UC Riverside
UC Riverside Electronic Theses and Dissertations

Title

Gaining Insight into the Formica Social Supergene: Variation in Function and Clues on Evolutionary History

Permalink

https://escholarship.org/uc/item/31k8n4h8

Author

McGuire, Darin Justin

Publication Date

2021

Copyright Information

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Gaining Insight into the *Formica* Social Supergene:
Variation in Function and Clues on Evolutionary History

A Thesis submitted in partial satisfaction
of the requirements for the degree of

Master of Science

in

Entomology

by

Darin Justin McGuire

December 2021

Thesis Committee:
    Dr. Jessica Purcell, Chairperson
    Dr. Hollis Woodard
    Dr. Kieran Samuk

The Thesis of Darin Justin McGuire is approved by:

_____

_____

_____

Committee Chairperson

University of California, Riverside

ACKNOWLEDGEMENTS

ABSTRACT OF THE THESIS

Gaining Insight into the *Formica* Social Supergene:
Variation in Function and Clues on Evolutionary History

by

Darin Justin McGuire

Master of Science, Graduate Program in Entomology
University of California, Riverside, December 2021
Dr Jessica Purcell, Chairperson

Supergenes are chromosomal regions of tightly linked loci experiencing reduced recombination that are inherited as a single unit when heterozygous. Polygenic morphological or behavioral traits can be maintained as distinct morphs, polymorphisms, within a population due to alternative supergene haplotypes that rarely engage in recombination with one another. Supergenes have been found in a variety of organisms resulting in various types of polymorphisms including mimicry complexes in butterflies, mating morphs in birds, and alternative social organization in ants. *Formica* is a genus of ants known for having alternative supergene haplotypes on chromosome 3, resulting in either monogyne or polygyne colonies. Monogyne colonies are single family units headed by one queen, with the associated haplotype being Sm. Polygyne colonies can have multiple queens, with the alternative supergene haplotype, Sp, being associated with this form of social organization. This system, first described in *Formica selysi*, is now known to be present in multiple other Formica species. As we continue to expand our knowledge of the *Formica* supergene, two questions emerge: is the function of the supergene the same across the genus and what genomic events contributed to the evolution of the supergene? To answer the first question, I utilized SNP data from 280 workers

of *Formica neoclara,* sampled from 32 colonies and 8 transect locations spanning from California to British Columbia. I determined that *F. neoclara* is socially polymorphic, with monogyne and polygyne colony assignment stemming from three metrics: inferred queen number, average colony relatedness, and opposing homozygosity. This social polymorphism is due to the supergene found in other *Formica* species through principal component analysis as well as a genome wide association study. Finally, I analyzed the population structure of *F. neoclara*, showing that populations from all sampled regions are experiencing limited gene flow with one another; a finding supported by expected heterozygosity values, principal component analysis, and isolation by distance analysis. Unlike *F. selysi*, in which all individuals within polygyne colonies must have an Sp haplotype, polygyne colonies of *F. neoclara* can consist of workers with all three genotypes (Sm/Sm, Sm/Sp, Sp/Sp), showing that the mode of action of the supergene is not identical across the genus. To answer the second question, regarding the evolution of the supergene, I performed comparative linkage mapping with *F. selysi* and an outgroup genus *Polyergus*. I constructed a linkage map for *Polyergus* using RADseq data from 83 workers from a single colony, with *F. selysi* serving as a reference. While *Polyergus* has synteny with *F. selysi* in many regions of the genome, including a large portion of chromosome 3, there are also many regions with inversions. By performing comparative linkage mapping between genera, we can better understand the timing and conditions surrounding the evolution of the supergene. This thesis, in exploring the elements of the supergene both within a genus and between genera, provides a glimpse into the function and evolution of this social supergene.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER III.

LIST OF TABLES

Chapter I. Introduction

Selectively advantageous phenotypes are determined in part by the genotype; in terms of complex and polymorphic traits, these are often polygenic in nature. Therefore, it would be of maximum benefit to inherit and pass on multiple advantageous alleles as a single unit without the risk of crossover reshuffling and disrupting established favorable combinations. Supergenes are regions of chromosomes with multiple tightly linked loci experiencing reduced recombination. Inherited as a single Mendelian unit when heterozygous, supergenes can result in balanced polymorphisms within populations with alternative haplotypes (Schwander et al., 2014; Thompson & Jiggins, 2014). Potentially spanning megabases (Huang & Wang, 2014; Yan et al., 2020), supergenes can include hundreds of genes within their nonrecombining regions, occupying a large fraction of an organism's genome (Riehle et al., 2017). The term 'supergene' was first coined by Darlington and Mather (1949), though conceptual ideas were being synthesized before this (Fisher, 1930; Dobzhansky & Sturtevant, 1938), being refined into the definition widely accepted today (Dobzhansky, 1970; Schwander et al., 2014; Thompson & Jiggins, 2014). While multiple mechanisms of supergene evolution have been postulated (Mather, 1955; Ford, 1965), we have lacked the tools to test these hypotheses until relatively recently. Nearly one hundred years after a recognizable supergene concept was introduced, new molecular techniques allow us to empirically study supergenes.

Supergenes often exist as balanced polymorphisms within populations, in which the resulting morphs are maintained at a stable frequency in a panmictic population over several generations (Dobzhansky & Pavlovsky, 1960, Dobzhansky, 1970). In fact, sex chromosomes found in a wide range of taxa (Bull, 1983; Charlesworth, 1996) can be classified as supergenes.

Sex chromosomes are thought to evolve from autosomes via mutations that cause preferential development into a male or female; restricted recombination between proto-sex chromosomes would then prevent less-fit recombinants from developing (Charlesworth, 1996; Lewis, 1942). This restricted recombination is often the result of chromosomal inversions, which can link sex-determining genes with alleles that benefit the life history of a particular sex, potentially accumulating multiple alleles in the nonrecombining region (Charlesworth & Charlesworth, 1980; Rice, 1987; Charlesworth et al., 2005; Rubenstein et al., 2019). Autosomal supergenes function similarly to sex chromosomes, they can result in various types of polymorphisms besides sex-related traits. The evolution of supergenes is still an ongoing field of study, with various mechanisms being proposed, such as the "eroded strata" model. In this model, following an event that reduces recombination, the functionally important regions of the supergene are maintained by selection despite rare recombination between the haplotypes (Brelsford et al., 2020). Regardless of their mechanism of evolution, all supergenes start with an initial critical event which results in reduced recombination. Following this event, alleles that form advantageous combinations with the initial loci are then incorporated into the low recombining region, usually through subsequent inversion events, introgression, transposable elements, rare recombination events, and gene conversion (Turner, 1967; Jay et al., 2018; Brelsford et al., 2020; Thompson & Jiggins, 2014; Kirkpatrick, 2010; Brelsford et al., 2020; Lisachov et al., 2015). As previously stated, the evolution of supergenes can result in occurrence of two or more clearly defined morphs in a population: a polymorphism. Potential mechanisms for maintaining a balanced polymorphism in a population where a supergene arises include recessive lethal alleles (Huang & Wang, 2014), disassortative mating (Jay et al., 2021; Tuttle et al., 2016), and frequency dependent selection (Dobzhansky, 1970). Differences between individuals within a population

frequently have a genetic basis. Origins to the differences between males and females remained elusive until the concept of sex chromosomes was elucidated. As time and research into supergenes progresses, we are increasingly discovering instances where supergenes are at the root of polymorphisms within populations.

All ants are eusocial (Peters et al., 2017) and they vary wildly in their ways that their societies are organized. Two common types of social forms are monogyny, colonies headed by single queens, and polygyny, colonies with multiple queens (Hölldobler & Wilson 1990). Monogyny is generally accepted as the ancestral form in ants (Hughes et al., 2008). Aside from queen number, these two social forms also differ in reproductive strategy and worker morphology. Within polygyne colonies, workers and queens are generally smaller than their monogyne counterparts, though colony size is larger in terms of the number of individuals (Keller, 1993; Rosset & Chapuisat, 2007). Moreover, monogyny tends to be associated with long range dispersal while polygyny with short range dispersal or budding (Hölldobler & Wilson, 1990; Sundström, 1995). Budding is a colony founding process in which one or more queens leaves their natal nest, usually with a group of workers, to start a new nesting unit in close proximity to the natal nest (Hölldobler & Wilson, 1990). The two social forms often occur within the same geographic location and coexist as interbreeding populations (Tschinkel, 2006; Gotzek & Ross, 2007). Two ant genera frequently studied due to their social polymorphism are *Solenopsis* and *Formica*.

In *Solenopsis invicta*, the fire ant, the monogyne and polygyne social forms can have differences in terms of worker size, queen size, density, reproduction, and dispersal strategy (Macom & Porter, 1996; Greenberg et al., 1985) in addition to queen number. Seeing as how *S.*

invicta, displays a polymorphism in terms of social organization, it may be no surprise that this species harbors a supergene related to social organization (Wang et al., 2013). Initially described using allozymes (Ross, 1996), researchers identified a genetic polymorphism in Gp-9, a gene that encodes an odorant binding protein (Krieger and Ross, 2005). Later research revealed that Gp-9 is part of a much larger supergene that spans a large portion of the chromosome; the supergene haplotypes in *S. invicta* are now called SB and Sb, respectively (Huang and Wang, 2014). *S. invicta* supergene haplotypes (SB and Sb) are associated with queen number, SB is the monogyne associated haplotype while Sb is polygyne associated (Wang et al., 2013). The *Solenopsis* supergene experiences reduced recombination due to an inversion (Wang et al., 2013) and potentially, accumulation of repetitive elements as well (Huang and Wang., 2014; Stolle et al., 2019). Monogyne colonies only possess SB/SB workers and only tolerate a single SB/SB queen. Polygyne colonies have both SB/SB and SB/Sb workers and will accept multiple queens as long as they have the SB/Sb genotype; polygyne queens without an Sb haplotype are culled via a green beard effect (Keller & Ross, 1998; Ross & Keller, 1998). The green beard operates within polygyne colonies via chemical cues associated with supergene status in queens (Trible & Ross, 2015). Interestingly, Sb/Sb queens are usually non-viable due to recessive lethal alleles (Keller & Ross, 1998); this frequently draws comparisons to the analogous Y-chromosome in mammals (Wang et al., 2013) as the Sb haplotype usually exists only in the heterozygous state in females and the hemizygous state in males (Huang & Wang, 2014). Other reports have shown the presence of Sb homozygous workers and brood in polygyne colonies (Fritz et al., 2006). Sb homozygosity appears to be lethal or deleterious mostly in queens (Hallar et al., 2007), having a negative impact on energy reserves (Ross & Keller, 1998). Not only does *S. invicta* harbor a social supergene but five other socially polymorphic *Solenopsis* species have the same supergene, all

experiencing reduced recombination between haplotypes (Yan et al., 2020). Additionally, the Sb

haplotype has a significantly longer DNA sequence than SB and appears to be experiencing

degenerative expansion in three *Solenopsis* species, including *S. invicta* (Stolle et al., 2019).

