國立臺灣大學生命科學院生態學與演化生物學研究所

碩士論文

Graduate Institute of Ecology and Evolutionary Biology
College of Life Science
National Taiwan University

Master Thesis

榕小蜂科的化學感受器基因家族演化

The evolution of chemosensory gene families in fig wasps (Agaonidae)

羅百尉

Bai-Wei Lo

指導教授：王弘毅 博士

Advisor: Hurng-Yi Wang, Ph.D.

中華民國 108 年 8 月

August, 2019

doi:10.6342/NTU201903609

誌謝

能完成這篇論文，我首先要感謝我的指導教授王弘毅老師對我的細心指導，以及所有 R638 成員。很榮幸能夠待過呂雪梅老師和李承叡老師的實驗室，在那我結識到很多志同道合的朋友，也學到了很多東西。我要特別謝謝文艷玲師姊，她帶領我認識基因體學跟生物資訊學的方法。感謝可文亞老師出借 server，以及林恭平學弟幫我搞定一堆裝到吐血的軟體。感謝殷利鐸師兄在百忙之中幫忙我處理多項樣本，感謝 Philipp Brand 先生分享昆蟲嗅覺相關的研究方法。感謝我的好友譚文皓、黃偉傑、李其儒跟我討論實驗和抓 bug。

本實驗的樣本取得要感謝這些人:新品莊園賴加進學長、文化大學謝佳宏學長、山水裡張尚平先生、苗栗農改場林孟均研究員、神山愛玉冰、撒拉伯民宿。

最後我要感謝我的家人，感謝我父親無條件的支持我做想做的事。是我母親啟迪我對生態研究的熱誠。感謝我的女朋友涵茹，作為一個比我有太多美感天賦的人，多次挽救過我的研討會海報外觀。妳身為一個研究生前輩，總是給予我很多有用的建議。

doi:10.6342/NTU201903609

中文摘要

化學受器是昆蟲主要的感覺器官，昆蟲利用嗅覺來感知食物、配偶、宿主、掠食者等外在因子。榕小蜂科的物種丟失了大多數的化學受器基因，可能與其封閉的特殊生活史有關。即使如此，榕小蜂的嗅覺在其與榕樹的互利共生關係中至關重要，榕小蜂主要透過辨識榕樹隱頭果氣味，來達到專一性授粉。本論文透過定序兩種榕小蜂的基因體跟轉錄體，研究小蜂的化學受器基因如何與其宿主共演化。本文第一章使用粒線體基因標記，進行系統發生以及族群遺傳分析，進而釐清兩研究物種 *Wiebesia pumilae*, *W.* sp3 和他們的近親 *W.* sp1 之間的親緣關係、天然分佈位置、自然宿主，並揭露 *W.* sp3 有數個族群源於人為引入，且在其入侵地發生多次獨立轉換宿主，由原本共生的愛玉 (*Ficus pumila var. awkeotsang*)轉至薜荔(*Ficus pumila var. pumila*)。經由回顧前人在入侵族群做的物候調查，我提出一個可能機轉解釋榕小蜂的宿主轉換在不同榕樹性別系統之最終演化結果。本文第二章提供詳細的生物資訊流程描述組裝及注釋兩研究物種的基因體以及粒線體基因體，並進行種間基因演化速率分析。

本文第三章探討不同演化層面和尺度下，榕小蜂嗅覺基因與其宿主的共演化。在精細演化尺度，利用近期與宿主達成共演化的 *W. pumilae* 和 *W.* sp3 的基因體與轉錄體，發現在科內基因拷貝數保守的嗅覺基因出現種間差異表達現象。而在跨越整個榕小蜂科的廣大演化尺度，藉由比較 *W. pumilae*、*W.* sp3 和 *Ceratosolen solmsi* 三物種，發現在嗅覺受器 (olfactory receptor)出現各支系獨有的適應性基因串聯重複 (adaptive tandem gene duplication)，極有可能是與宿主榕樹氣味長期共適應、共演化的結果。最後我比較與不同性別系統榕樹共生之榕小蜂，發現三個在榕小蜂科中出現縮減的嗅覺基因家族，於雌雄同株榕樹的小蜂 *Elisabethiella stueckenbergi* 都出現較多的基因數目擴張，可能與雌雄同株榕樹小蜂具有更頻繁的宿主轉換有關。

關鍵字：榕小蜂、嗅覺、共演化、基因體組裝、寄主專一性

doi:10.6342/NTU201903609

ABSTRACT

Pollinating fig wasps (Agaonidae) have one of the most reduced chemosensory genes in insects, which is probably associated with specialized life cycle in obligate mutualism. On the other hand, olfaction plays a crucial role in maintaining host specificity in the fig-fig wasp coevolution. In this thesis, I sequenced genomic and transcriptomic data from two fig wasp species to understand how reduced chemosensory genes maintain host-specificity during species divergence. The first chapter describes the evolutionary relationships of the two studied fig wasps (*Wiebesia pumilae* and *W.* sp3), their close species (*W.* sp1), and their associated hosts (*Ficus pumila var. pumila* and *Ficus pumila var. awkeotsang*), which revealed that while originally an endemic species, recent human intervention had resulted in introduced populations along with recurrent host-shifting in *W.* sp3. Possible mechanism for distinct co-pollinator pattern seen in different fig sexual systems was also proposed. The second chapter provides bioinformatics pipelines to generate high quality nuclear genomes and mitochondrial genomes of both species using next generation sequencing, and assess the evolutionary rates in protein-coding genes between them.

The final chapter characterizes chemosensory gene evolution of fig wasp from multiple evolutionary perspectives. For fine scale evolution, utilizing the genome and transcriptome of *W.* sp3 and *W. pumilae*, both of which codiverged recently with their host, I discovered that regulatory changes at copy-number conservative chemosensory genes are associated with local coadaptations. For large scale evolution, by comparing the two *Wiebesia* species with *Ceratosolen solmsi*, I found that lineage-specific adaptive tandem gene duplications in olfactory receptors (OR) family may drive phenotypic

doi:10.6342/NTU201903609

coevolution with figs. For olfactory evolution in wasps belonging to different sexual systems of hosts, larger expanded gene families were found in the ancestrally contracted gene families: OR, gustatory receptor (GR) and odorant-binding protein (OBP) in the monoecious fig wasp *Elisabethiella stueckenbergi*, possibly reflecting differences in host-shifting frequency.

Keywords: fig wasp, olfaction, coevolution, genome assembly, host specificity

doi:10.6342/NTU201903609

CONTENTS

doi:10.6342/NTU201903609

doi:10.6342/NTU201903609

doi:10.6342/NTU201903609

doi:10.6342/NTU201903609

LIST OF FIGURES

doi:10.6342/NTU201903609

doi:10.6342/NTU201903609

LIST OF TABLES

doi:10.6342/NTU201903609

Chapter 1 Genetic evidence revealed pollinator sharing via host-switching in pollinators associated with *Ficus pumila*

1.1 Introduction

The obligate mutualism between figs and fig wasps is a classic co-evolutionary system that had been studied extensively (Cruaud et al., 2012; Machado, Robbins, Gilbert, & Herre, 2005; Wiebes, 1979) . Figs and pollinating fig wasps were once thought to evolve under strict co-speciation, which was known as the one-to-one rule. However, species delimination according to molecular evidence in the last two decades had rejected the universality of strict co-speciation (Yan Chen, Compton, Liu, & Chen, 2012; Cook & Segar, 2010; L.-Y. Yang et al., 2015)s. One common scenario that deviates from the one-to-one rule is when the fig tree is widely distributed and has multiple geographically isolated subspecies/varieties (Rasplus, 1996), which is the case in *Ficus pumila*.

Ficus pumila is a functional dioecious fig that occur across a wide range of latitudes in East Asia. Two varieties of *F. pumila* had been described. The first being the widely distributed nominate variety: *F. pumila. var. pumila*, commonly known as the creeping-fig. The other is *F. pumila var. awkeotsang* (Makino) Corner, commonly known as the jelly-fig. Jelly-fig is most likely to be an endemic variety originated in Taiwan, but had been cultivated in Southern China due to its agricultural value (Chen et al., 2012). The two varieties differ in their leaf shapes, syconia size, phenology, and habitant preferences (Hsieh et al., 1993). Both varieties can be found in Taiwan, but are mostly allopatric or parapatric. Creeping fig occur in lowlands while jelly fig mostly occurs in altitudes between 1200-1900m (Hsieh et al., 1993).

Two large scale molecular studies revealed the phylogeography and diversity of the

doi:10.6342/NTU201903609

pollinators associated with *F. pumila* (Chen et al., 2012; Wang et al., 2013). Chen et al., 2012 reported three cryptic species associated with *F. p. var. pumila* in China. These species are *W.* sp1 that occupies north of the Wuyi mountains; *W.* sp2 at the south of Wuyi mountains, and the mysterious *W.* sp3 that discontinuously scattered in Zhoushan Islands and Fujian. Chen et al., 2012 proposed that the observed distribution of the three species is a result of multiple ice age refugia followed by post-ice age expansions. On the other hand, Wang et al., 2013 revealed that the two *F. pumila* varieties in Taiwan are pollinated by two cryptic species respectively. So far the relationship between all the reported pollinating fig wasp species associated with the two *F. pumila* varieties remains unclear. A combined study containing pollinators of *F. p. var. pumila* and *F. p. var. awkeotsang* is crucial in understanding the full picture of their coevolutionary history.

Here, using mitochondrial cytochrome oxidase I (COI) sequences from published literatures, I provide a phylogenetic and population genetic study on pollinating fig wasps of *F. pumila*. My results resolve the relationships between each *F. pumila* variety and its associated pollinators, while revealing that natural pollinator of *F. p. var. awkeotsang* had shifted its host into *F. p. var. pumila* in several introduced regions. Finally, by reviewing past literatures regarding phenology of *F. pumila* and their pollinators, I provide insights into the mechanism that lead to different co-pollinator pattern observed in monoecious and dioecious figs.

1.2 Material and Methods

1.2.1 Phylogeny and population genetics

Mitochondrial cytochrome oxidase subunit I (COI) sequences belonging to pollinators of *F. pumila* were downloaded from National Center for Biotechnology Information (NCBI) nucleotide database. Keywords used for searching were "*Wiebesia*"

doi:10.6342/NTU201903609

and “cytochrome oxidase subunit I”. Reference publications of searched results were examined; only sequences from studies that featured pollinators associated with *F. pumila* were kept. Individual’s haplotype with collection site were used for estimating population genetic statistics including genetic diversities and fixation indexes (FSTs). All sequences were used in clustering analysis including phylogenetic reconstruction, principal coordinate analysis (PCoA), and TCS haplotype network. Four additional sequences from *W. brusi*, *W. frustrate*, *W. punctatae*, and *W. boldinghi* (AF200412-AF200414, JN103288) were used as outgroup in phylogeny analysis.

The sequences were aligned by codon using ClustalW (Larkin et al., 2007) implemented in MEGA7 (Kumar, Stecher, & Tamura, 2016). Neighbor-joining tree (Nei & Gojobori, 1986; Saitou & Nei, 1987) with 1000 bootstrap replicates and pairwise genetic distance matrices were calculated using MEGA7. TCS networks (Clement, Snell, Walke, Posada, & Crandall, 2002) were drew using PopART 1.7 (Leigh & Bryant, 2015). The haplotype diversities, nucleotide diversities, and FSTs of each population were calculated using DNAsp v6 (Rozas et al., 2017). For species that were distributed across the Taiwan Strait, analysis of molecular variation (AMOVA) (L Excoffier, Smouse, & Quattro, 1992) implemented in Arlequin 3.5 (Laurent Excoffier & Lischer, 2010) was used to determine the composition of genetic variation.

1.2.2 Species distribution modeling

Species distribution modeling was used to assess the potential invading areas of *W.* sp3 in mainland. Since *W.* sp3 was the sole known pollinator of *F. p. var. awkeotsang*, occurrence data of *F. p. var. awkeotsang* from Global Biodiversity Information Facility (GBIF) were used as occurrence data of *W.* sp3. GBIF data was cleaned after manually examined under google earth; points that are not in Taiwan or in cities (representing cultivated sites, which did not guarantee to have associated *W.* sp3 or simply data error)

doi:10.6342/NTU201903609

were discarded. Cleaned GBIF data as well as coordinates from invading sites described in 1.2.1 were used as occurrence data in species distribution modeling. 19 abiotic environmental data were downloaded from WorldClim (Fick & Hijmans, 2017) using R package raster. Species distribution was predicted by Bioclim and MaxEnt (Phillips, Anderson, & Schapire, 2006).

1.3 Results

591 COI sequences belonging to pollinators associated with *F. pumila* were downloaded from NCBI. 581 of them represent individuals and has information of collection sites (accession numbers: JN183988-JN184049, KC579186-KC579356); 10 sequences only represent unique haplotypes within population (HQ398108-HQ398117). After alignment with ClustalW, an overlapping 908 bp region was used for downstream analysis.

Neighbor-joining phylogeny showed that the wasps are consisted of three distinct clades/species (Figure 2). The three species were *W.* sp1 that was described in Chen et al., 2012; *W. pumilae*, which includes *W.* sp2 from Chen et al., 2012 and creeping-fig wasp in Wang et al., 2013. The third species is *W.* sp3, which included *W.* sp3 from Chen et al., 2012 and jelly-fig wasp in Wang et al., 2013 (Figure 1, Figure 2). According to this result, *W.* sp3 is a shared pollinator of *F. p. var. pumila* and *F. p. var. awkeotsang* (Figure 1). Haplotype JN184049 of *W.* sp3 is the sister clade to all other *W.* sp3 individuals. JN184049 was collected from a cultivated jelly fig in China. Mean genetic distance between JN184049 and other *W.* sp3 was 0.0115, which was greater than the largest within species distance in *W.* sp1 (0.0112) and *W. pumilae* (0.0111).

Both *W. pumilae* and *W.* sp3 were distributed in Taiwan and mainland East Asia. Considering the limited migration capability of fig wasps, vicariance between Taiwan

doi:10.6342/NTU201903609

and mainland since the last glacial maximum should result in genetic divergence between the two groups. To test the prediction, AMOVA was used to decompose the genetic variation within the two species. AMOVA results showed that in *W. pumilae*, 42.99% of overall genetic variation can be explained by population differentiation across the Taiwan Strait (Table 2), indicating the rule of vicariance in shaping the genetic landscape within species. On the other hand, only 12.42% of the overall genetic variation are found between Taiwan and mainland populations in *W.* sp3.

TCS haplotype network tree was used to visualize the genetic distance within *W. pumilae* and *W.* sp3. Consistent with the finding in AMOVA, *W. pumilae* samples from Taiwan is divergent from mainland samples and nearly constitute one single clade. In contrast, *W.* sp3 haplotypes from mainland are mixed with haplotypes from Taiwan. Particularly, nearly all the haplotypes from Zhoushan Islands of mainland had identical or closely-related haplotype to one Taiwanese population: JML. Population JML came from a cultivated strain in a governmental agricultural research center (Miaoli District Agricultural Research and Extension Station), and had zero genetic diversity, a result of generations of inbreeding.

Population genetic statistics revealed that *W.* sp3 populations from Zhoushan Islands had extremely low nucleotide and haplotype diversity when compared to other *W.* sp3 populations (Table 1), a pattern typically seen in introduced populations. Both *W.* sp1 and *W.* sp3 occurred in Zhoushan Islands. The pattern of genetic differentiation in local populations is distinct between the two species. Mean F_{ST} of *W.* sp1 in Zhoushan Islands was 0.246, while no detectable population differentiation (F_{ST}=0) were found in co-occurring *W.* sp3 populations, suggesting recent bottleneck.

95 occurrence coordinates were retrieved from GBIF, 85 of them past manual data cleaning, representing 63 unique sites (Figure 5). The final occurrence data used in sdm

doi:10.6342/NTU201903609

were these 63 sites and the 12 sites reported in Chen et al., 2012. Results of sdm from both Bioclim and Maxent showed that the south eastern shore of mainland are at risk of *W. sp3* invasions (Figure 6).

1.4 Discussion

The results show that the natural pollinator of *F. p. var. pumila* are the sibling species *W. pumilae* and *W.* sp1, while the natural pollinator of *F. p. var. awkeotsang* is *W.* sp3. Current distribution of the three fig wasp species indicates that they arose through allopatric speciation under vicariance, as suggested by Chen et al., 2012. Previous studies had reported that speciation of pollinating fig wasp usually precedes their fig host (Edward Allen Herre, Jandér, & Machado, 2008). In the case of *F. pumila* and their pollinators, Wang et al., 2013 did found that the estimated divergence time between *W. pumilae* and *W.* sp3 was indeed longer than corresponding *F. p. var. pumila* and *F. p. var. awkeotsang*. While duplication (speciation of wasps but not figs) was seen in *F. p. var. pumila*.

Most *W.* sp3 individuals found in mainland are recently introduced from Taiwan. Evidence of recent introduction included genetic bottleneck, non-monophyletic origins, and low genetic differentiation with the source populations. Most of the haplotypes in Zhoushan Islands can be traced back to the strain from Miaoli District Agricultural Research and Extension Station, which makes sense since the Station was responsible for providing jelly-fig pollinator strains to farmers in Taiwan. Those wasps were probably then shipped across the strait in failed attempts to cultivate jelly-figs. Results from sdm suggest that *W.* sp3 has similar environmental preference to its sibling species, and can potentially invade areas that are occupied by *F. p. var. pumila*. Only one *W.* sp3 individual (JN184049) is distantly related to Taiwanese haplotypes. This sample is

doi:10.6342/NTU201903609

peculiar not only for its divergence but also because it is the only *W.* sp3 sample in mainland that was collected from cultivated *F. p. var. awkeotsang*. Whether this individual represent hidden population structure or a result of sequencing error can only be answered by more complete sampling.

1.4.1 How the introduced fig wasp species affect fig-fig wasp mutualism?

Independent host switching events from *F. p. var. awkeotsang* to *F. p. var. pumila* was seen in mainland *W.* sp3 populations. How did *W.* sp3 manage to compete with native fig wasps if they shared nearly identical niche, and what are the long term evolutionary consequences? Phenological study of *F. p. var. pumila* in Zhoushan Islands discovered that the emergence time of the two fig wasp species are almost non-overlapping (Figure 7) (Liu et al., 2014), with *W.* sp1 emerges first followed by *W.* sp3. Unknowing that *W.* sp3 is a recent invader from another variety, the researchers argued that this phenomenon is the consequence of long term competitive exclusion between two co-pollinators. However, flowering time of male *F. p. var. pumila* and *F. p. var. awkeotsang* in Fujian (Figure 8) (Yong Chen, Li, & Ma, 2002) respectively matches the pollinator emergence time of *W.* sp1 and *W.* sp3 in Zhoushan. The matching of phenology between Zhoushan fig wasps and their natural hosts suggests a non-mutually exclusive explanation to the observed phenological separation, which is that phenology of male fig is a plastic phenotype, and can be determined by pollinators reside within it. Indeed, Liu et al., 2014 observed that *F. p. var. pumila* occupied by *W.* sp3 tend to also produce receptive male syconia perfectly matches the time of *W.* sp3 emergence. Phenotypic plasticity of male tree phenology allows the once invaded *W.* sp3 to exploit fig resources separated from local fig wasps. Phenological separation reduces competition between local and invading species, thus facilitating the colonization.

doi:10.6342/NTU201903609

1.4.2 Evolutionary outcome of co-pollinator caused by host shifting

Long term evolutionary fates of co-pollinators arise by host shifts is unclear. However, altered male phenology due to non-native pollinator colonization can disrupt the synchrony between male and female flowering time in dioecious figs. For example given the female flowering time of *F. p. var. pumila* in spring (Figure 8), *W.* sp3 are unable to pollinate female flowers due to its late emergence (Figure 7), leading to fitness cost in male fig that harbors *W.* sp3. Meanwhile, if there is heritability in flowering time (which should be), natural selection will act against fig trees that are phenologically more suitable for non-native pollinator, due to the fitness deficit the pollinators bring upon the hosts. This will possibly result in the extinction of non-native pollinator in the long run, for the suitable hosts were wiped out by natural selection.

By contrast, such force of negative selection due to disruption of synchrony do not exist in monoecious figs, since female and male flowers all reside within the same syconia. What's more, separated phenology can produce gene flow barrier among figs harboring different species of fig wasps, which may lead to co-speciation in the long run. Analysis of co-pollinator relationships in monoecious and dioecious figs had revealed that while co-pollinators in dioecious fig constitutes nearly exclusively of sibling taxa, distantly related co-pollinators can often be found in monoecious figs (L.-Y. Yang et al., 2015). The authors proposed two non-mutually exclusive hypotheses to explain the phenomenon: (1) dioecious fig wasp has more specificity with less host shifting; and (2) pollinator experienced more cost in fitness during host shifting in dioecious fig wasps. In this study, I provide evidence showing that dioecious fig wasp does shift host, thus countering hypothesis 1. On the other hand, flowering time plasticity followed by disruption in male/female synchrony provides a possible mechanism of fitness cost specific to dioecious fig wasp mentioned in hypothesis 2.

doi:10.6342/NTU201903609

Chapter 2 The nuclear and mitochondrial genome of *Wiebesia* sp3 and *Wiebesia pumilae*

2.1 Introduction

2.1.1 Applying genomics in fig wasp study

Fig-fig wasp mutualism is a classic example of coevolution, and had been extensively studied for decades. Studies of figs and their associated fig wasps range from natural history, stability of mutualism, community assembly, and speciation (E. A. Herre et al., 2008). The closed system of fig syconia offers perfect opportunities for testing iconic evolutionary theories. Two such theories were sex allocation (Fellowes, Compton, & Cook, 1999) and parasite virulence evolution (E. A. Herre, 1993). In both cases, the fitness consequences, precision of adaptation and their association with fig wasp population structure are well-established. Although not affecting the power and beauty of both classic theory, clearly there are several knowledge gaps in understanding the full picture of population structure and its effects on fig wasps. For instance, exact mechanism (proximate cause) and evolutionary origin (ultimate cause) of sex allocation remained unknown. On the other hand, E. A. Herre et al., 2008 stated that differential selection pressure owing to parasite virulence in different wasp groups "provide remarkable opportunities to study the factors that shape fundamental processes in molecular evolution, …" Such statement provoked applying molecular tools on the classic system to answer fundamental evolutionary questions.

Another unsolved problem regarding fig-fig wasp coevolution is fine-scale co-speciation. Figs and fig wasps were roughly one to one in broad picture (*Ficus* sections vs agaonidae genus) (Cruaud et al., 2012). However, strict co-speciation rule does not apply in fine scale. There is mounting evidence that speciation of fig wasps is

doi:10.6342/NTU201903609

generally faster than the associated fig hosts (Jackson, Machado, Robbins, & Herre, 2008; Machado et al., 2005). The pattern of gene flow in both hosts and pollinators during and after pollinator speciation, mechanism of reproductive isolation, and co-adaptation under fine-scale co-speciation are largely unknown.

Genomic study is a starting point to bridge the gap and to explore all these evolutionary questions. The only published genome belonging to agaonidae is *Ceratosolen solmsi*, pollinator of *Ficus hispida* (Xiao et al., 2013). *C. solmsi* genome provides first glimpse into how long term specialized mutualistic life style shaped its evolution. For example, *C. solmsi* has a very compact protein-coding gene set, with reduction in gene families associated with detoxification and chemosensory systems. It is also the *C. solmsi* partial mitogenome that elucidated a potential premating isolation system caused by *Wolbachia* (Xiao et al., 2012). Such studies showed the importance of genomic study on understanding the evolutionary history of organisms.

2.1.2 Study species and system

One potential system for studying speciation, local adaptation, and population structure effects on fig wasp are the pollinators associated with *Ficus pumila*. *F. pumila* is a widely distributed fig species in East Asia. There are two described, phylogenetically distinct varieties in Taiwan (H.-Y. Wang et al., 2013). The endemic *F. p. var. awkeotsang* (jelly-fig) and the nominate *F. p. var. pumila* (creeping-fig). Jelly-fig is an economically important crop in Taiwan, for its dried drupes produce edible jelly, a popular dessert. Creeping-fig is a common gardening plant that had been widely cultivated, but it is also native to Taiwan.

Genetic study had shown that the two varieties and the three associated wasps (see chapter 1) fitted the classic scenario of which wasp speciation precedes fig speciation. In this case, *W. pumilae* and *W.* sp1, pollinators of *F. p. var. pumila* came into speciation

doi:10.6342/NTU201903609

through duplication (speciation of wasps but not their host), while there is no detectable genetic differentiation between their hosts (Yan Chen et al., 2012). Also, the estimated divergence time between *F. p. var. awkeotsang* and *F. p. var. pumila* is also much shorter than their pollinators (H.-Y. Wang et al., 2013). The different stages of recent codivergence makes them suitable model for fine-scale cospeciation and coadaption study.

The two fig varieties differed in several ecological traits, which possibly results in differential selection pressure in their pollinators during and after codivergence. These includes habitant preferences (Hsieh et al., 1993), population structure (Yong Chen et al., 2002), and floral scents (Y. Y. Chen & Wu, 2010; You-ling Chen et al., 2016). Jelly-fig occurred at altitudes around 1200-1900m, while creeping-fig occur in lowlands. Transplantation of jelly-fig to lowlands proved viable for the plants, but collapse of pollinating fig wasp population was often noted by farmers. Phenotypic adaptation to cold environment and signatures of selection on mitochondrial COI gene in jelly-fig wasp were also noted (H.-Y. Wang et al., 2013). For population structural effects, although there are currently no ecological data about sex ratio or nematode virulence in these species, the huge gall flower number differences between two fig varieties suggest that there is a foundress number variation between species, making these species potential system for revisiting the two classic theory at molecular level. My study of fig wasp olfactory coadaptation with host floral scent is presented in chapter 3.

Here I present the draft genome of two *Wiebesia* species. *W.* sp3 (jelly-fig wasp), pollinator of *F. p. var. awkeotsang*, was *de novo* assembled with a combined approach using long read contigs as backbone and short paired-end reads for polishing. *W. pumilae* (creeping-fig wasp) was assembled by a mapping-based approach using jelly-fig wasp genome as a reference. In addition, I reconstructed the near complete

doi:10.6342/NTU201903609

mitochondrial genomes of both species from short reads, representing the most complete mitochondrial genomes in fig wasp. The genomes shed light on questions regarding the ecology, coevolution, and cospeciation of fig wasps.

2.2 Material and Methods

2.2.1 Biological materials for genomic DNA sequencing

Near D phase male jelly-fig syconia were collected in December 2017, from a jelly-fig farmland in Pingtung, Taiwan (altitude: 670M, 22.45 N, 120.43 E). The farmland was surrounded by natural forest, and the pollinating wasps were shared among cultivated and wild figs. Male creeping-fig syconia were collected in December 2017 from Chiayi, Taiwan (altitude: 90M, 23.33 N, 120.28 E).

The collected syconia were brought back to lab, which were then split in half by knife for the wasps to emerge. Living fig wasps were collected once they leave the galls, and were washed with double-distilled water. The fresh samples were stored in 100% ethanol before DNA extraction. Fig wasps are haplodiploid, which means that male wasps came from unfertilized eggs, and only contains genetic information of all the foundresses from previous generation. In order to reduce the genetic diversity in samples, male individuals of jelly-fig wasp from one single syconium were pooled for DNA extraction.

