran M, Robertson AJ, Qiu X. Differential gene expression of two extreme honey bee (*Apis mellifera*) colonies showing *Varroa* tolerance and susceptibility. Insect Mol Biol. 2016;25(3):272–82.

46. Troczka BJ, Homem RA, Reid R, Beadle K, Kohler M, Zaworra M, et al. Identification and functional characterisation of a novel N-cyanoamidine neonicotinoid metabolising cytochrome P450, CYP9Q6, from the buff-tailed bumblebee Bombus terrestris. Insect Biochem Mol Biol. 2019;111:103171.
47. Hu Y-T, Wu T-C, Yang E-C, Wu P-C, Lin P-T, Wu Y-L. Regulation of genes related to immune signaling and detoxification in *Apis mellifera* by an inhibitor of histone deacetylation. Sci Rep. 2017;7(1):1–14.
48. Shi T, Burton S, Zhu Y, Wang Y, Xu S, Yu L. Effects of field-realistic concentrations of carbendazim on survival and physiology in forager honey bees (Hymenoptera: Apidae). J Insect Sci. 2018;18(4):6.
49. Stone JR, Yang S. Hydrogen peroxide: a signaling messenger. Antioxid Redox Signal. 2006;8(3–4):243–70.
50. Cox-Foster DL, Stehr JE. Induction and localization of FAD-glucose dehydrogenase (GLD) during encapsulation of abiotic implants in *Manduca sexta* larvae. J Insect Physiol. 1994;40(3):235–49.
51. Jayaram A, Pradeep ANR, Awasthi AK, Murthy GN, Ponnuvel KM, Sasibhushan S, et al. Coregulation of host–response genes in integument: switchover of gene expression correlation pattern and impaired immune responses induced by dipteran parasite infection in the silkworm, *Bombyx mori*. J Appl Genet. 2014;55(2):209–21.
52. Felton G, Donato K, Broadway R, Duffey S. Impact of oxidized plant phenolics on the nutritional quality of dietar protein to a noctuid herbivore, *Spodoptera exigua*. J Insect Physiol. 1992;38(4):277–85.
53. Bolton JL, Trush MA, Penning TM, Dryhurst G, Monks TJ. Role of quinones in toxicology. Chem Res Toxicol. 2000;13(3):135–60.
54. Xie C, Zhong D, Chen X. Identification of the ortho-benzoquinone intermediate of 5-O-caffeoylquinic acid in vitro and in vivo: comparison of bioactivation under normal and pathological situations. Drug Metab Dispos. 2012;40(8):1628–40.
55. Elbert A, Haas M, Springer B, Thielert W, Nauen R. Applied aspects of neonicotinoid uses in crop protection. Pest Manag Sci Former Pestic Sci. 2008;64(11):1099–105.
56. Whitehorn PR, O'Connor S, Wackers FL, Goulson D. Neonicotinoid pesticide reduces bumble bee colony growth and queen production. Science. 2012 Apr 20;336(6079):351–2.
57. Bindschadler M, McGrath JL. Sheet migration by wounded monolayers as an emergent property of single-cell dynamics. J Cell Sci. 2007 Mar 1;120(5):876–84.
58. Perez-Moreno M, Jamora C, Fuchs E. Sticky business: orchestrating cellular signals at adherens junctions. Cell. 2003;112(4):535–48.
59. Hartsock A, Nelson WJ. Adherens and tight junctions: structure, function and connections to the actin cytoskeleton. Biochim Biophys Acta BBA-Biomembr. 2008;1778(3):660–9.
60. Koch H, Woodward J, Langat MK, Brown MJ, Stevenson PC. Flagellum removal by a nectar metabolite inhibits infectivity of a bumblebee parasite. Curr Biol. 2019;29(20):3494–500.
61. Blackmore S, Wortley AH, Skvarla JJ, Rowley JR. Pollen wall development in flowering plants. New Phytol. 2007 May 1;174(3):483–98.

