EEOB 563 Final Paper – *Ficus petiolaris* fig wasps' Olfactory Receptor (OR) gene family evolution

Devin Molnau

4/18/2019

Introduction:

The fig tree and its fig wasps pollinators are an example of a highly coevolved mutualistic system with each fig species producing a unique blend of volatile chemicals to attract its specific wasp pollinator to receptive inflorescences, and each pollinator species recognizing the volatile blend of its specific host (Weiblen, 2002). This specificity is not typically a result of rare chemical compounds being produced, but rather unique blends and ratios of common volatiles. The pattern of volatiles is sensed by pollinating fig wasps using the five major chemosensory gene families in insects. Of these gene families, it is the odorant receptors (OR), ionotropic receptors (IR), odorant binding proteins (OBP) and chemosensory proteins (CSP) in antennal tissues that are involved in the detection of volatile molecules. The presence of these four olfactory multi-gene families has been confirmed in an Old World pollinating fig wasp (*Ceratosolen solmsi*), with whole genome sequencing revealing a substantial contraction in OR and OBP gene diversity in this host specialist relative to other Hymenoptera (ants, bees, and wasps) (J.-H. Xiao et al., 2013).

Of all of the families, the odorant receptor (OR) family is the largest and most diverse. Insects typically have anywhere from 60-400 OR gene family members, of which can be divided into three distinct groups: ligand-selective ORs that respond to general odorants, ligand-sensitive ORs that are pheromone specific, and a single highly conserved obligate co-receptor known as Orco (Nakagawa, Pellegrino, Sato, Vosshall, & Touhara, 2012). As an example, in the well-studied Hymenopteran model system, the parasitic wasp *Nasonia vitripennis,* has 301 odorant receptors (75 of which are pseudogenes). Whereas, in contrast, the Old World fig wasp *Ceratosolen solmsi* has only 46 OR family members and 2 pseudogenes (J.-H. Xiao et al., 2013).

Table 2
Comprehensive list of chemosensory system genes among insects

	Fig Wasp	Wasp Model	Bee Model	Ant Model				Human Louse
	C. solmsi	*N. vitripennis*	*A. mellifera*	*S. invicta*	*A. pisum*	*T. castaneum*	*P. xylostella*	*P. humanus*
Gr	6 (1)	58 (11)	13 (3)	NA	77 (2)	215 (25)	26	6
Or	46 (2)	301 (75)	174 (1)	297	79 (10)	307 (42)	83	10
Ir	11	10 (1)	9	NA	11	23	49	10
OBP	7	90	21	18	15 (1)	50 (1)	64	5 (1)
CSP	7	9	6	14	11 (1)	20 (1)	20	7 (1)

Xiao et al., 2013

ORs are expressed in the olfactory neurons of the antennae's olfactory sensilla (primarily the basiconic and trichoid sensilla) and are characterized by their seven transmembrane (7-TM) domain with an intracellular N-terminus and extracellular C-terminus (Engsontia, Sangket, Chotigeat, & Satasook, 2014). The one to three odorant receptors function with a highly conserved odorant receptor coreceptor (Orco) to form a ligand-gated cation channel, binding lipophilic volatiles such as aromatics, terpenes, and

fatty acid derivatives (Montagné, De Fouchier, Newcomb, & Jacquin-Joly, 2015) as well as polar ligands such as esters and alcohols (Silbering et al., 2011).