While Wang et al. (2013) described the first social supergene in the ant species *S. invicta*, they

also anticipated that social supergenes would exist in other social insects. Wang et al. (2013)

were correct in their prediction; an ancient and widespread social supergene is present in the

ant genus *Formica* (Brelsford et al., 2020). The extensive work done with *S. invicta,* and its social

form associated supergene, displays how supergenes can result in complex polymorphisms in

social form and reproductive strategy. It follows then that supergenes in other socially

polymorphic taxa are not too far-fetched.

      Formica is a widespread genus of ants garnering much attention due to their social

organization. A social supergene initially described in *F. selysi* (Purcell et al., 2014) was

subsequently found to be present in additional *Formica* species (Brelsford et al., 2020; Purcell et

al., 2021). Found on chromosome 3, the supergene shares no gene content with the *S. invicta*

supergene, despite being analogous in structure and function (Purcell et al., 2014). The *Formica*

supergene likely has a long history of rare recombination or gene conversion (Brelsford et al.,

2020), displaying highly differentiated SNPs on the respective haplotypes: Sm and Sp (Purcell et

al., 2014). In *F. selysi*, the Sm haplotype is associated with monogyny while the Sp haplotype is

associated with polygyny. Monogyne colonies are composed of Sm/Sm workers and queens, and

Sm males exclusively. All workers and queens within polygyne colonies have at least one Sp

haplotype; Sm males have not been recorded from polygyne colonies (Purcell et al., 2014; Avril

et al., 2020). There is evidence of a maternal effect killing in which Sm/Sm worker and Sm male

eggs laid by Sm/Sp queens fail to hatch, regardless of worker presence or genotype (Avril et al., 2020). Additionally, *F. selysi* queens differ in their mating habits depending on genotype, gynes harboring the respective supergene haplotypes prefer to mate with males with the same haplotype (Avril et al., 2019). Despite generally preferring assortative mating, all four potential mating combinations are possible in *F. selysi*, with no evidence of genetic incompatibilities (Fontcuberta et al., 2021; Avril et al., 2019). Unlike the fire ant system, *F. selysi* does not appear to harbor recessive lethal alleles on the polygyne associated supergene haplotype (Wang et al., 2013; Purcell et al., 2014) nor does it appear to have a green beard (Keller & Ross, 1998; Avril et al., 2020). This means that both homozygotes (Sm/Sm and Sp/Sp) are viable (Avril et al., 2019) and that the Sp haplotype can recombine, giving it an opportunity to rid itself of deleterious mutations. Recombination within the respective haplotypes may explain why the *Formica* supergene is one of the oldest autosomal supergenes described, spanning about 30 million years of evolutionary history (Brelsford et al., 2020; Purcell et al., 2021). This differs from other supergene systems where one haplotype can rarely or never occur in a homozygous state, leaving it unable to recombine and purge deleterious alleles. For example, the *S. invicta* supergene system is relatively young, being less than half a million years old (Yan et al., 2020). Study of the social supergene initially described in *F. selysi* (Purcell et al., 2014), has expanded to multiple *Formica* species (Brelsford et al., 2020; Purcell et al., 2021), with more on the way.

Formica neoclara is an ant native to western North America, frequently found in sandy soils (Gregg, 1963) in disturbed habitats (Wheeler & Wheeler, 1963). This ant is documented as being both potentially beneficial and deleterious to insect management programs, as it has been observed tending various aphids and membracids, while preying on lepidopterous defoliators

(Capinera, 1980; Capinera & Roltsch, 1981). Aside from studies into the life history and habits of

F. neoclara, no studies so far have focused on the genetics or social organization of this ant. In

thesis chapter II, I explore the social supergene in the socially polymorphic ant *Formica neoclara*.

Using RADseq data, I show that *F. neoclara* is socially polymorphic and it has a social supergene

on chromosome three, like previously studied *Formica* species. However, unlike previously

described *Formica*, *F. neoclara* differs in polygynous colony-level genotype distributions.

Formica neoclara aptly illustrates how, despite sharing a supergene, species can differ in their

supergene correlated polymorphisms. Unlike previously described *Formica*, *F. neoclara* polygyne

colonies tolerate individuals without an Sp haplotype, with some colonies harboring all three

genotypes (Sm/Sm, Sm/Sp, and Sp/Sp). This suggests that the *F. neoclara* supergene differs from

F. selysi, at least in terms of selfish genetic elements (Avril et al., 2020). Gaining insight into the

function of the *Formica* supergene outside of previously described species illustrates how

syntenic regions within a genus can exhibit functional differences - polygyne colonies of *F.

neoclara* possess Sm homozygous workers, unlike polygyne colonies of *F. selysi*. Observing

functional differences within a genus harboring a widespread supergene naturally raises

questions about the evolution of said supergene. Exploring the shared genetic elements

between *Formica* and the closely related genus *Polyergus*, we may better comprehend the

evolution of the *Formica* supergene.

 Formica ants often have alternate social forms, monogyne and polygyne colonies, within

species (Hölldobler & Wilson 1990). Regardless of queen number, social insect colonies are

prime targets for parasites; as they offer protection, food, and, depending on the form of

parasitism, brood care. Social parasites within Hymenoptera are organisms that exploit the

parental nature of their hosts for their own reproductive fitness (Bono et al., 2007; Sapp et al., 2020). Social parasites are known in several hymenopteran clades such as bumblebees (LHomme & Hines, 2018), paper wasps (Cervo, 2006), and ants (de la Mora et al., 2020; Rabeling 2021). One such social parasite genus in particular, *Polyergus*, has garnered much scientific attention. *Polyergus* is a genus of ant known for parasitizing various *Formica* species, obtaining their host pupae through well-documented raids (Talbot, 1967; Hasegawa & Yamaguchi, 1994; Trager, 2013). A new *Polyergus* queen will enter a host Formica nest, assassinate the queen, assume her position, then use the existing workforce to rear her brood (Topoff et al., 1988; Johnson et al., 2001). Since *Polyergus* species lack the capability to feed themselves or rear brood (King & Trager, 2007), they require a new supply of hosts periodically, hence they engage in raids. *Formica* and its sister genus *Iberoformica* form a clade that is sister to *Polyergus*, separated by about 33 million years of evolutionary history (Borowiec et al., 2021). Since the *Formica* supergene is ancient, potentially predating the split between the taxa (Brelsford et al., 2020; Borowiec et al., 2021), investigating the genome of *Polyergus* provides a prime opportunity to explore the timing and synteny in the evolution of the *Formica* supergene. In thesis chapter III, I explore the synteny between *Polyergus* and *Formica* by constructing a linkage map of a Polyergus species using *F. selysi* as a reference. I then visualize regions of chromosomal synteny and inversions by Oxford grid.

While the evolution and complete function of supergenes within many systems remain to be elucidated, understanding of these systems continues to advance. Studying the supergenes within a specific species, such as *F. neoclara*, allows us to better understand how supergene genotypes shape phenotypes. Comparative analyses of linkage maps across genera,

such as with *Polyergus* and *F. selysi*, allow us to investigate the evolutionary history of the

supergene polymorphism. This thesis aims to contribute to the current working knowledge of

the *Formica* social supergene.

References

Avril, A., Purcell, J., Béniguel, S., & Chapuisat, M. (2020). Maternal effect killing by a supergene controlling ant social organization. *Proceedings of the National Academy of Sciences*, *117*(29), 17130–17134. https://doi.org/10.1073/pnas.2003282117

Avril, A., Purcell, J., Brelsford, A., & Chapuisat, M. (2018). Asymmetric assortative mating and queen polyandry are linked to a supergene controlling ant social organization. *Molecular Ecology*, *28*(6), 1428–1438. https://doi.org/10.1111/mec.14793

Avril, A., Zahnd, S., Djordjevic, J., & Chapuisat, M. (2019). No mate preference associated with the supergene controlling social organization in Alpine Silver Ants. *Journal of Evolutionary Biology*. https://doi.org/10.1111/jeb.13479

Bono, J. M., Blatrix, R., Antolin, M. F., & Herbers, J. M. (2007). Pirate ants (*Polyergus breviceps*) and sympatric hosts (*Formica occulta* and *Formica sp. cf. argentea*): Host specificity and Coevolutionary Dynamics. *Biological Journal of the Linnean Society*, *91*(4), 565–572. https://doi.org/10.1111/j.1095-8312.2007.00817.x

Borowiec, M. L., Cover, S. P., & Rabeling, C. (2021). The evolution of social parasitism in *Formica* ants revealed by a global phylogeny. *Proceedings of the National Academy of Sciences*, *118*(38). https://doi.org/10.1073/pnas.2026029118

Brelsford, A., Purcell, J., Avril, A., Tran Van, P., Zhang, J., Brütsch, T., Sundström, L., Helanterä, H., & Chapuisat, M. (2020). An ancient and eroded social supergene is widespread across *Formica* ants. *Current Biology*, *30*(2). https://doi.org/10.1016/j.cub.2019.11.032

Bull, J. J. (1983). *Evolution of sex determining mechanisms*. Benjamin/Cummings Publishing Company.

