Population genomics of the yellow crazy ant and its intracellular microorganisms

2021

LEE CHIH CHI

Contents

List of Figures

List of Tables

Chapter 1: General introduction

1.1 Biological invasion, contemporary evolution, and invasive ants

Nonindigenous species usually disperse with assistance by human-mediated transportations to new environments where they have never experienced. The process of biological invasion therefore represents an excellent opportunity for exploring organismal evolution over a contemporary timescale. Understanding how a nonindigenous species adapts novel environments and successfully sustains therefore provide insights into the ongoing progress of the evolution. Ants are emerging as successful invaders due to their social traits that contribute to ecological dominance across native communities in the introduced range [1]. Such social traits include polygynous (nest headed by multiple reproductive queens), polydomous (a colony inhabiting in multiple spatially separated yet socially connected nests), dependent foundation (e.g. budding), and formation of supercolony (i.e. a multiple physically separated colonies but reduce or lack territoriality among colonies) [1]. Ant species with supercolony structure have advantages of increased nest densities and colony size [2], and meanwhile reduce intraspecific competition [3], which make ants readily establish their initial colonies in the introduced areas.

For the long-term survival after the initial establishment, regional environmental factors may further influence the fate of the ant colony. For example, the yellow crazy ant (*Anoplolepis gracilipes*) populations decline without human intervention after invasion in Australia [4]. Another example is that the temperature serves as an abiotic limit for post-establishment invasion success in Argentine ants (*Linepithema humile*) [5]. While the influence of regional environmental factors plays a critical role during the biological invasion process, most of current studies of nonindigenous ants have been focused on recent invasion events without concerning whether local adaptations occur after the population establishment [6].

1.2 The yellow crazy ant

To understand evolutionary consequences after the nonindigenous ant's establishment, in this thesis I used the yellow crazy ant (*Anoplolepis gracilipes*) as the model system due to it's global distribution and long invasion history [7]. The yellow crazy ant is distributed primarily in tropical Asia and tropical Indo-Pacific islands [7]. Although the native range of *A. gracilipes* remains undetermined, Southeast Asia has been a general consensus [7,8]. Yellow crazy ants extend their territories into humid subtropical climate regions and form sustainable populations [7]. This ant is similar to most invasive ant species and possess polygynous, polydomous, and often forms

supercolonies. Like many other invasive ant species, independent foundation (e.g., nuptial flight) is rare in *A. gracilipes*, and colony reproduction is primarily through dependent foundation (e.g., budding) [8-12]. Dispersal by female alates seem rather limited, and intranidal mating is believed to be this ant's main reproductive strategy [9]. Such a reproductive strategy restricts gene flow among different nests of *A. gracilipes* and even results in within-supercolony divergence [11]. Behavioral assays demonstrate that the yellow crazy ant from different supercolonies tend to be highly aggressive toward each other [9,10,12], suggesting that this ant is at an initial process of speciation [12]. Altogether, the yellow crazy ant appears to be a suitable model system in studying how evolution shapes invasive ant after initial establishment.

1.3 Wolbachia: a parasite or a mutualistic endosymbiont

The interplay between intercellular microorganisms and its host may shape the host population structure and hence influences the consequence of biological invasion. *Wolbachia* is an endosymbiont commonly found in arthropods and filarial nematodes [13]. This bacterium could manipulate the reproduction of its arthropod hosts to benefit itself as a reproductive parasite. The reproductive manipulations include cytoplasmic incompatibility, male killing, feminization, and induction of parthenogenesis [13,14]. Evidence has been increasingly accumulated in recently to suggest that *Wolbachia* confers benefits to their hosts and acts as a mutualistic partner that is involved in either nutritional provisioning or host pathogen defense [14]. The role of *Wolbachia* in ants, however, is relatively under-described yet likely varies across species. Most studies have found limited evidence supporting the existence of *Wolbachia*-induced cytoplasmic incompatibility, parthenogenesis, and feminization as viable reproductive manipulation in ant [15]. Previous studies have shown that *A. gracilipes* collected from several Indo-Pacific islands and Australia including Christmas Island share identical *Wolbachia wsp* genotype, and that the bacterium seems to persist at a high prevalence across these populations [16,17]. While prevalence of *Wolbachia* may persist at high prevalence in the yellow crazy ant, the role of *Wolbachia* in this ant remains under-described. As effects of *Wolbachia* on ants are poorly understood, the investigation of whether *Wolbachia* is involved in colonization process of *A. gracilipes* is expected to open a gate for understanding the process of microorganism homogenization at the global scale.

1.4 Viruses: potential threats accompany biological invasions

The disease-mediated invasion may promote the success of invasive speceis via parasite spillover [18]. For example, the transfer of parapoxvirus from North American grey squirrel (*Sciurus carolinensis*) to the native squirrel red squirrel (*Sciurus vulgaris*) assists the grey squirrel in replacing the red squirrel in UK [19]. Viruses in invasive

ants such as the red imported fire ant (*Solenopsis invicta*) and the Argentine ant (*Linepithema humile*) likely have spread with their hosts as a hitchhiker [20,21]. Despite low prevalence [22], pathogens that invasive ants carry during colonization may be harmful to the local ant community [22]. Some of common features of invasive ants, such as polygyne and formation of supercolony, allow these ants to act as a pathogen reservoir in the introduced range as loss of nest boundary would facilitate horizonal transmission of a broad spectrum of pathogens over a large geographical scale, as the pathogen buffering mechanism predicts [23]. The invasive ants potentially carry various of co-introduced pathogens to the local community [23], if any. However, little is known about pathogens (e.g., viruses) in *A. gracilipes*. The investigation of virome in the yellow crazy ant will enhance our knowledge about the relationship of microorganisms and its host during the biological invasions.

1.5 Objectives and outline of this thesis

In this thesis, I attempt to solve enigmas during the globalization of invasive ant. I asked questions such as 1) does an invasive ant has local adaptation(s) in response to environmental factors, 2) what roles do microorganisms play during *A. gracilipes* colonization, and 3) what does pathogen community look like in the yellow crazy ant? I used high throughput sequencing including whole genome (DNA) and transcriptome (RNA) sequencing to generate sequence data that allow me to conduct comprehensive analysis on the host, yellow crazy ant and its intracellular microorganisms. Samples were collected from 13 geographic regions across Southeast and East Asia, Sri Lanka, several Pacific islands, and Australia. Chapter 2 is devoted to characterizing the mitochondrial genome evolution in *A. gracilipes*. I showed that mitochondrial genome of the ant possesses tribe-level gene rearrangements. Chapter 3 analyzed the population genomics and indicated the occurrence of multiple colonization waves in *A. gracilipes*. The yellow crazy ant in Southeast Asia likely have originated from at least three ancestral supercolonies. Subsequent genome scan detected 17 genomic outlier regions that are potentially associated with adaptive evolution of this ant. In Chapter 4, the prevalence and genetic diversity of *Wolbachia* were assessed, and near fixation of single *Wolbachia* strain in *A. gracilipes* across a broad geographical range was found. Ant-*Wolbachia* interactions and coevolution is evident as informed by the finding of a strong concordance between phylogenies of the endosymbiont and host mitochondrial genome. The final part of this thesis (Chapter 5 &6) is to explore the RNA virome in *A. gracilipes*, which leads to a discovery of multiple novel Picornaviruses and XXXX. Further analyses support that most of these pathogens may have been co-introduced with the host ant. This thesis provides baseline information that practically fills the knowledge gaps regarding how colonization success has been shaped in this ant via evolutionary and microorganism perspective.

1.6 References

1. Bertelsmeier, C.; Ollier, S.; Liebhold, A.; Keller, L. Recent human history governs global ant invasion dynamics. *Nat. Ecol. Evol.* **2017**, *1*, 0184.
2. Abbott, K.L. Supercolonies of the invasive yellow crazy ant, Anoplolepis gracilipes, on an oceanic island: forager activity patterns, density and biomass. *Insectes Sociaux* **2005**, *52*, 266-273.
3. Thomas, M.L.; Tsutsui, N.D.; Holway, D.A. Intraspecific competition influences the symmetry and intensity of aggression in the Argentine ant. *Behav. Ecol.* **2005**, *16*, 472-481.
4. Cooling, M.; Hoffmann, B.D. Here today, gone tomorrow: declines and local extinctions of invasive ant populations in the absence of intervention. *Biol. Invasions* **2015**, *17*, 3351-3357.
5. Holway, D.A.; Suarez, A.V.; Case, T.J. Role of abiotic factors in governing susceptibility to invasion: a test with Argentine ants. *Ecology* **2002**, *83*, 1610-1619.
6. Bertelsmeier, C.; Keller, L. Bridgehead effects and role of adaptive evolution in invasive populations. *Trends Ecol. Evol.* **2018**, *33*, 527-534.
7. Wetterer, J.K. Worldwide distribution and potential spread of the long-legged ant, *Anoplolepis gracilipes* (Hymenoptera: Formicidae). *Sociobiology* **2005**, *45*, 77-97.
8. Ito, F.; Asfiya, W.; Kojima, J.I. Discovery of independent-founding solitary queens in the yellow crazy ant *Anoplolepis gracilipes* in East Java, Indonesia (Hymenoptera: Formicidae). *Entomol. Sci.* **2016**, *19*, 312-314.
9. Drescher, J.; Blüthgen, N.; Schmitt, T.; Bühler, J.; Feldhaar, H. Societies drifting apart? Behavioural, genetic and chemical differentiation between supercolonies in the yellow crazy ant *Anoplolepis gracilipes*. *PLoS ONE* **2010**, *5*, e13581.
10. Thomas, M.L.; Becker, K.; Abbott, K.; Feldhaar, H. Supercolony mosaics: two different invasions by the yellow crazy ant, Anoplolepis gracilipes, on Christmas Island, Indian Ocean. *Biol. Invasions* **2010**, *12*, 677-687.
11. Gruber, M.A.; Hoffmann, B.D.; Ritchie, P.A.; Lester, P.J. Recent behavioural and population genetic divergence of an invasive ant in a novel environment. *Divers. Distrib.* **2012**, *18*, 323-333.
12. Drescher, J.; Blüthgen, N.; Feldhaar, H. Population structure and intraspecific aggression in the invasive ant species Anoplolepis gracilipes in Malaysian Borneo. *Mol. Ecol.* **2007**, *16*, 1453-1465.
13. Werren, J.H.; Baldo, L.; Clark, M.E. *Wolbachia*: Master manipulators of invertebrate biology. *Nat. Rev. Microbiol.* **2008**, *6*, 741–751.
14. Zug, R.; Hammerstein, P. Bad guys turned nice? A critical assessment of *Wolbachia* mutualisms in arthropod hosts. *Biol. Rev.* **2015**, *90*, 89–111.
15. Pontieri, L.; Schmidt, A.M.; Singh, R.; Pedersen, J.S.; Linksvayer, T.A. Artificial selection on ant female caste ratio uncovers a link between female-biased sex ratios and infection by *Wolbachia* endosymbionts. *J. Evol. Biol.* **2017**, *30*, 225–234.
16. Tseng, S.P.; Hsu, P.W.; Lee, C.C.; Wetterer, J.K.; Hugel, S.; Wu, L.H.; Lee, C.Y.; Yoshimura, T.; Yang, C.C.S. Evidence for common horizontal transmission of

Wolbachia among ants and ant crickets: Kleptoparasitism added to the list. *Microorganisms* **2020**, *8*, 805.
17. Sebastien, A.; Gruber, M.A.M.; Lester, P.J. Prevalence and genetic diversity of three bacterial endosymbionts (*Wolbachia*, *Arsenophonus*, and *Rhizobiales*) associated with the invasive yellow crazy ant (*Anoplolepis gracilipes*). *Insectes Soc.* **2012**, *59*, 33–40.
18. Young, H.S.; Parker, I.M.; Gilbert, G.S.; Guerra, A.S.; Nunn, C.L. Introduced species, disease ecology, and biodiversity–disease relationships. *Trends Ecol. Evol.* **2017**, *32*, 41-54.
19. Collins, L. M.; Warnock, N. D.; Tosh, D. G.; McInnes, C.; Everest, D.; Montgomery, W. I.; ...; Reid, N. Squirrelpox virus: assessing prevalence, transmission and environmental degradation. *PLoS One* **2014**, *9*, e89521.
20. Valles, S.M.; Varone, L.; Ramírez, L.; Briano, J. Multiplex detection of Solenopsis invicta viruses-1,-2, and-3. *J. Virol. Methods* **2009**, *162*, 276-279.
21. Sébastien, A.; Lester, P.J.; Hall, R.J.; Wang, J.; Moore, N.E.; Gruber, M.A. Invasive ants carry novel viruses in their new range and form reservoirs for a honeybee pathogen. *Biol. Lett.* **2015**, *11*, 20150610.
22. Yang, C.C.; Yu, Y.C.; Valles, S.M.; Oi, D.H.; Chen, Y.C.; Shoemaker, D.; Wu, W.J.; Shih, C.J. Loss of microbial (pathogen) infections associated with recent invasions of the red imported fire ant *Solenopsis invicta*. *Biol. Invasions* **2010**, *12*, 3307-3318.
23. Cremer, S. Pathogens and disease defense of invasive ants. *Curr. Opin. Insect. Sci.* **2019**, *33*, 63-68.

Chapter 2: Evolution of mitochondrial genome in the yellow crazy ant, *Anoplolepis gracilipes*

2.1 Introduction

The yellow crazy ant (*Anoplolepis gracilipes*) is a widespread invasive species of great conservation concern [1]. This ant is polygynous (multiple-queens) and often forms supercolonies [2]. Colonies usually reproduce by dependent foundation although independent foundation also was observed [3]. The mitochondrial DNA has been used as a genetic marker to characterize the population structure of this ant [4,5], however, no complete mitochondrial genome is currently available. Here, I present the first complete mitogenome for this species.

2.2 Materials and methods

A specimen of *A. gracilipes* was collected from Hsinchu City, Taiwan (24°46'43.43"N, 120°56'30.24"E) and preserved in 70% EtOH. I extracted high molecular weight genomic DNA from a single virgin queen via the Phenol-Chloroform purification method. I sequenced DNA on the Illumina HiSeq 2500 platform through PCR-free library construction with the average library insert size as 250 bp in paired-end mode (2 x 125 bp).

For *de novo* assembly and annotation, I removed Illumina adapters from raw sequence reads with Trimmomatic v0.36 [6] and then conducted *de novo* mitogenome assembly using NOVOPlasty v2.6.4. I used the *Camponotus atrox* mitogenome as the seed for de novo assembly [7]. I annotated protein coding genes (PCGs), rRNAs and tRNAs using MITOS [8], OrfFinder [9], and ARWEN [10].

I inferred the phylogenetic relationship of 16 ants and two bees, using the concatenated nucleotide sequences of all 13 PCGs. The phylogenetic tree was constructed by the Bayesian inference and Maximum likelihood methods under GTR+I+Γ and GTRGAMMA models, respectively. The Bayesian inference method conducted using MrBayes v3.2.5 [11] with 1,000,000 generations that first 25% of generations as burnin, and sampled in every 100 generations. The Maximum likelihood methods used RAxML v8.2.11 with 1,000 bootstrap replicates [12]. The mitogenome accession numbers for the tree construction are listed as follows: *Apis cerana* (GQ162109), *Atta texana* (MF417380), *Bombus consobrinus* (MF995069), *Camponotus atrox* (KT159775), *Cardiocondyla obscurior* (KX951753), *Formica fusca* (LN607805), *F. selysi* (KP670862), *Leptomyrmex pallens* (KC160533), *Linepithema humile*

(KX146468), *Myrmica scabrinodis* (LN607806), *Polyrhachis dives* (NC_030790), *Pristomyrmex punctatus* (NC_015075), *Solenopsis geminata* (NC_014669), *S. invicta* (NC_014672), *S. richteri* (HQ215539), *Vollenhovia emeryi* (NC_030176), *W. auropunctata* (KX146469).

2.3 Results and discussion

Similar to other ant mitogenomes, the complete mitogenome of *A. gracilipes* (GenBank: MH122734) is 16,943 bp. Average read coverage of the mitogenome assembly was 4,849 per nucleotide per base, providing ample depth for correctness. The nucleotide composition is AT-biased (72%). The mitogenome contains 13 PCGs, two rRNAs and 22 tRNAs, typical for most animals (Table 2.1.). The tRNAs, ranging in size from 59-75 bp, are similar to other ants (circa 54-90 bp). The gene order (GO) has three rearrangements compared to the ancestral insect (ancestor GO :: A. gracilipes GO; trnR trnN trnS1 trnE :: trnR trnS1 trnN trnE; -nad4 -nad4L trnT -trnP :: -nad4 trnT -nad4L -trnP; CR trnI -trnQ trnM nad2 :: CR trnM trnI -trnQ nad2). The first two GO rearrangements differed from two Formica species [13,14], whereas all the three rearrangements differed from *Camponotus atrox* and *Polyrhachis dives* [15,16]. Such differences in GO rearrangement are consistent with the tribe-level classification of these species where *Anoplolepis*, Formica and *Camponotus*/*Polyrhachis* belong to Plagiolepidini, Formicini and Camponitini, respectively [17]. All PCGs use ATN as the start codon (N, any nucleotide) and TAA as the stop codon. The control region presumably corresponds to the single largest non-coding AT-rich region (893 bp, A+T 93%). The result of phylogeny analysis showed both Bayesian inference and maximum likelihood computation agree with the current phylogenetic placement of *Anoplolepis* in the Formicinae (Figure 2.1).

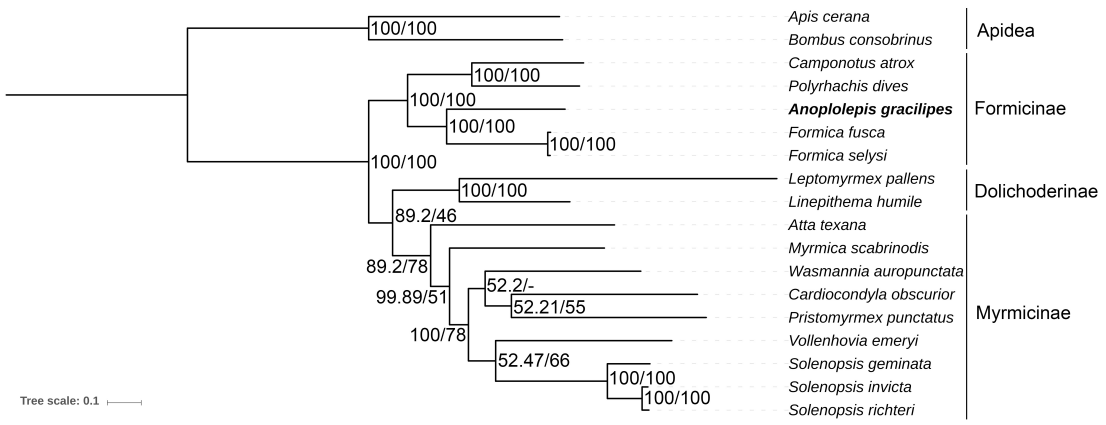


Figure 2.1 Molecular phylogeny of *Anoplolepis gracilipes* and 17 other Hymenoptera species (15 ants and two bees) based on the concatenated nucleotide sequences of 13 PCGs. The numbers at each node indicate the posterior probability and the bootstrap value resulting from the analyses. Note that only Bayesian probability is shown at the

branch node of *Wasmannia auropunctata* as slight differences in placement of this species is observed between the two phylogenetic methods. I rooted the phylogenetic tree at *A. cerana* and *B. consobrinus* (two bees) based on their exclusion from Formicidae (ants).

Table 2.1 Mitochondrial genome organization of *Anoplolepis gracilipes*

Gene	Strand	Position		Spacer (+)	Size (bp)	Start	Stop	Anti
		Start	Stop	overlap (-)		codon	codon	codon
cox1	J	1	1530	24	1548	ATG	TAA	
trnL2(tta)	J	1555	1620	-15	66			TAA
cox2	J	1606	2301	34	696	ATT	TAA	
trnK(aaa)	J	2336	2408	0	73			TTT
trnD(gac)	J	2409	2483	0	75			GTC
atp8	J	2484	2645	-19	162	ATA	TAA	
atp6	J	2627	3307	30	681	ATG	TAA	
cox3	J	3338	4162	39	825	ATA	TAA	
trnG(gga)	J	4202	4271	-3	70			TCC
nad3	J	4269	4625	55	357	ATA	TAA	
trnA(gca)	J	4681	4755	5	75			TGC
trnR(cga)	J	4761	4830	20	70			TCG
trnS1(aga)	J	4851	4909	29	59			TCT
trnN(aac)	J	4939	5007	20	69			GTT
trnE(gaa)	J	5028	5097	3	70			TTC
trnF(ttc)	N	5101	5170	17	70			GAA
nad5	N	5188	6852	-3	1665	ATT	TAA	
trnH(cac)	N	6850	6917	96	68			GTG
nad4	N	7014	8405	-15	1392	ATA	TAA	
trnT(aca)	J	8391	8459	52	69			TGT
nad4l	N	8512	8877	-46	366	ATA	TAA	
trnP(cca)	N	8832	8899	65	68			TGG
nad6	J	8965	9549	-5	585	ATC	TAA	
cob	J	9545	10678	58	1134	ATT	TAA	
trnS2(tca)	J	10737	10808	164	72			TGA
nad1	N	10973	11929	-3	957	ATA	TAA	
trnL1(cta)	N	11927	11996	-23	70			TAG
l-rRNA	N	11974	13413	32	1440			
trnV(gta)	N	13446	13514	0	69			TAC
s-rRNA	N	13515	14321	0	807			
CR		14322	15214	0	893			
trnM(atg)	J	15215	15288	50	74			CAT
trnI(atc)	J	15339	15410	13	72			GAT
trnQ(caa)	N	15424	15495	54	72			TTG
nad2	J	15550	16569	68	1020	ATT	TAA	
trnW(tga)	J	16638	16711	9	74			TCA
trnC(tgc)	N	16721	16789	27	69			GCA
trnY(tac)	N	16817	16883	60	67			GTA

2.4 References

1. O'Dowd D.J.; Green P.T.; Lake P.S. Invasional 'meltdown'on an oceanic island. *Ecol. Lett.* **2003**, *6*:812-817.
2. Hoffmann B.D.; Hagedorn H. Quantification of supercolonial traits in the yellow crazy ant, Anoplolepis gracilipes. *J. Insect Sci*. **2014**, *14*,25.
3. Ito F.; Asfiya W.; Kojima Ji. Discovery of independent-founding solitary queens in the yellow crazy ant Anoplolepis gracilipes in east Java, Indonesia (Hymenoptera: Formicidae). *Entomol. Sci.* **2016**, *19*,312-314.
4. Gruber M.A.; Hoffmann B.D.; Ritchie P.A.; Lester P.J. Recent behavioural and population genetic divergence of an invasive ant in a novel environment. *Divers. Distrib.* **2012**, *18*,323-333.
5. Thomas, M.L.; Becker, K.; Abbott, K.; Feldhaar, H. Supercolony mosaics: two different invasions by the yellow crazy ant, Anoplolepis gracilipes, on Christmas Island, Indian Ocean. *Biol. Invasions* **2010**, *12*, 677-687.
6. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114-2120.
7. Dierckxsens N.; Mardulyn P.; Smits G. Novoplasty: De novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* **2016**, *45*,e18-e18.
8. Bernt M.; Donath A.; Jühling F.; Externbrink F.; Florentz C.; Fritzsch G.; Pütz J.; Middendorf M.; Stadler P.F. MITOS: Improved de novo metazoan mitochondrial genome annotation. *Mol. Phylogen. Evol.* **2013**, *69*, 313-319.
9. Coordinators NR. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **2016**, *44*, D7.
10. Laslett D.; Canbäck B. Arwen: A program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. *Bioinformatics* **2007**, *24*, 172-175.
11. Ronquist, F.; Huelsenbeck, J.P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **2003**, *19*, 1572-1574.
12. Stamatakis A. Raxml version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **2014**, *30*,1312-1313.
13. Babbucci M.; Basso A.; Scupola A.; Patarnello T.; Negrisolo E. Is it an ant or a butterfly? Convergent evolution in the mitochondrial gene order of Hymenoptera and lepidoptera. *Genome Biol. Evol.* **2014**, *6*,3326-3343.
14. Yang S.; Li X.; Cai L-G; Qian Z-Q. Characterization of the complete mitochondrial genome of Formica selysi (Insecta: Hymenoptera: Formicidae: Formicinae). Mitochondr DNA Part A. **2016**, *27*,3378-3380.
15. Kim M.J.; Hong E.J.; Kim I. Complete mitochondrial genome of Camponotus atrox (Hymenoptera: Formicidae): A new tRNA arrangement in Hymenoptera. *Genome* **2015**, *59*, 59-74.
16. Liu J.-H.; Jia P.-F.; Fu J.-Q.; Dan W.-L.; Yang L.-Y.; Wang Q.-M.; Li Z.-N. Characterization of mitochondrial genome and phylogenetic implications for chinese black ant, Polyrhachis dives (Hymenoptera: Formicidae). *Mitochondr DNA Part B* **2017**, *2*,679-680.
17. Blaimer B.B.; Brady S.G.; Schultz T.R.; Lloyd M.W.; Fisher B.L.; Ward P.S. Phylogenomic methods outperform traditional multi-locus approaches in

resolving deep evolutionary history: A case study of Formicine ants. *BMC Evol. Biol.* **2015**, *15*,271.

Chapter 3: Population genomics of *Anoplolepis gracilipes*

3.1 Introduction

Human-mediated jump dispersal occasionally spreads nonindigenous species to novel environments where they never experienced. How a nonindigenous species adapts novel environments and successfully sustained therefore represent an excellent opportunity for understanding the evolution progress [1]. Ants are successful invaders that their social traits as part of reasons to dominate in new environments [2]. Many invasive ants are polygynous (multiple reproductive queens within a colony), polydomous, dependent foundation (e.g. budding), and often form supercolonies (i.e. a multiple physically separated colonies but reduce or lack territoriality among colonies) [2]. With the large colony size, a supercolony has advantage to increase nest densities with rapid population growth rate [3] and reduces the consumption of intraspecific competition [4]. The genomic analysis of the red imported fire ant (*Solenopsis invicta*) supported social traits are crucial in successful invasion of ants [5]. Beside social traits, other biotic and abiotic factors may further influence the successful invasion in ants such as the released from nature enemies [6] or temperatures tolerance [7,8]. Mutations that are neutral or even deleterious in the native range may prove to be advantageous in the novel environment of the introduced range [9]. For example, a tropical ant, *Wasmannia auropunctata*, evolves thermal and cold temperature tolerance in native environment before human media long distance jump dispersal [7,8] suggest the genetic variant in native range is one of crucial mechanisms during biological invasion.

The yellow crazy ant, *Anoplolepis gracilipes* is a widespread invasive ant that major distributed in tropical Asia and Indo-Pacific islands. Although the native range of *A. gracilipes* remains undetermined, it putative originate from Southeast Asia [10,11]. Yellow crazy ants extend their territories into humid subtropical climate regions and form a sustainable population [10]. Common social traits associate with invasive ants present in *Anoplolepsi gracilipes* including polygynous, polydomous, budding dispersal, and form a supercolony. The high density of supercolonies and social traits could be one of factors determined the dominance of *A. gracilipes* in invasion area [3,12]. However, previous study in Australia [13] reported *A. gracilipes* population declines without human intervention that indicates other confounding factors besides social traits may determine the consequence of *A. gracilipes* sustain in long time-scale. For example, the gene flow through individual migration or multiple introductions generally maintain or enhance the genetic diversity that may facilitate the rapid evolution during evolution [1]. In contrast to most other ant species, nuptial flight of supercolony-forming is rare or absent in *A. gracilipes*, and mating often takes place

within the nest or supercolony [14,15]. The consequence of this mating strategy may limit the gene flow between supercoloies [14,15]. Previous study suggests the gene flow is limit or absent between two sympatric supercolonies in Christmas Island, Indian Ocean [16], the scarce gene flow even drives a supercolony divergence after single biological invasion [17]. This mating strategy providing an excellent opportunity to explore the how the yellow crazy ant dominate in various of introduced region with limit gene flow.

In this study, our objective was to investigate the population structure and evolution consequence of *A. gracilipes* at a global scope. I hypothesized that *A. gracilipes* may have distinct divergent populations at large geographic regions even with frequent anthropogenic dispersal. With the double digest restriction site-associated DNA sequencing (ddRAD-seq), I obtained single nucleotide polymorphisms (SNPs) across 13 geographic regions for analyzing population structure, phylogenomic relationship, and gene flows. I next inquisitive about whether genome regions under positive selection during the colonization for potential local adaptions. I used genome outlier approaches for identifying SNP loci that link to genome regions under positive selection. I further pinpointed genes potentially associated with adaptive evolution. This is the first population genomic study with large geographic scale in this notorious invasive ant.

