# Molecular mechanisms of mutualistic and antagonistic interactions in a plant–pollinator association

by Wang, R., Yang, Y., Jing, Y., Segar, S.T., Zhang, Y., Wang, G., Chen, J., Liu, Q.F., Chen, S., Chen, Y. and Cruaud, A.




Wang, R., Yang, Y., Jing, Y., Segar, S.T., Zhang, Y., Wang, G., Chen, J., Liu, Q.F., Chen, S., Chen, Y. and Cruaud, A. 2021. Molecular mechanisms of mutualistic and antagonistic interactions in a plant–pollinator association. *Nature Ecology & Evolution*, pp.1-13.

17 May 2021

# Molecular mechanisms of mutualistic and antagonistic interactions in a plant-pollinator association

Rong Wang[1,20†], Yang Yang[1†], Yi Jing[2†], Simon T. Segar[3†], Yu Zhang[1], Gang Wang[4], Jin Chen[4], Qing-Feng Liu[2], Shan Chen[1], Yan Chen[5], Astrid Cruaud[6], Yuan-Yuan Ding[1], Derek Dunn[7], Qiang Gao[2], Philip M. Gilmartin[8,9], Kai Jiang[1], Finn Kjellberg[10], Hong-Qing Li[11], Yuan-Yuan Li[1], Jian-Quan Liu[12], Min Liu[13], Carlos A. Machado[14], Ray Ming[15], Jean-Yves Rasplus[6], Xin Tong[1], Ping Wen[4], Huan-Ming Yang[2], Jing-Jun Yang[1], Ye Yin[2], Xing-Tan Zhang[16], Yuan-Ye Zhang[17], Hui Yu[18,19*], Zhen Yue[2*], Stephen G. Compton[9*], Xiao-Yong Chen[1,20*]

[1] Zhejiang Tiantong Forest Ecosystem National Observation and Research Station, School of Ecological and Environmental Sciences, East China Normal University, Shanghai 200241, China

[2]BGI Genomics, BGI-Shenzhen, Shenzhen 518083, China

[3]Agriculture & Environment Department, Harper Adams University, Newport, TF10 8NB, United Kingdom

[4]CAS Key Laboratory of Tropical Forest Ecology, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Mengla, Yunnan Province 666303, China

[5]Ecological Security and Protection Key Laboratory of Sichuan Province, Mianyang Normal University, Mianyang 621000, China

[6]INRA, UMR1062 CBGP, F-34988 Montferrier-sur-Lez, France

[7]College of Life Sciences, Northwest University, Xi'an, Shaanxi 710069, China

[8] School of Biology, University of Hull, Hull HU6 7RX, UK

[9]School of Biology, University of Leeds, Leeds LS2 9JT, UK

[10]CEFE, CNRS, Univ Montpellier, Univ Paul Valéry Montpellier, EPHE, IRD, France

[11]School of Life Sciences, East China Normal University, Shanghai 200241, China

[12]Key Laboratory of Bio-resource and Eco-environment of Ministry of Education, College of Life Sciences, Sichuan University, Chengdu 610065, China.

[13]School of Life Sciences, Guangzhou University, Guangzhou 510006, China

[14]Department of Biology, University of Maryland, College Park, MD 20742, The United States of America

[15]FAFU and UIUC-SIB Joint Center for Genomics and Biotechnology, Fujian Agriculture and Forestry University, Fuzhou 350002, China; Department of Plant Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA.

[16]Center for Genomics and Biotechnology, Fujian Provincial Key Laboratory of Haixia Applied Plant Systems Biology, Key Laboratory of Genetics, Breeding and Multiple Utilization of Corps, Ministry of Education, Fujian Agriculture and Forestry University, Fuzhou 350002, China.

[17]Key Laboratory of the Ministry of Education for Coastal and Wetland Ecosystems, College of the Environment and Ecology, Xiamen University, Xiamen, Fujian 361102, China

[18]Key Laboratory of Plant Resource Conservation and Sustainable Utilization, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou 510650, China

[19]School of Life Sciences, Qufu Normal University, Qufu Shandong, 273165, China

[20]Shanghai Institute of Pollution Control and Ecological Security, Shanghai 200092, China

† These authors contributed equally to this work.

* Co-corresponding authors.

E-mail: xychen@des.ecnu.edu.cn (X.-Y.C.); yuhui@scib.ac.cn (H.Y.);

S.G.A.Compton@leeds.ac.uk (S.G.C.); yuezhen@bgi.com (Z.Y.)

**Abstract**

Many insects metamorphose from antagonistic larvae into mutualistic adult pollinators, with reciprocal adaptation leading to specialized insect-plant associations. It remains unknown how such interactions are established at molecular level. Here we assembled high-quality genomes of a fig species, *Ficus pumila* var. *pumila*, and its specific pollinating wasp, *Wiebesia pumilae*. We combined multi-omics with validation experiments to reveal molecular mechanisms underlying this specialized interaction. In the plant, we identified the specific compound attracting pollinators and validated the function of several key genes regulating its biosynthesis. In the pollinator, we found a highly reduced number of odorant-binding protein (OBP) genes and an OBP mainly binding the attractant. During antagonistic interaction, we found similar chemical profiles and turnovers throughout the development of galled ovules and seeds, and a significant contraction of detoxification-related gene families in the pollinator. Our study detects some key genes bridging coevolved mutualists, establishing expectations for more diffuse insect-pollinator systems.

**Keywords**

Multi-omics, plant-pollinator mutualism, insect-host identification, pollinator adaptation to host plant, gene for gene coadaptation

## Introduction

Evolutionary adaptation fuels the genetic diversification of living organisms, driving speciation and emergent biodiversity[1,2]. However, in contrast to adaptation to abiotic conditions[3-5], it remains unclear how species adapt to reciprocally evolving biotic factors at the molecular level. This reflects the difficulty of identifying the traits linking interspecific interactions in a dynamic selective landscape. The high diversity of phytophagous insects and angiosperms is believed to be the result of coevolution, in part driven by ongoing insect-plant arms races[6,7]. Many herbivorous insects are also responsible for mediating gene flow between plant populations, often occurring as both antagonistic (i.e., herbivorous) larvae and mutualistic pollinating adults[8]. Selection by multiple agents associated with herbivorous/pollinating insects acts on floral traits to both deter herbivores and attract pollinators[9], making it difficult to separate mechanistic processes in many plant-pollinator systems. Tightly co-evolved species often have more apparent interacting traits, which provide an excellent testing ground for exploring coadaptation.

The obligate mutualisms comprising ~800 species from the genus *Ficus* (Moraceae) and their host specific pollinating wasps (fig wasps; Hymenoptera, Agaonidae) form a classical example of coevolution and contribute greatly to ecosystem functioning, biodiversity and agriculture[10,11]. Both mutualists have evolved strict correspondence in morphological, metabolic and life history traits[10,12]. The plants reward the larvae of pollinating wasps with nutrition and protection, and each mutualist wasp species is both pollinator and herbivore[12]. Each individual wasp

96 spends the majority of its lifespan at the larval stage (from three weeks up to nine
97 months) and develops inside a single galled ovule of a female floret located inside the
98 enclosed inflorescences characteristic of the genus (figs or 'syconia')[13-15] (Fig. 1a).
99 There are two predominant types of breeding systems in *Ficus* species, monoecy and
100 dioecy[16]. In monoecious figs, each fig produces female florets that can be either
101 pollinated or galled by pollinator larvae. In dioecious species, only the female florets
102 (feeder florets) in figs of functional male trees support the development of pollinator
103 offspring; figs growing on female trees attract pollinators to fertilize the female florets
104 (seed florets) that do not support wasp development (Fig. 1a). Upon locating host figs,
105 adult female wasps must crawl through a narrow passage usually lined by bracts (the
106 ostiole), into a dark central lumen where they typically remain trapped following
107 oviposition and/or pollination. Short lived (usually shorter than three days) adult
108 wasps do not feed[10].

109 Central to mediating these species-specific interactions are plant-emitted volatile
110 organic compounds (VOCs), which guide adult female wasps to precisely identify and
111 locate host figs[13,14,17-21]. Moreover, the high-quality genomes of a *Ficus* species and its
112 pollinating wasp (*Ficus microcarpa* and *Eupristina verticillata*) have been recently
113 reported[11], which create a basis for exploring how these pollinators identify host figs
114 at the molecular level. However, to date the key attractive VOCs have only been
115 explicitly identified in a small number of fig-pollinator mutualisms[17,18], and the
116 underlying molecular mechanisms determining host-specific signaling and insect
117 attraction remain unknown.

Once the problem of host identification has been overcome, pollinator larvae must also survive and develop under a set of unique conditions inside galled ovules that support their development (Fig. 1a). While figs can defend against herbivores from a wide range of taxa[22], it is unclear how pollinator larvae cope with plant defensive chemicals inside the galled ovules during this antagonistic phase of mutualism. One possible explanation is that galling behavior may activate the plant reproductive program in galled tissues, so that galling insects can avoid the strong chemical defenses induced by stress reaction when they utilize plant nutrients[23-25]. To test whether the reproductive program is activated in galled ovules, it is necessary to compare between the chemical profiles of galled ovules and seeds. We also expect that such adaptation to a specialized environment must leave molecular footprints in pollinator genome, for example contracted detoxification-related gene families[26,27].

Here we focused on a fig-pollinator mutualism comprising a dioecious *Ficus* species *Ficus pumila* var. *pumila*[28] and its specific pollinator *Wiebesia pumilae*[29] (Fig. 1a). We used multi-omics in combination with validation experiments to unravel the key molecular mechanisms contributing to the antagonistic and the mutualistic interactions in this system. We determined the specific attractive VOC and several key genes relevant to its biosynthesis. We identified the corresponding responses in the odorant-binding protein (OBP) genes in the pollinator genome and an OBP mainly binding the attractant. During the antagonistic phase, we found a similar turnover of secondary metabolites when female florets developed into either galled ovules or seeds, and almost identical chemical profiles between these two tissues. It implies that

140 the galled ovules may develop like seeds. A contraction of detoxification-related gene
141 families was found in the pollinator genome, providing insights into the fig-pollinator
142 coadaptation during antagonistic interaction.

143 **Results**

144 **Assembly of genomes and evolution**

145 To provide high-quality reference genomes for transcriptomic and proteomic
146 analyses, we assembled genomes of *F. pumila* var. *pumila* and *W. pumilae* using a
147 combination of Illumina and PacBio sequencing technologies (Supplementary Table 1;
148 see Methods). The assembled genomes were 315.7 Mb (contig N50 of 2.3 Mb) for the
149 plant and 318.2 Mb (contig N50 of 10.9 Mb) for the pollinator (Table 1 and
150 Supplementary Table 2). Using the uniquely mapped reads produced by the
151 high-throughput chromatin conformation capture (Hi-C) technique (Supplementary
152 Tables 1, 2), we generated Hi-C-based physical maps composed of 13 and 6
153 pseudo-chromosomes, with 96.6% (305 Mb) and 99.8% (318 Mb) of the assembled
154 genomes anchored to the pseudo-chromosomes (Table 1 and Supplementary Fig. 1).
155 The scaffold N50 of the assembled genomes were 22.4 Mb and 59.4 Mb, and the
156 pseudo-chromosomes included 97.1% (27,378) and 99.8% (12,292) of protein-coding
157 genes (Table 1). Genome annotation results showed that the structures and functions
158 of 25,905 and 12,305 protein-coding genes were annotated in the two genomes
159 (Supplementary Figs. 2 and Supplementary Tables 3-5). BUSCO quality analysis of
160 annotation showed that 92.4% of 1375 conserved plant genes and 91.3% of 4,415

161 Hymenoptera genes have complete coverage (Supplementary Table 3).

162 The protein-coding genes of *F. pumila* var. *pumila* and *W. pumilae* were clustered
163 into 15,631 and 7,969 gene families (Supplementary Fig. 3 and Supplementary Table
164 6). Analysis of comparative genomics using the genomes of 13 Angiosperm species
165 including *F. pumila* var. *pumila* and three congeneric species (*Ficus hispida*[11], *Ficus*
166 *microcarpa*[11] and *Ficus carica*[16]) showed that in the common ancestors of the four
167 *Ficus* species, 1,473 gene families had contracted and 888 gene families had
168 expanded. Phylogenetic reconstruction revealed that *F. hispida* is more closely related
169 to *F. pumila var. pumila* than the other two *Ficus* species (Supplementary Fig. 4a). In
170 the analysis of comparative genomics using the genomes of 11 arthropod species
171 containing *W. pumilae* and two other pollinator wasp species (*Ceratosolen solmsi*[11]
172 and *Eupristina verticillata*[27]), we found 48 expanded and 1261 contracted gene
173 families in the common ancestors of three pollinating wasp species. We recovered a
174 group containing and *E. verticillata* and *W. pumilae* with *C. solmsi* as its sibling
175 (Supplementary Fig. 4b). There was no evidence for recent whole-genome duplication
176 in the plant, and only a few small segments (total length of 1.3 Mb) were found to be
177 duplicated in the pollinator genome (Supplementary Fig. 5).

