

<https://doi.org/10.1038/s42003-025-08733-1>

# Demography and adaptation of a species specific pollinator associated with an invasive fig tree



Li-Hua Wu<sup>1,11</sup>, Jie Zou<sup>1,11</sup>, Yu-Ting Jiang<sup>1,11</sup>, Ling Lu<sup>1</sup>, Kai Jiang<sup>1,2</sup>, Kai-Jian Zhang<sup>3</sup>, Wei-Chao Liu<sup>1</sup>, Wei-Min Xiang<sup>1</sup>, Gang Wang<sup>4</sup>, Simon T. Segar<sup>5</sup>, Simon van Noort<sup>6,7</sup>, Yuan-Yuan Li<sup>1</sup>, Yuan-Ye Zhang<sup>8</sup>, Xiao-Yong Chen<sup>1</sup>, Stephen G. Compton<sup>9</sup> & Rong Wang<sup>1,10</sup>✉

Some fig species introduced outside of their native range have become invasive when colonized by their obligate pollinating wasps, but how these pollinators migrated and adapted to novel environments are less studied. Here, we focus on *Eupristina verticillata*, the obligate pollinating wasp of an invasive fig tree species (*Ficus microcarpa*), to uncover its demography and the molecular basis for adaptations to novel environments. We find that only one of the three cryptic species colonized in the sampling locations outside of its native range. This dominant cryptic species migrated simultaneously from the native range to the Americas and to the Mediterranean c. 130 years ago. Moreover, selective sweep analyses reveal several positively selected genes associated with adaptations to the nonnative range. Genome-wide association detect a nonsynonymous substitution in a dopamine N-acetyltransferase gene significantly linked with brood size. Our study outlines the route to colonization and genetic adaptations of an invasive mutualism.

Invasive plants pose a major threat to the maintenance of global biodiversity and can cause severe economic damage<sup>1</sup>. Release from natural enemies and competitors can lead to ecological domination by nonnatives and ecological homogenization<sup>2</sup>. The field of invasion ecology seeks to determine the traits and conditions that lead to expansion in the nonnative range<sup>3</sup>. Successful establishment and invasion of introduced plants often depends on their mutualistic organisms (e.g., pollinators, seed dispersers, and rhizobia), as well as intrinsic factors including fast growth, self-compatibility, and anemophily<sup>2,4,5</sup>. Plants with generalist mutualists are considered to have a high risk of invasion due to their ability to form new interactions with local species<sup>6,7</sup>. While the invasive potential of more specific or even obligate interactions is generally considered to be lower, examples of obligate mutualists facilitating the invasion of introduced plants have been reported recently<sup>8</sup>. Little is known about how these obligate mutualists have expanded beyond their native ranges and how have adapted to novel environments. This is unfortunate because these tight interactions are defined by

conserved and easily quantifiable traits, allowing straightforward tests of ecological theory that can be scaled up to more diffuse interactions.

The introduction/migration of obligate mutualists (initially absent in the introduced ranges of host plants) is likely to be uncoupled from their host plants. This is because most host plants were deliberately introduced outside their native ranges for human purposes (e.g., horticulture and forestry)<sup>9</sup>, while pollinator introductions have been largely accidental. In addition, in the absence of obligate mutualists, closely related species of the obligate mutualists may switch hosts and facilitate plant invasion<sup>10</sup>. Thus, precise identification of mutualistic organisms and analysis of population demography using molecular markers are necessary to reveal how introduced plants gain mutualists and become invasive (and how we might control them)<sup>11</sup>. An obligate mutualist species may disperse from a single native population at a single time, leading to a strong founder effect limiting its potential for adaptation<sup>12</sup>. Alternatively, introduction from different populations and/or multiple times can cause genetic admixture,

<sup>1</sup>Zhejiang Tiantong Forest Ecosystem National Observation and Research Station, Shanghai Key Lab for Urban Ecological Processes and Eco-Restoration, School of Ecological and Environmental Sciences, East China Normal University, Shanghai, China. <sup>2</sup>Shanghai Key Laboratory of Plant Functional Genomics and Resources, Shanghai Chenshan Botanical Garden, Shanghai, China. <sup>3</sup>Novogene Bioinformatics Institute, Beijing, China. <sup>4</sup>CAS Key Laboratory of Tropical Forest Ecology, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Mengla, Yunnan Province, China. <sup>5</sup>Agriculture & Environment Department, Harper Adams University, Newport, UK. <sup>6</sup>Research and Exhibitions Department, South African Museum, Iziko Museums of South Africa, Cape Town, South Africa. <sup>7</sup>Department of Biological Sciences, University of Cape Town, Cape Town, South Africa. <sup>8</sup>Key Laboratory of the Ministry of Education for Coastal and Wetland Ecosystems, College of the Environment and Ecology, Xiamen University, Xiamen, Fujian Province, China. <sup>9</sup>School of Biology, University of Leeds, Leeds, UK. <sup>10</sup>Institute of Eco-Chongming, Shanghai, China. <sup>11</sup>These authors contributed equally: Li-Hua Wu, Jie Zou, Yu-Ting Jiang. ✉e-mail: [rwang@des.ecnu.edu.cn](mailto:rwang@des.ecnu.edu.cn)

likely generating novel genotypes that may assist adaptation to nonnative range<sup>13</sup>.

Two prevailing mechanisms are invoked to explain the adaptation of an alien species to novel environments: rapid evolution and pre-adaptation<sup>5,14</sup>. Heterogeneous environments across the nonnative range generate strong and differentiated selection<sup>15</sup>. High genetic diversity, high phenotypic plasticity, and frequent introgression by hybridization with local relatives (though it may also result in maladaptive phenotypes) contribute to rapid adaptation and genetically differentiated lineages<sup>16–18</sup>. In contrast, pre-adaptation assumes that invasive species have already evolved functional traits enabling their adaptation to a wide range of conditions found outside of the native range. Under a scenario of pre-adaptation strong genetic differentiation among nonnative populations is not expected to occur rapidly<sup>19</sup>.

Control of invasive species requires an understanding of how they adapt to novel conditions through stress tolerance (e.g., drought, flooding, climate, pathogens and natural enemies, and competition)<sup>20–22</sup>. Adaptation frequently has a genetic basis<sup>23,24</sup>. However, the molecular basis of adaptation to nonnative range is known for few invasive species<sup>25,26</sup>. In particular, the adaptation of obligate mutualists relies on efficient utilization of their hosts. Indeed, adaptation has yet to be explored for the obligate mutualists of invasive host plants. This is unfortunate because fitness and invasive potential of introduced plants can increase dramatically upon the successful integration of their obligate mutualists into the local ecosystem. As biology has entered the genomics era, genomic data covering most/all functional genes have made detection of key adaptation-related genes possible and population genomic analyses can identify migration routes and adaptation mechanisms.

The angiosperm plant genus *Ficus* (Moraceae; fig species; *c.* 800 species) and their obligate pollinating wasps (Hymenoptera, Agaonidae) form an obligate nursery pollination mutualism<sup>27</sup>. Each fig species produces enclosed inflorescences (figs) that are pollinated by one or only a few pollinating wasp species. Wasps can only oviposit in the ovules growing inside the figs of their obligate host plants. Therefore, introductions of fig species without their specific pollinating wasps are considered low risk, though mature trees with figs transported together have been recorded in a few early introductions (e.g., introduction of *Ficus microcarpa* from the Philippines to Hawaii<sup>8</sup>). Many fig species have been introduced as fruit crops<sup>28</sup> and ornamental plants<sup>29,30</sup>. However, with the emergence of specific pollinating wasps in the introduced ranges, a few introduced fig species undergo sexual reproduction and have become invasive<sup>31–33</sup>. In addition, host shifts have also occurred in several introduced fig species, e.g., the introduced populations of *Ficus macrophylla* are pollinated by the pollinating wasps of *Ficus rubiginosa*<sup>34</sup>.

In monoecious figs, the female pollinating wasps lay eggs into ovules by inserting their ovipositors into the styles of female florets with drastically varying lengths (Fig. 1a). Pollinating wasps with longer ovipositors can lay eggs into more female florets, increasing their brood size in the host figs. Therefore, ovipositor length is considered as a key trait determining oviposition range and utilization efficiency of the resources from host figs, important for optimizing fitness<sup>35</sup>. Exploring the genes associated with this trait should improve our understanding of how pollinating wasps interact with their hosts. However, in the presence of parasitoids, the advantage of having a long ovipositor is greatly reduced. This is because female florets with longer styles are generally located near the fig wall, where galled ovules are under high risk of being attacked by parasitoids that lay eggs from the outside of figs<sup>35</sup>. In the absence of parasitoids (e.g., in the introduced range), selection on ovipositor length should be relaxed, and the host exploitation potential of the pollinating wasps is predicted to increase by evolving slightly longer ovipositors.

*F. microcarpa* is a monoecious fig species found naturally in tropical and subtropical areas of Asia and northern Australia (Fig. 1b). *F. microcarpa* is only pollinated by an obligate pollinating wasp, *Eupristina verticillata*, which is a complex containing at least three cryptic species<sup>36</sup>. In the last 200 years, *F. microcarpa* has been widely planted as ornamental and street tree in warm temperate and tropical areas (mainly covering the Mediterranean and

the tropical areas in the Americas) outside of its native range<sup>8,33</sup>. Such intentional planting has mostly used seeds or branch cuttings, preventing the introduction of *E. verticillata*, though transportation of mature trees with figs (e.g., from the Philippines to Hawaii<sup>8</sup>) has also been reported in a few cases of introduction. However, *E. verticillata* has been recorded in many introduced *F. microcarpa* populations since the 1970s, facilitating sexual reproduction in introduced *F. microcarpa* trees. Though closely related pollinating wasps might have switched hosts to *F. microcarpa*, all of our field observations have detected *E. verticillata* as the unique pollinating wasp species inside *F. microcarpa* figs in both the native and the nonnative ranges<sup>37</sup>. With the help of long-distance seed dispersers like birds, the population density of this strangler fig tree has increased locally, threatening biodiversity and damaging urban infrastructure<sup>8,33,38</sup> (Fig. 1b). It is unclear how *E. verticillata* emigrated from its native range and rapidly adapted to its new cooler and drier climate. Moreover, while very common in the native range of *F. microcarpa*, parasitoids are rare in the regions where the tree has been introduced. The range of ovules used by female *E. verticillata* has expanded significantly in the nonnative range<sup>35</sup>. These findings suggest that the ovipositors of female *E. verticillata* are getting longer under directional selection in the absence of parasitoids.

In this study, we conducted large-scale sampling and resequenced individuals of *E. verticillata* in both the native and nonnative ranges of this pollinating wasp (Supplementary Table 1; Fig. 1b). Based on the reference genome of this species (diploid females and monoploid males, N chromosomes = 6 for monoploid genome, genome size = 387 Mb, and heterozygosity = 1.1%)<sup>39,40</sup>, we obtained single nucleotide polymorphism (SNP) data and performed population genomic analyses (Fig. 2a). We aimed to (1) determine the composition of cryptic species within *E. verticillata* based on the previous study (Sun et al. 2011) (where three cryptic species were detected); (2) explore colonization from the native range to the nonnative range; and (3) reveal the genes potentially associated with adaptation to the nonnative range. Given the centrality of introduced species to the erosion of biodiversity, we believe that addressing these aims in a tractable system like figs and wasps will enhance our general understanding of invasive potential.