3.2 Materials and methods

3.2.1 Sample Collection, DNA Isolation and ddRAD-seq

A total of 73 colonies were collected across 13 countries/regions (Figure 1, Table S1) from 2012 to 2019. Ant workers were preserved in 95% EtOH until DNA extraction. A single worker per colony was randomly selected for DNA purification using the QIAamp DNA Mini Kit (Qiagen, Valencia, CA, USA). The ddRAD-seq was employed following protocol described in Perterson et. al., 2012 with slight modifications [18]. In brief, DNA quantity was first measured using the Qubit dsDNA BR assay kit (Thermo Fisher Scientific, Waltham, MA, USA) and took 50-100ng DNA per sample. The restriction enzyme digestion included HpyCH4IV (New England BioLabs, Ipswich, MA, USA) and BfaI (New England BioLabs, Ipswich, MA, USA) that required to be incubated with DNA sample at 37°C for 3 hours. The following clean-up was carried out using AMPure XP magnetic beads (Beckman Coulter, Pasadena, CA, USA). Both 3' and 5' end adaptors (including 5 bp inline barcode) were ligated using T4 DNA ligase (New England BioLabs, Ipswich, MA, USA) at 23°C for 30 min, and then inactivated at 65°C for 20 minutes. Ligated DNA was mixed into a DNA pool at one to one ratio based on the amount of original DNA input. The DNA pool was purified and concentrated using AMPure XP magnetic beads (Beckman Coulter, Pasadena, CA, USA). Size selection (~550-~700 bp) of the condensed DNA was performed in 2%

agarose gel (Invirtogen, Carlsbad, CA, USA), and DNA fragments were purified using the FastGene gel/PCR extraction kit (Nippon genetics, Tokyo, Japan). The 12 cycles of two steps PCR enrichment was conducted to add complete Illumina adapter and index using Phusion high-fidelity PCR kit (New England BioLabs, Ipswich, MA, USA) with the following conditions: 98°C for 30 sec., 12 cycles of 98°C for 10 sec. and 72°C for 30 sec., with the final extension at 72°C for 10 min., and hold at 4°C. The DNA library was purified using AMPure XP magnetic beads (Beckman Coulter, Pasadena, CA, USA) again before being mixed with 10% PhiX. The finalized library was then sequenced by Macrogen Japan (Kyoto, Japan) under the single lane of Illumina HiSeqX platform in pair-end (2x150 bp) mode.

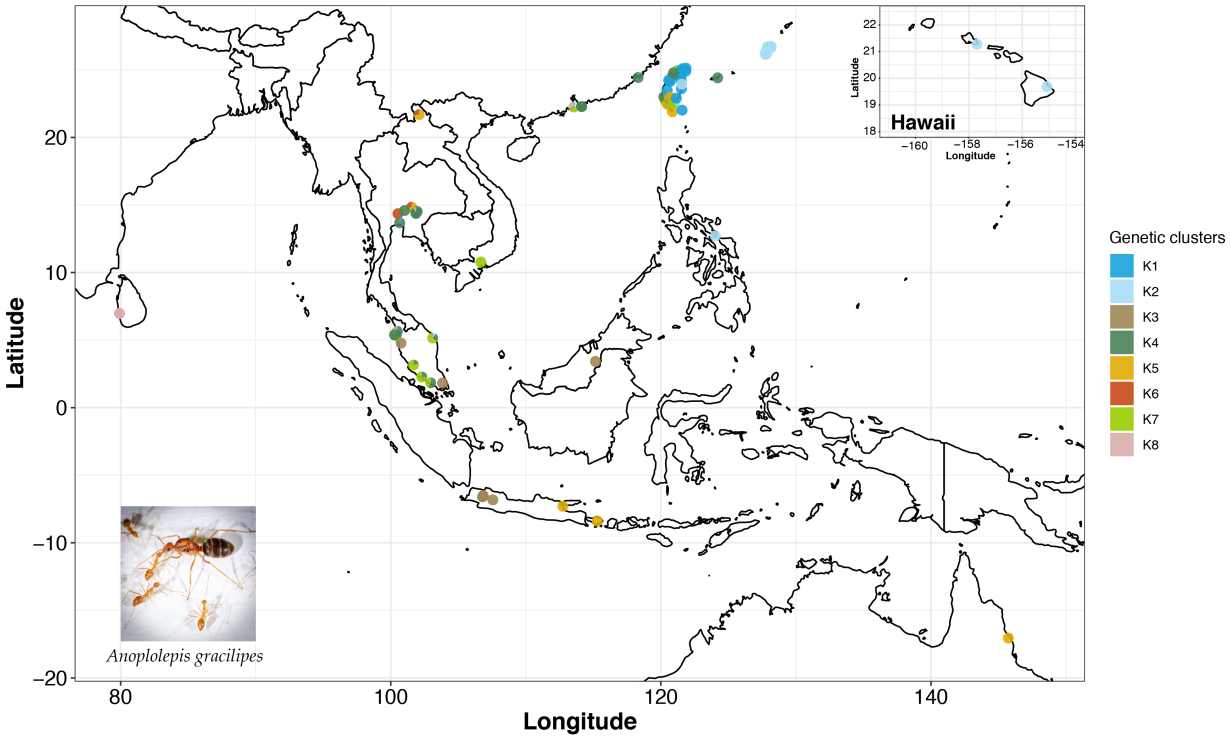


Figure 3.1 Geographic distribution of 73 yellow crazy ant colonies in this study. The center of pie chart indicates the sample location and colors with percentage of ancestral genetic clusters be assigned to each *A. gracilipes* colony. I choose 8 genetic clusters (K=8) for representing the *A. gracilipes* population structure. (Photo credit: Hung-Wei Hsu)

3.2.2 Variant Calling and SNP Filtering

The quality of raw reads by Illumina sequencing was first examined using FastQC [19]. For the insertion size of ddRAD-seq library construction in range of ~550-700 bp, 3′ adapters should not be present in any sequences. Reads containing 3′ adapters therefore indicated the insertion size was < 300 bp, which was out of our target size range and thus were discarded. Reads were trimmed by 1 bp on their 5′ end for both R1 and R2 reads to standardize the start position for barcode demultiplexing, and by 30 bp from the 3′ end of R2 reads to remove low quality bases. Using Cutadapt v2.8

[20], the demultiplexing was performed by implementing the anchored primer method using 4 bp barcode at 5′ end with restriction enzyme recognition sequence as an index (-g ^<barcode><enzyme recognition sequence>; 7 bp in total; for both R1 and R2 reads), and no error was allowed (-e 0.01, --no-indels). A reference-based method was used to identify nucleotide variants. Reads were mapped to the *Anoplolepis gracilipes* "core" genome (see 2.3) using BWA v0.7.17 [21], and SNPs were called using gstacks implemented in Stacks2 v2.5 [22]. To minimize false positive variant, mapped SNP loci without proper pairs, extremely rare alleles (frequency < 0.05), and alleles present in only a single individual were discarded (--rm-unpaired-reads, --min-maf 0.05, --min-mac 3), using population program in Stacks2 v2.5 [22]. To reduce the influence of linkage disequilibrium, a single SNP was kept per locus (--write-single-snp). To reduce effects of missing loci, SNP loci presented in more than 80% of all samples were retained for further analyses handling by populations program (-R 0.8) in Stacks2 v2.5 [22]. To ensure the accuracy of genotype calling, SNPs with read depth less than 10 (--minDP 10) were removed to prevent the inconsistent calling [23,24]. Since the nucleotide polymorphism calls from extremely high coverage loci may cause by PCR artifacts of over amplification on certain loci or mapping errors derive from repetitive sequences [23,24], the SNP loci with more than three times the standard deviation of SNP read depth (--maxDP 104) were excluded using VCFtools v0.1.15 [25].

3.2.3 Reference Genome *de novo* Assembly

The yellow crazy ant was previously reported to harbor *Wolbachia* and that the endosymbiont appears to be nearly fixed in populations across the ant's geographical range [26,27]. To exclude *Wolbachia*-origin SNPs or other potential contaminants, the reference-based SNP calling method was used in this study. Due to the current lack of the yellow crazy ant reference genome, a contig-level *A. gracilipes* "core" genome was generated using the following method. A yellow crazy ant virgin queen was collected from Hsinchu City, Taiwan (24°46′43.43″ N, 120°56′30.24″ E) in 2014, and its high molecular weight DNA was purified via the Phenol-Chloroform purification method. An PCR-free Illumina sequencing library was constructed and sequenced by Sequencing Technology Company (Taipei, Taiwan) through Illumina HiSeq 2500 platform (average library insert size: 250 bp; paired-end read length: 2x125 bp).

Illumina adapters, low-quality bases (base call accuracy < 99.8%, q <28), and short reads (< 36 bp after trimming) were trimmed using Trimmomatic v0.36 [28]. The whole genome *de novo* assembly was conducted using IDBA-UD [29] with a minimum read support of 10 (--min_support 10; other parameters: default). Contigs sizes smaller than 250 bp were excluded from further analyses. The *A. gracilipes* genome contigs were then separated from others via mega-BLASTn [30] against a Formicinae ant *Camponotus floridanus* genome v7.5 (GenBank: GCA_003227725.1). Contigs with e-value lower than 1e-20 were treated as our "core" reference genome. The completeness

of genome assembly was evaluated using BUSCO v4.0.5 [31], which measures the proportion of highly conserved, single-copy orthologs derived from 40 Hymenoptera species (Hymenoptera_odb10, 5,991 BUSCO orthologs). The assembly statistics were calculated using QUAST v5.0.2 [32].

3.2.4 SNP Datasets for Population Genomics Analyses

A total of four SNP datasets were generated for different population genomic analyses. First, the original dataset (dataset 0) comprised all SNP loci identified by gstacks, and was used for kinship analysis (see 2.5). Second, the all-qualified-loci dataset (dataset 1) included all qualified SNP loci based on all 73 individuals in our collection. This dataset was used in determination of genetic structure, kinship analysis, reconstruction of phylogenomic tree (see 2.5) and detection of outlier loci (see 2.7). Third, the neutral-loci dataset comprised SNP loci that were a subset of the all-qualified-loci dataset after the removal of loci potentially under selection (see 2.7). This dataset was seemingly more unbiased in revealing population structure and gene flow of a species [33]. The neutral-loci dataset (dataset 2) thus was employed to estimate genetic clustering of *A. gracilipes* (see 2.5) and migrant estimation (see 2.6). Lastly, individuals were separated in two sub-populations based on the results of the kinship coefficient and genetic clustering (see 2.5) for identification of putative adaptive loci between the two distinct genetic clusters (Clade I and III, see 2.5). The two-genetic-clusters dataset (dataset 3) was essentially the neutral-loci dataset with predefined population structure.

3.2.5 Summary Statistics, Population Genetic Structure and Phylogenomic Analysis

Population genetic summary statistics were calculated based on the all-qualified-loci dataset (dataset 1) using populations program implemented in Stacks2 v2.5 [22]. The statistics included genetic diversity (as expressed by the average proportion of nucleotide differences between all possible pairs of sequences obtained from a population (pi)), percentage of polymorphic SNP loci, observed heterozygosity and expected heterozygosity for all geographical populations.

The kinship coefficient was one of relatedness analyses to recognize sibling relationship by genetic similarity, and expressed by probability of a pair of randomly sampled homologous alleles being identical by descent [34]. The kinship coefficient > 0.354 represents the twin pairs like genetic similarity in diploid species, and as the full sibling in range 0.177-0.354 [34]. To determine the supercolony identity across a board geographical range, the original and all-qualified-loci dataset were used to perform kinship inference using KING algorithms (kingship-based inference for GWASs) implemented in VCFtools v0.1.15 (--relatedness2) [25,34]. The result of kinship coefficient was subsequently clustered and plotting by pheatmap package in R.

Inference of genetic clusters of species with broad geographic distribution could be achieved by assigning samples into a series of presumed genetic clusters (K). For unbiased estimating the population structure, both all-qualified-loci (dataset 1) and neutral-loci datasets (dataset 2) were used for genetic clustering analysis. The population cluster was calculated using Bayesian-based discriminant analysis of principle components (DAPC) estimation implemented in adegenet v2.1.0 package in R [35], and maximum likelihood approach implemented in ADMIXTURE v1.3 [36], respectively. The number of genetic clusters was calculated from K=1 to K=20. The optimal genetic cluster (*K*) was determined using cross-validation procedure with default values (5-fold) and 10-fold cross-validation and Bayesian-based DAPC methods which retained 65 out of 72 principal components (PCs) for sequential *k-means* clustering associated with Bayesian Information Criterion (BIC).

To understand the phylogeographic structure of sampled *A. gracilipes* colonies, the maximum-likelihood tree based on the all-qualified-loci dataset (dataset 1) was reconstructed using RAxML-NG v0.1.6 [37]. The confidence for the phylogeny tree was assessed using 1,000 bootstrap replications with the weighted Robinson-Foulds (WRF) distance (< 3% as the threshold) [38]. As SNPs were potentially scattered throughout the *A. gracilipes* genome, estimation of the most appropriate mutation model was impractical due to varying mutation rates across different genomic regions. I thus used the Jukes-Cantor model when reconstructing the phylogenomic. Phylogenomic trees were visualized using iTOL v5 [39].

3.2.6 Estimation of Barriers to Gene Flow

To estimate the pattern of gene flow among geographical populations of *A. gracilipes*, the estimated effective migration surfaces (EEMS) method [40] was employed to identify how genomic diversity of *A. gracilipes* is geographically structured, which informs genetic dissimilarity and the spatial pattern of gene flow. The EEMS method was based on a stepping-stone model, where migration occurring between neighboring demes were modeled in a number of dense regular grids [40]. The expected genetic dissimilarity between two individuals was calculated over all possible migration histories and routes between their demes on the grids. The missing data were imputed based on the average genotype for a given locus using bed2diffs_v2 in the EEMS. The deme size was set as 2,000 from South Asia to West-Pacific region. Individuals from Hawaii (n=2) were excluded from this analysis due to the lack of samples collected from most Pacific region that may bias the analysis. The MCMC iterations were set as 5,000,000 with a 2,000,000 burn-in and thinning every 9,999 iterations. The evaluation of the MCMC convergence and plotting of EEMS results were conducted using the R package rEEMSplots.

3.2.7 Detection of Loci under Positive Selection

Two approaches were conducted to scan our SNP datasets for candidate outlier loci. First, I employed a F_{ST}-based outlier approach implemented in BayeScan v2.1 [41], with parameters set as follows: a burn-in period with 50,000 iterations, a thinning interval of 10, a sample size of 95,000 and 20 pilot runs with 5,000 iterations each, for a total of 100,000 iterations. The Bayesian method assumed the demes have evolved independently from an ancestral gene pool [42]. Therefore, the BayeScan analysis was performed on the two-genetic-clusters dataset (dataset 3) as this dataset was constructed with predefined population structure based on genetic clusters (dataset 3, also see results 3.2: Clade I (n=55) and III (n=17)). Note that I excluded the sample Thailand_01 from the BayeScan analysis because our objective was to determine outlier loci associated with adaptive evolution during colonization, the ambiguous assignment of this particular individual when K=2, and that this genetic singleton could not represent a population (see 3.2). Second, a PCA-based statistic implemented in R package pcadapt v4.3.3 [43] was utilized to detect outlier loci based on the all-qualified-loci dataset (without predefined population; dataset 1). Two principal components were retained. The false discovery rate (FDR) of both methods was set at 1% with q-value < 0.01 (BayeScan) and q-value < 0.01 after Bonferroni correction (pcadapt).

To reduce the false positive outlier identification, only those SNP loci with positive "outlier signal" revealed by both approaches were regarded as candidate outliers. Genes that are located nearest the SNP outliers were identified by comparing the genome contig in which the outlier was located against NCBI GenBank non-redundant (nr) database using BLASTx searches. To obtain the neutral-loci dataset used in analysis of genetic clustering, SNP loci with $\log_{10}$(Probability) > 1 identified by BayeScan were regarded as potentially non-neutral loci and then excluded from the dataset. The remaining SNPs without signal for natural selection comprised the neutral-loci dataset.

3.3 Results

3.3.1 ddRAD-seq and de novo genome assembly

The *A. gracilipes* "core" genome consisted of 23,616 contigs with a total length of 204,485,205 bp. The assembly summary statistics indicated the N50 of the assembly was 16,560 bp, and the percentage of GC was 33.78 with average coverage depth as 48 reads per nucleotide. The BUSCO score showed that 80.2% (4,808 out of 5,991 complete BUSCOS) of *A. gracilipes* core genome matched the existing single copy Hymenoptera gene orthologs.

The initial reference genome based on variant calling revealed 46,563 SNP loci in 73 *A. gracilipes* workers. Among these loci, 33,885 to 35,987 SNPs failed to pass the

filters thus were excluded in our analysis (Table 3.1). As the result, I obtained a total of 11,476 SNP loci with the average coverage of 37.65 (minimum mean depth: 23.54, maximum mean depth: 55.83; Figure 3.2), which were defined as the all-qualified-SNP dataset. Another two datasets were further constructed: the neutral-loci dataset composed a subset of loci from the all-qualified-SNP dataset from which 133 loci were excluded due to the presence of signal indicative of potentially under selection (see Results 3.3.4); the two-genetic-cluster dataset with predefined populations (phylogenomic Clade I and Clade III) for investigating genome outliers. Detailed information of all SNPs dataset for subsequent evolutionary analysis was summarized in Table 3.1.

Table 3.1 SNP datasets obtained from ddRAD-seq in *Anoplelopis gracilipes*

No.	Dataset	SNP loci	Individuals	Analysis: program
0	Original	46,563	73	Kinship analysis: VCFtools v0.1.15
1	All qualified loci	11,476	73	Genetic diversity: Stacks2 v2.5 Kinship analysis: VCFtools v0.1.15 Population structure: ADMIXTURE v2.3, DAPC in adegenet v 2.1.0 Phylogeny: RAxML-NG v0.1.6 Outlier: pcadapt v4.3.4
2	Neutral loci	11,357	73	Population structure: ADMIXTURE v2.3, DAPC in adegenet v 2.1.0 Gene flow: EEMS
3	Two Genetic Clusters	12,678	72 (55+17)	Outlier: BayeScan v2.1

Data subset of No. 1-3 were filtered from the original one (No. 0) by following common parameters: SNP loci presented in at least 80% individuals, alleles appeared in at least 2 individuals, locus with minimum minor allele frequency 0.05, and kept single SNP per locus.

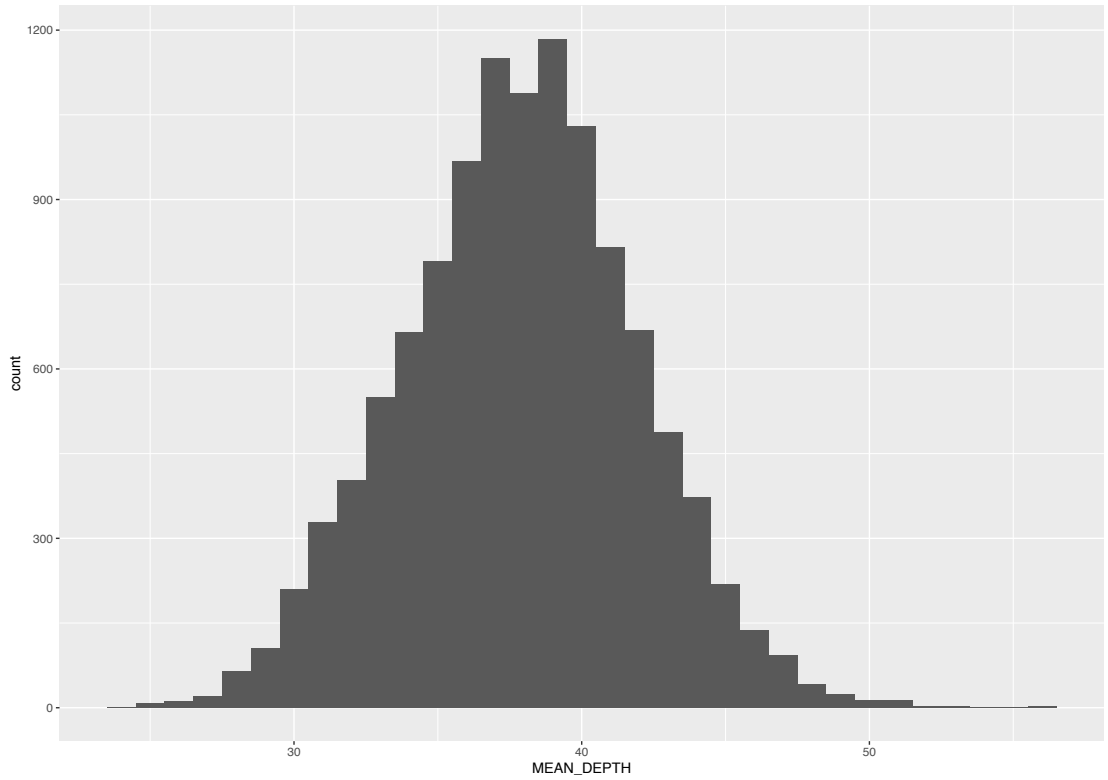


Figure 3.2 The coverage (mean depth) distribution of 11,476 qualified SNP loci from 73 *A. gracilipes* individuals. The average coverage is 37.65, minimum mean depth is 23.54, and maximum mean depth is 55.83.

3.3.2 Summary Statistics and Population Structure in Anoplolepis gracilipes

Result of summary statistics for genetic diversity, percentage of polymorphic SNP loci, observed and expected heterozygosity were listed in Table 2. One notable pattern is that the observed heterozygosity was greater than expected heterozygosity in all geographical populations.

The result of kinship analysis identified the presence of three major or common families (pairwise kinship coefficient >0.354) in all *A. gracilipes* workers. A sample from Thailand (Thailand_01) possessed a negative value of kinship against other samples, suggesting genetic distinctness or unique colony structure compared to others. Two major famlies comprised 55 and 50 individuals based on the origin dataset and all-qualified-loci, respectively. These individuals included *A. gracilipes* from Thailand, Malaysia, Taiwan, Okinawa, Philippines, Vietnam, China coast (Hong Kong, Zhuhai, and Kinmen), and Sri Lanka. The other family included *A. gracilipes* from Malaysia, Borneo, Indonesia, Laos, Thailand, Australia, and Southern part of Taiwan (Figure 3.3).

The results of ADMIXTURE based on the neutral-loci dataset and all-qualified-loci dataset were similar (Figure 3.4), and thus provided the first insights into "optimal" number of genetic clusters (K) to explain the population structure of *A. gracilipes*. The cross-validation error of the ADMIXTURE estimation seems unstable between two

datasets after K=10. The cross-validation error of the neutral-loci dataset was lowest when K=6 (10-fold CV error: 0.108), but only slightly higher than that of K=3 (10-fold CV error: 0.11) and 8 (10- fold CV error: 0.112; Figure S3), whereas the find.clusters DAPC analysis produced the lowest BIC at K=8 (Figure 3.5) for both datasets. According to the results of the two clustering methods, I chose K=8 for visualizing the population structure of *A. gracilipes* in Figure 1.

Table 3.2 Summary statistic of each geographical locations based on 11,476 SNP loci (dataset 1).

Sample region	Number of colonies	Pi	% polymorphic loci	H_o	H_e
n>5					
Thailand	7	0.423	91.89	0.625	0.384
Malaysia	10	0.418	94.53	0.676	0.394
Taiwan	26	0.402	96.95	0.683	0.393
Okinawa	9	0.382	75.65	0.700	0.358
Indonesia	6	0.372	71.30	0.624	0.334
n<5					
Southeast coastal China*	4	0.440	75.09	0.691	0.360
Vietnam	3	0.441	73.76	0.689	0.356
Sri Lanka	2	0.460	68.66	0.681	0.342
Hawaii	2	0.552	69.79	0.696	0.348
Borneo	1	0.616	61.55	0.616	0.308
Australia	1	0.641	64.11	0.641	0.321
Laos	1	0.642	64.22	0.642	0.321
Philippines	1	0.696	69.65	0.697	0.348

Note. *Sample from Southeast coastal China include Hong Kong, Zhuhai, and Kinmem Island, Taiwan. Pi: genetic diversity; H_o: observed heterozygosity; H_e: expected heterozygosity.

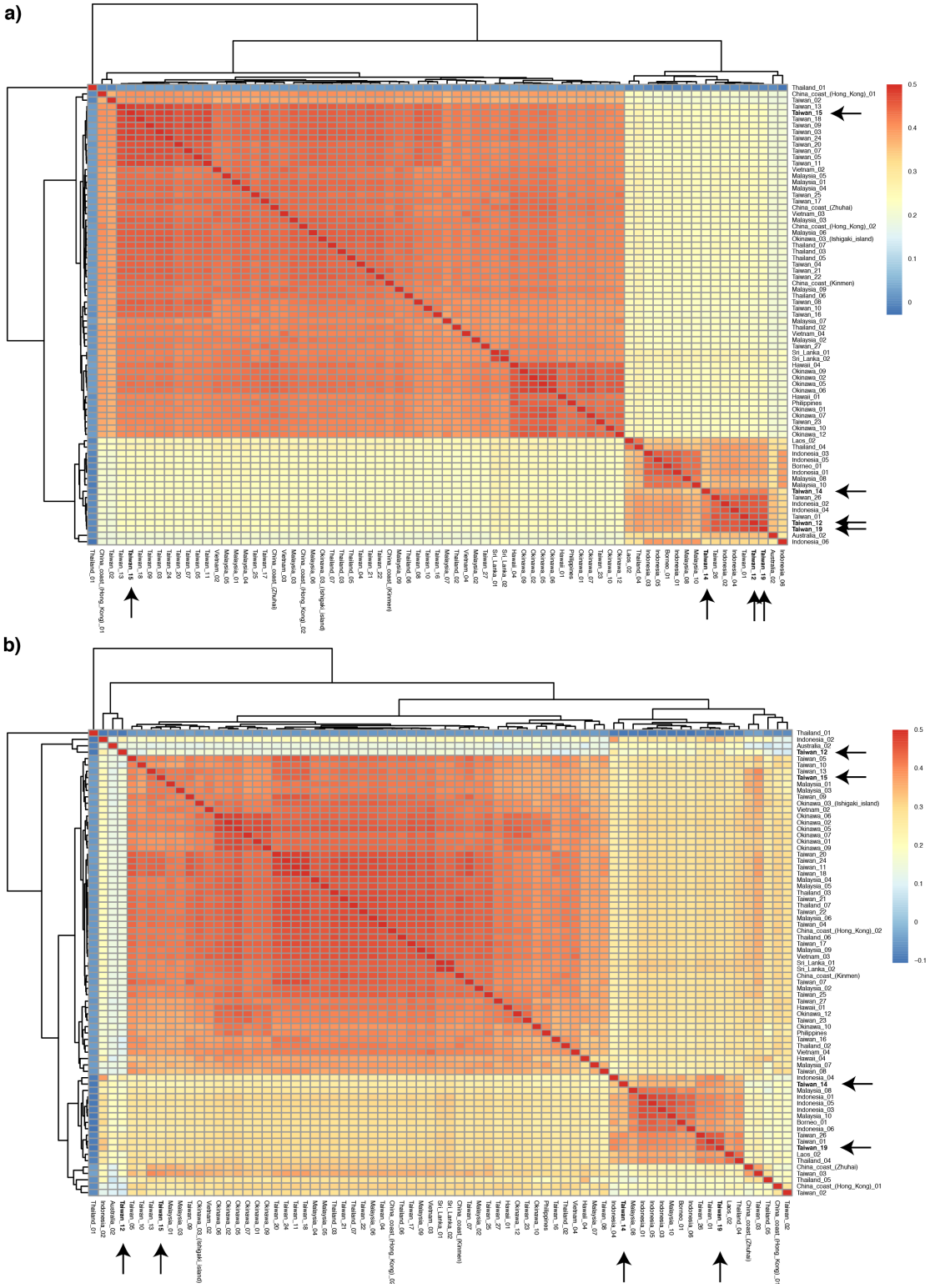


Figure 3.3 The pairwise kinship coefficient of *A. gracilipes* based on KING algorithms calculated by VCFtools v0.1.15 with a) original SNP dataset (dataset 0) and b) all qualified SNPs (dataset 1). Arrows indicated ant colony collected from small geographical regions (between colonies geographical distance 3.8-10.6 km). Note the order of ant colonies are different in a) and b).