62. Vanderplanck M, Gilles H, Nonclercq D, Duez P, Gerbaux P. Asteraceae paradox: Chemical and mechanical protection of *Taraxacum* pollen. Insects. 2020;11(5):304.
63. Zhu KY, Merzendorfer H, Zhang W, Zhang J, Muthukrishnan S. Biosynthesis, turnover, and functions of chitin in insects. Annu Rev Entomol. 2016;61:177–96.
64. Kuraishi T, Binggeli O, Opota O, Buchon N, Lemaitre B. Genetic evidence for a protective role of the peritrophic matrix against intestinal bacterial infection in Drosophila melanogaster. Proc Natl Acad Sci. 2011;108(38):15966–71.
65. Shibata T, Maki K, Hadano J, Fujikawa T, Kitazaki K, Koshiba T, et al. Crosslinking of a peritrophic matrix protein protects gut epithelia from bacterial exotoxins. PLoS Pathog. 2015;11(10):e1005244.
66. Nakata R, Yoshinaga N, Teraishi M, Okumoto Y, Mori N. An easy, inexpensive, and sensitive method for the quantification of chitin in insect peritrophic membrane by image processing. Biosci Biotechnol Biochem. 2019;83(9):1624–9.
67. Kearns CA, Thomson JD. The Natural History of Bumblebees: A Sourcebook for Investigations. Boulder, CO: University Press of Colorado; 2001.
68. Shipp J, Whitfield G, Papadopoulos A. Effectiveness of the bumble bee *Bombus impatiens* Cr. (Hymenoptera: Apidae), as a pollinator of greenhouse sweet pepper. Sci Hortic. 1994;57:29–39.
69. Whittington R, Winston ML. Comparison and examination of *Bombus occidentalis* and *Bombus impatiens* (Hymenoptera: Apidae) in tomato greenhouses. J Econ Entomol. 2004;97(4):1384–9.
70. Velthuis HH, Van Doorn A. A century of advances in bumblebee domestication and the economic and environmental aspects of its commercialization for pollination. Apidologie. 2006;37(4):421–51.
71. Sadd BM, Barribeau SM, Bloch G, De Graaf DC, Dearden P, Elsik CG, et al. The genomes of two key bumblebee species with primitive eusocial organization. Genome Biol. 2015;16(1):1–32.
72. Schmid-Hempel P, Durrer S. Parasites, floral resources and reproduction in natural populations of bumblebees. Oikos. 1991;62:342–50.
73. Durrer S, Schmid-Hempel P. Shared use of flowers leads to horizontal pathogen transmission. Proc R Soc B. 1994;258:299–302.
74. Gegear RJ, Otterstatter MC, Thomson JD. Does parasitic infection impair the ability of bumblebees to learn flower-handling techniques? Anim Behav. 2005;70(1):209–15.
75. Gegear RJ, Otterstatter MC, Thomson JD. Bumble-bee foragers infected by a gut parasite have an impaired ability to utilize floral information. Proc R Soc Lond B. 2006;273:1073–8.
76. Shykoff JA, Schmid-Hempel P. Parasites delay worker reproduction in bumblebees: consequences for eusociality. Behav Ecol. 1991 Oct 1;2(3):242–8.
77. Goulson D, O'connor S, Park KJ. The impacts of predators and parasites on wild bumblebee colonies. Ecol Entomol. 2018;43(2):168–81.
78. Brown MJF, Moret Y, Schmid-Hempel P. Activation of host constitutive immune defence by an intestinal trypanosome parasite of bumble bees. Parasitology. 2003 Mar;126(03):253–60.