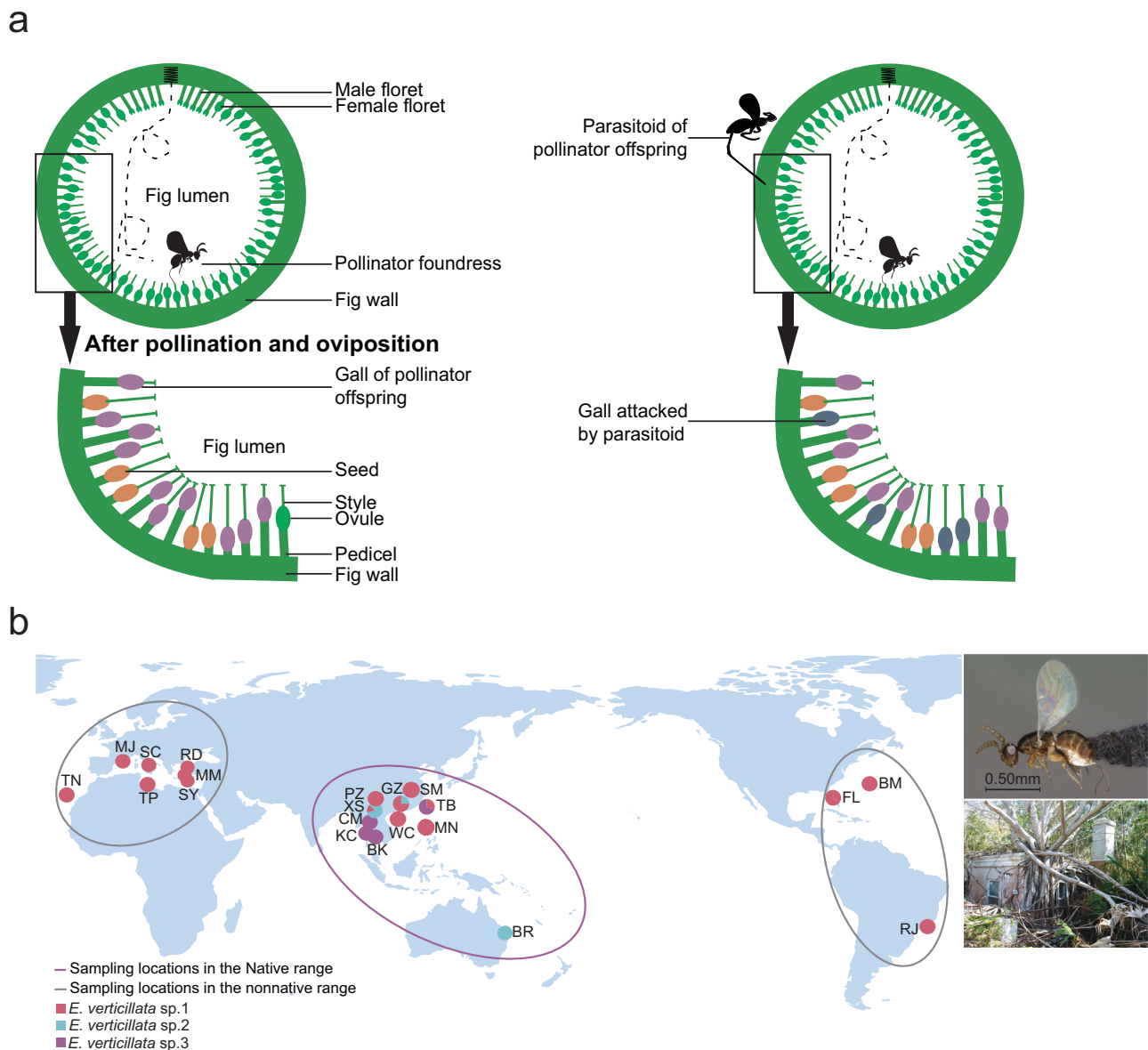
## Results

### Genome resequencing

We collected a total of 127 mature females of *E. verticillata* from 21 sampling locations worldwide, of which 76 and 51 individuals were sampled in its native (11 sampling locations) and nonnative ranges (10 sampling locations), respectively (Fig. 1b; Supplementary Table 2). Using the Illumina platform, we obtained 1.21 TB clean pair-end sequencing data from our sampled 127 individuals (Supplementary Table 2). After mapping clean reads to the reference *E. verticillata* genome<sup>39</sup>, we found extremely low mapping ratio (< 30%) in 26 samples from 4 sampling locations (BK, CM, KC, and TB) in the native range (due to the high genetic divergence between these samples and the reference genome, see below) as well as low average sequencing depth (< 10 ×) and 4 × genome coverage (< 70%) in 6 samples from 4 sampling locations (GZ, PZ, BM, and RJ) in both ranges. We therefore removed these 32 samples for the analysis of genetic structure. For the 95 samples retained, we detected high mapping ratio (mean ± S.E.: 96.1 ± 0.7%), average mapping depth (21.5 ± 0.4 ×), and 4 × genome coverage (88.4 ± 0.8%) (Supplementary Table 2). A total of 16,484,084 SNVs (single nucleotide variants) were detected from these samples, from which various sets of SNPs were generated for the following genomic analyses (see Fig. 2a for the filtering conditions in each analysis).

### Composition of cryptic species

To determine the composition of cryptic species, we uncovered the phylogenetic relationships among the 127 samples using their *COI* gene sequences along with all the 24 *COI* gene sequences reported in Sun et al.<sup>36</sup> as reference sequences (Supplementary Table 3). The phylogenetic tree constructed using the maximum likelihood (ML) approach showed that all our samples were divided into three distinct clades, with all reference sequences



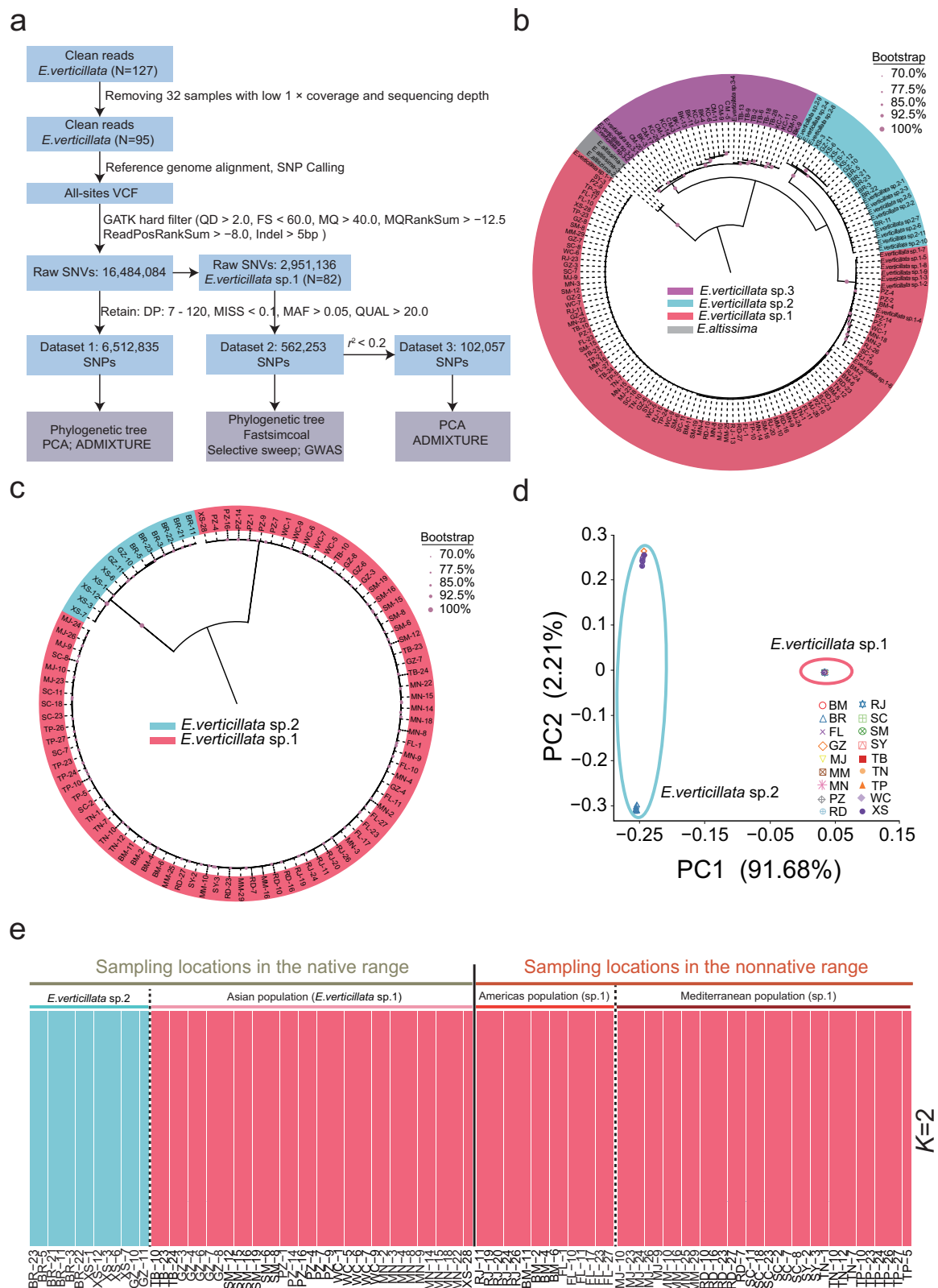
**Fig. 1 | Spatial stratification in monoecious figs and sampling distribution.** Schematic spatial stratification of contents in monoecious figs (a), sampling locations and species/genetic composition (b). **a** Pollinator foundresses enter into the lumen of a fig and lay eggs into the ovules of female florets (which become galls) by inserting their ovipositors into the styles, and thus foundresses with longer ovipositors can utilize female florets with longer styles and leave more offspring (the left figure). At the presence of parasitoids, this advantage will be weakened, because female florets with longer styles are generally located near the fig wall, where galls are under high risk of being attacked by parasitoids that lay eggs from the outside of figs

(the right figure). This figure is modified from Fig. 1 in Wang et al.<sup>35</sup>. **b** The sampling locations in the native and the nonnative ranges are highlighted by different colors, and composition of cryptic species and different genetic groups in *E. verticillata* sp.1 in each sampling location were illustrated by a pie chart (see Supplementary Table 2 for details); two photos show a mature female *E. verticillata* and a *F. microcarpa* tree destroying a house (photos were taken by Simon van Noort and Stephen G. Compton); the map operations were performed using ArcGIS v10.8 (<http://www.esri.com/>).

from each reported cryptic species assigned into a single clade (Fig. 2b), providing clear evidence for the presence of three cryptic species. We found that one cryptic species (*E. verticillata* sp.3) contained only the 26 samples with an extremely low mapping ratio, which diverged from the common ancestors of the other two cryptic species (Fig. 2b; Supplementary Table 2). Intriguingly, all samples from the nonnative range (51) belonged to a single cryptic species (*E. verticillata* sp.1), which also included 37 samples from the native range, and the remaining 13 samples (from BR, GZ, and XS) were assigned to the other cryptic species (*E. verticillata* sp.2) (Fig. 2b; Supplementary Table 2).