Sex ratio is highly female-biased in agaonid wasps, and because creeping-fig bear much less gall flower per syconia than jelly-fig, collecting enough male individuals from one syconium to yield enough DNA was unsuccessful. As a consequence, female creeping-fig wasps from one syconium were used instead. Genomic DNA was extracted from pooled individuals using Purgene kit (Qiagen, USA) using protocol suggested by the manufacturer.

doi:10.6342/NTU201903609

2.2.2 Biological materials for RNA sequencing

The following samples from five different life stages of each species were sequenced for gene prediction purposes: (1) adult female, (2) adult male, (3) pupal female, (4) pupal male, (5) larva. All samples were freshly taken and stored in liquid nitrogen before RNA extraction. RNA was extracted using TRIzol (ThermoFisher, USA) following the manufacturer's protocols. Detailed information about collection site, sequencing depth, can be seen in Table 3.

2.2.3 Library construction and sequencing

Library preparation and sequencing were performed by Novogene, Beijing, China. Separate libraries that had insert size around 300-500bp were prepared for jelly-fig wasp and creeping-fig wasp DNA. A total of 75'957'927 and 92'826'608 150bp paired-end reads were generated for each species using an Illumina Hiseq 4000 platform (Table 3). The same extracted DNA from jelly-fig wasp was used in creating library for single molecule long read sequencing (SMRT). Single molecule sequencing was done on a PacBio Sequel platform and yielded a total of 28 Gb filtered polymerase read base. RNA libraries with 400bp insert size were constructed and sequenced on Illumina Hiseq 4000 platform (Table 3).

2.2.4 Sequencing quality check

Raw Illumina pair-end (PE) reads from genomic sequencing of both species were quality trimmed by Trimmomatic v0.36 (Bolger, Lohse, & Usadel, 2014). The head and tail of reads were trimmed if they were below a quality score of 3. Reads with an average per base quality under 15 in a 4 base sliding window would be cut. Reads that were less than 36bp were discarded. Quality assessment was done by FastQC v0.10.1 (Andrews, 2010) before and after Trimmomatic trimming.

doi:10.6342/NTU201903609

2.2.5 Taxonomic validation

Studies had reported that creeping-fig syconia may sometimes contain jelly-fig wasp individuals, although inviability and failure to leave syconia as adults were also noted (Jiang, 2011; H.-Y. Wang et al., 2013). The location where creeping-fig wasp were collected in this study is far from natural habitants of jelly-figs, so the chance of pooling individuals from different species was low. Several large scale molecular genetic surveys had confirmed fixed difference between the studied species in the mitochondrial cytochrome c oxidase subunit I (COI) gene (Yan Chen et al., 2012; H.-Y. Wang et al., 2013). The high coverage of mitochondrial sequences within short reads library also made it suitable to test if individuals from different species were pooled, since every individual are highly likely to be sequenced. To validate the purity of pooled individuals, trimmed PE reads of creeping-fig wasps were mapped to the published COI sequence (National Center for Biotechnology Information (NCBI) accession: JN184048.1) (Yan Chen et al., 2012) with BWA v0.7.10-r789 (H. Li, 2013) mem. The mapped bam file was then manually examined under the genome browser Integrative Genomics Viewer (IGV) (J. T. Robinson et al., 2011), to see if there were any foreign reads present.

2.2.6 Read decontamination and *de novo* assembly of jelly-fig wasp

SMRT reads of jelly-fig wasp were first assembled using wtdbg v1.1.006 (Ruan, n.d.). The parameters used were " -k 0 -p 21 -S 4 --edge-min 4 ". SMRT read-based and PE read-based polish were applied to increase the accuracy of the initial assembly. For SMRT read-based polish, the reads were mapped on to the contigs by pbalign ("Pacific Biosciences," n.d.) , and polished using arrow ("Pacific Biosciences," n.d.). PE reads were aligned to arrow-polished contigs using BWA mem, and polished by pilon

doi:10.6342/NTU201903609

v1.22 (Walker et al., 2014).

The initial draft contained contigs from symbiotic organisms that was revealed by FastQC report of reads (Figure 9) and the unusual bimodal distributed GC content of the initial assembly (Figure 11). Identification of contamination, and reads partition according to their biological source into bins were done using blobtools v1.01 (Laetsch & Blaxter, 2017). Briefly, PE and SMRT reads were mapped to the polished contigs using BWA mem and blasr ("Pacific Biosciences," n.d.) respectively. Polished contigs were blast (Altschul, Gish, Miller, Myers, & Lipman, 1990) against NCBI nucleotide database ("Database resources of the National Center for Biotechnology Information," 2012), with parameter "-max_target_seqs= 10". Given the blast result, each contig was assigned to a taxonomic position at every taxonomic rank by blobtools using the parameter "-x bestsum ". Blobtools then outputs blobplot based on the taxon, GC content, and coverage (based on PE reads) of contigs. Criteria for filtering the wasp reads were based on the blobplot. Contigs that matches at least one of the following criteria get past filtration. First, contigs that were assigned to the phylum Arthropoda. Second, contigs that had a GC content < 0.4 and mean coverage >13. The reads that mapped to those passed contigs were extracted using samtools v1.7 (H. Li, 2011) as decontaminated reads. Decontaminated PE reads of jelly-fig wasp were used to estimate the genome size based on 21mer counts using jellyfish 2.0 (Marçais & Kingsford, 2011). Plot of kmer frequency was draw using R (Figure 12).

Decontaminated jelly-fig wasp SMRT reads were assembled by wtdbg. To see if the decontamination leaded to over-cleaning and loss of genomic information, the decontaminated assembly was first polished the same way as described above. Busco v3 (Simão, Waterhouse, Ioannidis, Kriventseva, & Zdobnov, 2015) genome-mode analysis using hymenoptera_odb9 (Zdobnov et al., 2017) dataset were performed on the

doi:10.6342/NTU201903609

polished initial and polished decontaminated assembly to compare the integrity of each assembly.

For further polishing the assembly, pilon was run in an iterative way until the number of changes in each round became saturated (total six times). According to pilon report, the pilon-polished genome still contained some level of misassembly that cannot be corrected. Thus, I used GATK v4.0.3 (Van der Auwera et al., 2013) on the alignment to perform variant calling based on the GATK Best Practices. I then update the assembly according to the called variants using GATK.

2.2.7 Mapping-based assembly of creeping-fig wasp

To characterize the amount of symbiotic DNA within the creeping-fig wasp library, the trimmed short reads were first *de novo* assembled with minia (-k 61). The assembly was then evaluated using blobtools. For mapping-based assembly, the jelly-fig wasp genome, which was *de novo* assembled, served as a backbone. PE reads of creeping-fig wasp were mapped to jelly-fig wasp genome using BWA mem. Variants from alignment file were called using samtools. Consensus sequence in fastq format was generated by bcftools v1.3 (H. Li, 2011) with a upper and a lower coverage cutoff. The minimal and maximum coverage for calling are a third and twice the mean coverage, as recommended by PSMC (H. Li & Durbin, 2011). The sequence fastq was converted to fasta by seqtk v1.2 (H. Li, n.d.). Because bcftools did not output indel information to the consensus, pilon was applied iteratively to further polish the creeping-fig wasp assembly.

To evaluate the assemblies, Quast v4.6.3 (Gurevich, Saveliev, Vyahhi, & Tesler, 2013) was used to calculate the assembly statistics; blobtools were used for detecting contamination, and busco for evaluating the completeness of orthologs found in the assembly. Contaminations detected by blobtools at this stage were excluded before

doi:10.6342/NTU201903609

annotation.

2.2.8 Repeat annotation

Repeat searching based on homology was done by RepeatMasker v4.0 (Smit & Hubley, n.d.-a), the database used was order Hymenoptera from Repbase (Bao, Kojima, & Kohany, 2015). A *de novo* repeat library was constructed using RepeatModeler v1.0.11 (Smit & Hubley, n.d.-b), and analyzed by RepeatMasker. RepeatProteinMask provided by ReapeatMasker was used to identify transposable elements. Repeat-masked genome were created by bamtools v2.5.1 (Barnett, Garrison, Quinlan, Strömberg, & Marth, 2011).

2.2.9 Gene prediction

The genes were predicted in several ways. For transcriptome based prediction, RNAseq reads from five life stages were mapped to the genome by hisat2 v2.1 (Kim, Langmead, & Salzberg, 2015). Each library's alignment was assembled independently using stringtie v1.3 (Pertea et al., 2015). The assembled transcripts of each library were combined by stringtie merge command. Transdecoder v5.3 (Haas et al., 2013) was used to predict the ORF of transcripts, with the parameter "retain_blastp_hits", which took account of blastp results of Uniprot (UniProt Consortium, 2012) database.

For homologous gene prediction, gene sets from five close hymenoptera species, C. solmsi (Xiao et al., 2013), Nasonia vitripennis (Werren et al., 2010), Copidosoma floridanum (i5K Consortium, 2013), Trichogramma pretiosum (Lindsey et al., 2018), and Apis mellifera (Consortium, 2006) were downloaded from NCBI ftp. Non-redundant gene set for each species was generated by taking the longest isoform as representative of each gene using a custom Perl script, which was then aligned them to the genome using genblasta v1.0.4 (She, Chu, Wang, Pei, & Chen, 2009). The aligned

doi:10.6342/NTU201903609

region, along with their downstream and upstream 3 Kbp, was sent to genewise v2.2 (Birney, Clamp, & Durbin, 2004) for gene model prediction.

For *ab initio* gene prediction, augustus v3.3.1 (Stanke & Morgenstern, 2005), SNAP (Korf, 2004), geneid v1.3 (Blanco, Parra, & Guigó, 2007), fgenesh server (Solovyev, Kosarev, Seledsov, & Vorobyev, 2006), and glimmerHMM v3.0.1 (Majoros, Pertea, & Salzberg, 2004) were applied to repeat-masked assemblies. Augustus was run under training parameters generated by the genome mode of busco using hymenoptera odb9 dataset. Intron hints generated from RNAseq data were also provided to augustus. Geneid and Fgenesh were run using built-in *Nasonia* parameters. GlimmerHMM and SNAP were trained by *C. solmsi* gene set.

Protein alignment, RNAseq alignment, Transdecoder-predicted ORF, and *ab initio* predictions were sent to Evidencemodeler v1.1.1 (EVM) (Haas et al., 2008) for consensus gene calling. Different combinations of evidence weight were tested for each EVM run, and the result of each run were evaluated by the number of predicted genes, average gene length, and busco.

After getting a primary result of EVM. Genes that were identified as complete buscos from this first consensus gene set were extracted. These genes served as high confident gene models to re-train trainable *ab initio* predictors. A second round of EVM was done using the new re-trained predictions, along with other evidence used in the first round to generate a final EVM consensus gene set.

2.2.10 Manual gene curation

Web Apollo v2.1.0 (Lee et al., 2013) were used to manually curate the gene predictions. The gene models predicted by EVM, genewise, *ab initio* software, and transdecoder were loaded into Web Apollo as tracks. The manual curation mainly focused on fixing gene fragmentation or gene fusions, and eliminate those genes that

doi:10.6342/NTU201903609

have low evidence support. The curation started from EVM predicted models, while taking transdecoder, genewise, augustus and other evidence into account. Renaming of the manually curated genes according to their coordinates was done using scripts provided by Maker2 v2.31.0 (Holt & Yandell, 2011).

2.2.11 Functional annotation

Blast2GO (Gotz et al., 2008) v5.2.1 pipeline was applied for functional annotation, which integrates blast and interproscan (Jones et al., 2014) annotation results and produces gene ontology (GO) annotation. Gene sets of the two species were blasted against the UniprotKB/Swiss-prot database. Interproscan were used to search the following databases: CDD-3.16,Coils-2.2.1 (Marchler-Bauer et al., 2015), Gene3D-4.2.0 (Lees et al., 2012), Hamap-2018_03 (Lima et al., 2009), MobiDBLite-1.5 (Necci, Piovesan, Dosztányi, & Tosatto, 2017), Pfam-31.0 (Finn et al., 2016), PIRSF-3.02 (C. H. Wu et al., 2004), PRINTS-42.0 (Attwood et al., 2012), ProDom-2006.1 (Bru et al., 2004), ProSitePatterns-2018_02 (Sigrist et al., 2012), ProSiteProfiles-2018_02, SFLD-3 (Akiva et al., 2014), SMART-7.1 (Letunic, Doerks, & Bork, 2012), SUPERFAMILY-1.75 (de Lima Morais et al., 2011), and TIGRFAM-15.0 (Haft et al., 2012). Online server of BlastKOALA (Minoru Kanehisa, Sato, & Morishima, 2016) were used to map the gene sets to KEGG (M Kanehisa & Goto, 2000) database.

2.2.12 Manual annotation of interested genes

Manual annotation of genes involved in chemosensory (see 2.6.1) and venoms were done using homology-based approach. The only Chalcid wasp whose venom profile had been well-studied is the model species *N. vitripennis* (Sim & Wheeler, 2016; Werren et al., 2010). Blastp searches were done for the genes of jelly-fig wasp and

doi:10.6342/NTU201903609

creeping-fig wasp using *N. vitripennis* venom genes as database, with an e-value cutoff at 1e-15. Additional annotation of signal peptide and transmembrane helix was done using SignalP (Petersen, Brunak, von Heijne, & Nielsen, 2011) and TMHMM (Krogh, Larsson, von Heijne, & Sonnhammer, 2001) online server.

2.2.13 Mitochondrial genome assembly and annotation

An initial attempt to retrieve the mitochondrial genome of jelly-fig wasp from *de novo* assembly turned out unsuccessful. Alternatively, mapping based iterative assembly with MITObim v1.9.1 (Hahn, Bachmann, & Chevreux, 2013) were used to assemble the mitochondrial genome from PE reads for both species. For computational efficiency, 5’000’000 PE reads were randomly sampled from Trimmomatic trimmed data using seqtk. The reads were then transformed into interleaved format using bash commands. The reads and a seed sequence of COI from each species (NCBI accession of seed sequences: JN184048.1 (Yan Chen et al., 2012) for creeping-fig wasp; HQ398112.1 (Liu et al., 2014) for jelly-fig wasp) were given to MITObim and assembled with the parameter “-paired”. To validate the assembly, the input reads were mapped back to the assembled contigs with BWA mem and manually checked under IGV. Regions near both ends where there was no coverage support were manually trimmed. The trimmed contig then served as a new seed to repeat the assembly process. The second round assembly result was trimmed as mentioned above then polished with pilon. The assembled mitochondrial genomes were annotated using MITOS2 server (Bernt et al., 2013).

To study the evolution of mitochondrial synteny in chalcid wasps, mitochondrial genomes of all the families of chalcid wasps available on NCBI were downloaded and examined manually, which included *C. solmsi* (Agaonidae) (Xiao et al., 2012), *N. vitripennis* (Pteromalidae) (Oliveira, Raychoudhury, Lavrov, & Werren, 2008),

doi:10.6342/NTU201903609

Encarsia formosa (Aphelinidae) (Zhu et al., 2018), *Eupelmus* sp. (Eupelmidae) (Tang et al., 2019), *Eurytoma* sp. (Eurytomidae) (Tang et al., 2019), *Torymus* sp. (Torymidae) (Tang et al., 2019), *Brachymeria* sp. (Chalcididae) (Tang et al., 2019), and *Gonatocerus* sp. (Mymaridea). The gene arrangements events were inferred based on maximum parsimony and their known phylogeny (Munro et al., 2011).

2.2.14 Orthology

Orthology between jelly-fig wasp, creeping-fig wasp, and *C. solmsi* were designated by best blast reciprocal hits (BBRH) of amino acid sequences implemented in proteinortho v5.16 (Lechner et al., 2011). To determine orthology in a broader evolutionary scale, the online service of OrthoDB was used to assign IDs from OrthoDB v9 to the gene sets of both species. OrthoDB online service was also used to plot the shared orthology between jelly-fig wasp, creeping-fig wasp, other Chalcidoidea species, and *A. mellifera*.

2.2.15 Evolutionary rates between *Wiebesia* species

All 9352 1:1 orthologs of creeping-fig wasp and jelly-fig wasp were aligned by codon using the aligner prank v.14 (Löytynoja, 2014), which left 9218 sequences with complete triplet ORF. The alignments were filtered by Gblocks 0.91b (Castresana, 2000) under the parameter: -t=c -e=-gb1 -b4=6 -b5=a. dN /dS ratio of filtered alignments were calculated using pairwise codeml from_paml 4.9 (Z. Yang, n.d.), with parameter: CodonFreq = 0. The two species are so closely related that many genes had no synonymous substitutions. To adjust for this matter, for any gene pair that has dS value equals to 0, an adjusted dN/ dS ratio was calculated using the genome wide median dS. Gene pairs that had dS > 5 were excluded from downstream analysis to prevent poor alignments or saturation of mutations. A list of genes that had adjusted dN / dS > 1 were

doi:10.6342/NTU201903609

created for GO enrichment analysis.

To detect signatures of positive selection, codeml site model was tested on all gene pairs using model M1a (neutral evolution) vs M2a (positive selection), and M7 vs M8. P values were calculated using log likelihood ratio test. Correction for multiple testing were done with Benjamini & Hochberg method (Benjamini & Hochberg, 1995) using R, with false discovery rate (FDR) set to 0.05. GO enrichment analysis were performed on the dN /dS > 1, M1a vs M2a, and M7 vs M8 significant sets using Blast2GO, with a FDR cutoff 0.05. GOs used for enrichment analysis were based on the GO annotation of jelly-fig wasp gene set.

2.3 Results

2.3.1 Sequencing quality check and validation of taxonomy

28.0 Gbp of SMRT reads were sequenced from jelly-fig wasp. Mean read length is 9,863 bp. Median read length is 7,823 bp. 22.8 Gbp of jelly fig wasp pair-end reads were sequenced, 98.50% of the reads past Trimmomatic in pairs. For creeping-fig wasp PE reads, 99.57% out of the total 28.0 Gbp reads past in pairs. GC content plot from fastQC report revealed multimodal distribution which suggests that different degree of foreign DNA is present in both species (Figure 9 & Figure 10). Manual examination of reads mapped to COI sequences under IGV did not find any jelly-fig wasp haplotype, indicating no evidence of jelly-fig wasp contamination in creeping-fig wasp library.

2.3.2 Blobtools and *de novo* assembly of jelly-fig wasp

The initial *de novo* assembly of jelly-fig wasp by wtdbg yielded a genome with the size of 418Mbp, which is larger than the published *C. solmsi* genome (294M). A bimodal distributed GC content was seen in quast report, suggesting contamination of foreign DNA (Figure 11). Blobplot draw by blobtools revealed several sources of

doi:10.6342/NTU201903609

contaminations, including 6 near-full sized prokaryotic OTU assembly, other less well assembled bacterial species (Table 4), and some eukaryotes single cellular and multicellular parasites. (Figure 13). After applying decontamination the raw libraries with blobtools, 47.9% of the PE reads and 69.0% of the SMRT reads past the filtering. The genome size estimated from kmer count is 334M (Figure 12). The average coverage of SMRT reads is 57.8x, and 32.7x for PE reads in jelly-fig wasp.

Assembly of decontaminated SMRT reads yield a 332.9Mbp genome. Busco report of the second round assembly and the initial assembly showed that only 5 busco genes (0.1%) went from complete to fragmented after the decontamination. After the iterative polishing steps, 2 additional contamination contigs were identified by blobtools. One belonged to proteobacteria, and the other from amoeba. 24 more contigs were identified as contaminations after manual curation and functional annotation. All these contaminations were removed from the final assembly.

The final assembly is 323.8Mbp (Table 5), that spread across 230 contigs (N50=18.6 Mbp), with low GC content (30.16%). The rate of complete Hymenoptera buscos in genome mode is 93.7%. The assembly statistic and busco report suggest that this genome is comparable with the genomes and gene sets of other Chalcidoidea species. (Table 6). Around 97% of the genome had been assembled according to the estimated size.

2.3.3 Genome assembly of creeping-fig wasp

The average coverage of creeping-fig PE reads mapped to jelly-fig wasp genome is 75X. I assembled a genome that is 330 Mbp and span across 188 contigs (Table 5). The rate of complete busco is 93.3%.

doi:10.6342/NTU201903609

2.3.4 Repeat annotations and Gene predictions

Around 10% of repetitive sequences were identified and masked in jelly-fig wasp (details in Table 7) and creeping-fig wasp. The final consensus gene set generated by EVM contained 13342 and 13684 protein coding genes for jelly-fig and creeping-fig wasp, with mean cds length at 1358 and 1296. After manual curation, 10072 and 10037 genes were left, mean gene length were 1725 and 1728. Complete percentage of Busco genes of the two species went from 93.0% to 95.5% and 92.8% to 95.7%. The gene counts and length were close to the previous published agaonid wasp genome *C. solmsi*. Summary of gene models predicted by each program were at Table 8.

2.3.5 Functional annotation

Of all the 10072/10037 protein-coding genes in jelly-fig wasp and creeping-fig wasp, 8309/8252 had UniProt annotation, 9528/9498 had interproscan match, 8552/8252 had at least one GO annotation from Blast2GO and 5606/5576 had KEGG annotation (Figure 14). Manual annotation of venom proteins found 458 and 423 genes in jelly-fig and creeping-fig wasp, 214 and 196 of them had signal peptide.

2.3.6 Mitochondrial genome

The iterative mapping based assembly resulted near complete mitochondrial genome of jelly-fig wasp and creeping-fig wasp. The length of the two mitochondrial genomes is 16797 and 13630. This is the most complete mitochondrial genome ever sequenced within the fig wasp family. All 13 protein coding genes (PCGs) were present in both genomes, and all tRNAs and rRNAs were recovered in jelly-fig wasp (Table 9). After comparing the synteny with other published mitochondrial genome in Chalcidoidea, a unique gene arrangement of tRNA K and tRNA D was identified. The synteny pattern is found exclusively on the two species, and is not seen in other chalcid

doi:10.6342/NTU201903609

species including the fig wasp *C. solmsi*. Three types of syteny of tRNA K and tRNA D were discovered in Chalcidoidea. The possible evolutionary routes that lead to the current synteny pattern in Agaonidae were summarized in Figure 15.

2.3.7 Orthology

9352 single-copy orthologs were identified by BBRH between jelly-fig wasp and creeping-fig wasp. 8186 of them are among the two and *C. solmsi*. 9149 jelly-fig wasp and 9187 creeping-fig wasp genes were assigned to a known OrthoDB ID by the OrthoDB server. The expansion and contraction of genes compared with other Chalcidoidea genomes and European honey bee were presented in Figure 16.

2.3.8 Evolutionary rate between *Wiebesia* species

643 out of 9352 single-copy ortholog pairs had an adjusted dN/dS ratio over 1. Several GOs that are associated with serine protease and respiratory chain complex IV were enriched in the dN/dS > 1 set given a 0.05 significant FDR level (Table 10). 279 and 270 genes were significant at 0.05 p-value in M1a vs M2a and M7 vs M8 tests. After correction for multiple testing, 57 and 50 genes had FDR over 0.05. GO ID: 0004252, serine-type endopeptidase activity is enriched in the two site tests before FDR correction. No GO is enriched after FDR correction for site tests.

2.4 Discussion

2.4.1 *de novo* genome assembly and symbionts DNA removal

I sequenced the complete genome of jelly-fig wasp and creeping-fig wasp, which is the second and third genome that belong to pollinating fig wasps (Agaonidae). The N50 statistics, number of contigs and busco results indicated that these two genomes are comparable with other published genomes.

The initial assembly of jelly-fig wasp turned out to be the holobiome of wasp and

doi:10.6342/NTU201903609

symbionts. My samples were collected directly from wild populations and the DNA was extracted from whole body tissue. Most other wasp genome projects used laboratory inbreeding strains and often treated with antibiotics for some time to eliminate microbiomes. That is nearly impossible for fig wasps due to its mutualistic life cycle and very short adult life span. The genome project of fig wasp *C. solmsi* also used whole body tissue from male individuals for DNA extraction, but did not report such serious issue of contamination. Microbiome abundance may differ between taxa, as seen in my case of jelly-fig wasp and creeping-fig wasp. Removing abdominal segment prior to DNA extraction may reduce the amount of contaminations from microbes in sequencing.

Blobtools was highly successful in filtering the contaminations. The continuity of long read assembly was the key that the origin taxon of contigs can be confidently recovered by blast. One precious byproduct is the high quality microbiome data, which provides opportunity for further research on the holobiome of fig wasps.

The proportion of DNAs that came from symbionts varies in short reads and long reads sequencing results. Size selection during long read library preparation process is probably the reason that caused this differences. The original proportion of microbe DNA in short read library may be even more than observed in raw reads due to PCR bias, since most of the symbiotic prokaryotes found have high GC content.

Another interesting finding is that the microbiome composition and abundance seem to differ greatly between jelly-fig wasp and creeping-fig wasp, as shown by the per sequence GC content. I estimated the amount of contamination in creeping–fig wasp short read library using the following equation:

doi:10.6342/NTU201903609

$$1 - \frac{mean\ \ depth \times genome\ \ size}{total\ \ sequenced\ \ nucleotides}$$

In contrast to the over 50% of symbionts' DNA presented in the jelly-fig wasp short read sequencing (estimated by blobtools), the contamination presented in creeping-fig wasp was less than 10%. There are two non-mutually exclusive hypotheses to explain this drastic difference. One is that the difference lies between species, the other is that the microbiome is sex specific. A recent unpublished sequencing project from my lab used female jelly-fig wasps seemed to support both hypotheses. GC content of the raw reads in that library suggest a microbiome composition similar to female creeping-fig wasps, but the amount of foreign DNA is more abundant, which was similar to male jelly-fig wasps.

A few symbiont's contigs that were not found by blobtools were discovered later during the annotation phase. Some were found by homology based functional annotation like blast. A common property of symbiont contigs that can be seen during manual annotation is that the coding regions were absent of RNA coverage. The protocol of RNA library construction used poly T to extract messenger RNA, so the prokaryotic symbiont RNA were not present in RNAseq library.

2.4.2 Genome assembly using a closely related reference

My sequencing and assembly strategy of creeping-fig wasp was mapping-based assembly. It is a cost-effective way and require little computational resources. Also, it can produce highly contiguous assembly, with only one short read library needed. Disadvantages of mapping-based assembly include that unique sequences in the genome and sequences failed to be assembled in the reference may not present in the final results, and the synteny may not be totally correct due to divergence. By applying pilon

doi:10.6342/NTU201903609

iteratively, some of the mentioned issues can be fixed due to the local reassembly property. Actually, some paralogs are only seen in creeping-fig wasp and seemed not to be due to annotation bias. Nevertheless, the number of complete busco genes in both assembled genome and annotated gene set suggest that most of the coding regions are covered in the mapping-based assembly. An alternative approach for making use of closely related species genome is reference-guided *de novo* assembly (Lischer & Shimizu, 2017), which utilizes the power of De Bruijn graph based assembly and coordinates from close-related genome. While reference-guided *de novo* assembly can yield unique sequences not present in the reference, it requires libraries of different insert sizes including mate-pairs. The AT-rich property of fig wasp may also cause trouble for solving De Bruijn graph and result in a more fragmented assembly.