178 **Attractive compound forming fig-pollinator identification**

179 At the receptive stage, figs release VOCs containing critical compound(s)
180 attracting their pollinating wasps (Fig. 1a). To determine the attractive compound(s),
181 we collected VOCs from functional male and female figs of *F. pumila* var. *pumila* at

the pre-repetitive and the receptive stages using the dynamic headspace sampling (DHS) approach, and identified a total of 70 compounds (Fig. 1a and Supplementary Tables 7, 8). Only three (linalool, nonanal and decanal) of these compounds were found to elicit physiological responses of adult females of *W. pumilae* (Fig. 1b, c), of which only decanal was emitted exclusively at the receptive stage (Supplementary Table 8). We then conducted behavioral preference tests among the three compounds using 50 female pollinating wasps in each testing group. The wasps showed a significantly greater preference for decanal than the control and a significantly reduced preference for nonanal than the control, with a similar preference between decanal and a nonanal-decanal blend (Fig. 1d and Supplementary Table 9). These results demonstrate that the VOC compound decanal, emitted by *F. pumila* var. *pumila* figs at the receptive stage, functions to attract the pollinating wasp *W. pumilae*.

**Molecular mechanisms of specific host identification**

To identify the molecular mechanisms underlying the behavioral responses of *W. pumilae* to the VOCs emitted by its host figs, we annotated the four gene families involved in insect olfaction[30]. Across these gene families, *W. pumilae*, *E. verticillata* and *C. solmsi* consistently have lower numbers of genes, and, in particular, the number of odorant-binding protein (OBP) genes is significantly lower than less host-specific insects (Fig. 2a). Phylogenetic and synteny analysis including genomes of the three pollinating wasp species and the distantly related *Nasonia vitripennis* showed that most OBP genes in the pollinating wasp species displayed strong

homology and that the small number of OBP genes resulted from gene loss and infrequent tandem duplication (Supplementary Figs. 6a, 7a). There were apparent differences in motif structure among OBPs in six of the ten syntenic blocks (Supplementary Fig. 6b). The general contraction in OBP genes and frequent changes in motif structure of homologous OBPs among pollinating wasp species may be expected given their high host specificity and different VOC cues used for detecting host figs.

Among the 12 OBP genes of *W. pumilae*, transcriptome and proteome evidence showed that all genes were transcribed but only seven are translated into detectable proteins in adult females (Fig. 2b, Supplementary Fig. 8 and Supplementary Table 7). There were no proteins with significant differences in quantity (PSDs) and differentially expressed genes (DEGs) between the control and the VOCs-contacting treatment (Supplementary Table 10).

To explore functions of *W. pumilae* OBPs, we predicted motif structures of OBPs and compared them with the OBPs in *Adelphocoris lineolatus*[31] and *Culex quinquefasciatus*[32], known to have decanal or nonanal binding activity. Among the seven OBPs with detectable protein products, WpumOBP2 shows similar structure to the known decanal-binding protein and WpumOBP11 is similar to the known nonanal-binding protein (Fig. 2c and Supplementary Fig. 9). To validate the functions of WpumOBP2 and WpumOBP11, we produced the recombinant proteins for these two OBPs and measured their binding affinity to decanal and nonanal using surface plasmon resonance (SPR) experiments. Consistent with the prediction, the

experiments revealed considerably lower $K_D$ (representing much higher binding affinity) of WpumOPB2 to decanal than to nonanal and far lower $K_D$ of WpumOBP11 to nonanal than to decanal, and thus demonstrate the high binding affinity of these two OBPs to the corresponding compounds (Fig 2d, Supplementary Fig. 10 and Supplementary Table 11). Therefore, these results provide solid evidence that WpumOBP2 is the main binding protein to the attractant, and pollination of *F. pumila var. pumila* by *W. pumilae* is initiated by the binding of decanal with WpumOPB2.

**Regulation of gene expression in attractant biosynthesis**

To identify the tissue for attractant emission within figs, we measured the concentration of decanal emitted by ostiolar tissues and female florets at the receptive stage from both sexes of *F. pumila* var. *pumila* (Fig. 1a) using DHS, as previous studies from other species suggested VOCs are mainly released from these tissues[13,20]. The concentration of collected decanal in ostiolar tissues was 3.13 ± 1.11 pg/g, which was 9.1 times as that in female florets (0.34 ± 0.05 pg/g) (Pairwise T Test: df=9, t=6.02, p=0.002). Thus, the results revealed that decanal was predominantly emitted by ostiolar tissues at a similar concentration between sexes (T Test: df=4, t=0.20, p=0.858 in ostiolar tissues; df=4, t=0.24, p=0.826 in female florets).

To identify key genes involved in the biosynthesis of decanal, we conducted transcriptome and proteome analysis on ostiolar tissues collected at the pre-receptive and the receptive stages (Supplementary Table 7). The biosynthesis of decanal and nonanal is involved in the pathways of fatty acid biosynthesis (ko00061), elongation

(ko00062) and metabolism (ko00071 and ko00592) (Fig. 3a)[33,34]. Genes in these pathways showed similar patterns of expression between transcriptome and proteome data (Supplementary Fig. 11 and Supplementary Table 12). Comparing the receptive with the pre-receptive stage, we detected a total of eight PSDs (Fig. 3b), likely facilitating the biosynthesis of decanal and suppressing the biosynthesis of nonanal at the receptive stage (Fig. 3a). Down-regulated PSDs included two ACSLs (long-chain acyl-CoA synthetase) and one HACD (very-long-chain (3R)-3-hydroxyacyl-CoA dehydratase), while up-regulated PSDs comprised an ALDH (acetaldehyde dehydrogenase), an ADH (alcohol dehydrogenase), two LOX2Ss (lipoxygenase) and one HPL (hydroperoxide lyase) (Fig. 3b). To validate the function of key genes (the two ACSLs, the ALDH and the ADH) in decanal biosynthesis, we produced the recombinant proteins of these genes and conducted *in vitro* enzyme activity assay (see Methods). The final products of the *in vitro* reactions identified by LC–MS or GC–MS are consistent with the standards (Fig. 3 c-e). These results validate the enzyme activity of the two ASCLs in synthesizing hexadecanoyl-CoA as well as the ALDH and the ADH in synthesizing decanal and decanol.

To understand the transcriptional regulation of decanal biosynthesis, we conducted co-expression network analysis and found one module containing two key genes (*FpumACSL10* and *FpumALDH1*) and four potential regulating transcription factors (two *HD-ZIP*s, one *bHLH* and one *bZIP*) (Fig. 3f and Supplementary Table 13). Cis-element detection analysis revealed one G-box motif upstream of *FpumACSL10* and six G-box and one HD-Zip motifs upstream of *FpumALDH1*

(Supplementary Table 14). As G-box binds to transcription factor families of bZIPs and bHLHs and HD-Zip binds to HD-ZIPs[35,36], we hypothesized that expression of *FpumACSL10* is regulated by the bHLH and the bZIP and all above four transcription factors regulate the expression of *FpumALDH1*. To test this hypothesis, we obtained qualified polyclonal antibodies for the four transcription factors and performed ChIP-qPCR experiments. High % input and fold enrichment values showed that the bHLH and the bZIP could bind to the promoter region of *FpumACSL10* and all the four transcription factors could bind to the promoter region of *FpumALDH1* (Fig. 3g, h), providing evidence for our hypothesis.

**Metabolic and genomic signature of antagonistic interaction**

To understand the mechanisms of antagonistic interaction between figs of *F. pumila* var. *pumila* and larvae of *W. pumilae*, we analyzed chemical profiles of different tissue types of female and functional male figs at the receptive and the post-receptive stage using metabolomic data (Supplementary Table 7; see Methods). We focused on the secondary metabolites associated with plant chemical defenses (SMCDs)[22,37-39], comprising some terpenoids (triterpenes and sesquiterpenes) and phenylpropanoids (including their precursors and their derivatives) (Supplementary Fig. 12). Metabolomic analysis revealed 736 SMCDs (108 terpenoids and 628 phenylpropanoids) (Supplementary Table 15). While we found significant differences in chemical profiles between two types of tissues and between different fig development stages, there were few differences between female and functional male

292 figs (Fig. 4a). No secondary metabolites with significant difference in quantity
293 (SMSDs) were found between feeder and seed florets at the receptive stage, and there
294 were only three SMSDs between galled ovules and seeds at the post-receptive stage
295 (Fig. 4b). Remarkably, we found similar changes of SMSDs in both the feeder
296 floret-galled ovule and the seed floret-seed transitions (Fig. 4c and Supplementary Fig.
297 13a). Besides SMCDs, galled ovules and seeds shared similar overall chemical
298 profiles (Supplementary Fig. 14). These results showed similar chemical changes and
299 profiles in the development of female florets no matter they were parasitized by
300 pollinator larvae (becoming galled ovules) or not (developing into seeds).
301 As might be expected from organisms that spend most their lives in a specific
302 environment, contraction of three gene families crucial to the detoxification of plant
303 defensive chemicals[40] (CYP450s, glutathione s-transferases (GSTs) and
304 carboxylesterases (CCEs) gene families) was found in the genomes of *W. pumilae*, *E.*
305 *verticillata* and *C. solmsi* (Fig. 4d and Supplementary Fig. 15). Such contraction was
306 mainly caused by gene loss and infrequent tandem duplication (Supplementary Fig.
307 7b-d), and most of the detoxification-related genes in the three pollinating wasp
308 species were in the same monophyletic groups (Supplementary Fig. 16). Ten out of
309 the 56 detoxification-related genes in *W. pumilae* was at a high level (read counts >
310 200) and was significantly upregulated at the larval stage compared to the adult stage
311 (Supplementary Fig. 17 and Supplementary Table 16). These metabolic and genomic
312 signatures provide a molecular basis for further exploring the mechanisms of
313 fig-pollinator coadaptation during their antagonistic interaction.

## Discussion

Reciprocal selection on signaling and defense traits has shaped the molecular constraints governing how antagonistic larvae develop into mutualistic adult pollinators[8,41]. In this study, our novel combination of classic electrophysiological experiments and multi-omics approaches has illuminated some key mechanisms forming the coadaptation in a pair of fig-pollinator mutualists. We identified the attractive VOC, detected that host identification by the specific pollinators may be linked to their reduced number of OBP genes, and validated an OPB mainly binding the attractant. We identified the key genes involved in the regulation of both attractant and repellent biosynthesis in the plant: from facilitating the synthesis of the repellent to favoring the production of the attractant. Surprisingly, matched changes in SMCDs occurred across the transitions from i) floret to galled ovule and ii) floret to seed, and almost identical profiles of SMCDs were found in galled ovules and seeds. As for the pollinator, we detected a contraction of detoxification-related gene families.

Previous studies have mainly focused on the dominant components or the bouquet of components in the VOCs emitted by figs[14,17]. In contrast our results showed that only one VOC of relative low concentration attracts the focal pollinating wasp species, addressing the importance of detailing the complete spectrum of VOCs. Moreover, the attractive VOC (an aldehyde) in our focal species is distinct from the attractants found in other *Ficus* species (usually terpenes)[11,13,14,17,18]. Such a dramatic difference indicates deep divergence among congeners in the recognition of VOC attractants[42,43], providing the basis for adaptive radiations in both *Ficus* and their

pollinating wasps[44]. In addition, similar concentration of the attractant emitted by different sexes of figs supports the intersexual mimicry hypothesis in *Ficus* species[21], which argues that any changes in biosynthesis of attractant VOCs in female figs may cause loss of sexual reproduction[21].

Similar chemical changes in the development of galled ovules and seeds and almost identical SMCD profiles in these tissues showed that the occupancy of pollinator larvae activates the reproductive program of galled ovules. This suggests that galling strategy may help pollinator larvae avoid the potential chemical sanctions when they exploit nutrients of host plants. This is likely to result from either pollinator larvae manipulating plant physiology or changes triggered by host figs once the feeder florets are galled. Chemical mimicry of fruits and seeds has been reported in other galling insects[24,25], while many studies also suggest that figs have evolved to accommodate pollinator larvae[10,15]. Other possibilities, such as pollination before oviposition combined with minimal initial interference of pollinator larvae can be largely excluded, because most galled ovules were not pollinated (Wang R., personal observations). Furthermore, we collected figs at the middle (four weeks after the entrance of pollinator foundresses) of the post-receptive stage (generally lasting 8-10 weeks). Future research should perform bioassays to determine the chemicals inducing the development of galled ovules and the specific secondary metabolites defending against pollinator larvae. This will reveal how pollinator larvae activate the reproductive programs of host plants and why they can only utilize feeder florets.