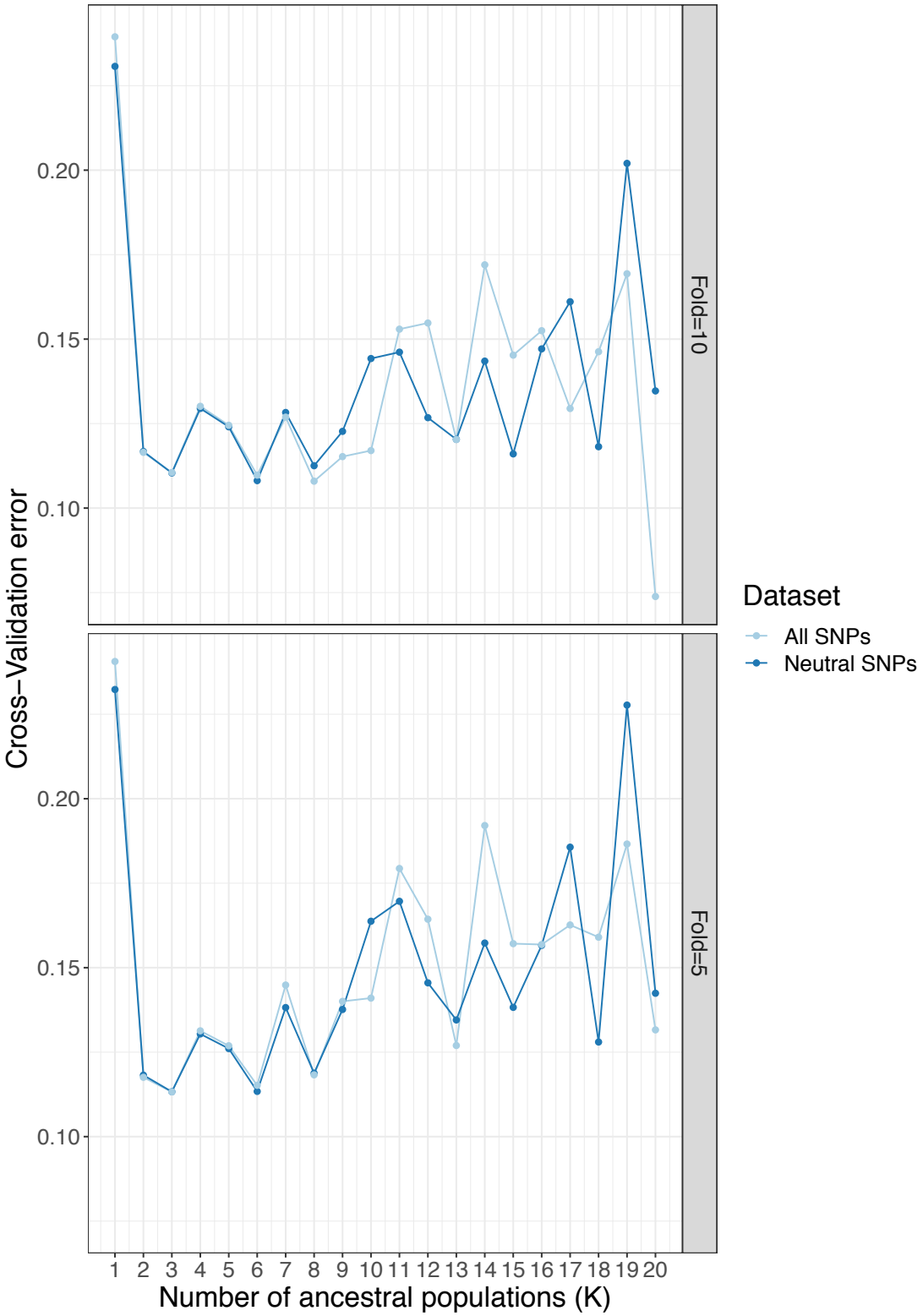
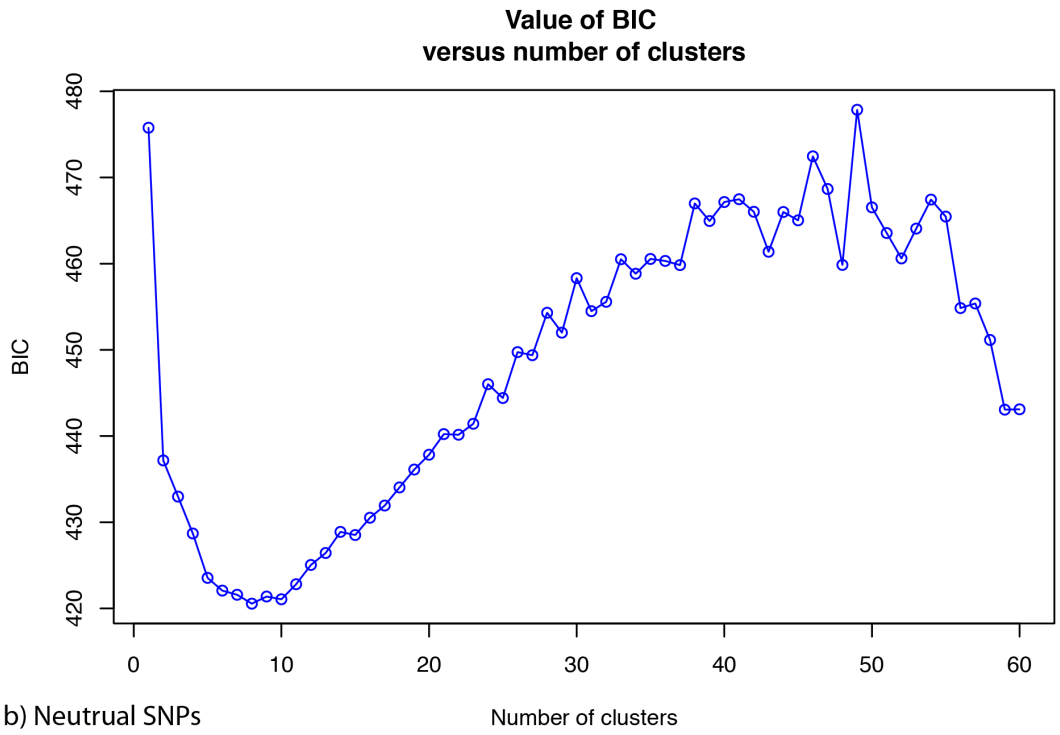


Figure 3.4 Cross validation error of ADMIXTURE estimation from K=1-20 in 5 and 10 folds. The lower CV error presented when K=3, 6, 8 in both folds of each datasets.

a) All quilified SNPs



b) Neutrual SNPs

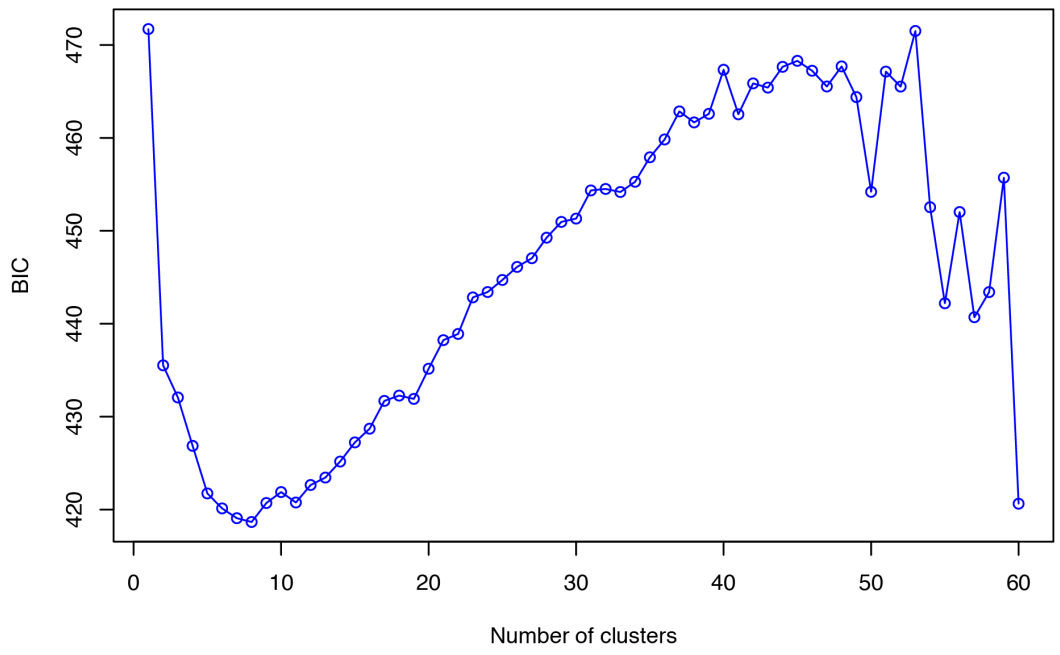


Figure 3.5 Bayesian Information Criterion (BIC) value based on *k-means* with 65 principal components in series number of clusters (K) using find.clusters in adegenet. The lowest BIC value for both (a) all qualified SNPs, and (b) neutral SNPs presented when K=8.

Assignment of individuals to hypothetical clusters changed as the number of K increased in ADMIXTURE (Figure 3.6). When K=2, individuals from Indonesia, Borneo, Australia, Malaysia, Laos, and partial samples of southern Taiwan were grouped into a single genetic cluster (Figure 3.6, phylogenomic Clade III) and other individuals (from South, Southeast, East Asia to Indo-Pacific islands) were clustered into another genetic cluster (Figure 3.6, phylogenomic Clade I). One particular sample, namely Thailand_01, was assigned to one of the two genetic groups. The Thailand_01 was represented by a single, independent genetic cluster from others when K=4. In the same strata, a "Pacific" genetic cluster (Figure 2, light blue) emerged from others and this cluster included individuals from Okinawa, Hawaii, Philippines, and eastern Taiwan. A genetic cluster (K1; dark blue) composed of some samples in Taiwan only (Figure 3.1, Figure 3.7) was separated from others when K=5. When K=6 individuals from Malaysia, Borneo, and western part of Java were separated from individuals of eastern part of Java, Australia and southern part of Taiwan. The K=8 possessed a similar clustering pattern to that of K=6 with a notable exception involving but *A. gracilipes* from Sri Lanka that formed its own genetic group. Interestingly, several *A. gracilipes* individuals from Ishigaki island (Japan), Taiwan, China, Vietnam, Thailand and Malaysia were characterized with an admixture signal and could not be assigned into a single, major genetic cluster.

The maximum likelihood phylogenomic tree was roughly divided into three major clades (Clade I~III) with each receiving long branch length and 100% bootstraps support (Figure 3.6). Clade I and III then were subdivided into two and three subclades, respectively, with each supported by >90 bootstraps values. *Anoplolepis gracilipes* in Clade I displayed negligible geographical differentiation except those among certain small islands (e.g., Sri Lanka, Okinawa, and Hawaii). Clade I subsequent separated Sri Lanka (Clade I-1) with others from Thailand, Malaysia, Vietnam, Philippines, coastal of China (Hong Kong, Zhuhai, and Kinmen), Okinawa, Hawaii, and Taiwan (Clade I-2). In Clade III, most of individuals from Malaysia, Borneo, and western part of Java formed a unique genetic subcalde (Clade III-1), whereas colonies near the eastern part of Java, Australia and the southern part of Taiwan were grouped in another subclade (Clade III-2). Clade III-3 composed of individuals from Thailand and Laos. Inspection of geographic distribution of each genetic cluster revealed that virtually all *A. gracilipes* samples in Clade III were distributed in the tropical regions, while some ants in Clade I were distributed in the subtropical regions (Figure 3.1, Figure 3.8).

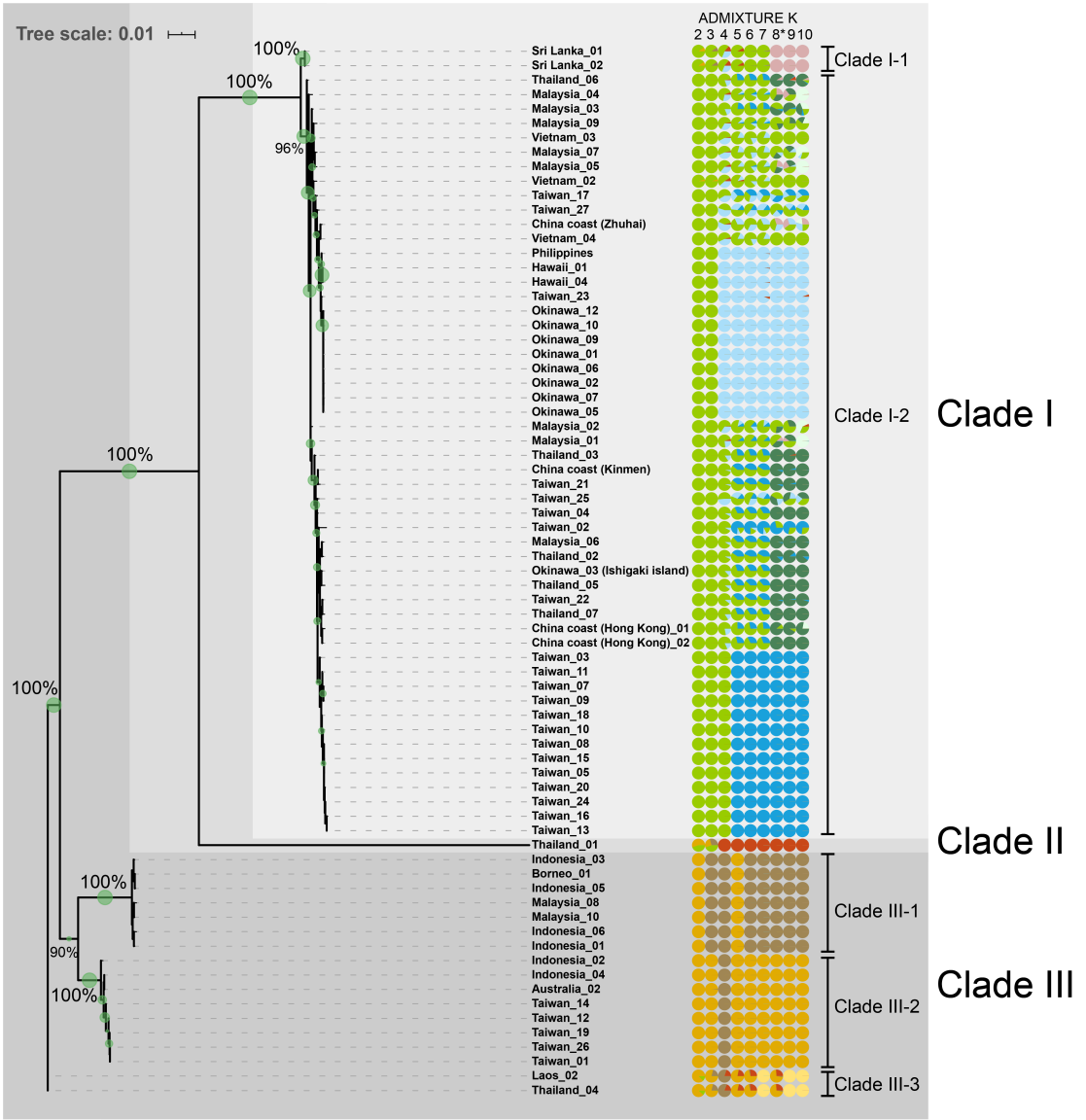


Figure 3.6 Maximum likelihood un-rooted tree and genetic cluster assignment. The SNP tree based on JC model with 11,476 *A. grecilipes* SNP loci implemented in RAxML-NG. Node supported by over 60 percent bootstraps were denoted by green circle. The circle size indicated the relative bootstraps value from 60-100%. The supporting bootstrap value of major clades denoted on nodes in percentage. The probability of genetic group assignment was calculated in ADMIXTURE with 11,357 neutral SNP loci. The probability of genetic group assignment (K=2-10) was indicated in the colored pie chart following individual name.

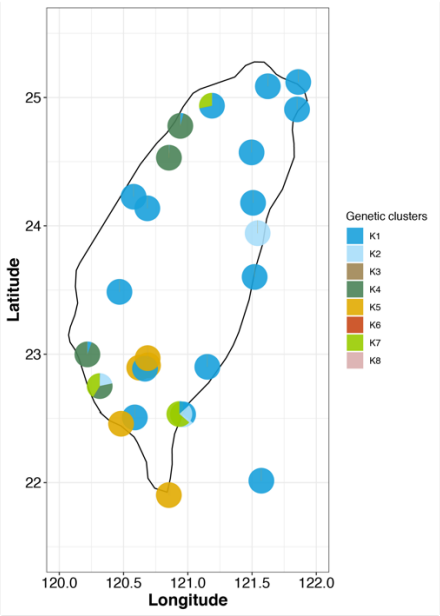


Figure 3.7. Visualizing the geographic distribution of genetic clusters of 26 yellow crazy ant colonies in Taiwan. The center of pie chart indicates the sample location and colors with percentage of ancestral genetic clusters be assigned to each *A. gracilipes* colony.

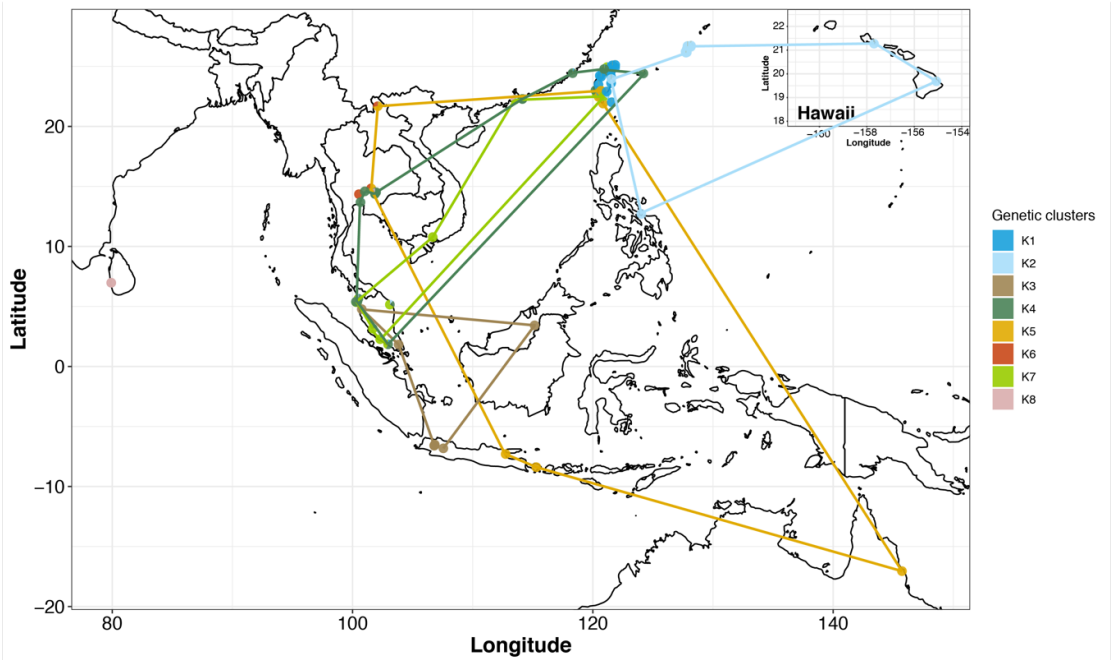


Figure 3.8. Visualizing the geographic distribution of 8 genetic clusters by connecting individuals with the same genetic cluster.

3.3.3 Genetic Barriers

Results of effective migration estimated by the EEMS method confirmed the presence of gene flow corridors and barriers in *A. gracilipes* across our sampled regions

(Figure 3.9). The main migration corridor involved Sri Lanka, Malaysia, Thailand, China and Taiwan, and then extended to Philippines and Okinawa. The strongest genetic barrier resided at the tip of the Malay Peninsula, which represented an geographical separation of the two major phylogenomic clades (Clade I and Clade III). Another major genetic barrier was resided in Java, Indonesia separating *A. gracilipes* in the Eastern part of Java (clade III-2) from the Western part of Java (clade III-1).

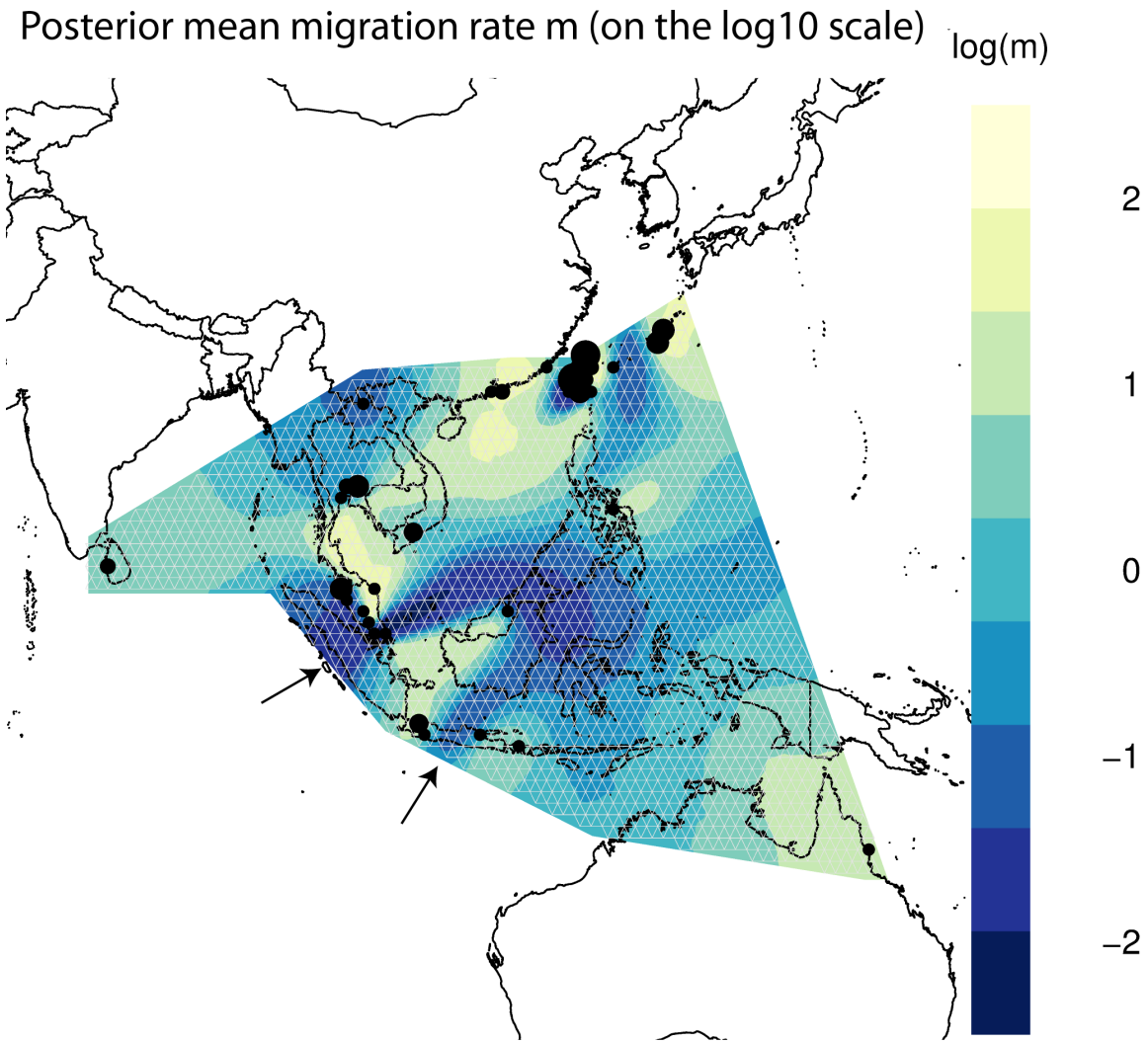


Figure 3.9 Effective migration rate (m) inferred by EEMS for all qualified 11,476 SNP loci. Dark blue represents areas of low migration relative to the average, and light yellow are areas of higher migration. Black dots represent the *A. gracilipes* sample location and the size is the sample size. Arrows indicate genetic barriers.

3.3.4 Outliers and potential under positive selection genes

A total of 133 outliers from the two-genetic-cluster datasets (12,678 SNP loci) were detected by BayeScan ($\log_{10}$(posterior probability >1)), including 123 loci with q-value

< 0.01. The PCA outlier analysis by pcadapt on all 73 individuals with all qualified 11,476 SNPs (dataset 1) pinpointed 974 outliers after Bonferroni correction (q-value < 0.01) (Figure 3.10). I found that 20 loci were inferred as outliers by the two outlier loci detection approaches. Among the 20 loci, three of them were found to be physically linked to a nearby SNP loci (within 300 bp) thus denoted as the same genomic region. I compared the 17 outlier presenting genome contigs against the NCBI nr database, and 12 of which were found to link to functional genes such as dynein heavy chain protein, coatomer subunit protein, calmodulin binding transcription activator, Wnt-11b like protein, WD repeat-containing protein, and transposons (Table 3.3)

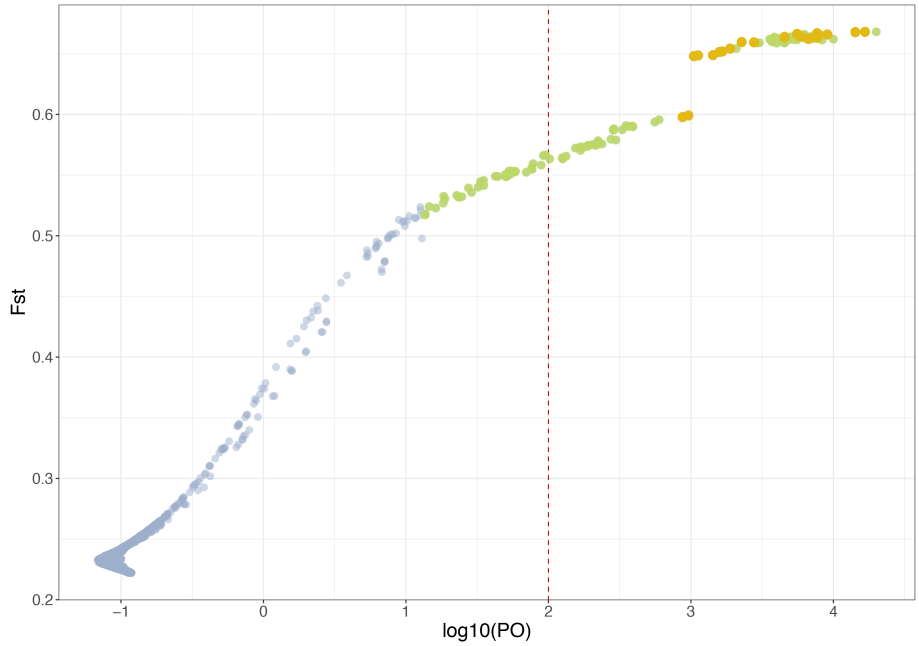


Figure 3.10 Fst-outlier detection by BayeScan on 12,678 SNP loci. The SNP outlier loci identified by both outlier detection approaches were denoted by yellow dots. Those SNP loci with q-value < 0.01 pinpointed by BayeScan was colored by green. The red dash line indicated the posterior probability =0.99.

Table 3.3 Annotation of protein coding genes near 12 outlier located genomic regions

Scaffold ID	SNP Position	Gene description	Function	Top hit species	E-value
scaffold_142	3,490	Dynein heavy chain	Cytoskeleton	*Formica exsecta*	2e-46
scaffold_377	2,750	Coatomer subunit beta	Cytosolic protein	*Camponotus floridanus*	0
scaffold_640	7,202	Hypothetical protein EAG_09986	Unknown	*Camponotus floridanus*	3e-42
scaffold_1041	18,035	Calmodulin-binding transcription activator	Calcium signaling pathway	*Camponotus floridanus*	5e-08
scaffold_1600	13,976	Hypothetical protein EAG_11459	Unknown	*Camponotus floridanus*	1e-12
scaffold_3006	9,678	Wnt-11b-like protein	Wnt pathway	*Formica exsecta*	1e-5
scaffold_6148	6,081; 6,163	Reverse transcriptase	Transposable element	*Lasius niger*	4e-140
scaffold_6613	13,748	Gag-pol polyprotein	Transposable element	*Lasius niger*	0
scaffold_10968	3,151	WD repeat-containing protein	Cilia function	*Camponotus floridanus*	0
Scaffold_11683	3,962; 4,167	Abscission/NoCut checkpoint regulator	Cell cycle	*Camponotus floridanus*	8e-88
scaffold_19074	527	PiggyBac	Transposable element	*Hyposmocoma kahamanoa*	3e-40
scaffold_22321	1,505	Reverse transcriptase	Transposable element	*Wasmannia auropunctata*	1e-35

3.4 Discussion

Unlike many widespread invasive ant species, the invasion history and consequences associated with the colonization of yellow crazy ants received little attention. In this study, I assessed the population genetic structure and patterns of gene flow among multiple geographical populations of *A. gracilipes*, and also attempted to uncovered potential genomic consequences associated with invasion as I identified several potential genomic regions contributing to population divergence between the tropical and subtropical populations of *A. gracilipes*. I also showed no evidence of genomic admixture following secondary contact after population divergence, thus providing new insights into the role of population divergence in shaping species' invasion success.

3.4.1 Distribution of Genetic Clusters and Gene Flow

Our results of inter-colony kinship coefficient indicate *A. gracilipes* collected in this study may have derived from three common families. I further disentangled the demographic history of *A. gracilipes* based on the information of branch length that represented the evolutionary time for variant accumulation. Our phylogenomic analysis revealed that all sampled *A. gracilipes* individuals are separated into three long-branch major clades (Clade I-III), and two subclade Calde III-1 and III-2 with high (>90%) bootstrap support. No geographical concordance is observed for multiple short-branches within the subclades (e.g., short branches in Clade I-1, I-2, III-1, III-2, and III-3). All these data suggest the current distribution of *A. gracilipes* may have been attributable to multiple dispersal waves.

3.4.1.1 Putative origin and ancient dispersals

Our results provide several lines of evidence supporting the continental portion of Southeast Asia as the putative range for the yellow crazy ant. First, Thailand represents the only region that harbors all the three families of *A. gracilipes* (Figure 3.6). The phylogenomic analysis also shows that individuals from Thailand could be found in one of the three major clades, with a single individual (Thailand_01) possessing deep genetically divergence from all other samples. Second, all individuals from Thailand in Clade I-2 are basal to all other subclades that comprise those ant colonies from Pacific islands (Okinawa, Hawaii). Third, the genetic diversity (pi) is highest in Thailand (sample size n > 5). All these data imply that *A. gracilipes* in Thailand are characterized by a complex genetic structure and high genetic diversity. Furthermore, a long branch of the phylogenomic Clade II suggests *A. gracilipes* in Thailand have a relatively ancient population demographic history. Similar patterns of long branch and population differentiation was also found in mitochondrial DNA and *Wolbachia* (*w*Agra) of this ant in an earlier study [27]. The co-cladogenesis of nuclear and mitochondrial genome as well as endosymbiont support that *A. gracilipes* most likely originate from Thailand or other parts of Southeast Asia nearby.

The next strata of ancient dispersal may occur from Malaysia, where the second highest genetic diversity was observed, to Indonesia. Our results from the EEMS analysis identified a genetic barrier at the tip of Malay Peninsula, which separates phylogenomic Clade I and Clade III. Interestingly, a similar divergence pattern also is evident in Asian black-spined toad that reflected the dispersal pathway of toad may from Malay Peninsula to Indonesia [44]. The other genetic barrier is found to be located at Java separating individuals settled in eastern part of Java (Clade III-2) form individuals in western part of Java (Clade III-1). Although such genetic barrier is rarely described in insects, a previous study shows a pattern of sequential divergence in a freshwater fish (*Rasbora lateristerata*) from west to east at the Java island [45]. Besides,

the prehistorical human dispersal patterns reveal several ancient human movements from continental Southeast Asia to Malaysia and Indonesia [46]. The yellow crazy ant is general thought dispersal via human-mediate transportation [47], it is possible *A. gracilipes* occasionally transport with early human colonization. While timing of population expansion may vary across species in these studies, the coincidence of *A. gracilipes* following a comparable population divergence pattern with numerous species in Southeast Asia suggests a strong likelihood of the continental Southeast Asia as the putative origin of this ant.