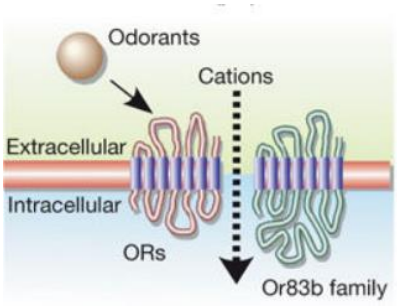


Fig 1. Schematic model of ligand gated ion channel OR+Orco (Or83b) complex (Sato et al., 2008)

Although it is known that the odorant receptor family plays a very important role in the recognition of the receptive fig host, the history of the gene family is not well understood. In terms of a phylogenetic study, it is unclear which of these subfamilies have been retained, completely lost, or expanded upon in light of their fig specialization when compared to other Hymenoptera. Keep in mind that there are also parasites in the system that must recognize the same bouquet of the receptive host as the pollinators. In theory, even though these wasps are not closely related at all to one another, the wasps must recognize the same volatiles in order to locate the host. If the odorant receptor gene families of the parasites and the pollinators of the system are compared phylogenetically, we would expect convergence on which subfamilies have been retained, lost, reduced, or expanded upon. Also, when compared to Hymenopteran species that have not specialized to either pollinate or parasitize the fig, one would also expect that the fig specialists will have comparatively suffered severe gene reductions due to their limited life history.

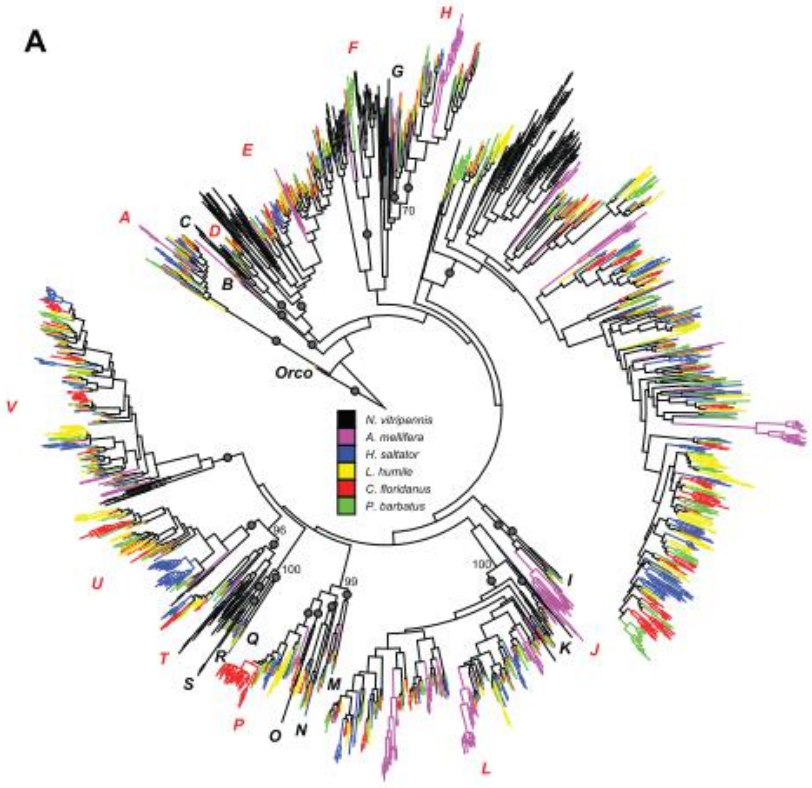
Methods:

To test this, pollinators and parasites of the fig species *Ficus petiolaris* were collected in Baja California, Mexico, August 2017 and stored in RNAlater at 4 degrees Celsius. 150 bp paired end sequencing was done on each fig wasp species in the system on an Illumina HiSeq3000 with a targeted insert length of ~425 bp and 30X coverage. Fig wasps sequenced are *Pegoscapus, Idarnes LO1, Idarnes SO1, Idarnes SO2, Heterandrium 1, Heterandrium 2, Ficicola* (aka *Aepocerus*) and *Physothorax.* Sequences were then trimmed and assembled using Plananus which specializes in assembling highly heterozygotic sequences(Kajitani et al., 2014). The odorant receptors were identified using iterative BLASTX searches with known reference Hymenopteran OR proteins as a database(Mcginnis & Madden, 2004). BLASTX will be using a cutoff of 0.01 and ran via the command-line. The query hits were sorted and culled using the methods described by Zhou et al (Zhou et al., 2012). The Hymenopteran sequences used to build the OR reference database consisted of OR proteins from *Apis mellifera* (western honey bee), *Nasonia vitripennis* (non-fig-related parasitic wasp), *Camponotus floridanus* (Florida Carpenter ant), *Harpegnathos saltator* (Indian jumping ant), *Linepithema humile* (argentine ant), and *Pogonomyrmex barbatus* (harvester ant)(H. M. Robertson, Gadau, & Wanner, 2010; Hugh M Robertson & Wanner, 2006; C. D. Smith et al., 2011; C. R. Smith et al., 2011; Zhou et al., 2012a). Putative genes were verified using the gene prediction software AUGUSTUS and *Nasonia vitripennis* as the fig wasps' model organism(Stanke, Keller, Gunduz, Hayes, & Waack, 2006).