2.4.3 Reference-based RNA assembly

I used hisat2-stringtie pipeline to do reference-based RNA assembly. Which was very fast compared to the classic tophat-cufflink pipeline. One pitfall of reference-based RNA assembly is that neighboring genes can be assembled into fused gene. Thankfully, this error does not affect the ORFs. However, because ORF predictors such as transdecoder by default only predict the longest and most supported ORF per gene, the ORF that came from the shorter gene in the fused gene will be ignored. Also, EVM require *ab initio* predictions or “other evidence” such as transdecoder-predicted ORFs to provide intron/exon model and initiate its evidence weighing process. If *ab initio* predictors also failed to predict gene model in the region where the short gene lied, it will result in an odd situation that even though you did sequence the very mRNA from one gene, you are still unable to predict it from the genome using EVM. One way to solve the fused gene problem is to make use of *de novo* assembled transcriptome by mapping them back to the genome using spliced aligner (Singh et al., 2017), and feed it

doi:10.6342/NTU201903609

as “other evidence” to EVM.

2.4.4 Performance of *ab initio* predictors

While some *ab initio* predictors can only be trained by built-in species parameters (GeneID, fgenesh), some can be trained with the user’s own datasets (augustus, glimmerHMM, SNAP). Three types of training options were available for my project. Reference-based RNA assembled genes, busco-predicted parameters, and published gene sets of closely-related species. Busco-predicted parameter and RNAseq assembled gene all seem quite effective for training when their prediction results were evaluated by busco. Using an iterative training strategy, I successfully improved the predictions of each predictors and produced a more accurate consensus gene set.

2.4.5 Quality of gene predications

An astonishing outcome of gene prediction is that the mapping-based assembled creeping-fig wasp genome had a more complete busco gene set than jelly-fig wasp. After evaluating the completeness of all the evidence used by EVM through busco, I found that the RNAseq is more complete in creeping-fig wasp than jelly-fig wasp, which was likely the deterministic factor of the final gene prediction outcome. One method that can potentially increase the prediction is cross-referencing the two close species. Using spliced alignment programs such as PASA and align each species’ gene set onto another’s genome, then used it in manual curation, may greatly increase the sensitivity of predictions in the final curated gene sets.

2.4.6 Mitochondrial genome assembly

I tried several *de novo* assembly strategies using different sequencing libraries at hand, but all failed to produce anything close to an actual mt genome. The methods I attempted were *de novo* long read assembly (mecat, wtdbg, smartdenovo), short read *de*

doi:10.6342/NTU201903609

novo assembly (minia), and *de novo* transcriptome assembly (trinity). The contigs that contain mt PCGs or rRNAs are either fragmented, or had several weird tandem repetitive genes. Only the mapping-based iterative assembly method using short reads worked, but the yielded contig was much shorter than expected. After mapping the reads back to the contig and examining it under genome browser, I found out that at both ends of the contigs there were regions without any coverage support, which I assume were misassembly. So I cut off those regions, use the trimmed contig as a new "seed" sequence to re-run MITObim, which produce a much more complete assembly. Nuclear mitochondrial DNA segments (NUMTs) were probably the reason of misassembly. It has been reported that jelly-fig wasp processes more NUMTs than creeping-fig wasp (Yan Chen, Liu, Compton, & Chen, 2014), but whether this biological phenomenon is associated with my struggle in mt genome assembly remains unclear.

2.4.7 Gene rearrangements in mitochondrial genome

Mitochondrial synteny in Chalcidoidea is known to evolve at very high rate. A novel synteny pattern of tRNA K and tRNA D only found in *Wiebesia* was discovered. *C. solmsi* is one of the only two member of Chalcidoidea that process the insect ancestral pattern of tRNA K, tRNA D synteny, (L. Chen et al., 2018; Xiao, Jia, Murphy, & Huang, 2011; Xiao et al., 2012), the other one is the Mymarid wasp *Gonatocerus* sp.. Mymaridae is the sister group to all other chalcid wasps (Munro et al., 2011), so the identical synteny pattern probably came from homoplasy. I proposed several possible routes that lead to the current pattern in Figure 15.

2.4.8 Evolutionary rate between two species

The two species of wasp I studied diverged through allopatric cospeciation. Creeping-fig and its wasp came to secondary contact with jelly-fig after crossing the

doi:10.6342/NTU201903609

land bridge of Taiwan Strait sometime during the ice age. Multiple barriers of gene flow between the wasps and figs were observed today, including habitant isolation, discrete flowering time, asymmetrical host preference, failure to reach maturity in wrong host, and post-pollination inviability of seeds (Yong Chen et al., 2002; Jiang, 2011). Selection against hybrid, if any, can only happen in narrow secondary contact zones due to habitant isolation. Thus, divergence in the two fig wasp genome observed in my study should mostly came from drift and local adaptation, including coevolution with host figs (Cook & Segar, 2010).

2.4.9 Mito-nuclear coevolution

2 major groups of GOs were enriched in the dN/dS > 1 gene sets. One group is the GOs that are associated with respiratory chain complex IV, and the other is serine-type peptidase. It has been reported that nuclear-encoded oxidative phosphorylation (OXPHOs) genes in Hymenoptera evolved at faster rates compared to other major endopterygota insect orders, with higher amino acid substitution rate and dN/dS ratio (Y. Li et al., 2017). Genome project of 3 *Nasonia* species revealed that nuclear OXPHOs pathway not only had significantly elevated dN/dS ratio in species comparisons, but they are also contributed to post-zygotic, cyto-nuclear incompatibility between *N. vitripennis* and *N. giraulti* (Werren et al., 2010). In addition, dN/dS ratio of mitochondrial encoded COI between creeping-fig wasp and jelly-fig wasp were found to be elevated and deviated from neutral expectations (H.-Y. Wang et al., 2013). This background information along with the elevated dN/dS in respiratory chain complex IV suggest the role of cyto-nuclear coevolution in shaping the genomic landscape of fig wasps, and may potentially cause post-zygotic incompatibility when the two species hybrid.

doi:10.6342/NTU201903609

2.4.10 Rapid evolution and changes of expression in protease genes

Serine-type peptidase is the only GO that is enriched in both the M1a vs M2a and M7 vs M8 selection test before correction for multiple testing, and is also enriched in dN/dS > 1 and differentially expressed genes across two species of female adults (for cross-species expression analysis, see 3.2.8 for methods and 3.3.5 for results). The large number of genes belonging to this GO probably provide the statistical power needed to detect those signatures of selection. Biological function of this category of genes is very diverse in insects, which include development, digestion, venom, and immunity (Kanost & Jiang, 2015; Martinson, Hackett, Machado, & Arnold, 2015; Zou, Lopez, Kanost, Evans, & Jiang, 2006). To find out the exact function of each gene is very difficult for non-model organisms, but the expression profile can provide some clues. Serine-type peptidase that are expressed in adult females are less likely to be involved in development, nor in digestion since the adults possibly don't eat. That leaves venom and immunity as the most likely function of these differentially expressed genes.

Unlike their parasitoid ancestors, female pollinating fig wasps use their venom to induce gall formation during oviposition (Grandi, 1920). Coevolution or evolutionary arm race due to conflict of interests between fig wasp and fig host may explain the rapid evolutionary and expressional changes seen between species. I listed a crude annotation of possible venom genes, although the annotation probably overpredict the number of actual genes, due to the similarity between venom protease and non-venom protease genes. Also, many venom genes in parasitoid wasps derived from non-venom progenitors (Martinson, Mrinalini, Kelkar, Chang, & Werren, 2017), and will not be found by homology-based approaches. Further study on the venom composition using venom gland transcriptome and proteome can provide insights into the mechanism of gall formation and stability of mutualism.

doi:10.6342/NTU201903609

The two wasps occupy different habitats, and are likely to face threats from different pathogens/symbionts, as shown in the different microbiome abundance. Also, the number of gall flowers per syconium and average number of foundress differed between the two fig hosts (Yong Chen et al., 2002). According to the theory of parasite virulence evolution (E. A. Herre, 1993), the two wasp species face different intensity of parasite virulence, which can also affect selection pressure on immune genes. A detailed annotation of insect innate immunity pathways, and expressional study of nematode infection is needed to further explore the hypothesis. Other coevolving traits such as diet adaptation (D.-D. Wu, Wang, Irwin, & Zhang, 2009) are also possible in contributing selection to serine-type peptidase that are not expressed in adult females.

doi:10.6342/NTU201903609

Chapter 3 The evolution of chemosensory genes in pollinating fig wasps

3.1 Introduction

3.1.1 Basis of host recognition in pollinating fig wasps

Fig (*Ficus* spp.) and pollinating fig wasp (family Agaonidae) mutualism is an ancient, diverse and the most studied nursery pollination system (Galil & Eisikowitch, 1971). Pollination behavior of fig wasps not only determines the gene flow of host figs, but also the gene flow of wasp themselves. Pollinator attraction and host preference that coped well with one another create a positive feedback that exaggerates both phenotype toward higher host specificity, a phenomenon similar to Fisherian runaway selection (Herre et al., 2008). Similar to the case of sexual selection, premating isolation resulting from pollinator attraction/host preference interaction can contribute to speciation and in the long run, influence the diversity of the taxa involved (Herre, 1996). Like most nursery pollinators, pollinating fig wasps are attracted primary by floral scents (Hossaert-McKey, Soler, Schatz, & Proffit, 2010). Diversification of floral volatile compounds is associate with differential host preferences in fig wasps, and in many cases contributes to pre-mating gene flow barrier between close species (Grison-Pigé, Bessière, & Hossaert-McKey, 2002; Souto-Vilarós et al., 2018; Anthony B. Ware, Kaye, Compton, & Van Noort, 1993).

In contrast to the association of floral scent profiles and pollinator attraction in figs, which was extensively studied, only a handful of studies had focused on the intrinsic basis of host preference in pollinating fig wasps. Host preference is a complex behavioral trait, and probably involves multiple genes from different pathways. However, it is likely that the reception genes of scents and its associates--the

doi:10.6342/NTU201903609

chemosensory gene families, plays an important role in host preference evolution, as it is seen in multiple cases of chemosensory speciation (Brand et al., 2015; Matsuo, Sugaya, Yasukawa, Aigaki, & Fuyama, 2007; McBride, 2007; Shiao et al., 2015). The first sequenced agaonid wasp genome of *Ceratosolen solmsi* provides first glimpse into chemosensory system of pollinating fig wasps. Most of the chemosensory gene families in *C. solmsi* experienced massive reduction in repertoire numbers, which was suggested might be due to the simple environment encountered in mutualistic life style. Other studies had focused on the role of expressional modifications in chemosensory genes of pollinating fig wasps during host sensing (N. Wang et al., 2014; Zeng & Yu, 2018). However, genetic basis behind host preference during co-diversification in fig-fig wasp mutualism is still unclear. Therefore, a comparative study on fig wasps' chemosensory gene families from diverged lineages is necessary.

3.1.2 Study system and design

I provide a comparative study of chemosensory genes on four pollinating fig wasp species: *Elisabethiella stueckenbergi*, *Ceratosloen solmsi*, *Wiebesia pumilae*, and *Wiebesia* sp3. My study aimed at three different aspect of pollinating fig wasp evolution. In the first part, I explore the fine scale evolutionary changes during fig-fig wasp co-speciation using two species from genus *Wiebesia*. *W. pumilae* (creeping-fig wasp) and *W.* sp3 (jelly-fig wasp) are two closely-related, recently diverged species (0.23-0.36 mya). The two species arose allopatrically in South China and Taiwan (see Chapter 1). Their speciation is accompanied by the co-cladogenesis of their host figs: *Ficus pumila* var. *pumila* (creeping-fig), and *F. p.* var. *awkeotsang* (jelly-fig) which are two sibling lineages. Previous studies shown that the two figs are different in compositions of floral volatile compounds, and the wasps were attracted by the scent of their natural host (Y. Y. Chen & Wu, 2010; You-ling Chen et al., 2016). Asymmetrical affinity to each other's

doi:10.6342/NTU201903609

host was found in these two wasps. Both empirical (see Chapter 1) and behavioral study (Jiang, 2011) pointed out that jelly-fig wasp can be attracted to the female and male syconia of creeping-fig; while creeping-fig wasp do not respond to the male syconia of jelly-fig. The recent co-speciation of the two lineage pairs as well as ample chemical ecological and behavioral data provides a great opportunity to study the underlying mechanism of host preference in fine evolutionary scale. My study uses whole-genome and whole transcriptome data to capture evolution of chemosensory gene families at both amino acid sequence, copy number variation, and gene regulatory level.

In the second part, I focused on the long term effect of co-evolution on chemosensory genes in pollinating fig wasps. Co-phylogeny between major sections of *Ficus* and their associated agaonid genus shown that these co-cladogenesis were ancient and stable (Cruaud et al., 2012). To understand how the highly efficient and specific host-sensing co-adaption is accomplished in the long run given such contracted chemosensory gene families, I compared the chemosensory genes of two *Wiebesia* with *C. solmsi*, while testing the hypothesis of adaptive gene duplications. I also explore the association between local transposable element (TE) densities and gene duplication of olfactory receptors (OR), as TE has been reported to affect OR duplication in various insects (Harrison et al., 2018; McKenzie & Kronauer, 2018; Schrader et al., 2014; S. Yang et al., 2008).

Lastly, I studied the relationships between host sexual systems and chemosensory gene profiles. Ancestral state of fig is monoecy (Jousselin, Rasplus, & Kjellberg, 2003), while functional dioecy (hereafter: dioecy for short) evolved parallelly at least two times (Edward Allen Herre, Jandér, & Machado, 2008). Monoecious and dioecious fig/fig wasp differed in many ecological and speciation properties, including effective pollinating distance (Compton, Ellwood, Davis, & Welch, 2000; Nason, Herre, &

doi:10.6342/NTU201903609

Hamrick, 1998) and mutualism stability (L.-Y. Yang et al., 2015, and see 1.4.2). Dioecy evolved independently in the host figs of *C. solmsi* and *Wiebesia*. The convergent evolution provides great opportunity to examine the effect of sexual system on pollinating fig wasp's olfaction while excluding the bias caused by shared ancestry. I included transcriptome data from a monoecious fig pollinating wasp *E. stueckenbergi* and compared the evolution of chemosensory gene family size across the four species.

3.2 Material and Methods

3.2.1 Sample collection and RNA sequencing

Four adult female libraries of each *Wiebesia* species were sequenced as biological replicates for cross-species differential expression analysis. Adult female wasp samples were collected from D-phase syconia. The syconia were brought back to the lab and kept for the wasp to emerge. By doing so, the collected female wasps were not exposed to odors from receptive stage syconia. All samples were freshly taken and stored in liquid nitrogen before RNA extraction. RNA was extracted using TRIzol (ThermoFisher, USA) following the manufacturer's protocols. RNA libraries with 400bp insert size were constructed and sequenced on Illumina Hiseq 4000 platform by Novogene, Beijing, China. Detailed information about collection site, sequencing depth, can be seen in Table 3.

3.2.2 *de novo* transcriptome assembly of *E. stueckenbergi*

E. stueckenbergi transcriptome was downloaded from NCBI SRA database (SRR1502962). *de novo* transcriptome assembly was done by Trinity using default parameters. The assembly was evaluated by the perl script TrinityStats.pl provided by Trinity and by the transcript mode of busco using hymenoptera odb9 and endopterygota odb9 datasets.

doi:10.6342/NTU201903609

3.2.3 Reconstruction of fig wasp phylogeny

Protein sequences of *C. solmsi*, *C. floridanum* (Encyrtidae), and *N. vitripennis* (Pteromalidae) were downloaded from NCBI. Transcript isoforms were filtered so that only one longest isoform were left as a representative of each gene. Busco protein mode was run on gene sets of these 3 species as well the annotated gene sets of jelly-fig wasp and creeping-fig wasp (chapter 2) using the dataset hymenoptera odb9. Busco transcript mode was also run on the transcriptome of *E. stueckenbergi*; the results were filtered so that busco duplicates due to isoforms were viewed as single copy busco genes. Busco genes that are single copy in all six species were extracted and aligned with MAFFT. The alignments were then filtered with trimal using default parameters. All aligned busco genes were concatenated with catfasta2phyml.pl. Maximum likelihood tree of the concatenated busco genes was constructed using RAXML (-# 20 -m PROTGAMMAAUTO -T 10 -p 12345) with 100 bootstraps. The resulting tree was visualized under MEGA7.

3.2.4 Annotation of chemosensory genes

To annotate chemosensory genes in jelly-fig wasp and creeping-fig wasp, chemosensory genes of insect olfactory receptors (OR) (Zhou et al., 2015), gustatory receptors (GR) (Zhou et al., 2015), chemosensory proteins (CSP) (Kulmuni & Havukainen, 2013), odorant binding proteins (OBP) (Forêt & Maleszka, 2006; Vieira et al., 2012) and ionotropic glutamate receptors (iGluRs) (Croset et al., 2010) were collected from literatures. Chemosensory genes were annotated using two approaches. The first one is the homology-based gene prediction using genewise (see 2.2.9) based published sequences of chemosensory genes. The second approach is using blastp to search both the EVM predicted (see 2.2.9) and manual curated gene sets of jelly-fig

doi:10.6342/NTU201903609

wasp and creeping-fig wasp. Genes retrieved from the two approaches were manually combined using Web Apollo.

For annotation of ionotropic receptors (IR), the annotated iGluRs were aligned with the query sequences using MAFFT 7.407 (Katoh & Standley, 2013) with parameter: "--localpair --maxiterate 1000", and a maximum likelihood tree was made using RAXML 8.2.12 (Stamatakis, 2014) with parameter: "-# 20 -m PROTGAMMAAUTO" and 100 bootstrap value, genes clusters with known IRs were found through manually examine the phylogeny. The chemosensory gene annotations of *C. solmsi* were based on Zhou et al., 2015 (OR, GR), and my manual annotation (IR, CSP, OBP) from blast.

For annotation of chemosensory genes from transcriptome assembly of *E. stuenckenbergi*, tblastn (-evalue 1e-6) was used to search for chemosensory candidates from collected chemosensory sequences. Open reading frame of the resulting candidates were estimated using orfpredictor (Min, Butler, Storms, & Tsang, 2005). The predicted peptide sequences were reconfirmed using blastp (-evalue 1e-6). Finally, I manually pick up the isoform containing the longest cds for each gene as representatives.

3.2.5 Phylogenetic analysis of chemosensory genes

Each of the annotated chemosensory gene family of jelly-fig wasp, creeping-fig wasp, *C. solmsi*, *E. stueckenbergi*, and the outgroup *N. vitripennis* were aligned using MAFFT (--localpair --maxiterate 1000). A maximum likelihood tree was constructed with RAXML using parameters mentioned above. The tree was visualized and labeled using MEGA7 (Kumar et al., 2016).

3.2.6 Selection analysis

The phylogeny of fig wasps revealed lineage-specific gene expansions in OR. To see whether gene duplication events in fig wasps were associated with positive selection,

doi:10.6342/NTU201903609

monophyletic ortho-groups were identified based on best blast reciprocal hits employed in proteinortho and the maximum likelihood tree. Due to the theoretical incompleteness of transcriptome data, only annotations came from genomic data were involved in this part of the study. The ortho-groups were consisted of at least three sequence and two species. A simple ortho-group is consisted of 1:1 orthologs from the three species only, while a complex ortho-group contains paralogs. To determine the frequency of positive selection, branch-site tests were performed on all the ortho-groups with aBSREL (Smith et al., 2015). Fisher exact test was used to test whether branches under positive selection is significantly more common in complex ortho-groups than in simple ortho-groups.

3.2.7 Localization of transposable elements and odorant receptors within jelly fig wasp genome

Transposable elements (TE) had been reported to promote gene duplication in various species (Bailey, Liu, & Eichler, 2003; Harrison et al., 2018; McKenzie & Kronauer, 2018; Schrader et al., 2014; S. Yang et al., 2008). To see how local density of transposable elements is associated with recent lineage specific expansions of ORs, sliding window TE density was calculated on the jelly-fig wasp contigs that are longer than 1 million bps. TE annotations were based on the methods of (McKenzie & Kronauer, 2018). Specifically, *de novo* TE annotations from Repeatmodeler belonging to classified TE families were used. TE density along each contig was calculated in 10kbp sliding windows with 5kbp step using a custom R script. TE densities of flanking 10kbp region of OR loci were calculated. TE densities of OR loci from single copy ortho-groups were compared with OR loci from orthogroups with paralogs.

doi:10.6342/NTU201903609

3.2.8 Cross-species differential expression analysis on two *Wiebesia* species

The quality of RNA reads was assessed by fastQC. To prevent bias in differential expression analysis (Williams, Baccarella, Parrish, & Kim, 2016), no quality control trimming was done. For cross-species differential expression analysis, only single-copy orthologs were included (see 2.5.1). Transcript per million reads (TPM) of each gene in each library were estimated using Perl script "align_and_estimate_abundance.pl" provided by Trinity (Haas et al., 2013), with the abundance estimation program kallisto v0.44 (Bray, Pimentel, Melsted, & Pachter, 2016). Each species used their own sequence as templates for abundance estimation. Because the reference templates came from different species and had different gene length, TPM values were adjusted using the 'scaledTPM' method from Bioconductor package 'tximport' (Soneson, Love, & Robinson, 2016). Cross sample normalization was then performed using TMN method (M. D. Robinson, McCarthy, & Smyth, 2010) by Perl script "abundance_estimates_to_matrix.pl" (provided by Trinity). Biological replicates validation was done by correlation matrix, hierarchical clustering and principal component analysis (PCA) using R.

Differential expression analysis was performed using the bioconductor package edgeR (Ritchie et al., 2015; M. D. Robinson et al., 2010) (minimum log fold change = 4, false discovery rate (FDR) < 0.05). Significantly differentially expressed genes were then put to Blast2GO for GO enrichment analysis. GO annotation were based on functional annotation of jelly fig wasp. The reference set for GO enrichment test are all the genes that had counts per million (CPM) value larger than 1 in at least 2 samples. *p* value was calculated using Fisher exact test, with a cutoff FDR < 0.05. Enriched GOs

doi:10.6342/NTU201903609

were reduced to the most specific using Blast2GO.

3.3 Results

3.3.1 *de novo* transcriptome assembly

de novo assembly of *E. stuenckenbergi* using Trinity yielded 39697 genes and 51693 transcripts in a total of 59748971bps. Contig N50 based on longest isoform per gene is 1746bp. Median contig length based on longest isoform per gene is 390bp. Busco transcript mode reported 79.7% and 87.6% of complete buscos when searched using hymenoptera odb9 and endopterygota odb9 datasets respectively.

3.3.2 Phylogeny of studied fig wasps

2687 busco genes within the hymenoptera odb9 dataset were single copy in all six chalcid wasp species. A maximum likelihood tree based on concatenated single copy genes constructed by RAXML showed that *E. stuenckenbergi* is the sister group to the two *Wiebesia* species (Figure 17). All the branches have 100% bootstrap supports. This finding is consistent with the previous study (Cruaud et al., 2012). Phylogeny of fig wasps is incongruent with the phylogeny of their fig hosts. In fig wasp phylogeny, *E. stuenckenbergi* is sister to *W.* sp3 and *W. pumilae*; while in fig phylogeny, *F. pumila* (host of the two *Wiebesia* species) is more closely related to *F. hispica* (host of *C. solmsi*).

3.3.3 Evolution of chemosensory gene families in fig wasps

Repertoire sizes of chemosensory genes are quite conservative across the four fig wasps (Table 12). The gene family OR, GR, and OBP had reduction in size compared with the outgroup parasitoid wasp species *N. vitripennis*. However, given that *E. stueckenbergi* annotation was based on transcriptome data, which was much more incomplete compared with genomic data (As a reference, the complete busco value of

doi:10.6342/NTU201903609

all three other species' gene set are over 90% Table 6) the actual gene size should outnumber the observed (Brand & Ramírez, 2017). Even though *E. stueckenbergi* has a theoretically incomplete annotation, number of genes inside OR, GR, and OBP, the three reduced gene families, still outnumber the other three species. Phylogeny of OR, GR and OBP shown that the larger family size in *E. stueckenbergi* was due to more rapid gene expansion after the ancestral gene family size reduction (Figure 20 & Figure 21).

OR genes had an elevated mean dN/dS compared with non-chemosensory genes between the two *Wiebesia* species (Table 13). There is a dramatic difference when comparing the numbers of single-copy orthologous ORs in the two *Wiebesia* species and the three agaonid species with genomic data. 55 ORs are single-copy orthologous between *W.* sp3 and *W. pumilae*, while only 33 of them are single-copy ortholog among the three agaonid wasps (Table 13), indicating genus-specific gene expansions. Phylogenetic tree of ORs revealed that the major expansions came from local tandem repetitive blocks (Figure 18) Test for branches under positive selection revealed an significantly ($p < 0.001$) excess number of branches that are under selection in ortho-groups with paralogs than in single copy genes (Table 14). This pattern is consistent with the model of adaptive gene duplication.

3.3.4 TE density and gene duplication in OR

In *W.* sp3, TE densities around OR loci (mean=0.040) are not significantly larger than the genome-wide TE densities (mean=0.039) calculated from sliding window. There is also no enrichment of OR containing windows in windows with extremely high TE values (TE density > 95 percentiles). OR paralogs loci and OR loci that are single-copy in all fig wasp species with genome showed no significant difference in level of OR density (Figure 19). These results found no evidence supporting that tandem

doi:10.6342/NTU201903609

duplication of OR, whether recent or ancient, was facilitated by local transposable elements.

3.3.5 Cross species differential expression analysis

I performed PCA and hierarchical clustering on the TMN normalized expression values to validate biological replicates. PCA showed that the two species can be separated on PC1, which accounted for 41.6% of the total variation (Figure 22). Jelly-fig wasp and creeping-fig wasp nearly formed two distinct clades on hierarchical clustering (Figure 23).

8701 genes were expressed (CPM>1) in at least two samples. After DE analysis with edgeR, 432 genes were found to be differentially expressed between jelly-fig wasp and creeping-fig wasp. 118 were upregulated in jelly-fig wasp and 314 in creeping-fig wasp. GO enrichment analysis performed for the differentially expressed genes showed overrepresentation in GOs that are involved in olfaction, cellular signal transduction, serine type endopeptidase, and cuticle structure (Table 15). Chemosensory genes that were differentially expressed were summarized at Table 16. Notably, the odorant receptor coreceptor (Orco) was among the DE genes.

3.4 Discussion

Phylogenetic relationship of the four studied fig wasp is congruent with our current understand of their phylogeny. The mismatch between fig wasp species tree and species tree of their host probably reflects ancient host shifting events. One may argue the observed evolutionary changes of chemosensory genes in this study was not because of co-evolution with host scent, but some other kind of ecological factor. However, pollinating fig wasp spend almost its entire life time inside gall until emerge as adult, and the adult have only few days of life span. The extreme contraction of chemosensory

doi:10.6342/NTU201903609

genes showed there is reduced demand on olfactory functions as nursery pollinators (Xiao et al., 2013). Only on two occasions in their life cycle does olfaction plays a major role (N. Wang et al., 2014), which are localization of females done by males and host sensing done by females. Sexual selection due to mate preference may drive the divergence of chemosensory genes. However, sexual dimorphism in antenna morphology of pollinating fig wasps shows that olfactory function of the two sexes are under quite different selection pressure. The antennas are highly degenerated in males, with reduced structure and much fewer sensilla (Meng, Huang, Xiao, & Bu, 2016). On the other hand, elaborate antennal structure in females (Meng et al., 2016), repeated convergent evolution of the only known sensilla chaetica (an elongated multiparous plate sensilla found in male non-agaonid chalcid wasps that are assumed to evolve for better sensory sensitivity) in females chalcid wasps (A B Ware & Compton, 1992), and the diversity of host floral scents all suggest the importance of host sensing in agaonid wasps. Thus, I argue that host specificity is the major selective force in driving the evolution of chemosensory genes in agaonid wasps.