Moreover, we investigated the genetic structure using genomic SNP data (6,512,835 SNPs, Dataset 1; of which 1,615,537 were found to be heterozygous at least in one sample) to further test the assignment of the

95 samples (Fig. 2a). The ML phylogenomic tree revealed two major clades that were completely consistent with the clade assignments using *COI* gene sequences (82 and 13 samples belonging to *E. verticillata* sp.1 and sp.2, respectively) (Fig. 2c). The first axis of principal component analysis (PCA) accounted for 91.7% of genetic variations and also assigned the samples of *E. verticillata* sp.1 and sp.2 into two different groups (Fig. 2d). Genetic clustering analysis using ADMIXTURE divided the samples from different cryptic species into two distinct clusters when  $K = 2$  (Fig. 2e), consistent with the clade assignments using *COI* gene sequences and SNPs. Because the cross-validation (CV) errors of different  $K$  values (from 3 to 7) did not differ meaningfully (Supplementary Fig. 1a), we performed additional genetic clustering analysis from  $K = 3$  to  $K = 8$ . The results revealed the absence of genetic admixture between the two clusters with increasing  $K$  value



**Fig. 2 | Variant filtering workflow and delimitation of cryptic species.** Filtering workflow from single-nucleotide variants (SNVs) to single nucleotide polymorphism (SNP) datasets for different analyses (**a**), and results showing the composition of cryptic species using both sequences of *COI* gene (**b**) and whole-genome SNPs (**c–e**). **a** Datasets and the downstream analyses are highlighted by blue and purple boxes (see Methods). DP: read depth; MISS: missing rate in samples; MAF: minor allele frequency; QD: quality by depth; FS: FisherStrand annotation to indicate strand bias; MQ: mapping quality; MQRankSum: rank-sum test for mapping qualities; ReadPosRankSum: rank-sum test for relative positioning of reference versus alternate alleles within reads;  $r^2$ : linkage

disequilibrium. **b** GenBank accession IDs for the downloaded *COI* sequences of the three cryptic species and the outgroup species (*E. altissima*) used in the maximum likelihood phylogenetic (ML) tree are listed in Supplementary Table 3. **c–e** Results from ML tree, principal component analysis (PCA) and Admixture ( $K = 2$ ) show that only *E. verticillata* sp.1 were found in the sampling locations in the nonnative range, with one (Asian) and two (The Americas and the Mediterranean) populations in the sampling locations in the native and nonnative ranges, respectively. Full names of sampling locations are listed in Supplementary Table 2.



(Supplementary Fig. 1b). Therefore, consistent with the results based on *COI* gene sequences, the genetic structure of the 95 samples revealed the presence of two cryptic species, of which *E. verticillata* sp.2 was only present in the native range whereas *E. verticillata* sp.1 was found in both the native and the nonnative (the Mediterranean and Americas) ranges. In addition, we also conducted SNP calling using the genomic sequencing data from 121 samples (including the 26 samples assigned to *E. verticillata* sp.3), and a total of 3,431,957 SNPs were identified (see Supplementary Fig. 1c). We then constructed a ML phylogenomic tree based on 133,039 fourfold degenerate (4DTV sites) and found consistent clade assignments with those using *COI* gene sequences, supporting the presence of three cryptic species (Supplementary Fig. 1c).

The percentage of heterozygous sites per sample ranged from 0.04% to 4.81% (means  $\pm$  S.E.:  $1.10 \pm 0.13\%$ ). The observed heterozygosity per site ranged from 0.01 to 0.27, which was lower than the expected heterozygosity (Supplementary Fig. 2), suggesting high level of inbreeding in *E. verticillata*, consistent with the results from other pollinating wasp species<sup>41</sup>.

### Genetic structure of *E. verticillata* sp.1

To explore the genetic structure of *E. verticillata* sp.1, we performed PCA, genetic clustering, and phylogenomic analysis using Datasets 2 and 3 (Fig. 2a). The first two axes of PCA assigned the 82 samples into three groups (Fig. 3a). The largest group (47 samples, genetic group I) contained all samples from the sampling locations in the nonnative range except two from Florida (FL) and two samples from the sampling locations in the native range (Manila (MN)), and the second largest group included 26 samples from the sampling locations in the native range (genetic group II), with the other 7 samples from the sampling locations in the native range (from the Manila sampling locations) and 2 samples from FL assigned into the smallest group (genetic group III) (Fig. 3a; Supplementary Table 2).

We found consistent results from the maximum likelihood phylogenetic tree, where the two major clades comprised the samples belonging to genetic group I and the other two genetic groups (with samples of genetic group III forming a subclade), respectively (Fig. 3b). As similar CV errors of *K* values from 4 to 8 were detected (Supplementary Fig. 3a), we conducted genetic clustering analysis using these different *K* values. When *K* = 4, ADMIXTURE divided the samples of genetic group I into two clusters and those of the other two genetic groups into two distinct clusters (Fig. 3c). These results showed the presence of three distinct genetic groups within *E. verticillata* sp.1. Our sampling locations in the nonnative range were dominated by genetic group I that was rare in the sampling locations in the native range (present only in MN), and only one nonnative sampling site in Florida contained samples from both genetic groups I and II.

We found that 533,772 of the 562,253 SNPs in Dataset 2 were heterozygous at least in one sample, and the percentage of heterozygous sites per sample ranged from 1.12% to 24.06% (means  $\pm$  S.E.:  $9.74 \pm 0.60\%$ ). The observed heterozygosity per site ranged from 0.02 to 0.50, which was lower than expected heterozygosity (Supplementary Fig. 3b).

### Demographic history and migration route of *E. verticillata* sp.1

To determine the most likely scenario of the emigration of *E. verticillata* sp.1 from the sampling locations in the native range to those in the nonnative range, we utilized the composite likelihood approach using Fastsimcoal based on Dataset 2. We developed 10 demographic models (setting a single ancestral population) with only population split events or the combination of split and admixture events at the presence/absence of migration from Asian to the Americas and the Mediterranean populations (Supplementary Fig. 4a; see Methods). By comparing the Akaike Information Criterion (AIC) value of the best simulation run in each model, we found that model 1 (three modern populations derived from one population split event in the ancestor population without migration among modern populations) was the best-fitting model (Supplementary Table 4). Results of parameter estimation showed that both the American and the Mediterranean populations diverged in parallel from their common ancestor population 520.5 (95% CI: 482.3–558.7) generations ago, with similar effective population sizes that

were far smaller than those of the Asian population (Fig. 3d; Supplementary Fig. 4b; Supplementary Table 5). As *E. verticillata* normally experiences four generations per year<sup>42</sup>, these two nonnative populations of *E. verticillata* were estimated to be formed at c. 130 years ago (at c. 1880s), earlier than the first record of this species in its nonnative range (at 1930s)<sup>8</sup>.

### Positively selected genes associated with adaptation to the nonnative range

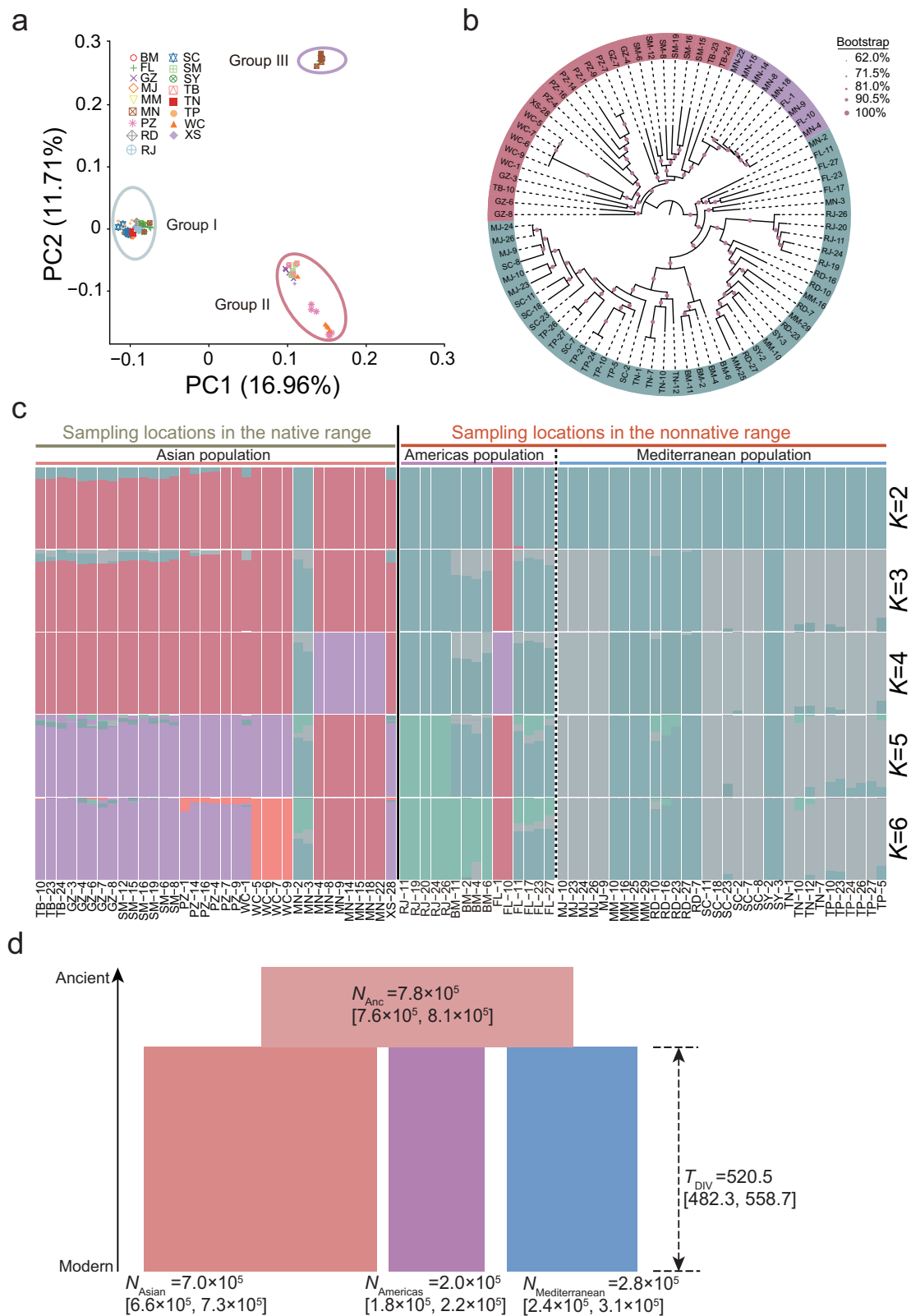
To further explore the genes likely contributing to the adaptation of genetic group I of *E. verticillata* sp.1 to the sampling locations in the nonnative range we performed selective sweep analysis using four tests ( $F_{ST}$ ,  $\pi$  ratio, XP-CLR, and Tajima's *D*) to screen the genes with signatures of positive selection in genetic group I when compared with the other two genetic groups (based on Dataset 2). Using standards for choosing the genomic regions falling into the top 5% values for  $F_{ST}$ ,  $\pi$  ratio, and XP-CLR, and negative values for Tajima's *D*, we found signals of positive selection in 946, 1060, 2626, and 496 genes, respectively (Fig. 4a–d). A total of 75 positively selected genes were supported by all four statistics (Fig. 4e; Supplementary Table 6).