In the event that the assembly does not yield reasonable sized proteins, I will use some older transcriptomic data that may yield better results. Again, odorant receptors will be identified using BLASTX on a remote machine using reference Hymenopteran OR proteins to build a database. All queries will be culled and then putative genes verified using AUGUSTUS which will give us a putative protein sequence which can then be aligned with the references.

These putative proteins were aligned with the known Hymenopteran OR genes using MAFFT and a phylogenetic tree built using maximum likelihood methods in RAxML using the Le-Gascuel (LG) model and

100 bootstrap iterations (Katoh, Kuma, Toh, & Miyata, 2005; Stamatakis, 2014). These species are in no way closely related, but all recognize the host *F.petiolaris* and this should be reflected in the genes clades that are retained or expanded upon. Ideally, time abiding, I will also run build this tree in BEAST using BEAUti and compare the log likelihood of the maximum likelihood to its Bayesian generated counterpart (Drummond, Suchard, Xie, & Rambaut, 2012). Clades will be assessed using the nomenclature established by Zhou et al. in their 2012 paper (see below) (Zhou et al., 2012a). A basic maximum likelihood tree will be built using the reference hymenopteran OR sequence to make sure my attempt to recreate the methods was reasonable.



B

Number of *Or* genes in well-supported subfamilies

	Nv	Am	Hs	Lh	Cf	Pb
Orco	1	1	1	1	1	1
A	0	3	15	12	8	10
B	0	1	1	1	1	1
C	0	1	1	1	1	1
D	10	0	3	5	5	3
E	32	6	22	31	35	24
F	20	1	1	1	1	8
G	1	3	0	3	3	5
H	1	13	9	6	9	13
9-exon	64	34	118	119	114	139
I	2	1	1	0	0	1
J	0	21	2	2	2	2
K	1	2	2	2	2	2
L	7	57	53	37	41	42
M	0	1	3	3	3	4
N	1	0	3	6	5	3
O	2	0	0	1	2	0
P	0	5	9	10	36	7
Q	0	1	3	2	0	1
R	0	0	2	1	1	1
S	4	0	1	1	0	1
T	21	2	9	8	8	5
U	4	1	35	29	25	25
V	9	6	50	43	47	39

Figure 2. Phylogenetic relationships of Hymenoptera *Or* genes. (A) A maximum-likelihood tree of hymenopteran *Or* genes estimated by using RAxML with Le-Gascuel (LG) model. Reliability of internal nodes was evaluated by 100 bootstrap replicates. Grey round dots indicate well-supported subfamilies (bootstrap value ≥80). Bootstrap values ≥70 are shown for relationships among subfamilies. *Ors* in different species are color-coded as following: *N. vitripennis*, black; *A. mellifera*, purple; *H. saltator*, blue; *L. humile*, yellow; *P. barbatus*, green; and *C. floridanus*, red. Subfamilies with rapid changes in gene copy numbers are highlighted in red. (B) Numbers of hymenopteran *Or* genes in well supported subfamilies.
doi:10.1371/journal.pgen.1002930.g002

Results:

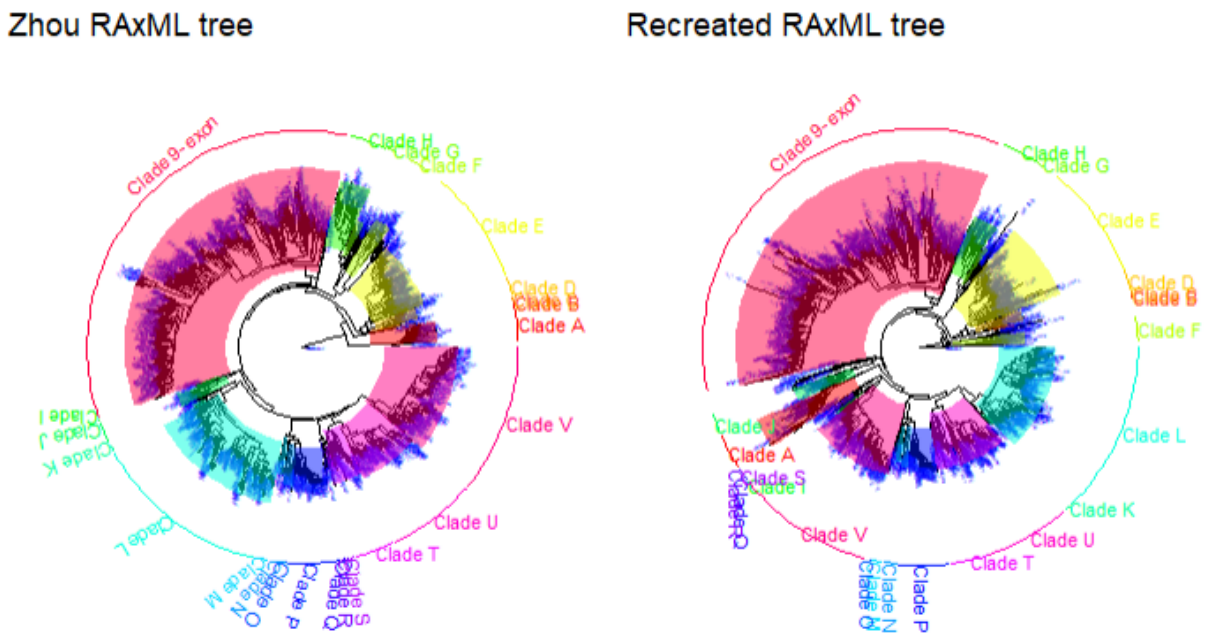


Figure 1: Recreating the reference hymenopteran odorant receptor tree from Zhou et al(Zhou et al., 2012b). Tree of odorant receptors built using MUSCLE and RAxML, bootstrap iteration 100 times and LG4X as a model. The trees are colors according to the clades described in the Zhou paper. The phylogeny includes all known odorant receptors sequences from reference hymenopteran species: from *Apis mellifera* (western honey bee), *Nasonia vitripennis* (non-fig-related parasitic wasp), *Camponotus floridanus* (Florida Carpenter ant), *Harpegnathos saltator* (Indian jumping ant), *Linepithema humile* (argentine ant), and *Pogonomyrmex barbatus* (harvester ant)(H. M. Robertson et al., 2010; Hugh M Robertson & Wanner, 2006; C. D. Smith et al., 2011; C. R. Smith et al., 2011; Zhou et al., 2012a).

Odorant Receptors in the fig wasps of *Ficus petiolaris*							
Pegoscapus	**Aepocerus**	**Het1**	**Het2**	**Idarnes SO1**	**Idarnes SO2**	**Idarnes LO1**	**Physothorax**
8	**13**	**13**	**21**	**35**	**9**	**39**	**18**

Table 1: Table of identified odorant receptor protein sequences in pollinators and non-pollinators in the *Ficus petiolaris* system.