3.4.1 On the fine-scale evolution between jelly-fig and creeping-fig wasp

I used jelly-fig wasp and creeping-fig wasp as model to study fine-scale evolution of host preference for the following reasons. First, the host preference phenotypes of the two wasp species are well documented, both floral scent of host and responsive behavior had been studied (Y. Y. Chen & Wu, 2010; You-ling Chen et al., 2016; Jiang, 2011). Second, it is a case of allopatric cospeciation (H.-Y. Wang et al., 2013), meaning that most evolutionary changes came from genetic drift and local adaptation, including co-evolution with hosts after geographic isolation. Under this evolutionary framework, the divergence and differential expression in females found in chemosensory genes are highly likely the underlying genes involved in host preference behavior.

doi:10.6342/NTU201903609

Overall, the sizes of chemosensory gene families in jelly-fig wasp and creeping-fig wasp were similar to the published *C. solmsi* genome. Elevated mean dN/dS ratio were observed in OR gene family. 9 out of the 55 ortholog-pairs had dN/dS larger than 1. Directional positive selection due to co-adaptation to host scents may explain the observed changes, but relaxation of selection is also likely. Olfaction-associated GOs were enriched in the cross-species DE gene set, and all differentially expressed chemosensory genes were single copy in all three agaonid wasps. Orco, the ancient and conservative OR co-receptor that forms dimer complex with all other OR genes (Sato et al., 2008), is among one of the DE genes. My finding suggested that expression modifications on certain conservative genes is important in determining host preferencer. More chemosensory genes were up-regulated in creeping-fig wasp, a possible explanation to the increased sensitivity and host discrimination ability compared with jelly-fig wasp.

3.4.2 Long-term effects of co-cladogenesis

Most chemosensory gene families were conservative in orthology except OR. Genus-specific gene expansions accompanied by increased frequency of positive selection in OR was discovered. Gene duplication, followed by neo-functionalization is a classic model of novel gene function evolution (birth-death model) (Nei, Gu, & Sitnikova, 1997). Birth-death model in chemosensory genes had been discovered in many insect taxa, and is generally considered to be associated with important lineage-specific adaptations (Engsontia, Sangket, Chotigeat, & Satasook, 2014). In hymenoptera, such pattern was discovered in ants (Engsontia, Sangket, Robertson, & Satasook, 2015) and corbiculate bees (Brand & Ramírez, 2017). The findings in fig wasp OR likely represents adaptive evolution owing to co-adaption with host scents. Figs and pollinating fig wasps were roughly congruent in phylogeny in broad picture.

doi:10.6342/NTU201903609

Main sections of *Ficus* formed stable relationship with major genus of Agaonidae (Cruaud et al., 2012; Machado et al., 2005). The genus-specific expansions of OR are associated with these ancient co-radiation events. It is possible that adaptation to host scents during co-radiation were the cause of genus-specific expansions. However, since the study only include data from two genera, diverse taxa sampling is needed to test whether the OR expansions are indeed associated with the deep co-radiation event or the number of duplicated genes just gradually accumulate over time.

Unlike ants, the OR loci do not contain larger amounts of TEs in fig wasps. This result is unsurprising, since OR gene family experience great contraction in the lineage leading to the common ancestor of fig wasps. Fig wasp also has a very low proportion of repetitive DNA within their genome. Whether the low proportion of repeats directly contributes to a smaller gene set by lowering the gene duplication frequency or it is the confounding result of the same evolutionary force is not clear. Meanwhile, ORs that are recently duplicated do not harbor more TEs than other ORs, suggesting that OR expansion in fig wasps are mostly independent from TE involvement.

Combing the findings in fine-scale expressional and broad-scale evolution analysis, I propose a model of chemosensory gene evolution in pollinating fig wasps. Only a few chemosensory genes were left after the transition from ancestral parasitoid to nursey pollination life style in the common ancestor of pollinating fig wasps. Loss of chemosensory genes could be driven by relaxation of selection or positive selection. Some chemosensory genes evolved under the classic gene birth-death model during co-diversification with hosts. The other group that remained conservative in copy numbers and evolved slowly is the core gene set. Core gene set probably served in important biological functions (such as Orco), and were under strong purifying selection. Evolution of core gene set mainly based on regulatory changes to alternate phenotypes.

doi:10.6342/NTU201903609

3.4.3 Host sexual system and sizes of chemosensory gene family

The study of sexual system and fig wasp olfaction not only help us understand the co-evolutionary history and adaptation, but it may also shed light on how and why dioecy evolved multiple times in fig-fig wasp mutualism. However, the result here is just a preliminary study, since transcriptome data is too fragmented to make too much inferences. Nevertheless, the trend of larger post-contraction expansions in monoecious fig wasp is consistent in all the three contracted chemosensory gene families. Based on the results of broad-scale coevolution, chemosensory gene family expansions may reflect co-adaption with hosts. Faster accumulation of gene expansions, therefore, could mean that the scent of their mutualistic hosts changed more rapidly. This scenario is consistent with the hypothesis 2 mentioned in 1.4.2, which stated that monoecious fig wasps can have more frequent host shifting. Future study using genomic data from multiple monoecious fig wasp lineages that are closer to *Wiebesia* and *Ceratosolen* can provide better resolution of this issue.

doi:10.6342/NTU201903609

Chapter 4 Conclusion

In the first chapter, I discovered that there are in fact only three pollinator species associated with the widespread *Ficus pumila*. *Wiebesia pumilae, W.* sp1, and *W.* sp3 are all co-pollinators of *F. p. var. pumila*. However, the natural host of *W.* sp3 is *F. p. var. awkeotsang*, and *W.* sp3 populations in mainland were recently introduced by human activities. The south eastern shore of China are potential habitants for these invaders. Phenological separation of native and introduced wasp communities can be explained by phenotypic plasticity of host flowering time. Such plasticity may further explain the different co-pollinator pattern seen among monoecious and dioecious figs.

In the second chapter, I present the high quality nuclear and mitogenome of *W.* sp3 and *W. pumilae*. Mitogenomes of the two species reveal rapid evolution of synteny in fig wasp family. There are signatures of positive selection on serine type protease and OXPHOS complex IV genes. The former may reflect evolutionary arm races with fig hosts and pathogens. The latter may be explained by mito-nuclear coevolution, which is a likely mechanism for the rapid speciation rate of fig wasps.

In the third chapter, I explore the evolution of fig wasp chemosensory systems. I found that regulatory changes are responsible for fine-scale coevolution, while adaptive tandem gene duplication in olfactory receptors is associated with broad-scale coevolution. Preliminary study on fig wasp chemosensory system and host sexual system showed rapid gene expansions that may be associated with frequent host-switching in monecious fig wasps.

doi:10.6342/NTU201903609

REFERENCE

Akiva, E., Brown, S., Almonacid, D. E., Barber, A. E., Custer, A. F., Hicks, M. A., … Babbitt, P. C. (2014). The Structure–Function Linkage Database. *Nucleic Acids Research*, *42*(D1), D521–D530. https://doi.org/10.1093/nar/gkt1130

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Andrews, S. (2010). FastQC A Quality Control tool for High Throughput Sequence Data. Retrieved June 1, 2018, from https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Attwood, T. K., Coletta, A., Muirhead, G., Pavlopoulou, A., Philippou, P. B., Popov, I., … Mitchell, A. L. (2012). The PRINTS database: a fine-grained protein sequence annotation and analysis resource--its status in 2012. *Database*, *2012*(0), bas019-bas019. https://doi.org/10.1093/database/bas019

Bailey, J. A., Liu, G., & Eichler, E. E. (2003). An Alu Transposition Model for the Origin and Expansion of Human Segmental Duplications. *The American Journal of Human Genetics*, *73*(4), 823–834. https://doi.org/10.1086/378594

Bao, W., Kojima, K. K., & Kohany, O. (2015). Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA*, *6*(1), 11. https://doi.org/10.1186/s13100-015-0041-9

Barnett, D. W., Garrison, E. K., Quinlan, A. R., Strömberg, M. P., & Marth, G. T. (2011). BamTools: a C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics (Oxford, England)*, *27*(12), 1691–1692. https://doi.org/10.1093/bioinformatics/btr174

doi:10.6342/NTU201903609

Benjamini, Y., & Hochberg, Y. (1995). *Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing*. *Source: Journal of the Royal Statistical Society. Series B (Methodological)* (Vol. 57). Retrieved from http://engr.case.edu/ray_soumya/mlrg/controlling_fdr_benjamini95.pdf

Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., … Stadler, P. F. (2013). MITOS: Improved de novo metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution*, *69*(2), 313–319. https://doi.org/10.1016/j.ympev.2012.08.023

Birney, E., Clamp, M., & Durbin, R. (2004). GeneWise and Genomewise. *Genome Research*, *14*(5), 988–995. https://doi.org/10.1101/gr.1865504

Blanco, E., Parra, G., & Guigó, R. (2007). Using geneid to Identify Genes. In *Current Protocols in Bioinformatics* (Vol. 18, p. 4.3.1-4.3.28). Hoboken, NJ, USA: John Wiley & Sons, Inc. https://doi.org/10.1002/0471250953.bi0403s18

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Brand, P., & Ramírez, S. R. (2017). The Evolutionary Dynamics of the Odorant Receptor Gene Family in Corbiculate Bees. *Genome Biology and Evolution*, *9*(8), 2023–2036. https://doi.org/10.1093/gbe/evx149

Brand, P., Ramírez, S. R., Leese, F., Quezada-Euan, J. J. G., Tollrian, R., & Eltz, T. (2015). Rapid evolution of chemosensory receptor genes in a pair of sibling species of orchid bees (Apidae: Euglossini). *BMC Evolutionary Biology*, *15*(1), 176. https://doi.org/10.1186/s12862-015-0451-9

Bray, N. L., Pimentel, H., Melsted, P., & Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology*, *34*(5), 525–527.

doi:10.6342/NTU201903609

https://doi.org/10.1038/nbt.3519

Bru, C., Courcelle, E., Carrère, S., Beausse, Y., Dalmar, S., & Kahn, D. (2004). The ProDom database of protein domain families: more emphasis on 3D. *Nucleic Acids Research*, *33*(Database issue), D212–D215. https://doi.org/10.1093/nar/gki034

Castresana, J. (2000). Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. *Molecular Biology and Evolution*, *17*(4), 540–552. https://doi.org/10.1093/oxfordjournals.molbev.a026334

Chen, L., Chen, P.-Y., Xue, X.-F., Hua, H.-Q., Li, Y.-X., Zhang, F., & Wei, S.-J. (2018). Extensive gene rearrangements in the mitochondrial genomes of two egg parasitoids, Trichogramma japonicum and Trichogramma ostriniae (Hymenoptera: Chalcidoidea: Trichogrammatidae). *Scientific Reports*, *8*(1), 7034. https://doi.org/10.1038/s41598-018-25338-3

Chen, Y., Compton, S. G., Liu, M., & Chen, X.-Y. (2012). Fig trees at the northern limit of their range: the distributions of cryptic pollinators indicate multiple glacial refugia. *Molecular Ecology*, *21*(7), 1687–1701. https://doi.org/10.1111/j.1365-294X.2012.05491.x

Chen, Y., Huang, M., Wu, W., Wang, A., BAO, T., Zheng, C., … Tu, S. (2016). The floral scent of Ficus pumila var. pumila and its effect on the choosing behavior of pollinating wasps of Wiebesia pumilae. *Acta Ecologica Sinica*, *36*(5), 321–326. https://doi.org/10.1016/j.chnaes.2016.06.008

Chen, Y., Li, H. Q., & Ma, W. L. (2002). The reproductive character of Ficus pumila var. pumilar, F.pumila var. awkeotsang and their pollinators. *Acta Phytoecologica Sinica*, *26*(1), 58–63. Retrieved from http://scholar.googleusercontent.com/scholar?q=cache:rpDc96SYpZYJ:scholar.google.com/+pollination+ficus+pumila&hl=zh-TW&as_sdt=0,5

doi:10.6342/NTU201903609

Chen, Y., Liu, M., Compton, S. G., & Chen, X.-Y. (2014). Distribution of nuclear mitochondrial pseudogenes in three pollinator fig wasps associated with Ficus pumila. *Acta Oecologica*, *57*, 142–149. https://doi.org/10.1016/j.actao.2013.10.001

Chen, Y. Y., & Wu, W. W. (2010). Volatile compounds from the syconia of ficus awkeotsang makino and their attractiveness to pollinator wasps. *Shengtai Xuebao/ Acta Ecologica Sinica*, *30*(8), 2212–2219. Retrieved from http://www.ecologica.cn/stxb/ch/reader/view_abstract.aspx?file_no=stxb20090917 1221

Clement, M., Snell, Q., Walke, P., Posada, D., & Crandall, K. (2002). TCS: estimating gene genealogies. In *Proceedings 16th International Parallel and Distributed Processing Symposium* (p. 7 pp). IEEE. https://doi.org/10.1109/IPDPS.2002.1016585

Compton, S. G., Ellwood, M. D. F., Davis, A. J., & Welch, K. (2000). The Flight Heights of Chalcid Wasps (Hymenoptera, Chalcidoidea) in a Lowland Bornean Rain Forest: Fig Wasps are the High Fliers1. *Biotropica*, *32*(3), 515–522. https://doi.org/10.1111/j.1744-7429.2000.tb00497.x

Consortium, T. H. G. S. (2006). Insights into social insects from the genome of the honeybee Apis mellifera. *Nature*, *443*(7114), 931–949. https://doi.org/10.1038/nature05260

Cook, J. M., & Segar, S. T. (2010). Speciation in fig wasps. *Ecological Entomology*, *35*, 54–66. https://doi.org/10.1111/j.1365-2311.2009.01148.x

Croset, V., Rytz, R., Cummins, S. F., Budd, A., Brawand, D., Kaessmann, H., … Benton, R. (2010). Ancient protostome origin of chemosensory ionotropic glutamate receptors and the evolution of insect taste and olfaction. *PLoS Genetics*, *6*(8), e1001064. https://doi.org/10.1371/journal.pgen.1001064

doi:10.6342/NTU201903609

Cruaud, A., Ronsted, N., Chantarasuwan, B., Chou, L. S., Clement, W. L., Couloux, A., … Savolainen, V. (2012). An Extreme case of plant-insect codiversification: Figs and fig-pollinating wasps. *Systematic Biology*, *61*(6), 1029–1047. https://doi.org/10.1093/sysbio/sys068

Database resources of the National Center for Biotechnology Information. (2012). *Nucleic Acids Research*, *41*(D1), D8–D20. https://doi.org/10.1093/nar/gks1189

de Lima Morais, D. A., Fang, H., Rackham, O. J. L., Wilson, D., Pethica, R., Chothia, C., & Gough, J. (2011). SUPERFAMILY 1.75 including a domain-centric gene ontology method. *Nucleic Acids Research*, *39*(Database), D427–D434. https://doi.org/10.1093/nar/gkq1130

Engsontia, P., Sangket, U., Chotigeat, W., & Satasook, C. (2014). Molecular Evolution of the Odorant and Gustatory Receptor Genes in Lepidopteran Insects: Implications for Their Adaptation and Speciation. *Journal of Molecular Evolution*, *79*(1–2), 21–39. https://doi.org/10.1007/s00239-014-9633-0

Engsontia, P., Sangket, U., Robertson, H. M., & Satasook, C. (2015). Diversification of the ant odorant receptor gene family and positive selection on candidate cuticular hydrocarbon receptors. *BMC Research Notes*, *8*(1), 380. https://doi.org/10.1186/s13104-015-1371-x

Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, *10*(3), 564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

Excoffier, L., Smouse, P. E., & Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, *131*(2), 479–491. Retrieved from

doi:10.6342/NTU201903609

http://www.ncbi.nlm.nih.gov/pubmed/1644282

Fellowes, M. D. E., Compton, S. G., & Cook, J. M. (n.d.). Sex Allocation and Local Mate Competition in Old World Non-Pollinating Fig Wasps. *Behavioral Ecology and Sociobiology*. Springer. https://doi.org/10.2307/4601647

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*(12), 4302–4315. https://doi.org/10.1002/joc.5086

Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., … Bateman, A. (2016). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Research*, *44*(D1), D279–D285. https://doi.org/10.1093/nar/gkv1344

Forêt, S., & Maleszka, R. (2006). Function and evolution of a gene family encoding odorant binding-like proteins in a social insect, the honey bee (Apis mellifera). *Genome Research*, *16*(11), 1404–1413. https://doi.org/10.1101/gr.5075706

Galil, J., & Eisikowitch, D. (1971). Studies on mutualistic symbiosis between syconia and sycophilous wasps in monoecious figs. *New Phytologist*, *70*(4), 773–787. https://doi.org/10.1111/j.1469-8137.1971.tb02578.x

Gotz, S., Garcia-Gomez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., … Conesa, A. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research*, *36*(10), 3420–3435. https://doi.org/10.1093/nar/gkn176

Grandi, G. (1920). Studio morfologico e biologico della Blastophaga psenes (L.). Estudio morfológico y biológico de Blastophaga psenes (L.). *Bollettino Del Laboratorio Di Zoologia Generale Ed Agraria*, *14*, 60–204.

Grison-Pigé, L., Bessière, J.-M., & Hossaert-McKey, M. (2002). Specific Attraction of

doi:10.6342/NTU201903609

Fig-Pollinating Wasps: Role of Volatile Compounds Released by Tropical Figs. *Journal of Chemical Ecology*, *28*(2), 283–295. https://doi.org/10.1023/A:1017930023741

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics*, *29*(8), 1072–1075. https://doi.org/10.1093/bioinformatics/btt086

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., … Regev, A. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, *8*(8), 1494–1512. https://doi.org/10.1038/nprot.2013.084

Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., … Wortman, J. R. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biology*, *9*(1), R7. https://doi.org/10.1186/gb-2008-9-1-r7

Haft, D. H., Selengut, J. D., Richter, R. A., Harkins, D., Basu, M. K., & Beck, E. (2012). TIGRFAMs and Genome Properties in 2013. *Nucleic Acids Research*, *41*(D1), D387–D395. https://doi.org/10.1093/nar/gks1234

Hahn, C., Bachmann, L., & Chevreux, B. (2013). Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. *Nucleic Acids Research*, *41*(13), e129–e129. https://doi.org/10.1093/nar/gkt371

Harrison, M. C., Jongepier, E., Robertson, H. M., Arning, N., Bitard-Feildel, T., Chao, H., … Bornberg-Bauer, E. (2018). Hemimetabolous genomes reveal molecular basis of termite eusociality. *Nature Ecology & Evolution*, *2*(3), 557–566. https://doi.org/10.1038/s41559-017-0459-1

doi:10.6342/NTU201903609

Herre, E. A. (1993). Population Structure and the Evolution of Virulence in Nematode Parasites of Fig Wasps. *Science*, *259*(5100), 1442–1445. https://doi.org/10.1126/science.259.5100.1442

Herre, E. A. (1996). An Overview of Studies on a Community of Panamanian Figs. *Journal of Biogeography*, *23*(4), 593–607. Retrieved from http://www.jstor.org/stable/2845802

Herre, E. A., Jandér, K. C., & Machado, C. A. (2008). Evolutionary Ecology of Figs and Their Associates: Recent Progress and Outstanding Puzzles. *Annual Review of Ecology, Evolution, and Systematics*, *39*(1), 439–458. https://doi.org/10.1146/annurev.ecolsys.37.091305.110232

Holt, C., & Yandell, M. (2011). MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics*, *12*(1), 491. https://doi.org/10.1186/1471-2105-12-491

Hossaert-McKey, M., Soler, C., Schatz, B., & Proffit, M. (2010). Floral scents: their roles in nursery pollination mutualisms. *Chemoecology*, *20*(2), 75–88. https://doi.org/10.1007/s00049-010-0043-5

Hsieh, C. F., Huang, T. C., Li, Z. Y., Lo, H. C., Ohashi, H., Shen, C. F., … Yang, K. C. (1993). *Flora of Taiwan, Vol. 2* (2nd ed.). Taipei.

i5K Consortium, i5K. (2013). The i5K Initiative: advancing arthropod genomics for knowledge, human health, agriculture, and the environment. *The Journal of Heredity*, *104*(5), 595–600. https://doi.org/10.1093/jhered/est050

Jackson, A. P., Machado, C. A., Robbins, N., & Herre, E. A. (2008). Multi-locus phylogenetic analysis of neotropical figs does not support co-speciation with the pollinators: The importance of systematic scale in fig/wasp cophylogenetic studies. *SYMBIOSIS*, *45*, xx–xx. Retrieved from http://evolve.zoo.ox.ac.uk/software

doi:10.6342/NTU201903609

Jiang, S.-W. (2011). *Morphological differences between pollinating fig wasps of Ficus pumila L. var. pumila and var. awkeotsang (Makino) Corner and their asymmetric host specificity*. National Taiwan University.

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., … Hunter, S. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics (Oxford, England)*, *30*(9), 1236–1240. https://doi.org/10.1093/bioinformatics/btu031

Jousselin, E., Rasplus, J.-Y., & Kjellberg, F. (2003). Convergence and coevolution in a mutualism: evidence from a molecular phylogeny of ficus. *Evolution*, *57*(6), 1255–1269. https://doi.org/10.1111/j.0014-3820.2003.tb00334.x

Kanehisa, M., & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, *28*(1), 27–30. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10592173

Kanehisa, M., Sato, Y., & Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *Journal of Molecular Biology*, *428*(4), 726–731. https://doi.org/10.1016/j.jmb.2015.11.006

Kanost, M. R., & Jiang, H. (2015). Clip-domain serine proteases as immune factors in insect hemolymph. *Current Opinion in Insect Science*, *11*, 47–55. https://doi.org/10.1016/j.cois.2015.09.003

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772–780. https://doi.org/10.1093/molbev/mst010

Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nature Methods*, *12*(4), 357–360.

doi:10.6342/NTU201903609

https://doi.org/10.1038/nmeth.3317

Korf, I. (2004). Gene finding in novel genomes. *BMC Bioinformatics*, *5*(1), 59. https://doi.org/10.1186/1471-2105-5-59

Krogh, A., Larsson, B., von Heijne, G., & Sonnhammer, E. L. . (2001). Predicting transmembrane protein topology with a hidden markov model: application to complete genomes11Edited by F. Cohen. *Journal of Molecular Biology*, *305*(3), 567–580. https://doi.org/10.1006/jmbi.2000.4315

Kulmuni, J., & Havukainen, H. (2013). Insights into the evolution of the CSP gene family through the integration of evolutionary analysis and comparative protein modeling. *PloS One*, *8*(5), e63688. https://doi.org/10.1371/journal.pone.0063688

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution*, *33*(7), 1870–1874. https://doi.org/10.1093/molbev/msw054

Laetsch, D. R., & Blaxter, M. L. (2017). BlobTools: Interrogation of genome assemblies. *F1000Research*, *6*, 1287. https://doi.org/10.12688/f1000research.12232.1

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., … Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, *23*(21), 2947–2948. https://doi.org/10.1093/bioinformatics/btm404

Lechner, M., Findeiß, S., Steiner, L., Marz, M., Stadler, P. F., & Prohaska, S. J. (2011). Proteinortho: Detection of (Co-)orthologs in large-scale analysis. *BMC Bioinformatics*, *12*(1), 124. https://doi.org/10.1186/1471-2105-12-124

Lee, E., Helt, G. A., Reese, J. T., Munoz-Torres, M. C., Childers, C. P., Buels, R. M., … Lewis, S. E. (2013). Web Apollo: a web-based genomic annotation editing platform. *Genome Biology*, *14*(8), R93. https://doi.org/10.1186/gb-2013-14-8-r93

doi:10.6342/NTU201903609

Lees, J., Yeats, C., Perkins, J., Sillitoe, I., Rentzsch, R., Dessailly, B. H., & Orengo, C. (2012). Gene3D: a domain-based resource for comparative genomics, functional annotation and protein network analysis. *Nucleic Acids Research*, *40*(Database issue), D465-71. https://doi.org/10.1093/nar/gkr1181

Leigh, J. W., & Bryant, D. (2015). popart : full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, *6*(9), 1110–1116. https://doi.org/10.1111/2041-210X.12410

Letunic, I., Doerks, T., & Bork, P. (2012). SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Research*, *40*(D1), D302–D305. https://doi.org/10.1093/nar/gkr931

Li, H. (n.d.). seqtk: Toolkit for processing sequences in FASTA/Q formats. Retrieved from https://github.com/lh3/seqtk

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics (Oxford, England)*, *27*(21), 2987–2993. https://doi.org/10.1093/bioinformatics/btr509

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Retrieved from http://arxiv.org/abs/1303.3997

Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, *475*(7357), 493–496. https://doi.org/10.1038/nature10231

Li, Y., Zhang, R., Liu, S., Donath, A., Peters, R. S., Ware, J., … Zhou, X. (2017). The molecular evolutionary dynamics of oxidative phosphorylation (OXPHOS) genes in Hymenoptera. *BMC Evolutionary Biology*, *17*(1), 269. https://doi.org/10.1186/s12862-017-1111-z

doi:10.6342/NTU201903609

Lima, T., Auchincloss, A. H., Coudert, E., Keller, G., Michoud, K., Rivoire, C., … Bairoch, A. (2009). HAMAP: a database of completely sequenced microbial proteome sets and manually curated microbial protein families in UniProtKB/Swiss-Prot. *Nucleic Acids Research, 37*(Database issue), D471-8. https://doi.org/10.1093/nar/gkn661

Lindsey, A. R. I., Kelkar, Y. D., Wu, X., Sun, D., Martinson, E. O., Yan, Z., … Werren, J. H. (2018). Comparative genomics of the miniature wasp and pest control agent Trichogramma pretiosum. *BMC Biology, 16*(1), 54. https://doi.org/10.1186/s12915-018-0520-9

Lischer, H. E. L., & Shimizu, K. K. (2017). Reference-guided de novo assembly approach improves genome reconstruction for related species. *BMC Bioinformatics, 18*(1), 474. https://doi.org/10.1186/s12859-017-1911-6

Liu, M., Zhao, R., Chen, Y., Zhang, J., Compton, S. G., & Chen, X.-Y. (2014). Competitive Exclusion among Fig Wasps Achieved via Entrainment of Host Plant Flowering Phenology. *PLoS ONE, 9*(5), e97783. https://doi.org/10.1371/journal.pone.0097783

Löytynoja, A. (2014). Phylogeny-aware alignment with PRANK. In *Methods in molecular biology (Clifton, N.J.)* (Vol. 1079, pp. 155–170). https://doi.org/10.1007/978-1-62703-646-7_10