Capinera, J. L. (1980). A trail pheromone from silk produced by larvae of the range caterpillar *Hemileuca oliviae* (Lepidoptera: Saturniidae) and observations on aggregation behavior. *Journal of Chemical Ecology*, *6*(3), 655–664. https://doi.org/10.1007/bf00987676

Capinera, J. L., & Roltsch, W. J. (1981). The Predatory Ant Formica neoclara: Effect on Within-Field Bean Aphid Distribution, and Activity in Relation to Thermal Conditions. *Journal of the Kansas Entomological Society*, *54*(3), 578–586.

Cervo, R. (2006). *Polistes* wasps and their social parasites: an overview. *Annales Zoologici Fennici*, *43*(5/6), 531–549.

Charlesworth, D., & Charlesworth, B. (1980). Sex differences in fitness and selection for centric fusions between sex-chromosomes and Autosomes. *Genetical Research*, *35*(2), 205–214. https://doi.org/10.1017/s0016672300014051

Charlesworth, D., Charlesworth, B., & Marais, G. (2005). Steps in the evolution of heteromorphic sex chromosomes. *Heredity*, *95*(2), 118–128. https://doi.org/10.1038/sj.hdy.6800697

Darlington, C. D., & Mather, K. (1949). *The elements of Genetics*. George Allen & Unwin Ltd: London.

de la Mora, A., Sankovitz, M., & Purcell, J. (2020). Ants (Hymenoptera: Formicidae) as host and intruder: recent advances and future directions in the study of exploitative strategies. *Myrmecological News*, *30*, 53–71. https://doi.org/https://doi.org/10.25849/myrmecol.news_030:053

Dobzhansky, T. (1970). *Genetics of the evolutionary process*. Columbia Univ. Pr.

Dobzhansky, T., & Pavlovsky, O. (1960). How stable is balanced polymorphism? *Proceedings of the National Academy of Sciences*, *46*(1), 41–47. https://doi.org/10.1073/pnas.46.1.41

Dobzhansky, T., & Sturtevant, A. H. (1938). Inversions in the chromosomes of *Drosophila pseudoobscura*. *Genetics*, *23*(1), 28–64. https://doi.org/10.1093/genetics/23.1.28

Fisher, R. A. (1930). *The genetical theory of natural selection*. Oxford University Press, Oxford.

Fontcuberta, A., De Gasperin, O., Avril, A., Dind, S., & Chapuisat, M. (2021). Disentangling the mechanisms linking dispersal and sociality in supergene-mediated ant social forms. *Proceedings of the Royal Society B: Biological Sciences*, *288*(1949). https://doi.org/10.1098/rspb.2021.0118

Ford, E. B. (1965). *Genetic polymorphism*. Faber & Faber: London.

Gotzek, D., & Ross, K. G. (2007). Genetic regulation of colony social organization in fire ants: An integrative overview. *The Quarterly Review of Biology*, *82*(3), 201–226. https://doi.org/10.1086/519965

Greenberg, L., Fletcher, D. J. C., & Vinson, S. B. (1985). Differences in Worker Size and Mound Distribution in Monogynous and Polygynous Colonies of the Fire Ant *Solenopsis invicta* Buren. *Journal of the Kansas Entomological Society*, *58*(1), 9–18.

Gregg, R. E. (1963). *The ants of Colorado with reference to their ecology, taxonomy, and geographic distribution*. University of Colorado Press.

Hasegawa, E., & Yamaguchi, T. (1994). Raiding behavior of the Japanese slave-making ant *Polyergus samurai*. *Insectes Sociaux*, *41*(3), 279–289. https://doi.org/10.1007/bf01242299

Hölldobler Bert, & Wilson, E. O. (1990). *The ants*. The Belknap Press of Harvard University Press.

Huang, Y.-C., & Wang, J. (2014). Did the fire ant supergene evolve selfishly or socially? *BioEssays*, *36*(2), 200–208. https://doi.org/10.1002/bies.201300103

Hughes, W. O., Oldroyd, B. P., Beekman, M., & Ratnieks, F. L. (2008). Ancestral monogamy shows kin selection is key to the evolution of eusociality. *Science*, *320*(5880), 1213–1216. https://doi.org/10.1126/science.1156108

Jay, P., Chouteau, M., Whibley, A., Bastide, H., Parrinello, H., Llaurens, V., & Joron, M. (2021). Mutation load at a mimicry supergene sheds new light on the evolution of inversion polymorphisms. *Nature Genetics*, *53*(3), 288–293. https://doi.org/10.1038/s41588-020-00771-1

Jay, P., Whibley, A., Frézal, L., Rodríguez de Cara, M. Á., Nowell, R. W., Mallet, J., Dasmahapatra, K. K., & Joron, M. (2018). Supergene evolution triggered by the introgression of a chromosomal inversion. *Current Biology*, *28*(11). https://doi.org/10.1016/j.cub.2018.04.072

Johnson, C., Vander Meer, R., & Lavine, B. (2001). Changes in the Cuticular Hydrocarbon Profile of the Slave-Maker Ant Queen, *Polyergus breviceps* Emery, After Killing a *Formica* Host Queen (Hymenoptera: Formicidae). *Journal of Chemical Ecology*, *27*, 1787–1804. https://doi.org/https://doi.org/10.1023/A:1010456608626

Joron, M., Frezal, L., Jones, R. T., Chamberlain, N. L., Lee, S. F., Haag, C. R., Whibley, A., Becuwe, M., Baxter, S. W., Ferguson, L., Wilkinson, P. A., Salazar, C., Davidson, C., Clark, R., Quail, M. A., Beasley, H., Glithero, R., Lloyd, C., Sims, S., … ffrench-Constant, R. H. (2011). Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature*, *477*(7363), 203–206. https://doi.org/10.1038/nature10341

Keller, L. (1993). The assessment of reproductive success of Queens in ants and other social insects. *Oikos*, *67*(1), 177. https://doi.org/10.2307/3545107

Keller, L., & Ross, K. G. (1998). Selfish genes: A green beard in the red fire ant. *Nature*, *394*(6693), 573–575. https://doi.org/10.1038/29064

King, J. R., & Trager, J. C. (2007). Natural history of the slave making ant, *Polyergus lucidus*, sensu lato in northern Florida and its three *Formica pallidefulva* group hosts. *Journal of Insect Science*, *7*(42), 1–14. https://doi.org/10.1673/031.007.4201

Kirkpatrick, M. (2010). How and why chromosome inversions evolve. *PLoS Biology*, *8*(9). https://doi.org/10.1371/journal.pbio.1000501

Kunte, K., Zhang, W., Tenger-Trolander, A., Palmer, D. H., Martin, A., Reed, R. D., Mullen, S. P., & Kronforst, M. R. (2014). *doublesex* is a mimicry supergene. *Nature*, *507*(7491), 229–232. https://doi.org/10.1038/nature13112

Lewis, D. (1942). The evolution of sex in flowering plants. *Biological Reviews*, *17*(1), 46–67. https://doi.org/10.1111/j.1469-185x.1942.tb00431.x

Lhomme, P., & Hines, H. M. (2018). Ecology and Evolution of Cuckoo Bumble Bees. *Annals of the Entomological Society of America*, *112*(3), 122–140. https://doi.org/10.1093/aesa/say031

Lisachov, A. P., Zadesenets, K. S., Rubtsov, N. B., & Borodin, P. M. (2015). Sex chromosome synapsis and recombination in male guppies. *Zebrafish*, *12*(2), 174–180. https://doi.org/10.1089/zeb.2014.1000

Macom, T. E., & Porter, S. D. (1996). Comparison of polygyne and monogyne red imported fire ant (Hymenoptera: Formicidae) population densities. *Annals of the Entomological Society of America*, *89*(4), 535–543. https://doi.org/10.1093/aesa/89.4.535

Mather, K. (1955). Polymorphism as an outcome of disruptive selection. *Evolution*, *9*(1), 52. https://doi.org/10.2307/2405357

Peters, R. S., Krogmann, L., Mayer, C., Donath, A., Gunkel, S., Meusemann, K., Kozlov, A., Podsiadlowski, L., Petersen, M., Lanfear, R., Diez, P. A., Heraty, J., Kjer, K. M., Klopfstein, S., Meier, R., Polidori, C., Schmitt, T., Liu, S., Zhou, X., … Niehuis, O. (2017). Evolutionary history of the Hymenoptera. *Current Biology*, *27*(7), 1013–1018. https://doi.org/10.1016/j.cub.2017.01.027

Purcell, J., Brelsford, A., Wurm, Y., Perrin, N., & Chapuisat, M. (2014). Convergent genetic architecture underlies social organization in ants. *Current Biology*, *24*(22), 2728–2732. https://doi.org/10.1016/j.cub.2014.09.071

Purcell, J., Lagunas-Robles, G., Rabeling, C., Borowiec, M. L., & Brelsford, A. (2021). The maintenance of polymorphism in an ancient social supergene. *Molecular Ecology*. https://doi.org/10.1111/mec.16196

Rabeling, C. (2021). Social parasitism. *Encyclopedia of Social Insects*, 836–858. https://doi.org/10.1007/978-3-030-28102-1_175

Rice, W. R. (1987). Genetic hitchhiking and the evolution of reduced genetic activity of the Y sex chromosome. *Genetics*, *116*(1), 161–167. https://doi.org/10.1093/genetics/116.1.161

Riehle, M. M., Bukhari, T., Gneme, A., Guelbeogo, W. M., Coulibaly, B., Fofana, A., Pain, A., Bischoff, E., Renaud, F., Beavogui, A. H., Traore, S. F., Sagnon, N. F., & Vernick, K. D. (2017). The *Anopheles gambiae* 2LA chromosome inversion is associated with susceptibility to *Plasmodium falciparum* in Africa. *ELife*, *6*. https://doi.org/10.7554/elife.25813

Ross, K. G., & Keller, L. (1998). Genetic control of social organization in an ant. *Proceedings of the National Academy of Sciences*, *95*(24), 14232–14237. https://doi.org/10.1073/pnas.95.24.14232

Rubenstein, D. R., Ågren, J. A., Carbone, L., Elde, N. C., Hoekstra, H. E., Kapheim, K. M., Keller, L., Moreau, C. S., Toth, A. L., Yeaman, S., & Hofmann, H. A. (2019). Coevolution of Genome Architecture and social behavior. *Trends in Ecology & Evolution*, *34*(9), 844–855. https://doi.org/10.1016/j.tree.2019.04.011

Sapp, J. R., Yost, J., & Lyon, B. E. (2020). The socially parasitic ant *Polyergus mexicanus* has host-associated genetic population structure and related neighbouring colonies. *Molecular Ecology*, *29*(11), 2050–2062. https://doi.org/10.1111/mec.15468

Schwander, T., Libbrecht, R., & Keller, L. (2014). Supergenes and complex phenotypes. *Current Biology*, *24*(7). https://doi.org/10.1016/j.cub.2014.01.056

Stolle, E., Pracana, R., Howard, P., Paris, C. I., Brown, S. J., Castillo-Carrillo, C., Rossiter, S. J., & Wurm, Y. (2019). Degenerative expansion of a young supergene. *Molecular Biology and Evolution*, *36*(3), 553–561. https://doi.org/10.1093/molbev/msy236

Sundström, L. (1995). Dispersal polymorphism and physiological condition of males and females in the ant, *Formica truncorum*. *Behavioral Ecology*, *6*(2), 132–139. https://doi.org/10.1093/beheco/6.2.132

Sundström, L., Seppä, P., & Pamilo, P. (2005). Genetic population structure and dispersal patterns in *Formica* ants — a review. *Annales Zoologici Fennici*, *42*(3), 163–177.