Among these 75 genes, we found three genes (*spitz* (*spi*), GATOR complex protein *WDR24* (*wdr24*), and Protein E(sev)2B (*drk*) that were relevant to the development of insect organs (as their functions have been validated experimentally in *Drosophila melanogaster* and *Leptinotarsa decemlineata*). Specifically *spi* plays an important role in the development of insect genitals<sup>43</sup>, while *wdr24* and *drk* participate in TOR signaling pathways<sup>44,45</sup> (Supplementary Table 6). These three genes are likely to be linked with the longer ovipositors of genetic group I (Fig. 5a; Supplementary Table 6). Moreover, three positively selected genes (glucose dehydrogenase (*Gld*), octopamine receptors (*OctRs*), and peroxisome biogenesis factor *PEX14* (*PEX14*) participate in sperm storage and utilization in female *D. melanogaster*<sup>46</sup>, maintenance of ovarian phenotype, and regulation of hormone levels, ovarian development of female *D. melanogaster*<sup>47</sup>, and egg numbers of female *Nilaparvata lugens*<sup>48</sup> (Supplementary Table 6). Given that the brood size (represented by the number of offspring) of *E. verticillata* was significantly higher in the nonnative range than in the native range (Supplementary Fig. 5a; Supplementary Table 6), these genes may be related to the change of brood size within *E. verticillata* sp.1.

As the mean monthly temperatures in the sampling locations in the nonnative range are lower than the sampling locations in the native range (Supplementary Fig. 5b), *E. verticillata* sp.1 in the nonnative range is expected to suffer a degree of cold stress. We found two positively selected genes (ionotropic receptor 25a (*Ir25a*)) that can mediate cool sensing of *D. melanogaster* and regulate circadian clock reset by temperature<sup>49,50</sup> (Supplementary Table 6). Furthermore, three positively selected genes (peroxisomal acyl-coenzyme A oxidase 3 (*ACOX3*), ATP synthase subunit O, mitochondrial (*ATPsynO*), and organic solute carrier partner 1 (*OSCP1*)) involved in fatty acid metabolism of *Caenorhabditis elegans*<sup>51</sup> and ATP synthesis<sup>52</sup>, energy metabolism and ATP production of *D. melanogaster*<sup>53</sup> (Supplementary Table 6). These four genes thus are likely to assist genetic group I in achieving cold tolerance. Therefore, the genes mentioned above revealed the molecular footprints likely contributing to the adaptation of genetic group I to the nonnative range. To validate the annotation of positively selected genes, we selected the gene *Gld* by synthesizing its recombinant protein and conducted in vitro enzymatic activity assay (see Methods). The product of the in vitro reactions identified by UPLC was identical to the standards of D-glucono-1,5-lactone (Fig. 4f), confirming its enzymatic function.

### Genome-wide association study (GWAS) on ovipositor length

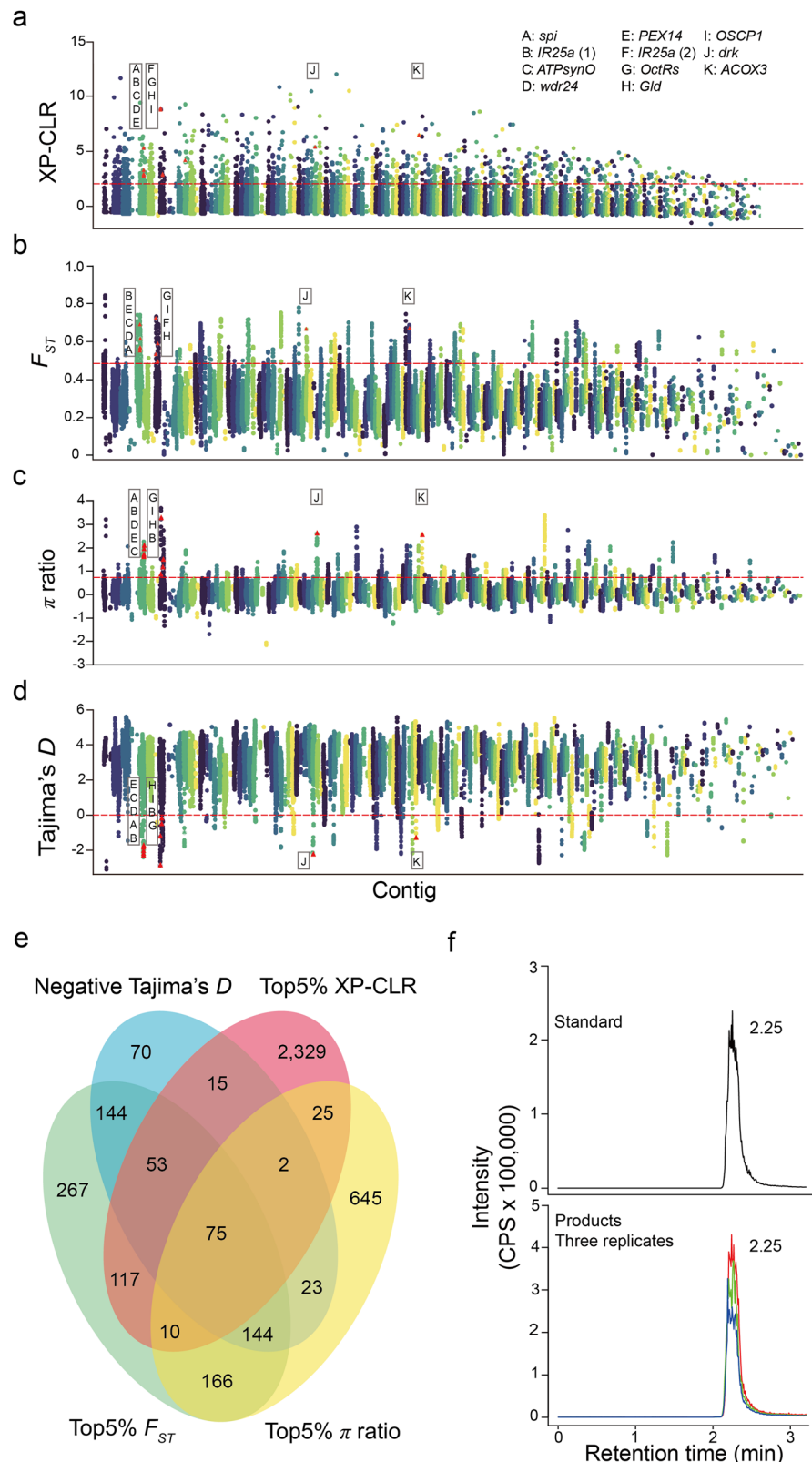
The ovipositor length of pollinating fig wasps is tightly related with their ability to oviposit eggs into the ovules with various lengths of styles inside figs (Fig. 1a). Variation in this trait has been considered to reflect how many ovules wasps can utilize and by association to be tightly associated with brood size<sup>35</sup>. However, it is still unclear whether the ovipositor length of the predominant genetic group was longer in comparison to the other groups. Further, the molecular basis of variation in ovipositor length is unknown.



**Fig. 3 | Genetic structure and demographic history of *E. verticillata* sp.1.** Genetic structure of *E. verticillata* sp.1 based on whole-genome SNPs (a–c) and results of demographic analysis (d). **a** PCA of genetic variance among all samples. **b** The phylogenetic tree constructed using the maximum likelihood based on whole-genome SNPs. **c** Genomic ancestry proportions of each sample from  $K=2$  to  $K=6$

using ADMIXTURE. **d** Demography under the best-fitting model. Parameter estimations and 95% confidence intervals are shown for the effective population sizes of the ancestral ( $N_{Anc}$ ) and the derived three populations ( $N_{Asian}$ ,  $N_{Americas}$ , and  $N_{Mediterranean}$ ) as well as the time when population split occurred among the derived populations ( $T_{DIV}$  (years)) (see Supplementary Table 5).

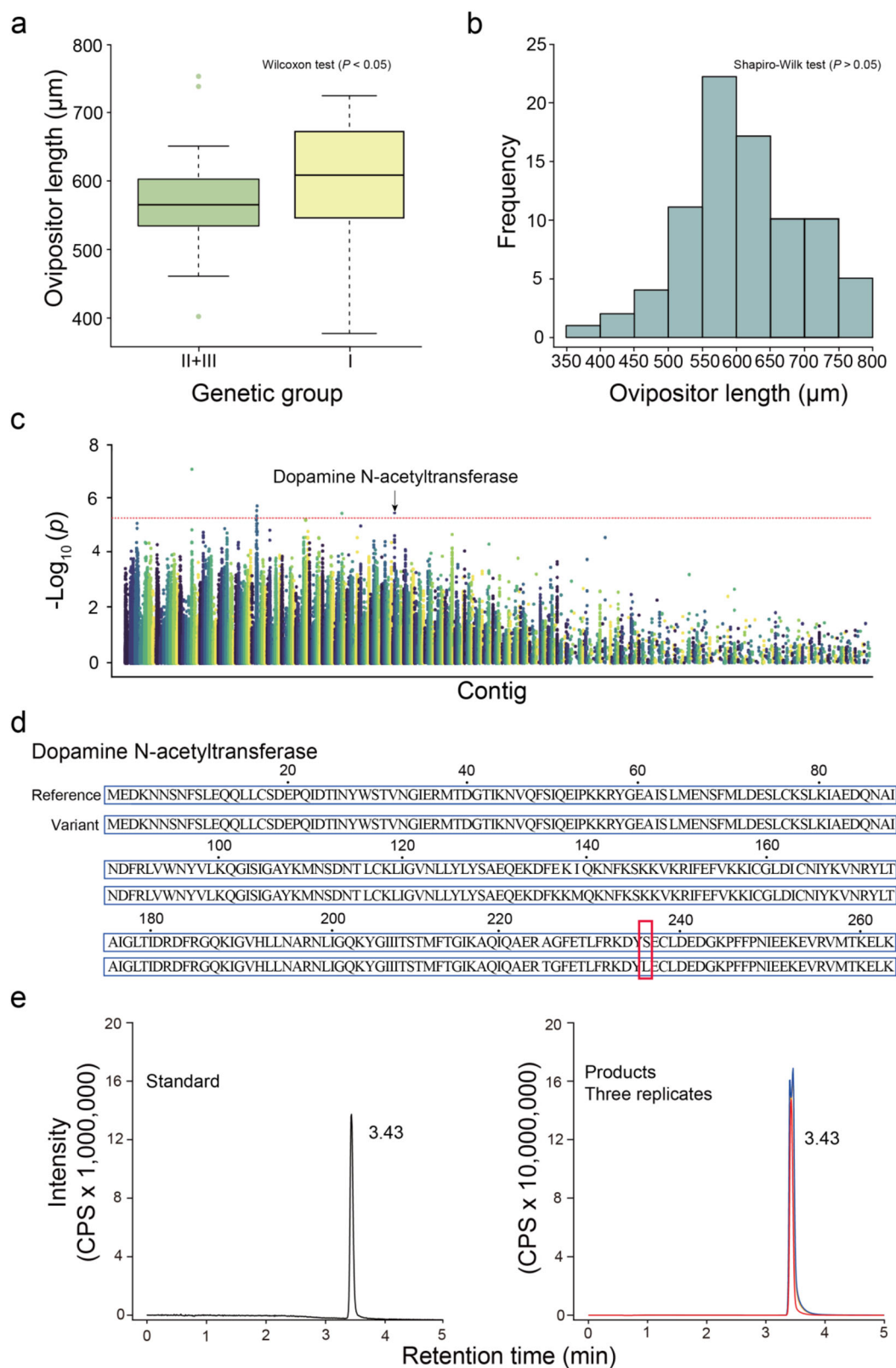
**Fig. 4 | Results of selective sweep analysis and functional validation of the *Gld* gene.** Results of selective sweep analyses (a–e) and in vitro functional validation of the glucose dehydrogenase (*Gld*) (f). **a–d** Manhattan plots showing the distribution of XP-CLR,  $F_{ST}$ ,  $\pi$  ratio and Tajima's  $D$  values in each contig. The non-overlapping windows with top 5% values for XP-CLR,  $F_{ST}$  and  $\pi$  ratio (above the red dashed lines) and negative values for Tajima's  $D$  were considered (below the red dashed line) as regions under positive selection. The genes that are supported by all the four statistics and relevant with adaptation to the nonnative range are marked (Supplementary Table 6). **e** Venn diagram of positively selected genes revealed by the four statistics (Supplementary Table 6). **f** Standard of D-glucono-1,5-lactone (at the concentration of 100 ng/ml) and reaction products (treatments with enzyme added for three replicates) are identified using UPLC (see Methods).