3.4.1.2 Contemporary invasions

Most *A. gracilipes* samples in our study are assigned to the Clade I with genomic admixture but show no apparent geographical concordance. In Clade I, I found a "Pacific" genetic clade (denoted by light blue in Figure 3.1, 3.6) in which most of the ants occupy in Philippines, Eastern Taiwan, Okinawa, and Hawaii. A most parsimonious explanation for such a widespread distribution of genetically similar populations across these islands is human-mediated dispersal, as have been reported in many invasive ants that frequently travel as a hitch-hiker with human transportation and commerce activities [48].

The finding of a single genetic cluster in both Okinawa and Hawaii implies that single invasion or multiple invasions involving genetically similar founder populations is likely responsible to the current infestation of *A. gracilipes* in these islands. Unlike these two islands, the invasion history of this ant in Taiwan may have been more complex. I found the presence of five genetic clusters (K1, K2, K4, K5, and K7) in Taiwan including: 1) a widespread genetic cluster (dark blue, K1) to which most colonies are assigned to, but no other *A. gracilipes* outside Taiwan, 2) five colonies were assigned to the same genetic cluster (K2) with *A. gracilipes* collected from China, Vietnam, Thailand, and Malaysia 3) a "Pacific" genetic cluster (K4) with those samples collected from Eastern Taiwan. 4) Four individuals have a pattern of genomic admixture that could not be consistently assigned to any single genetic cluster. 5) The genetic cluster (K5) in phylogenomic Clade III-2 is found in southern Taiwan, indicating the occurrence of multiple invasions by *A. gracilipes* to Taiwan. Multiple invasions and admixture of genetically distinct populations have been shown to compensate potential drawbacks associated with loss of genetic diversity [9], thus largely contributing to the persistence and possibly long-term success of an invasive population; Although individuals from Clade I and III coexist in some areas sympatrically (e.g., distributions of individuals assigned to genetic cluster K1 and K5 overlap, no evidence of genomic admixture of individuals between the two clusters is found (Figure 3.3 and Figure 3.6). This suggests genetic isolation of *A. gracilipes* clades, which is consistent to another introduced population in Christmas Island where ants from different clades remain genetically distinct even after secondary contact [16]. One

potential reason for such a genetic isolation is that intraspecific aggression behavior towards individuals belonging to different supercolonies as a result of nestmate recognition precludes inter-supercolony exchange of nestmates and thus gene flow. Indeed, a high level of aggression towards workers and reproductives (e.g., virgin queens and males) from different supercolonies was reported in this ant [14,15]. These studies also reported a positive correlation of aggression and genetic divergence between different supercolonies, reinforcing the possibility of both genetic divergence and aggressive interactions disrupting the gene flow between supercolonies [14,15]. As the different genetic clusters mirror the presence of genetic divergence, absence of gene flow between the two divergent genetic clusters in sympatric areas suggests speciation may be expected as a major genetic consequence.

3.4.2 Positive Selection Genes

One of potential mechanisms for long-term persistence of an invasive species is local adaptation. Our population genomic analysis revealed that *A. gracilipes* tend to have tropical (phylogenomic Clade II, III) and subtropical clusters (phylogenomic Clade I). Coincidentally, one of protein coding genes that are physically linked to the identified SNP outliers is associated with Ca2+ signaling pathway [49] in which it encodes calmodulin-binding transcription activator (CAMTA). The CAMTA proteins are transcription factor involved in light response [50], courtship [51], and potentially associated with insecticide resistance [52] in drosophila. In plants, the CAMTA protein family is involved in responses to pathogen defense [53,54] and temperature tolerance [55]. While the CAMTA protein family in ants remains poorly described, all known functions of the proteins to date are highly associated with organismal responses to dynamic environmental factors. The outlier SNP locus near the CAMTA gene thus indicates the potential role of this gene in local adaptation during the colonization of yellow crazy ant into environments with different climate regimes.

Several genes adjacent to the identified SNP outlier loci in *A. gracilipes* are reported essential in other taxa. For example, the dyneins are a class of motor protein involved in ciliary and flagellar motility, organelle transport, and chromosome segregation of eukaryotic cells [56]. While it may be challenging to identity the role of these genes in the process of local adaptations in *A. gracilipes* owing to their essential role in the cell function, I notice the dynein heavy chain also was identified as an outlier when comparing introduced and native populations of another highly invasive ant, *Solenopsis invicta* [5]. This therefore opens a new avenue to research into the potential of function of this gene, especially in the context of success invasion in social insect.

Finally, transposable elements are known to play a crucial role in adaptive evolution and can be activated by both biotic and abiotic stress [9]. Our results revealed the presence of several transposable element-related genes in the candidate

genome regions that potentially are under positive selection. The movement of transposable elements may create *de novo* mutations that are believed to potentially contribute to rapid adaptation [57]. In ants, transposable elements ae found to facilitate generation of genomic diversity [58], and may promote adaptation of invasive ants to the novel environment [59]. Although evidence for such a mechanism is not yet clearly demonstrated in *A. gracilipes*, transposable elements have higher insertion frequencies in the invasive Asian tiger mosquito compared to those in native range [60]. Our result enhances the confidence of transposable element may associate with local adaptation in invasive species.

In present study, I showed that the current distribution of *A. gracilipes* most likely derived from multiple waves of population dispersal that originated from three common families. One of evolutionary consequence of the observed population divergence and limit gene flow of *A. gracilipes* between sympatric families (clusters) suggest this species may be a first step of speciation. I further provide information on candidate genome regions with genes under positive selection and future research effort should be focused on characterize how these genes potentially contribute to adaptive evolution of the yellow crazy ant.

3.5 References

1. Colautti, R.I.; Lau, J.A. Contemporary evolution during invasion: evidence for differentiation, natural selection, and local adaptation. *Mol. Ecol.* **2015**, *24*, 1999-2017.
2. Bertelsmeier, C.; Ollier, S.; Liebhold, A.; Keller, L. Recent human history governs global ant invasion dynamics. *Nat. Ecol. Evol.* **2017**, *1*, 0184.
3. Abbott, K.L. Supercolonies of the invasive yellow crazy ant, Anoplolepis gracilipes, on an oceanic island: forager activity patterns, density and biomass. *Insectes Sociaux* **2005**, *52*, 266-273.
4. Thomas, M.L.; Tsutsui, N.D.; Holway, D.A. Intraspecific competition influences the symmetry and intensity of aggression in the Argentine ant. *Behav. Ecol.* **2005**, *16*, 472-481.
5. Privman, E.; Cohen, P.; Cohanim, A.B.; Riba-Grognuz, O.; Shoemaker, D.; Keller, L. Positive selection on sociobiological traits in invasive fire ants. *Mol. Ecol.* **2018**, *27*, 3116-3130.
6. Yang, C.C.; Yu, Y.C.; Valles, S.M.; Oi, D.H.; Chen, Y.C.; Shoemaker, D.; Wu, W.J.; Shih, C.J. Loss of microbial (pathogen) infections associated with recent invasions of the red imported fire ant Solenopsis invicta. *Biol. Invasions* **2010**, *12*, 3307-3318.
7. Rey, O.; Estoup, A.; Vonshak, M.; Loiseau, A.; Blanchet, S.; Calcaterra, L.; Chifflet, L.; Rossi, J.P.; Kergoat, G.J.; Foucaud, J.; Orivel, J. Where do adaptive shifts occur during invasion? A multidisciplinary approach to unravelling cold adaptation in a tropical ant species invading the Mediterranean area. *Ecol. Lett.* **2012**, *15*, 1266-1275.

8. Foucaud, J.; Rey, O.; Robert, S.; Crespin, L.; Orivel, J.; Facon, B.; Loiseau, A.; Jourdan, H.; Kenne, M.; Masse, P.S.M.; Tindo, M. Thermotolerance adaptation to human-modified habitats occurs in the native range of the invasive ant Wasmannia auropunctata before long-distance dispersal. *Evol. Appl.* **2013**, *6*, 721-734.
9. Estoup, A.; Ravigné, V.; Hufbauer, R.; Vitalis, R.; Gautier, M.; Facon, B. Is there a genetic paradox of biological invasion? *Annu. Rev. Ecol. Evol. Syst.* **2016**, *47*, 51-72.
10. Wetterer, J.K. Worldwide distribution and potential spread of the long-legged ant, *Anoplolepis gracilipes* (Hymenoptera: Formicidae). *Sociobiology* **2005**, *45*, 77-97.
11. Ito, F.; Asfiya, W.; Kojima, J.I. Discovery of independent-founding solitary queens in the yellow crazy ant *Anoplolepis gracilipes* in East Java, Indonesia (Hymenoptera: Formicidae). *Entomol. Sci.* **2016**, *19*, 312-314.
12. O'Dowd, D.J.; Green, P.T.; Lake, P.S. Invasional 'meltdown'on an oceanic island. *Ecology Lett.* **2003**, *6*, 812-817.
13. Cooling, M.; Hoffmann, B.D. Here today, gone tomorrow: declines and local extinctions of invasive ant populations in the absence of intervention. *Biol. Invasions* **2015**, *17*, 3351-3357.
14. Drescher, J.; Blüthgen, N.; Feldhaar, H. Population structure and intraspecific aggression in the invasive ant species Anoplolepis gracilipes in Malaysian Borneo. *Mol. Ecol.* **2007**, *16*, 1453-1465.
15. Drescher, J.; Blüthgen, N.; Schmitt, T.; Bühler, J.; Feldhaar, H. Societies drifting apart? Behavioural, genetic and chemical differentiation between supercolonies in the yellow crazy ant *Anoplolepis gracilipes*. *PLoS ONE* **2010**, *5*, e13581.
16. Thomas, M.L.; Becker, K.; Abbott, K.; Feldhaar, H. Supercolony mosaics: two different invasions by the yellow crazy ant, Anoplolepis gracilipes, on Christmas Island, Indian Ocean. *Biol. Invasions* **2010**, *12*, 677-687.
17. Gruber, M.A.; Hoffmann, B.D.; Ritchie, P.A.; Lester, P.J. Recent behavioural and population genetic divergence of an invasive ant in a novel environment. *Divers. Distrib.* **2012**, *18*, 323-333.
18. Peterson, B.K.; Weber, J.N.; Kay, E.H.; Fisher, H.S.; Hoekstra, H.E. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE* **2012**, *7*, e37135.
19. FastQC. Available online: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed on 26 08 2020).
20. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **2011**, *17*, 10-12.
21. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics* **2009**, *25*,1754-60.
22. Rochette, N.C.; Rivera-Colón, A.G.; Catchen, J.M. Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Mol. Ecol.* **2019**, *28*, 4737-4754.
23. Rochette, N.C.; Catchen, J.M. Deriving genotypes from RAD-seq short-read data using Stacks. *Nat. Protoc.* **2017**, *12*, 2640-2659.
24. O'Leary, S.J.; Puritz, J.B.; Willis, S.C.; Hollenbeck, C.M.; Portnoy, D.S. These aren't the loci you'e looking for: Principles of effective SNP filtering for molecular ecologists. *Mol. Ecol.* **2018**, *27*, 3193-3206.

25. Danecek, P.; Auton, A.; Abecasis, G.; Albers, C.A.; Banks, E.; DePristo, M.A.; Handsaker, R.E.; Lunter, G.; Marth, G.T.; Sherry, S.T.; McVean, G. The variant call format and VCFtools. *Bioinformatics* **2011**, *27*, 2156-2158.
26. Sebastien, A.; Gruber, M.A.M.; Lester, P.J. Prevalence and genetic diversity of three bacterial endosymbionts (*Wolbachia*, *Arsenophonus*, and *Rhizobiales*) associated with the invasive yellow crazy ant (*Anoplolepis gracilipes*). *Insectes Soc.* **2012**, *59*, 33-40.
27. Lee, C.C.; Lin C.Y.; Tseng S.P; Matsuura K.; Yang C.C.S. Ongoing co-evolution of Wolbachia and a widespread invasive ant, Anoplolepis gracilipes. *Microorganisms* **2020**, *8*, 1569.
28. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114-2120.
29. Peng, Y.; Leung, H.C.; Yiu, S.M.; Chin, F.Y. IDBA-UD: a *de novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **2012**, *28*, 1420-1428.
30. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T.L. BLAST+: architecture and applications. *BMC Bioinformatics* **2009**, *10*, 421.
31. Seppey, M.; Manni, M.; Zdobnov, E.M. BUSCO: assessing genome assembly and annotation completeness. In: *Gene Prediction Methods in Molecular Biology*; Kollmar M., Eds.; Humana: New York, NY, 2019; Volume 1962, pp. 227-245.
32. Gurevich, A.; Saveliev, V.; Vyahhi, N.; Tesler, G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **2013**, *29*, 1072-1075.
33. Holderegger, R.; Kamm, U.; Gugerli, F. Adaptive vs. neutral genetic diversity: implications for landscape genetics. *Landsc. Ecol*. **2006**, *21*, 797-807.
34. Manichaikul, A.; Mychaleckyj, J.C.; Rich, S.S.; Daly, K.; Sale, M.; Chen, W.M. Robust relationship inference in genome-wide association studies. *Bioinformatics* **2010**, *26*, 2867-2873.
35. Jombart, T. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **2008**, *24*, 1403-1405.
36. Alexander, D.H.; Novembre, J.; Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **2009**, *19*, 1655-1664.
37. Kozlov, A.M.; Darriba, D.; Flouri, T.; Morel, B.; Stamatakis, A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **2019**, *35*, 4453-4455.
38. Pattengale, N.D.; Alipour, M.; Bininda-Emonds, O.R.; Moret, B.M.; Stamatakis, A. How many bootstrap replicates are necessary? *J. Comput. Biol.* **2010**, *17*, 337-354.
39. Letunic I.; Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res*. **2019**, *47*, W256-W259.
40. Petkova, D.; Novembre, J.; Stephens, M. Visualizing spatial population structure with estimated effective migration surfaces. *Nat. Genet.* **2016**, *48*, 94-100.
41. Foll, M.; Gaggiotti, O. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* **2008**,*180*, 977-993.
42. Lotterhos, K.E.; Whitlock, M.C. Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. *Mol. Ecol*. **2014**. *23*, 2178-2192.

43. Luu, K.; Bazin, E.; Blum, M.G. pcadapt: an R package to perform genome scans for selection based on principal component analysis. *Mol. Ecol. Resour*. **2017**, *17*, 67-77.
44. Othman, S.N.; Chen, Y.H.; Chuang, M.F.; Andersen, D.; Jang, Y.; Borzée, A. Impact of the Mid-Pleistocene Revolution and Anthropogenic Factors on the Dispersion of Asian Black-Spined Toads (Duttaphrynus melanostictus). *Animals* **2020**, *10*, 1157.
45. Kusuma, W.E.; Ratmuangkhwang, S.; Kumazawa, Y. Molecular phylogeny and historical biogeography of the Indonesian freshwater fish Rasbora lateristriata species complex (Actinopterygii: Cyprinidae): cryptic species and west-to-east divergences. M*ol. Phylogenetics Evol.* **2016**, *105*, 212-223.
46. McColl, H.; Racimo, F.; Vinner, L.; Demeter, F.; Gakuhari, T.; Moreno-Mayar, J.V.; Van Driem, G.; Wilken, U.G.; Seguin-Orlando, A.; De la Fuente Castro, C.; Wasef, S. The prehistoric peopling of Southeast Asia. *Science* **2018**, *361*, 88-92.
47. Holway, D.A.; Lach, L.; Suarez, A.V.; Tsutsui, N.D.; Case, T.J. The causes and consequences of ant invasions. *Annu. Rev. Ecol. Syst.* **2002**, *33*, 181-233.
48. Ascunce, M.S.; Yang, C.C.; Oakey, J.; Calcaterra, L.; Wu, W.J.; Shih, C.J.; Goudet, J.; Ross, K.G.; Shoemaker, D. Global invasion history of the fire ant *Solenopsis invicta*. *Science* **2011**, *331*, 1066-1068.
49. Finkler, A.; Ashery-Padan, R.; Fromm, H. CAMTAs: calmodulin-binding transcription activators from plants to human. *FEBS Lett.* **2007**, *581*, 3893-3898.
50. Han, J.; Gong, P.; Reddig, K.; Mitra, M.; Guo, P.; Li, H.S. The fly CAMTA transcription factor potentiates deactivation of rhodopsin, a G protein-coupled light receptor. *Cell* **2006**, *127*, 847-858.
51. Sato, K.; Ahsan, M.T.; Ote, M.; Koganezawa, M.; Yamamoto, D. Calmodulin-binding transcription factor shapes the male courtship song in Drosophila. *PLoS Genet*. **2019**, *15*, e1008309.
52. Huang, J.; Sun, W.; Seong, K.M.; Mittapalli, O.; Ojo, J.; Coates, B.; Paige, K.N.; Clark, J.M.; Pittendrigh, B.R. Dietary antioxidant vitamin C influences the evolutionary path of insecticide resistance in Drosophila melanogaster. *Pestic. Biochem. Phys*. **2020**, *168*, 104631.
53. Galon, Y.; Nave, R.; Boyce, J.M.; Nachmias, D.; Knight, M.R.; Fromm, H. Calmodulin-binding transcription activator (CAMTA) 3 mediates biotic defense responses in Arabidopsis. *FEBS Lett*. **2008**, *582*, 943-948.
54. Du, L.; Ali, G.S.; Simons, K.A.; Hou, J.; Yang, T.; Reddy, A.S.N.; Poovaiah, B.W. Ca 2+/calmodulin regulates salicylic-acid-mediated plant immunity. *Nature* **2009**, *457*, 1154-1158.
55. Kim, Y.; Park, S.; Gilmour, S.J.; Thomashow, M.F. Roles of CAMTA transcription factors and salicylic acid in configuring the low-temperature transcriptome and freezing tolerance of A rabidopsis. Plant J. **2013**, 75, 364-376.
56. Mazumdar, M.; Mikami, A.; Gee, M.A.; Vallee, R.B. In vitro motility from recombinant dynein heavy chain. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 6552-6556.
57. Stapley, J.; Santure, A.W.; Dennis, S.R. Transposable elements as agents of rapid adaptation may explain the genetic paradox of invasive species. *Mol. Ecol.* **2015**, *24*, 2241-2252.

58. Lee, C.C.; Wang, J. Rapid expansion of a highly germline-expressed mariner element acquired by horizontal transfer in the fire ant genome. *Genome Biol. Evol.* **2018**, *10*, 3262-3278.
59. Schrader, L.; Kim, J.W.; Ence, D.; Zimin, A.; Klein, A.; Wyschetzki, K.; Weichselgartner, T.; Kemena, C.; Stökl, J.; Schultner, E.; Wurm, Y. Transposable element islands facilitate adaptation to novel environments in an invasive species. *Nat. commun.* **2014**, *5*, 1-10.
60. Goubert, C.; Henri, H.; Minard, G.; Valiente Moro, C.; Mavingui, P.; Vieira, C.; Boulesteix, M. High-throughput sequencing of transposable element insertions suggests adaptive evolution of the invasive Asian tiger mosquito towards temperate environments. *Mol. Ecol.* **2017**, *26*, 3968-3981.

Chapter 4: Coevolution of *Wolbachia* and the yellow crazy ant

4.1 Introduction

Wolbachia is an endosymbiont commonly found in arthropods and filarial nematodes [1]. This bacterium could manipulate the reproduction of its arthropod hosts to benefit itself as a reproductive parasite. The reproductive manipulations include cytoplasmic incompatibility, male killing, feminization, and induction of parthenogenesis [1,2]. Evidence has been increasingly accumulated recently to suggest that *Wolbachia* confers benefits to their hosts and acts as a mutualistic partner that is involved in either nutritional provisioning or host pathogen defense [2]. In more extreme cases *Wolbachia* can be an obligate mutualist with fitness costs otherwise being significant when *Wolbachia* is removed/cured from the hosts [2,3]. Such intimate interaction between the symbiont and host would lead to coevolution and high infection rate of *Wolbachia* (e.g., 90–100%) within the host population. For example, the coevolutionary pattern (e.g., co-divergence or co-speciation) is commonly observed in filarial nematodes [4] and bedbugs [5] as *Wolbachia* functions as a nutritional mutualist [5,6] and often persists within and among species at 100% prevalence [5,7]. While infections up to 100% have been occasionally reported in *Wolbachia* acting as a reproductive parasite in arthropods [8,9], a high level of *Wolbachia* prevalence within arthropod populations seems uncommon [10,11]. The role of *Wolbachia* in ants, however, is relatively under-described yet likely varies across species. Most studies have found limited evidence supporting the existence of *Wolbachia*-induced cytoplasmic incompatibility, parthenogenesis, and feminization as viable reproductive manipulation in ant [12] (but see [13] for cytoplasmic incompatibility).

The yellow crazy ant, *Anoplolepis gracilipes*, is a widespread invasive ant that has posed significant threats to local biodiversity and ecosystem sustainability in most of its introduced range [14]. This ant is polygynous (nest headed by multiple queens), polydomous (a colony inhabiting in multiple spatially separated yet socially connected nests), and often forms supercolonies (a supercolony comprises of physically separated nests that are mutually tolerated to each other). Like many other invasive ant species, independent foundation (e.g., nuptial flight) is rare in *A. gracilipes*, and colony reproduction is primarily through dependent foundation (e.g., budding) [15,16]. The limited female dispersal ability and potential intranidal mating have been shown to restrict gene flow among different nests of *A. gracilipes* even within the same supercolony and thus result in within-supercolony divergence [17]. Consequently, these life history traits may also facilitate divergence of *Wolbachia* in the ant, especially given the fact that this endosymbiont is maternally transmitted.

Both vertical and horizontal transmissions were reported in *Wolbachia* [1]. Previous studies and our preliminary data have shown that *A. gracilipes* collected from several Indo-Pacific islands and Australia including Christmas Island share identical *Wolbachia wsp* genotype [18,19], and that the bacterium seems to persist at a high prevalence across these populations [18,19]. Coupled with a recent study that revealed no evidence of horizontal transfer of *Wolbachia* between *A. gracilipes* and its closely associated kleptoparasitic ant crickets [18], it is likely that *Wolbachia* may have been primarily transferred vertically in *A. gracilipes*, thus providing an excellent opportunity to explore *Wolbachia*–host coevolutionary history.

In the present study, I conducted survey of *Wolbachia* prevalence and infection status (single or multiple strains infections) in *A. gracilipes* at a global scale (a total of 12 geographic regions including Southeast and East Asia, Sri Lanka, several Pacific islands, and Australia). I also assessed and compared phylogenies of *Wolbachia* (based on genomic single nucleotide polymorphisms (SNPs)) and partial sequence of mitochondrial DNA of *A. gracilipes* to examine whether the coevolution between *Wolbachia* and the ant holds true. Lastly, I tested whether there is signature of natural selection in *Wolbachia* to reflect potential ant–*Wolbachia* coevolutionary history.

4.2 Materials and methods

4.2.1 Sample Collection, DNA Extraction and Molecular Genetic Assays

Wolbachia prevalence was assessed based on 80 yellow crazy ant colonies from a total of 12 geographical regions across Southeast and East Asia, Sri Lanka, several Pacific islands, and Australia collected from 2012 to 2019 (Figure 4.1). The collected specimens were preserved in 95% ethanol upon DNA extraction. To examine *Wolbachia* infection and identify the respective *Wolbachia* strain, three random adult worker ants per ant colony were selected. The whole genomic DNA was purified individually from the whole ant body using QIAamp DNA Mini Kit (Qiagen, Valencia, CA, USA) following the manufacture's instruction. The multi-locus sequence typing system (MLST; *hcpA*, *ftsZ*, *gatB*, *coxA*, and *fbpA*) [20] and partial *Wolbachia* surface protein gene (*wsp*) were amplified using the standard polymerase chain reaction (PCR) with inclusion of proper positive control and blank (ddH_2O) in every batch of PCR reaction. Colonies were only considered to be *Wolbachia*-infected if at least one of the three assayed worker ants displayed an amplificon of the *wsp* gene with the expected size. The MLST loci were amplified following the protocols described in the PubMLST [21], whereas the *wsp* gene was amplified using PCR conditions reported in [22]. Both PCR reactions were performed using EmeraldAmp® MAX PCR Master Mix (Takara, Shiga, Japan). Only one of the three randomly selected ant workers from each *Wolbachia*-positive colony was subject to Sanger sequencing and further genetic analyses (e.g., determination of mitochondrial haplotype and ddRAD sequencing, see

below for more details). I assigned sequence type (ST; a unique series of alleles) for alleles of each of the MLST loci based on comparison against those deposited in the MLST database. The sequencing electropherogram of each gene was manually checked by naked eyes to identify potential infection by multiple strains of *Wolbachia*.

Host mitochondrial cytochrome C oxidase subunit I (*COI*) haplotype was identified from the same *A. gracilipes* workers subjected to Sanger sequencing for MLST loci and the *wsp* gene. Partial mitochondrial *COI* gene was amplified with primers designed based on the *A. gracilipes* mitochondrial genome [23] using software Primer3-Plus [24]. The amplification was performed using EmeraldAmp® MAX PCR Master Mix (Takara, Shiga, Japan). PCR primers and conditions were listed in Table S2. The amplicons were purified and then sent to Sanger sequencing.

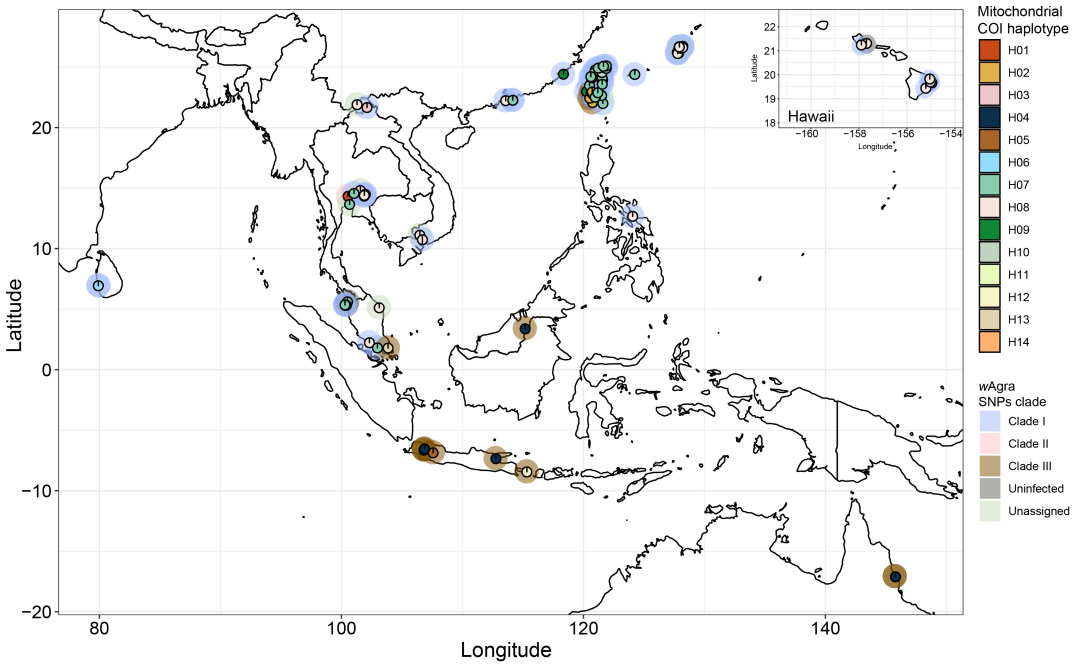


Figure 4.1 Geographic distribution, mitochondrial COI haplotype (H01-H14, inner circle) and wAgra SNPs clade (outer circle) of all yellow crazy ant colonies in this study.

4.2.2 *Wolbachia* Draft Genome and ddRAD-seq

4.2.2.1 *de novo* Assembly of *Wolbachia* Draft Genome

The MLST loci are generally conserved and may possess insufficient resolution for reconstructing recent evolutionary history [25,26]. A *Wolbachia* draft genome was therefore assembled as a reference for uncovering more genetic variations. A female alate of *A. gracilipes* was collected from Taiwan (Hsinchu City, 24°46′43.43″ N,

120°56′30.24″ E) in 2014, and its high molecular weight DNA (a whole-body extraction) was purified via the standard phenol–chloroform purification method. The whole genomic DNA was sequenced by Sequencing Technology Company (Taipei, Taiwan) on the Illumina HiSeq 2500 platform with PCR-free library construction (average library insert size: 250 bp; paired-end read length: 2 × 125 bp). Illumina adapters, low-quality bases (quality score <28; base call accuracy <99.8%), and short reads (<36 bp after trimming) were trimmed using Trimmomatic v. 0.36 [27]. The whole genome de novo assembly was conducted using IDBA-UD [28] with the minimum 10 reads supported (--min_support 10; other parameters: default). The *Wolbachia* sequences were then separated from host genome via mega-BLASTn [29] against two complete supergroup A *Wolbachia* genomes [*w*Ri, host: *Drosophila simulans* Riverside (GenBank accession: GCA_000022285.1); *w*Mel, host: *Drosophila melanogaster* (GenBank accession: GCF_000008025.1)]. The contigs with e-value lower than 1e-20 were kept as our *Wolbachia* draft genome (hereinafter referred to as *w*Agra) except those with length <250 bp which were excluded from the analysis. The completeness of genome assembly was evaluated using the BUSCO v. 4.0.5 [30], which measures the proportion of highly conserved, single copy orthologs derived from 2,327 proteobacterial species (proteobacteria_odb10, 219 BUSCO orthologs). The BUSCO analysis was also conducted on other supergroup A *Wolbachia* genomes (*w*Mel, *w*Ri, *w*Ha, and *w*Au) for comparison purposes. The assembly statistics were calculated using QUAST [31].