Talbot, M. (1967). Slave-raids of the ant *Polyergus lucidus* Mayr. *Psyche: A Journal of Entomology*, *74*(4), 299–313. https://doi.org/10.1155/1967/24519

Thompson, M. J., & Jiggins, C. D. (2014). Supergenes and their role in evolution. *Heredity*, *113*(1), 1–8. https://doi.org/10.1038/hdy.2014.20

Topoff, H., Cover, S., Greenberg, L., Goodloe, L., & Sherman, P. (1988). Colony founding by Queens of the obligatory slave-making ant, *Polyergus breviceps*: The role of the Dufour's gland. *Ethology*, *78*(3), 209–218. https://doi.org/10.1111/j.1439-0310.1988.tb00231.x

Trager, J. C. (2013). Global revision of the dulotic ant genus *Polyergus* (Hymenoptera: Formicidae, Formicinae, formicini). *Zootaxa*, *3722*(4), 501. https://doi.org/10.11646/zootaxa.3722.4.5

Trible, W., & Ross, K. G. (2015). Chemical Communication of queen supergene status in an ant. *Journal of Evolutionary Biology*, *29*(3), 502–513. https://doi.org/10.1111/jeb.12799

Tschinkel, W. R. (2006). *The fire ants*. Belknap Press of Harvard University Press.

Turner, J. R. G. (1967). On supergenes. I. The evolution of supergenes. *The American Naturalist*, *101*(919), 195–221. https://doi.org/10.1086/282485

Tuttle, E. M., Bergland, A. O., Korody, M. L., Brewer, M. S., Newhouse, D. J., Minx, P., Stager, M., Betuel, A., Cheviron, Z. A., Warren, W. C., Gonser, R. A., & Balakrishnan, C. N. (2016).

Divergence and functional degradation of a sex chromosome-like supergene. *Current Biology*, *26*(3), 344–350. https://doi.org/10.1016/j.cub.2015.11.069

Wang, J., Wurm, Y., Nipitwattanaphon, M., Riba-Grognuz, O., Huang, Y.-C., Shoemaker, D. W., & Keller, L. (2013). A Y-like social chromosome causes alternative colony organization in fire ants. *Nature*, *493*(7434), 664–668. https://doi.org/10.1038/nature11832

Wheeler, G. C., & Wheeler, J. (1963). *The ants of North Dakota*. University of North Dakota Press.

Yan, Z., Martin, S. H., Gotzek, D., Arsenault, S. V., Duchen, P., Helleu, Q., Riba-Grognuz, O., Hunt, B. G., Salamin, N., Shoemaker, D. W., Ross, K. G., & Keller, L. (2020). Evolution of a supergene that regulates a trans-species social polymorphism. *Nature Ecology & Evolution*, *4*(2), 240–249. https://doi.org/10.1038/s41559-019-1081-1

Chapter II. A Novel Distribution of Supergene Genotypes is Present in the Socially Polymorphic Ant *Formica neoclara*

Abstract

Supergenes are chromosomal regions with tightly linked clusters of alleles that control compound phenotypic traits. Supergenes have been demonstrated to contribute to the maintenance of polymorphisms within populations in traits as diverse as mimetic wing coloration in butterflies, mating strategies in birds, and malarial susceptibility in mosquitoes. A large supergene also underlies variation in social organization in *Formica* ants. Alternative supergene haplotypes are associated with the presence of either a single queen (monogyny) or multiple queens (polygyny) within colonies. Here, we assess the social structure and supergene status of the North American species *Formica neoclara*. We sequenced a subset of the genome in 280 individuals sampled in populations distributed from California to northern British Columbia using ddRADseq. We determined that *F. neoclara* is socially polymorphic in queen number and we show that the social polymorphism is associated with alternative haplotypes at the social supergene. Intriguingly, polygyne colonies can harbor workers that are homozygous for both haplotypes as well as heterozygotes. This colony genetic composition contrasts with other *Formica* species, in which almost all individuals in polygyne colonies have the polygyne-associated haplotype. The social polymorphism is present in widely distributed and genetically subdivided populations of *F. neoclara*. In studying this system in *F. neoclara*, we expand our understanding of the functional evolution of supergene haplotypes as they diverge in different lineages.

Introduction

Stable multilocus genetic polymorphisms often underlie complex phenotypic variation within populations (Turner, 1967; Joron et al., 2011; Schwander et al., 2014; Thompson & Jiggins, 2014). Such coadapted gene complexes are present in many organisms (Turner, 1967), playing a role in mimicry in butterflies (Turner, 1967; Joron et al., 2011; Kunte et al., 2014), mating morphs in birds (Küpper et al., 2015; Tuttle et al., 2016), and malaria susceptibility in mosquitoes (Riehle et al., 2017). These linked functional mutations, designated as supergenes, occur in regions of suppressed recombination (Schwander et al, 2014; Charlesworth, 2015) that can act as a single Mendelian element when heterozygous (Thompson and Jiggins, 2014; Kunte et al, 2014). Supergenes allow for the unified control of compound phenotypes (Schwander et al, 2014), providing a genetic mechanism to maintain balanced polymorphisms within populations (Thompson and Jiggins, 2014). A benefit of supergenes lies in their architecture; these clusters of tightly linked functional mutations often prevent disadvantageous intermediate phenotypes (Thompson and Jiggins, 2014) through reduced recombination (Schwander et al, 2014). As supergenes are widespread, many organisms can serve as models of study. Supergenes have been explored in studies involving the evolution of phenotypic diversity, such as the divergence of geographic races of *Heliconius* butterflies (Joron et al., 2006). Supergenes have also garnered scientific attention for their role in polymorphisms within populations (Thompson and Jiggins, 2014; Dobzhansky & Sturtevant, 1938), including social organization in ant species (Wang et al, 2013; Huang & Wang, 2013; Purcell et al, 2014; Avril et al, 2019).

Independent and distinct supergenes that underlie a polymorphism in colony queen

number were initially described in two ant species, *Solenopsis invicta* and *Formica selysi*

(Hymenoptera: Formicidae) (Wang et al. 2013; Purcell et al. 2014). Monogyne colonies are

headed by a single queen, whereas polygyne colonies have multiple queens, wherein nestmates

have lower genetic relatedness (Ross & Fletcher, 1985; Beye et al, 1998; Avril et al, 2018). These

large supergenes that span most of the chromosome were subsequently found in other *Formica*

(Brelsford et al., 2020; Purcell et al. 2021) and *Solenopsis* species (Yan et al. 2020), meaning that

they likely predate speciation of at least some species in these genera. Intriguingly, both

supergene polymorphisms are partly maintained by selfish genetic mechanisms, but the precise

mechanisms are different in each system. In *S. invicta*, the supergene haplotype found

exclusively in polygyne colonies (Sb) selfishly promotes its propagation via a green-beard effect

(Huang & Wang, 2013), in which heterozygous workers actively kill joining queens that lack the

polygyne-associated haplotype (Keller and Ross, 1998). Certain alleles on the respective

haplotypes also appear to bias their own transmission via meiotic drive (Ross and Shoemaker

2018). The selfish genetic mechanism in *Formica selysi* also favors the polygyne-associated

haplotype (Sp) through maternal effect killing (Avril et al, 2020), wherein offspring of

heterozygous queens only survive if they have an Sp haplotype.

There are some notable differences between *Solenopsis* and *Formica* species in the

distribution of supergene genotypes in colonies. In both cases, monogyne colonies contain

exclusively one supergene haplotype, SB in *S. invicta* and Sm in *F. selysi* (Wang et al., 2013;

Purcell et al. 2014). In contrast, polygyne *S. invicta* colonies possess SB/SB and SB/Sb workers

and SB/Sb queens. Sb/Sb females rarely reach adulthood (Keller and Ross, 1998; Deheer et al,

1998) due to one or more recessive lethal alleles on the Sb social supergene (Wang et al, 2013).

Polygyne *F. selysi* colonies do not contain Sm/Sm homozygotes (workers or queens), but they do

have Sp/Sp and Sm/Sp workers and queens (Purcell et al. 2014; Avril et al. 2018). We are now

beginning to look at the distribution of supergene haplotypes in other *Formica* species (e.g.,

Brelsford et al. 2020). Understanding the evolutionary history and any changes in the mode of

action in supergenes found in multiple species will provide novel insights into the processes that

shape complex phenotypic and multi-locus genetic polymorphisms.

Formica neoclara is an ant species found throughout western North America. Workers

forage in trees (Paulson and Akre, 1991), where they search for prey and tend honeydew-

producing insects (Capinera and Roltsch, 1981). Past research has focused on the natural history

and agricultural impact of *F. neoclara*, including nesting habits and Sternorrhyncha control on

crops (Paulson and Akre, 1991; Capinera and Roltsch, 1981; Wheeler & Wheeler, 1944). Despite

its broad range and impact on crops, the social organization and population structure of *F.*

neoclara remain largely unknown.