79. Gillespie S. Factors affecting parasite prevalence among wild bumblebees. Ecol Entomol. 2010;35(6):737–47.
80. Funk VA. Systematics, evolution, and biogeography of Compositae. International Association for Plant Taxonomy; 2009.
81. Reagon M, Snow AA. Cultivated *Helianthus annuus* (Asteraceae) volunteers as a genetic "bridge" to weedy sunflower populations in North America. Am J Bot. 2006;93(1):127–33.
82. Holcomb R. Agricultural Statistics 2015. Washington D.C.: USDA: National Agricultural Statistics Service (NASS); 2015 Jun. Report No.: ISSN: 1949–1522.
83. Strange K, Krautgartner R, Lefebvre L, Rehder LE, Boshnakova M, Dobrescu M, et al. USDA Foreign Agricultural Service Gain Report: Oilseeds and Products Annual 2016. USDA FAS; 2016 Apr. Report No.: AU1603.
84. Figueroa LL, Blinder M, Grincavitch C, Jelinek A, Mann EK, Merva LA, et al. Bee pathogen transmission dynamics: deposition, persistence and acquisition on flowers. Proc R Soc B. 2019;286(1903):20190603.
85. Manson JS, Otterstatter MC, Thomson JD. Consumption of a nectar alkaloid reduces pathogen load in bumble bees. Oecologia. 2010 Jan 1;162(1):81–9.
86. Richardson LL, Adler LS, Leonard AS, Andicoechea J, Regan KH, Anthony WE, et al. Secondary metabolites in floral nectar reduce parasite infections in bumblebees. Proc R Soc B. 2015 Mar 22;282(1803):20142471.
87. Kearns CA, Inouye DW. Techniques for pollination biologists. Niwot, CO: University Press of Colorado; 1993.
88. Schmid-Hempel P, Schmid-Hempel R. Transmission of a pathogen in *Bombus terrestris*, with a note on division of labour in social insects. Behav Ecol Sociobiol. 1993 Nov 1;33(5):319–27.
89. LoCascio GM, Pasquale R, Amponsah E, Irwin RE, Adler LS. Effect of timing and exposure of sunflower pollen on a common gut pathogen of bumble bees. Ecol Entomol. 2019;44(5):702–10.
90. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2020. Available from: https://www.R-project.org/
91. Lenth R. emmeans: Estimated Marginal Means, aka Least-Squares Means. [Internet]. 2020. Available from: https://CRAN.R-project.org/package=emmeans
92. Nooten SS, Rehan SM. Historical changes in bumble bee body size and range shift of declining species. Biodivers Conserv. 2020;29(2):451–67.
93. Venables WN, Ripley BD. Modern Applied Statistics with S [Internet]. Fourth. New York: Springer; 2002. Available from: http://www.stats.ox.ac.uk/pub/MASS4/
94. Pulpitel T, Pernice M, Simpson SJ, Ponton F. Tissue-specific immune gene expression in the migratory locust, Locusta Migratoria. Insects. 2015;6(2):368–80.
95. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods. 2015;12(4):357–60.

96. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30(7):923–30.
97. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):1–21.
98. Clarke R, Ressom HW, Wang A, Xuan J, Liu MC, Gehan EA, et al. The properties of high-dimensional data spaces: implications for exploring gene and protein expression data. Nat Rev Cancer. 2008;8(1):37–49.
99. Bzdok D, Altman N, Krzywinski M. Points of significance: statistics versus machine learning. Nature Publishing Group; 2018.
100. Chapman RW, Reading BJ, Sullivan CV. Ovary transcriptome profiling via artificial intelligence reveals a transcriptomic fingerprint predicting egg quality in striped bass, Morone saxatilis. PLoS One. 2014;9(5):e96818.
101. Eibe F, Mark A. Hall, Ian H. Witten. The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques." Fourth Edition. Morgan Kaufmann; 2016.
102. Shannon C. The lattice theory of information. Trans IRE Prof Group Inf Theory. 1953;1(1):105–7.
103. Platt J. Sequential minimal optimization: A fast algorithm for training support vector machines. 1998;
104. Chen L, Zhang Y-H, Huang T, Cai Y-D. Gene expression profiling gut microbiota in different races of humans. Sci Rep. 2016;6(1):1–11.
105. Schilling J, Nepomuceno A, Schaff JE, Muddiman DC, Daniels HV, Reading BJ. Compartment proteomics analysis of white perch (Morone americana) ovary using support vector machines. J Proteome Res. 2014;13(3):1515–26.
106. Schilling J, Nepomuceno AI, Planchart A, Yoder JA, Kelly RM, Muddiman DC, et al. Machine learning reveals sex-specific 17β-estradiol-responsive expression patterns in white perch (Morone americana) plasma proteins. Proteomics. 2015;15(15):2678–90.
107. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841–2.
108. Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, et al. The GeneCards suite: from gene data mining to disease genome sequence analyses. Curr Protoc Bioinforma. 2016;54(1):1–30.
109. Consortium U. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res. 2019;47(D1):D506–15.