The pollinating wasp species have evolved specializations in OBP and

detoxification-related genes, probably because they are host specific and spend most of their lives inside galled ovules (though some detoxification-related genes are not only involved in detoxification but also important for the general life cycle of insets). Such specializations facilitate the maintenance of host-specificity, but conservation of some OBP genes among pollinating wasps (Supplementary Figs. 6, 7) and their fast rates of evolution due to short generation time also offer opportunities for host shift, which is considered as a frequently occurred event throughout the evolutionary history of fig-pollinator mutualism[19,45,46]. Moreover, selection to maximize pollinator fitness may drive rapid adaptive changes in fig traits like floral scents, and such reciprocal selection has occurred in some generalized plant-pollinator systems[47].

Ongoing global changes are causing rapid evolution and phenotypic changes in many plants, leading to mismatches between key traits bridging plants and their pollinators[4,48]. Erosion of these links can result in the collapse of long-evolved mutualisms and a loss of biodiversity, but may also lead to the rewiring of host association networks[4,49,50]. Limitations to our knowledge of molecular determination in plant-pollinator interactions have made predictions about future changes in biodiversity and ecosystem functioning largely speculative. Our findings offer an example of gene for gene coadaptation that extends beyond the existing phenotype-based models of mutualism persistence[51] and place trait-based multi-omics at the center of the ecological and evolutionary research concerning interacting species in more diffuse systems.

## ACKNOWLEDGMENTS

We thank Yong-Jin Wang, Qi-Chong Zhu, Qiong Sun, Qian-Ya Li, Jing-Wen Wang, Tong-Lei Xu, Yue Chen, Fang-Lu Wei, Guo-Chun Shen, Xiu-Lian Wen and Li-Sha Li for their kind helps in field experiments, Xiu-Lian Wen, Li-Sha Li and the Public Technology Service Center of XTBG (CAS) for assisting active VOCs analysis, and Xiu-Guang Mao, Pan-Yu Hua and Da-Yong Zhang for constructive suggestions in data analysis. This work is supported by NSFC grants 31630008 and 31870356 (X.-Y.C.), and 31870359 (G.W.); and a Talents 1000 Fellowship of Shaanxi Province (D.D.). S.S. acknowledges departmental support from Harper Adams University.

## Author contributions

X.-Y.C. and R.W. conceived and designed the study. R.W., Y.Yang., S.G.C., S.S., H.Y. and Z.Y conducted the experiments and analyzed data. Y.J., Q.F.L., H.Y., Y.Z., G.W., J.C., R.M., S.C., Y.C., D.D., H.Q.L., M.L., Y.-Y.D., Y.-Y.L, X.T., P.W., J.J.Y., X.-T.Z., Q.G., J.-Y.Y., Y.Yin, K.J., and H.M.Y. contributed to data acquisition and data analyses. R.W., S.S., S.G.C., J.Q.L., J.-Y.R., F.K., C.A.M, A.C., P.M.G, Y.-Y.Z. and X.-Y.C. edited the manuscript. All authors contributed to writing the manuscript.

## Competing interests

The authors declare no competing interests.

## Data availability

The data that support the findings of this study have been deposited in the CNSA (https://db.cngb.org/cnsa/) of CNGBdb with accession code CNP0000674.

## Methods

**Genome assembly and annotation.** Genomic DNA was extracted from leaves of a female *F. pumila* var. *pumila* individual nearby Zhejiang Tiantong Forest Ecosystem National Observation and Research Station (TINAS) (E 121°47', N 29°48'), Ningbo, China, and from c. 500 adult female pollinators of *W. pumilae* emerged from five figs on a functional male tree in South China Botanic Garden (SCBG) (E 113°11', N 23°11'), Guangzhou, China. Six pair-end and mate-pair libraries were prepared with varying insert sizes (Supplementary Table 1) for sequencing on an Illumina Hiseq 4000 platform. We also carried out PacBio single-molecule real-time sequencing of 20kb SMRTbell libraries using a PacBio Sequel platform. Based on the Illumina pair-end sequencing data, the genome sizes of both species were estimated by counting k-mer frequency using Jellyfish version 2.1.3[52].

*De novo* genome assembly was conducted using MECAT version 1.2[53]. The initial contig was polished twice based on raw PacBio data and then corrected twice using Illumina paired-end reads with pilon version 1.22[54]. Redundans version 0.13c[55] was used to exclude redundant contigs, and we removed contaminative sequences by searching against the NCBI nucleotide sequences database (ftp://ftp.ncbi.nlm.nih.gov/blast/db/ FASTA/) using megablast[56] with e-value $\leq 1e^{-5}$. Gap filling was implemented with PBJelly[57] after scaffolding based on Illumina

mate-pair reads using BESST version 2.2.7[58].

To further improve the quality of genome assembly of both species, we used high-throughput chromatin conformation capture (Hi-C) technique to scaffold contigs into pseudo-chromosomes. We constructed Hi-C libraries using the protocol described by Belton et al.[59]. Fresh leaves sampled from the same *F. pumila* var. *pumila* individual used in above sequencing and adult female pollinators from SCBG were cross-linked by 4% formaldehyde solution, followed by an overnight digestion with a 4-cutter restriction enzyme MboI (400 units) at 37°C, preparation DNA ends with biotin-14-dCTP and blunt-end ligation of the cross-linked fragments. Then, the proximal chromatin DNA was re-ligated by ligation enzyme, and the nuclear complexes were reverse cross-linked by proteinase K. After that, we extracted and purified DNA and removed biotin from non-ligated fragment ends using T4 DNA polymerase. The following steps including end reparation, enrichment of biotin-labeled Hi-C samples, and ligation by Illumina paired-end (PE) sequencing adapters, and then the Hi-C library (insert size of 350 bp) was amplified by PCR and sequenced on an Illumina NovaSeq 6000 platform. High quality data checked by HiC-Pro[60] were mapped to genome using BWA, with extraction of uniquely mapped reads for pseudo-chromosome clustering and assembly using Juicer[61] and 3D-DNA[62].

Following genome assembly, we assessed completeness using BUSCO version 3.0.3[63] and Iso-Seq full-length transcripts. The high-quality full-length transcripts were mapped to genome assemblies using GMAP version 2014-12-21[64], setting a cutoff of aligned coverage at 0.85 and aligned identity at 0.9. The quality of genome

446 assembly was further tested by mapping Illumina paired-end reads to the genome
447 assemblies using BWA with the depth of coverage calculated using BamDeal version
448 0.19 (https://github.com/BGI-shenzhen/BamDeal/). For each species, Iso-Seq
449 sequencing was performed using a PacBio Sequel platform, based on two SMRTbell
450 libraries with insert sizes of 0 - 5kb and 4.5kb - 10kb established by full-length
451 complementary DNA (cDNA). We used fresh leaves, young stems from fertile and
452 sterile branchlets and figs at different developmental stages for the plant, and adult
453 males and females for the pollinator.

454 Genome annotation includes repeat identification (including tandem repeats (TRs)
455 and transposable elements (TEs)), annotation of non-coding RNAs (ncRNA) and gene
456 prediction and annotation. When annotating repeat sequences, TRs were identified
457 using Tandem Repeats Finder (TRF) version 4.07[65], and TEs were searched against
458 Repbase 21.01[66] and the transposable element protein database using RepeatMasker
459 version 4.0.6 (http://www.repeatmasker.org/) and RepeatProteinMask in
460 RepeatMasker. LTR_Finder[67], PILER[68] and RepeatScout[69] were used to create a *de*
461 *novo* TE library, and the combined non-redundant library was classified by running
462 RepeatMasker again.

463 To annotate ncRNAs, tRNAscan-SE version 1.3.1[70] was used to identify tRNA
464 and their secondary structures. While small nuclear RNA (snRNA) and microRNAs
465 (miRNAs) were searched for using INFERNAL version 1.1.1[71] in the Rfam database
466 version 12.0[72], followed by the detection of rRNAs by aligning with plant or
467 invertebrate rRNA sequences using BLASTN (E-value $\leq 1e^{-5}$).

Gene model prediction was conducted using the MAKER pipeline version 2.31.10[73]. The Iso-Seq full-length transcripts, RNA-seq transcripts (assembled using Hisat2 version 2.0.1[74] and StringTie version 1.3.3[75]), the protein sequences of related species and protein sequences from Swiss-Prot database (https://www.uniprot.org) were included in the analysis. Ab-initio gene prediction was performed with the gene predictors SNAP[76] and AUGUSTUS[77]. The MAKER pipeline was run for two (for the plant) and three (for the pollinator) iterations for training and the final trained hidden Markov model (HMM) was used for annotation. JBrowse version 1.12.3[78] was used to examine the gene models following each iteration. The gene models with the presence of a PFAM domain or with AED $\leq$ 0.6 for *W. pumilae* and AED $<$ 1 for *F. pumila* var. *pumila* were retained. BUSCO was used to evaluate the completeness of gene annotation for both genomes.

After determining gene models, functions of protein-coding genes were annotated using DIAMOND version 0.8.23[79] by aligning them to NCBI NR database, Swiss-Prot[80] and KEGG[81] databases. Motifs and domains in protein sequences were annotated using InterProScan version 5.16-55.0[82] via searching public databases. Gene Ontology terms for each gene were assigned using Blast2GO version 3.3[83].

**Comparative genomics.** To analyze the evolutionary characters of our studied genomes, we first carried out gene family clustering. The genome and annotation data of 13 other angiosperm species and 11 other arthropod species were downloaded (Supplementary Table 6). The gene models with open reading frames shorter than

90bp were removed, and only the longest transcript was chosen to represent each gene. Gene family clustering was performed using OrthoMCL version 10-148[84] for the plant and TreeFam pipeline version 0.5.1[85] for the pollinator.

We then determined the phylogenetic relationships among the plants and among the insects in the species pools used in gene family clustering. Corresponding coding sequences (CDSs) were aligned based on the protein sequences of all single-copy orthologs using MUSCLE version 3.8.31[86], and codon position 2 of aligned CDSs were concatenated into a super gene using an in-house Perl script with a filtration of ambiguously aligned positions using trimAI version 1.4.1[87]. After that, phylogenetic trees were reconstructed using PhyML version 3.0[88] using a GTR substitution model with a gamma distribution and 100 bootstrap replicates. PAML version 4.9[89] was used to estimate divergence time, setting 10,000 MCMC generations with a sampling frequency of 5,000 and a burn-in of 5,000,000 iterations. Overall substitution rate was assessed using BASEML setting a REV substitution model.

Gene family expansion and contraction was analyzed using CAFE version 2.1[90], which employed a stochastic birth-and-death process to model the evolution of gene family sizes over a phylogeny. The birth-and-death parameter ($\lambda$) was estimated using 10,000 Monte Carlo random samples. We then used family-wise method to statistically test if a gene family experienced significant expansion/contraction, and gene families with conditional P-values less than 0.05 were considered to have accelerated rates of gains and losses.

We then tested whether the genomes of *F. pumila* var. *pumila* experienced whole

genome duplication (WGD). Syntenic blocks were identified using MCscan version 0.8[91], and the rate of transversions on fourfold degenerate synonymous sites (4DTv) was calculated using the HKY substitution model to uncover potential speciation or WGD events occurring in evolutionary history of the plant. For *W. pumilae*, we tested for genomic segmental duplications (SDs). The self-alignment was performed using BLASTZ version 1.02[92], and a non-redundant set of SDs was obtained using WGAC version 1.3[93].

**Annotation of specific gene families and analysis of their evolution.** To test whether the contraction specific gene families in *W. pumilae*, *E. verticillata* and *C. solmsi* contributes to the wasps' host-specificity and detoxification ability, we conducted a detailed annotation in chemosensory gene families (OBPs, CSPs, ORs and IRs) and detoxification gene families including CYP450s, GSTs and CCEs. The homologous genes of *N. vitripennis*, *Apis mellifera* and *Drosophila melanogaster* were used as queries to search the genome assembly of *W. pumilae* using TBLASTN at a criterion of E-value $\leq 1e^{-5}$, and gene structures of identified genes were predicted using GeneWise version 2.4.1[94] with pseudogenes masked. We repeated this process iteratively until no more genes were detected. Additional genes from the MAKER annotation were also included if they included corresponding InterPro domains. All gene structures were manually checked and corrected if necessary, on the basis of full-length transcripts, RNA-seq transcripts and homologous proteins in JBrowse. We used Binomial Distribution One-tailed test to examine gene family

532 expansion/contraction among the compared species without considering their
533 evolutionary relationships.
534 To reveal the evolutionary history of OBP, CYP450, CCE and GST gene families,
535 syntenic blocks were identified across the genomes of the three pollinating wasp
536 species and *N. vitripennis* using MCscan (https://github.com/tanghaibao/jcvi/wiki
537 /MCscan-(Python-version)). We then constructed neighbor-joining phylogenetic trees
538 to verify homologous genes among these insect species, using TreeBeST version
539 1.9.2[95] using a JTT model and 1000 bootstraps.