Machado, C. A., Robbins, N., Gilbert, M. T. P., & Herre, E. A. (2005). Critical review of host specificity and its coevolutionary implications in the fig/fig-wasp mutualism. *Proceedings of the National Academy of Sciences of the United States of America, 102 Suppl 1*(Suppl 1), 6558–6565. https://doi.org/10.1073/pnas.0501840102

Majoros, W. H., Pertea, M., & Salzberg, S. L. (2004). TigrScan and GlimmerHMM:

doi:10.6342/NTU201903609

two open source ab initio eukaryotic gene-finders. *Bioinformatics*, *20*(16), 2878–2879. https://doi.org/10.1093/bioinformatics/bth315

Marçais, G., & Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*, *27*(6), 764–770. https://doi.org/10.1093/bioinformatics/btr011

Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Geer, L. Y., … Bryant, S. H. (2015). CDD: NCBI's conserved domain database. *Nucleic Acids Research*, *43*(D1), D222–D226. https://doi.org/10.1093/nar/gku1221

Martinson, E. O., Hackett, J. D., Machado, C. A., & Arnold, A. E. (2015). Metatranscriptome Analysis of Fig Flowers Provides Insights into Potential Mechanisms for Mutualism Stability and Gall Induction. *PLOS ONE*, *10*(6), e0130745. https://doi.org/10.1371/journal.pone.0130745

Martinson, E. O., Mrinalini, Kelkar, Y. D., Chang, C.-H., & Werren, J. H. (2017). The Evolution of Venom by Co-option of Single-Copy Genes. *Current Biology*, *27*(13), 2007–2013.e8. https://doi.org/10.1016/j.cub.2017.05.032

Matsuo, T., Sugaya, S., Yasukawa, J., Aigaki, T., & Fuyama, Y. (2007). Odorant-Binding Proteins OBP57d and OBP57e Affect Taste Perception and Host-Plant Preference in Drosophila sechellia. *PLoS Biology*, *5*(5), e118. https://doi.org/10.1371/journal.pbio.0050118

McBride, C. S. (2007). Rapid evolution of smell and taste receptor genes during host specialization in Drosophila sechellia. *Proceedings of the National Academy of Sciences*, *104*(12), 4996–5001. https://doi.org/10.1073/pnas.0608424104

McKenzie, S. K., & Kronauer, D. J. C. (2018). The genomic architecture and molecular evolution of ant odorant receptors. *Genome Research*, *28*(11), 1757–1765. https://doi.org/10.1101/gr.237123.118

doi:10.6342/NTU201903609

Meng, J., Huang, D., Xiao, J., & Bu, W. (2016). Antennal Sensilla of Fig Wasps (Hymenoptera: Agaonidae): Function-Driven Elaboration in Females and Degeneration in Males. *Annals of the Entomological Society of America*, *109*(1), 99–105. https://doi.org/10.1093/aesa/sav084

Min, X. J., Butler, G., Storms, R., & Tsang, A. (2005). OrfPredictor: predicting protein-coding regions in EST-derived sequences. *Nucleic Acids Research*, *33*(Web Server), W677–W680. https://doi.org/10.1093/nar/gki394

Munro, J. B., Heraty, J. M., Burks, R. A., Hawks, D., Mottern, J., Cruaud, A., … Jansta, P. (2011). A Molecular Phylogeny of the Chalcidoidea (Hymenoptera). *PLoS ONE*, *6*(11), e27023. https://doi.org/10.1371/journal.pone.0027023

Nason, J. D., Herre, E. A., & Hamrick, J. L. (1998). The breeding structure of a tropical keystone plant resource. *Nature*, *391*(6668), 685–687. https://doi.org/10.1038/35607

Necci, M., Piovesan, D., Dosztányi, Z., & Tosatto, S. C. E. (2017). MobiDB-lite: Fast and highly specific consensus prediction of intrinsic disorder in proteins. *Bioinformatics*, *33*(9), btx015. https://doi.org/10.1093/bioinformatics/btx015

Nei, M., & Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution*, *3*(5), 418–426. https://doi.org/10.1093/oxfordjournals.molbev.a040410

Nei, M., Gu, X., & Sitnikova, T. (1997). Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(15), 7799–7806. https://doi.org/10.1073/PNAS.94.15.7799

Oliveira, D. C. S. G., Raychoudhury, R., Lavrov, D. V., & Werren, J. H. (2008). Rapidly Evolving Mitochondrial Genome and Directional Selection in

doi:10.6342/NTU201903609

Mitochondrial Genes in the Parasitic Wasp Nasonia (Hymenoptera: Pteromalidae). *Molecular Biology and Evolution*, *25*(10), 2167–2180. https://doi.org/10.1093/molbev/msn159

Pacific Biosciences. (n.d.). Retrieved from https://github.com/PacificBiosciences/

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T., & Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology*, *33*(3), 290–295. https://doi.org/10.1038/nbt.3122

Petersen, T. N., Brunak, S., von Heijne, G., & Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods*, *8*(10), 785–786. https://doi.org/10.1038/nmeth.1701

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, *190*, 231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026

Rago, A., Gilbert, D. G., Choi, J.-H., Sackton, T. B., Wang, X., Kelkar, Y. D., … Colbourne, J. K. (2016). OGS2: genome re-annotation of the jewel wasp Nasonia vitripennis. *BMC Genomics*, *17*(1), 678. https://doi.org/10.1186/s12864-016-2886-9

Rasplus, J. Y. (1996). The one-to-one species specificity of the Ficus-Agaoninae mutualism: how casual? In *The Biodiversity of African Plants* (pp. 639–649). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-009-0285-5_78

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, *43*(7), e47–e47. https://doi.org/10.1093/nar/gkv007

doi:10.6342/NTU201903609

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology*, *29*(1), 24–26. https://doi.org/10.1038/nbt.1754

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)*, *26*(1), 139–140. https://doi.org/10.1093/bioinformatics/btp616

Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., & Sánchez-Gracia, A. (2017). DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. *Molecular Biology and Evolution*, *34*(12), 3299–3302. https://doi.org/10.1093/molbev/msx248

Ruan, J. (n.d.). A fuzzy Bruijn graph approach to long noisy reads assembly. Retrieved from https://github.com/ruanjue/wtdbg

Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, *4*(4), 406–425. https://doi.org/10.1093/oxfordjournals.molbev.a040454

Sato, K., Pellegrino, M., Nakagawa, T., Nakagawa, T., Vosshall, L. B., & Touhara, K. (2008). Insect olfactory receptors are heteromeric ligand-gated ion channels. *Nature*, *452*(7190), 1002–1006. https://doi.org/10.1038/nature06850

Schrader, L., Kim, J. W., Ence, D., Zimin, A., Klein, A., Wyschetzki, K., … Oettler, J. (2014). Transposable element islands facilitate adaptation to novel environments in an invasive species. *Nature Communications*, *5*(1), 5495. https://doi.org/10.1038/ncomms6495

She, R., Chu, J. S.-C., Wang, K., Pei, J., & Chen, N. (2009). GenBlastA: enabling BLAST to identify homologous gene sequences. *Genome Research*, *19*(1),

doi:10.6342/NTU201903609

143–149. https://doi.org/10.1101/gr.082081.108

Shiao, M.-S., Chang, J.-M., Fan, W.-L., Lu, M.-Y. J., Notredame, C., Fang, S., … Li, W.-H. (2015). Expression Divergence of Chemosensory Genes between *Drosophila sechellia* and Its Sibling Species and Its Implications for Host Shift. *Genome Biology and Evolution*, *7*(10), 2843–2858. https://doi.org/10.1093/gbe/evv183

Sigrist, C. J. A., de Castro, E., Cerutti, L., Cuche, B. A., Hulo, N., Bridge, A., … Xenarios, I. (2012). New and continuing developments at PROSITE. *Nucleic Acids Research*, *41*(D1), D344–D347. https://doi.org/10.1093/nar/gks1067

Sim, A. D., & Wheeler, D. (2016). The venom gland transcriptome of the parasitoid wasp Nasonia vitripennis highlights the importance of novel genes in venom function. *BMC Genomics*, *17*(1), 571. https://doi.org/10.1186/s12864-016-2924-7

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Singh, R., Lawal, H. M., Schilde, C., Glöckner, G., Barton, G. J., Schaap, P., & Cole, C. (2017). Improved annotation with de novo transcriptome assembly in four social amoeba species. *BMC Genomics*, *18*(1), 120. https://doi.org/10.1186/s12864-017-3505-0

Smit, A., & Hubley, R. (n.d.-a). RepeatMasker Open-4.0. Retrieved from http://www.repeatmasker.org

Smit, A., & Hubley, R. (n.d.-b). RepeatModeler Open-1.0. Retrieved from http://www.repeatmasker.org

Smith, M. D., Wertheim, J. O., Weaver, S., Murrell, B., Scheffler, K., & Kosakovsky

doi:10.6342/NTU201903609

Pond, S. L. (2015). Less Is More: An Adaptive Branch-Site Random Effects Model for Efficient Detection of Episodic Diversifying Selection. *Molecular Biology and Evolution*, *32*(5), 1342–1353. https://doi.org/10.1093/molbev/msv022

Solovyev, V., Kosarev, P., Seledsov, I., & Vorobyev, D. (2006). Automatic annotation of eukaryotic genes, pseudogenes and promoters. *Genome Biology*, *7 Suppl 1*(Suppl 1), S10.1-12. https://doi.org/10.1186/gb-2006-7-s1-s10

Soneson, C., Love, M. I., & Robinson, M. D. (2016). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research*, *4*, 1521. https://doi.org/10.12688/f1000research.7563.2

Souto-Vilarós, D., Proffit, M., Buatois, B., Rindos, M., Sisol, M., Kuyaiva, T., … Segar, S. T. (2018). Pollination along an elevational gradient mediated both by floral scent and pollinator compatibility in the fig and fig-wasp mutualism. *Journal of Ecology*, *106*(6), 2256–2273. https://doi.org/10.1111/1365-2745.12995

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics (Oxford, England)*, *30*(9), 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Stanke, M., & Morgenstern, B. (2005). AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Research*, *33*(Web Server issue), W465-7. https://doi.org/10.1093/nar/gki458

Tang, P., Zhu, J., Zheng, B., Wei, S., Sharkey, M., Chen, X., & Vogler, A. P. (2019). Mitochondrial phylogenomics of the Hymenoptera. *Molecular Phylogenetics and Evolution*, *131*, 8–18. https://doi.org/10.1016/j.ympev.2018.10.040

UniProt Consortium. (2012). Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Research*, *40*(D1), D71–D75. https://doi.org/10.1093/nar/gkr981

doi:10.6342/NTU201903609

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., … DePristo, M. A. (2013). From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. In *Current Protocols in Bioinformatics* (Vol. 43, p. 11.10.1-11.10.33). Hoboken, NJ, USA: John Wiley & Sons, Inc. https://doi.org/10.1002/0471250953.bi1110s43

Vieira, F. G., Forêt, S., He, X., Rozas, J., Field, L. M., & Zhou, J.-J. (2012). Unique features of odorant-binding proteins of the parasitoid wasp Nasonia vitripennis revealed by genome annotation and comparative analyses. *PloS One*, *7*(8), e43034. https://doi.org/10.1371/journal.pone.0043034

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., … Earl, A. M. (2014). Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLoS ONE*, *9*(11), e112963. https://doi.org/10.1371/journal.pone.0112963

Wang, H.-Y., Hsieh, C.-H., Huang, C.-G., Kong, S.-W., Chang, H.-C., Lee, H.-H., … Wu, W.-J. (2013). Genetic and physiological data suggest demographic and adaptive responses in complex interactions between populations of figs (*Ficus pumila*) and their pollinating wasps (*Wiebesia pumilae*). *Molecular Ecology*, *22*(14), 3814–3832. https://doi.org/10.1111/mec.12336

Wang, N., Wang, N. X., Niu, L. M., Bian, S. N., Xiao, J. H., & Huang, D. W. (2014). Odorant-binding protein (OBP) genes affect host specificity in a fig-pollinator mutualistic system. *Insect Molecular Biology*, *23*(5), 621–631. https://doi.org/10.1111/imb.12110

Ware, A. B., & Compton, S. G. (1992). Repeated evolution of elongate multiporous plate sensilla in female fig wasps (Hymenoptera: Agaonidae: Agaoninae). In *Proc*

doi:10.6342/NTU201903609

K Ned Akad Wet (Vol. 95, pp. 275–292).

Ware, A. B., Kaye, P. T., Compton, S. G., & Van Noort, S. (1993). Fig volatiles: Their role in attracting pollinators and maintaining pollinator specificity. *Plant Systematics and Evolution*, *186*(3–4), 147–156. https://doi.org/10.1007/BF00940794

Werren, J. H., Richards, S., Desjardins, C. A., Niehuis, O., Gadau, J., Colbourne, J. K., … Gibbs, R. A. (2010). Functional and evolutionary insights from the genomes of three parasitoid Nasonia species. *Science (New York, N.Y.)*, *327*(5963), 343–348. https://doi.org/10.1126/science.1178028

Wiebes, J. T. (1979). Co-Evolution of Figs and their Insect Pollinators. *Annual Review of Ecology and Systematics*, *10*(1), 1–12. https://doi.org/10.1146/annurev.es.10.110179.000245

Williams, C. R., Baccarella, A., Parrish, J. Z., & Kim, C. C. (2016). Trimming of sequence reads alters RNA-Seq gene expression estimates. *BMC Bioinformatics*, *17*(1), 103. https://doi.org/10.1186/s12859-016-0956-2

Wu, C. H., Nikolskaya, A., Huang, H., Yeh, L.-S. L., Natale, D. A., Vinayaka, C. R., … Barker, W. C. (2004). PIRSF: family classification system at the Protein Information Resource. *Nucleic Acids Research*, *32*(90001), 112D–114. https://doi.org/10.1093/nar/gkh097

Wu, D.-D., Wang, G.-D., Irwin, D. M., & Zhang, Y.-P. (2009). A Profound Role for the Expansion of Trypsin-Like Serine Protease Family in the Evolution of Hematophagy in Mosquito. *Molecular Biology and Evolution*, *26*(10), 2333–2341. https://doi.org/10.1093/molbev/msp139

Xiao, J.-H., Jia, J.-G., Murphy, R. W., & Huang, D.-W. (2011). Rapid evolution of the mitochondrial genome in Chalcidoid wasps (Hymenoptera: Chalcidoidea) driven

doi:10.6342/NTU201903609

by parasitic lifestyles. *PloS One*, *6*(11), e26645. https://doi.org/10.1371/journal.pone.0026645

Xiao, J.-H., Wang, N.-X., Murphy, R. W., Cook, J., Jia, L.-Y., & Huang, D.-W. (2012). Wolbachia infection and dramatic intraspecific mitochondrial DNA divergence in a fig wasp. *Evolution*, *66*(6), 1907–1916. https://doi.org/10.1111/j.1558-5646.2011.01561.x

Xiao, J.-H., Yue, Z., Jia, L.-Y., Yang, X.-H., Niu, L.-H., Wang, Z., … Huang, D.-W. (2013). Obligate mutualism within a host drives the extreme specialization of a fig wasp genome. *Genome Biology*, *14*(12), R141. https://doi.org/10.1186/gb-2013-14-12-r141

Yang, L.-Y., Machado, C. A., Dang, X.-D., Peng, Y.-Q., Yang, D.-R., Zhang, D.-Y., & Liao, W.-J. (2015). The incidence and pattern of copollinator diversification in dioecious and monoecious figs. *Evolution; International Journal of Organic Evolution*, *69*(2), 294–304. https://doi.org/10.1111/evo.12584

Yang, S., Arguello, J. R., Li, X., Ding, Y., Zhou, Q., Chen, Y., … Wang, W. (2008). Repetitive Element-Mediated Recombination as a Mechanism for New Gene Origination in Drosophila. *PLoS Genetics*, *4*(1), e3. https://doi.org/10.1371/journal.pgen.0040003

Yang, Z. (n.d.). PAML 4: Phylogenetic Analysis by Maximum Likelihood. https://doi.org/10.1093/molbev/msm088

Zdobnov, E. M., Tegenfeldt, F., Kuznetsov, D., Waterhouse, R. M., Simão, F. A., Ioannidis, P., … Kriventseva, E. V. (2017). OrthoDB v9.1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. *Nucleic Acids Research*, *45*(D1), D744–D749. https://doi.org/10.1093/nar/gkw1119

doi:10.6342/NTU201903609

Zeng, J., & Yu, H. (2018). Transcriptome analysis of genes involved in the response of a pollinator fig wasp to volatile organic compounds from its host figs. *Acta Oecologica*, *90*, 91–98. https://doi.org/10.1016/J.ACTAO.2018.01.003

Zhou, X., Rokas, A., Berger, S. L., Liebig, J., Ray, A., & Zwiebel, L. J. (2015). Chemoreceptor Evolution in Hymenoptera and Its Implications for the Evolution of Eusociality. *Genome Biology and Evolution*, *7*(8), 2407–2416. https://doi.org/10.1093/gbe/evv149

Zhu, J., Tang, P., Zheng, B.-Y., Wu, Q., Wei, S., & Chen, X. (2018). The first two mitochondrial genomes of the family Aphelinidae with novel gene orders and phylogenetic implications. *International Journal of Biological Macromolecules*, *118*(Pt A), 386–396. https://doi.org/10.1016/j.ijbiomac.2018.06.087

Zou, Z., Lopez, D. L., Kanost, M. R., Evans, J. D., & Jiang, H. (2006). Comparative analysis of serine protease-related genes in the honey bee genome: possible involvement in embryonic development and innate immunity. *Insect Molecular Biology*, *15*(5), 603–614. https://doi.org/10.1111/j.1365-2583.2006.00684.x

doi:10.6342/NTU201903609

FIGURES

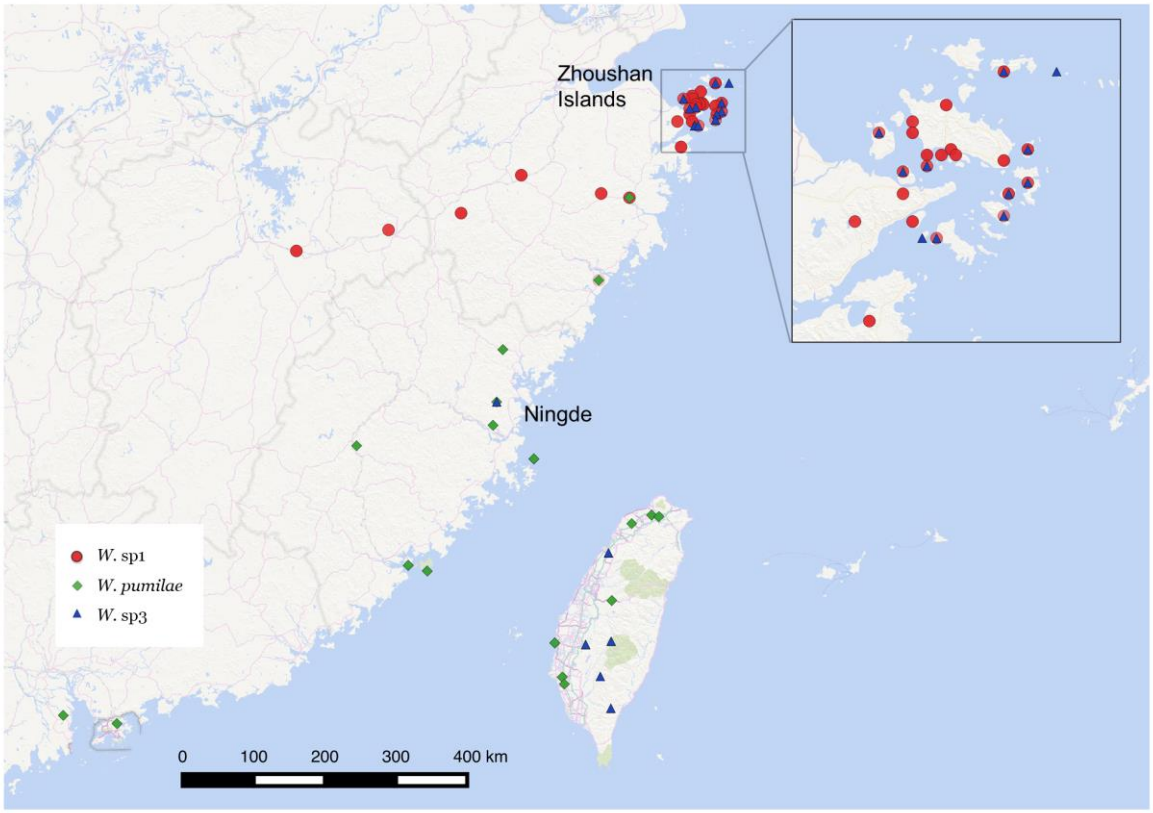

Figure 1 Geographical distribution of *Wiebesia* samples from published literatures. *W.* sp3 occur in Taiwan and two non-adjacent sites in mainland. All the *W.* sp3 in Taiwan and one individual in Ningde (JN184049) was found within *Ficus pumila var. awkeotsang*. All the other wasp samples came from *Ficus pumila var. pumila*.

doi:10.6342/NTU201903609

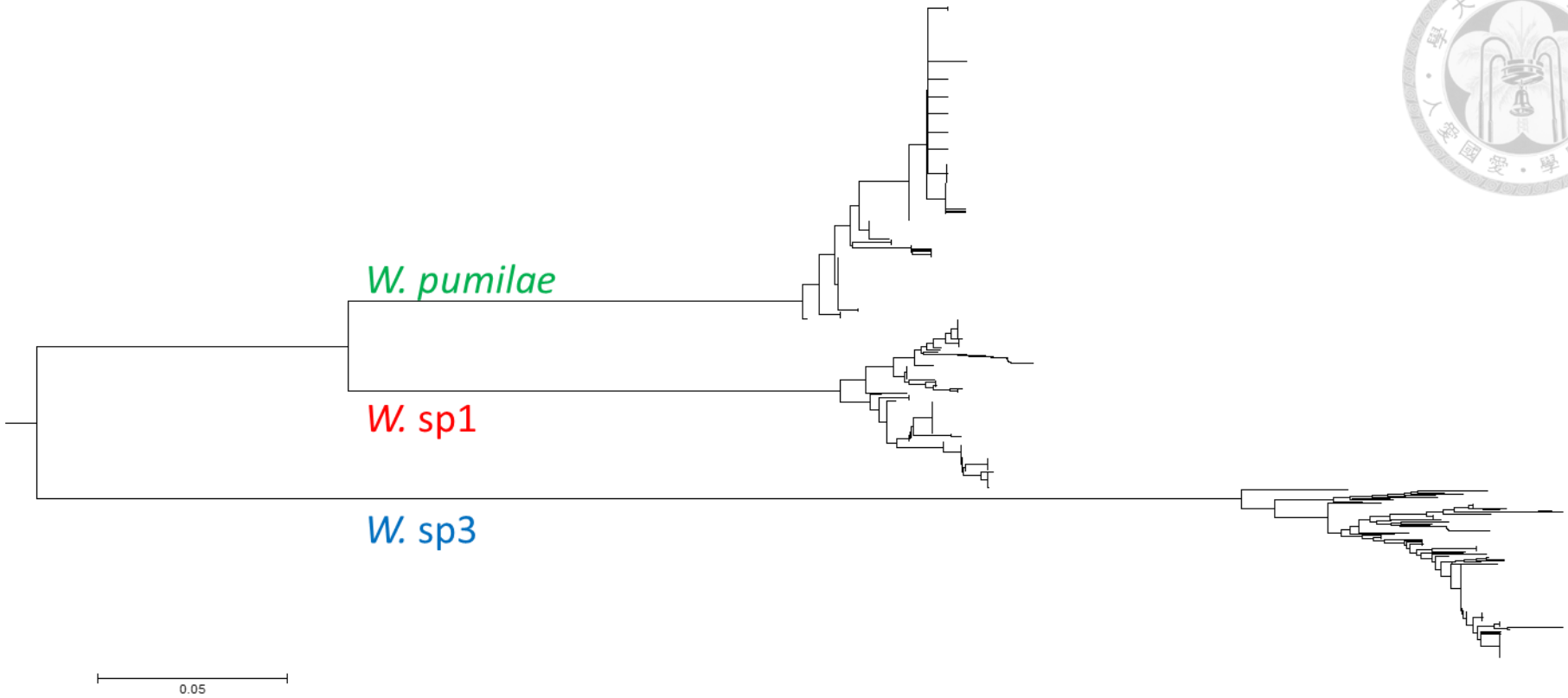

Figure 2 Neighbor-joining phylogeny of pollinators associated with *Ficus pumila*.

doi:10.6342/NTU201903609

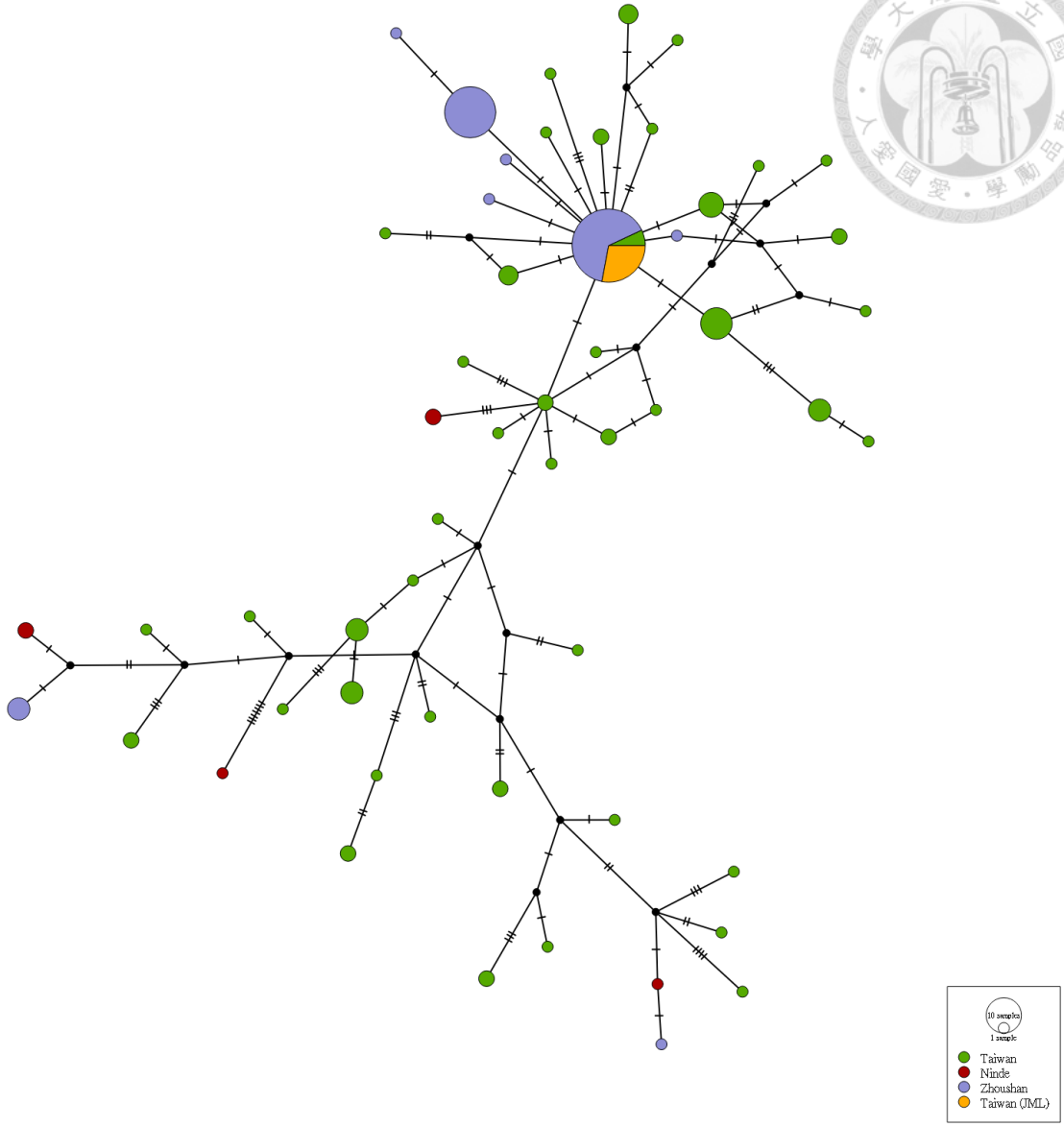

Figure 3 TCS network of *W.* sp3. Haplotype from mainland (red and purple) does not display divergence from Taiwanese populations (green and orange). Most haplotypes from Zhoushan (purple) are identical or cloesly-related to a cultivated strain in Taiwan (JML, orange).