Figures

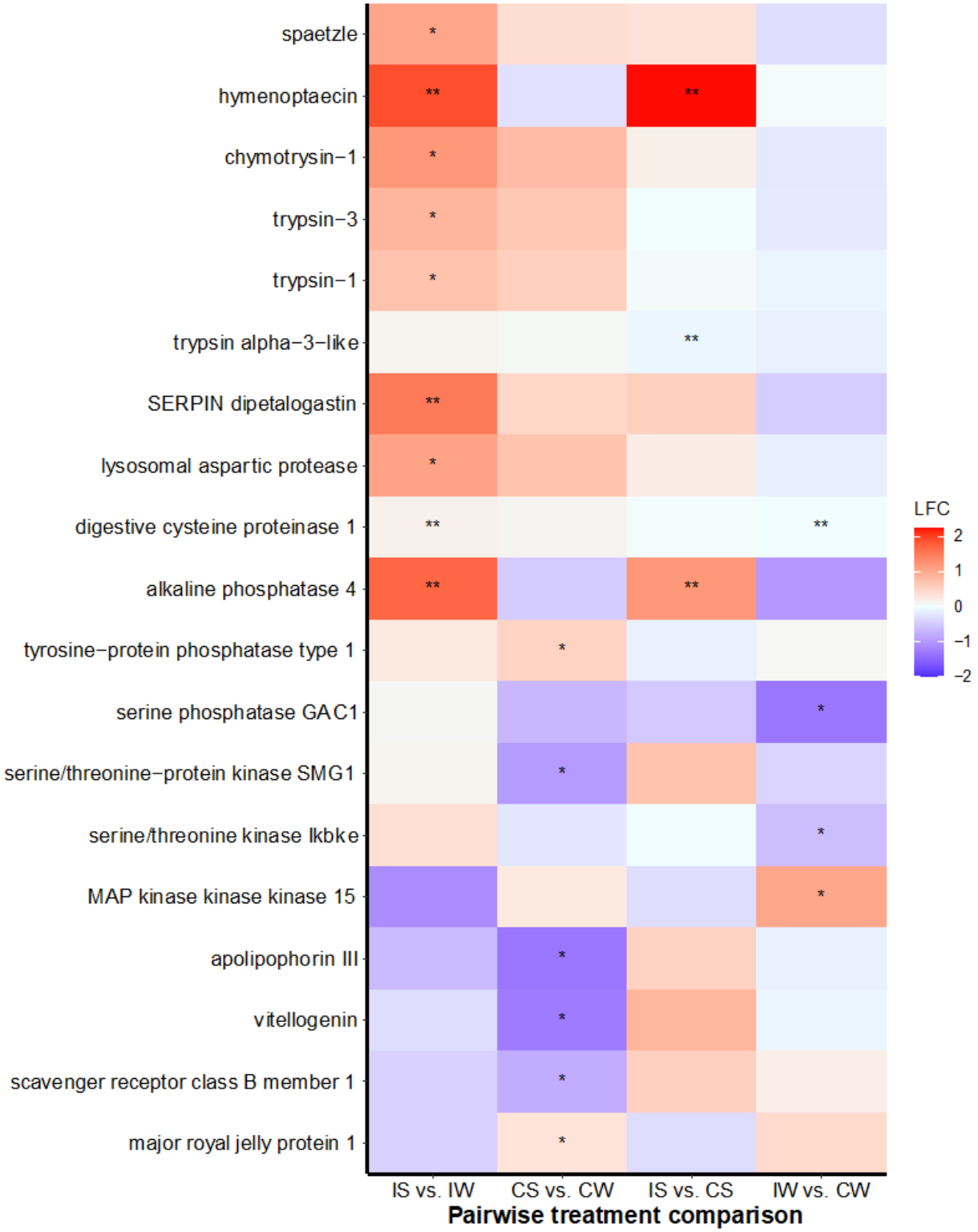

Figure 1

Gene expression profiles of putative immune transcripts in Bombus impatiens workers inoculated with the gut protozoan parasite Crithidia bombi (I) or inoculated with a sham control (C) and fed either sunflower (S) or wildflower (W) pollen. Colors indicate shrunken log fold changes estimated using a negative binomial model. Double asterisks indicate differentially expressed transcripts based on negative

binomial DESeq2 model and FDR < 0.05. Single asterisk indicates important differentially expressed transcripts based on machine learning analysis.