540 **VOC collection and component analysis.** To reveal the composition of volatile
541 organic compounds (VOCs) emitted by figs of *F. pumila* var. *pumila* at different
542 developmental stages, we collected VOCs from figs at both pre-receptive and
543 receptive stages (Fig. 1a) in spring 2018 using dynamic headspace sampling (DHS)
544 techniques[96]. After a careful search, we chose ten mature *F. pumila* var. *pumila* trees
545 comprising five females and five functional males (Supplementary Table 8) nearby
546 TINAS, within the natural range of the plant. Three figs were labeled on each selected
547 individual. At either fig developmental stage (from early to middle April for
548 pre-receptive stage and late April for receptive stage), we extracted the VOCs emitted
549 by each labeled fig into an activated porapak adsorption tube (150 mg) during
550 8:00-11:00 am, using a protocol identical to Tholl et al. (2006)[96]. Each adsorption tube
551 was then eluted three times using 300 μl n-hexane and stored at -20 .
552 VOCs emitted by figs were then separated and identified using a coupled Gas

553 Chromatography-Mass Spectrometry (GC–MS) system (HP 7890A-5975C, Agilent,
554 US)[97]. For each sample, 1.8 μl of eluate VOC extract, concentrated using nitrogen,
555 was injected in split mode with a split ratio of 10 : 1 at 250 . Helium (1 mL/min)
556 was used as carrier gas in a HP-5ms (30 m × 250 μm × 0.25 μm, Agilent, US) GC
557 column. We set the oven ramp at 40°C for 1 min, and then 3 °C/min to 140 °C for 1
558 min, followed by 5 °C/min to 230 °C for 3 mins. Ionization was conducted by
559 electron impact (70 eV, source temperature 230 ). The MS quadrupole was heated to
560 150 °C, with the scanned mass range setting as from 40 to 550 m/z. Compound
561 identification was implemented by matching the mass spectra with NIST 08 MS
562 libraries. We then calculated the relative proportions of all compounds emitted by figs
563 at each developmental stage.
564 To evaluate the difference in the concentration of decanal between ostiolar tissues
565 and female florets, we sampled figs at receptive stages from three female and three
566 functional male individuals and identified the composition of VOCs emitted from
567 these two types of tissues using the same approach mentioned above. The decanal
568 concentration in each type of tissues in a plant individual was quantified by
569 comparing its peak area with the internal standard (decyl acetate).

570 **Electrophysiological responses of pollinating wasps.** To narrow the range of
571 candidate VOCs attracting *W. pumilae*, we tested the electrophysiological responses of
572 the pollinators to the collected VOCs, using gas chromatography-electroantennogram
573 detection (GC–EAD). We used a system coupling a custom-made EAG[98] with a GC

(Trace GC 2000, Themo Finnigan, US). We injected 1.8 μl of concentrated VOC extract eluate into the GC to separate different compounds. The GC conditions were identical to those used for the GC–MS component analysis, except that the oven ramp was set to 50 °C for 2 mins, and then to 10 °C/min up to 280 °C for 1 min. After GC–FID (flame ionization detector) quantification, outflow from the GC column was delivered to the EAG as the stimulus through a custom, 40 cm long heated (at 250 °C) transfer line with a clean, wet, and static-free airflow. The stimulus was then puffed to the antenna of an adult female pollinator (collected from figs in TINAS) fixed onto the EAG with both ends of the antennae connected with prepared glass electrodes linking the probes of EAG to the potentiometric amplifiers.

This experiment was repeated 5 times (i.e. antennae of 5 adult female pollinators), and the EAD signal was recorded using a HP 34465A digital multimeter (Keysight, US). Both EAD and FID signal data were aligned to verify the tentative compounds stimulating the adult female pollinator, using the software IO Libraries Suite 16 (Agilent, US) and BenchVue (Keysight, US). These tentative effective compounds were identified by matching the chromatographs with the results of component analysis using GC–MS.

We further tested the electrophysiological response of adult female pollinators (collected from figs in SCBG) to the synthesized standard of each tentative compound (TIC, JPN; TRC, CAN; Sigma-Aldrich, US), following the same procedures as above. A compound was determined as truly effective only when it was confirmed by the experiments using both eluate of VOC extracts and synthesized standard.

**Behavioral preference of pollinating wasps.** To test the behavioral preference of *W. pumilae* to different tentative effective VOCs, we used a Y-tube olfactometer (stem 8 cm, arms 9 cm, at an angle of 55°, internal diameter of 1.5 cm) following the methods described by Wang et al.[19]. We placed the synthesized standard of each tentative effective VOC in the glass container, connecting one arm of the olfactometer to this treatment of n-hexane and blends of putative stimuli compounds and the other arm to the controls (only n-hexane) (Supplementary Table 9). VOCs were passed from both arms to the stem through equal flow rates of cleaned and humidified airflow created by an air pump system with an activated charcoal filter and distilled water. To avoid visual distractions to the pollinators, we placed the olfactometer in the center of a white table illuminated using three 40-W cool white fluorescent tubes above the arms.

Each healthy adult female pollinator (collected from figs in SCBG) was tested independently with an observation for 5 mins in the olfactometer, and its behavior was assigned to one of the three choices: (1) towards the treatment (the insect went 1 cm past the Y junction (decision line) and stayed there more than 1 min); (2) towards the control; and (3) no choice (the insect did not reach the decision line within 5 mins). For each treatment-control combination, we repeated this experiment 50 times (i.e. 50 adult female pollinators) and compared the proportions of different choices (towards the treatment and towards the control) using GLMs assuming binomial distribution of residuals to examine the preference of *W. pumilae*.

**Sample collection for comparative transcriptome, proteome and metabolome.** To

617 reveal the molecular mechanisms forming the specific pollinator-host identification
618 based on both transcriptomic and proteomic data, in spring 2017 and 2018, after
619 collecting several pre-receptive and receptive figs from the ten mature individuals of
620 the plants used in VOCs collection (Supplementary Table 7), we dissected each
621 sampled fig to gather ostiolar tissues with bracts and female florets. The total sample
622 size therefore was 40 for the plant (a type of tissues collected from a single tree at
623 each fig developmental stage is one sample, with 20 samples for each type of tissues).
624 In spring 2018, we sampled at least 5 figs at the mature stage from each of the five
625 functional male mature individuals used for VOC collection (Supplementary Table 7).
626 Each sampled mature fig was dissected into halves in a Teflon bag, and then each half
627 was rapidly moved into a Teflon bag containing only clean air filtered by activated
628 charcoal (as a control) or clean air and a receptive fig (as a treatment), to test whether
629 differential expression occurred in some chemosensory genes when adult females
630 were exposed to attractive VOCs. We then collected all adult females of *W. pumilae*
631 emerging from the sampled figs according to the identity of functional male trees (a
632 total of 10 samples with at least 100 adult female wasps in each sample). All sampled
633 fig tissues and adult female pollinators were first stored in liquid nitrogen for 72 hours
634 and then moved into a refrigerator at -80 .

635 To unravel how pollinator larvae adapt to the environments inside galled ovules
636 using metabolomes, we sampled several receptive and post-receptive figs from the ten
637 plant individuals (Supplementary Table 7) and collected ostiolar tissues (20 samples),
638 female florets (10 samples), galled ovules (5 samples) and seeds (5 samples) in spring

2020. For clearly distinguishing galled ovules and seeds from the female florets that were neither pollinated nor utilized by pollinators, the post-receptive figs were sampled four weeks after the entrance of adult female pollinators.

**RNA-seq for *F. pumila* var. *pumila* and *W. pumilae*.** After generating PCR-based libraries and sequencing on a BGISEQ500 platform (BGI, CHN), low quality reads were filtered using SOAPnuke version 1.5.6[99]. The acquired clean reads were then mapped to the genome assemblies of our studied species using Bowtie version 2.2.5[100] and gene expression were quantified by RSEM version 1.2.12[101].

**Quantitative proteomes for *F. pumila* var. *pumila* and *W. pumilae*.** We identified and quantified proteins for ostiolar tissues (sampled in 2017) using iTRAQ (isobaric tags for relative and absolute quantitation)-based method. The strategy of quantifying proteomes was conducted according to the methods described by Tian et al. (2013)[102]. After total protein extraction, peptide labeling was performed using an iTRAQ Reagent 8-plex Kit according to the manufacturer's protocol. Extraction was followed by peptide fractionation, and the peptides separated from LC-20AD nano-HPLC (Shimadzu, JPN) were transferred into the tandem mass spectrometry Q EXACTIVE (MS/MS) (Thermo Fisher Scientific, US) for data-dependent acquisition (DDA) detection. After converting the raw MS/MS data into MGF format using Proteome Discoverer version 1.2 (Thermo Fisher Scientific, US), the exported data in MGF format were searched using Mascot version 2.3 (Matrix Science, US) against the protein-coding sequences from our gene prediction. Quantification of proteins was

660 achieved using IQuant[103], which uses the Mascot Percolator algorithm[104] to improve
661 the results of peptide identification and the principle of parsimony to assemble
662 proteomes. All the proteins with a false discovery rate (FDR)[105] of less than 1% were
663 retained for further analyses.

664 We used a DIA (data independent acquisition) approach to identify and quantify
665 proteins in female florets and adult female pollinators (collected in 2018). Procedures
666 identical to iTRAQ were first performed on the total protein extraction, peptide
667 fractionation and peptides separation. Then, to create reference spectra for DIA, we
668 first conducted DDA on a Q-EXACTIVE HF mass spectrometer (Thermo Fisher
669 Scientific, US) coupled with an Ultimate 3000 RSLCnano system (Thermo Fisher
670 Scientific, US) after a further peptide separation on an in-house packed nano-LC
671 column (150 μm × 30 cm, 1.8 μm, 100 Å). Then, using the same instruments, DIA
672 was performed following a brief procedure that consisted of a survey scan at 120,000
673 resolution from 400 to 1,250 m/z (MIT 50 ms), followed by scanning in DIA isolation
674 windows setting17 m/z with loop count 50 at 30,000 resolution (automatic gain
675 control target $3\times e^6$ and auto MIT). The DDA spectra were identified by searching
676 against the database of protein-coding sequences using the MaxQuant version 1.5.3.30
677 [106](Cox and Mann, 2008) at the FDR level of 1% with the minimum peptide length of
678 7. Based on the spectrogram database of DDA spectra, peptides and proteins in DIA
679 data were identified and quantified using Spectronaut[107], employing the mProphet
680 approach and setting iRT for retention time prediction. A target-decoy model was used
681 to verify the quantification results at an FDR level of 1%.

**Measurement of metabolomes of different types of tissues.** Chromatographic separation of metabolites was performed using an Ultra-Performance Liquid Chromatography (UPLC) System (Waters, UK), with an ACQUITY UPLC HSS T3 column (100mm*2.1mm, 1.8μm) (Waters, UK) being used for the reversed phase separation and setting oven temperature at 50° C and flow rate of 0.4 ml/min. After separation, gradient elution was conducted as following procedure: 0～2 min, 100% mobile phase A (water + 0.1% formic acid); 2~11 min, 0% to 100% mobile phase B (acetonitrile + 0.1% formic acid); 11~13 min, 100% B; 13~15 min, 0% to 100% A. The injection volume for each sample was 10 μl. Then, the eluted metabolites were identified in both positive and negative ion modes using a high-resolution tandem mass spectrometer Xevo G2 XS QTOF (Waters, UK). The capillary and sampling cone voltages were set at 3.0 kV and 40.0 V for positive ion mode and at 2.0 kV and 40.0 V for negative ion mode. Mass spectrometry data were acquired in Centroid MSE mode, setting the TOF mass range from 50 to 1200 Da and the scan time of 0.2 s. For MS/MS detection, all precursors were fragmented at 20-40 eV with the scan time of 0.2 s. A quality control (QC) sample (pooling all samples together) was used after every 10 samples. Peak alignment, peak picking and quantitation of each metabolite were performed using Progenesis QI version 2.2, and the quality control based on LOESS signal correction[108] was conducted using QC samples.

**Comparative transcriptome, proteome and metabolome analysis.** We carried out differential expression/concentration analysis for transcriptomes, proteomes and

703 metabolomes. This allowed us to anchor the key genes contributing to the
704 attractant-induced host specificity and adaptation of pollinator larvae to plant
705 chemical defenses.

706 For transcriptomes, differential expression were tested between ostiolar tissues
707 and female florets (feeder and seed florets)/galled ovules/seeds at each of the three fig
708 developmental stages (pre-receptive, receptive and post-receptive stages), and in
709 either type of fig tissues between any two of these three stages and between sexes at
710 each stage. For the pollinator, differential expression was conducted between
711 contacting attractive VOC(s) vs. not contacting and between adults and larvae. We
712 performed comparisons using DEseq2 version 1.4.5[109] based on negative binomial
713 distributions. *P*-values were corrected using a Benjamini-Hochberg (BH) method for
714 multiple tests. The differentially expressed genes (DEGs) with a fold change $\geq 2$ and
715 an adjusted p-value $\leq 0.05$ were considered as statistically significant.