4.2.2.2 ddRAD-seq and *Wolbachia* SNP Filtering

The restriction site associated DNA sequencing (RAD-seq) has been demonstrated as a powerful population genomic approach for recovering high number of single nucleotide polymorphisms (SNPs) in various systems, especially for non-model organisms [32]. I conducted double digest RAD-seq (ddRAD-seq) on the same individuals mentioned above following the protocol described in Peterson et al. [33] with slight modifications. The restriction enzymes HpyCH4IV (New England BioLabs, Ipswich, MA, USA) and BfaI (New England BioLabs, Ipswich, MA, USA) were used for DNA digestion. In brief, 50–100ng DNA digested with enzymes for 3 h and ligated to 5 bp inline barcodes and partial Illumina adapter at both ends with T4 DNA ligase (New England BioLabs, Ipswich, MA, USA) at 23 °C for 30 min. Ligated DNA was mixed into a DNA pool at one to one ratio. The DNA mixture was purified and concentrated using AMPure XP magnetic beads (Beckman Coulter, Pasadena, CA, USA). I then performed size selection (550–700 bp) on the condensed DNA in 2% agarose gel (Invitrogen, Carlsbad, CA, USA) and purified the selected DNA fragments using the FastGene gel/PCR extraction kit (Nippon genetics, Tokyo, Japan). The complete Illumina adapter and index were added through 12 cycles of two-step PCR enrichment (98 °C for 30 s, 12 cycles of 98 °C for 10 s and 72 °C for 30 s, with the final extension at 72 °C for 10 min, and hold at 4 °C) with Phusion high-fidelity PCR kit

(New England BioLabs, Ipswich, MA, USA). The PCR library was purified using AMPure XP magnetic beads (Beckman Coulter, Pasadena, CA, USA). The final library was mixed with 10% PhiX and sequenced in the Illumina HiSeqX platform in pair-end (2 × 150) mode by Macrogen Japan (Kyoto, Japan).

Quality of generated raw reads was first examined using FastQC v. 0.11.8 [34]. To keep read quality at the same read length for SNP calling, reads containing 3′ adapters were discarded. Reads were trimmed by 1 bp on their 5′ end for both R1 and R2 reads, and by 30 bp from their 3′ end of R2 reads. Trimmed reads then were demultiplexed using 5′ anchored primer method without allowance for error (-e 0.01, --no-indels) performed in Cutadapt v. 2.8 [35]. Variant calling was conducted using a reference-based method. Reads were mapped to the *w*Agra genome using BWA v. 0.7.17 [36], and variants were called using gstacks implemented in Stacks2 v. 2.5 [37]. To minimize false positive variant call, mapped SNP loci without proper pairs were discarded (--rm-unpaired-reads). SNP loci present in more than 80% of all the samples (-R 0.8) were kept using populations program in Stacks2 v. 2.5 [37]. Employing VCFtools v. 0.1.15 [38], the SNPs with read depth less than 3 (--minDP 3) or over three times the standard deviation of SNP read depth (--maxDP 90) were identified and thus excluded from further analyses.

To verify if the SNPs I recovered were actually located in the *w*Agra genome instead of the host genome (e.g., horizontal gene transfer from *Wolbachia* to host [39]), I selected 10 polymorphic SNP loci and conducted locus-specific PCR amplification on both *Wolbachia*-infected and *Wolbachia*-free colonies. Primers were designed using Primer3-Plus [24], and the PCR conditions were accessible in the Table S2.

4.2.3. Phylogenomic Tree Reconstruction and Mitochondrial Network Analysis

The concatenated SNPs were generated by populations program in Stacks2 v. 2.5 [37]. The SNPs were located in various genomic regions and might have diverse evolutionary models. The Jukes–Cantor model was therefore used in the analysis. The maximum likelihood phylogenetic tree was reconstructed using RAxML-NG v. 0.9.0 [40] with 1,000 bootstrap replications. The mitochondrial *COI* sequences were aligned in MEGA 7 [41]. The haplotypes and network analysis were conducted in POPART [42] using the median-joining network analysis.

4.2.4. Wolbachia Transcriptome Analysis

The genomic regions containing polymorphic SNPs in *w*Agra were manually examined using IGV v. 2.3.71 [43]. The SNP-containing open reading frames (ORF) were annotated using NCBI ORFfinder, and potential gene function was assigned using BLASTp query against NCBI non-redundant protein sequences (nr) database.

The synonymous and non-synonymous mutations were manually examined in each ORF.

To verify if the non-synonymous SNP containing *Wolbachia* ORFs were expressed in *w*Agra, the transcriptome sequencing by RNA-Seq on *w*Agra was carried out. The RNA pool-seq was conducted on a mixture of total RNA from two adult worker ants of *A. gracilipes* collected in 2018. One adult worker ant was collected from Penang, Malaysia (5°21′32.4″N, 100°18′09.0″E), whereas the other was from Okinawa, Japan (26°40′19.2″N, 128°00′41.0″E). The entire worker ants were soaked in TRIzol™ RNA Extraction Reagent (Invitrogen, Carlsbad, CA, USA) and homogenized with a pestle, with the rest of the procedure following the standard TRIzol RNA extraction protocol. The RNA quality and quantity were measured with Nanodrop spectrophotometers, and then mixed at one to one ratio of the total RNA amount. The mixed sample was submitted to the Eurofins Genomics (Tokyo, Japan) where the total RNA was purified using polyA selection method, and sequencing library was constructed with the TruSeq® Stranded mRNA Library Prep (insert size: 200 bp). The library was sequenced using Illumina HiSeq 4000 platform in the pair-end mode (2 × 100 bp). Read adapters and low-quality bases (base call accuracy <99.8%; q <28) were trimmed using Trimmomatic v. 0.36 [27] and mapped onto the *w*Agra genome used HISAT2 v2.1.0 [44]. The alignment was sorted in SAMtools v. 1.9 [45] and visualized using IGV v. 2.3.71 [43]. The RNA reads in genomic regions of MLST and *wsp* gene were examined for confirming expression level of the selected *Wolbachia* genes (see Results 3.4).

4.2.5 Test for Signature of Selection

To detect the signal of selection, two *Wolbachia* genes with non-synonymous mutations (a hypothetical protein, HP_12890 and *RluA*-like gene, see Section 3.4 for more details) were amplified and sequenced for all *Wolbachia* infected colonies. Sequences were aligned for identification of haplotype identity, and the pairwise codon-based Z-test was performed in MEGA7 for the detection of selection [41].

4.3 Results

4.3.1. Prevalence and Strain Identity of Wolbachia

I screened the presence of *Wolbachia* in a total of 80 *A. gracilipes* colonies from 12 geographical regions and revealed that the infection rates ranged from 77.78% (*n* = 7/9; Malaysia), 83.33% (*n* = 5/6; Hawaii), to 100% (other geographic regions; Table 4.1), resulting in a 96.25% overall infection rate (*n* = 77/80). *Wolbachia* detected in our *A. gracilipes* samples shared identical sequence at both MLST loci (ST52; *gatB* 22, *coxA* 2, *hcpA* 51, *ftsZ* 32, and *fbpA* 36) and the *wsp* gene (*wsp* 55, HVR1 39, HVR2 1, HVR3 44,

and HVR4 40). All the sequences indicated that this *Wolbachia* strain belongs to supergroup A. I found no evidence for co-infection with more than one *Wolbachia*-variant (strain) in all infected individuals, suggesting single *Wolbachia* infection in this ant.

Table 4.1 *Wolbachia* prevalence in *Anoplolepis gracilipes* across all sampled regions.

Collection region	Infected colony	Uninfected colony	Infection rate
Okinawa	12	0	100.00%
Hawaii	5	1	83.33%
Taiwan	26	0	100.00%
Southeast coastal China (Hong Kong and Zhuhai) and Kinmen Island, Taiwan	4	0	100.00%
Laos and Southwest China (Yunnan)	2	0	100.00%
Vietnam	3	0	100.00%
Thailand	7	0	100.00%
Sri Lanka	2	0	100.00%
Philippines	1	0	100.00%
Malaysia	7	2	77.78%
Indonesia	6	0	100.00%
Australia	2	0	100.00%
Total	77	3	96.25%

4.3.2. *Wolbachia* Genome and SNP discovery

I assembled the draft genome of *Wolbachia* in *A. gracilipes* (*w*Agra) as the reference genome (GenBank accession: JACRYZ000000000). The draft genome assembly (96 contigs) contains 1,202,684 bp, a comparable size to other supergroup A *Wolbachia* genomes (genome size: 1.26–1.8 Mb) and within the range of other *Wolbachia* belonging to other supergroups discovered to date (supergroup B-F; 0.96–1.8 Mb). Similar to other supergroup A *Wolbachia*, GC content in the *w*Agra assembly is 35.23% (35.2–35.3%); it, however, differed from the rest of contigs with the majority of which deriving from the host ant genome (GC: 33.78%). The average genome coverage was 278.22 reads per nucleotide.

The BUSCO score of the *w*Agra genome was 85.4%, as 187 out of 219 complete and single copy proteobacteria orthologs were found in our assembly. The same BUSCO analysis was used to calculate other complete *Wolbachia* supergroup A genomes. The BUSCO scores were 84.5% in *w*Mel and *w*Ri, 84% in *w*Ha, and 84.9% in *w*Au, suggesting that our assembly, despite being fragmented, has recovered most of the

proteobacteria orthologs. Thus, the assembly could serve as the reference genome with robustness to reveal the population genomic variations (Table 4.2).

Table 4.2 The statistics of *Wolbachia* genome assembly.

Strain	Host species	Super-group	No. of contig	Genome size (Mb)	GC%	BUSCO score (%)	RefSeq Assembly Accession
*w*Au	*Drosophila simulans*	A	1	1.26	35.2	84.9	GCF_000953315.1
*w*Ha	*Drosophila simulans*	A	1	1.29	35.3	84	GCF_000376605.1
*w*Mel	*Drosophila melanogaster*	A	1	1.27	35.2	84.5	GCF_000008025.1
*w*Ri	*Drosophila simulans*	A	1	1.45	35.2	84.5	GCF_000022285.1
*w*Agra	*Anoplolepis gracilipes*	A	96	1.2	35.2	85.4	JACRYZ000000000 (This study)

The BUSCO scores for all genomes were calculated with identical parameters using proteobacteria_obd10.

Our SNP calling and filtering yielded 129 SNP loci from 68 *w*Agra-infected individuals (a total of 12 individual were excluded from the analysis, and these individuals included 3 *Wolbachia*-free individuals and 9 individuals that failed to pass our SNP filtering). I found 53 SNP loci that possessed polymorphisms. To verify if the discovered SNPs are located in the *Wolbachia* genome, I randomly selected 10 SNP loci and PCR-amplified on both *Wolbachia*-infected and *Wolbachia*-free colonies. All 10 polymorphic SNP loci were successfully amplified only from *Wolbachia*-infected samples but not *Wolbachia*-free ones, suggesting that these SNPs originated from the *w*Agra genome.

4.3.3. Phylogenomic Tree and Mitochondrial Network

I compared the phylogenetic trees based on 129 *Wolbachia*-origin SNPs and the network analysis based on partial mitochondrial *COI* sequence of the host (854 bp; GenBank accessions: MT899004-MT899082) and found a high degree of concordance (Figure 4.2). *w*Agra from Thailand (Thailand_01) was among the most genetically divergent (Figure 4.2a), coinciding with the fact that mitochondrial *COI* haplotype from the same individual was characterized with the most nucleotide polymorphism among all the samples (Figure 4.2b). Individuals harboring mitochondrial haplotypes H06-H11 were grouped together in *Wolbachia* SNPs Clade I (Figure 4.2; blue square), whereas the remaining samples including those from Indonesia, Australia, Malaysia (Borneo), and part of southern Taiwan were clustered in *Wolbachia* SNPs Clade III (Figure 4.2; brown square). To exclude the possibility of the observed similar topology resulting from amplification of the *Wolbachia COI* instead of the host mitochondrial *COI*, I used BLASTn (task: blastn-short) to compare *COI* primers against the *w*Agra

genome, and the result indicated no *Wolbachia* amplification was available. I also compared mitochondrial *COI* sequences against NCBI database with BLASTn and verified the *COI* sequences I amplified indeed derived from *A. gracilipes*.

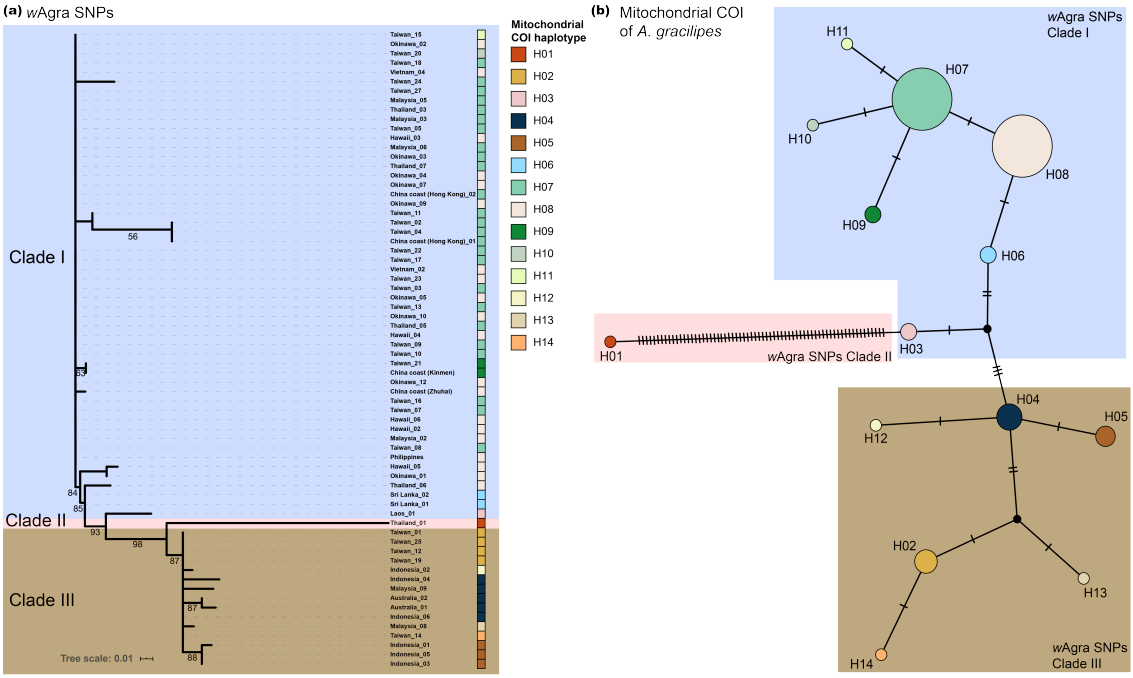


Figure 4.2 (a) *Wolbachia* (*w*Agra) SNPs phylogenetic tree based on Maximum Likelihood method. The mitochondrial COI haplotype of each sample is indicated in color boxes following individual name. Numbers at node indicate bootstrap support values (1,000 replicates). Tree branches with < 50% bootstrap supporting value were removed. (b) The median-joining network analysis of *A. gracilipes* mitochondrial COI haplotypes. The color shading depicts the *w*Agra clade displayed in (a). Circle area is proportional to the number of individuals carrying a given haplotype.

4.3.4. Expression of MLST, *wsp*, and Non-sysnonmous ORFs

Alignment of the *Wolbachia* SNPs revealed that eight distinct SNP loci contributed to the separation of the *Wolbachia* SNPs Clade I and Clade III. To further dissect the SNP polymorphisms associated with selection and/or demography, I analyzed the synonymous and non-synonymous mutations of these SNP loci. Seven of the eight SNP loci were located in ORFs, and three out of these seven were non-synonymous mutations, with two of which each residing in one of two *Wolbachia* hypothetical proteins and one in *RluA*-like gene. The *RluA* is a pseudouridine synthase gene involved in RNA binding and has been known to participate in the cellular information processing [46].

To verify if the genes with non-synonymous mutation express in *w*Agra, I examined the transcriptome from a pooled RNA sample prepared from two yellow crazy ant workers. In total, 32,771,008 paired RNA-seq reads passed the quality control. Among these, 11,975 (0.04%) of reads mapped onto the *w*Agra draft genome assembly. First, I found the MLST genes had fewer reads (*coxA*: 0, *hcpA*: 2, *gatB*: 4, *fbpA*: 8, and *ftsZ*: 26 reads per gene) than the *wsp* gene (73 reads). I then inspected the non-synonymous mutation containing ORFs and found that the hypothetical protein (scaffold_12890: 3,880-5,208 bp, hereinafter referred to as HP_12890; GenBank accession: MT896046) possessed 70 RNA reads, which is comparable to that of the *wsp* gene. Another hypothetical protein (scaffold_2912: 3397-3,696 bp) had only 9 reads, whereas no evidence of expression of the *RluA*-like gene (GenBank accession: MT896124) was detected.

4.3.5. Adaptive Selections

I sequenced partial fragment of the HP_12890 and *RluA*-like gene from all *Wolbachia*-infected colonies. The HP_12890 possessed four haplotypes (Figure 4.3a; GenBank accessions: MT895969- MT896045) and *RluA*-like gene had two haplotypes (Figure 4.3b; GenBank accessions: MT896047- MT896123). Both genes harbored SNP(s) that were detected by ddRAD-seq. Moreover, haplotype distribution of the HP_12890 was generally concordant to the *Wolbachia* SNPs clades. For example, the haplotype I of HP_12890 was found in most of individuals belonging to *Wolbachia* SNPs Clade I (Figure 3, highlighted in olive green square, mitochondrial H06-H11); while haplotype II of HP_12890 was restricted to individuals harboring mitochondrial *COI* haplotype V (H05). The haplotype III of HP_12890 is associated with individuals bearing mitochondrial *COI* haplotype H02-H04 and H12-H14, and the haplotype IV of HP_12890 was only harbored in Thailand_01 with a unique mitochondrial *COI* haplotype (H01; Figure 4.3). Among the four HP_12890 haplotypes, I found three polymorphic sites with all conveyed by non-synonymous mutations.

A similar pattern was found in the haplotype distribution of the *RluA*-like gene, with individuals from *Wolbachia* SNPs Clade I generally harboring one haplotype (*RluA*_H01) and those from Clade III harboring the other haplotype (*RluA*_H02). One exception, however, involved an individual in China (Yunnan) with the *RluA*-like haplotype that was found in the individual belonging to *Wolbachia* SNPs Clade III but harboring mitochondrial *COI* haplotype H08 (*Wolbachia* SNPs Clade I). The only difference between the two *RluA*-like gene haplotypes was a non-synonymous mutation.

Results of the Z-test on the HP_12890 indicated the signature of positive selection (dN > dS; $p = 0.043$) on the HP_12890 haplotype IV in Thailand (Thailand_01) compared to the most common HP_12890 haplotype (haplotype I, mitochondrial *COI*

haplotype H06-H11). Signal of positive selection was not detected in the *RluA*-like gene ORF (scaffold_2786:4747-5,286 bp; $p = 0.16$). I also found no evidence for purifying selection among all haplotype pairs for both genes (dN < dS; $p = 1$).

In addition, I compared HP_12890 against NCBI nucleotide collection (nr/nt) database with mega-BLASTn, and showed that orthologs of HP_12890 were present in *Wolbachia* of blowflies (*Chrysomya megacephala*; nucleotide identity: 93% (704/758); GenBank accession: CP021120.1) and the Asian citrus psyllid (*Diaphorina citri*; nucleotide identity: 91% (700/770); GenBank accession: CP051608.1). The results of interspecific Z-test indicated evidence of purifying selection for the HP_12890 (dN < dS; $p < 0.001$), while there was no evidence of positive selection between species (dN > dS; $p = 1$). A similar pattern was found for the *RluA*-like gene: I compared *Wolbachia* in hosts, including the horn fly (*Haematobia irritans*; *w*Irr; nucleotide identity: 96% (541/566); GenBank accession: CP037426.1) and fruit fly (*Drosophila melanogaster*; *w*Mel; nucleotide identity: 92% (523/567); GenBank accession: CP042445.1), and the interspecific Z-test was significant for purifying selection (dN < dS; $p < 0.001$) but not positive selection (dN > dS; $p = 1$).

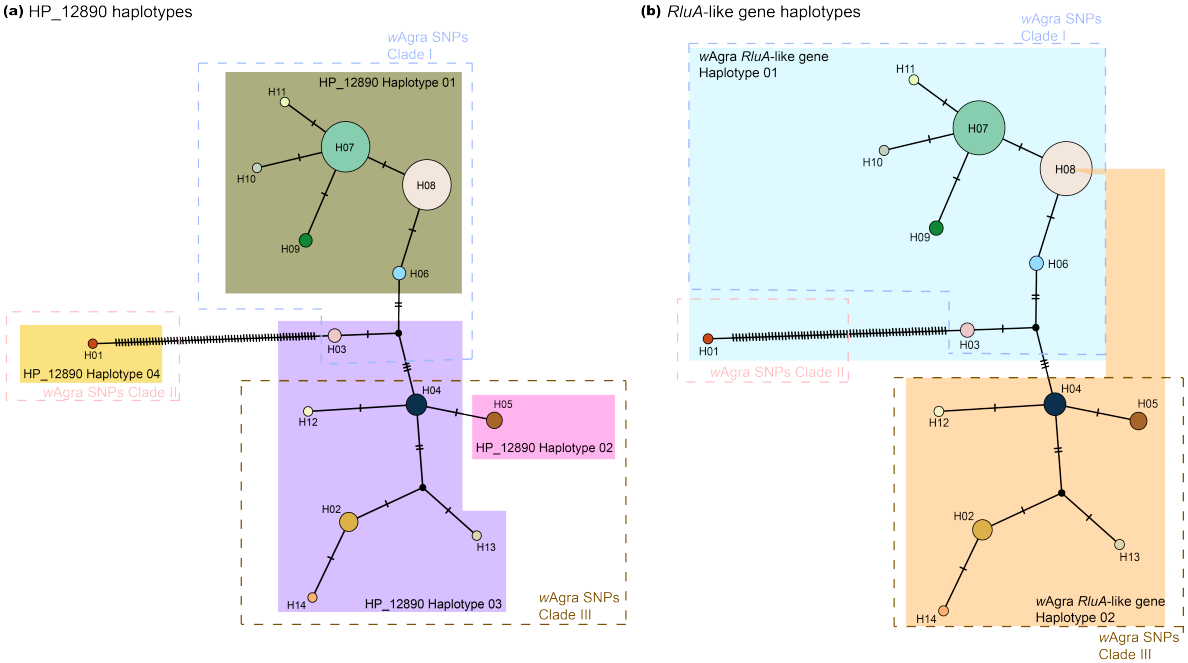


Figure 4.3 The *A. gracilipes* mitochondrial COI haplotypes (circles in color) and their corresponding *Wolbachia* (a) HP_12890, and (b) *RluA*-like gene haplotypes. Dashed lines indicate *w*Agra SNP clades as shown in Figure 2a. Circle area is proportional to the number of individuals carrying a given haplotype.

4.4 Discussion

4.4.1. Origin of *Wolbachia* in Yellow Crazy Ant

Our study extended the findings of Tseng et al. [18] and Sebastien et al. [19] showing that all yellow crazy ants across broad geographical regions harbor *Wolbachia* with the identical sequences of both housekeeping (e.g., MLST) and fast-evolving genes (e.g., *wsp*), suggesting that evolutionary history of *w*Agra should be recent and that *w*Agra may have derived from a single ancestral population. The MLST strain (ST52) in the yellow crazy ant (Formicinae) was also reported in an ant *Lophomyrmex* sp. (Myrmicinae) in Thailand [47]. Considering the fact that divergence time of *Lophomyrmex* sp. and *A. gracilipes* is approximately 111 MYA (CI: 99-126 MYA) [48], the identical MLST may imply the occurrence of horizontal transfer between *Lophomyrmex* sp. and *A. gracilipes*. Since phylogenies of *Wolbachia* and the ant's mitochondrial *COI* mirror to each other, it is likely that there may have been no additional horizontal transfer of *Wolbachia* involving other ant species bearing different MLST strains along the evolutionary time. Tseng et al. [18] showed that horizontal transfer of *Wolbachia* is prevailing between ant and ant guests with an intimate ecological association (e.g., longhorn crazy ant, *Paratrechina longicornis* and its host-specific ant cricket, *Myrmecophilous americanus*), except *A. gracilipes* and its closely-associated, host-specific ant cricket, *M. albicinctus*. Combined with near fixation of *Wolbachia* in all sampled *A. gracilipes* populations, vertical transmission, rather than horizontal, appears to be the predominant route for the spread of *Wolbachia* within this ant species.

4.4.2. High Infection Rate with a Single *Wolbachia* Strain

One major finding of the present study is that *Wolbachia* is highly prevalent across sampled colonies of *A. gracilipes* in virtually all geographic regions. In some cases, *Wolbachia* infection rates in ants were found to be 100%, but most of these studies either were conducted at a rather local scale (e.g., single collection site) [49] or that the infection rates fluctuate temporally and spatially [22,47,50–56]. A similar *Wolbachia* prevalence pattern has been reported in the parasitic wasp [57], bedbugs [5], and filarial nematodes [3,7]. These hosts apparently rely on *Wolbachia* for essential functions such as nutritional provision, metabolism facilitation, or oocyte maturation and thus engage in an exclusive mutualistic relationship with *Wolbachia*. Hence, the observed high prevalence of *Wolbachia* in *A. gracilipes* may imply an intimate association between *Wolbachia* and *A. gracilipes*, although more empirical tests are needed.

The selection pressures exerted by natural enemies often lead to an increase of particular inherited symbionts in insect populations [58,59]. For example, high

infection of *Spiroplasma* is observed in *Drosophila* populations with the presence of parasitic nematodes [60]. While natural enemies of *A. gracilipes* and their pathogenicity remain poorly described (as compared to other major pest insects), several potential pathogens (i.e., Anoplolepis gracilipes virus 1, Anoplolepis gracilipes virus 2, and TR44839 virus) have been reported in Cooling et al. [61] and Lee et al. [62]. Evidence is emerging to suggest some of which can be pathogenic and associated with fitness cost of the host [63]. TR44839 virus was found to be highly prevalent in *A. gracilipes* populations in Okinawa (Japan), Taiwan, and Penang (Malaysia) [64], it, however, appears to persist in low viral titers (Hsu et al., unpublished data). Another two dicistroviruses (Anoplolepis gracilipes virus 1 and Anoplolepis gracilipes virus 2) were also reported in the ant collected from Taiwan and Malaysia [62]. High prevalence of *Wolbachia* may potentially be maintained by benefits of endosymbiotic-driven defense that provides *A. gracilipes* protection against these viral pathogens, as this is the case frequently observed in flies [65] and mosquitos [66].

Another likely explanation for the high prevalence of *Wolbachia* in *A. gracilipes* is "Jekyll and Hyde" infection of *Wolbachia* [2,67]. *Wolbachia* may act as beneficial symbiont and reproductive parasite simultaneously, which is termed "Jekyll and Hyde" infection [2,67], and this may largely facilitate the endosymbiont spread in the host population [67,68]. For example, rapid spread and high infection frequency of *Wolbachia* in California *Drosophila* populations were shown to be associated with the dynamic interaction between parasitic and mutualistic life modes of *Wolbachia* [69]. The reproductive manipulation of *Wolbachia* in ants seems uncommon (but see [13], with the effects, if any, being highly species-dependent). For example, *Wolbachia* infection possesses negligible influence on sex ratio in Formica ants [70,71] yet drives the Pharaoh ant *Monomorium pharaonis* toward female-biased [12]. Interestingly, the same *Wolbachia* strain was also found to promote colony growth and early colony reproduction of the Pharaoh ant [72], providing arguably the first case of "Jekyll and Hyde" infection of *Wolbachia* in ant. Whether *w*Agra manipulates reproduction and/or enhance productivity of *A. gracilipes* remains an open question, and empirical evidence is now being generated.

Host life-history style may represent another potential reason for the extraordinarily high prevalence of *Wolbachia* in *A. gracilipes*. For example, *Wolbachia* infection rate in beetles is associated with reproduction mode (e.g., parthenogenesis), mobility, geographical distribution, body size, and/or population connectivity (e.g., fragmented or isolated) [73]. In ants, species with polygynous colony structure (multiple reproductive queens) and budding as major colony reproduction mode generally harbor a higher *Wolbachia* infection rate, and these traits may have facilitated the persistence of *Wolbachia* in the host [74]. We, however, failed to find evidence for

such support, as other ants with a similar life history style were never reported to have an even comparable level of *Wolbachia* prevalence observed in *A. gracilipes*.

4.4.3. Evidence for *Wolbachia*–Host Coevolution

While it remains uncertain that *Wolbachia* are in parasitic or mutualist association with *A. gracilipes*, *Wolbachia*–host coevolution should be expected, given the extremely high *Wolbachia* infection in the ant and the high level of concordance between *Wolbachia* SNPs tree and geographical distribution of mitochondrial *COI* haplotypes of *A. gracilipes*.