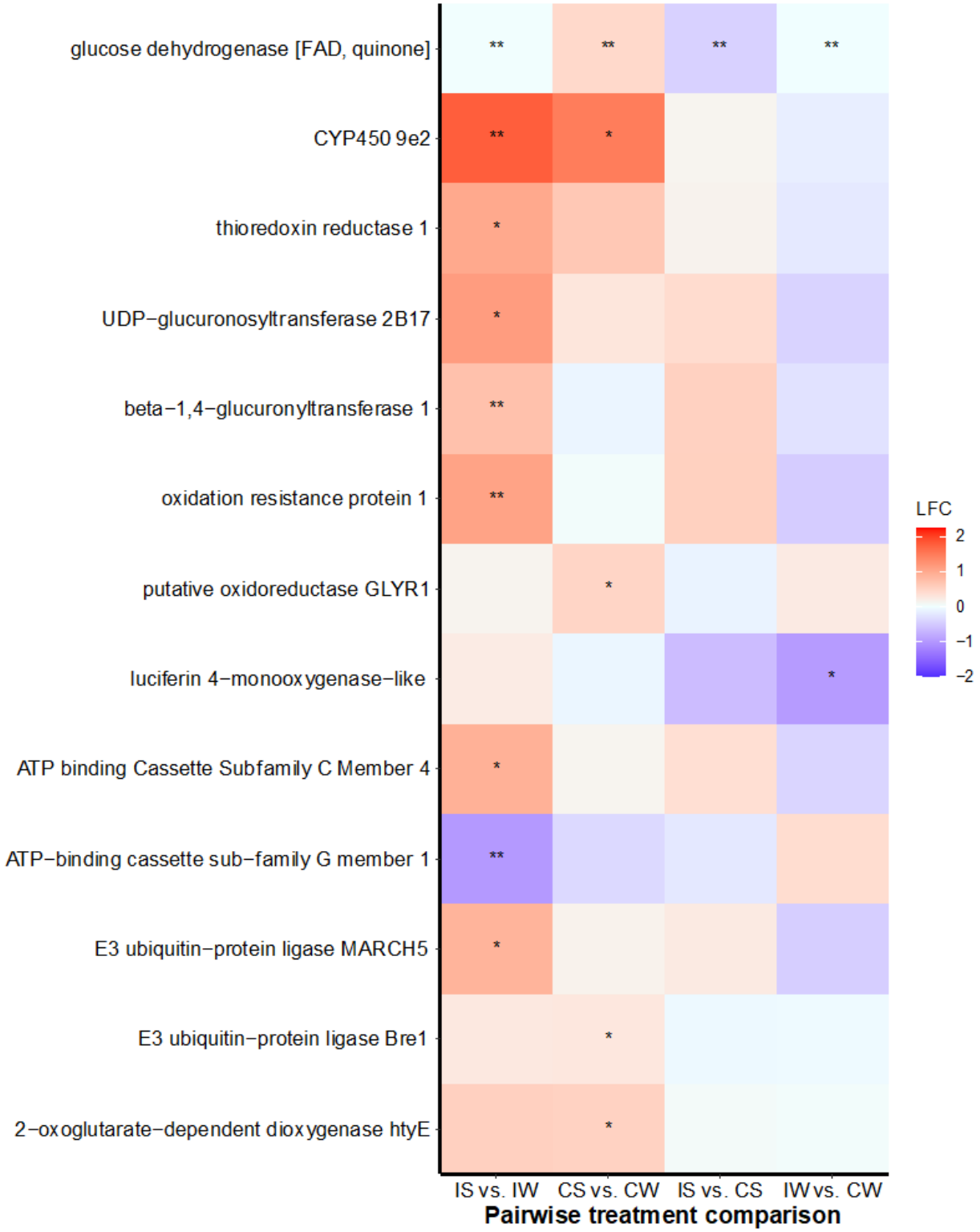

Figure 2

Gene expression profiles of putative detoxification transcripts in Bombus impatiens workers inoculated with the gut protozoan parasite Crithidia bombi (I) or inoculated with a sham control (C) and fed either sunflower (S) or wildflower (W) pollen. Colors and symbols as in Figure 1.