Discussion:

I was able to mimic the methods to the Zhou paper and create a reasonable looking maximum likelihood tree of the reference Hymenoptera (figure 1). BLASTX was performed on fig wasps transcriptomes using these reference Hymenoptera OR proteins as a database. The fasta sequence was extracted from contigs where the match was found and AUGUSTUS created putative protein sequences using Nasonia vitripennis as a model. From this output, it was determined that there was 8 OR proteins in Pegoscapus, 13 OR proteins in Aepocerus, 13 OR proteins in Heterandrium sp.1, 21 OR proteins in Heterandrium sp.2, 35 OR proteins in Idarnes SO1, 9 OR proteins in Idarnes SO2, 39 OR proteins in Idarnes LO1 and 18 OR proteins

in Physothorax (table 1). It is noteworthy that all fig wasps have a lower odorant receptor count than reference Hymenoptera, with their counts more similar to the old-world fig wasp *Ceratosolen solmsi*: *Nasonia vitripennis* (301), *Apis mellifera*(174) ,*S. invicta (*274) , *C.solmsi* (46) (J. H. Xiao et al., 2013). There does seem to be a suspicious drop in odorant receptors in Idarnes SO2 compared to the other Idarnes species which makes me think that Idarnes SO2 might be misidentified. The blasting done for this project was not iterative and therefore that could be many missing OR sequences however. The transcriptomes used were not built using antennae, thus there are most likely some OR missing and will require more thorough investigation. That being said, we would have to be missing quite a bit for us to completely ignore the current trend of fig wasps olfactory gene reductions as a result of their lifestyle.

Below is a species tree showing the families of wasps, their lifestyles, and when they diverged(Peters et al., 2018). The species associated with *F. petiolaris* are *Pegoscapus (*Agaonidae*), Idarnes* (Sycophaginae), *Heterandrium* (Pteromalidae, subfamily Otitesellinae), and *Aepocerus* (Pteromalidae, subfamily Otitesellinae), *Physothorax* wasps(Torymidae). All are accounted for in the species tree below except for *Idarnes* from the Sycophaginae family, of which, there has been some contention on where *Idarnes* falls and if it is the sister taxa to the pollinator or not. It is interesting that the fig wasp parasitoid *Physothorax*, which is more distantly related to the pollinator than *Nasonia vitripennis*, has significantly less odorant receptors. *Heterandrium* is actually part of the same family as *N.vitipennis* and similarly has this apparent reduction of OR proteins. This may be do to the fig wasps' limited lifestyle. Pollinators live only a few days and simply move from their birth fig to the fig that they pollinate, never stopping to eat or explore. Non-pollinators do have a slightly more varied lifestyle, which can be seen by their slightly larger OR families (excluding the suspicious *Idarnes SO1*). Although I am not able to say which gene subfamilies are important yet and not if there is convergent evolution on those subfamilies, we can say there appears to be an overall reduction of OR protein members as a result of the fig wasps of *F. petiolaris* being associated with the fig.