Here, we investigate *F. neoclara* populations distributed from California to Northern

British Columbia to determine whether the species is socially polymorphic and, if so, whether

colony queen number is under genetic control throughout its range. Further, we investigate the

genetic structure of populations across the range of this species to determine whether the

population likely expanded recently or whether geographically distant populations have a long

history of isolation from one another. Overall, this study will begin to uncover similarities and

differences in social polymorphisms and their genetic bases in ant species with distinct

evolutionary histories.

Methods

Field Sampling, DNA Extraction, and Sequencing

We collected *F. neoclara* workers from colonies and along transects in Alberta, British Columbia, California, and Idaho in June-July 2016. Whenever possible, we sampled at least eight workers from each colony. The transects consisted of collecting the first *Formica* ant that we observed every hundred meters along a road or trail in a chosen location, up to eight individuals. We frequently sampled individuals from different species at each stop along the transect. We stored samples in 100% ethanol.

We extracted DNA from the head and thorax of workers using a QIAGEN DNeasy Blood & Tissue Kit, following the insect tissue protocol with several modifications. Specifically, we manually ground the tissue in a tube while immersed in liquid nitrogen, used alternatively sourced spin columns (BPI-tech.com) and 70% ethanol for the second DNA wash, and eluted the DNA in 30 μL of buffer AE. We then used a double-digest restriction site associated DNA sequencing (RADseq) approach to sequence samples (for protocol, see Brelsford et al. 2016). Briefly, we digested the DNA using restriction enzymes MseI and SbfI (New England Biolabs Inc.) and ligated barcoded adapters. Next, we removed small DNA fragments using a mix of Sera-Mag SpeedBeadsTM Magnetic Carboxylate-Modified Particles (Thermo Fisher Scientific, cat. #65152105050250) and PEG/NaCl buffer (as in Rohland & Reich, 2012). We then amplified each sample in four separate PCR reactions, pooled all PCR products, and did a final round of small fragment removal using the Sera-Mag bead mixture. We sequenced 288 ant workers (8 were technical replicates of one colony and were removed from subsequent analyses) in three pooled libraries that contained

additional samples of other species not used in this analysis. Details of each library are provided in Table 1.

Bioinformatics

We demultiplexed reads across each of the three batches using the process_radtags (version 1.4) command in Stacks, with default parameters (Catchen et al., 2011). To merge paired end reads and remove the adapter sequence, we used PEAR (Zhang et al., 2014). We then aligned reads to the *Formica selysi* reference genome (Brelsford et al. 2020) using Bowtie2 (Langmead and Salzberg, 2012) and called genetic variants across the sample using *Samtools mpileup* (Li, 2009).

We initially filtered genotypes using *VCFtools* (v 0.1.13, Danecek et al., 2011) for missing data (to remove loci that were present in fewer than 80% of samples, --max-missing 0.8), read depth (--minDP 7), and sites with a minor allele frequency less than 0.05 (*--maf 0.05*). This filtering resulted in 342 retained loci in 280 workers.

We assessed colony composition using multiple metrics, allowing us to come to a consensus to infer colony queen number. The *COLONY* program (Jones & Wang, 2010) allowed us to infer the queen number of 32 colonies. We separated colonies by region (Alberta, California, British Columbia, Idaho) and ran *COLONY* once for each region. We excluded colonies with fewer than six workers (three in total: GCRC7, BHSC2, FRLC6) from colony-level analyses. Additionally, we excluded SNPs on chromosome 3 from the *COLONY* analysis to ensure that our assessment of colony queen number would be independent of the expected supergene (Purcell et al. 2021). After inferring queen number, we estimated the average relatedness among workers for the 32 colonies using several estimators. Relatedness calculations include the Ajk

statistic (--relatedness, Yang et al., 2010) and kinship-based inference for genome-wide

association studies (KING) ф (--relatedness2, Manichaikul et al., 2010) available with *VCFtools* as

well as the Huang diploid A estimator available on the Polyrelatedness program (e 14 0; Huang

et al., 2015). The unadjusted Ajk statistic is the genomic relationship of each pair of subjects j

and k, calculated from SNPs. Estimates of relationship use individuals in the sample as a base so

that the average relationship between all pairs of individuals is 0. Thus, the expectation for

output values is 0 for individuals within populations and 1 for individuals within themselves

(Yang et al., 2010). KING uses only markers with genotype data for both individuals, outputting

kinship coefficients, ф. Values of ф have a maximum of 0.5, with values above 0.354 being

considered duplicates or monozygotic twins (Manichaikul et al., 2010). The Huang estimator

uses a method of moments approach, equating sample moments with population moments, to

output pairwise relatedness values. Several factors can decrease the certainty of relatedness

estimator values (Huang et al., 2015). Therefore, we used these three relatedness estimators

jointly to account for shortcomings within the individual estimators. The Ajk statistic can be

impacted by rare variants, and allele frequencies near 0 or 1 make the equation unstable. ф

loses reliability when individuals are from a mix of close and distant populations (Manichaikul et

al., 2010), which can be an issue in large geographic scale analyses such as these. We note that

the Huang estimator is impacted by missingness, with individuals with higher levels of missing

data inflating their own and population mean pairwise relatedness estimations (Fig. II.4S). In

addition to these relatedness estimators, we calculated the pairwise proportion of identity by

descent between individuals (plink --genome, v1.07, Purcell et al., 2007). Finally, we used

opposing homozygosity as an inference of parentage for each colony (following Lagunas-Robles

et al., 2021). We calculated opposing homozygosity for the respective colonies by counting the

loci for which homozygotes were present for both the reference and alternative alleles within a colony, for bi-allelic single-nucleotide polymorphisms (SNPs). We inferred monogyne colonies as those with one queen identified by COLONY, higher average relatedness, and lower opposing homozygosity. We inferred polygyne colonies as those with more than one queen identified by COLONY, lower average relatedness, and higher opposing homozygosity. When different estimators resulted in conflicting signals, we considered the colonies to have undetermined ('ambiguous') social structure.

We assessed the association between the social polymorphism and the supergene region using two complementary approaches. First, after determining the colonies' putative social form, we assessed the supergene genotypes of individuals within said colonies. We assigned genotypes based on their position on a principal component analysis (PCA) of markers from chromosome 3 (Fig. II.2). The known region of suppressed recombination on chromosome 3, which spans from 2-12.5 Mbp (Purcell et al., 2021), was analyzed in *plink* (*--pca --allow-extra-chr*, Purcell et al., 2007)*. We determined that individuals within a colony having an inbreeding coefficient, F_{IS}, value above zero were homozygous, while those with a F_{IS} value below zero were heterozygous (--het, *VCFtools*). To distinguish the putative Sm/Sm and Sp/Sp homozygotes, we compared the SNPs of individuals from each cluster to the *F. selysi* Sm reference genome. One group of homozygotes had a higher proportion of reference alleles and was determined to represent the Sm/Sm workers. Based on clusters in the PCA of the low recombining region of chromosome 3 and an assessment of heterozygosity, we assigned genotypes to individual *F. neoclara* workers.

Second, we performed a genome-wide efficient mixed model association (GEMMA) analysis to test for an association between each locus and the inferred social form of each colony. We ran two analyses: one for all regions (N = 215 individuals included) and one for colonies from Alberta, Canada only (N = 85 individuals); the latter analysis reduced the effects of population structure on the analysis. We excluded workers from transect samples and from colonies labeled as ambiguous in social form from these analyses (65 individuals in total). *Beagle* (v 5.1, Browning et al., 2018) was used to impute missing genotypes within the *F. neoclara* genetic data. *GEMMA* (Zhou & Stephens, 2012) was used to estimate a relatedness matrix (-gk 1) and then fit a linear mixed model to each SNP (*-k -lmm 1*). We then visualized output data from this process via a Manhattan plot (Fig. II.1S).

To observe whether the geographically distant populations show signs of historic isolation or recent expansion, we utilized heterozygosity data for multiple analyses. We used *VCFtools* to assess the Hardy-Weinberg equilibrium (--hardy) of British Columbia and California populations using all available markers (N = 342). We calculated expected heterozygosity at variable sites (--site-pi, *VCFtools*) for each population as the average nucleotide diversity per variable site on all chromosomes except chromosome 3 of one individual per colony (the individual with the least missing data, ranging from 0-9.32%) and all transect samples. Lastly, we performed a pairwise isolation by distance (IBD) analysis on the 32 colonies. Like previous colony-level analyses, we excluded colonies with fewer than six individuals. We calculated the Weir and Cockerham (1984) F_{ST} between each colony using the --weir-fst-pop command in *VCFtools*. We calculated the distance between colonies using the *Imap* package (v1.32, Wallace,

2012) in r (R Core Team, 2021). We then plotted the log-transformed distance by Rousset's

(1997) genetic distance (Fig. II.4C).

Results

Assigning social form to colonies

Using multiple complementary metrics, we determined the colony queen number for 29

out of 32 colonies. Colonies with multiple queens generally had relatively high levels of

opposing homozygosity and relatively low within colony relatedness estimates, while colonies

with one queen had little or no opposing homozygosity and high within colony relatedness

(Figure II.1). In these cases, the majority of metrics supported this inference (Table II.S1). In

four cases, colonies exhibited intermediate values and social structure was not determined

("ambiguous" colonies, grey dots in Figure 1). Both social forms are present in all well-sampled

regions.

Identifying supergene genotypes of individuals

Independent of the social structure assessment, we determined whether there were

long-range haplotypes in *F. neoclara* on chromosome 3, which harbors the social supergene in *F.

selysi.* We performed a PCA for all individuals using chromosome 3 markers, and we observed

discrete clusters of individuals along PC1 and separation within clusters based on geographic

origin along PC2 and PC3 (Fig. II.2). Along PC1, in particular, we observe great differences in

allele frequency between clusters. Individuals shown in brown are in the central cluster along

PC1 and have excess heterozygosity, as determined by F_{IS} values on chromosome 3, suggesting

that they are heterozygous for two distinct supergene variants. Of the two remaining clusters,

the leftmost cluster aligns consistently to the *F. selysi* Sm/Sm reference alleles (green cluster,

Fig. II.2), while the rightmost cluster more often is homozygous for the alternate allele (yellow

cluster). On this basis, we inferred that individuals in the left cluster (green) are homozygous for

the *F. neoclara* Sm, individuals in the center cluster (brown) are heterozygous (Sm/Sp), and

individuals in the right cluster (yellow) are homozygous for the alternative haplotype (Sp).