716 For proteomes, in ostiolar tissues, we tested the proteins with significant
717 difference in quantity (PSDs) between the pre-receptive and the receptive stages and
718 between sexes at each stage using IQuant, and PSDs were defined as fold changes in
719 protein abundance $\geq 1.2$ and Q-value $\leq 0.05$. In female florets and adult female
720 pollinators, PSDs were analyzed using MSstats[110] at criterions of fold changes $\geq 2$ and
721 Q-value $\leq 0.05$.

722 For metabolomes, to examine whether there were significant differences in
723 profile of secondary metabolites associated with chemical defenses (SMCDs) between
724 different types of tissues and between the receptive and the post-receptive stages, we

first carried out enrichment analysis to enrich all relevant secondary metabolites into the pathways associated with plant chemical defenses and then clustered all samples into different categories using PLS-DA model in metaX[111]. Data were $log_2$-transformed and scaled by Pareto scaling. Secondary metabolites with significant difference in quantity (SMSDs) were defined as VIP (variable importance for the projection calculated based on the first two axes from PLS-DA model) ≥1, fold change ≥ 1.2 or ≤ 0.83 and Q-value ≤ 0.05. In addition, we performed PLS-DA model to test the difference in the entire profile of secondary metabolites between different types of tissues and between different fig developmental stages.

**Motif analysis.** We conducted motif analysis to check whether the OBPs in the same syntenic blocks among the three pollinating wasp species have similar motif structure using MEME Suite 5.0.4[112]. Motifs with E-value ≤ 0.05 were used for inter-specific comparisons. To predict the most likely OBPs related to the identification of specific attractant and repellent, we created a dataset consisting of all OBPs in *W. pumilae* and the specific OBPs that had been verified to bind to decanal and nonanal in *Adelphocoris lineolatus*[33] and *Culex quinquefasciatus*[34] as references, and performed motif analysis.

***In vitro* functional characterization of key genes.** The full-length of open reading frame (ORF) of the four key genes (for the plant) and of the two OBPs (for the pollinator) (Supplementary Table 17) was confirmed by RT–PCR and was then cloned into pET-28a (MilliporeSigma, US). After checking sequences by Sanger sequencing,

746 these genes were expressed in *E. coli* strains BL21 (DE3) and Rosetta (DE3). The
747 recombinant proteins produced were purified (purity > 90%) using modified
748 nickel-nitrilotriacetic acid agarose (Thermo Fisher Scientific, US).

749 We measured the affinities of the two OBPs to different substrates using surface
750 plasmon resonance (SPR) on a Biacore T200 system (GE Healthcare). OBPs were
751 reconstituted in sterile PBS and were diluted in 10mM sodium acetate trihydrate (pH
752 = 4.5) to the concentration of 20ug/ml. Then, each OBP was immobilized by the
753 amine coupling method on a CM5 sensor chip according to the manufacturer's
754 protocol (GE Healthcare). Analytes (decanal and nonanal) were diluted in running
755 buffer (5% DMSO-PBS-P) to the concentration ranging from 0 to 1000 μM and were
756 injected through channels at a flow rate of 20 μl/min. Using BIAevaluation (GE
757 Healthcare), both steady state affinity model and 1:1 binding model were performed
758 to quantify the binding affinity ($K_D$).

759 For enzyme activity assays of the four key genes of the plant, we used the
760 reaction system (500 μl) mainly composed of 50 mM Tris-HCl (pH 7.4), 0.4~1.0 mM
761 substrate(s) (Supplementary Table 17), 2M dimethyl sulfoxide (for the ADH and the
762 ALDH)/10% triton X-100 (for the two ASCLs), and 10 μl of purified protein (0.2
763 mg/ml). After 60 min of incubation at room temperature, we collected the reaction
764 products by headspace solid-phase microextraction for the ALDH and the ADH
765 (which were analyzed by GC–MS) and by extraction using diethyl ether for the two
766 ACSLs (which were analyzed by LC–MS). These experiments were repeated for three
767 time. In addition, three replications of negative controls (only adding the subtrates and

bovine serum albumin) were conducted, and no reaction products were detected.

**Cis-element detection and co-expression network analysis.** To test the regulatory mechanisms in the biosynthesis of attractant and repellent emitted by figs of *F. pumila* var. *pumila*, we first scanned the binding motifs present in the 2-kb promoter sequences upstream of key plant genes using PlantCARE[116]. Then, weighted undirected co-expression networks were conducted using the R package WGCNA[113] with a soft thresholding power of 8. Modules containing genes with correlated expression patterns were identified by gene clustering based on the topological overlap matrix[114] and by cutting the resulting dendrogram using the cutreeDynamic approach in the R package The Dynamic Tree Cut[115]. Genes with kME values larger than 0.95 were selected as hub genes. We checked whether some modules containing both some key plant genes and the transcription factors predicted to bind to them. This allowed us to uncover the likely regulatory mechanisms.

**ChIP-qPCR.** The open reading frame of each of the four transcription factors (*FpumHD-ZIP1*, *FpumHD-ZIP2*, *FpumbZIP1* and *FpumbHLH1* (Supplementary Table 13)) was cloned into the pET-28a to generate the fusion plasmid encoding the 6 His-tagged fusion protein. This plasmid was transformed into *E. coli* strain Rosetta (DE3), which were cultured and induced by 0.8 mM isopropyl-β-D-thiogalactoside (IPTG) at 37 °C. The induced cells were then sonicated for supernatant collection, and the purified recombinant proteins were obtained using a His-tag Protein Purification Kit (Beyotime Biotechnology, CHN). The purified proteins were used to immunize

789 rabbits for 52 days to acquire polyclonal antibody (ABclonal Biotechnology, CHN).
790 We successfully obtained the qualified antibodies for all the four transcription factors
791 for ChIP-qPCR experiments.
792 ChIP-qPCRs were then conducted for the two transcription factors with qualified
793 antibodies to examine if it can bind the putative target genes by model prediction. The
794 ChIP assay was performed based on the protocols described in Gendrel et al.,
795 (2005)[117]. Approximately 3.0 g ostiolar tissues from figs at receptive stages were
796 treated using 1% formaldehyde to crosslink and fix the DNA-protein complexes. The
797 cells of sampled tissues were lysed, and each antibody was used to immunoprecipitate
798 the antigen transcription factor with its binding DNA fragments. The DNA in the ChIP
799 products was applied in qPCR with primer pairs designed for the promoters of
800 putative target genes in a QuantStudio$^{TM}$ 5 real-time PCR detection system (Thermo
801 Fisher Scientific, US). Each qPCR reaction was performed in triplicates, and the cycle
802 thresholds (Cts values) of ChIP products were compared with those of input samples
803 and negative controls (only using IgG) for calculating % input and fold enrichment (%
804 input (ChIP)/ % input (negative control)). We failed to obtain the Ct values for
805 negative controls by the end of 35 qPCR cycles, and we therefore used the Ct value of
806 35 for each negative control when calculating % input and fold enrichment.
807

808 **Additional information**

809 All supplemental figures and tables are included in supplementary information.
810

## References


1. Lamichhaney, S. et al. Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature* **518**, 371–375 (2015).
2. Simões, M. et al. The evolving theory of evolutionary radiations. *Trends Ecol. Evol.* **31**, 27–34 (2016).
3. Arnegard, M. E. et al. Genetics of ecological divergence during speciation. *Nature* **511**, 307–311 (2014).
4. Hoffmann, A. A. & Sgrò, C. M. Climate change and evolutionary adaptation. *Nature* **470**, 479–485 (2011).
5. Olsen, J. L. et al. The genome of the seagrass *Zostera marina* reveals angiosperm adaptation to the sea. *Nature* **530**, 331–335 (2016).
6. Becerra, J. X., Nogeb, K. & Venable, D. L. Macroevolutionary chemical escalation in an ancient plant–herbivore arms race. *Proc. Natl. Acad. Sci. USA* **106**, 18062–18066 (2009).
7. Edger, P. P. et al. The butterfly plant arms-race escalated by gene and genome duplications. *Proc. Natl. Acad. Sci. USA* **112**, 8362–8366 (2015).
8. Adler, L. S. & Bronstein, J. L. Attracting antagonists: does floral nectar increase leaf herbivory? *Ecology* **85**, 1519–1526 (2004).
9. McCall, A. C. & Irwin, R. E. Florivory: the intersection of pollination and herbivory. *Ecol. Lett.* **9**, 1351–1365 (2006).
10. Cook, J. M. & Rasplus, J.-Y. Mutualists with attitude: coevolving fig wasps and figs. *Trends Ecol. Evol.* **18**, 241–248 (2003).

11. Zhang, X. et al. Genomes of the banyan tree and pollinator wasp provide insights into fig-wasp coevolution. *Cell* **183**, 1–15 (2020).

12. Herre, E. A., Jandér, K. C. & Machado, C. A. Evolutionary ecology of figs and their associates: recent progress and outstanding puzzles. *Annu. Rev. Ecol. Evol. Syst.* **39**, 439–458 (2008).

13. Souza, C. D. et al. Diversity of fig glands is associated with nursery mutualism in fig trees. *Am. J. Bot.* **102**, 1564–1577 (2015).

14. Souto-Vilarós, D. et al. Pollination along an elevational gradient mediated both by floral scent and pollinator compatibility in the fig and fig-wasp mutualism. *J. Ecol. Evol.* **106**, 2256–2273 (2018).

15. Wang, R. et al. Loss of top-down biotic interactions changes the relative benefits for obligate mutualists. *Proc. R. Soc. B.* **286**, 20182501 (2019).

16. Mori K. et al. Identification of RAN1 orthologue associated with sex determination through whole genome sequencing analysis in fig (*Ficus carica L.*). *Sci*. *Rep*. **7**, 41124 (2017).

17. Proffit M. et al. Chemical signal is in the blend: bases of plant-pollinator encounter in a highly specialized interaction. *Sci. Rep.* **10**, 10071 (2020).

18. Chen, C. et al. Private channel: a single unusual compound assures specific pollinator attraction in *Ficus semicordata*. *Funct. Ecol.* **23**, 941–950 (2009).

19. Wang, G., Cannon, C. H. & Chen, J. Pollinator sharing and gene flow among closely related sympatric dioecious fig taxa. *Proc. R. Soc. B.* **283**, 20152963 (2016).

855 20. Yu, H. et al. De novo transcriptome sequencing in *Ficus hirta* Vahl. (Moraceae) to
856 investigate gene regulation involved in the biosynthesis of pollinator attracting
857 volatiles. *Tree Genet. Genomes* **11**, 91 (2015).
858 21. Soler, C. C. L., Proffit, M., Bessière, J.-M., Hossaert -McKey, M. & Schatz, B.
859 Evidence for intersexual chemical mimicry in a dioicous plant. *Ecol. Lett.* **15**,
860 978–985 (2012).
861 22. Volf, M. et al. Community structure of insect herbivores is driven by
862 conservatism, escalation and divergence of defensive traits in *Ficus*. *Ecol. Lett.*
863 **21**, 83–92 (2018).
864 23. Martinson, E.O., Hackett, J.D., Machado, C.A. & Arnold A.E. Metatranscriptome
865 analysis of fig flowers provides insights into potential mechanisms for mutualism
866 stability and gall induction. *PLoS ONE* **10**, e0130745 (2015).
867 24. Zhang, H. et al. Leaf-mining by *Phyllonorycter blancardella* reprograms the
868 host-leaf transcriptome to modulate phytohormones associated with nutrient
869 mobilization and plant defense. *J. Insect Physiol.* **84**, 114–127 (2016).
870 25. Schultz, J.C., Edger, P.P., Body, M. & Appel, H.M. A galling insect activates plant
871 reproductive programs during gall development. *Sci. Rep.* **9**, 1833 (2019).
872 26. The *Nasonia* Genome Working Group. Functional and evolutionary insights from
873 the genomes of three parasitoid *Nasonia* species. *Science* **327**, 343–348 (2010).
874 27. Xiao, J.-H. et al. Obligate mutualism within a host drives the extreme
875 specialization of a fig wasp genome. *Genome Biol.* **14**, R141 (2013).
876 28. Ohri, D. & Khoshoo, T. N. Nuclear DNA contents in the genus *Ficus* (Moraceae).