doi:10.6342/NTU201903609

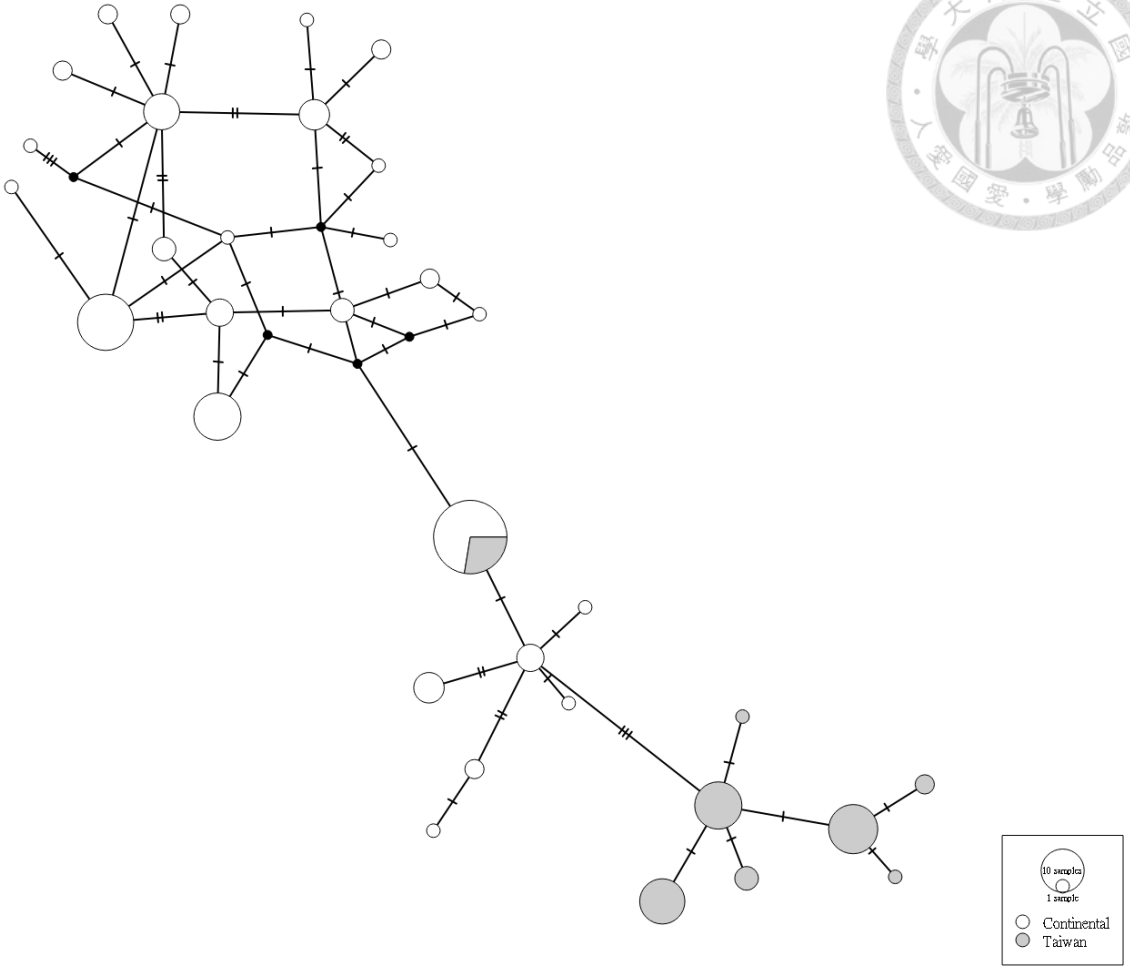

Figure 4 TCS network of *W. pumilae*. Haplotypes from Taiwan (grey) almost form one single clade, and have less genetic diversity compared with mainland populations (white). The divergence and differences in diversity between Taiwan and mainland population suggest that *W. pumilae* was originated in mainland.

doi:10.6342/NTU201903609

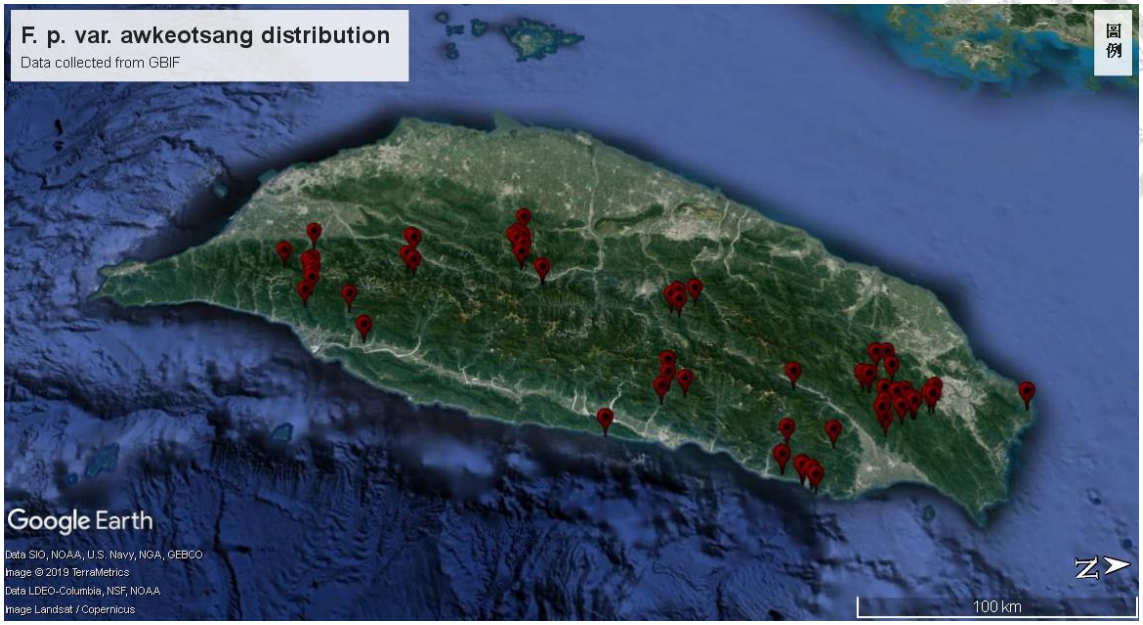

Figure 5 Occurrence sites of *Ficus pumila var. awkeotsang* from GBIF. Site that occurred outside of Taiwan or inside cities were manually excluded.

doi:10.6342/NTU201903609

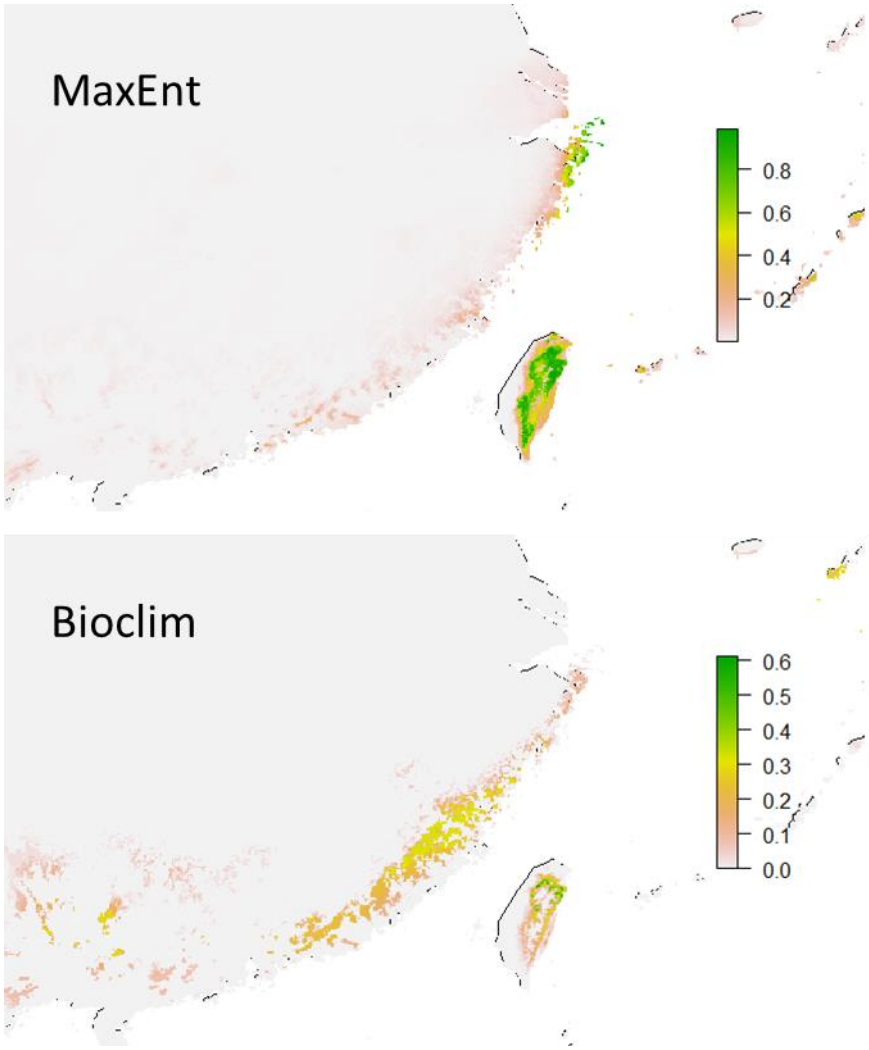

Figure 6 Results of species distribution modeling of *Wiebesia* sp3. Both model predicts that south eastern shore of China is a potential suitable niche for *W.* sp3.

doi:10.6342/NTU201903609

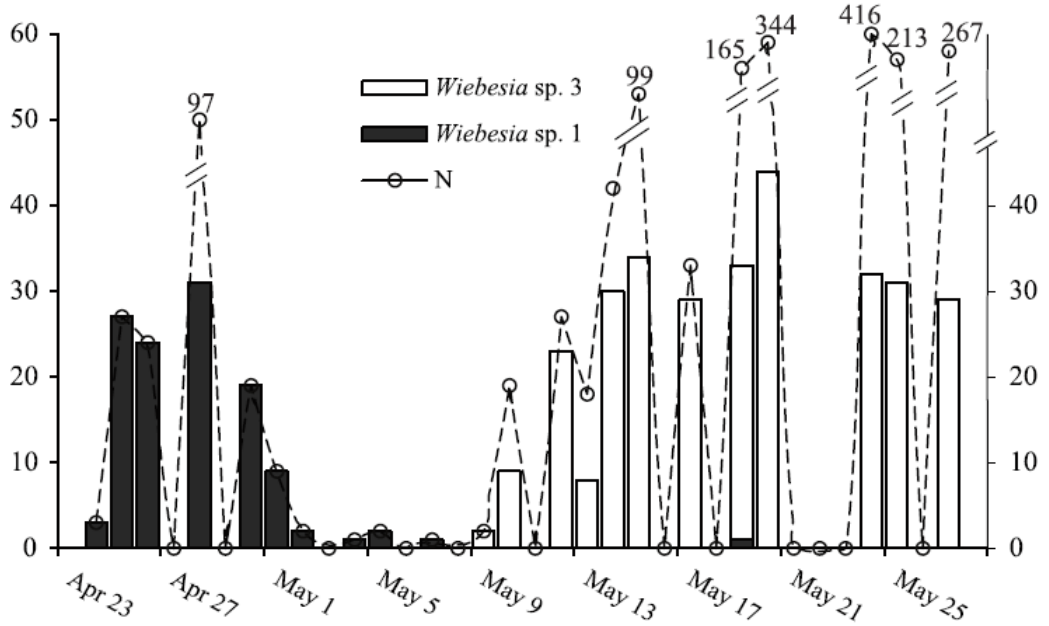

Figure 7 The time and number of female fig wasps capture in Taohua, Zhoushan Islands. Figure from Liu et al., 2014. Dashed line represented number of trapped individuals while solid line represented sampled individual for species identification using COI.

doi:10.6342/NTU201903609

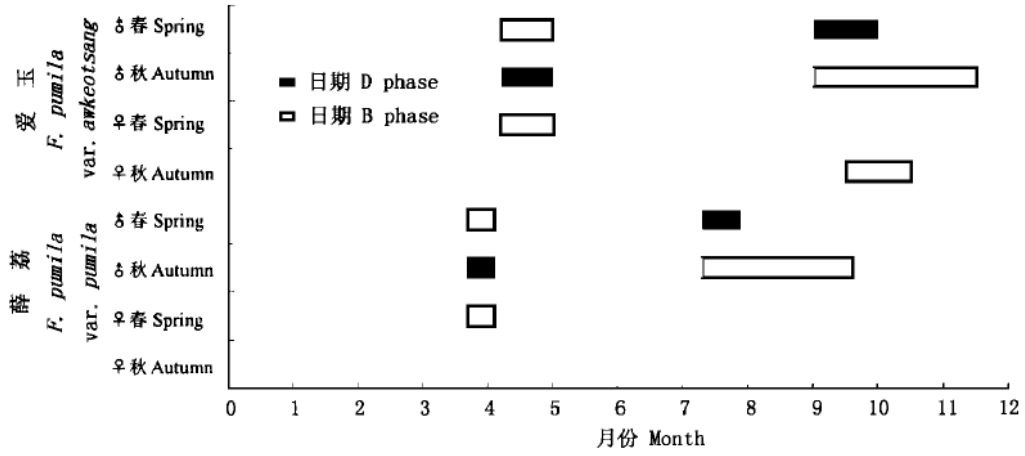

Figure 8 Phenology of *F. p. var. pumila* and *F. p. var. awkeotsang* in Fujian. Figure from Yong Chen, Li, & Ma, 2002. Noted that the D phase and B phase within variety in spring time perfectly matches one another. A delay in spring time D phase in *F. p. var. pumila* can cause mismatch in male and female flowering time thus affecting pollination.

doi:10.6342/NTU201903609

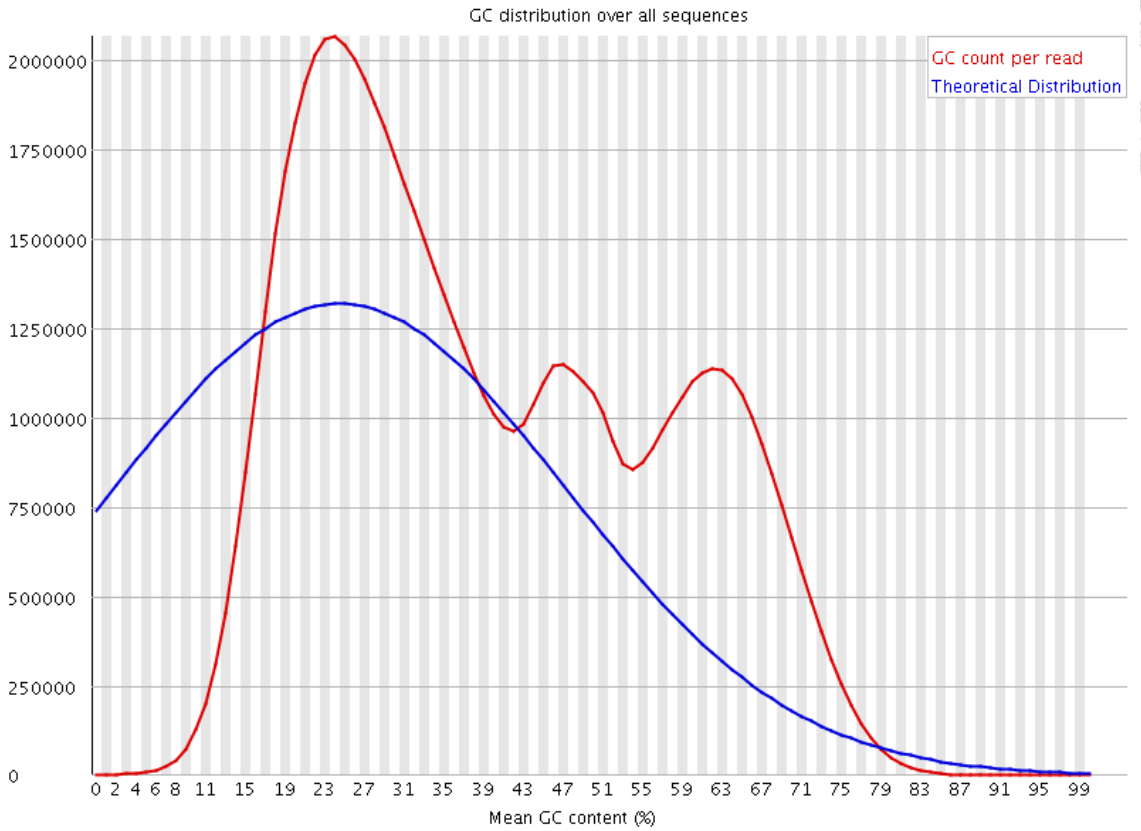

Figure 9 Per sequence GC content of jelly-fig wasp pair-end reads. The highest peak at 25% GC is the peak from fig wasp. X axis is mean GC content of read. Y axis is the total number of reads given the GC content.

doi:10.6342/NTU201903609

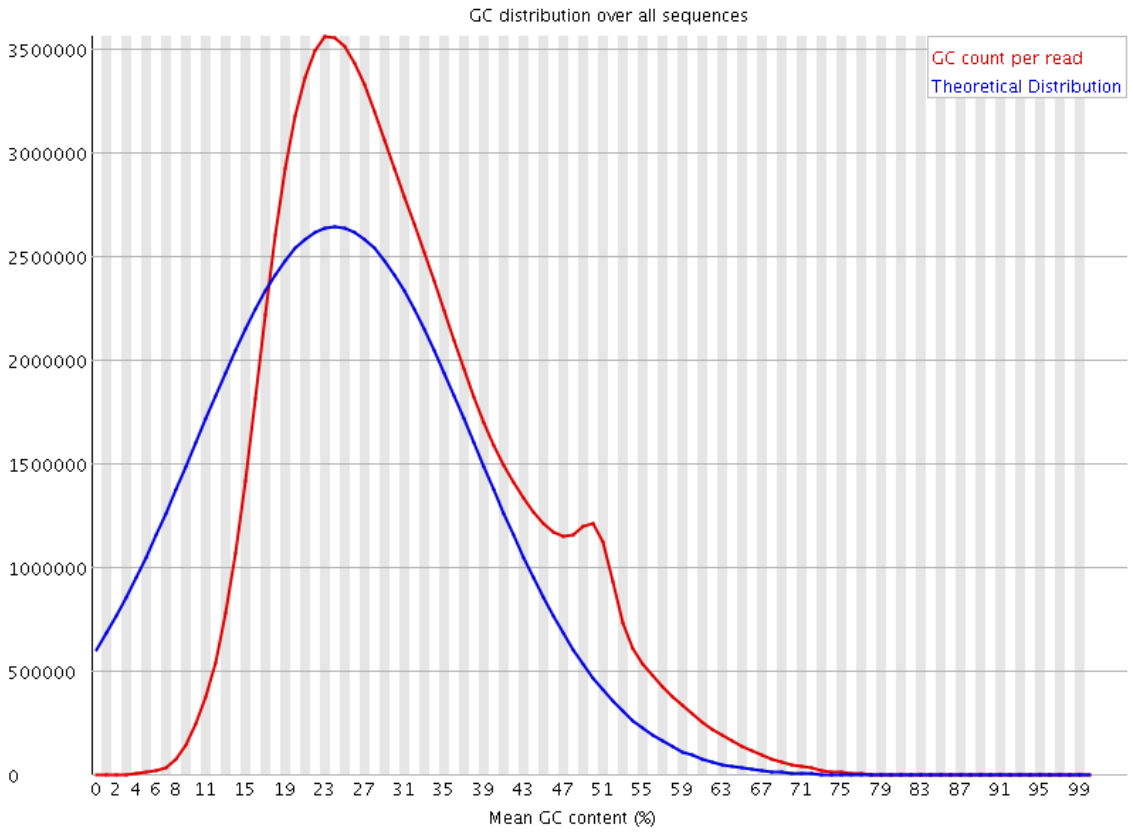

Figure 10 Per sequence GC content of creeping-fig wasp pair-end reads. X axis is mean GC content of read. Y axis is the total number of reads given the GC content.

doi:10.6342/NTU201903609

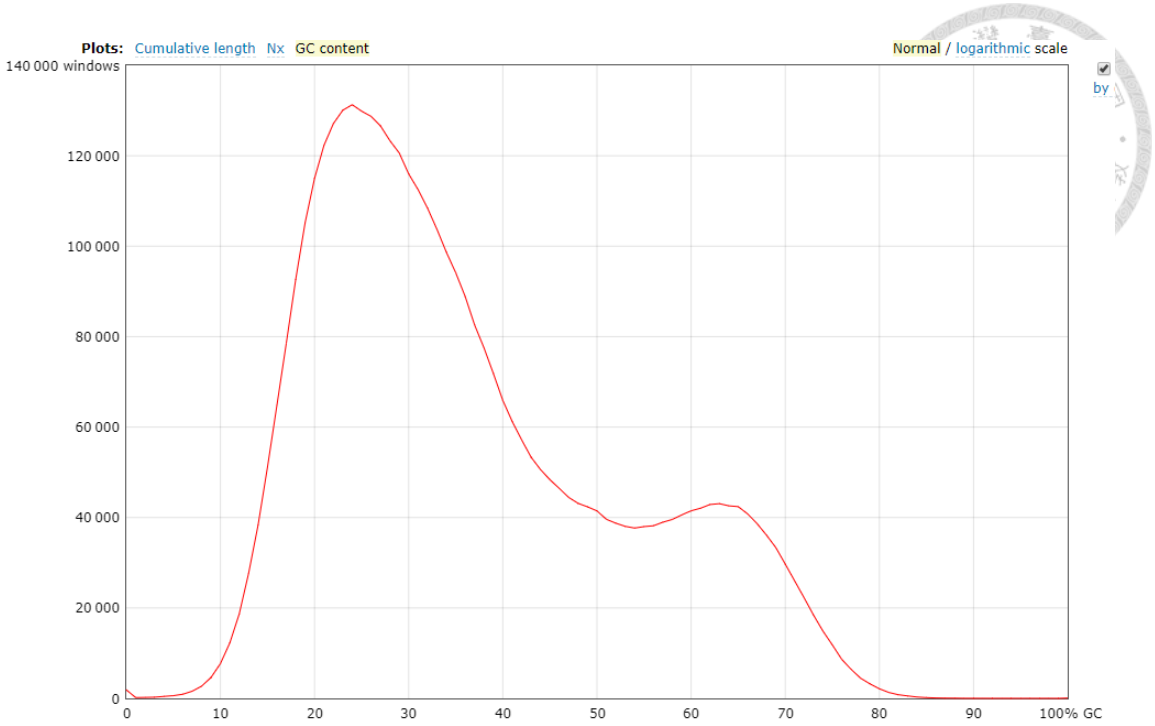

Figure 11 Per window GC content of initial jelly-fig wasp assembly. Contigs were broken into 100bp window. X axis is the average window GC content. Y axis is the number of windows at given GC content.

doi:10.6342/NTU201903609

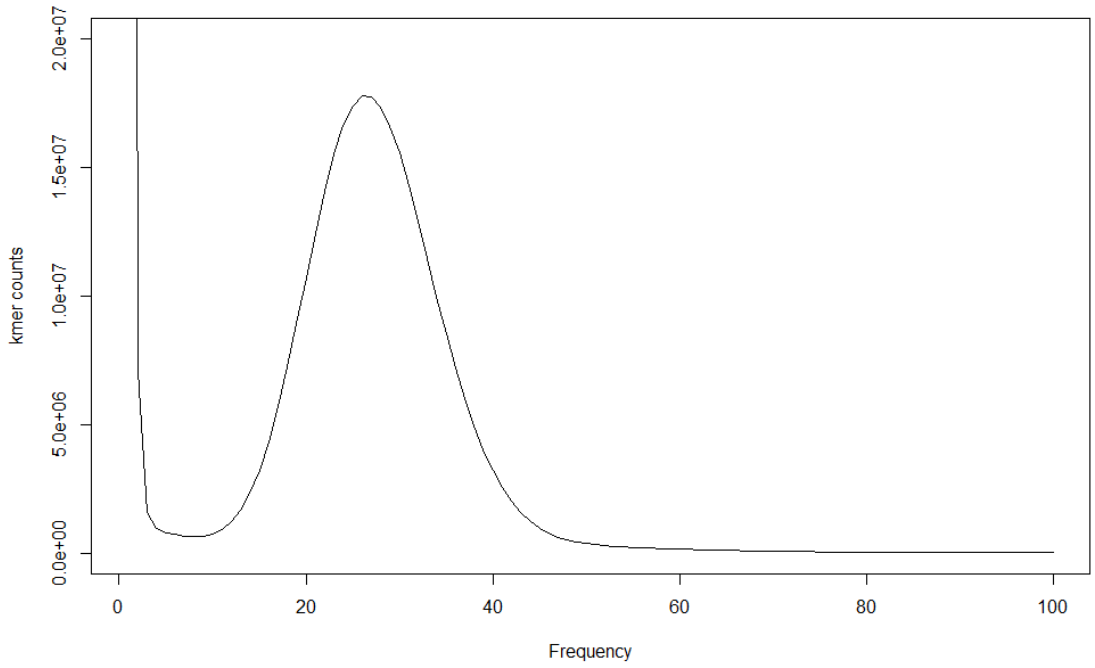

Figure 12 Kmer analysis using 21mer counts of decontaminated jelly-fig wasp pair end reads. X axis is the number of times one given 21mer is observed in the data, Y axis is the total number of 21mers in a given frequency.