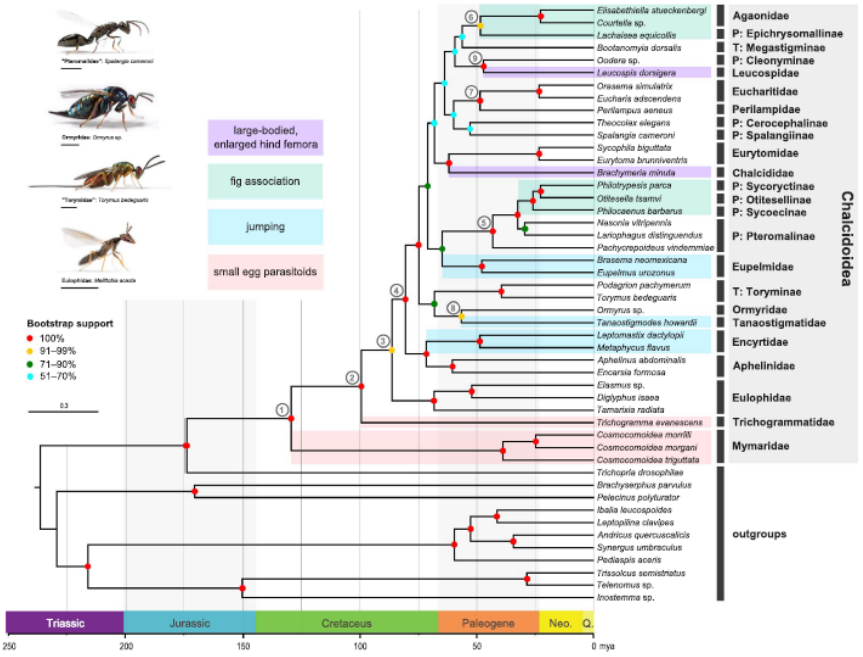


Fig. 1. Phylogenetic relationships and divergence time estimates of Chalcidoidea. The tree was inferred under the maximum likelihood optimality criterion, analyzing 1,469,006 amino acid sites and applying a combination of protein domain- and gene-specific substitution models. Divergence times were estimated with an independent-rates molecular clock approach and considering nine validated fossils. Divergence times shown are those from the analysis having the highest score out of 1,023 dating analyses, which comprise all possible combinations of fossils included and eight different settings for a subset of these fossils. The dating analyses were scored with an *a posteriori* evaluation considering the consistency of the results with the fossilization rates (see methods for more details). P = Pteromalidae, T = Torymidae. Scale bars represent 1 mm. Images by O. Niehuis, with assistance from R.S. Peters. Nodes with circled numbers are referred to in the main text.

Unfortunately, due to time constraints, only the odorant receptor maximum likelihood tree of the reference Hymenoptera was built in time for this paper. Hopefully the tree with the additional pollinator and non-pollinators of *F.petiolaris* will be incorporated in time for my presentation. The genomes that were assembled also appeared to be too fragmented to yield full odorant receptor proteins and thus I switched to using transcriptomes of these wasps. The transcriptomes used were not built using antennae and most likely are missing numerous OR proteins. The odorant receptor proteins that were found can be a part of iterative and more thorough blasts later on when I have longer genomic reads and/or antennae transcriptomes. In addition to this more extensive blasting, I would like to use HMMER to identify any additional OR based on their profile. It does help to have some idea what members are in what clade as it, just incase I would need to break the HMMER search apart by profile. Also, I do want to create a Bayesian tree and compare this to my maximum likelihood tree. I do know of conserved regions in my protein and would like to incorporate this into the maximum likelihood model as well in the future.

Sources:

Drummond, A. J., Suchard, M. A., Xie, D., & Rambaut, A. (2012). Bayesian P hylogenetics with BEAUti and the BEAST 1 . 7 Research article, *29*(8), 1969–1973. https://doi.org/10.1093/molbev/mss075

Engsontia, P., Sangket, U., Chotigeat, W., & Satasook, C. (2014). Molecular evolution of the odorant and gustatory receptor genes in lepidopteran insects: Implications for their adaptation and speciation. *Journal of Molecular Evolution*, *79*(1–2), 21–39. https://doi.org/10.1007/s00239-014-9633-0

Kajitani, R., Toshimoto, K., Noguchi, H., Toyoda, A., Ogura, Y., Okuno, M., … Itoh, T. (2014). Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Research*, *24*, 1384–1395. https://doi.org/10.1101/gr.170720.113.Freely

Katoh, K., Kuma, K., Toh, H., & Miyata, T. (2005). MAFFT version 5 : improvement in accuracy of multiple sequence alignment, *33*(2), 511–518. https://doi.org/10.1093/nar/gki198

Mcginnis, S., & Madden, T. L. (2004). BLAST : at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Research*, *32*, 20–25. https://doi.org/10.1093/nar/gkh435

Montagné, N., De Fouchier, A., Newcomb, R. D., & Jacquin-Joly, E. (2015). Advances in the identification and characterization of olfactory receptors in insects. *Progress in Molecular Biology and Translational Science*, *130*, 55–80. https://doi.org/10.1016/bs.pmbts.2014.11.003