Using linear mixed models, we found that the ovipositor length of samples from genetic group I of *E. verticillata* sp.1 ( $638.53 \pm 12.39$   $\mu$ m,  $N = 47$ ) were significantly longer than those from the other two genetic groups ( $580.10 \pm 12.41$   $\mu$ m;  $N = 35$ , Fig. 5a; Supplementary Table 7). This difference may explain the higher brood size of genetic group I when the parasitoids of *E. verticillata* was rare, thereby facilitating its adaptation to the nonnative range. The ovipositor lengths of our samples were independent of their body

lengths, the size of their natal figs, and the number of galled ovules in the natal figs (Supplementary Table 7).

To explore the genes associated with ovipositor length, we performed GWAS to link the ovipositor lengths of samples from *E. verticillata* sp.1 (which were normally distributed; Shapiro–Wilk test:  $W = 0.971$ ,  $P = 0.06$ ) (Fig. 5b) with their genomic SNPs (Dataset 2 in Fig. 2a). The GWAS based on ovipositor length (at the corrected threshold of  $-\log_{10}(P) = 5.41$ ) resulted



**Fig. 5 | Results of Genome-wide association study (GWAS) using ovipositor length.** Distribution of ovipositor length in samples of *E. verticillata* sp.1 (**a**, **b**), results of genome-wide association studies (GWAS) (**c**, **d**), and results of in vitro functional assay of the variant dopamine N-acetyltransferase (*Dat*) (**e**). **a** Boxplot showing the difference in ovipositor length between the samples from the native and the nonnative range. The center lines, box edges, and whiskers represent medians, the 25% and 75% quartiles, and the upper and lower distribution of 1.5 times of quartile range. **b** Histogram illustrating the frequency distribution of ovipositor

length. **c** Manhattan plot showing the SNPs significantly associated with the variation of ovipositor length in each contig (Supplementary Table 8). **d** Amino acid sequences in the reference and the variant dopamine N-acetyltransferase. The variant amino acid caused by the SNP located in exon is highlighted by a red frame. **e** Standard of N-acetyldopamine (at the concentration of 100 ng/ml, the left plot) and reaction products (treatments with enzyme added for three replicates, the right plot) are identified using UPLC (see Methods).



in a total of six associated SNPs (Fig. 5c; Supplementary Fig. 5c). Among these SNPs, one was a nonsynonymous site located in the exon of a dopamine N-acetyltransferase gene (Fig. 5d), which has been reported to participate in modulating cuticle pigmentation and female fecundity<sup>54</sup> (Supplementary Table 8). To validate the function of the variant, we conducted *in vitro* enzymatic activity assay (see Methods). The product of the *in vitro* reactions identified by UPLC was identical to the standards of N-acetyldopamine (Fig. 5e), confirming its enzymatic function. The other five SNPs were located in intergenic regions (Supplementary Table 8). No functional genes were in linkage disequilibrium with these six SNPs because  $r^2$  decreased rapidly with genomic distance (Supplementary Fig. 5d).

## Discussion

Exploring the adaptation of obligate mutualists to novel environments can improve our understanding of biological invasion<sup>2</sup>. Here, we focused on the obligate pollinating fig wasp of an invasive *Ficus* species, finding that only one of the three identified cryptic species of the pollinator successfully colonized in the sampling locations in the nonnative range. Interestingly, a single genetic group within this cryptic species occupied most of the nonnative range, most likely colonizing the Americas and Mediterranean independently around 130 years ago. Furthermore, selective sweep analyses revealed a few positively selected genes that may contribute to survival in the nonnative range, and we identified a SNP located in the exotic region of a gene (dopamine N-acetyltransferase) significantly linked with ovipositor length, a key trait related with pollinator brood size<sup>35</sup> (Fig. 1a). Together, our results illustrate how the mutualist dispersed from its native range and pinpoint the molecular basis of adaptation.

Consistent with previous study<sup>36</sup>, our results support the presence of three cryptic species within *E. verticillata*. The low mapping ratio to the reference genome for the cryptic wasp species mainly found in Thailand is probably due to the deep genetic divergence from the reference species genome, which was sampled in Southeast China<sup>39</sup> (likely to be *E. verticillata* sp.1) (Fig. 1b). The fact that only one cryptic species was sampled in the sampling locations in the nonnative range may be a consequence of either founder effects caused by a limited number of migration events<sup>55</sup>, or due to competitive exclusion of other species<sup>56</sup>. The presence of only two genetic groups in *E. verticillata* sp.1 in the nonnative range, and limited evidence for deliberate intentional introductions suggest that accidental introductions by humans were the most likely translocation model. Moreover, though the number of our sampling locations were not sufficient to completely cover the native range, the fact that the population split recovered was earlier than the first record of *E. verticillata* in the nonnative range suggests a very small initial population size. Therefore, our results support the scenario that most translocated *F. microcarpa* trees did not carry figs with pollinating wasps and that *E. verticillata* populations sampled in the nonnative range likely came from a single part of the native range. Meanwhile, much lower observed heterozygosity than expected heterozygosity is likely due to low numbers of foundress wasps per fig leading to high levels of inbreeding<sup>57</sup>.

The inferred migration route of *E. verticillata* sp.1 largely agrees with the earliest reports of *F. microcarpa* being introduced into the USA from the Philippines<sup>31</sup>, and its early presence in the Mediterranean might have been overlooked. Moreover, frequent immigrations can effectively relieve founder effects and form a large effective population size<sup>58</sup>, but as the nonnative populations were established quite recently, it is difficult to measure gene flow between the native and the nonnative populations based on our genomic data. Nevertheless, because *F. microcarpa* has been widely used as a street tree, migrations of the pollinator were likely via international trade<sup>8</sup>. Due to the host specificity of *E. verticillata*, prevention of introducing *F. microcarpa* trees with figs should be an efficient measure to control further invasion of this plant in its introduced areas far away from the invaded places.

Though rare introductions may cause the predominance of genetic group I in the sampling locations in the nonnative range, this genetic group shows adaptation to local climatic conditions, and such adaptation is often

tightly linked with reproduction and responses to stresses<sup>59,60</sup>. Ovipositor length is a key trait relevant to fig wasps' brood size, and it determines oviposition range of female pollinating wasps (i.e., how many female florets can be used inside host figs)<sup>35</sup>. Our results provide evidence that the longer ovipositors in genetic group I are related to higher fecundity than other genetic groups (Supplementary Fig. 5a). However, genetic group I was rare in the sampling locations in the native range. Although this genetic group may utilize more ovules using long styles, the ovules with longer styles are normally closer to the fig wall and larvae living in these galled ovules suffer a higher risk of being attacked by parasitoids that lay eggs from the outside surface of figs (Fig. 1a)<sup>35</sup>. Therefore, the fitness of genetic group I may be restricted in the native range (where their parasitoids were abundant) but far increased in the nonnative range (where parasitoids were rare or even absent in most areas)<sup>37</sup>. Notably, GWAS detected a nonsynonymous SNP in the gene dopamine N-acetyltransferase, which has been considered to contribute to organ development<sup>54</sup>. Our functional validation showed that the variant did not change the function of this gene, and it is possible that this genetic variation facilitates ovipositor growth by elevating the efficiency of enzyme.

Seven positively selected genes supported by all selective sweep analyses were likely relevant to pollinating wasps' reproduction, suggesting that the utilization of host figs play a crucial role in the adaptation to the nonnative range, consistent with other insect species<sup>61,62</sup>. We also found four positively selected genes likely to contribute to cold tolerance, reflecting the cooler climate in the sampling locations in the nonnative range<sup>8</sup>. As other climate and environment factors vary drastically among sampling locations, it is difficult to evaluate the importance of other positively selected genes in adaptation. Furthermore, the maintenance of genetic homogeneity across different sampling locations in the nonnative range, and the short time of immigration into the nonnative range, mean that the selected genes underwent selection in advance of the global introduction of *F. microcarpa*, supporting the hypothesis that pre-adaptation, rather than rapid evolution, enables the colonization and expansion of genetic group I.

Whether a species adapts to dramatic environmental changes determines its fate in the Anthropocene<sup>63,64</sup>. Exploring the mechanisms underlying range expansions of alien species can provide not only insights into how they survive under novel selection regimes, but also the genetic resources that can be exploited for improving species' resistance against various stresses induced by global changes<sup>65–67</sup>. Our study unravels how the obligate mutualists of invasive plants migrate and adapt to novel environments, highlighting the necessity of identifying key genes facilitating biological invasions.

## Methods

### Sampling

Between December 2010 and July 2018, we carried out field surveys to sample mature females of *E. verticillata* from 21 locations in the tropical and subtropical areas around the world (Fig. 1b). Given that *E. verticillata* is not only indigenous to the natural distribution range of *F. microcarpa* (from Asia to Australia) but also can naturally disperse to the adjacent planted sites of this fig species<sup>8,33,36</sup>, we divided all sampling locations from Asia and Australia into the native range of *E. verticillata* and allocated all sampling locations (where *F. microcarpa* is introduced) from the Americas and the Mediterranean into its nonnative range (Fig. 1b; Supplementary Table 2).