doi:10.6342/NTU201903609

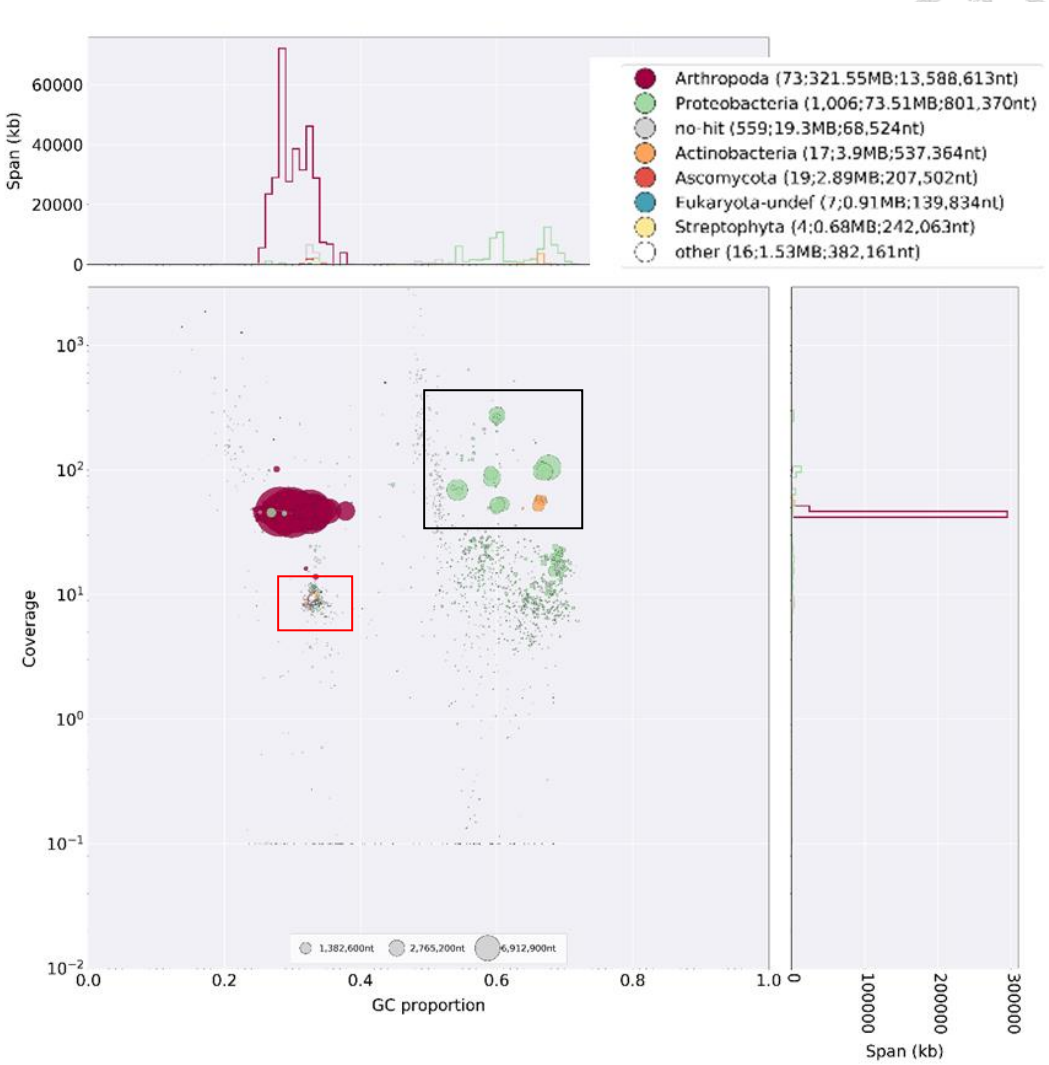

Figure 13 Blobplot of jelly-fig wasp initial assembly, coverage was based on PE reads. The circle size represent length of the contig. Color of the circle represents the phylum that contig belong. Most of the prokaryotic contigs have higher GC content than eukaryotic contigs. Black square contains six well assembled prokaryotic species/OTU; red square contains contigs of eukaryotic symbionts.

doi:10.6342/NTU201903609

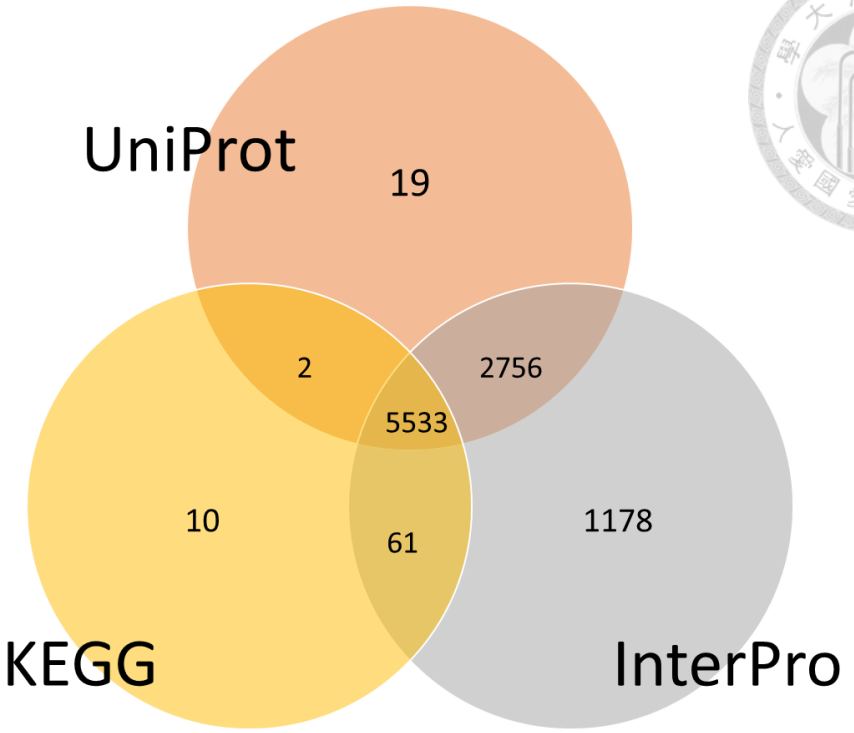

Figure 14 Functional annotation of jelly-fig wasp using three databases. Number represent protein-coding genes annotated by the database.

doi:10.6342/NTU201903609

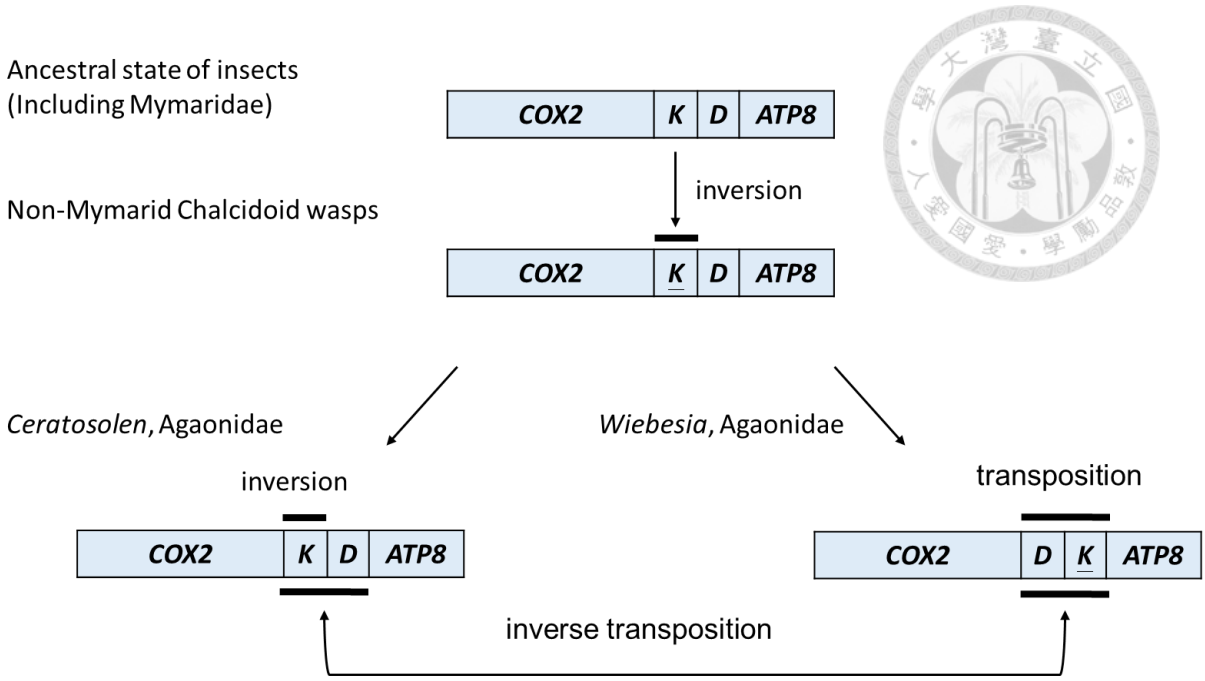

Figure 15 Evolution of mitochondrial tRNA K and tRNA D synteny in Chalcidoidea. Three possible evolutionary routes can lead to this current synteny. Chalcidoidea -> *Ceratosolen* -> *Wiebesia*, Chalcidoidea -> *Wiebesia* -> *Ceratosolen*, and Chalcidoidea -> *Ceratosolen*; Chalcidoidea -> *Wiebesia*. All three routes take equal number of changes. Genes that are underlined means that they are on the reverse strand. D, K, COX2 and ATP8 denote tRNA D, tRNA K, cytochrome c oxidase subunit 2, and ATP synthase 8.

doi:10.6342/NTU201903609

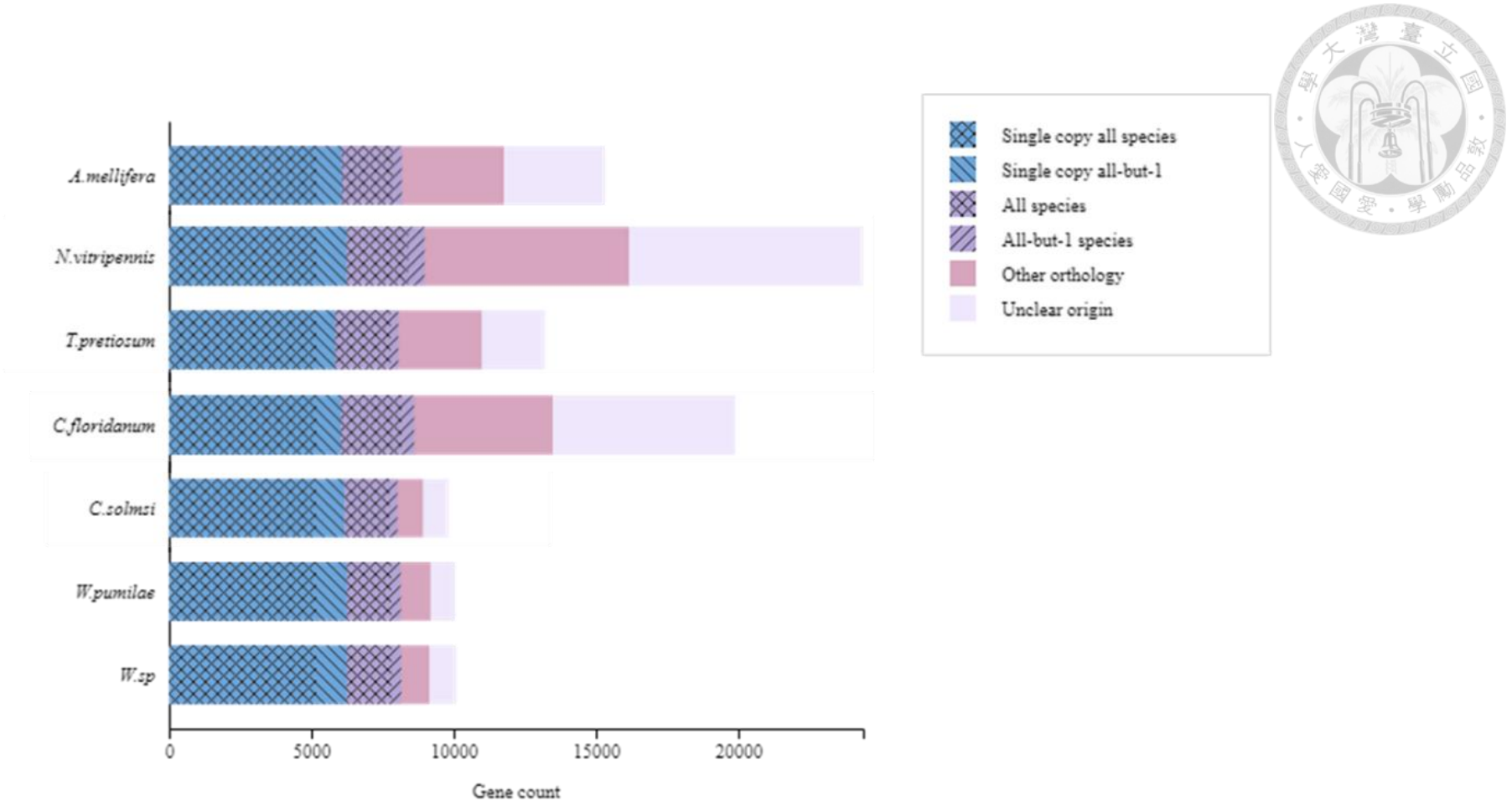

Figure 16 Gene expansion and contraction of 7 Hymenoptera species.

doi:10.6342/NTU201903609

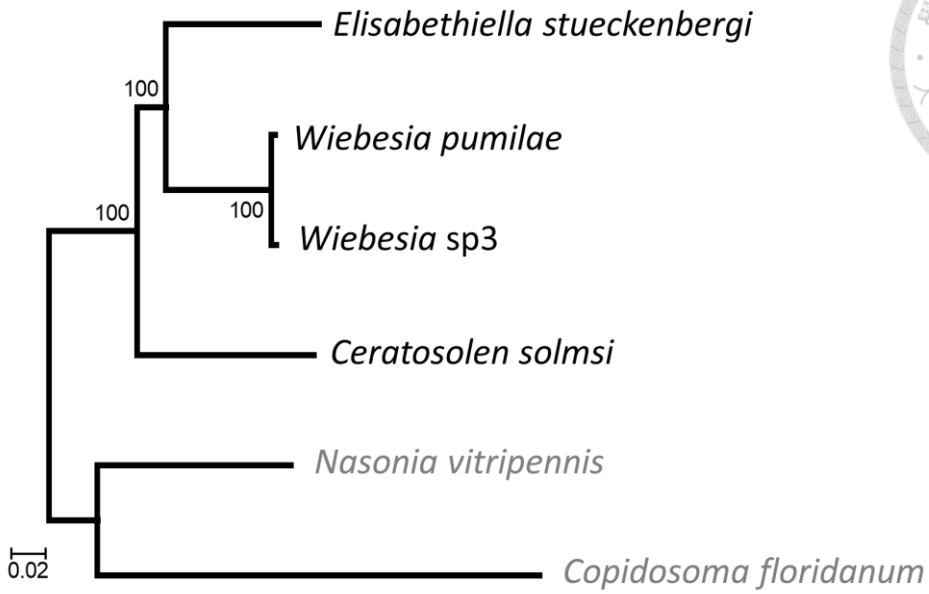

Figure 17 Maximum likelihood phylogenetic tree of fig wasps. The tree was inferred by concatenating all the shared single copy busco genes' amino acid sequence.

doi:10.6342/NTU201903609

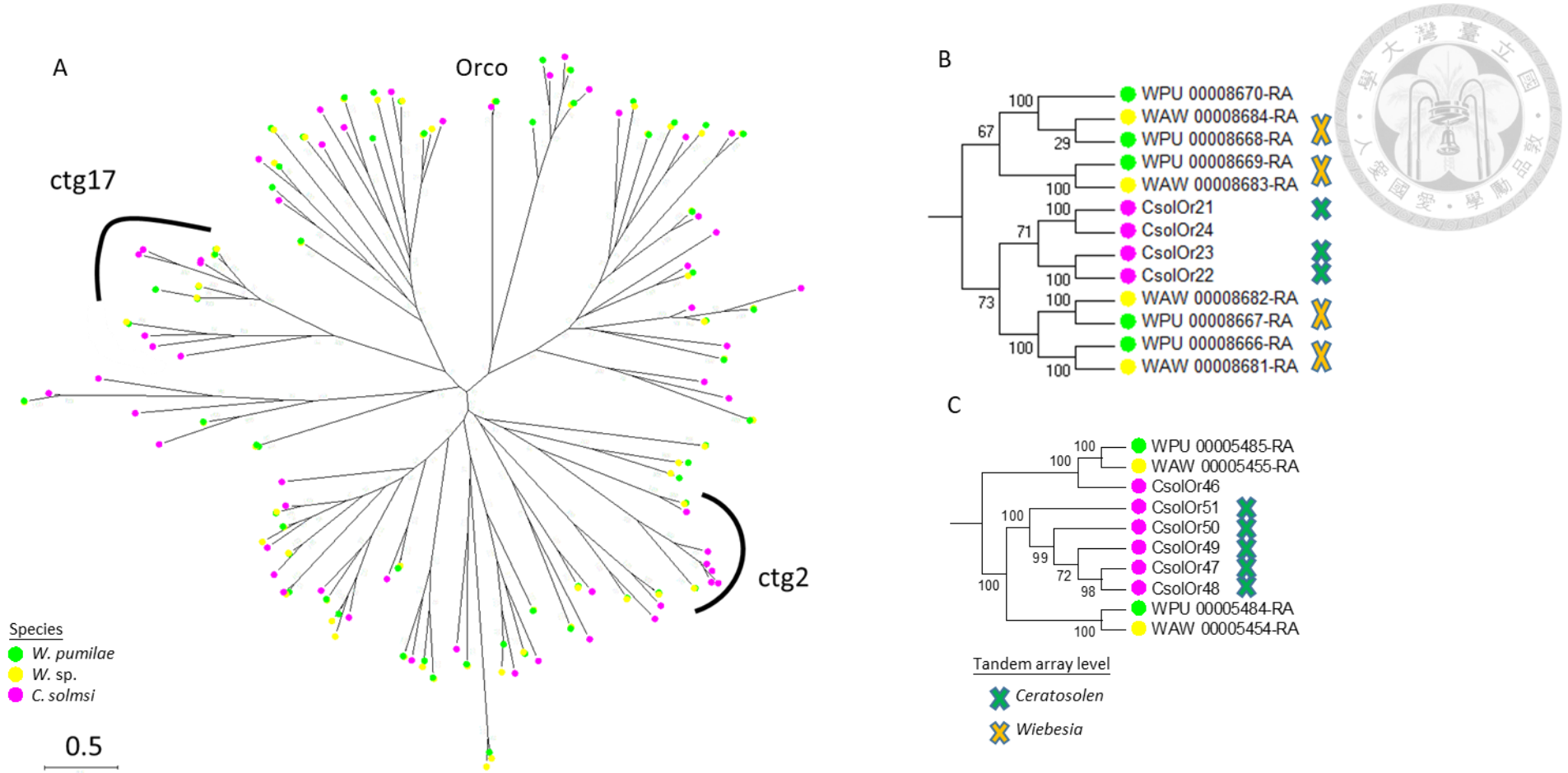

Figure 18 Maximum likelihood tree of OR genes from three agaonid wasps (A). ctg number represents the contig where jelly-fig wasp genes are located. The subtree of ctg17 (B) and ctg2 (C) shown genus-specific gene expansions mainly caused by tandem duplications.

doi:10.6342/NTU201903609

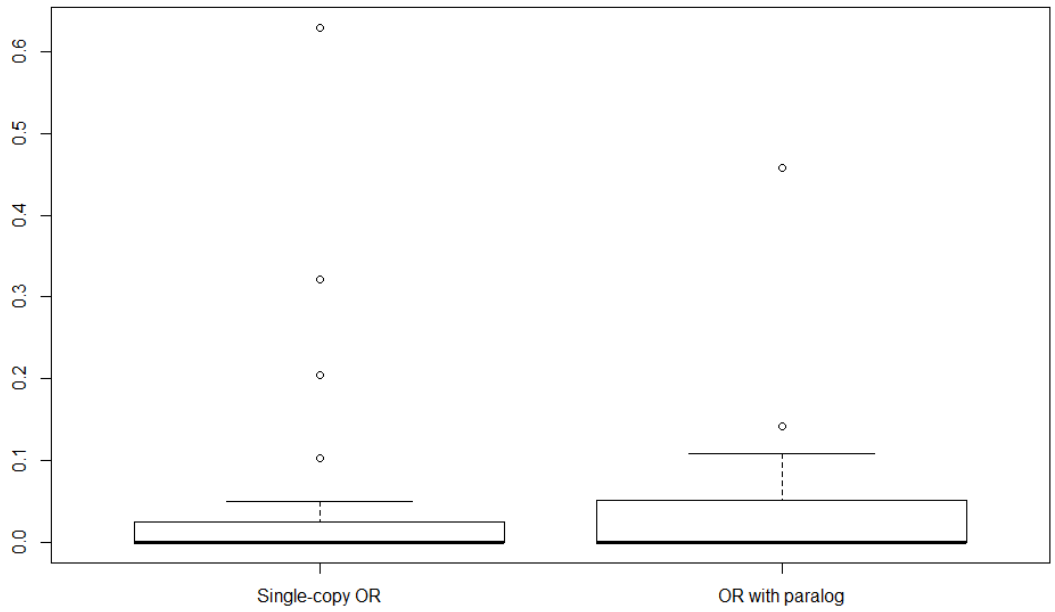

Figure 19 Boxplot of local transposable element densities around olfactory receptors that are single copy or recently duplicated. There is no significant differences between the two groups.

doi:10.6342/NTU201903609

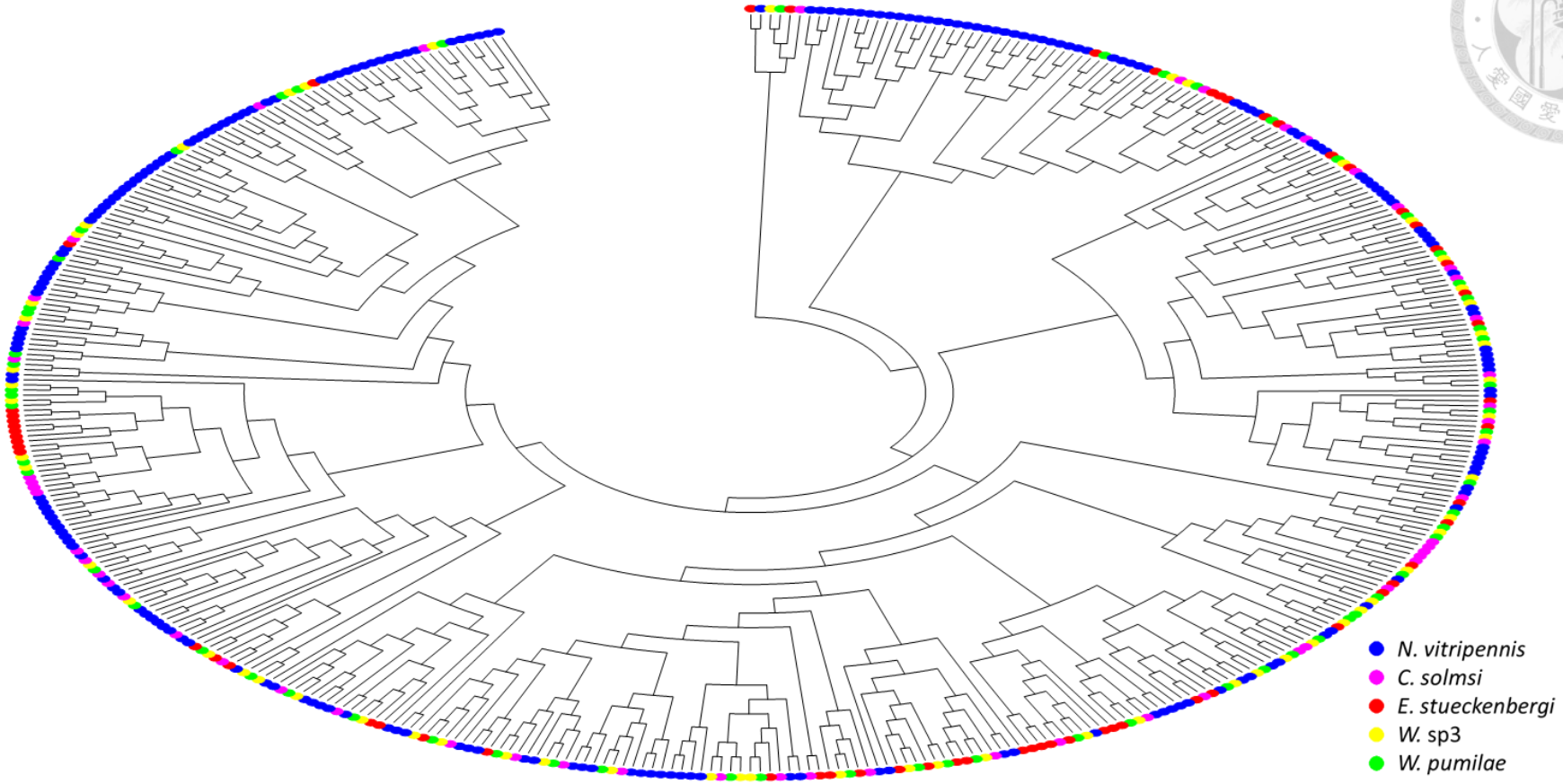

Figure 20 Maximum likelihood tree of OR.