Looking across PC2 and PC3, we observe signatures of geographic variation within each

supergene genotype cluster (shapes). In complement, we performed a genome-wide efficient

mixed model association (GEMMA) to identify SNPs associated with variation in colony queen

number. When we restrict the analysis to the socially polymorphic Alberta population (85

individuals from 11 colonies), we see a strong association between five SNPs on chromosome 3

and colony social form (the presence or absence of multiple queens, inferred by *COLONY*). These

SNPs lie within 7.7Mb and 12.6Mb (Fig. II.S1). We did not detect significant SNPs elsewhere in

the genome. When we analyze the data of all non-socially ambiguous populations (222

individuals from 28 colonies), we detect one significant SNP on chromosome 3 at 12.1Mbp. We

posit that the substantial genetic variation between populations affects the signal-to-noise ratio

of the latter analysis.

Genotypic distribution within regions and colonies

Both monogyne and polygyne colonies were found across the broad geographical

sample tested in this study. Every colony determined to be monogyne is composed of Sm/Sm

workers exclusively, except for the single colony in Idaho (pocC4). In contrast, every colony

independently determined to be polygyne harbors at least three workers with the Sp haplotype.

Interestingly, all polygyne colonies have Sm/Sp individuals present and frequently contain individuals with all three supergene genotypes (Sm/Sm, Sm/Sp, Sp/Sp). The solitary colony from Idaho was assessed to be monogyne (Table II.1S), with nestmate relatedness values (0.832 according to the Huang estimator) likely to be inflated due to the small sample size from this region. Genotype distributions of the three ambiguous colonies range from Sm/Sm exclusively to all workers having an Sp haplotype (Fig. II.2S). Alberta, British Columbia, and California possess the three respective genotypes: Sm/Sm, Sm/Sp, and Sp/Sp. Although the Sm/Sm genotype is associated with monogyne colonies, it is also found in polygyne colonies (Figs. II.3-4). The presence of all three genotypes in polygynous *F. neoclara* colonies contrasts with previously studied ant species with social supergenes (Fig. II.5).

Population Genetic Structure

Expected heterozygosity values from Alberta (sites from the plains and Rocky mountains grouped separately on the map), northern and southern British Columbia, California, and Idaho range from 0.15 to 0.27, with an average of 0.235 (Fig. 4A). This pattern is not consistent with a recent population expansion in these parts of the species range. The PCA utilizing all markers except those on chromosome 3 reveals clustering by region of origin as well, further supporting the inference that these populations are genetically distinct (Fig. 4B). Isolation by distance (IBD) analysis utilized the 32 colonies with six or more worker samples. The Pairwise F_{ST} values between colonies ranged from 0.014 to 0.405, with a mean of 0.245 (Fig. 4C). Pairwise distances between colonies were also variable, ranging from 3.54 meters up to 2200 kilometers, with an average of 964 kilometers. The r^2 for geographic distance by genetic distance is 0.408.

Discussion

Formica neoclara exhibits a social polymorphism in queen number across its range. The supergene underlying queen number variation in multiple *Formica* species (Purcell et al., 2014, 2021; Brelsford et al. 2020) is also present and associated with colony queen number in *F. neoclara* (Figs. II.3, II.5). However, the distribution of haplotypes within nests is notably different (Fig. II.5). In both *F. selysi* and *F. neoclara*, individuals in single queen colonies are all homozygous for the monogyne-associated haplotype, Sm. The difference between the species is observed in polygyne colonies. In *F. selysi,* every individual in a multiple queen colony harbors at least one copy of the polygyne-associated haplotype Sp (queen and worker genotypes include Sp/Sm and Sp/Sp; Purcell et al., 2014; Avril et al., 2019). In contrast, *F. neoclara* polygyne colonies can harbor individuals lacking the Sp allele, with some colonies containing all three possible genotypes (Sm/Sm, Sm/Sp, and Sp/Sp). Out of 18 polygyne colonies sampled, we never detected a multiple queen colony with exclusively Sm/Sm individuals. This pattern suggests that the association between the supergene and colony queen number is present in *F. neoclara,* as in other *Formica* species, despite differences in haplotype distribution within colonies*.*

The distribution of genotypes within polygyne colonies raise several questions about how the genetic and phenotypic polymorphisms are maintained in *F. neoclara.* Finding Sm/Sm workers in polygyne nests in all populations suggests that the 'maternal effect killing' selfish genetic mechanism found in *F. selysi* is not operating in *F. neoclara* (Avril et al., 2020). Likewise, finding Sp/Sp workers in polygyne nests suggests that the Sp haplotype is not a recessive lethal allele. In the convergently-evolved fire ant supergene, the polygyne-associated haplotype, Sb, has recessive lethal alleles, such that Sb/Sb individuals rarely survive to adulthood and never

reproduce (Wang et al. 2013; Keller & Ross, 1998; Ross 1997). Further research is needed to understand what selective pressures maintain the genetic polymorphism and prevent either haplotype from sweeping to fixation in these populations. Given that some of the well-studied mechanisms found in *S. invicta* and *F. selysi* appear to be weak or absent, it is unclear what factors maintain this genetic polymorphism in *F. neoclara*.

GEMMA analysis of all populations reveals that at least one SNP on chromosome 3 (12.2Mbp) is associated with social form. Restricting the analysis to colonies from Alberta shows that five SNPs on chromosome 3 are correlated with social form. A handful of genes associated with or around regions of the supergene (chromosome 3) are conserved across multiple *Formica* species, with *Knockout* standing out as a strong candidate gene (Brelsford et al., 2020; Purcell et al., 2021). While the positions of our five significant SNPs from the Alberta population, and the single SNP from all populations, do not fall within the base pair ranges given for candidate genes identified in other *Formica* (Purcell et al., 2021), they are nonetheless within the supergene region of chromosome 3.

To assess colony social structure, we employed a method that evaluates the opposing homozygosity of biallelic RAD loci and the nestmate relatedness in parallel (Fig. II.1). We have used variations of this method in several other species (Pierce et al., in prep; Lagunas-Robles et al., 2021), but the present study spans the most massive spatial scale. Most sampled colonies either exhibited a relatively low number of opposing homozygotes and a high level of relatedness, suggesting that workers are all daughters of a single queen, or had relatively high opposing homozygosity paired with low relatedness, suggesting that workers are produced by multiple queens. In theory, we should never detect opposing homozygosity in workers produced

by singly-mated monogyne queens. However, we note that rare genotyping errors (especially non-detection of one allele in truly heterozygous individuals) can generate a small number of loci that exhibit apparent opposing homozygosity. Polyandry likely occurs at a relatively low frequency in *F. neoclara*, as has been detected in *F. selysi* (Avril et al., 2018) as well as *F. aquilonia* (Pamilo, 1993) and *F. truncorum* (Sundtröm, 1994). Colonies of multiply-mated monogyne queens may be classified as ambiguous or polygyne under our method of assessing opposing homozygosity and relatedness. Polyandry may be occurring in monogyne colonies of *F. neoclara*. For example, ETHC3 and RVCC1, colonies labeled as ambiguous with low opposing homozygosity and relatively low average relatedness (Table II.1S), may be instances of multiply-mated queens. Colonies labeled as monogyne may also be polyandrous, such as FRLC4, which has elevated opposing homozygosity values despite all other metrics supporting monogyny (Fig. 1, Table II.1S). The pairwise relatedness between half-siblings is lower than between full siblings, and opposing homozygosity occurs at loci for which the queen is heterozygous, and each mate harbors an alternative allele. However, opposing homozygosity levels are still expected to be relatively low compared to that found in polygyne colonies. Alongside multiply mated monogynous queens, reproductive skew within polygyne colonies may also lead to ambiguous calls in social form because it would elevate average relatedness if the majority of sampled nestmate workers are full siblings. For example, ETHC2, labeled ambiguous, exhibits higher levels of average relatedness, moderate levels of opposing homozygosity, and the presence of the Sp haplotype in workers.

There are several other facets of our dataset that could influence the classification of parentage in colonies from our dataset. First, our sample covered a large geographic scale, but

F. neoclara population densities tended to be low. As a result, some geographically isolated sites were represented with just a single colony in our dataset. Relatedness values for colonies within populations with few samples are likely to be biased upward. For example, two relatively isolated colonies, pocC4 and HROC2, appear to display elevated relatedness values. The sole colony collected in Idaho, pocC4, appears to be monogyne in terms of relatedness and opposing homozygosity yet has Sm/Sm and Sm/Sp individuals and two queens according to *COLONY*. Likewise, HROC2, a relatively isolated colony from California, exhibits elevated relatedness values despite apparently having two queens and high opposing homozygosity (Table II.1S). On a technical note, our dataset also includes individuals sequenced in single-end and paired-end reads in different batches (Table II.1). Samples sequenced in 2016 were represent two of three colonies called as ambiguous (ETHC2 and ETHC3), which could reflect a difference in missing data or genotyping error rate. However, we verified that there was no pervasive batch effect in the data used in our analyses.

The principal component analysis using chromosome 3 markers revealed some population structure in both the Sp and the Sm haplotypes at the continental scale. Performing a principal component analysis for all markers except those on chromosome 3 yielded strong signals of geographic population structure. We see distinct clustering by region, with principal components 1 and 2 apparently separating the clusters by latitude and longitude, respectively (Fig. 4B). This structure, combined with the discoveries that genome-wide expected heterozygosity is high across our spatially distant localities and F_{ST} is elevated between populations, suggests that these populations likely have a long history of independence, with gene flow occurring rarely or slowly at this scale. Given the latitudinal distribution of our

sampling sites, from 39.3° N to 58.8° N, we initially expected that we might find evidence of a

recent expansion from one or more southern refugia following the last glacial maximum.