In each sampling location, we surveyed 5–10 *F. microcarpa* trees (with the minimum interval being 50 m) and collected at least 5 mature figs for each surveyed tree. To ensure genetic independence among samples, we only haphazardly selected one fig for each surveyed tree and collected all mature *E. verticillata* individuals by dissecting all galled ovules in this fig. We then randomly chose one mature female of *E. verticillata* from this fig for sequencing (i.e., one sample per tree) (Supplementary Table 1). To exclude the possibility of collecting nonpollinating fig wasps that are common in the native range of *F. microcarpa* but are rare in its introduced range<sup>35</sup>, all samples were carefully identified under a Nikon SMZ25 microscope (Nikon, Japan).

## DNA extraction and high-throughput sequencing

We extracted the genomic DNA of our samples using the E.Z.N.A.<sup>®</sup> Tissue DNA kit (OMEGA, USA) according to the manufacturer's instructions. The quality and concentration of extracted genomic DNA were then measured via NanoDrop2000 Ultra Micro Spectrophotometer (Thermo Scientific, USA), with the integrity of DNA being assessed by electrophoresis using a 1% agarose gel.

For each sample, an Illumina pair-end library with the insert size of 350 bp was prepared, and 150 bp paired-end sequencing (PE 150) was performed using an Illumina HiSeq 4000 platform. The quality control of raw data was performed using fastp v0.19.7<sup>68</sup>, which first trimmed adapters and then removed low-quality reads according to the following filtering criteria: (1) the proportion of unidentified nucleotides (N) larger than 10%; and (2) more than 50% nucleotides having a Phred score  $\leq 5$ . The yielded clean data ( $N = 127$ ) were used for data mapping to the reference genome.

## SNP calling and annotation

We mapped the clean reads of each sample ( $N = 127$ ) to the high-quality reference genome assembly of *E. verticillata*<sup>39</sup> (GSA database (<https://ngdc.cncb.ac.cn>) under the accession number GWHALOE000000000) using BWA v0.7.8<sup>69</sup>, setting “mem -t 4 -k 32 -M”. To ensure that sufficient sequencing data were available for SNP calling, the samples with a mapping ratio lower than 70% were excluded from all the analyses of genetic structure. After the data mapping, we used SAMtools v1.9<sup>70</sup> for reads sorting and the removal of duplicated reads to generate the Binary Alignment Map (BAM) files. To guarantee the accuracy and the representativity of variation detection, we removed the samples with average sequencing depth  $< 10 \times$  and  $4 \times$  genome coverage  $< 70\%$ . SNVs were called for the retained samples ( $N = 95$ , and 32 samples were excluded) using the HaplotypeCaller tool in GATK v4.1<sup>71</sup>, based on the filtering criteria of  $QD > 2.0$ ,  $FS < 60.0$ ,  $MQ > 40.0$ ,  $MQRankSum > -12.5$ , and  $ReadPosRankSum > -8.0$  (Fig. 2a). In addition, the SNVs within 5 bp of an indel were also removed. We further identified SNPs according to the standards of  $7 \leq DP \leq 120$ ,  $MISS < 0.1$ ,  $MAF > 0.05$ , and  $QUAL > 20.0$ . All genes with SNPs were functionally annotated using ANNOVAR<sup>72</sup>. Using the same approach of SNP calling, various datasets of SNPs were generated for different analyses ( $N = 95$  for identification of cryptic species, and  $N = 82$  population genomic analysis for *E. verticillata* sp.1) (see Fig. 2a).

## Identification of cryptic species based on COI gene sequences

To test whether multiple cryptic species within *E. verticillata* existed within our samples and uncover the composition of cryptic species, we adopted the mitochondrial gene cytochrome c oxidase I (COI), which is widely employed as a DNA barcoding gene in species identification of obligate pollinating wasps<sup>73</sup>. The COI gene of each sample was amplified using the universal primer LepF1 and LepR1<sup>74</sup>. Bidirectional Sanger sequencing of the amplified products was performed using an ABI 3730 Genetic Analyser (Applied Biosystems<sup>™</sup>, Foster City, California). To identify cryptic species based on the published DNA sequences, we haphazardly chose the COI gene sequences of three individuals belonging to each of the three previously characterized cryptic species<sup>36</sup> (a total of 24 COI gene sequences were selected) (Supplementary Table 3). After sequence alignment, we constructed a phylogenetic tree using the maximum likelihood (ML) approach in IQ-TREE v2.3.6<sup>75</sup> with the TIM3 + F + G4 model and 1000 bootstraps. To determine the evolutionary relationships among different cryptic species, three COI gene sequences from *Eupristina altissima*<sup>76</sup>, a closely related species to *E. verticillata*, were selected as the outgroup (Supplementary Table 3).

## Genetic structure based on genomic SNP data

As COI gene alone may not accurately reflect species boundaries due to the possibility of pseudogenes, the chance of introgression and the fact that mitochondrial DNA does not recombine, it is necessary to use genomic data to provide additional evidence for species delimitation. We investigated the genetic structure based on genomic SNP data of all retained samples

( $N = 95$ ) after the filtering mentioned above (Fig. 2a). An ML tree was constructed to unravel the phylogenetic relationships among samples using IQ-TREE<sup>75</sup> with the TVMe+I + R5 model and setting 1000 bootstraps. The phylogenetic tree was visualized using Figtree v1.4. We carried out principal component analysis (PCA) using GCTA v1.24.2<sup>77</sup> to calculate eigenvectors and eigenvalues, and the top two principal components were extracted for the assignment of our samples. Genetic clustering analysis was performed using ADMIXTURE v1.23<sup>78</sup>. The optimal number of clusters ( $K$ ) was determined by running ADMIXTURE from  $K = 1$  to  $K = 8$  and choosing the  $K$  value with the minimum cross-validation error. When assigning samples into different genetic clusters, the standard errors of parameters were calculated using bootstrapping (bootstrap = 200). The percentage of heterozygous, the expected and observed heterozygosity were calculated using Plink v1.07<sup>79</sup>.

For the dominant cryptic species within *E. verticillata*, we used phylogenetic analysis (following the parameters: TVM + F + R9 model, and bootstraps = 1000), PCA, and genetic clustering analysis (with the same parameter settings as mentioned above) to delineate genetic structure for determining the comparisons in selective sweep analyses. We measured linkage disequilibrium (LD) using PopLDdecay v3.40<sup>80</sup>. We then used PLINK to remove SNPs with linkage disequilibrium coefficient ( $r^2$ ) larger than 0.2, with the nonoverlapping slide window size of 10 kb and the step size of 1 SNP. The filtered SNPs were used for PCA and genetic clustering analysis, and the phylogenetic analysis was based on all SNPs (Fig. 2a).

## Demographic history and migration route

To determine the most likely demographic history of *E. verticillata* sp.1 and its migration route from the sampling locations in the native range to those in the nonnative range, we employed the composite likelihood approach using Fastsimcoal v2.8<sup>81</sup>. This analysis included all SNPs detected in this cryptic species (Dataset 2 in Fig. 2a). The VCF files were used to generate the site frequency spectrum (SFS) using easySFS.py (<https://github.com/isaacovercast/easySFS#easysfs>), with ‘-i -p --preview’ first and then ‘-i -p --proj 60, 24, 54’ for maximizing the number of segregating sites during the convert from the VCF to SFS. The SFS files were used as the observed data for Fastsimcoal simulations. We then designed 10 candidate demographic models to elucidate the evolutionary process forming the three modern populations (Asia, the Americas, and the Mediterranean populations) from their common-ancestor population (Supplementary Fig. 4a). In model 1, we assumed that all three modern populations were derived from one common-ancestor population with a splitting event at the same time without post-divergence migration among modern populations. In models 2–5, we assumed that the modern populations were generated by two population split events or one admixture event after an ancient population split. Models 6–10 kept the same settings as the first five models, but with additional parameters considering the migration from Asian to the Americas and the Mediterranean populations (6–8) and migration from source populations to admixture populations (9–10).

We conducted simulations by running each model for 100 times to find the best run (the smallest difference between the maximum possible likelihood and the obtained likelihood) using the code ‘-m -o -C 10 -n 2000000 -L 60 -s0 -M -c 48 -B 48 -q’. The best-fitting model was determined according to the Akaike Information Criterion (AIC)<sup>82</sup> values with the best run of each model. Then, bootstrap parameters with 95% confidence intervals were estimated according to the 100 simulations based on the \*.pv file using the parameters from the best-fitting model. We set the mutation rate as  $2.8 \times 10^{-8}$  per site per year<sup>83</sup> and the generation time of 0.25 year<sup>42</sup>. Specific parameters were set for effective population sizes, time of splitting/admixture event, and migration rate between populations (Supplementary Fig. 4a).

## Measurement of key traits and climate data extraction

We measured the ovipositor lengths (ovipositor sheaths were removed prior to measurement) and body length of all samples ( $N = 127$ ) using a dissecting microscope (SMZ-25 (Nikon, Japan)) following the approach used in Nefdt

and Compton<sup>84</sup>. The lengths and widths of the natal figs ( $N = 127$ ) where our samples were collected were measured to the nearest 0.2 mm using a dissecting microscope with an eyepiece graticule, and the volumes of the figs were calculated using the formula for an ellipsoid, which was their approximate shape<sup>8</sup>. In addition, we counted the number of female florets and galled ovules in each fig. For the samples belonging to the dominant cryptic species ( $N = 82$ ), we tested if ovipositor length in the predominant genetic group in the sampling locations in the nonnative range was significantly different from that in other genetic groups. We also examined if ovipositor length was independent of body length, fig size, and number of galled ovules. In addition, we tested the correlation between the number of galled ovules and the number of female florets. All statistical analyses were performed using linear mixed models (LMMs) in R package ‘nlme’ v3.1<sup>85</sup> setting sampling locations as the random effect.

We then tested whether or not the brood size of *E. verticillata* females varied between the sampling locations in the native and the nonnative ranges. We selected figs from Wang et al., (2019) (where all galled ovules in each fig were dissected and all fig wasps (both pollinating wasps and non-pollinating fig wasps) were identified) following the standards: (1) figs only containing *E. verticillata* offspring; and (2) figs including remains of only a single *E. verticillata* foundress. We obtained a total of 94 figs (52 and 42 in the native and the nonnative ranges (Supplementary Table 1)) with only *E. verticillata* offspring and only one *E. verticillata* foundress. As a proxy for brood size, we counted the number of *E. verticillata* offspring in these 94 figs, and tested the difference between sampling locations in different ranges using LMMs.

To compare the temperature between the sampling locations in the native and the nonnative ranges (which may be associated with the adaptation of the dominant cryptic species), we downloaded the climate data from 1970 to 2000 with a resolution of 2.5 min ( $\sim 4.5 \text{ km}^2$ ) from the climate database WorldClim v2.1 (<https://www.worldclim.org/data/worldclim21.html>) and extracted the mean monthly temperatures of all sampling locations using the R package ‘raster’.