doi:10.6342/NTU201903609

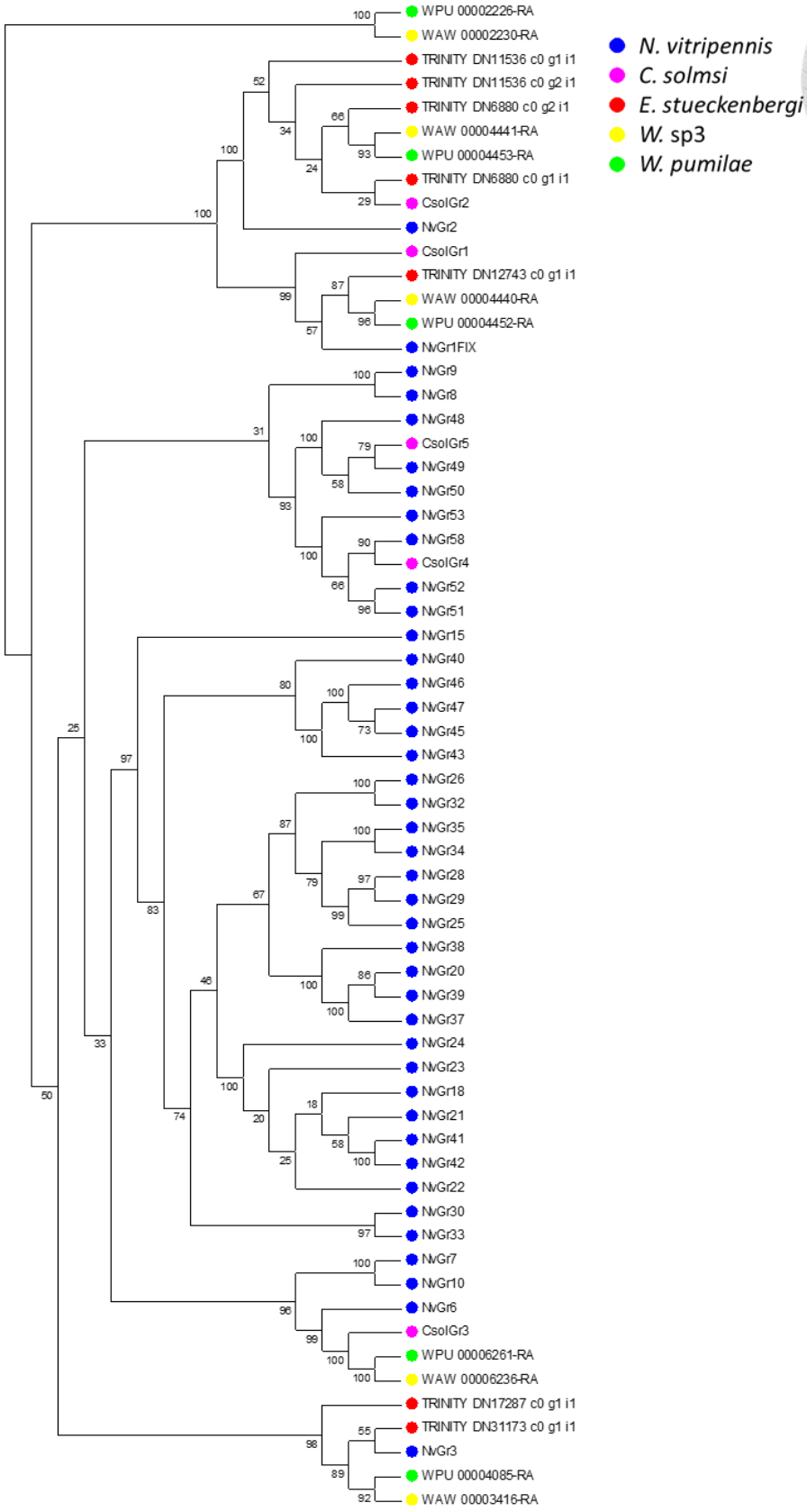

Figure 21 Maximum likelihood tree of GR

doi:10.6342/NTU201903609

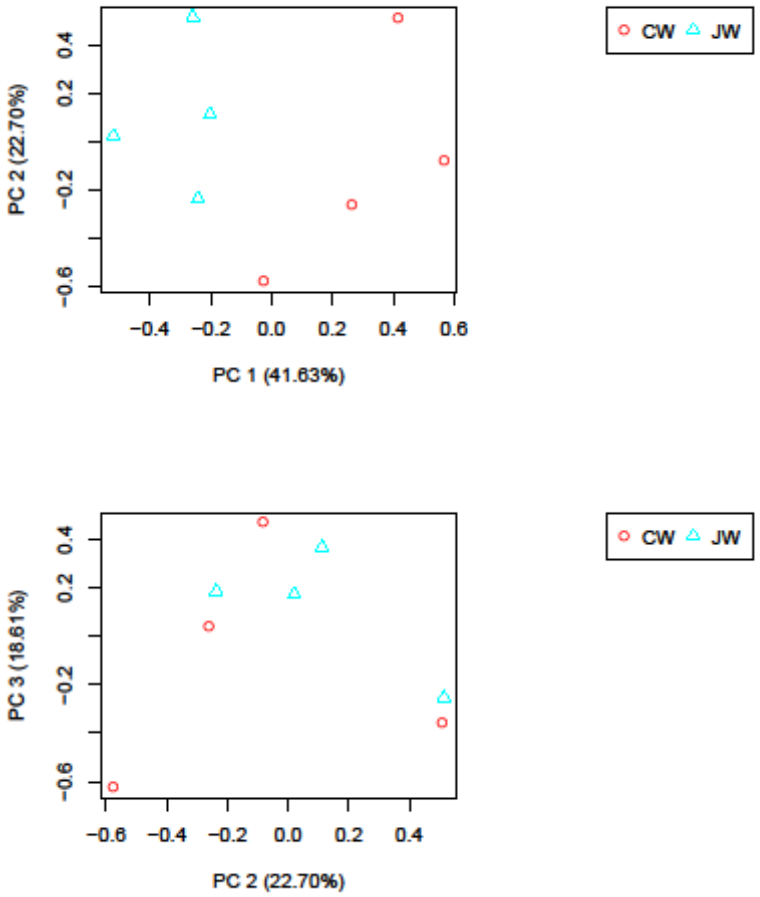

Figure 22 Principal component analysis of TMN normalized expression levels of each samples for cross-species differential expression analysis. CW: creeping-fig wasp; JW: jelly-fig wasp.

doi:10.6342/NTU201903609

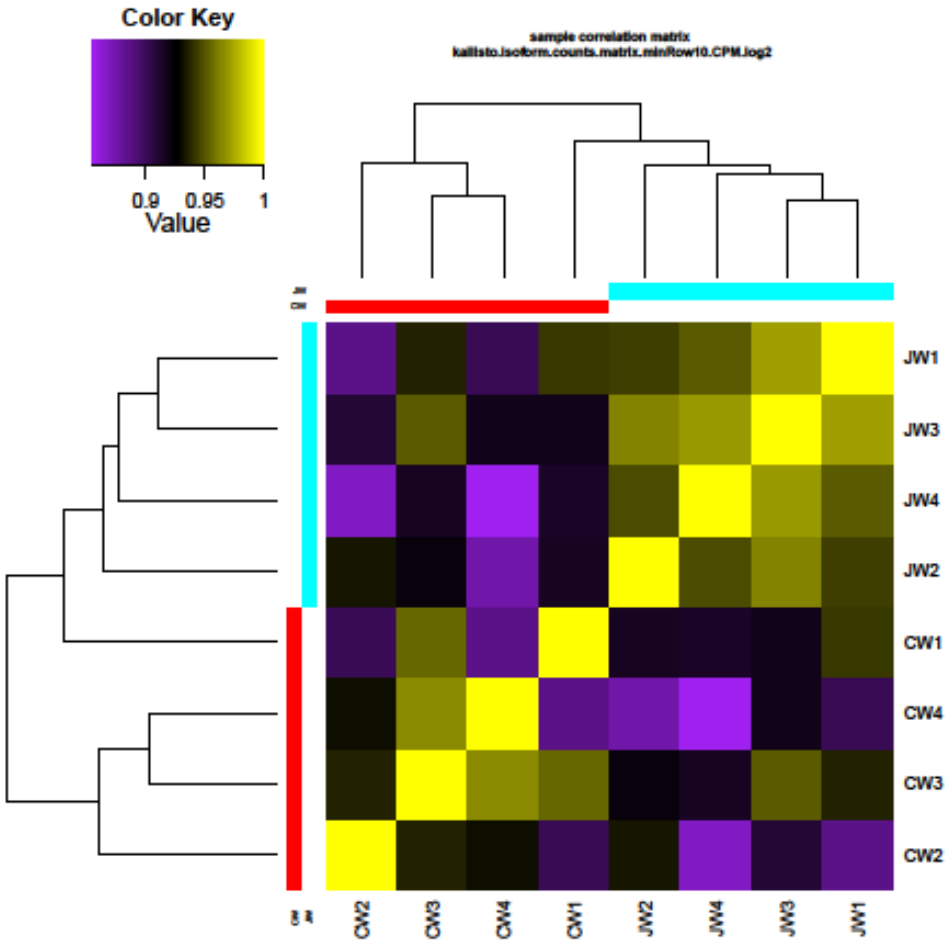

Figure 23 Correlation matrix of TMN normalized expression levels of each samples for cross-species differential expression analysis.

doi:10.6342/NTU201903609

TABLES

Table 1. Haplotype and nucleotide diversities of *W.* sp3 populations.

Group	*W.* sp3								
	Ningde	Zhoushan Islands			Taiwan				
	ND	JT	FD	DJ	JCY	JKS	JNT	JTT	JML (cultivated)
Number of wasps	6	5	10	10	19	20	23	15	12
Number of haplotypes	4	2	2	2	11	15	14	8	1
Haplotype diversity	0.867	0.6	0.533	0.2	0.912	0.963	0.949	0.895	0
Nucleotide diversity (π)	1.08%	0.07%	0.06%	0.02%	0.68%	0.71%	0.59%	0.65%	0.00%

doi:10.6342/NTU201903609

Table 2. AMOVA results of *W. pumilae* and *W.* sp3.

	W. pumilae	***W.* sp3**
	Percentage of variation	
Among groups (across Taiwan Strait)	42.99%	12.42%
Among populations within groups	21.96%	2.58%
Within populations	35.04%	82.99%

doi:10.6342/NTU201903609

Table 3 Information of all sequenced short read libraries

Library name	Description	Collection site	Sequenced bps
JW_DNA	Adult male	屏東	22.8G
CW_DNA	Adult female	嘉義	27.8G
JW_RNA_AF1	Adult female	南投	7.6G
JW_RNA_AF2	Adult female	南投	14.5G
JW_RNA_AF3	Adult female	南投	6.2G
JW_RNA_AF4	Adult female	南投	6.8G
JW_RNA_AM	Adult male	南投	7.8G
JW_RNA_PF	Pupa female	南投	9.9G
JW_RNA_PM	Pupa male	南投	8.5G
JW_RNA_L	Larva both sex	南投	3.2G
CW_RNA_AF1	Adult female	馬祖	7.4G
CW_RNA_AF2	Adult female	馬祖	15.7G
CW_RNA_AF3	Adult female	嘉義	6.7G
CW_RNA_AF4	Adult female	嘉義	5.3G
CW_RNA_AM	Adult male	馬祖	6.3G
CW_RNA_PF	Pupa female	馬祖	7.6G
CW_RNA_PM	Pupa male	馬祖	7.7G
CW_RNA_L	Larva both sex	花蓮	4.2G

JW=jelly-fig wasp; CW= creeping-fig wasp.

doi:10.6342/NTU201903609

Table 4 Prokaryotic contigs that are over 0.5Mbp in the initial jelly-fig wasp assembly. Taxa were assigned by blobtools using blastn results of NCBI nt database.

Contig	Length	GC content	Coverage	Order	Family	Species
ctg16	6801905	0.6763	105.047	Burkholderiales	Alcaligenaceae	*Achromobacter xylosoxidans*
ctg19	4795225	0.5421	69.335	Enterobacterales	Erwiniaceae	*Pantoea ananatis*
ctg20	3455062	0.593	87.05	Enterobacterales	Yersiniaceae	*Serratia marcescens*
ctg21	2975796	0.6706	97.292	Burkholderiales	Burkholderiaceae	*Burkholderia cenocepacia*
ctg23	3637676	0.6669	98.868	Burkholderiales	Burkholderiaceae	*Burkholderia cenocepacia*
ctg28	2340241	0.6011	52.69	Rhizobiales	Rhizobiaceae	*Agrobacterium tumefaciens*
ctg33	2025596	0.6087	53.357	Rhizobiales	Rhizobiaceae	*Agrobacterium tumefaciens*
ctg35	1990835	0.5915	94.671	Enterobacterales	Yersiniaceae	*Serratia marcescens*
ctg39	2879527	0.6005	275.071	Rhizobiales	Rhizobiaceae	*Agrobacterium tumefaciens*
ctg44	1361023	0.5994	49.146	Rhizobiales	Rhizobiaceae	*Rhizobium leguminosarum*
ctg50	1039403	0.2689	45.608	Burkholderiales	Burkholderiaceae	*Burkholderia cenocepacia*
ctg51	850052	0.6731	92.343	Burkholderiales	Burkholderiaceae	*Burkholderia cenocepacia*
ctg56	801370	0.6017	263.746	Rhizobiales	Rhizobiaceae	*Agrobacterium tumefaciens*
ctg57	641224	0.5452	67.83	Enterobacterales	Erwiniaceae	*Pantoea* sp. At-9b
ctg58	1544814	0.6614	51.937	Corynebacteriales	Tsukamurellaceae	*Tsukamurella paurometabola*
ctg59	537364	0.669	55.893	Corynebacteriales	Nocardiaceae*	*Tsukamurella paurometabola*
ctg63	932981	0.6032	272.206	Rhizobiales	Rhizobiaceae	*Agrobacterium tumefaciens*
ctg73	1382185	0.6844	15.478	Xanthomonadales	Xanthomonadaceae	*Xanthomonas citri*
ctg79	732604	0.694	17.718	Xanthomonadales	Xanthomonadaceae	*Xanthomonas citri*
ctg97	550369	0.6832	20.191	Xanthomonadales	Xanthomonadaceae	*Xanthomonas citri*

*Blobtools assigns each taxonomic rank independently base on blast results, contradiction between taxonomic rank can happen.

doi:10.6342/NTU201903609

Table 5 Assembly statistics of jelly-fig wasp and creeping-fig wasp

	Creeping-fig wasp	***Jelly-fig wasp***
Contigs (>= 0 bp)	188	230
Contigs (>= 5000 bp)	169	221
Contigs (>= 10000 bp)	121	133
Contigs (>= 25000 bp)	67	72
Contigs (>= 50000 bp)	49	49
Total length (>= 0 bp)	330288206	323808067
Total length (>= 5000 bp)	330287547	323766451
Total length (>= 10000 bp)	330218436	323112828
Total length (>= 25000 bp)	329877261	322141903
Total length (>= 50000 bp)	329026082	321366603
Largest contig	33617662	33052186
GC (%)	30.48	30.16
N50	19339076	18609986
N75	7249810	6255745
L50	7	7
L75	14	15
# N's per 100 kbp	1226.25	0

doi:10.6342/NTU201903609

Table 6 Statistics of other Chalcidoidea genomes and BUSCO result of gene sets.

Species	**Sequencing platform**	**Assembly level**	**Scaffolds**	**Total size**	**N50**	**Gene count**	**BUSCO (gene set mode)[a]**
Nasonia vitripennis	Sanger/ Illumina	Pseudochromosome	6169	295.781M	897131	24388[b]	C:96.6%[S:95.7%,D:0.9%],F:2.1%,M:1.3%
Copidosoma floridanum	Illumina	Scaffold	4840	553.956M	1210516	17308	C:93.3%[S:91.1%,D:2.2%],F:3.7%,M:3.0%
Ceratosolen solmsi	Illumina	Scaffold	7397	277.940M	9558897	11412	C:93.9%[S:93.3%,D:0.6%],F:3.0%,M:3.1%
***Wiebesia* sp3 (Jelly-fig wasp)**	PacBio	Scaffold	230	323.808M	18609986	10072	C:95.5%[S:95.0%,D:0.5%],F:2.0%,M:2.5%

[a]Classification of BUSCO genes include complete (C), complete single-copy (S), complete duplicated (D), fragmented (F), and missing (M).

BUSCO was run using hymenoptera odb9 dataset. [b]Based on the *N. vitripennis* official gene set (OGS) 2.0 (Rago et al., 2016)

doi:10.6342/NTU201903609

Table 7 Repetitive DNA contents in jelly-fig wasp

	Repeatmodeler	**Repbase**	**RepeatProtein-Mask**	**Combined**	**Percentage**
DNA	3852826	1459065	224966	4711271	1.45%
LINE	722260	145464	165194	897960	0.28%
LTR	1524146	949446	3171361	4175138	1.29%
Low complexity	2048667	2121798	2190975	2220568	0.69%
Other			246	246	0.00%
SINE	71052			71052	0.02%
Simple repeat	9213329	9485476	9743499	9848319	3.04%
Unknown	11554309	28810		11583026	3.58%
rRNA		29902		29902	0.01%
tRNA		2204		2204	0.00%
Total	28986589	14222165	15496241	33539686	10.36%

DNA: DNA transposons; LINE: long interspersed nuclear elements; LTR: long terminal repeats; SINE: short interspersed nuclear elements.

doi:10.6342/NTU201903609

Table 8 Summary of jelly-fig wasp gene models predicted by each gene annotation software.

Gene prediction pipeline	**Predicted genes**	**BUSCO results**
hisat2+stringtie+transdecoder	20162	C:90.1%,F:3.2%,M:6.7%
GeneWise/*N. vitripennis*	20313	C:82.1%,F:4.7%,M:13.2%
GeneWise/*A. mellifera*	15544	C:84.4%,F:6.7%,M:8.9%
GeneWise/*C. solmsi*	18510	C:91.0%,F:4.9%,M:4.1%
GeneWise/*C. floridum*	22633	C:88.1%,F:6.8%,M:5.1%
GeneWise/*T. pretiosum*	23246	C:83.4%,F:6.2%,M:10.4%
SNAP (trained by *C. solmsi*)	73718	C:43.6%,F:26.2%,M:30.2%
SNAP (trained iteratively)	42731	C:49.6%,F:24.8%,M:25.6%
GlimmerHMM (trained by *C. solmsi*)	28149	C:26.2%,F:22.0%,M:51.8%
GlimmerHMM (trained iteratively)	37372	C:42.9%,F:28.1%,M:29.0%
Augustus (trained by BUSCO)	31374	C:71.7%,F:14.5%,M:13.8%
Augustus (trained iteratively)	24273	C:75.6%,F:13.5%,M:10.9%
GeneID	15300	C:54.4%,F:27.5%,M:18.1%
FgenesH	48716	C:48.8%,F:27.5%,M:23.7%
EVM (1st round)	14738	C:93.0%,F:2.5%,M:4.5%
EVM (2nd round)	13342	C:93.0%,F:2.5%,M:4.5%
Apollo (manually curated)	10072	C:95.5%,F:2.0%,M:2.5%

doi:10.6342/NTU201903609

Table 9 Annotation of jelly-fig wasp and creeping-fig wasp mitochondrial genomes.

Gene	Strand	Jelly-fig wasp				Creeping-fig wasp			
		Start	Stop	Length	Start/Stop Codons	Start	Stop	Length	Start/Stop Codons
trnM(cat)	+	149	213	65					
trnV(tac)	+	216	282	67					
rrnS	+	283	1080	798					
trnA(tgc)	+	1105	1170	66					
trnQ(ttg)	+	1199	1258	60					
rrnL	+	1715	3106	1392		1	371	371	
trnL1(tag)	+	3111	3181	71		377	447	71	
nad1	+	3179	4117	939	ATA/TAA	445	1383	939	ATA/TAA
trnS2(tga)	-	4121	4187	67		1386	1452	67	
trnR(tcg)	-	4192	4249	58		1457	1514	58	
cob	-	4255	5382	1128	ATG/TAA	1517	2644	1128	ATG/TAA
nad6	-	5382	5885	504	ATT/TAA	2644	3147	504	ATT/TAA
trnP(tgg)	+	5959	6027	69		3211	3279	69	
trnT(tgt)	-	6036	6100	65		3288	3353	66	
nad4l	+	6140	6415	276	ATT/TAA	3390	3677	288	ATT/TAA
nad4	+	6409	7744	1336	ATG/T(AA)	3671	5006	1336	ATG/T(AA)
trnH(gtg)	+	7733	7798	66		4995	5061	67	
nad5	+	7838	9511	1674	ATT/TAA	5133	6806	1674	ATT/TAA
trnF(gaa)	+	9512	9579	68		6807	6874	68	
trnE(ttc)	-	9583	9649	67		6880	6947	68	
cox1	+	9787	11322	1536	ATA/TAA	7084	8616	1533	ATT/TAA
trnL2(taa)	+	11325	11388	64		8618	8682	65	
cox2	+	11389	12069	681	ATT/TAA	8683	9363	681	ATT/TAA
trnD(gtc)	+	12074	12138	65		9368	9433	66	
trnK(ttt)	-	12146	12213	68		9442	9509	68	
atp8	+	12280	12444	165	ATT/TAA	9573	9737	165	ATT/TAA
atp6	+	12432	13103	672	ATG/TAA	9725	10396	672	ATG/TAA
cox3	+	13110	13886	777	ATG/TAA	10403	11179	777	ATG/TAA
trnG(tcc)	+	13892	13959	68		11185	11252	68	
nad3	+	13979	14320	342	ATT/TAG	11273	11614	342	ATT/TAA
trnC(gca)	+	14321	14386	66		11615	11681	67	
trnY(gta)	-	14407	14474	68		11727	11793	67	
trnS1(tct)	-	14476	14537	62		11796	11857	62	
trnN(gtt)	-	14540	14608	69		11860	11928	69	
trnW(tca)	-	14610	14675	66		11930	11995	66	
nad2	-	14674	15672	999	ATT/TAA	11994	12992	999	ATT/TAA
trnI(gat)	+	15709	15778	70		13029	13098	70	

doi:10.6342/NTU201903609

Table 10 Gene ontologies enriched in dN/dS > 1 dataset. All GOs significant at 0.05 FDR level were shown.

GO term	GO ID	FDR
Component		
Respiratory chain complex IV	0045277	0.0031
Process		
Mitochondrial electron transport, cytochrome c to oxygen	0006123	0.0075
Proteolysis	0006508	7.60E-07
Function		
Cytochrome-c oxidase activity	0004129	0.0031
Serine-type endopeptidase activity	0004252	9.47E-21

doi:10.6342/NTU201903609

Table 11 Candidate venom protein genes that are differentially expressed between jelly-fig wasp and creeping-fig wasp.

Gene	Signal peptide	TMM	Hit to *N. vitripennis* venom	Venom hit description	Hit to Uniprot	Uniprot hit description	Species
WAW_00000159	Y	N	Nasvi2EG000909	C1q-like venom protein	---NA---	---NA---	---NA---
WAW_00000705	Y	N	Nasvi2EG010351	Glucose dehydrogenase-like	DHGL_DROME	Glucose dehydrogenase	*Drosophila melanogaster*
WAW_00001574	Y	N	Nasvi2EG012510	Chitinase 5	CHIT1_HUMAN	Chitotriosidase-1	*Homo sapiens*
WAW_00001830	Y	N	Nasvi2EG009991	Acid phosphatase	ACPH1_APIME	Venom acid phosphatase Acph-1	*Apis mellifera*
WAW_00001901	Y	N	Nasvi2EG020296	Serine protease/CLIP	GD_DROME	Serine protease gd	*Drosophila melanogaster*
WAW_00002150	Y	Y	Nasvi2EG007615	Lipase	LIP3_DROME	Lipase 3	*Drosophila melanogaster*
WAW_00002351	Y	Y	Nasvi2EG007615	Lipase	LIP3_DROME	Lipase 3	*Drosophila melanogaster*
WAW_00002427	Y	N	Nasvi2EG012510	Chitinase 5	CHIA_BOVIN	Acidic mammalian chitinase	*Bos taurus*
WAW_00002519	Y	Y	Nasvi2EG006920	Arylsulphatase b	ARSB_MOUSE	Arylsulfatase B	*Mus musculus*
WAW_00002903	Y	N	Nasvi2EG007282	Serine protease	---NA---	---NA---	---NA---
WAW_00004350	Y	N	Nasvi2EG007282	Serine protease	TRYI_DROME	Trypsin iota	*Drosophila melanogaster*
WAW_00006004	Y	N	Nasvi2EG007282	Serine protease	CTR2_ANOGA	Chymotrypsin-2	*Anopheles gambiae*
WAW_00006031	Y	N	Nasvi2EG007282	Serine protease	SP4_POLDO	Venom serine protease	*Polistes dominula*
WAW_00006435	Y	N	Nasvi2EG011442	Serine protease	CTR2_ANOGA	Chymotrypsin-2	*Anopheles gambiae*
WAW_00008311	Y	N	Nasvi2EG022626	Serine protease/CUB	CTR2_VESOR	Chymotrypsin-2	*Vespa orientalis*
WAW_00008318	Y	N	Nasvi2EG022916	Serine protease	TRY3_SALSA	Trypsin-3 (Fragment)	*Salmo salar*
WAW_00008539	Y	N	Nasvi2EG022916	Serine protease	TRY5_ANOGA	Trypsin-5	*Anopheles gambiae*
WAW_00008541	Y	N	Nasvi2EG022914	Serine protease	HYPB_HYPLI	Hypodermin-B	*Hypoderma lineatum*
WAW_00008961	Y	Y	Nasvi2EG011442	Serine protease	TRYP_ASTAS	Trypsin-1	*Astacus astacus*
WAW_00009780	Y	N	Nasvi2EG011442	Serine protease	TRYP_ASTAS	Trypsin-1	*Astacus astacus*

doi:10.6342/NTU201903609

Table 12 Size of chemosensory gene family in agaonid wasps and N. vitripennis.

Species	**ORs**	**GRs**	**IRs**	**OBPs**	**CSPs**
***W. pumilae* (creeping-fig wasp)**	60	4	10	13	8
***W.* sp. (jelly-fig wasp)**	58	4	11	12	8
C. solmsi	55	6	11	10	8
E. stueckenbergi	60	7	12	17	6
N. vitripennis	207	47	10	69	10

doi:10.6342/NTU201903609

Table 13 Patterns of selection in chemosensory genes.

	ORs	**GRs**	**IRs**	**OBPs**	**CSPs**	**NCs[a]**
Single-copy orthologs in *Wiebesia*	55	4	10	12	8	9141
dN/dS>1	9	1	0	1	0	632
average dN/dS	0.4248	0.3826	0.2704	0.1088	0.0966	0.2315
average dN	0.0063	0.0051	0.0039	0.0018	0.0007	0.0028
average dS	0.0148	0.0134	0.0146	0.0168	0.007	0.0119
Single-copy orthologs in agaonid wasps	33	4	10	9	8	8182

[a]Non-chemosensory genes.

doi:10.6342/NTU201903609

Table 14 Number of branches under selection for OR ortho-groups.

	# of trees	# of branches without selection	# of branches under positive selection
Single-copy ortho-groups	33	97	2
Ortho-groups with paralogs	8	50	8

Fisher exact test $p = 0.0056$

doi:10.6342/NTU201903609

Table 15 Gene ontologies enriched in cross-species differential expressed gene set. All GOs significant at 0.05 FDR level were shown.

GO term	GO ID	FDR
Component		
Extracellular region	0005576	8.75E-06
Process		
Sensory perception of smell	0007608	1.66E-05
G protein-coupled receptor signaling pathway, coupled to cyclic nucleotide second messenger	0007187	0.0015
Function		
Serine-type endopeptidase activity	0004252	1.66E-05
Ion channel activity	0005216	4.74E-05
Olfactory receptor activity	0004984	3.72E-04
Odorant binding	0005549	4.95E-04
Ion gated channel activity	0022839	0.0015
Structural constituent of cuticle	0042302	0.0231

doi:10.6342/NTU201903609

Table 16 Chemosensory genes that are differentially expressed in jelly-fig wasp and creeping-fig wasp.

Gene name	Up regulated	Log fold change	Single-copy in Agaonidae
OR			
WAW_00001067	CW	4.42	Y
WAW_00002230 (Orco)	CW	2.07	Y
WAW_00002856	JW	-2.48	Y
WAW_00003378	JW	-3.19	Y
WAW_00006086	CW	5.18	Y
WAW_00006388	CW	2.15	Y
WAW_00006640	CW	5.43	Y
GR			
WAW_00004441	JW	-3.52	Y
WAW_00006236	CW	3.03	Y
IR			
WAW_00002511	CW	3.54	Y
WAW_00008616	CW	2.23	Y
WAW_00009024	CW	2.82	Y
CSP			
WAW_00005774	CW	5.09	Y
WAW_00005775	CW	3.10	Y
OBP			
WAW_00001780	CW	2.67	Y
WAW_00008197	JW	-4.90	Y
WAW_00009356	CW	2.56	Y

CW: creeping-fig wasp; JW: jelly-fig wasp

doi:10.6342/NTU201903609