877 *Plant Syst. Evol.* **156**, 1–4 (1987).
878 29. Chen, Y., Compton, S. G., Liu, M. & Chen, X.-Y. Fig trees at the northern limit of
879 their range: the distributions of cryptic pollinators indicate multiple glacial
880 refugia. *Mol. Ecol.* **21**, 1687–1701 (2012).
881 30. Chen, L.-G. et al. Binding affinity characterization of an antennae-enriched
882 chemosensory protein from the white-backed planthopper, *Sogatella furcifera*
883 (Horváth), with host plant volatiles. *Pestic. Biochem. Phys.* **152**, 1–7 (2018).
884 31. Gu, S.-H. et al. Functional characterization and immunolocalization of odorant
885 binding protein 1 in the lucerne plant bug, *Adelphocoris lineolatus* (GOEZE).
886 *Arch. Insect Biochem.* **77**, 81–98 (2011).
887 32. Leal, W. S. et al. Reverse and conventional chemical ecology approaches for the
888 development of oviposition attractants for *Culex* mosquitoes. *PLoS ONE* **3**, e3045
889 (2008).
890 33. Rizzo, W. B. et al. Fatty aldehyde and fatty alcohol metabolism: review and
891 importance for epidermal structure and function. BBA-Mol. *Cell Biol. L.* **1841**,
892 377–389 (2014).
893 34. Schwab, W., Davidovich Rikanati, R. & Lewinsohn, E. Biosynthesis of plant
894 derived flavor compounds. *Plant J.* **54**, 712–732 (2008).
895 35. Capella, M., Ribone, P. A., Arce, A. L. & Chan, R. L. *Arabidopsis thaliana*
896 HomeoBox 1 (AtHB1), a Homedomain-Leucine Zipper I (HD-Zip I) transcription
897 factor, is regulated by PHYTOCHROME-INTERACTING FACTOR 1 to
898 promote hypocotyl elongation. *New Phytol.* **207**, 669–682 (2015).

899 36. Jiang, W. et al. Two transcription factors TaPpm1 and TaPpb1 co-regulate
900 anthocyanin biosynthesis in purple pericarps of wheat. *J. Exp. Bot.* **69**, 2555–
901 2567 (2018).
902 37. Guan, R. et al. Draft genome of the living fossil *Ginkgo biloba*. *GigaScience* **5**,
903 49 (2016).
904 38. Mithöfer, A., & Boland, W. Plant defense against herbivores: chemical aspects.
905 *Annu. Rev. Plant Biol.* **63**, 431–450 (2012).
906 39. Salazar, D. et al. Origin and maintenance of chemical diversity in a species-rich
907 tropical tree lineage. *Nat. Ecol. Evol.* **2**, 983–990 (2018).
908 40. Després, L., David, J. P. & Gallet, C. The evolutionary ecology of insect
909 resistance to plant chemicals. *Trends Ecol. Evol.* **22**, 298–307 (2007).
910 41. Segar, S. T., Volf, M., Sisol, M., Pardikes, N. & Souto-Vilarós, A. D. Chemical
911 cues and genetic divergence in insects on plants: conceptual cross pollination
912 between mutualistic and antagonistic systems. *Curr. Opin. Insect Sci.* **32**, 83–90
913 (2019).
914 42. Cook, J. M. & Segar, S. T. Speciation in fig wasps. *Ecol. Entomol.* **35**, 54–66
915 (2010).
916 43. Cruaud, A. et al. An extreme case of plant-insect codiversification: figs and
917 fig-pollinating wasps. *Syst. Biol.* **61**, 1029–1047 (2012).
918 44. Yu, H. et al. Multiple parapatric pollinators have radiated across a continental fig
919 tree displaying clinal genetic variation. *Mol. Ecol.* **28**, 2391–2405 (2019).
920 45. Satler, J. D. et al. Inferring processes of coevolutionary diversification in a

community of Panamanian strangler figs and associated pollinating wasps. *Evolution* **73**, 2295–2311 (2019).

46. Wang, G. et al. Genomic evidence of prevalent hybridization throughout the evolutionary history of the fig-wasp pollination mutualism. *Nat. Commun.* **12**, 718 (2021).

47. Hoballah, M. E. et al. Single gene-mediated shift in pollinator attraction in Petunia. *Plant Cell* **19**, 779–790 (2007).

48. Potts, S. G. et al. Global pollinator declines: trends, impacts and drivers. *Trends Ecol. Evol.* **25**, 345–353 (2010).

49. Kiers, E. T., Palmer, T. M., Ives, A. R., Bruno, J. F. & Bronstein, J. L. Mutualisms in a changing world: an evolutionary perspective. *Ecol. Lett.* **13**, 1459–1474 (2010).

50. Segar, S. T., Fayle, T .M., Srivastava, D. S., Lewinsohn, T. M., Lewis, O. T., Novotny, V., Kitching, R. L. & Maunsell, S. C. The role of evolution in shaping ecological networks. *Trends Ecol. Evol.* **35**, 454-466 (2020).

51. Stoy, K. S., Gibson, A. K., Gerardo, N. M. & Morran, L. T. A need to consider the evolutionary genetics of host-symbiont mutualisms. *J Evol. Biol.* **33**, 1656–1668 (2020).

52. Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770 (2011).

53. Xiao, C.-L. et al. MECAT: fast mapping, error correction, and de novo assembly for single-molecule sequencing reads. *Nat. Methods* **14**, 1072–1074 (2017).

943 54. Walker, B. J. et al. Pilon: an integrated tool for comprehensive microbial variant
944 detection and genome assembly improvement. *PLoS One* **9**, e112963 (2014).
945 55. Pryszcz, L. P., & Gabaldón, T. Redundans: an assembly pipeline for highly
946 heterozygous genomes. *Nucleic Acids Res.* **44**, e113 (2016).
947 56. Zhang, Z., Schwartz, S., Wagner, L. & Miller, W. A greedy algorithm for aligning
948 DNA sequences. *J. Comput. Biol.* **7**, 203–214 (2000).
949 57. English, A. C. et al. Mind the gap: upgrading genomes with Pacific Biosciences
950 RS long-read sequencing technology. *PLoS ONE* **7**, e47768 (2012).
951 58. Sahlin, K., Chikhi, & Arvestad, R. L. Assembly scaffolding with
952 PE-contaminated mate-pair libraries. *Bioinformatics* **32**, 1925–1932 (2016).
953 59. Belton, J. M. et al. Hi-C: a comprehensive technique to capture the conformation
954 of genomes. *Methods* **58**, 268–276 (2012).
955 60. Servant, N. et al. (2015) HiC-Pro: an optimized and flexible pipeline for Hi-C
956 data processing. *Genome Biol.* **16**, 259.
957 61. Durand, N.C. et al. Juicer provides a one-click system for analyzing
958 loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98 (2016).
959 62. Dudchenko, O. et al. De novo assembly of the Aedes aegypti genome using Hi-C
960 yields chromosome-length scaffolds. *Science* **356**, 92–95 (2017).
961 63. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E.
962 M. BUSCO: assessing genome assembly and annotation completeness with
963 single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
964 64. Wu, T. D. & Watanabe, C. K. GMAP: a genomic mapping and alignment program

965 for mRNA and EST sequences. *Bioinformatics* **21**, 1859–1875 (2005).
966 65. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic*
967 *Acids Res.* **27**, 573–580 (1999).
968 66. Bao, W., Kojima, K. K. & Kohany, O. Repbase Update, a database of repetitive
969 elements in eukaryotic genomes. *Mobile DNA* **6**, 11 (2015).
970 67. Xu, Z. & Wang, H. LTR_FINDER: an efficient tool for the prediction of
971 full-length LTR retrotransposons. *Nucleic Acids Res.* **35**, W265–W268 (2007).
972 68. Edgar, R. C. & Myers, E. W. PILER: identification and classification of genomic
973 repeats. *Bioinformatics* **21**, i152–i158 (2005).
974 69. Price, A. L., Jones, N. C. & Pevzner, P. A. De novo identification of repeat
975 families in large genomes. *Bioinformatics* **21**, i351–i358 (2005).
976 70. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of
977 transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**, 955–964
978 (1997).
979 71. Nawrocki, E. P. & Eddy, S. R. Infernal 1.1: 100-fold faster RNA homology
980 searches. *Bioinformatics* **29**, 2933–2935 (2013).
981 72. Nawrocki, E.P. et al. Rfam 12.0: updates to the RNA families database. *Nucleic*
982 *Acids Res.* **43**, D130–D137 (2015).
983 73. Holt, C. & Yandell, M. MAKER2: an annotation pipeline and genome-database
984 management tool for second-generation genome projects. *BMC Bioinformatics* **12**,
985 491 (2011).
986 74. Kim, D., Langmead, B. & Salzberg, S.L. HISAT: a fast spliced aligner with low

memory requirements. *Nat. Methods* **12**, 357–360 (2015).

75. Pertea, M. et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).

76. Johnson, A. D. et al. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics* **24**, 2938–2939 (2008).

77. Stanke, M. & Morgenstern, B. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* **33**, W465–W467 (2005).

78. Buels, R. et al. JBrowse: a dynamic web platform for genome visualization and analysis. *Genome Biol.* **17**, 66 (2016).

79. Buchfink, B., Xie, C. & Huson, D.H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2015).

80. Bairoch, A. & Apweiler, R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* **28**, 45–48 (2000).

81. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* **44**, D457–D462 (2016).

82. Jones, P. et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).

83. Conesa, A. et al. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674–3676 (2005).

84. Li, L., Stoeckert, C. J., Jr. & Roos, D. S. OrthoMCL: identification of ortholog

groups for eukaryotic genomes. *Genome Res.* **13**, 2178–2189 (2003).

85. Li, H. et al. TreeFam: a curated database of phylogenetic trees of animal gene families. *Nucleic Acids Res.* **34**, D572–D580 (2006).

86. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).

87. Capella-Gutierrez, S., Silla-Martinez, J.M. & Gabaldon, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).

88. Guindon, S., Delsuc, F., Dufayard, J. F. & Gascuel, O. Gascuel, Estimating maximum likelihood phylogenies with PhyML. *Methods Mol. Biol.* **537**, 113–137 (2009).

89. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).

90. De Bie, T., Cristianini, N., Demuth, J. P. & Hahn, M. W. CAFE: a computational tool for the study of gene family evolution. *Bioinformatics* **22**, 1269–1271 (2006).

91. Tang, H. et al. Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. *Genome Res.* **18**, 1944–1954 (2008).

92. Schwartz, S. et al. Human-mouse alignments with BLASTZ. *Genome Res.* **13**, 103–107 (2003).

93. Bailey, J. A., Yavor, A. M., Massa, H. F., Trask, B. J. & Eichler, E. E. Segmental duplications: organization and impact within the current human genome project assembly. *Genome Res.* **11**, 1005–1017 (2001).

94. Birney, E., Clamp, M. & Durbin, R. GeneWise and Genomewise. *Genome Res.* **14**, 988–995 (2004).

95. Ruan, J., Li, H., Chen, Z. & Coghlan, A. TreeFam: 2008 Update. *Nucleic Acids Res.* **36**, D735–D740 (2008).

96. Tholl, D. et al. Practical approaches to plant volatile analysis. *Plant J.* **45**, 540–560 (2006).

97. Wen, B., Mei, Z., Zeng, C. & Liu, S. metaX: a flexible and comprehensive software for processing metabolomics data. *BMC Bioinformatics* **18**, 183 (2017).

98. Wen, P. et al. The sex pheromone of a globally invasive honey bee predator, the Asian eusocial hornet, *Vespa velutina*. *Sci. Rep.* **7**, 12956 (2017).

99. Chen, Y., Chen, Y., Shi, C. & Huang, Z. et al. SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. *Gigascience* **7**, 1–6 (2018).

100. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

101. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).

102. Tian, X., Chen, L., Wang, J., Qiao, J. & Zhang, W. Quantitative proteomics reveals dynamic responses of Synechocystis sp. PCC 6803 to next-generation biofuel butanol. *J. Proteomics* **78**, 326–345 (2013).

103. Wen, B., Zhou, R., Feng, Q., Wang, Q., Wang, J. & Liu, S. IQuant: an automated pipeline for quantitative proteomics based upon isobaric tags. *Proteomics* **14**,

2280–2285 (2014).

104. Brosch, M., Yu, L., Hubbard, T. & Choudhary, J. Accurate and sensitive peptide identification with Mascot Percolator. J. *Proteome Res.* **8**, 3176–3181 (2009).

105. Savitski, M. M., Wilhelm, M., Hahne, H., Kuster, B. & Bantscheff, M. A scalable approach for protein false discovery rate estimation in large proteomic data sets. *Mol. Cell. Proteomics* **14**, 2394–2404 (2015).

106. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).

107. Bruderer, R. et al. Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues. *Mol. Cell. Proteomics* **14**, 1400–1410 (2015).