Gene flow between different supercolonies of *A. gracilipes* is limited because a high level of aggression is expected between workers (and possibly between worker and gyne) originating from different supercolonies, and this pattern applies even within a fine geographical scale [16,17,64]. Combined with the fact that independent foundation is rare in *A. gracilipes* [15], spread of *Wolbachia* at a local scale is either discouraged or highly dependent on supercolony structure. Nevertheless, human-mediated jump dispersal has been the primary mode of spread for many invasive ants [75] including *A. gracilipes*. Indeed, historical records and our preliminary analyses (Lee et al., unpublished data) showed that *A. gracilipes* has been transported to multiple Indo-Pacific islands through anthropogenic activities [14,76]. Such a strong affinity with human-associated transportation is predicted to create a route for its associated symbionts, including *Wolbachia*, to spread across different biogeographical regions. Co-divergence of *Wolbachia* and mitochondrial *COI* phylogeny of *A. gracilipes* provides firm support to this prediction and suggests that this bacterium may have frequently spread with the host ant as a hitchhiker. An alternative possibility is that invasion success of this ant is at least partially associated with *Wolbachia* infection (e.g., *Wolbachia* as a potential nutritional mutualist in introduced populations of the ghost ant, *Tapinoma melanocephalum* [77]). The latter possibility is of particular interest in the context of invasion biology as our finding represents one of very few cases, especially in invasive ants, in which *Wolbachia* persists at high prevalence across the entire invasive range. Several factors—including drift, altered selection pressures, imperfect maternal transmission, or natural curing events—are attributable to the loss (or low prevalence) of *Wolbachia* infection during colonization of invasive ants [56,78,79], yet our results provide a new research avenue to investigate how interplay between *Wolbachia* and host shapes the invasion success of ant.

I note that one particular sample from Thailand (Thailand_01) possesses a deep genetic divergence in both *Wolbachia* and host mitochondria DNA from other samples of *A. gracilipes* (Figure 4.2), suggesting there is under-discovered yet great deal of genetic diversity in or somewhere near the western part of Thailand and that the populations in this region might be the origin of this invasive ant. Such a finding is

parallel to the Southeast Asian origin of *A. gracilipes* proposed by multiple studies [14,15] despite more comprehensive sampling and additional nuclear data needed. Furthermore, it has been demonstrated that endosymbionts could assist in disclosing the hidden and recent population demography of its host [80,81], combined with the fact that *Lophomyrmex* ants in Thailand and *A. gracilipes* share an identical MLST [47], characterization of genetic variation of *Wolbachia* in both ant species in the particular region would accelerate reconstructing the origin of *Wolbachia*, as well as their hosts through understanding the coevolutionary history.

4.4.4. Rapid Evolution of a Hypothetical *Wolbachia* Gene

The observed signature of co-divergence between *w*Agra and *A. gracilipes* raises a question of whether it is driven by the bacterium or simply reflects the demographic history of host population. I confirmed the presence of non-synonymous mutations in the two *Wolbachia* genes—one of which, namely *RluA*-like gene—is crucial for bacterial cellular functions [46]. Signal of purifying selection revealed by the interspecific Z-test supports the essential role of this gene in *Wolbachia*. While I observed nucleotide polymorphisms in *w*Agra across our sampled populations of *A. gracilipes*, no evidence for expression nor signature of positive selection suggests that these nucleotide polymorphisms are most likely associated with demographic history of the host such as bottleneck during population expansion or genetic drift, although I cannot rule out the possibility that divergence time was not sufficient to accumulate synonymous mutations or that our RNA sequencing depth is insufficient to detect the evidence for expression.

I fully understand that the read number here serves an indicator of "relative" abundance of gene expression because our RNA sequencing read depth likely is below the level where gene expression pattern can be precisely characterized. However, our transcriptome analysis verifies the expression of the HP_12890 in *w*Agra and shows that the HP_12890 possesses comparable RNA reads to the *wsp* gene that plays a key role in mediating *Wolbachia*–host interactions. One potential explanation for such high expression level is that the HP_12890 may interact intimately with its host *A. gracilipes*. Alternatively, it is likely that the HP_12890 acts as a selfish element (e.g., mobile elements), and the observed polymorphism may be the result of selfish gene replication, given that the selfish element usually can be highly expressed and spreads rapidly in the host genome [82]. We, however, did not detect the presence of multiple copies, a typical characteristic of selfish element, nor sequence polymorphisms within individuals in our analysis. Interestingly, the Z-test indicates signature of purifying selection in interspecific comparison but positive selection among intraspecific haplotypes for the HP_12890. This suggests that this gene may be functionally important and that the observed non-synonymous mutations may be associated with host demography and/or rapid evolution of *Wolbachia*.

This study reports a single *Wolbachia* strain that is nearly fixed in a globally distributed invasive ant species and adds to the existing literature showing *Wolbachia*–ant coevolution. More importantly, our data may represent a rare case involving *Wolbachia* as a potential mutualist in ants and raises an invasion biology-centered question: how this bacterium is maintained at high prevalence during the colonization of this ant. Genetic and functional characterization of *Wolbachia* and its phenotypic effect on *A. gracilipes* would offer additional insights into the dynamics of the *Wolbachia*–ant coevolution.

4.5 References

1. Werren, J.H.; Baldo, L.; Clark, M.E. *Wolbachia*: Master manipulators of invertebrate biology. *Nat. Rev. Microbiol.* **2008**, *6*, 741–751.
2. Zug, R.; Hammerstein, P. Bad guys turned nice? A critical assessment of *Wolbachia* mutualisms in arthropod hosts. *Biol. Rev.* **2015**, *90*, 89–111.
3. Landmann, F.; Foster, J.M.; Michalski, M.L.; Slatko, B.E.; Sullivan, W. Co-evolution between an endosymbiont and its nematode host: *Wolbachia* asymmetric posterior localization and AP polarity establishment. *PLoS Negl. Trop. Dis*. **2014**, *8*, e3096.
4. Lefoulon, E.; Bain, O.; Makepeace, B.L.; d'Haese, C.; Uni, S.; Martin, C.; Gavotte, L. Breakdown of coevolution between symbiotic bacteria *Wolbachia* and their filarial hosts. *PeerJ* **2016**, *4*, e1840.
5. Balvín, O.; Roth, S.; Talbot, B.; Reinhardt, K. Co-speciation in bedbug *Wolbachia* parallel the pattern in nematode hosts. *Sci. Rep*. **2018**, *8*, 8797.
6. Newton, I.L.; Rice, D.W. The Jekyll and Hyde symbiont: Could *Wolbachia* be a nutritional mutualist? *J. Bacteriol.* **2020**, *202*, e00589-19.
7. Taylor, M.J.; Bandi, C.; Hoerauf, A. *Wolbachia*. Bacterial endosymbionts of filarial nematodes. *Adv. Parasit.* **2005**, *60*, 245–284.
8. Riegler, M.; Stauffer, C. *Wolbachia* infections and superinfections in cytoplasmically incompatible populations of the European cherry fruit fly *Rhagoletis cerasi* (Diptera, Tephritidae). *Mol. Ecol*. **2002**, *11*, 2425–2434.
9. Schuler, H.; Köppler, K.; Daxböck-Horvath, S.; Rasool, B.; Krumböck, S.; Schwarz, D.; Hoffmeister, T.S.; Schlick-Steiner, B.C.; Steiner, F.M.; Telschow, A.; et al. The hitchhiker's guide to Europe: The infection dynamics of an ongoing *Wolbachia* invasion and mitochondrial selective sweep in *Rhagoletis cerasi*. *Mol. Ecol*. **2016**, *25*, 1595–1609.
10. Tagami, Y.; Miura, K. Distribution and prevalence of *Wolbachia* in Japanese populations of Lepidoptera. *Insect Mol. Ecol.* **2004**, *13*, 359–364.
11. Ahmed, M.Z.; Araujo-Jnr, E.V.; Welch, J.J.; Kawahara, A.Y. *Wolbachia* in butterflies and moths: Geographic structure in infection frequency. *Front. Zool.* **2015**, *12*, 16.
12. Pontieri, L.; Schmidt, A.M.; Singh, R.; Pedersen, J.S.; Linksvayer, T.A. Artificial selection on ant female caste ratio uncovers a link between female-biased sex ratios and infection by *Wolbachia* endosymbionts. *J. Evol. Biol.* **2017**, *30*, 225–234.

13. Ün, Ç.; Schultner, E.; Manzano-Marín, A.; Flórez, L.V.; Seifert, B.; Heinze, J.; Oettler, J. Cytoplasmic incompatibility between Old and New World populations of a tramp ant. *bioRxiv* doi: 10.1101/2020.03.13.988675
14. Wetterer, J.K. Worldwide distribution and potential spread of the long-legged ant, *Anoplolepis gracilipes* (Hymenoptera: Formicidae). *Sociobiology* **2005**, *45*, 77–97.
15. Ito, F.; Asfiya, W.; Kojima, J.I. Discovery of independent-founding solitary queens in the yellow crazy ant *Anoplolepis gracilipes* in East Java, Indonesia (Hymenoptera: Formicidae). *Entomol. Sci.* **2016**, *19*, 312–314.
16. Drescher, J.; Blüthgen, N.; Schmitt, T.; Bühler, J.; Feldhaar, H. Societies drifting apart? Behavioural, genetic and chemical differentiation between supercolonies in the yellow crazy ant *Anoplolepis gracilipes*. *PLoS ONE* **2010**, *5*, e13581.
17. Gruber, M.A.; Hoffmann, B.D.; Ritchie, P.A.; Lester, P.J. Recent behavioural and population genetic divergence of an invasive ant in a novel environment. *Divers. Distrib.* **2012**, *18*, 323–333.
18. Tseng, S.P.; Hsu, P.W.; Lee, C.C.; Wetterer, J.K.; Hugel, S.; Wu, L.H.; Lee, C.Y.; Yoshimura, T.; Yang, C.C.S. Evidence for common horizontal transmission of *Wolbachia* among ants and ant crickets: Kleptoparasitism added to the list. *Microorganisms* **2020**, *8*, 805.
19. Sebastien, A.; Gruber, M.A.M.; Lester, P.J. Prevalence and genetic diversity of three bacterial endosymbionts (*Wolbachia*, *Arsenophonus*, and *Rhizobiales*) associated with the invasive yellow crazy ant (*Anoplolepis gracilipes*). *Insectes Soc.* **2012**, *59*, 33–40.
20. Baldo, L.; Hotopp, J.C.D.; Jolley, K.A.; Bordenstein, S.R.; Biber, S.A.; Choudhury, R.R.; Hayashi, C.; Maiden, M.C.J.; Tettelin, H.; Werren, J.H. Multilocus sequence typing system for the endosymbiont *Wolbachia* pipientis. *Appl. Environ. Microbiol.* **2006**, *72*, 7098–7110.
21. PubMLST. Available online: https://pubmlst.org/wolbachia/info/protocols.shtml (accessed on 26 August 2020).
22. Shoemaker, D.D.; Ahrens, M.; Sheill, L.; Mescher, M.; Keller, L.; Ross, K.G. Distribution and prevalence of *Wolbachia* infections in native populations of fire ant *Solenopsis invicta* (Hymenoptera: Formicidae). *Environ. Entomol.* **2003**, *32*, 1329–1336.
23. Lee, C.C.; Wang, J.; Matsuura, K.; Yang, C.C.S. The complete mitochondrial genome of yellow crazy ant, *Anoplolepis gracilipes* (Hymenoptera: Formicidae). *Mitochondrial DNA B Resour*. **2018**, *3*, 622–623.
24. Untergasser, A.; Nijveen, H.; Rao, X.; Bisseling, T.; Geurts, R.; Leunissen, J.A. Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res.* **2007**, *35*, W71–W74.
25. Bleidorn, C.; Gerth, M. A critical re-evaluation of multilocus sequence typing (MLST) efforts in *Wolbachia*. *FEMS Microbiol. Ecol.* **2018**, *94*, fix163.
26. Kaur, R.; Siozios, S.; Miller, W.J.; Rota-Stabelli, O. Insertion sequence polymorphism and genomic rearrangements uncover hidden *Wolbachia* diversity in *Drosophila suzukii* and *D. subpulchrella*. *Sci. Rep.* **2017**, *7*, 14815.
27. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120.

28. Peng, Y.; Leung, H.C.; Yiu, S.M.; Chin, F.Y. IDBA-UD: A *de novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **2012**, *28*, 1420–1428.
29. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T.L. BLAST+: Architecture and applications. *BMC Bioinform.* **2009**, *10*, 421.
30. Seppey, M.; Manni, M.; Zdobnov, E.M. BUSCO: Assessing genome assembly and annotation completeness. In *Gene Prediction Methods in Molecular Biology*; Kollmar, M., Ed.; Humana: New York, NY, USA, 2019; Volume 1962, pp. 227–245.
31. Gurevich, A.; Saveliev, V.; Vyahhi, N.; Tesler, G. QUAST: Quality assessment tool for genome assemblies. *Bioinformatics* **2013**, *29*, 1072–1075.
32. Beichman, A.C.; Huerta-Sanchez, E.; Lohmueller, K.E. Using genomic data to infer historic population dynamics of nonmodel organisms. *Annu. Rev. Ecol. Evol. Syst.* **2018**, *49*, 433–456.
33. Peterson, B.K.; Weber, J.N.; Kay, E.H.; Fisher, H.S.; Hoekstra, H.E. Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE* **2012**, *7*, e37135.
34. FastQC. Available online: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed on 26 August 2020).
35. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **2011**, *17*, 10–12.
36. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics* **2009**, *25*,1754–1760.
37. Rochette, N.C.; Rivera-Colón, A.G.; Catchen, J.M. Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Mol. Ecol.* **2019**, *28*, 4737–4754.
38. Danecek, P.; Auton, A.; Abecasis, G.; Albers, C.A.; Banks, E.; DePristo, M.A.; Handsaker, R.E.; Lunter, G.; Marth, G.T.; Sherry, S.T.; et al. The variant call format and VCFtools. *Bioinformatics* **2011**, *27*, 2156–2158.
39. Dhaygude, K.; Nair, A.; Johansson, H.; Wurm, Y.; Sundström, L. The first draft genomes of the ant *Formica exsecta*, and its *Wolbachia* endosymbiont reveal extensive gene transfer from endosymbiont to host. *BMC Genom.* **2019**, *20*, 1–16.
40. Kozlov, A.M.; Darriba, D.; Flouri, T.; Morel, B.; Stamatakis, A. RAxML-NG: A fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **2019**, *35*, 4453–4455.
41. Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **2016**, *33*, 1870–1874.
42. Leigh, J.W.; Bryant, D. POPART: Full-feature software for haplotype network construction. *Methods Ecol. Evol.* **2015**, *6*, 1110–1116.
43. Thorvaldsdóttir, H.; Robinson, J.T.; Mesirov, J.P. Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Brief. Bioinform.* **2013**, *14*, 178–192.
44. Kim, D.; Paggi, J.M.; Park, C.; Bennett, C.; Salzberg, S.L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **2019**, *37*, 907–915.

45. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The sequence alignment/map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079.
46. Mao, M.; Yang, X.; Bennett, G.M. Evolution of host support for two ancient bacterial symbionts with differentially degraded genomes in a leafhopper host. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E11691–E11700.
47. Russell, J.A.; Goldman-Huertas, B.; Moreau, C.S.; Baldo, L.; Stahlhut, J.K.; Werren, J.H.; Pierce, N.E. Specialization and geographic isolation among *Wolbachia* symbionts from ants and lycaenid butterflies. *Evolution* **2009**, *63*, 624–640.
48. Hedges, S.B.; Marin, J.; Suleski, M.; Paymer, M.; Kumar, S. Tree of life reveals clock-like speciation and diversification. *Mol. Biol. Evol.* **2015**, *32*, 835–845.
49. Wenseleers, T.; Ito, F.; Van Borm, S.; Huybrechts, R.; Volckaert, F.; Billen, J. Widespread occurrence of the microorganism *Wolbachia* in ants. *P. Roy. Soc. B Biol. Sci.* **1998**, *265*, 1447–1452.
50. Shoemaker, D.D.; Ross, K.G.; Keller, L.; Vargo, E.; Werren, J.H. *Wolbachia* infections in native and introduced populations of fire ants (*Solenopsis* spp.). *Insect Mol. Biol.* **2000**, *9*, 661–673.
51. Van Borm, S.; Wenseleers, T.; Billen, J.; Boomsma, J.J. *Wolbachia* in leafcutter ants: A widespread symbiont that may induce male killing or incompatible matings. *J. Evol. Biol.* **2001**, *14*, 805–814.
52. Tsutsui, N.D.; Kauppinen, S.N.; Oyafuso, A.F.; Grosberg, R.K. The distribution and evolutionary history of *Wolbachia* infection in native and introduced populations of the invasive argentine ant (*Linepithema humile*). *Mol. Ecol.* **2003**, *12*, 3057–3068.
53. Frost, C.L.; Fernández-Marín, H.; Smith, J.E.; Hughes, W.O.H. Multiple gains and losses of *Wolbachia* symbionts across a tribe of fungus-growing ants. *Mol. Ecol.* **2010**, *19*, 4077–4085.
54. Ramalho, M.O.; Martins, C.; Silva, L.M.; Martins, V.G.; Bueno, O.C. Intracellular symbiotic bacteria of *Camponotus textor*, Forel (Hymenoptera, Formicidae). *Curr. Microbiol.* **2017**, *74*, 589–597.
55. Kelly, M.; Price, S.L.; de Oliveira Ramalho, M.; Moreau, C.S. Diversity of *Wolbachia* associated with the giant turtle ant, *Cephalotes atratus*. *Curr. Microbiol.* **2019**, *76*, 1330–1337.
56. Tseng, S.P.; Wetterer, J.K.; Suarez, A.V.; Lee, C.Y.; Yoshimura, T.; Shoemaker, D.; Yang, C.C.S. 2019. Genetic diversity and *Wolbachia* infection patterns in a globally distributed invasive ant. *Front. Genet.* **2019**, *10*, 838.
57. Dedeine, F.; Vavre, F.; Fleury, F.; Loppin, B.; Hochberg, M.E.; Boulétreau, M. Removing symbiotic *Wolbachia* bacteria specifically inhibits oogenesis in a parasitic wasp. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 6247–6252.
58. Vorburger, C.; Gouskov, A. Only helpful when required: A longevity cost of harbouring defensive symbionts. *J. Evol. Biol.* **2011**, *24*, 1611–1617.
59. Oliver, K.M.; Campos, J.; Moran, N.A.; Hunter, M.S. Population dynamics of defensive symbionts in aphids. *P. Roy. Soc. B Biol. Sci.* **2008**, *275*, 293–299.
60. Jaenike, J.; Brekke, T.D. Defensive endosymbionts: A cryptic trophic level in community ecology. *Ecol. Lett.* **2011**, *14*,150–155.
61. Cooling, M.; Gruber, M.A.M.; Hoffmann, B.D.; Sébastien, A.; Lester, P.J. A metatranscriptomic survey of the invasive yellow crazy ant, *Anoplolepis gracilipes*,

identifies several potential viral and bacterial pathogens and mutualists. *Insectes Soc*. **2017**, *64*, 197–207.
62. Lee, C.C.; Lin, C.Y.; Hsu, H.W.; Yang, C.C.S. Complete genome sequences of two novel dicistroviruses detected in the yellow crazy ant, *(Anoplolepis gracilipes). Arch. Vriol.* **2020**, *165*, 2715-2719.
63. Cooling, M.D.; Hoffmann, B.D.; Gruber, M.A.M.; Lester, P.J. Indirect evidence of pathogen-associated altered oocyte production in queens of the invasive yellow crazy ant, *Anoplolepis gracilipes*, in Arnhem Land, Australia. *Bull. Entomol. Res*. **2018**, *108*, 451–460.
64. Hsu, H.W.; Chiu, M.C.; Lee, C.C.; Lee, C.Y.; Yang, C.C.S. The association between virus prevalence and Intercolonial aggression levels in the yellow crazy ant, *Anoplolepis gracilipes* (Jerdon). *Insects* **2019**, *10*, 436.
65. Kaur, R.; Martinez, J.; Rota-Stabelli, O.; Jiggins, F.M.; Miller, W.J. Age, tissue, genotype and virus infection regulate *Wolbachia* levels in *Drosophila*. *Mol. Ecol.* **2020**, *29*, 2063–2079.
66. Frentiu, F.D.; Zakir, T.; Walker, T.; Popovici, J.; Pyke, A.T.; van den Hurk, A.; McGraw, E.A.; O'Neill, S.L. Limited dengue virus replication in field-collected *Aedes aegypti* mosquitoes infected with *Wolbachia*. *PLoS Negl. Trop. Dis*. **2014**, *8*, e2688.
67. Jiggins, F.M.; Hurst, G.D. Rapid insect evolution by symbiont transfer. *Science* **2011**, *332*, 185–186.
68. Himler, A.G.; Adachi-Hagimori, T.; Bergen, J.E.; Kozuch, A.; Kelly, S.E.; Tabashnik, B.E.; Chiel, E.; Duckworth, V.E.; Dennehy, T.J.; Zchori-Fein, E.; et al. Rapid spread of a bacterial symbiont in an invasive whitefly is driven by fitness benefits and female bias. *Science* **2011**, *332*, 254–256.
69. Weeks, A.R.; Turelli, M.; Harcombe, W.R.; Reynolds, K.T.; Hoffmann, A.A. From parasite to mutualist: Rapid evolution of *Wolbachia* in natural populations of *Drosophila*. *PLoS Biol*. **2007**, *5*, e114.
70. Keller, L.; Liautard, C.; Reuter, M.; Brown, W.D.; Sundström, L.; Chapuisat, M. Sex ratio and *Wolbachia* infection in the ant *Formica exsecta*. *Heredity* **2001**, *87*, 227–233.
71. Wenseleers, T.; Sundström, L.; Billen, J. Deleterious *Wolbachia* in the ant *Formica truncorum*. *P. Roy. Soc. B Biol. Sci.* **2002**, *269*, 623–629.
72. Singh, R.; Linksvayer, T.A. *Wolbachia*-infected ant colonies have increased reproductive investment and an accelerated life cycle. *J. Exp. Biol.* **2020**, *223*, jeb220079.
73. Kajtoch, Ł.; Kolasa, M.; Kubisz, D.; Gutowski, J.M.; Ścibior, R.; Mazur, M.A.; Holecová, M. Using host species traits to understand the *Wolbachia* infection distribution across terrestrial beetles. *Sci. Rep.* **2019**, *9*, 847.
74. Treanor, D.; Hughes, W.O.H. Limited female dispersal predicts the incidence of *Wolbachia* across ants (Hymenoptera: Formicidae). *J. Evol. Biol.* **2019**, *32*, 1163–1170.
75. Suarez, A.V.; Holway, D.A.; Case, T.J. Patterns of spread in biological invasions dominated by long-distance jump dispersal: Insights from Argentine ants. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 1095–1100.
76. Janicki, J.; Narula, N.; Ziegler, M.; Guénard, B.; Economo, E.P. Visualizing and interacting with large-volume biodiversity data using client–server web-mapping

applications: The design and implementation of antmaps.org. *Ecol. Inform.* **2016**, *32*, 185–193.
77. Cheng, D.; Chen, S.; Huang, Y.; Pierce, N.E.; Riegler, M.; Yang, F.; Zeng, L.; Lu, Y.; Liang, G.; Xu, Y. Symbiotic microbiota may reflect host adaptation by resident to invasive ant species. *PLoS Pathog*. **2019**, *15*, e1007942.
78. Reuter, M.; Pedersen, J.S.; Keller, L. Loss of *Wolbachia* infection during colonisation in the invasive Argentine ant *Linepithema humile*. *Heredity* **2005**, *94*, 364–369.
79. Yang, C.C.; Yu, Y.C.; Valles, S.M.; Oi, D.H.; Chen, Y.C.; Shoemaker, D.D.; Wu, W.J.; Shih, C.J. Loss of microbial (pathogen) infections associated with recent invasions of the red imported fire ant *Solenopsis invicta*. *Biol. Invasions* **2010**, *12*, 3307–3318.
80. Biek, R.; Drummond, A.J.; Poss, M. A virus reveals population structure and recent demographic history of its carnivore host. *Science* **2006**, *311*, 538–541.
81. Biek, R.; Henderson, J.C.; Waller, L.A.; Rupprecht, C.E.; Real, L.A. A high-resolution genetic signature of demographic and spatial expansion in epizootic rabies virus. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 7993–7998.
82. Lee, C.C.; Wang, J. Rapid expansion of a highly germline-expressed mariner element acquired by horizontal transfer in the fire ant genome. *Genome Biol. Evol.* **2018**, *10*, 3262–3278.

Chapter 5: Novel dicistroviruses in the yellow crazy ant

5.1 Introduction

High throughput sequencing (HTS) technique has accelerated discovery of new viral pathogens in a number of arthropod taxa. The family *Dicistroviridae* consists RNA viruses with a linear, positive-sense and single-stranded genome, which mainly infect arthropod hosts such as aphids, leafhoppers, flies, bees and ants [1]. Three genera were reported in *Dicistroviridae* including *Aparavirus*, *Cripavirus*, and *Triatovirus* [1]. To date, seven *Triatovirus*-like viruses were discovered and formed a monophyletic group with another five viruses that were officially classified as *Triatovirus*, suggesting the existence of a total of 12 *triatoviruses* [2-6]. The majority of these triatoviruses were found to infect insects including eight ant species [2-8]. While a previous study reported partial sequence of an *Aparavirus*-like virus in the invasive yellow crazy ant (*Anoplolepis gracilipes*) [9], diversity and prevalence of triatovirus is largely unknown. Here, I report two novel *Triatovirus*-like viruses in *A. gracilipes* detected by HTS, coupled with their phylogenetic placement among other *triatoviruses* and infection patterns of the two viruses from the Asia-Pacific region.

5.2 Materials and methods

I extracted RNA from two adult worker ants of *A. gracilipes* individually, with one collected from Penang, Malaysia (5°21'32.4"N, 100°18'09.0"E) and the other from Okinawa, Japan (26°40'19.2"N, 128°00'41.0"E). Briefly, the entire worker ant was soaked in TRIzol™ RNA Extraction Reagent (Invitrogen, Carlsbad, CA, USA) and homogenized with a pestle, with the rest of procedure following the standard TRIzol RNA extraction protocol. The RNA quality and quantity were measured with Nanodrop spectrophotometers, and then mixed in 1:1 ratio of the total RNA amount. I mixed RNA of the two workers to maximize recovery of diverse sequences, if any, given the facts that virus prevalence may vary across different geographic locations and that isolates of RNA virus from geographically distinct regions may differ in their sequence. The mixed sample was submitted to the Eurofins Genomics (Tokyo, Japan) where the total RNA was purified using polyA selection method, and the sequencing library was constructed with the TruSeq® Stranded mRNA Library Prep (insert size: 200bp). The library was sequenced using Illumina HiSeq 4000 platform in the pair-end mode (2x100 bp). The obtained pair-end read sequences were trimmed to remove adaptors and low-quality sequences (q <28) using Trimmomatic v0.36 [10], and assembled *de novo* into transcripts using Trinity v2.8.4 [11]. The resulting transcripts

were compared against NCBI GenBank non-redundant database using BLASTx search in Diamond v0.9.21 [12], and virus-like sequences were identified using MEGAN Community Edition v6.18.8 [13]. The putative open reading frames (ORFs) and conserved domains were identified using NCBI ORF finder [14] and NCBI Conserved Domains Search [15]. The preliminary taxonomic inference was performed using BLASTn and BLASTx that compared identified virus-like sequences against the NCBI standard databases (nucleotide collection (nr/nt) and non-redundant protein sequences (nr)).

To resolve the taxonomic classification of the viruses, I collected sequence information of up-to-date, well-classified dicistroviruses [2-6] containing full non-structural polyprotein (ORF1) sequences from GenBank and Virus-Host DB [7]. A total of 44 dicistroviruses (12 triatoviruses, 14 cripaviruses, 15 aparaviruses, and 3 additional dicistroviruses) were then obtained for phylogenetic comparisons. Amino acid sequences of the entire non-structural polyprotein (ORF1) were aligned using MUSCLE in MEGA 7. The best evolutionary model selection was performed using ModelTest-NG v0.1.6 [16], and the VT+I+G4 model was selected with the highest probability. Phylogenetic tree based on Maximum likelihood method was reconstructed using RAxML-NG v0.9.0 [17] with 2,000 bootstrap replicates and visualized using iTOL v5 [18] (Note that the final phylogenetic tree was re-rooted at the clade of Wenzhou picorna-like virus 34 and Beihai picorna-like virus 80 [6]).

Table 5.1 Primers of RNA-dependent RNA polymerase on *Anoplolepis gracilipes* viruses

Virus	Primer name	Nucleotide sequence (5'- 3')	Position (bp)	Annealing temperature	Anticipated size
Anoplolepis gracilipes virus 1	AgrV1_RdRp.for	GGACGAAAGAAAACCAATTCA	5007-5027	60°C	772 bp
	AgrV1_RdRp.rev	GGGACAAAGGACAAATCCAA	5779-5760		
Anoplolepis gracilipes virus 2	AgrV2_RdRp.for	AATGAATGGGATTCGGTTGA	2520-2539	60°C	764 bp
	AgrV2_RdRp.rev	CCCACACAGGTTCCTTGACT	3284-3265		

To confirm whether the viruses are actively replicating in *A. gracilipes*, the strand-specific primers (AgrV1_RdRp.for and AgrV2_RdRp.for; Table 5.1) were used for cDNA synthesis. PrimeScript™ 1st strand cDNA Synthesis kit (Takara) was employed to synthesize cDNA under conditions as follows: 30°C for 10 min., 42 °C for 60 min., and final inactivation at 95°C for 5 min. The cDNA then served as a template for PCR reaction carried out with EmeraldAmp® MAX PCR Master Mix (Takara) under cycling profile as follows: 94 °C for 3 min, 35 cycles of 94 °C for 30 s, 60 °C for 30 s, and 72 °C for 60 s, with a final extension step of 72 °C for 5 min. The PCR amplicons were confirmed by both gel electrophoresis and Sanger sequencing. To verify the occurrence of identified viruses in *A. gracilipes* in the field, a total of 84 colonies were collected

from three regions including Okinawa (Japan; n=18), Taiwan (n=27) and Penang (Malaysia; n=39). Three random selected worker ants per colony were pooled and homogenized in TRIzol, and the same RNA extraction procedure as mentioned above was carried out. Protocols for the cDNA synthesis and PCR reactions were identical to those mentioned above, except that oligo dT primer was utilized when synthesizing the cDNA. To confirm whether virus-like sequences were novel triatoviruses, pairwise comparison of the amino acid identities of the ORF2 (Capsid protein) among 14 triatoviruses were performed using BLASTp. The identity with the highest bit-score of each virus-pair was shown.