### Selective sweep analyses

To screen for positively selected genes likely associated with adaptation of the dominant cryptic species within *E. verticillata*, we compared the dominant genetic group ( $N = 47$ ) in the nonnative range with other genetic groups ( $N = 35$ ) using four selective sweep statistics ( $F_{ST}$ ,  $\pi$  ratio, Tajima’s  $D$  and the cross-population composite likelihood ratio (XP-CLR)).  $F_{ST}$ ,  $\pi$  ratio, and Tajima’s  $D$  were calculated using VCFtools v0.1.14<sup>86</sup>, with the sliding window size of 50 kb and the step size of 10 kb. Regions that with the top 5% values for  $F_{ST}$  and  $\pi$  ratio and negative values for Tajima’s  $D$  were considered as outliers. Analysis of XP-CLR was performed using XP-CLR<sup>87</sup>, with the parameter setting of “ld0.95, phased, maxsnps1000, size50000, and step 10000”. The windows with the top 5% values of XP-CLR score were considered as regions under selection. To reduce the false positives, we only used the windows containing at least 20 SNPs and kept the genes supported by all four statistics as the positively selected genes. We identified the function of these genes based on their annotation information.

### Genome-wide association studies (GWAS) on ovipositor length

To identify the SNPs associated with ovipositor length, we carried out GWAS based on the SNPs from all samples of the dominant cryptic species ( $N = 82$ ) using GEMMA v0.98.5<sup>88</sup>. We first tested the normality of data of two quantitative traits using Shapiro–Wilk test. Then, GWAS was performed under a linear mixed model setting relatedness among samples (the matrix of relatedness reflected by results of the phylogenomic analysis (Fig. 2b)) as the random effect.

To reduce false positive results, a significant  $P$ -value threshold was assessed and corrected using the Benjamini–Hochberg method in GEC (Genetic Type I error calculator) v1.0<sup>89</sup>. Manhattan Plot and Quantile–Quantile Plot (Q–Q plot) were drawn using R, and the model fit of GWAS was assessed according to the Q–Q plot. We then identified the position

types of the significantly associated SNPs (intergenic region, intron, exon, and upstream and downstream of a gene). Functional annotation of these candidate genes was performed using multiple databases (SwissProt, KEGG, GO, and NR) using BLASTP with the threshold of  $e < 10^{-5}$ .

### In vitro functional assay of key genes

The full lengths of the ORFs of two selected genes (the variant dopamine N-acetyltransferase (*Dat*) and the glucose dehydrogenase (*Gld*)) were further confirmed by reverse transcription PCRs and Sanger sequencing. These two genes were then synthesized and cloned into pET-22b (+) plasmids (MilliporeSigma), expressed in *E. coli* strains BL21 (DE3). The produced recombinant proteins were purified (purity > 90%) using modified nickel-nitrilotriacetic acid agarose (Thermo Fisher Scientific).

For enzymatic activity assays of the two genes, we used the reaction system (500  $\mu\text{l}$ ) mainly composed of 200 mM potassium phosphate buffer (pH 6.5), 0.01% (g/ml) bovine serum albumin (BSA), 0.5 mM substrate(s) (Supplementary Table 9), and 10  $\mu\text{l}$  of purified recombinant protein (0.4 mg/ml). After 5 min of incubation and 30 min of reaction at 35 °C, the reaction products were collected by extraction using methyl alcohol. The products were stored at  $-20^\circ\text{C}$  overnight and centrifuged with 13,523 g (12,000 rpm) using Centrifuge 5424 R (Eppendorf, Germany). All detections were performed on a UPLC instrument combined with a QTRAP® 6500 + MS system equipped with an electrospray ionization (ESI) source (AB SCIEX). For, a Kinetex C18 column (2.6  $\mu\text{m}$ ,  $2.1 \times 100 \text{ mm}$ , Phenomenex) and an ACQUITY UPLC® BEH Amide column (1.7  $\mu\text{m}$ ,  $2.1 \times 100 \text{ mm}$ , Waters) were used to separate N-acetyldopamine and D-glucono-1,5-lactone, respectively. The instrument parameters were optimized by infusion experiments with standards. The optimized ESI operating parameters for negative mode were: ion spray voltage, 4500 V; heated probe temperature, 400 °C, curtain Gas, 35 psi; cone temperature, 550 °C; probe gas flow, 55 psi; nebulizing gas, 55 psi. All analytes were detected using multiple reaction monitoring, and Analyst 1.6.3 software (AB SCIEX) was used for data acquisition and analysis. For each gene, the enzymatic activity assay was repeated three times, and three replications of negative controls (only adding the substrates and BSA into buffer) were conducted.

### Data availability

The source data underlying the graphs in the main figures are uploaded as Supplementary Data 1. The resequencing data used in our population genomic analysis are in the GSA database (China National Center for Bioinformation) under accession ID CRA027798, and all *COI* gene sequences are deposited in the GenBank [GenBank accession numbers PX106244–PX106370].

### Code availability

The code for the 10 custom candidate demographic models are available in the Figshare public repository (<https://doi.org/10.6084/m9.figshare.29432987.v2>).

Received: 24 September 2024; Accepted: 15 August 2025;

Published online: 26 September 2025

### References

- Pyšek, P. & Richardson, D. M. Invasive species, environmental change and management, and health. *Annu. Rev. Environ. Resour.* **35**, 25–55 (2010).
- Traveset, A. & Richardson, D. M. Mutualistic interactions and biological invasions. *Annu. Rev. Ecol. Evol. S.* **45**, 89–113 (2014).
- Guo, K. et al. Plant invasion and naturalization are influenced by genome size, ecology and economic use globally. *Nat. Commun.* **15**, 1330 (2024).
- Razanajatovo, M. et al. Autofertility and self-compatibility moderately benefit island colonization of plants. *Glob. Ecol. Biogeogr.* **28**, 341–352 (2019).



5. Winchell, K. M. et al. Genome-wide parallelism underlies contemporary adaptation in urban lizards. *Proc. Natl. Acad. Sci. USA* **120**, e2216789120 (2023).
6. Traveset, A. & Richardson, D. M. Biological invasions as disruptors of plant reproductive mutualisms. *Trends Ecol. Evol.* **21**, 208–216 (2006).
7. Nathan, P., Economo, E. P., Guénard, B., Simonsen, A. K. & Frederickson, M. E. Generalized mutualisms promote range expansion in both plant and ant partners. *Proc. R. Soc. B* **290**, 20231083 (2023).
8. Wang, R. et al. The fig wasp followers and colonists of a widely introduced fig tree, *Ficus microcarpa*. *Insect Conserv. Divers.* **8**, 322–336 (2015).
9. Palmér, C., Wallin, A., Persson, J., Aronsson, M. & Blennow, K. Effective communications on invasive alien species: identifying communication needs of Swedish domestic garden owners. *J. Environ. Manag.* **340**, 117995 (2023).
10. Aizen, M. A. & Torres, A. The invasion ecology of mutualism. *Annu. Rev. Ecol. Evol. S.* **55**, 41–63 (2024).
11. Hsin, K. T. et al. Allopatric lineage divergence of the east asian endemic herb *Conandron ramondiioides* inferred from low-copy nuclear and plastid markers. *Int. J. Mol. Sci.* **23**, 14932 (2022).
12. Dlugosch, K. M. & Parker, I. M. Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. *Mol. Ecol.* **17**, 431–449 (2008).
13. Qiao, H. M., Liu, W. W., Zhang, Y. H., Zhang, Y. Y. & Li, Q. Q. Genetic admixture accelerates invasion via provisioning rapid adaptive evolution. *Mol. Ecol.* **28**, 4012–4027 (2019).
14. Liu, W. W., Strong, D. R., Pennings, S. C. & Zhang, Y. H. Provenance-by-environment interaction of reproductive traits in the invasion of *Spartina alterniflora* in China. *Ecology* **98**, 1591–1599 (2017).
15. Parvizi, E., Vaughan, A. L., Dharmi, M. K. & Mcgaughan, A. Genomic signals of local adaptation across climatically heterogeneous habitats in an invasive tropical fruit fly (*Bactrocera tryoni*). *Heredity* **132**, 18–29 (2024).
16. Bailey, N. W. et al. A neglected conceptual problem regarding phenotypic plasticity's role in adaptive evolution: the importance of genetic covariance and social drive. *Evol. Lett.* **5**, 444–457 (2021).
17. Lemmen, K. D., Verhoeven, K. J. F. & Declerck, S. A. J. Experimental evidence of rapid heritable adaptation in the absence of initial standing genetic variation. *Funct. Ecol.* **36**, 226–238 (2022).
18. Wang, X. Y. et al. Genotypic diversity enhances invasive ability of *Spartina alterniflora*. *Mol. Ecol.* **21**, 2542–2551 (2012).
19. Lee, C. E. & Gelembiuk, G. W. Evolutionary origins of invasive populations. *Evol. Appl.* **1**, 427–448 (2008).
20. Jia, P. et al. Long-term invasion simplified soil seed bank and regenerated community in a coastal marsh wetland. *Ecol. Appl.* **34**, e2754 (2024).
21. Liu, Y. J. & van Kleunen, M. Responses of common and rare aliens and natives to nutrient availability and fluctuations. *J. Ecol.* **105**, 1111–1122 (2017).
22. Lu, X. M., Siemann, E., Shao, X., Wei, H. & Ding, J. Q. Climate warming affects biological invasions by shifting interactions of plants and herbivores. *Glob. Change Biol.* **19**, 2339–2347 (2013).
23. Xu, C. C., Li, Z. Y. & Wang, J. B. Linking heat and adaptive responses across temporal proteo-transcriptome and physiological traits of *Solidago canadensis*. *Environ. Exp. Bot.* **175**, 104035 (2020).
24. Hao, Y. et al. Genomic and phenotypic signatures provide insights into the wide adaptation of a global plant invader. *Plant Commun.* **5**, 100820 (2024).
25. Hayward, J., Horton, T. R. & Nuñez, M. A. Ectomycorrhizal fungal communities coinvasive with Pinaceae host plants in Argentina: Gringos bajo el bosque. *N. Phytol.* **208**, 497–506 (2015).
26. Smith, A. L. et al. Global gene flow releases invasive plants from environmental constraints on genetic diversity. *Proc. Natl. Acad. Sci. USA* **117**, 4218–4227 (2020).
27. Wang, R. et al. Molecular mechanisms of mutualistic and antagonistic interactions in a plant-pollinator association. *Nat. Ecol. Evol.* **5**, 974–986 (2021).
28. Martínez-Macias, K. J., Márquez-Guerrero, S. Y., Martínez-Sifuentes, A. R. & Segura-Castruita, M. A. Habitat suitability of fig (*Ficus carica* L.) in Mexico under current and future climates. *Agriculture* **12**, 1816 (2022).
29. Berg, C. C. Classification and distribution of *Ficus*. *Exp.* **45**, 605–611 (1989).
30. Wiebes, J. T. Taxonomy and host preference of Indo-Australian fig wasps of the genus *Ceratosolen* (Agaonidae). *Tijdschr. Entomol.* **106**, 1–112 (1963).
31. Beardsley, W. J. Chalcid wasps (Hymenoptera: Chalcidoidea) associated with fruit of *Ficus microcarpa* in Hawai'i. *Proc. Hawaii. Entomol. Soc.* **33**, 19–34 (1998).
32. Nadel, H., Frank, J. H. & Knight, R. J. Escapees and accomplices: the naturalization of exotic *Ficus* and their associated faunas in Florida. *Flor. Entomol.* **75**, 29–38 (1992).
33. Caughlin, T., Wheeler, J. H., Jankowski, J. & Lichstein, J. W. Urbanized landscapes favored by fig-eating birds increase invasive but not native juvenile strangler fig abundance. *Ecology* **93**, 1571–1580 (2012).
34. Verde, G. L., Speciale, M. & Cerasa, G. First record in Europe of seedlings of *Ficus macrophylla* F. *columnaris* (Moraceae) and of its pollinating wasp *Pleistodontes cf. imperialis* (Chalcidoidea Agaonidae). *Nat. Sicil.* **39**, 399–406 (2015).
35. Wang, R. et al. Loss of top-down biotic interactions changes the relative benefits for obligate mutualists. *Proc. Biol. Sci.* **286**, 20182501 (2019).
36. Sun, X. J., Xiao, J. H., Cook, J. M., Feng, G. & Huang, D. W. Comparisons of host mitochondrial, nuclear and endosymbiont bacterial genes reveal cryptic fig wasp species and the effects of *Wolbachia* on host mtDNA evolution and diversity. *BMC Evol. Biol.* **11**, 86 (2011).
37. Compton, S. G. et al. Host-parasitoid relationships within figs of an invasive fig tree: a fig wasp community structured by gall size. *Insect Conserv. Divers.* **11**, 341–351 (2018).
38. Berg, C. C. & Corner, E. J. H. Moraceae: ficeae. *Flora Males. 1, Spermatophyta* **17**, 1–702 (2005).
39. Zhang, X. T. et al. Genomes of the banyan tree and pollinator wasp provide insights into fig-wasp coevolution. *Cell* **183**, 875–889.e817 (2020).
40. Gokhman, V. E., Nugnes, F. & Bernardo, U. Chromosomes of *Eupristina verticillata* Waterston, 1921 and an overview of known karyotypes of chalcid wasps of the family Agaonidae (Hymenoptera). *J. Hymenopt. Res.* **71**, 157–161 (2019).
41. Yu, H. et al. Multiple parapatric pollinators have radiated across a continental fig tree displaying clinal genetic variation. *Mol. Ecol.* **28**, 2391–2405 (2019).
42. Yang, H. W., Tzeng, H. Y. & Chou, L. S. Phenology and pollinating wasp dynamics of *Ficus microcarpa* L.f.: adaptation to seasonality. *Bot. Stud.* **54**, 11 (2013).
43. Chen, E. H., Christiansen, A. E. & Baker, B. S. Allocation and specification of the genital disc precursor cells in *Drosophila*. *Dev. Biol.* **281**, 270–285 (2005).
44. Cai, W. L., Wei, Y. H., Jamik, M., Reich, J. & Lilly, M. A. The GATOR2 component Wdr24 regulates TORC1 activity and lysosome function. *PLoS Genet* **12**, e1006036 (2016).
45. Deng, P. et al. Silencing downstream of receptor kinase gene (*drk*) impairs larval-pupal ecdysis in *Leptinotarsa decemlineata* (Say). *J. Asia-Pac. Entomol.* **23**, 16 (2020).
46. Iida, K. & Cavener, D. R. Glucose dehydrogenase is required for normal sperm storage and utilization in female *Drosophila melanogaster*. *J. Exp. Biol.* **207**, 675–681 (2004).
47. Rohrbach, E. W., Knapp, E. M., Deshpande, S. A. & Krantz, D. E. Expression and potential regulatory functions of *Drosophila*