108. Dunn, W.B. et al. Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat. Protoc.* **6**, 1060–1083 (2011).

109. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

110. Choi, M. et al. MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. *Bioinformatics* **30**, 2524–2526 (2014).

111. Wen, X.-L., Wen, P., Dahlsjö, C. A. L., Sillam-Dussès, D. & Šobotník, J.

1075 Breaking the cipher: ant eavesdropping on the variational trail pheromone of its
1076 termite prey. *Proc. R. Soc. B.* **284**, 20170121 (2017).
1077 112.Bailey, T. L. & Elkan, C. Fitting a mixture model by expectation maximization to
1078 discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **2**, 28–36
1079 (1994).
1080 113.Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation
1081 network analysis. *BMC Bioinformatics* **9**, 559 (2008).
1082 114.Yip, A. M. & Horvath, S. Gene network interconnectedness and the generalized
1083 topological overlap measure. *BMC Bioinformatics* **8**, 22 (2007).
1084 115.Langfelder, P., Zhang, B. & Horvath, S. Defining clusters from a hierarchical
1085 cluster tree: the dynamic tree cut package for R. *Bioinformatics* **24**, 719–720
1086 (2008).
1087 116.Lescot, M. et al. PlantCARE, a database of plant cis-acting regulatory elements
1088 and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids*
1089 *Res.* **30**, 325–327 (2002).
1090 117.Gendrel, A.V., Lippman, Z., Martienssen, R. & Colot V. Profiling histone
1091 modification patterns in plants using genomic tiling microarrays. *Nat. Methods* **2**,
1092 213–218 (2005).

**Table 1 | Summary statistics for the assembly of *F. pumila var. pumila* and *W. pumilae* genomes.**

| Chromosome ID | *F. pumila var. pumila* | | *W. pumilae* | |
|---|---|---|---|---|
| | No. of genes | Length (bp) | No. of genes | Length (bp) |
| Chr1 | 1,697 | 20,463,500 | 816 | 21,315,831 |
| Chr2 | 1,871 | 21,202,951 | 2,076 | 59,985,216 |
| Chr3 | 2,335 | 23,199,346 | 2,631 | 66,440,284 |
| Chr4 | 2,412 | 23,721,380 | 2,225 | 54,409,331 |
| Chr5 | 3,327 | 31,603,922 | 2,281 | 59,419,729 |
| Chr6 | 1,649 | 20,816,579 | 2,263 | 55,968,755 |
| Chr7 | 2,070 | 23,331,000 | | |
| Chr8 | 2,097 | 21,006,959 | | |
| Chr9 | 1,856 | 21,788,500 | | |
| Chr10 | 1,740 | 20,107,953 | | |
| Chr11 | 1,798 | 22,360,920 | | |
| Chr12 | 2,000 | 20,847,995 | | |
| Chr13 | 2,526 | 34,592,857 | | |
| Number of contigs | 543 | | 102 | |
| Total length of contigs (Mb) | 315.7 | | 318.2 | |
| Contig N50 (Mb) | 2.3 | | 10.9 | |
| Anchored genome content (Mb) | 304.8 | | 317.5 | |
| Anchored rate | 96.6% | | 99.8% | |
| Scaffold N50 (Mb) | 22.4 | | 59.4 | |
| Number of genes | 28,187 | | 12,316 | |

## Figure legends

**Fig. 1 | Fig-pollinator mutualism between *F. pumila* var. *pumila* and *Wiebesia pumilae* and determination of the compound attracting *W. pumilae*. a,** Life cycle of *W. pumilae* based on four fig developmental stages (pre-receptive, receptive, post-receptive and mature stages). This *Ficus* species is dioecious with figs on female trees growing long-styled female florets (seed florets) that are not available for pollinator oviposition. Therefore, female trees only produce seeds, while figs on functional male trees contain both male florets and short-styled female florets (feeder florets) that can be used by female pollinators for oviposition to support the larvae of the pollinators. At the receptive stage, adult female pollinators are attracted by host-specific VOCs and enter figs only through ostiole (lined with bracts), either ovipositing into ovules of feeder florets in functional male figs or pollinating seed florets inside female figs. Pollinator larvae develop in induced galled ovules and both larvae and seeds grow during the post-receptive stage. At the mature stage, after mating with adult males, adult female pollinators leave their natal figs carrying pollen donated by mature male florets and search for receptive figs and complete the cycle. **b,** Electrophysiological responses of adult females of *W. pumilae* to the VOCs extracted from *F. pumila* var. *pumila* figs at receptive stage using GC–EAD. Each curve represents the response of a single female pollinator. **c,** Electrophysiological responses of adult female pollinators to the synthesized standard of each tentative VOC compound (each electroantennogram curve represents five overlapped replicates). **d,** Preference of adult female pollinators to different tentative compounds

using Y-tube olfactometer tests (Supplementary Table 9).

**Fig. 2 | Molecular mechanisms of the specific host identification of *W. pumilae*. a,** Numbers of genes in the four olfactory-related gene families (odorant-binding proteins (OBPs), olfactory receptors (ORs), chemosensory proteins (CSPs) and ionotropic receptor (IRs)) in different insect species. Significantly contracted families ($^{***}$: $p<0.001$) were shown for *W. pumilae* and *C. solmsi*, and species were ranked according to their phylogeny (Supplementary Fig. 4b). **b,** Transcription and translation of OBP genes of adult females of *W. pumilae* not contacting (as the control) and contacting the VOCs emitted by *F. pumila* var. *pumila* figs at the receptive stage (Supplementary Table 10). **c,** Motif analysis predicting the most likely *W. pumilae* OBPs that can bind to decanal and nonanal (Supplementary Fig. 9). **d,** The binding affinities ($K_D$) of the predicted OBPs to decanal and nonanal using surface plasmon resonance (SPR) experiments (Supplementary Fig. 10 and Supplementary Table 11). Lower $K_D$ indicates higher binding affinity, and error bars represent standard errors calculated by parameter estimation in steady state affinity model.

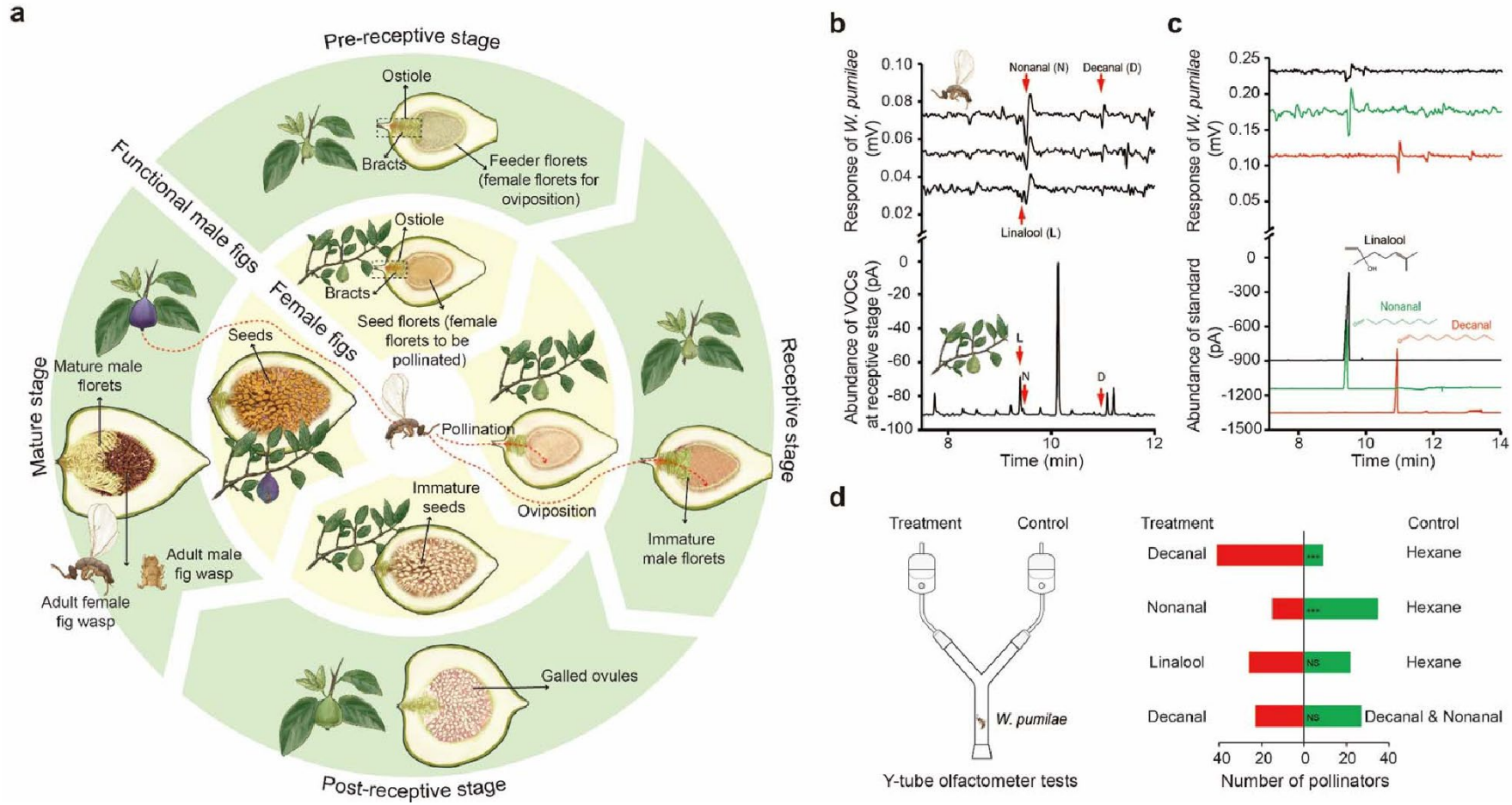
**Fig. 3 | Regulation of gene expression in attractant biosynthesis in figs of *F. pumila* var. *pumila*. a,** Pathways associated with biosynthesis of decanal and nonanal (fatty acid biosynthesis (ko00061), elongation (ko00062) and metabolism (ko00071 and ko00592)). **b,** Fold changes of all PSDs and their transcriptomic expression between receptive and pre-receptive stages in ostiolar types in proteomes

(Supplementary Table 12). $^{NS}$: p>0.05, $^{*}$: p<0.05, $^{**}$: p<0.01, $^{***}$: p<0.001. **c-e,** Results of *in vitro* functional characterization of the four key genes in the biosynthesis of decanal and nonanal. The peaks of synthesized standards and reaction products (treatments with enzyme added for three replicates) were shown for each key gene. Because there are two steps in the catalytic reaction of the two ACSLs, we showed the ion intensity of both the intermediate product (hexadecanoyl-AMP) and the final product (hexadecanoyl-CoA) separated by LC–MS. The reaction products of the ALDH and the ADH (decanal and decanol) were identified using GC–MS. **f,** Transcriptomic expression of genes in the co-expression module including two key genes and the transcription factors predicted to regulate the expression of these two key genes (Supplementary Tables 13 and 14). **g-h,** Results of ChIP–qPCRs (% input and fold enrichment) showing the evidence that the predicted transcription factors can bind to the promoter regions of *FpumACSL10* (FPUM_023966-RA) and *FpumALDH1* (see Supplementary Tables 12 and 13). Error bars represent standard errors of experimental results.

**Fig. 4 | Metabolic and genomic signature of antagonistic interaction between *F. pumila* var. *pumila* and *W. pumilae*. a,** Results of PLS-DA for terpenoids (triterpenes and sesquiterpenes) and phenylpropanoids. Each oval indicates the 95% confidence intervals of a sample group. **b,** Distribution of SMCDs across fates of female florets. No SMSDs between feeder and seed florets and only three SMSDs (two downregulated and one upregulated) between galled ovules and seeds were found in

the pathways related to plant chemical defenses (Supplementary Fig. 12). **c**, Largely matched turnover of SMSDs in feeder floret-galled ovule and seed floret-seed transitions (using Spearman's rank correlation tests). **d**, Numbers of genes in CYP450, CCE and GST gene families in different insect species. Significantly contracted families ($^{***}$: $p<0.001$) were shown for *W. pumilae* and *C. solmsi*.