5.3 Results and discussion

A total of 8,826,506 read sequences were assembled into 51,291 trinity 'genes' and 70,145 trinity transcripts. MEGAN analysis identified 22 virus-like genes (34 transcripts) with e-value < 1e-40. Only two of virus-like sequences contain complete viral conserved protein domains (RNA Helicase (Hel), RNA-dependent RNA polymerase (RdRp), and Capsid protein; Figure 5.1 and 5.2). The average coverage of reads per nucleotide for each virus-like genome is 50,371.26 and 42.63, respectively (Figure 5.1). The difference of coverage between two virus-like sequences likely results from differential viral titer between the two viruses in our sample. The two virus-like genomes are 10,662 and 8,238 nts in length (excluding the poly(A) tail), respectively, and both contain two open reading frames (ORFs) and a poly(A) tail at the 3'end (Figure 5.3), a characteristic feature from *Dicistroviridae* family members. For the first virus-like genome, ORF1 (nt 496-6,012) presents 5,517 nts, which encodes to a putative non-structural polyprotein containing RNA helicase (Hel; nt 2,137-2,491) and RNA-dependent RNA polymerase (RdRp; nt 4,975-5,812), respectively. ORF2 (nt 6,659-10,435) is 3,777 nts, which encodes to a putative structural protein containing a Capsid protein (CP; nt 9,434-9,977) domain. BLASTx and BLASTn analyses on non-structural polyprotein showed 54% amino acid identity and 68-75% nucleotide identity to the non-structural polyprotein of Solenopsis invicta virus 9 (MH727526.1) [3]. For the second virus-like genome, ORF1 (nt 135-4,151) contains Hel (nt 330-687) and RdRp (nt 3,072-3,966), respectively, whereas ORF2 (nt 4,688-8,056), a structural protein domain, contains CP (nt 6,932-7,592). The second virus-like genome had 64% amino acid identity and 69-75% nucleotide identity to the non-structural polyprotein of Myrmica scabrinodis virus 2 (MH477289.1) [2]. Results of RT-PCR confirmed the presence of sequences from the corresponding transcript as amplicons of anticipated size were recovered in the two *A. gracilipes* workers.

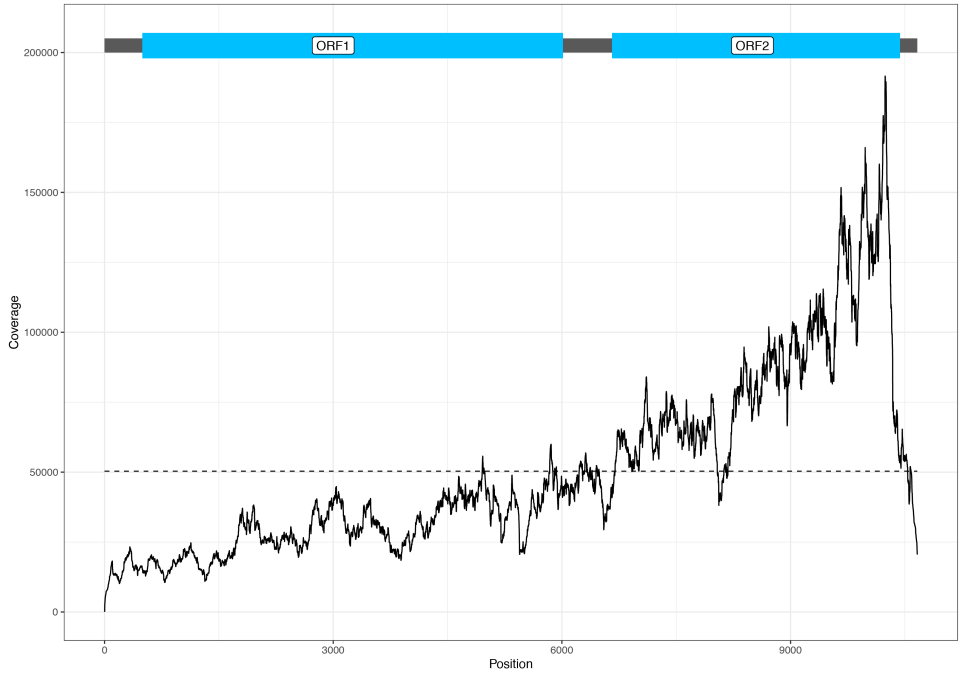
The first virus-like sequence was detected in both Taiwan (n=2) and Penang (n=7), resulting into 7.41% (2/27) and 17.95% (7/39) intercolonial infection rates, respectively (Table 5.2); whereas the second virus-like sequence was detected only in Penang (n=6)

with the prevalence of 15.38% (6/39). It is noteworthy that none of the two viruses was found in Okinawa, a pattern likely indicative of either a greater sample size required for positive detection or loss of microbes associated with invasion of the ant [19]. Furthermore, negative strands of both potential viral sequences were detected, implying active replication of the two "potential" viruses.

Table 5.2 Prevalence of the Anoplolepis gracilipes virus 1 and Anoplolepis gracilipes virus 2

Region (Country)	Number of colony	Anoplolepis gracilipes virus 1		Anoplolepis gracilipes virus 2	
		Number of infected colony	Prevalence (%)	Number of infected colony	Prevalence (%)
Okinawa (Japan)	18	0	0	0	0
Taiwan (Taiwan)	27	2	7.41	0	0
Penang (Malaysia)	39	7	17.95	6	15.38
Total	84	9	10.71	6	7.14

(a) Anoplolepis gracilipes virus 1 (MT108239)



(b) Anoplolepis gracilipes virus 2 (MT108240)

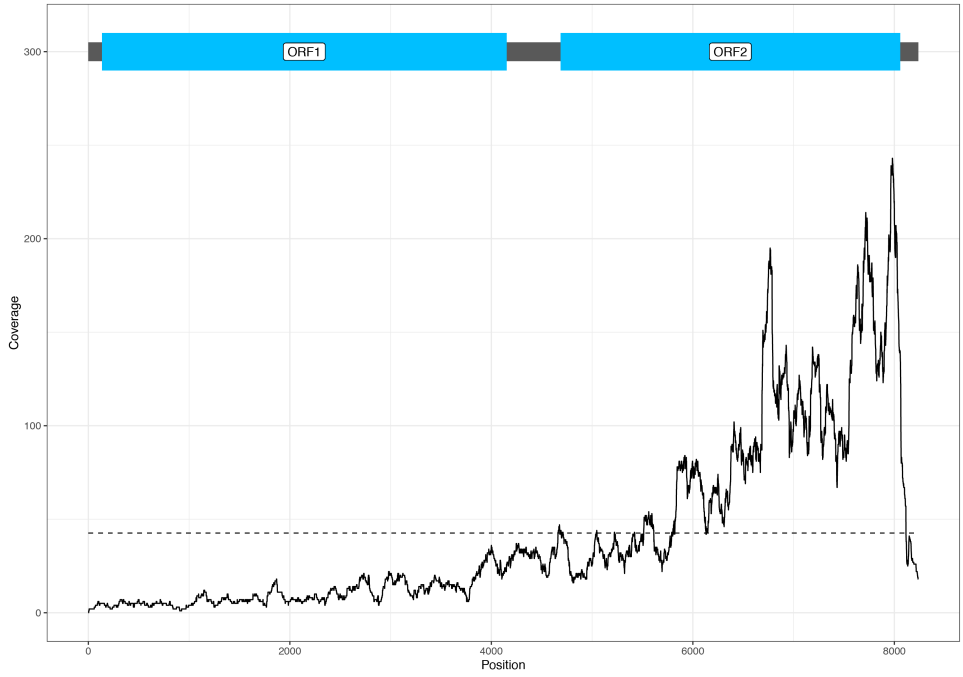


Figure 5.1 Reads count per nucleotide of two Anoplolepis gracilipes viruses in this study. The position begins from the 5′ end of the virus genome. The open reading frames were colored by blue. The dashed line indicates the average read coverage per nucleotide of the whole viral genome. Note the scale of the x-axis and y-axis are different between the two viruses.

(a) Anoplolepis gracilipes virus 1 (MT108239)

(b) Anoplolepis gracilipes virus 2 (MT108240)

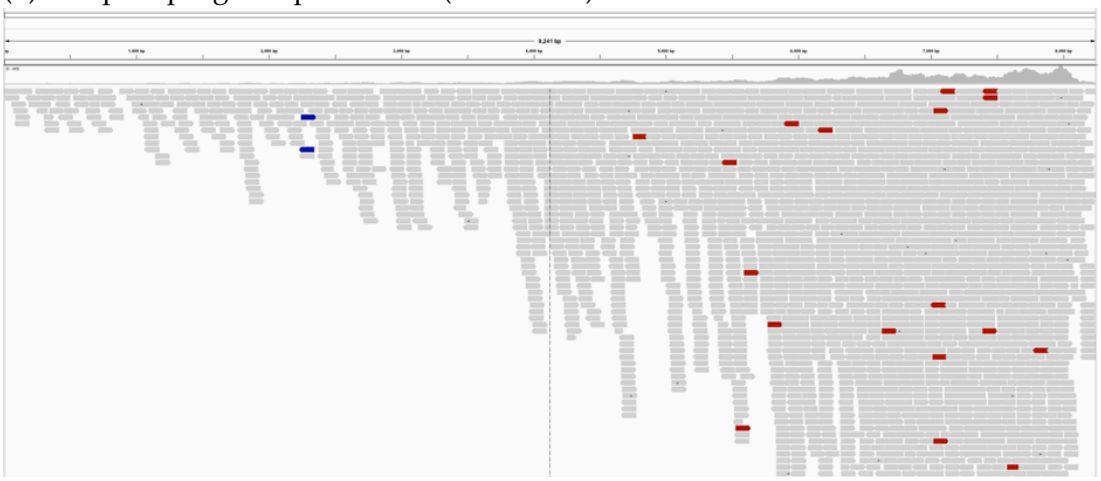


Figure 5.2 The IGV view of RNA-seq reads of two Anoplolepis gracilipes viruses. The position starts from the 5′ end of the virus genome. The read alignment showed no gap and variation indicate the correspondence of reads and consensus sequences.

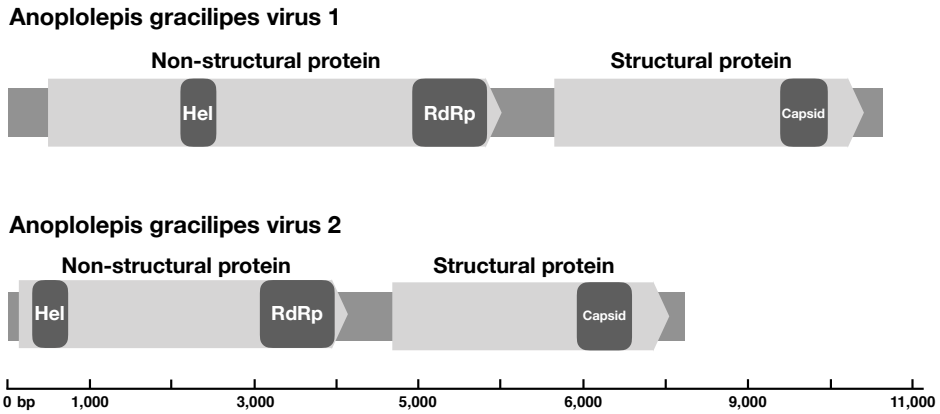


Figure 5.3 Genome structure of Anoplolepis gracilipes virus 1 and 2. Open reading frames are denoted as light gray rectangles, and the conserved domains were colored in dark gray rectangles.

Phylogenetic analyses based on amino acids of the entire non-structural polyprotein (ORF1) of the two virus-like sequences identified in the present study and additional 44 members of the family *Dicistroviridae* revealed that these two sequences are closely related to the members of the genus *Triatovirus* (Figure 5.2). The pair-wise amino acid identity of two virus-like sequences and other triatoviruses (maximum identity: 50% and 63%; Table S3) suggests that the two sequences shared a low similarity with any known viruses. Based on the analyses in the current study and the species demarcation criteria established by ICTV [20], I propose that the two viruses are novel *Triatovirus*-like viruses within *Dicistroviridae* family, and tentatively name these two viruses "Anoplolepis gracilipes virus 1" and "Anoplolepis gracilipes virus 2" (AgrV-1 and AgrV-2; Accession No. MT108239 and MT108240), respectively. This is the first report on discovery of novel viruses with description of its genome structures (two ORFs and conserve domains including RNA Helicase, RNA-dependent RNA polymerase, and Capsid protein) and complete sequences in the family *Dicistroviridae* from the invasive yellow crazy ant.

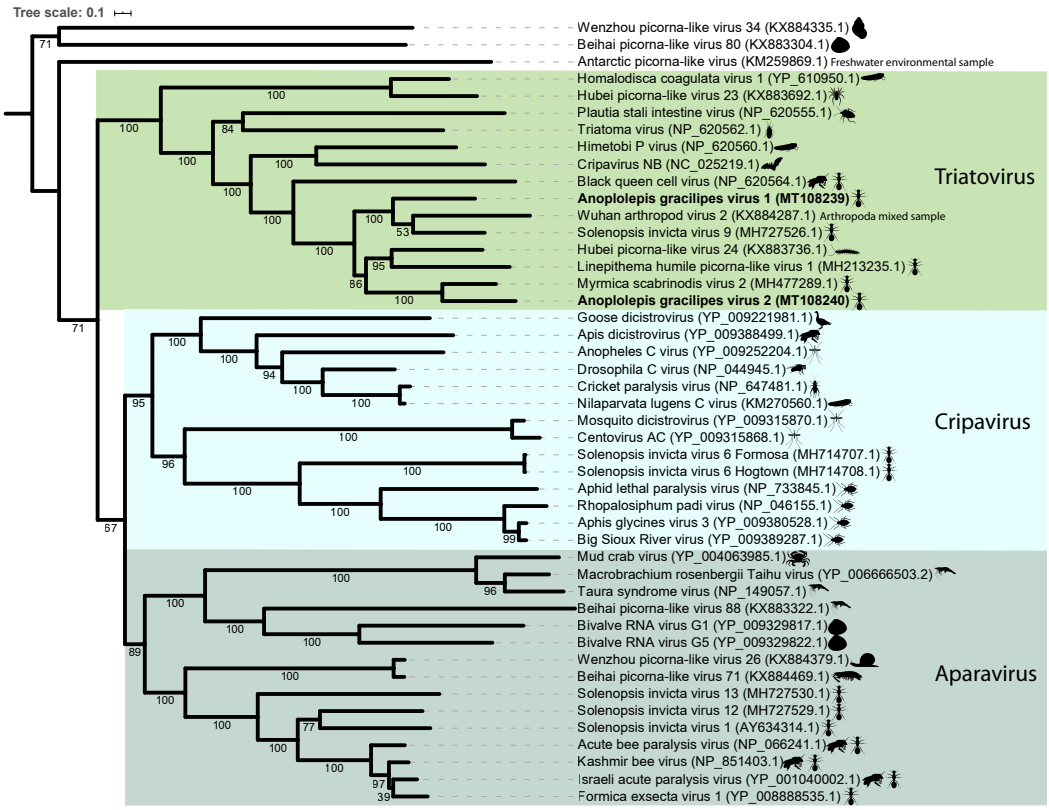


Figure 5.4 Phylogeny of Anoplolepis gracilipes virus 1, 2, and an additional 44 *Dicistroviridae* viruses based on non-structural polyprotein (ORF1). The phylogenetic tree was re-rooted at the clade of Wenzhou picorna-like virus 34 and Beihai picorna-like virus 80. The number at each node indicates the bootstrap probability (%). The accession number is provided in parenthesis. Silhouettes following the accession number indicate the host in which each virus was originally described.

5.4 References

1. Valles S.M.; Chen Y.; Firth A.E.; Guérin D.M.; Hashimoto Y.; Herrero S.; de Miranda J.R.; Ryabov E. ICTV Virus Taxonomy Profile: Dicistroviridae. *J. Gen. Virol.* **2017**, *98*,355-356.
2. Kleanthous E.; Olendraite I.; Lukhovitskaya N.I.; Firth A.E. Discovery of three RNA viruses using ant transcriptomic datasets. *Arch. Virol.* **2019**, *164*,643-647.
3. Valles S.M.; Rivers A.R. Nine new RNA viruses associated with the fire ant *Solenopsis invicta* from its native range. *Virus Genes* **2019**, *55*,368-380.
4. Viljakainen L.; Holmberg I.; Abril S.; Jurvansuu J. Viruses of invasive Argentine ants from the European Main supercolony: Characterization, interactions and evolution. *J. Gen. Virol.* **2018**, *25*,1129-40.
5. Reuter G.; Pankovics P.; Gyöngyi Z., Delwart E.; Boros Á. Novel dicistrovirus from bat guano. *Arch. Virol*. **2014**, *159*,3453-6.
6. Shi M.; Lin X.D.; Tian J.H.; Chen L.J.; Chen X.; Li C.X.; Qin X.C.; Li J.; Cao J.P.; Eden J.S.; Buchmann J. Redefining the invertebrate RNA virosphere. *Nature* **2016**, *540*, 539-43.
7. Mihara T.; Nishimura Y.; Shimizu Y.; Nishiyama H.; Yoshikawa G.; Uehara H.; Hingamp P.; Goto S.; Ogata H. Linking virus genomes with host taxonomy. *Viruses* **2016**, *8*, 66.
8. Payne A.N.; Shepherd T.F.; Rangel J. The detection of honey bee (*Apis mellifera*) associated viruses in ants. *Sci. Rep.* **2020**, *10*,2923.
9. Cooling M.; Gruber M.A.M.; Hoffmann B.D.; Sébastien A.; Lester P.J. A metatranscriptomic survey of the invasive yellow crazy ant, *Anoplolepis gracilipes*, identifies several potential viral and bacterial pathogens and mutualists. *Insect Soc.* **2017**, *64*, 197-207.
10. Bolger A.M.; Lohse M.; Usadel B. Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics* **2014**, *30*, 2114-2120.
11. Grabherr M.G.; Haas B.J.; Yassour M.; Levin J.Z.; Thompson D.A.; Amit I.; Adiconis X.; Fan L.; Raychowdhury R.; Zeng Q.; Chen Z.; Mauceli E.; Hacohen N.; Gnirke A.; Rhind N.; di Palma F.; Birren B.W.; Nusbaum C.; Lindblad-Toh K.; Friedman N.; Regev A. Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat Biotechnol.* **2011**, *29*,644-652.
12. Buchfink B.; Xie C.; Huson D.H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **2015**, *12*, 59.
13. Huson D.; Beier S.; Flade I.; Gorska A.; El-Hadidi M.; Mitra S.; Ruscheweyh H.; Rewati Tappu D. MEGAN Community Edition - Interactive exploration and analysis of large-scale microbiome sequencing data. *PLoS Comput. Biol*. **2016**, *12*, e1004957.

14. Rombel I.T.; Sykes K.F.; Rayner S.; Johnston S.A. ORF-FINDER: a vector for high-throughput gene identification. *Gene* **2002**, *282*, 33-41.
15. Marchler-Bauer A.; Derbyshire M.K.; Gonzales N.R.; Lu S.; Chitsaz F.; Geer L.Y.; Geer R.C.; He J.; Gwadz M.; Hurwitz D.I.; Lanczycki C.J.; Lu F.; Marchler G.H.; Song J.S.; Thanki N.; Wang Z.; Yamashita R.A.; Zhang D.; Zheng C.; Bryant S.H. CDD: NCBI's conserved domain database. *Nucleic Acids Res*. **2015**, *43*, D222-D226.
16. Darriba D.; Posada D.; Kozlov A.M.; Stamatakis A.; Morel B.; Flouri T. ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. *Mol. Biol. Evol.* **2020**, *37*,291-294.
17. Kozlov A.M.; Darriba D.; Flouri T.; Morel B.; Stamatakis A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **2019**, *35*, 4453-4455.
18. Letunic I.; Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res*. **2019**, *47*, W256-W259.
19. Yang C.C.; Yu Y.C.; Valles S.M.; Oi D.H.; Chen Y.C.; Shoemaker D.W.; Wu W.J.; Shih C.J. Loss of microbial (pathogen) infections associated with recent invasions of the red imported fire ant *Solenopsis invicta*. *Biol. Invasions* **2010**, *3*, 333-345.
20. Culley A.I.L.; Suttle C.A. Family dicistroviridae. In: King AMQA, Carstens EB, Lefkowitz EJ (eds) Virus Taxonomy, Classification and Nomenclature of Viruses. 9th Report of the ICTV Elsevier Academic Press, Amsterdam, 2012, pp 840-854.

Chapter 6: First polycipivirus in the yellow crazy ant

6.1 Introduction

The *Polycipiviridae* is a new virus family of *Picornavirus* [1,2]. The genome is composed of a positive-sense liner RNA of approximately 10-12 kb, including four 5'-proximal open reading frames (ORFs), and one long 3' ORF (ORF5). Three genera including *Sopolycivirus*, *Hupolycivirus*, and *Chipolycivirus* are currently under *Polycipiviridae* [1,2]. Most of the currently known hosts for polycipiviruses are arthropods, especially ants [3-10]. The Solenopsis invicta virus 2 (SINV-2) found in fire ants is a member of *Polycipiviridae*. Although little is known about the pathogeneses of viruses in ants, a previous study reveals the SINV-2 induces fitness cost in infected fire ants by reducing the queen reproduction that consequently impedes the colony founding [11]. Beside the pathogeneses, the previous observation reports the high genetic diversity exists in Linepithema humile polycipivirus 1 than other viruses in Argentine ant. The rapid replication of Linepithema humile polycipivirus 1 seems to form a quasispecies that are optimized to cope with immune responses of its host Argentine ants. The virus quasispecies may facilitate RNA viruses to adapt dynamic environments or incoming hosts [9,12]. The fact that majority of currently known polycipiviruses infect ants with a high level of genetic diversity elevates the importance of further study on polycipiviruses for understanding both virus evolution and the potential threat to the local ant community driven by spillover of invasive ant species.

The yellow crazy ant, *Anoplolepis gracilipes*, is a global distributed invasive ant. To our knowledge, only two dicistroviruses have been reported with complete genome sequences in this ant species [13], along with a dicistrovirus virus-like sequence [14]. The previous report of virus fragments in yellow crazy ant suggests the potential of the unrevealed virome in *A. gracilipes*. High throughput sequencing has accelerated the discovery of new viral pathogens in numbers of species. In this study, I investigated the transcriptome of *A. gracilipes* worker and characterized a new virus genome. Combined with the field survey across a board geographical range, I provide new evidence of polycipiviruses persist high genetic diversity in the yellow crazy ant and a new case of virus-host co-introduced event in invasive ants.

6.2 Materials and methods

6.2.1 Sample collection, RNA purification, and Sequencing

A total of 53 yellow crazy ant colonies collected in 2018 from Malaysia (n=17), Indonesia (n=4), Taiwan (n=27), and Okinawa (n=5). First, for identifying novel viruses, I reanalysis the dataset from [12] in which an RNA-pool sequencing of the combination of two randomly selected individuals from Penang, Malaysia (5°21'32.4"N 100°18'09.0"E) and the other from Okinawa, Japan (26°40'19.2"N 128°00'41.0"E). The RNA purification, in brief, the entire worker ant was soaked and homogenized with a pestle in TRIzol™ RNA Extraction Reagent (Invitrogen, Carlsbad, CA) then followed the standard TRIzol RNA extraction protocol. The RNA-pool was submitted to the Eurofins Genomics (Tokyo, Japan) where the total RNA was purified using the polyA selection method. The sequencing library was constructed with the TruSeq® Stranded mRNA Library Prep with the average 200 bp of the library insert size. The library was sequenced in the pair-end mode (2x100bp) using Illumina HiSeq 4000 platform. Second, for verifying a novel virus infects ants rather than contamination, and estimating the prevalence, I mixed 3-5 worker ants per colony and conducted RNA purification with the same procedure.

6.2.2 de novo Assembly and Virus Characterization

The pair-end reads obtained from Illumina sequencing were trimmed to remove adaptors and low-quality sequences (q<28) using Trimmomatic v0.36 [15]. The *de novo* assembly of transcripts was generated using Trinity v2.8.4 [16]. The assembly was compared against NCBI GenBank non-redundant (nr) protein database using BLASTx implemented in Diamond v0.9.21 [17] with e-value < 1e-40. The virus-like transcripts were identified using MEGAN Community Edition v6.18.8 [18]. The open reading frames (ORFs) of putative viral-like transcripts were pinpointed using NCBI ORF finder [19] and conserved domains were discriminated using HHpred [20] against Pfam [21] and PDB [22] databases. To confirm whether virus-like sequences were novel viruses, pairwise comparison of the amino acid identities of the entire ORF5 (non-structural polyprotein) against 28 polycipiviruses obtained from NCBI GenBank were performed using BLASTp.

6.2.3 Virus Verification

To verify the occurrence of the identified virus in *A. gracilipes* in the field, virus primers for PCR amplification were designed using Primer3 [23]. The cDNA synthesis with oligo dT primer used PrimeScriptTM 1st strand cDNA Synthesis kit (Takara) under conditions as follows: 30°C for 10 min., 42 °C for 60 min., and final inactivation at 95°C for 5 min. The cDNA then served as a template for PCR reaction targeted on RNA dependent RNA polymerase (RdRp) region with primers (AgrV3_RdRp.for: 5'-ACTGGGACTCTCCCGAAAAC-3' and AgrV3_RdRp.rev: 5'-TATGTCCATCAAGCGATCCA-3') and carried out with EmeraldAmp® MAX PCR Master Mix (Takara). The PCR cycling profile as follows: 94 °C for 3 min, 35 cycles of

94 °C for 30 s, 60 °C for 30 s, and 72 °C for 60 s, with a final extension step of 72 °C for 5 min. The PCR amplicons were confirmed by both gel electrophoresis and Sanger sequencing. The PCR used a verified virus infection sample and a host (*A. gracilipes*) gene *ef1-alpha* as the positive control with primers Agr_EF1a.for (5′-CCTGGGTGTTGGACAAACTT-3′) and Agr_EF1a.rev (5′-GCTCCTTCACCGAGATGTTC-3′).

6.2.4 Phylogenetic and Evolutionary Analysis

To resolve the taxonomic classification of the virus, I collected sequence information of up-to-date polycipiviruses containing full non-structural polyprotein (ORF5) sequences from GenBank. A total of 28 polycipiviruses (3 *Chipolycivirus*, 1 *Hupolycivirus*, 14 *Sopolycivirus*, and additional 10 unclassified polycipiviruses) were reconstructing the phylogeny with additional three Iflaviruses as the out-group. Amino acid sequences of the entire non-structural polyprotein (ORF5) were aligned using MUSCLE implemented in MEGA 7 [24]. The best evolutionary model calculated by ModelTest-NG v0.1.6 [25], and the LT+I+G4 model was determined by the Akaike information criterion and Bayesian information criterion. The phylogenetic tree based on the maximum likelihood method was reconstructed using RAxML-NG v0.9.0 [26] with 1,000 bootstrap replicates. The tree convergence verified using weight Robinson-Foulds (WRF) distance < 3% as the cutoff. The Bayesian inference method conducted using MrBayes v3.2.7 [27] with 1,000,000 generations that first 25% of generations as burnin, and sampled in every 500 generations. The substitution model for MrBayes [27] using VT+I+G4 that estimated by ModelTest-NG v0.1.6 [25]. The convergent diagnostic verified by a potential scale reduction factor (RSRF) =1 and estimated sample size (ESS) >100. The convergent phylogenetic tree then visualized in iTOL v5 [28] and rooted at Iflaviruses.

To understand the geological divergence of the novel virus, 13 nucleotide sequences of virus RdRp region from 11 ant colonies (Malaysia n=5, Taiwan n=2, Okinawa n=2, and Indonesia n=2), six Trinity assembled transcripts, and two Sopolycivirus (Lasius neglectus virus 1 (Genbank: ASK12200.1) and Myrmica scabrinodis virus 1 (Genbank: ASK12206.)) were aligned using MUSCLE implemented in MEGA 7 [24]. The best evolutionary model estimated by jModelTest v2.2.10 [29]. The substitution model used the GTR+I+G model determined by Akaike information criterion and Bayesian information criterion. The phylogenetic tree based on the maximum likelihood method reconstructed by MEGA 7 [24] with 1,000 bootstrap replicates and the Bayesian inference method used MrBayes v3.2.7 [27] with 1,000,000 generations that first 25% of generations as burnin, and sampled every 500 generations.

For investigating the virus genetic diversity and demographics, Tajima's D test of 13 field collected ant virus RdRp sequences was estimated using MEGA 7 [24]. The

nonsynonymous substitutions (dN) and synonymous substitutions (dS) were estimated using yn00 implemented in PAML-X [30].