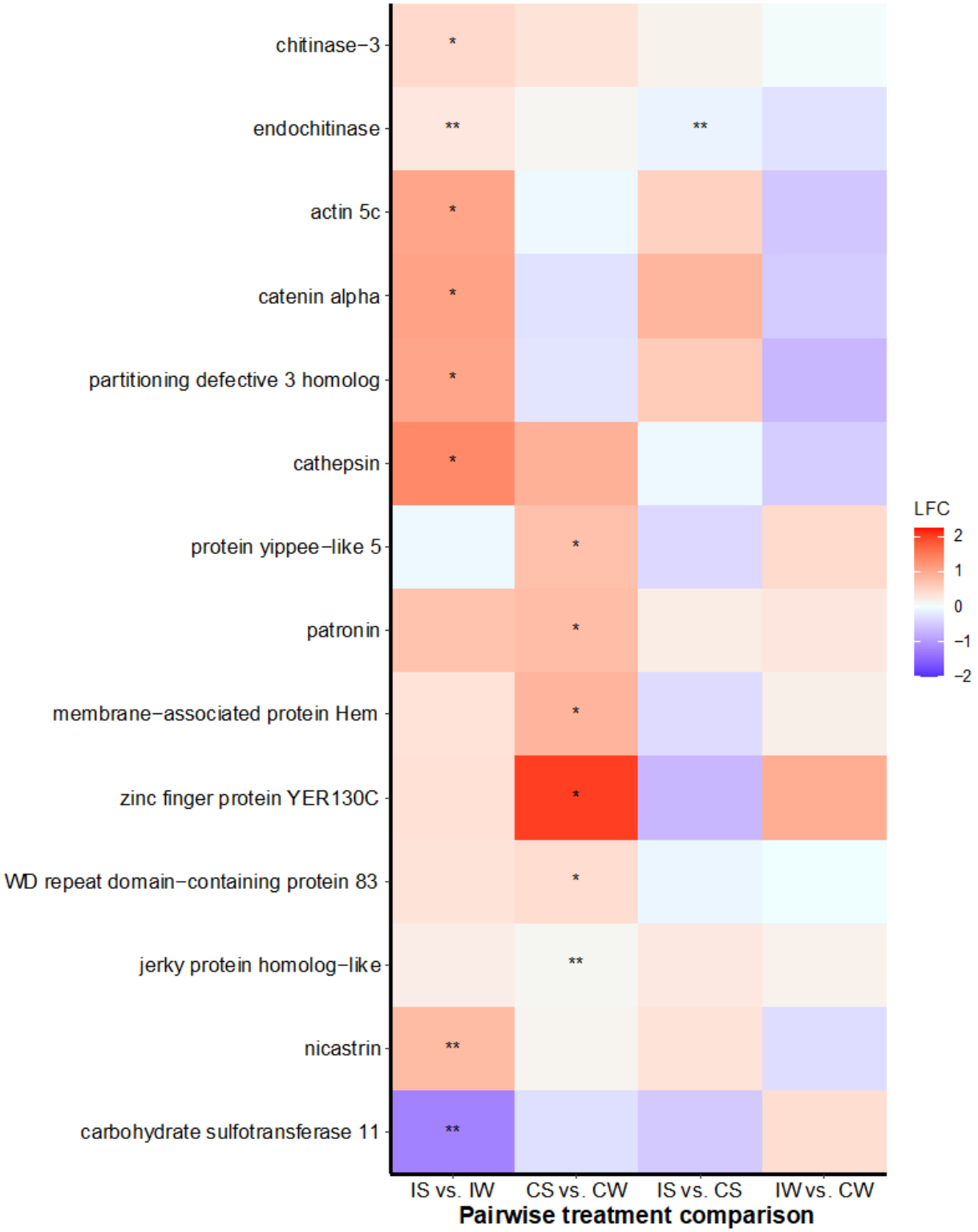

Figure 3

Gene expression profiles of gut morphology transcripts in Bombus impatiens workers inoculated with the gut protozoan parasite Crithidia bombi (I) or inoculated with a sham control (C) and fed either sunflower (S) or wildflower (W) pollen. Colors and symbols as in Figure 1.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- GiacominietalSupportingInformationBMCGenomics.docx