1168 -




1169

1170 Fig. 1.

1171

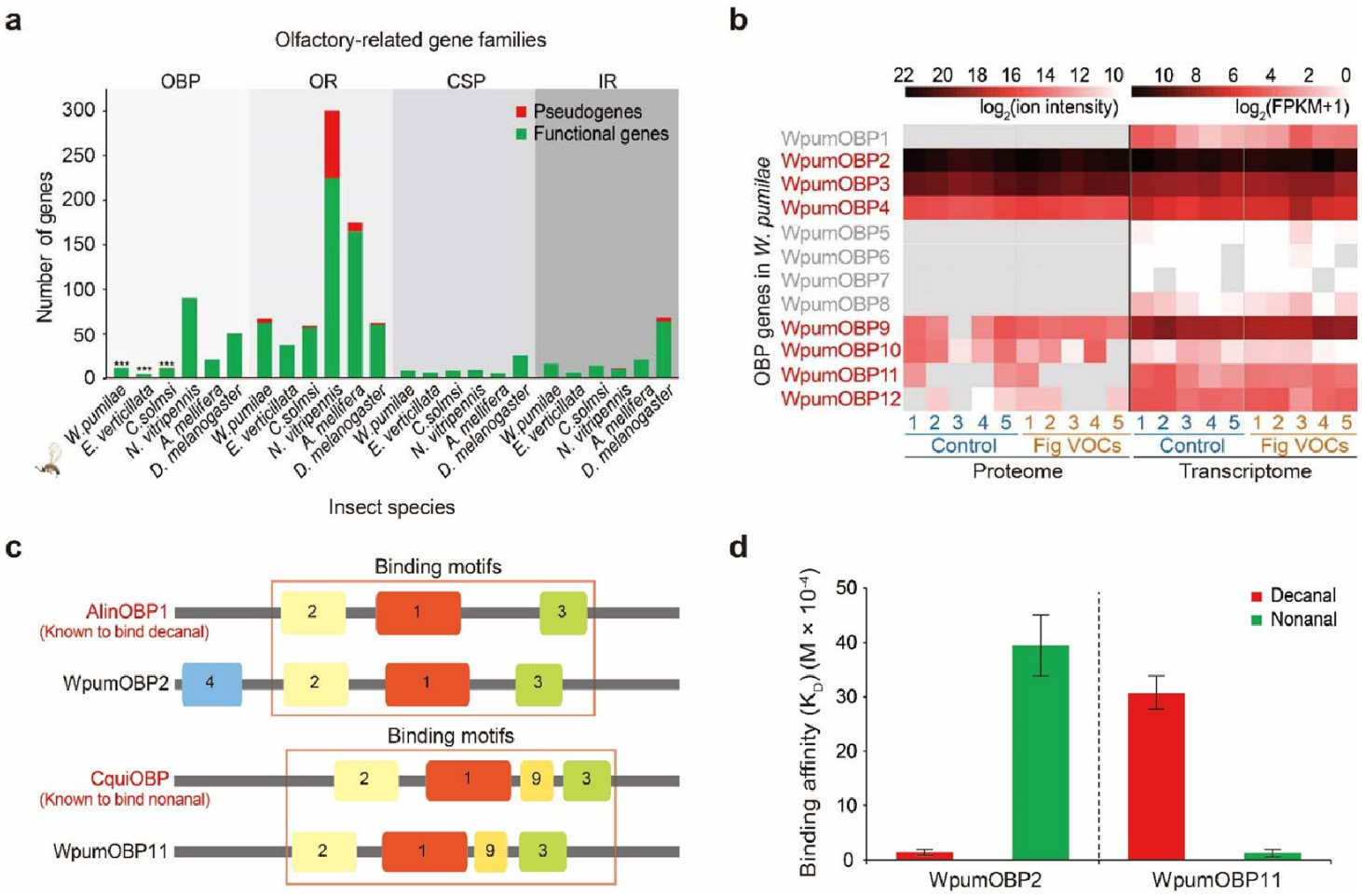
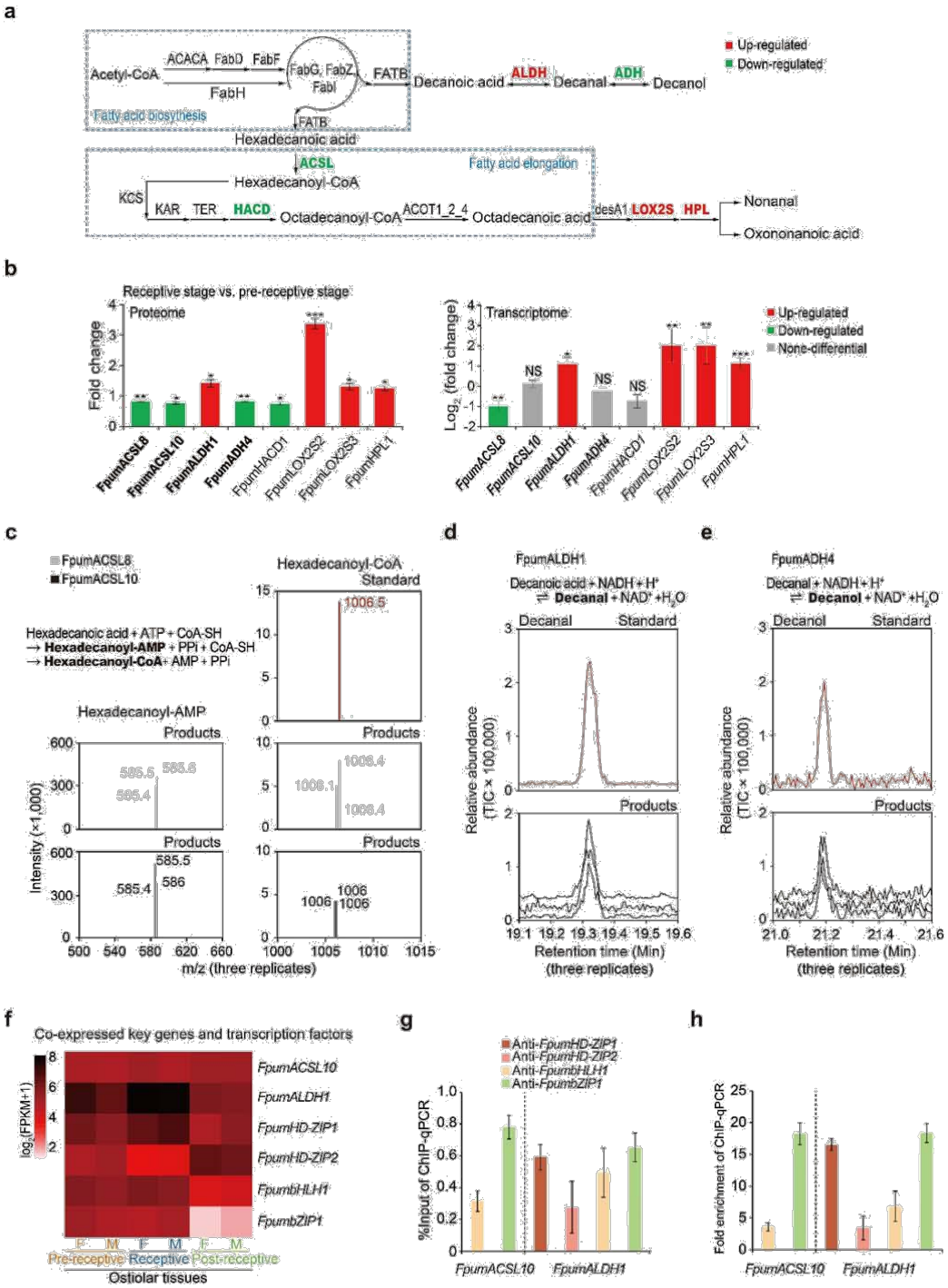

a
Olfactory-related gene families
OBP
OR
CSP
IR
Pseudogenes
Functional genes
Number of genes
300
250
200
150
100
50
0
***
***
***
W.pumilae
E. verticillata
C.solmsi
N. vitripennis
A. mellifera
D. melanogaster
Insect species
b
22 20 18 16 14 12 10
log2(ion intensity)
10 8 6 4 2 0
log2(FPKM+1)
OBP genes in W. pumilae
WpumOBP1
WpumOBP2
WpumOBP3
WpumOBP4
WpumOBP5
WpumOBP6
WpumOBP7
WpumOBP8
WpumOBP9
WpumOBP10
WpumOBP11
WpumOBP12
1 2 3 4 5 1 2 3 4 5 1 2 3 4 5 1 2 3 4 5
Control
Fig VOCs
Control
Fig VOCs
Proteome
Transcriptome
c
Binding motifs
AlinOBP1
(Known to bind decanal)
WpumOBP2
Binding motifs
CquiOBP
(Known to bind nonanal)
WpumOBP11
d
Binding affinity (KD) (M × 10-4)
50
40
30
20
10
0
Decanal
Nonanal
WpumOBP2
WpumOBP11

Fig. 2.




1174

1175 Fig. 3

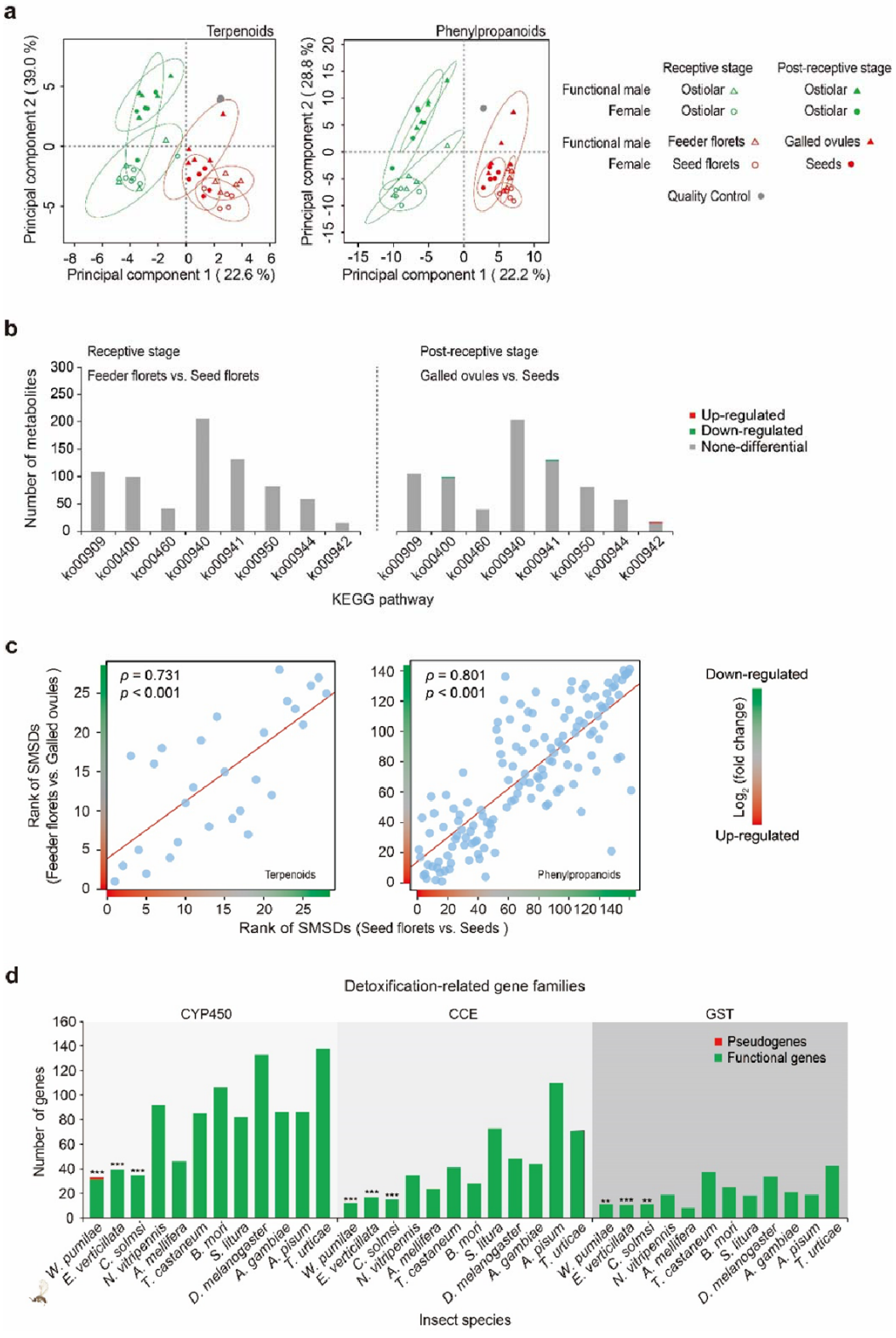

a
Terpenoids
Phenylpropanoids
Principal component 2 ( 39.0 %)
Principal component 1 ( 22.6 %)
Principal component 2 ( 28.8 %)
Principal component 1 ( 22.2 %)
Receptive stage
Post-receptive stage
Functional male
Female
Ostiolar
Feeder florets
Seed florets
Galled ovules
Seeds
Quality Control
b
Receptive stage
Feeder florets vs. Seed florets
Post-receptive stage
Galled ovules vs. Seeds
Number of metabolites
Up-regulated
Down-regulated
None-differential
ko00909
ko00400
ko00460
ko00940
ko00941
ko00950
ko00944
ko00942
KEGG pathway
c
p = 0.731
p < 0.001
p = 0.801
p < 0.001
Rank of SMSDs (Feeder florets vs. Galled ovules )
Terpenoids
Phenylpropanoids
Down-regulated
Log2 (fold change)
Up-regulated
Rank of SMSDs (Seed florets vs. Seeds )
d
Detoxification-related gene families
CYP450
CCE
GST
Pseudogenes
Functional genes
Number of genes
W. pumilae
E. verticillata
C. solmsi
N. vitripennis
A. mellifera
T. castaneum
B. mori
S. litura
D. melanogaster
A. gambiae
A. pisum
T. urticae
Insect species

Fig. 4