- octopamine receptors in the female reproductive tract. *G3 (Bethesda)* **14**, jkae012 (2024).
48. Liu, Y. Q., Dai, H., Bamu, A. F. & Lin, X. D. Peroxisome biogenesis factor PEX14 is crucial for survival and fecundity of female brown planthopper, *Nilaparvata lugens* (Stål). *Insect Biochem. Mol. Biol.* **170**, 104139 (2024).
49. Chen, C. et al. *Drosophila* ionotropic receptor 25a mediates circadian clock resetting by temperature. *Nature* **527**, 516–520 (2015).
50. Ni, L. et al. The ionotropic receptors IR21a and IR25a mediate cool sensing in *Drosophila*. *eLife* **5**, e13254 (2016).
51. Zhu, J. L. & Huang, X. Endocrine function of pheromones couples fat rationing and nutrient scarcity. *Sci. China Life Sci.* **65**, 1267–1269 (2022).
52. Huang, K. et al. RiboTag translomic profiling of *Drosophila* oenocytes under aging and induced oxidative stress. *BMC Genom.* **20**, 50 (2019).
53. Huu, N. T., Yoshida, H. & Yamaguchi, M. Tumor suppressor gene *OSCP1/NOR1* regulates apoptosis, proliferation, differentiation, and ROS generation during eye development of *Drosophila melanogaster*. *FEBS J.* **282**, 4727–4746 (2015).
54. Long, Y. H., Li, J. R., Zhao, T. F., Li, G. N. & Zhu, Y. A new arylalkylamine N-acetyltransferase in Silkworm (*Bombyx mori*) affects integument pigmentation. *Appl. Biochem. Biotechnol.* **175**, 3447–3457 (2015).
55. Roman, J. Diluting the founder effect: cryptic invasions expand a marine invader's range. *Proc. R. Soc. B.* **273**, 2453–2459 (2006).
56. Xue, Y., Lin, C., Wang, Y., Zhang, Y. & Ji, L. Ecological niche complexity of invasive and native cryptic species of the *Bemisia tabaci* species complex in China. *J. Pest Sci.* **95**, 1245–1259 (2022).
57. Liu, M., Zhang, J., Chen, Y., Compton, S. G. & Chen, X.-Y. Contrasting genetic responses to population fragmentation in a coevolving fig and fig wasp across a mainland–island archipelago. *Mol. Ecol.* **22**, 4384–4396 (2013).
58. Aoki, D. et al. Population genetics of recent natural colonization by the bull-headed shrike (*Lanius bucephalus*; Aves) suggests the importance of recurrent immigration on remote islands. *Biol. J. Linn. Soc.* **141**, 559–571 (2023).
59. Henery, M. L. et al. Evidence for a combination of pre-adapted traits and rapid adaptive change in the invasive plant *Centaurea stoebe*. *J. Ecol.* **98**, 800–813 (2010).
60. Lenz, M. et al. Non-native marine invertebrates are more tolerant towards environmental stress than taxonomically related native species: results from a globally replicated study. *Environ. Res.* **111**, 943–952 (2011).
61. Renault, D., Laparie, M., McCauley, S. J. & Bonte, D. Environmental adaptations, ecological filtering, and dispersal central to insect invasions. *Annu. Rev. Entomol.* **63**, 345–368 (2018).
62. Phillipsen, I. C. et al. Dispersal ability and habitat requirements determine landscape-level genetic patterns in desert aquatic insects. *Mol. Ecol.* **24**, 54–69 (2015).
63. Teixido, A. L. et al. Anthropogenic impacts on plant–animal mutualisms: a global synthesis for pollination and seed dispersal. *Biol. Conserv.* **266**, 109461 (2022).
64. Lewis, S. L. & Maslin, M. A. Defining the Anthropocene. *Nature* **519**, 171–180 (2015).
65. Hu, L. S. & Dong, Y. W. Multiple genetic sources facilitate the northward range expansion of an intertidal oyster along China's coast. *Ecol. Appl.* **34**, e2764 (2024).
66. Xie, H. J. et al. *ICE1* demethylation drives the range expansion of a plant invader through cold tolerance divergence. *Mol. Ecol.* **24**, 835–850 (2015).
67. Meraner, A. et al. Massive invasion of exotic *Barbus barbus* and introgressive hybridization with endemic *Barbus plebejus* in Northern Italy: where, how and why? *Mol. Ecol.* **22**, 5295–5312 (2013).
68. Chen, S. F., Zhou, Y. Q., Chen, Y. R. & Gu, J. FASTP: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
69. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
70. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
71. McKenna, A. et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
72. Wang, K., Li, M. Y. & Hakonarson, H. Annovar: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
73. Cruaud, A. et al. An extreme case of plant–insect codiversification: figs and fig-pollinating wasps. *Syst. Biol.* **61**, 1029–1047 (2012).
74. Hebert, P. D. N., Penton, E. H., Burns, J. M., Janzen, D. H. & Hallwachs, W. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proc. Natl. Acad. Sci. USA* **101**, 14812–14817 (2004).
75. Nguyen, L. T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: A Fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2014).
76. Zhang, T. et al. The evolution of parasitism from mutualism in wasps pollinating the fig, *Ficus microcarpa*, in Yunnan Province, China. *Proc. Natl. Acad. Sci. USA* **118**, e2021148118 (2021).
77. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
78. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
79. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
80. Zhang, C., Dong, S. S., Xu, J. Y., He, W. M. & Yang, T. L. PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* **35**, 1786–1788 (2019).
81. Excoffier, L. et al. Fastsimcoal2: demographic inference under complex evolutionary scenarios. *Bioinformatics* **37**, 4882–4885 (2021).
82. Johnson, J. B. & Omland, K. S. Model selection in ecology and evolution. *Trends Ecol. Evol.* **19**, 101–108 (2004).
83. Xu, X., Wang, B. S. & Yu, H. Intraspecific genomic divergence of a fig wasp species is due to geographical barrier and adaptation. *Front. Ecol. Evol.* **9**, 764828 (2021).
84. Nefdt, R. J. C. & Compton, S. G. Regulation of seed and pollinator production in the fig–fig wasp mutualism. *J. Anim. Ecol.* **65**, 170–182 (1996).
85. Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D. & Team, R. C. Linear and nonlinear mixed effects models. *R. Package Version* **3**, 1–89 (2007).
86. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
87. Chen, H., Patterson, N. & Reich, D. Population differentiation as a test for selective sweeps. *Genome Res.* **20**, 393–402 (2010).
88. Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* **44**, 821–824 (2012).
89. Li, M. X., Yeung, J. M. Y., Cherny, S. S. & Sham, P. C. Evaluating the effective numbers of independent tests and significant *p*-value thresholds in commercial genotyping arrays and public imputation reference datasets. *Hum. Genet.* **131**, 747–756 (2012).

## Acknowledgements

We thank Zhe-Ting Cai, Jing Cheng, Jun-Yin Deng, Qi-Chong Zhu, Qian-Ya Li, Rhett D. Harrison, Yan-Qiong Peng, Rodrigo A.S. Pereira, Alfredo Reyes-

Betancort, Lillian J.V. Rodriguez, Hui Yu for their kind helps in field sampling and analysis. This work is supported by the Shanghai Commission of Science and Technology grants 23DZ1204602 and TQ20240210 (R.W.) and the NSFC grants 32071667 (R.W.) and 32261123001 (X.-Y.C.).

### Author contributions

Conceptualization: R.W., X.Y.C., and S.G.C.; Methodology: R.W., Y.Y.Z., K.J.Z., J.Z., L.H.W., L.L., Y.T.J., K.J., S.T.S., and S.V.N.; Investigation: L.H.W., W.C.L., W.M.X., G.W., Y.Y.L.; Funding acquisition: R.W.; Project administration: X.Y.C., R.W.; Writing: all authors contribute to writing the manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s42003-025-08733-1>.

**Correspondence** and requests for materials should be addressed to Rong Wang.

**Peer review information** : *Communications Biology* thanks Jorge Duitama and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editors: Madhava Meegaskumbura and Michele Repetto.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025