Nakagawa, T., Pellegrino, M., Sato, K., Vosshall, L. B., & Touhara, K. (2012). Amino acid residues contributing to function of the heteromeric insect olfactory receptor complex. *PLoS ONE*, *7*(3). https://doi.org/10.1371/journal.pone.0032372

Peters, R. S., Niehuis, O., Gunkel, S., Bläser, M., Mayer, C., Podsiadlowski, L., … Krogmann, L. (2018). Molecular Phylogenetics and Evolution Transcriptome sequence-based phylogeny of chalcidoid wasps (Hymenoptera : Chalcidoidea) reveals a history of rapid radiations , convergence , and evolutionary success. *Molecular Phylogenetics and Evolution*, *120*(April 2017), 286–296.

https://doi.org/10.1016/j.ympev.2017.12.005

Robertson, H. M., Gadau, J., & Wanner, K. W. (2010). The insect chemoreceptor superfamily of the parasitoid jewel wasp Nasonia vitripennis. *Insect Molecular Biology*. https://doi.org/10.1111/j.1365-2583.2009.00979.x

Robertson, Hugh M, & Wanner, K. W. (2006). The chemoreceptor superfamily in the honey bee , Apis mellifera : Expansion of the odorant , but not gustatory , receptor family, 1395–1403. https://doi.org/10.1101/gr.5057506.

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

Silbering, A. F., Rytz, R., Grosjean, Y., Abuin, L., Ramdya, P., Jefferis, G. S. X. E., & Benton, R. (2011). Complementary Function and Integrated Wiring of the Evolutionarily Distinct Drosophila Olfactory Subsystems. *Journal of Neuroscience*, *31*(38), 13357–13375. https://doi.org/10.1523/JNEUROSCI.2360-11.2011

Smith, C. D., Zimin, A., Holt, C., Abouheif, E., Benton, R., Cash, E., … Fernandes, V. (2011). Draft genome of the globally widespread and invasive Argentine ant (Linepithema humile), *108*(14), 5673–5678. https://doi.org/10.1073/pnas.1008617108

Smith, C. R., Smith, C. D., Robertson, H. M., Helmkampf, M., Zimin, A., Yandell, M., … Graur, D. (2011). Draft genome of the red harvester ant Pogonomyrmex barbatus, *108*(14), 5667–5672. https://doi.org/10.1073/pnas.1007901108

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

Stanke, M., Keller, O., Gunduz, I., Hayes, A., & Waack, S. (2006). AUGUSTUS : ab initio prediction of alternative transcripts. *Nucleic Acids Research*, *34*(16), 435–439. https://doi.org/10.1093/nar/gkl200

Weiblen, G. D. (2002). How to be a fig wasp. *Annual Review of Entomology*, *47*, 299–330.

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

Xiao, J. H., Yue, Z., Jia, L. Y., Yang, X. H., Niu, L. H., Wang, Z., … Huang, D. W. (2013). Obligate mutualism within a host drives the extreme specialization of a fig wasp genome. *Genome Biology*, *14*(12), 1–18. https://doi.org/10.1186/gb-2013-14-12-r141

Zhou, X., Slone, J. D., Rokas, A., Berger, S. L., Liebig, J., Ray, A., … Zwiebel, L. J. (2012a). Phylogenetic and Transcriptomic Analysis of Chemosensory Receptors in a Pair of Divergent Ant Species Reveals Sex-Specific Signatures of Odor Coding. *PLoS Genetics*. https://doi.org/10.1371/journal.pgen.1002930

Zhou, X., Slone, J. D., Rokas, A., Berger, S. L., Liebig, J., Ray, A., … Zwiebel, L. J. (2012b). Phylogenetic and Transcriptomic Analysis of Chemosensory Receptors in a Pair of Divergent Ant Species Reveals Sex-Specific Signatures of Odor Coding. *PLoS Genetics*, *8*(8). https://doi.org/10.1371/journal.pgen.1002930