6.3 Results

6.3.1 Characterization of the Viral Genome

The result of BLASTx and MEGAN analysis revealed 34 virus-like transcripts with e-value < 1e-40 from a total of 70,145 trinity transcripts generated by Trinity. The virus-like transcripts included 11 *Dicistroviridae*, 3 *Inflaviridae*, 18 *Polycipiviridae*, and 2 unclassified *Riboviria* viruses. I found two of 18 *Polycipiviraidae* like transcripts identical in the length of 11,644bp (excluding poly-A tail), and 16 other *Polycipiviraidae* like transcripts possessed >80% nucleotide identity compared to each of two nearly identical transcripts. These two longest transcripts contained same features of five main ORFs: ORF1 (nucleotide (nt) 218-1063; 281 amino acids (aa)), ORF2 (nt 1060-1920; 286 aa), ORF3 (nt 1920-2747; 275 aa), ORF4 (nt 2744-4156; 470 aa), and ORF5 (nt 4438-11,196; 2252 aa). Besides, a short ORF2b (nt 1106-1432; 108 aa) located within the ORF2 (Figure 1). The alignment of the two longest transcripts presented 27 synonymous mutations in ORF5. The amino acid sequences of 6 ORFs were identical between the two longest transcripts. The results of the HHpred search predicted ORF1, 2, 3 and ORF4 were putative capsid proteins, and the ORF2b was putative a protein associated with protein translocation. The ORF5 contained RNA helicase (nt 6,613-6,922), Protease (nt 8,881-9,505), and RNA dependent RNA polymerase (9,649-11,158bp). The top hit of ORF5 compared against the NCBI nr database was Lasius neglectus virus 1 (Genbank: ASK12200.1) with amino acid identity as 58% (1,318/2,282). The pairwise comparison of the amino acid sequence of ORF5 against the other 27 current known polycipiviruses returned the same result that the highest pairwise identity was 58% when comparing to Lasius neglectus virus 1.

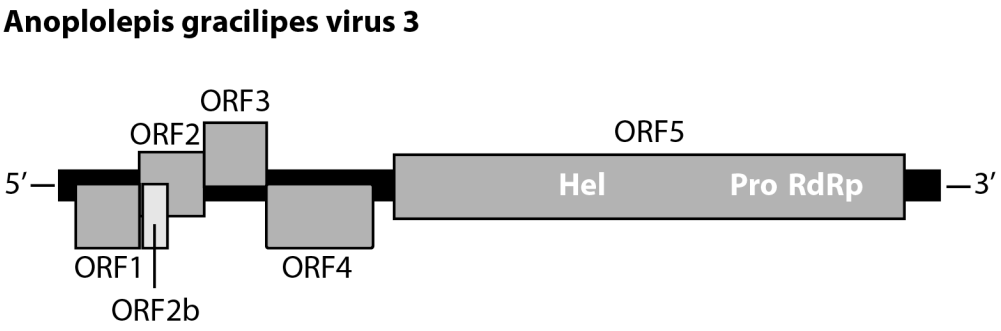


Figure 6.1 Genome structure of Anoplolepis gracilipes virus 3. ORFs are represented by grey rectangles with vertical offsets indicating reading frames (-1, 0, +1) relative to

the Hel-Pro-RdRp ORF (ORF5). Hel: RNA helicase, Pro: Protease, RdRp: RNA dependent RNA polymerase.

6.3.2 Phylogenetic Analysis and Virus Classification

Phylogenetic analysis based on amino acids of the entire ORF5 of the virus-like transcript and other 28 Polycipiviridae and 3 Iflaviridae viruses presented equivalent tree topology by both maximum likelihood and Bayesian inference methods. The virus-like transcript in the present study was closely related to other members in the genus of *Sopolycivirus* (Figure 6.2). Besides, our phylogenetic analysis meanwhile classified those previously published yet unassigned Polypiviridae viruses as follows: the genus of *Chipolycivirus*: Yongsan picorna-like virus 3 (host: mosquito, *Ades vexans nipponii*)[14], the genus of *Hupolycivirus*: Kandapolycivirus (host: bat, *Pteropus lylei*) [13], and Apple picorna-like virus 1 (host: apple leaf, *Malus domestica, Honeycrisp G.935*)[16].

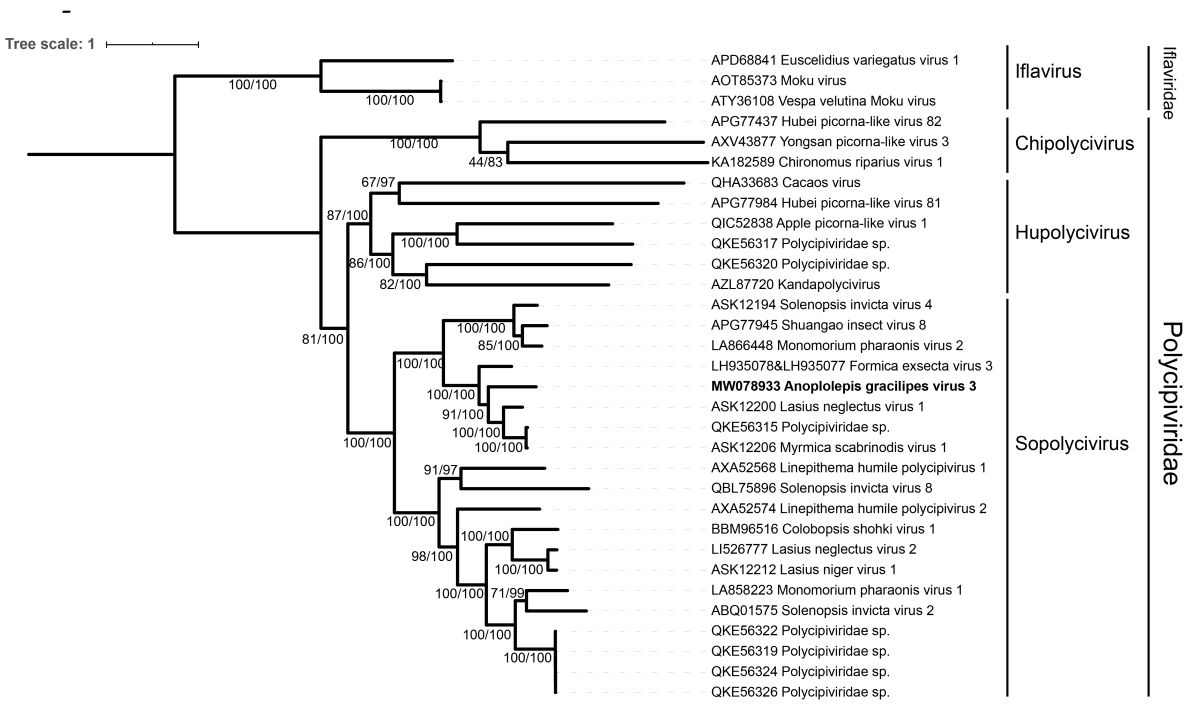


Figure 6.2 Phylogeny of Anoplolepis gracilipes virus 3 and other 29 *Polycipiviridae* and 3 *Iflaviridae* viruses based on the ORF5. The phylogenetic tree was constructed by the maximum likelihood and Bayesian inference methods. The numbers at each node indicate the bootstrap supporting value and the posterior probability. I rooted phylogenetic tree using three *Iflaviridae* viruses. The GenBank accessions denoted in the front of virus species.

6.3.3 Field prevalence and evolutionary analysis

The survey of the occurrence of a virus-like transcript in yellow crazy ant colonies across four countries revealed the highest prevalence reached 88% in Malaysia (n= 15/17), and followed up were 50% in Indonesia (n= 2/4), 44% in Taiwan (n=12/27), and 40% in Okinawa (n= 2/5).

Phylogenetic analysis based on the RdRp region (partial ORF5) of 6 Trinity transcripts and 13 virus sequence from 11 field colonies showed transcripts from our assembly were all closely related to virus strains from Malaysia rather than Okinawa, where two origins of our RNA-pool mixture. Interestingly, two viral sequences from the same ant colony (Indonesia_2-1 & Indonesia_2-2) were different strains. The overall Tajima's D was 0.3969 with the genetic diversity (pi) as 0.0796. The average nonsynonymous substitutions (dN) was 0.0115, synonymous substitutions (dS) was 0.4988, and the dN/dS=0.023.

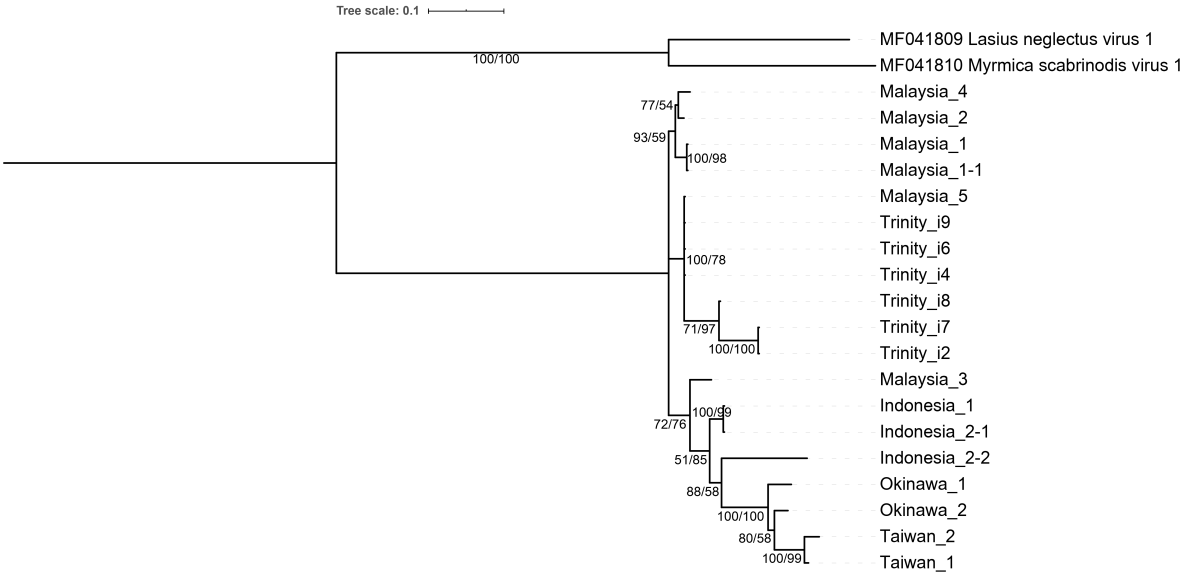


Figure 6.3 Phylogeny of Anoplolepis geacilipes virus 3 based on RdRp region from 11 filed sequences and 6 Trinity transcripts. The phylogenetic tree was constructed by the maximum likelihood and the Bayesian inference methods. The numbers at each node indicate the bootstrap supporting value and the posterior probability. The tree was rooted using Lasius neglectus virus 1 and Myrmica scabrinodis virus 1.

6.4 Discussion

I report a new virus with the complete genome sequence through a transcriptomic approach in the yellow crazy ant, which is the first virus in *Polycipiviridae* associated with this ant and meanwhile the third virus discovered in *A. gracilipes*. Our analysis of

the genomic structure (ORFs), pairwise amino acid identity, and phylogeny support this new virus is a novel member of *Sopolycivirus*. Based on the species demarcation criteria established by ICTV [2], no currently known Polycipiviridae virus [3-10] presented >90% amino acid identity to this new virus-like transcript basing on the ORF5, thus I propose a new virus and tentatively name "Anoplolepis gracilipes virus 3".

Our newly discovered Anoplolepis gracilipes virus 3 (hereafter: AgrV-3) contains 6 ORFs that putative ORFs function and orientation (Figure 1) resemble other polycipiviruses [1,2]. The phylogenetic analysis of the entire ORF5 compared to other 28 well-classified or unclassified polycipiviruses and 3 Inflaviruses by both maximum likelihood and Bayesian inference methods provides robust evidence in supporting AgrV-3 is a member of Sopolycivirus (Figure 2). The virus family *Polycipiviridae* is a relatively new (2017) viral classification under the order of *Picornavirales* [1]. This family is positive-sense single-stranded RNA viruses that infect arthropods (mainly discovered in ants) [3-10]. Interestingly, polycipiviruses in the genus of Hupolycivirus seems to be an exception that Apple picorna-like virus 1 was detected in the leaf tissue of apple trees [10], and Kandapolycivirus infects bats [7], suggesting the potential of exploitation of diverse host by polycipiviruses.

Our de novo assembly by Trinity constructed 2 full length (included 5'-end untranslated region, 6 ORFs, and polyA tail) and 16 contigs with high nucleotide identity (>80%) against each of two full-length contigs. As the assembly originates from an RNA-pool, I investigate those isoforms derived from Malaysia or Okinawa by comparing the RdRp region of 6 isoforms to those sequences from field ant colonies obtained by Sanger sequencing. Although the sample size is small, the geographical difference is revealed by the phylogeny (Figure 3). The phylogenetic tree, unexpectedly, implies all transcripts are closely related to those Malaysian origin viral strains. The result hints that multiple virus strains derived from the same viral species co-harbored in a single ant individual. Besides, an unusual case of two distinct AgrV-3 strains were detected in the same ant colony (Indonesia_2-1 & Indonesia_2-2), supporting the possibility of a single *A. gracilipes* colony can harbor multiple AgrV-3 strains. The quasispecies concept of the RNA virus is one of the potential explanations for the finding of multiple RNA virus strains replicating in a host. The rapid and error-prone replication of the RNA virus elevates the mutation rate and forms a "cloud" of variants. Theoretically, the diverse variants tend to adapt dynamic environments rapidly, including escaping from host immune or exploitation of a new host [4,12]. Multiple strains of AgrV-3 from an individual or an ant colony may be a consequence of error-prone replication. The positive value of Tajima's D supports this idea as the observed genetic diversity is greater than the expected genetic diversity for AgrV-3. Further analysis shows the rate of substitutions at silent sites (dS) is 0.4988, indicating

~50% of synonymous sites are polymorphic. Coincidently, our result is comparable to previous study on the Sopolycivirus (Linepithema humile polycipivirus 1) that presents the highest dS among other viruses (Kashmir bee virus and Linepithema humile C-virus 1) in Argentine ant [10]. Also, the result showed that AgrV-3 in yellow crazy ants possesses a similar dN/dS ratio (0.023) to Linepithema humile polycipivirus 1 (dN/dS =0.020). Although the dN/dS ratio below one suggests that purifying selection potentially acts on the RdRp region, it is not inconsistent to the quasispecies concept that the fitness landscape may currently have optimal fitness for AgrV-3 in yellow crazy ants. An interesting question arises concerning such a coincidence between two sopolyciviruses: whether the rapid replication and diverse variants are a common feature in Sopolycivirus? More investigation on this newly established virus family is needed in the near future.

Beside the AgrV-3, the dataset also includes two previously discovered dicistrovirus Anoplelepis gracilipes virus 1 and 2 [12], and other contigs resemble Dicistroviridae, Inflaviridae, and unclassified Riboviria viruses. The survey of our previous study reported the AgrV-1 and AgrV-2 are absent in Okinawa [13], suggesting all three viruses coinfect a single Malaysian origin worker ant. The social trait of the yellow crazy ant might associate with virus coinfection dynamics. The virus coinfection seems more common in polygyne colony compare to the monogyne colony in the fire ant [31], and the trophallaxis in polygyne colony may facilitate the food-borne spread of Sopolycivirus [32]. The rapid growth of polygyne supercolony may has a colony level pathogen buffering mechanism [33] and may lead to improved disease tolerance and as a reservoir for containing a broad pathogen spectrum.

The co-introduction of pathogens and invasive species is the first step of potential pathogen spillover to local community. Our field survey revealed the AgrV-3 presenting in Okinawa, where is believed as the introduced area of the A. gracilipes. Owing to major host of sopolyciviruses are ants, and our observation of potential quasispecies forming in AgrV-3, I speculate the AgrV-3 may have a cryptic impact on the local ant community in the introduced area by the nature of RNA virus may instantly be highly virulent to the incoming host. Our result provides the evidence of new case of virus-host co-introduced event accompany with biological invasion. The further investigation of AgrV-3 prevalence and pathogeneses on *A. gracilipes* and native ant species habits in A. gracilipes introduced area is needed for validating our hypothesis.

6.5 References

1. Olendraite, I.; Lukhovitskaya, N.I.; Porter, S.D.; Valles, S.M.; Firth, A.E. Polycipiviridae: a proposed new family of polycistronic picorna-like RNA viruses. *J. Gen. Virol.* **2017**, *98*, 2368.

2. Olendraite, I.; Brown, K.; Valles, S.M.; Firth, A.E.; Chen, Y.; Guérin, D.M.; Hashimoto, Y.; Herrero, S.; de Miranda, J.R.; Ryabov, E.; Consortium, I.R. ICTV virus taxonomy profile: Polycipiviridae. *J. Gen. Virol.* **2019**, *100*, 554.
3. Valles, S.M.; Strong, C.A.; Hashimoto, Y. A new positive-strand RNA virus with unique genome characteristics from the red imported fire ant, Solenopsis invicta. *Virology* **2007**, *365*, 457-463.
4. Viljakainen, L.; Holmberg, I.; Abril, S.; Jurvansuu, J. Viruses of invasive Argentine ants from the European main supercolony: characterization, interactions and evolution. *J. Gen. Virol.* **2018**, *99*, 1129-1140.
5. Shi, M.; Lin, X.D.; Tian, J.H.; Chen, L.J.; Chen, X.; Li, C.X.; Qin, X.C.; Li, J.; Cao, J.P.; Eden, J.S.; Buchmann, J. Redefining the invertebrate RNA virosphere. *Nature* **2016**, *540*, 539-543.
6. Valles, S.M.; Rivers, A.R. Nine new RNA viruses associated with the fire ant *Solenopsis invicta* from its native range. *Virus Genes* **2019**, *55*, 368-380.
7. Temmam, S.; Hul, V.; Bigot, T.; Hoem, T.; Gorman, C.; Duong, V.; Dussart, P.; Cappelle, J.; Eloit, M. A novel Polycipiviridae virus identified in Pteropus lylei stools. *Microbiol. Resour. Announc.* **2019**, *8*, e01662-18.
8. Sanborn, M.A.; Klein, T.A.; Kim, H.C.; Fung, C.K.; Figueroa, K.L.; Yang, Y.; Asafo-Adjei, E.A.; Jarman, R.G.; Hang, J. Metagenomic analysis reveals three novel and prevalent mosquito viruses from a single pool of Aedes vexans nipponii Collected in the Republic of Korea. *Viruses* **2019**, *11*, 222.
9. Fukasawa, F.; Hirai, M.; Takaki, Y.; Shimane, Y.; Thomas, C.E.; Urayama, S.I.; Nunoura, T.; Koyama, S. A new polycipivirus identified in Colobopsis shohki. *Arch. Virol.* **2020**, *165*, 761-763.
10. Wright, A.A.; Cross, A.R.; Harper, S.J. A bushel of viruses: Identification of seventeen novel putative viruses by RNA-seq in six apple trees. *PLoS One* **2020**, *15*, e0227669.
11. Manfredini, F.; Shoemaker, D.; Grozinger, C.M. Dynamic changes in host–virus interactions associated with colony founding and social environment in fire ant queens (Solenopsis invicta). *Ecol. Evol.* **2016**, *6*, 233-244.
12. Lauring, A.S.; Andino, R. Quasispecies theory and the behavior of RNA viruses. *PLoS Pathog*. **2010**, *6*, e1001005.
13. Lee, C.C.; Lin, C.Y.; Hsu, H.W.; Yang, C.C.S. Complete genome sequences of two novel dicistroviruses detected in the yellow crazy ant, *Anoplolepis gracilipes Arch. Vriol.* (in press).
14. Cooling, M.; Gruber, M.A.M.; Hoffmann, B.D.; Sébastien, A.; Lester, P.J. A metatranscriptomic survey of the invasive yellow crazy ant, Anoplolepis gracilipes, identifies several potential viral and bacterial pathogens and mutualists. *Insectes Soc.* **2017**, *64*, 197-207.
15. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114-2120.
16. Grabherr, M.G.; Haas, B.J.; Yassour, M.; Levin, J.Z.; Thompson, D.A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q.; Chen, Z. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **2011**, *29*, 644-652.
17. Buchfink, B.; Xie, C.; Huson, D.H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **2015**, *12*, 59-60.

18. Huson, D.H.; Beier, S.; Flade, I.; Górska, A.; El-Hadidi, M.; Mitra, S.; Ruscheweyh, H.J.; Tappu, R. MEGAN community edition-interactive exploration and analysis of large-scale microbiome sequencing data. *PLoS Comput. Biol.* **2016**, *12*, e1004957.
19. Rombel I.T.; Sykes K.F.; Rayner S.; Johnston S.A. ORF-FINDER: a vector for high-throughput gene identification. *Gene* **2002**, *282*, 33-41.
20. Söding, J.; Biegert, A.; Lupas, A.N. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* **2005**, *33*, W244-W248.
21. Finn, R.D.; Coggill, P.; Eberhardt, R.Y.; Eddy, S.R.; Mistry, J.; Mitchell, A.L.; Potter, S.C.; Punta, M.; Qureshi, M.; Sangrador-Vegas, A.; Salazar, G.A. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* **2016**, *44*, D279-D285.
22. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The protein data bank. *Nucleic Acids Res.* **2000**, *28*, 235-242.
23. Untergasser, A.; Cutcutache, I.; Koressaar, T.; Ye, J.; Faircloth, B.C.; Remm, M.; Rozen, S.G. Primer3—new capabilities and interfaces. *Nucleic Acids Res.* **2012**, *40*, e115-e115.
24. Kumar, S.; Stecher, G.; Tamura, K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **2016**, *33*, 1870-1874.
25. Darriba, D.; Posada, D.; Kozlov, A.M.; Stamatakis, A.; Morel, B.; Flouri, T. ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. *Mol. Biol. Evol*. **2020**, *37*, 291-294.
26. Kozlov, A.M.; Darriba, D.; Flouri, T.; Morel, B.; Stamatakis, A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **2019**, *35*, 4453-4455.
27. Ronquist, F.; Huelsenbeck, J.P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **2003**, *19*, 1572-1574.
28. Letunic, I.; Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* **2019**, *47*, W256-W259.
29. Darriba, D.; Taboada, G.L.; Doallo, R.; Posada, D. jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* **2012**, *9*, 772-772.
30. Xu, B.; Yang, Z. PAMLX: a graphical user interface for PAML. *Mol. Biol. Evol.* **2013**, *30*, 2723-2724.
31. Allen, C.; Valles, S.M.; Strong, C.A. Multiple virus infections occur in individual polygyne and monogyne Solenopsis invicta ants. *J. Invertebr. Pathol.* **2011**, *107*, 107-111.
32. Hashimoto, Y.; Valles, S.M. Infection characteristics of Solenopsis invicta virus 2 in the red imported fire ant, Solenopsis invicta. J. Invertebr. Pathol. **2008**, 99, 136-140.
33. Cremer, S.; Pull, C.D.; Fuerst, M.A. Social immunity: emergence and evolution of colony-level disease protection. *Annu. Rev. Entomol.* **2018**, *63*,105-123.

Chapter 7: Conclusion

In this thesis, I investigated yellow crazy ant and its intracellular microorganisms (Wolbachia, and three novel picornaviruses) through genome-wide approaches. The objective of this thesis is to investigate 1) the pattern of adaptive evolution during the ant's colonization and 2) the interaction and coevolution of the ant and its associated microorganisms. I summarized my results as follows:

In Chapter 2, I documented the *A. gracilipes* mitogenome using next-generation sequencing methods. The circular mitogenome of *A. gracilipes* was 16,943 bp included 13 protein-coding genes, two ribosomal RNA genes, 22 transfer RNAs, and a single large non-coding region of 893 bp. The base composition was AT-biased (72%). Three genomic rearrangements compared to ancestral insects were found. Phylogenetic analysis based on the concatenated nucleotide sequences of the 13 protein-coding genes supports *A. gracilipes* belonging to the Formicinae subfamily. I announce the *A. gracilipes* mitogenome as a DNA reference for further population genetic, phylogenetic and evolutionary analyses.

In Chapter 3, to investigate whether the adaptive evolution occurs during the colonization, I assessed the population structure of 73 ant colonies across 13 geographical regions (including Southeast Asia, East Asia, Australia, and several Indo-Pacific islands) based on 11,476 polymorphic genomic markers. The population structure analysis and migration surface estimation revealed admixture genetic groups form a gene flow corridor across Southeast, East Asia, and Indo-Pacific islands. I found a genetic barrier at the tip of the Malay Peninsula which separated *A. gracilipes* into two major phylogenomic clades that may origin from two ancestral supercolonies. One of phylogenomic clades restricted in the tropics and the other successfully sustained in the subtropics. Using outlier approaches, I further pinpointed 17 genome regions under positive selection that is likely linked to genes involved in several signaling pathways and transposable elements. Altogether, our results provide new insights for a potential adaptive evolution in an invasive ant.

In Chapter 4, I reported a nearly 100% prevalence of *Wolbachia* in global populations of the yellow crazy ant. To understand coevolutionary history between *Wolbachia* and *A. gracilipes*, I identified genomic SNPs in *Wolbachia* from the ant across 12 geographical regions and compared the phylogeny of *Wolbachia* based on the SNPs to patterns of the ant's mitochondrial DNA (mtDNA) variation. I revealed a strong concordance between phylogenies of *Wolbachia* and host mtDNA, providing immediate evidence of co-divergence. Among eight identified SNP loci separating the genetic clusters of *Wolbachia*, seven loci are located in potential protein-coding genes, and three of which are non-synonymous SNPs that may influence gene functions. I

found a *Wolbachia* hypothetical protein gene with signature of positive selection. These findings jointly allow us to characterize *Wolbachia*-ant coevolution and also raise a question about mechanism(s) underlying maintenance of high prevalence of *Wolbachia* during the colonization of this invasive ant.

In Chapter 5 and 6, I used the high throughput RNA sequencing and analyzed RNA virome in *A. gracilipes* and revealed virus-like transcripts related to Dicistroviridae, Inflaviridae, Polycipiviridae, and unclassified Riboviria viruses. I reported two novel dicistroviruses composed by 10,662 and 8,238 nucleotides in length, respectively, with both possessing 2 open reading frames with 3 conserved protein domains, as characteristic of the genus Triatovirus in the family Dicistroviridae. The two novel dicistroviruses were tentatively named "Anoplolepis gracilipes virus 1" and "Anoplolepis gracilipes virus 2", respectively. Moreover, I depicted a new polycipivirus genome composed of 11,644 nucleotides with possessing 6 open reading frames. The evidence of genome structures, phylogenetic analysis, and pairwise amino acid identity supported it is a new member of Sopolycivirus under Polycipiviridae. The polycipivirus was tentatively named as "Anoploleis gracilipes virus 3". The polycipiviruse is noticeable that its majority hosts are ants. The polycipiviruses are a few of the known viral pathogens that reduced ant fitness. I found the yellow crazy ants carried over all three novel viruses to introduced areas. The evolutionary analysis of the genetic diversity and the substitution rate exposed a high level of polymorphisms of Anoploleis gracilipes virus 3. The evidence of multiple virus stains in an ant colony providing a possibility of the quasispecies formation. Our results document new cases of a co-introduced pathogens with its host during the biological invasion.

Altogether, I revealed the population structure and history of the yellow crazy ant with several candidate genes associated with the signal of positive selection. The fixation of these genes may seem as a consequence of adaptive evolution during the ant colonization. A unique relationship of *Wolbachia* and ants was established and opens a new gate for investigating the role of *Wolbachia* in ants. Despite the lack of knowledge about pathogens in *A. gracilipes*, I reported multiple new viruses through comprehensive virome analysis. In addition, I provide new genome resources including a *A. gracilipes* draft genome, a complete mitochondrial genome of the yellow crazy ant, a new *Wolbachia* draft genome (strain: wAgra), and three complete virus genomes for further researches targeting on the yellow crazy ant and it associated microorganisms. This thesis provides baseline information that practically fills the knowledge gaps regarding how colonization success has been shaped in this ant via evolutionary and microorganism perspective.

Acknowledgement

First and foremost, I have to express my deepest gratitude to my two supervisors, Dr. Kenji Matsuura and Dr. Chin-Cheng Scotty Yang. Without their assistance and dedicated involvement in every step throughout the whole process, this thesis would never be accomplished. Dr. Kenji Matsuura has inspired and illuminated my motivation for scientific researches and provided crucial advice and valuable comments on my research topics. I am sincerely grateful to Dr. Chin-Cheng Scotty Yang, who never stops leading me and giving me research opportunities. Besides, I appreciate Dr. Chin-Cheng Scotty Yang's unwavering support with great patience at all times. I thank my committee, Drs. Ryohei Terauchi, Tsuyoshi Yoshimura, Yoshitaka Takano, and Takaaki Daimon for your valuable suggestions and comments throughout my thesis.

I am particularly grateful for my Bioinformatics mentor, Dr. John Wang. Without his support and encouragement, this thesis would not have been possible. I would like to express my appreciation to Dr. Chow-Yang Lee, Dr. Chung-Pin Lin, and Dr. Viet Dai Dang for their invaluable contribution to my field research. I am deeply indebted to Dr. Chun-Yi Lin, Dr. Hung-Wei Hsu for their invaluable contribution to my research.

I would like to thank the Japan-Taiwan Student Association for the scholarship that supports my study and life. I am particularly grateful to our secretaries Ms. Tomoko Nakanishi and Ms. Chikako Kitagawa for their reliable assistance and for providing me necessary and first-hand information. I extremely thank my tutors Mr. Naoto Idogawa and Mr. Ryoga Otake for their countless help.

I would like to extend my sincere thanks to Dr. Shu-Ping Tseng, Mr. Po-Wei Hsu, and Mr. Kunpeng Liu for their friendship, helpfulness, and joyfulness. Also, I thank all lab members in the Laboratory of Insect Ecology for useful suggestions. I would like to acknowledge Ms. Han-Chih Ho for her pilot works on this project.

Final, many thanks to my family and my friends for their relentless support. Thank you for never letting me down across all of the challenging times.