Instead, we see no clear latitudinal pattern in the distribution of expected heterozygosity and

population differentiation. Additionally, most colonies display elevated pairwise F_{ST} values, save

for pairwise comparisons of polygyne colonies in California and Northern British Columbia,

which are in close proximity to their neighboring colonies within their respective regions.

Monogyne colonies, even when in close proximity, display elevated F_{ST} values. Within our

sampled colonies, at least some allele frequency variance between populations is explained by

geographic distance (Fig. II.4C). In *F. selysi*, patterns of isolation by distance suggest restricted

dispersal for queens but not males (Avril et al., 2018). Future studies should investigate the

genetic and phenotypic differences between the geographic variants of the Sm and Sp

haplotypes using higher marker densities and additional field collection. This would provide an

ideal opportunity to understand how the evolutionary trajectories of supergene haplotypes,

which differ in the effective population size and, potentially, mode of transmission, diverge

within a widespread species. Future research may examine the functional genetic underpinnings

of social organization in *F. neoclara*.

Overall, *Formica neoclara* is socially polymorphic in queen number in all well-sampled

populations. This polymorphism is associated with divergent haplotypes at the previously

identified *Formica* social supergene. Interestingly, polygyne colonies frequently harbor Sm/Sm

workers, a pattern that has not been previously identified. As a result, this system offers an

exciting opportunity to examine epigenetic differences based on genotype and, independently,

social origin, at least for Sm/Sm individuals. In conclusion, our study clearly shows a novel axis of

variation in the *Formica* supergene evolution: the haplotypes must have some functional

differences among species, despite sharing a common evolutionary origin.

References

Avril A, Purcell J, Béniguel S, Chapuisat M. Maternal effect killing by a supergene controlling ant social organization. Proceedings of the National Academy of Sciences. 2020;117(29):17130–4.

Avril A, Purcell J, Brelsford A, Chapuisat M. Asymmetric assortative mating and queen polyandry are linked to a supergene controlling ant social organization. Molecular Ecology. 2018;28(6):1428–38.

Avril A, Zahnd S, Djordjevic J, Chapuisat M. No mate preference associated with the supergene controlling social organization in Alpine Silver Ants. Journal of Evolutionary Biology. 2019;32(7):742–8.

Beye M, Neumann P, Chapuisat M, Pamilo P, Moritz RF. Nestmate recognition and the genetic relatedness of nests in the ant *Formica pratensis*. Behavioral Ecology and Sociobiology. 1998;43(1):67–72.

Brelsford A, Purcell J, Avril A, Tran Van P, Zhang J, Brütsch T, et al. An ancient and eroded social supergene is widespread across *Formica* ants. Current Biology. 2020;30(2).

Browning BL, Zhou Y, Browning SR. A one-penny imputed genome from next-generation reference panels. The American Journal of Human Genetics. 2018;103(3):338–48.

Capinera JL, Roltsch WJ. The Predatory Ant *Formica neoclara*: Effect on Within-Field Bean Aphid Distribution, and Activity in Relation to Thermal Conditions. Journal of the Kansas Entomological Society. 1981;54(3):578–86.

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH. Stacks: Building and genotyping loci de novo from short-read sequences. G3 Genes|Genomes|Genetics. 2011;1(3):171–82.

Chapuisat M, Bocherens S, Rosset H. Variable queen number in ant colonies: No impact on queen turnover, inbreeding, and population genetic differentiation in the ant *Formica selysi*. Evolution. 2004;58(5):1064.

Charlesworth D. The status of supergenes in the 21st century: Recombination suppression in Batesian mimicry and sex chromosomes and other complex adaptations. Evolutionary Applications. 2015;9(1):74–90.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics. 2011;27(15):2156–8.

DeHeer CJ, Tscbinkel WR. The success of alternative reproductive tactics in monogyne populations of the ant *Solenopsis invicta*: Significance for transitions in social organization. Behavioral Ecology. 1998;9(2):130–5.

Fritz GN, Vander Meer RK, Preston CA. Selective male mortality in the red imported fire ant, *Solenopsis invicta*. Genetics. 2006;173(1):207–13.

Hölldobler B, Wilson EO. *The ants*. Cambridge: Harvard University Press; 1990.

Huang K, Ritland K, Dunn DW, Qi X, Guo S, Li B. Estimating relatedness in the presence of null alleles. Genetics. 2015;202(1):247–60.

Huang K, Ritland K, Guo S, Shattuck M, Li B. A pairwise relatedness estimator for polyploids. Molecular Ecology Resources. 2014;14(4):734–44.

Huang Y-C, Wang J. Did the fire ant supergene evolve selfishly or socially? BioEssays. 2013;36(2):200–8.

Jones OR, Wang J. Colony: A program for parentage and sibship inference from Multilocus
        Genotype Data. Molecular Ecology Resources. 2010;10(3):551–5.

Joron M, Frezal L, Jones RT, Chamberlain NL, Lee SF, Haag CR, et al. Chromosomal
        rearrangements maintain a polymorphic supergene controlling butterfly mimicry.
        Nature. 2011;477(7363):203–6.

Joron M, Papa R, Beltrán M, Chamberlain N, Mavárez J, Baxter S, et al. A conserved supergene
        locus controls colour pattern diversity in *Heliconius* butterflies. PLoS Biology. 2006;4(10).

Keller L, Ross KG. Selfish genes: A green beard in the red fire ant. Nature. 1998;394(6693):573–5.

Kunte K, Zhang W, Tenger-Trolander A, Palmer DH, Martin A, Reed RD, et al. *doublesex* is a
        mimicry supergene. Nature. 2014;507(7491):229–32.

Küpper C, Stocks M, Risse JE, dos Remedios N, Farrell LL, McRae SB, et al. A supergene
        determines highly divergent male reproductive morphs in the Ruff. Nature Genetics.
        2015;48(1):79–83.

Lagunas-Robles G, Purcell J, Brelsford A. Linked supergenes underlie split sex ratio and social
        organization in an ant. Proceedings of the National Academy of Sciences. 2021;118(46).

Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. Nature Methods.
        2012;9(4):357–9.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map
        format and SAMtools. Bioinformatics. 2009;25(16):2078–9.

Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen W-M. Robust relationship inference in genome-wide association studies. Bioinformatics. 2010;26(22):2867–73.

Pamilo P. Polyandry and allele frequency differences between the sexes in the ant *Formica aquilonia*. Heredity. 1993;70(5):472–80.

Paulson GS, Akre RD. Role of predaceous ants in Pear Psylla (Homoptera: Psyllidae) management: Estimating colony size and foraging range of *Formica neoclara* (Hymenoptera: Formicidae) through a mark-recapture technique. Journal of Economic Entomology. 1991;84(5):1437–40.

Purcell J, Brelsford A, Wurm Y, Perrin N, Chapuisat M. Convergent genetic architecture underlies social organization in ants. Current Biology. 2014;24(22):2728–32.

Purcell J, Lagunas-Robles G, Rabeling C, Borowiec ML, Brelsford A. The maintenance of polymorphism in an ancient social supergene. Molecular Ecology. 2021;30(23):6246-58

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. The American Journal of Human Genetics. 2007;81(3):559–75.

Riehle MM, Bukhari T, Gneme A, Guelbeogo WM, Coulibaly B, Fofana A, et al. The *Anopheles gambiae* 2LA chromosome inversion is associated with susceptibility to *Plasmodium falciparum* in Africa. eLife. 2017;6.

Ritland K. Estimators for pairwise relatedness and individual inbreeding coefficients. Genetical Research. 1996;67(2):175–85.

Rohland N, Reich D. Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. Genome Research. 2012;22(5):939–46.

Ross KG, Fletcher DJ. Comparative study of genetic and social structure in two forms of the fire Ant *Solenopsis invicta* (Hymenoptera: Formicidae). Behavioral Ecology and Sociobiology. 1985;17(4):349–56.

Ross KG, Shoemaker DW. Unexpected patterns of segregation distortion at a selfish supergene in the fire Ant *Solenopsis invicta*. BMC Genetics. 2018;19(1).

Ross KG, Vargo EL, Keller L. Social Evolution in a new environment: The case of introduced fire ants. Proceedings of the National Academy of Sciences. 1996;93(7):3021–5.

Ross KG. Multilocus evolution in fire ants: Effects of selection, gene flow and recombination. Genetics. 1997;145(4):961–74.

Rousset F. Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. Genetics. 1997;145(4):1219–28.

Schwander T, Libbrecht R, Keller L. Supergenes and complex phenotypes. Current Biology. 2014;24(7).

Stolle E, Pracana R, Howard P, Paris CI, Brown SJ, Castillo-Carrillo C, et al. Degenerative expansion of a young supergene. Molecular Biology and Evolution. 2018;36(3):553–61.

Sundström L. Sex ratio bias, relatedness asymmetry and queen mating frequency in ants. Nature. 1994;367(6460):266–8.

Thompson MJ, Jiggins CD. Supergenes and their role in evolution. Heredity. 2014;113(1):1–8.

Turner JR. On supergenes. i. the evolution of Supergenes. The American Naturalist. 1967;101(919):195–221.

Wang J, Wurm Y, Nipitwattanaphon M, Riba-Grognuz O, Huang Y-C, Shoemaker DW, et al. A Y-like social chromosome causes alternative colony organization in fire ants. Nature. 2013;493(7434):664–8.

Weir BS, Cockerham CC. Estimating F-statistics for the analysis of Population Structure. Evolution. 1984;38(6):1358.

Wheeler GC, Wheeler EW. The ants of North Dakota. North Dakota Historical Quarterly. 1944;11:231–71.

Yan Z, Martin SH, Gotzek D, Arsenault SV, Duchen P, Helleu Q, et al. Evolution of a supergene that regulates a trans-species social polymorphism. Nature Ecology & Evolution. 2020;4(2):240–9.

Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, et al. Common SNPs explain a large proportion of the heritability for human height. Nature Genetics. 2010;42(7):565–9.

Zhang J, Kobert K, Flouri T, Stamatakis A. Pear: A fast and accurate Illumina paired-end read merger. Bioinformatics. 2013;30(5):614–20.

Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. Nature Genetics. 2012;44(7):821–4.

Tables

Table II.1: Overview of libraries and samples.

Library year	# *F. neoclara* samples and # of total samples in the library	Sequencing facility	Sequence information	Average mean depth per individual, after filtering (range)
2016	75 of 1368	UC Berkeley Genomics Core	100 bp single-end reads, Illumina HiSeq 4000	123.0x (range 13-162.7)
2017	125 of 2629	UC Berkeley Genomics Core	150 bp paired-end reads, Illumina HiSeq 4000	68.1x (range 13.3-153.4)
2019	80 of 2348	Novogene	150 bp paired-end reads, Illumina HiSeq X 10	66.4x (range 13.8-134.1)

Table II.1S. Overview of consensus social form variables. The variables are queen number (inferred by *COLONY*), opposing homozygosity, average colony relatedness estimators (Ajk, KING, Huang), and average colony identity by descent. If the majority of these values agreed on one social form (monogyne or polygyne) the colony was labeled as such.

Colony	Region	Queen #	Opposing Homozygosity	AJK	KING	Huang	IBD	Social Form
ETHC1	Alberta	2	43	0.361244	0.210652	0.468476	0.468386	Polygyne
ETHC2	Alberta	1	18	0.525652	0.301461	0.6795	0.677714	Ambiguous
ETHC3	Alberta	1	20	0.516392	0.312056	0.667714	0.668136	Ambiguous
ETHC4	Alberta	1	0	0.689693	0.423655	0.854107	0.853643	Monogyne
ETHC5	Alberta	1	0	0.667018	0.384221	0.80825	0.807093	Monogyne
ETHC6	Alberta	1	33	0.451046	0.290043	0.586821	0.583354	Polygyne
ETHC7	Alberta	1	4	0.847787	0.401018	0.839393	0.838657	Monogyne
ETHC8	Alberta	3	46	0.372244	0.193056	0.443143	0.447471	Polygyne
CALC1	Alberta	1	5	0.767599	0.367085	0.73875	0.734511	Monogyne
CALC2	Alberta	1	41	0.415614	0.219902	0.4285	0.430968	Polygyne
RVCC1	Alberta	1	13	0.552622	0.333389	0.601107	0.592786	Ambiguous
RVPC1	Alberta	1	3	0.672572	0.345634	0.769	0.770475	Monogyne
SYLC1	Alberta	1	1	0.901961	0.368831	0.744	0.74528	Monogyne
SYLC5	Alberta	1	1	0.739002	0.382217	0.765857	0.764218	Monogyne
FRLC3	British Columbia	1	1	0.815752	0.387707	0.797786	0.79365	Monogyne
FRLC4	British Columbia	1	20	0.582716	0.343486	0.678643	0.676218	Monogyne
TOADC5	British Columbia	4	55	0.539146	0.198913	0.447536	0.451054	Polygyne
TOADC7	British Columbia	4	51	0.524667	0.179886	0.434333	0.413687	Polygyne
TOADC8	British Columbia	3	36	0.590429	0.212137	0.4866	0.488153	Polygyne
TOADC9	British Columbia	2	46	0.556905	0.22227	0.504769	0.435604	Polygyne
TOADC10	British Columbia	5	42	0.506351	0.206503	0.460714	0.459395	Polygyne
SHEC17	California	1	2	1.082954	0.337232	0.783905	0.781224	Monogyne
HROC2	California	2	30	1.061434	0.223977	0.662476	0.665724	Polygyne
RGOC22	California	2	38	0.808194	0.283722	0.612107	0.594307	Polygyne
RGOC25	California	3	18	0.677435	0.299934	0.608893	0.598382	Polygyne
RGOC30	California	4	39	0.731503	0.263564	0.562357	0.549486	Polygyne
RGOC31	California	4	25	0.751438	0.275231	0.578929	0.563493	Polygyne
RGOC32	California	7	38	0.69299	0.184254	0.509357	0.513479	Polygyne

RGOC33	California	5	45	0.714473	0.245003	0.534857	0.524564	Polygyne
RGOC34	California	3	37	0.966081	0.24939	0.574393	0.566514	Polygyne
RGOC36	California	2	38	0.759577	0.165084	0.467357	0.457614	Polygyne
pocC4	Idaho	2	6	0.886735	0.358112	0.831607	0.822818	Monogyne
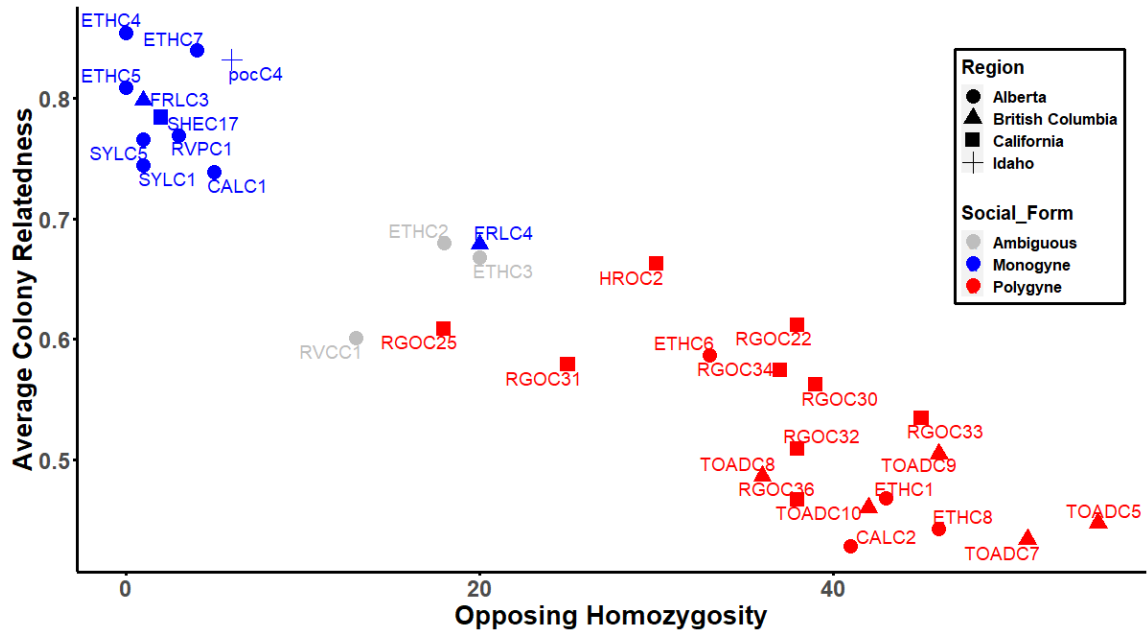
Figure II.1. Scatterplot displaying colony-level metrics of opposing homozygosity and average relatedness among nestmates (Huang estimator). We infer that the upper left cluster contains monogyne colonies (blue), whereas the lower right cluster contains polygyne colonies (red). Ambiguous colonies (social structure undetermined) are color-coded grey. Note that there is one colony outside of its cluster range: FRLC4. Despite the relatively high opposing homozygosity in this colony, other metrics suggest that just one queen is present.
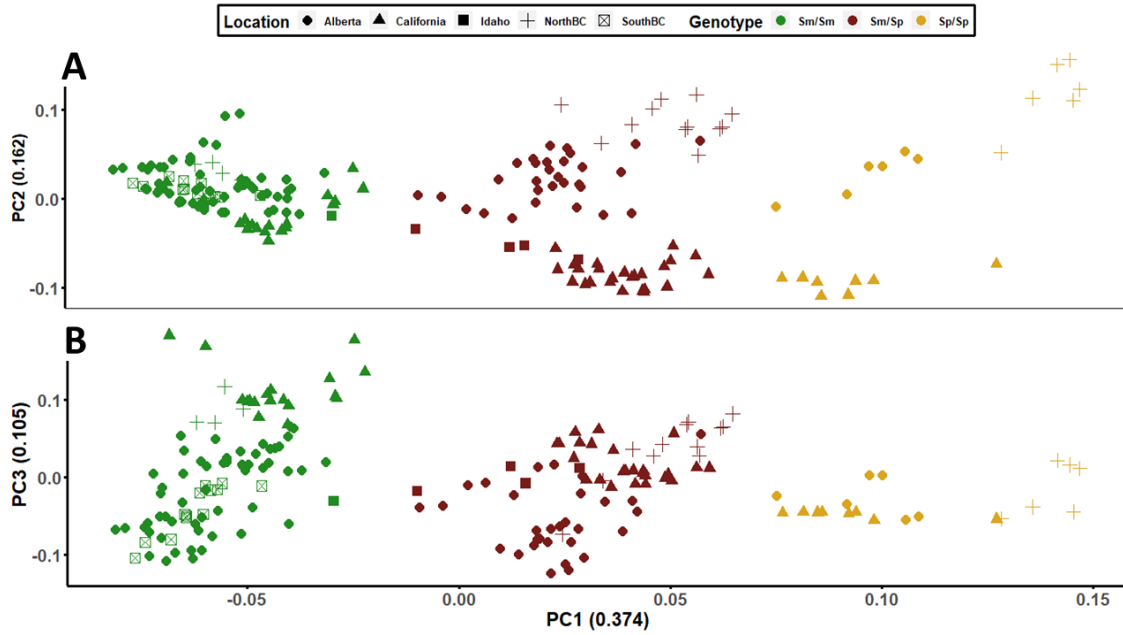
Figure II.2. Principal component analysis (PCA) for the low-recombining region of chromosome 3. The PCAs compare principal components 1 (PC1) and 2 (PC2) (**A**), and 1 (PC1) and 3 (PC3) (**B**). We identify one genotype cluster with excess heterozygosity in this region (brown) and two clusters that appear to be homozygous (green and yellow). To verify which cluster is homologous with the Sm/Sm genotype in *F. selysi,* we compared the three genotypes to the *F. selysi* reference genome (Sm/Sm). Individuals in the green cluster tend to be homozygous for the reference allele across the supergene, suggesting that they are Sm/Sm. Thus, the heterozygotes are Sm/Sp (brown), and the other clusters of homozygotes have region-specific versions of the Sp/Sp genotype (yellow). Point shapes are determined by sample location; circle = Alberta, triangle = California, filled square = Idaho, cross = Northern British Columbia, open square = Southern British Columbia.
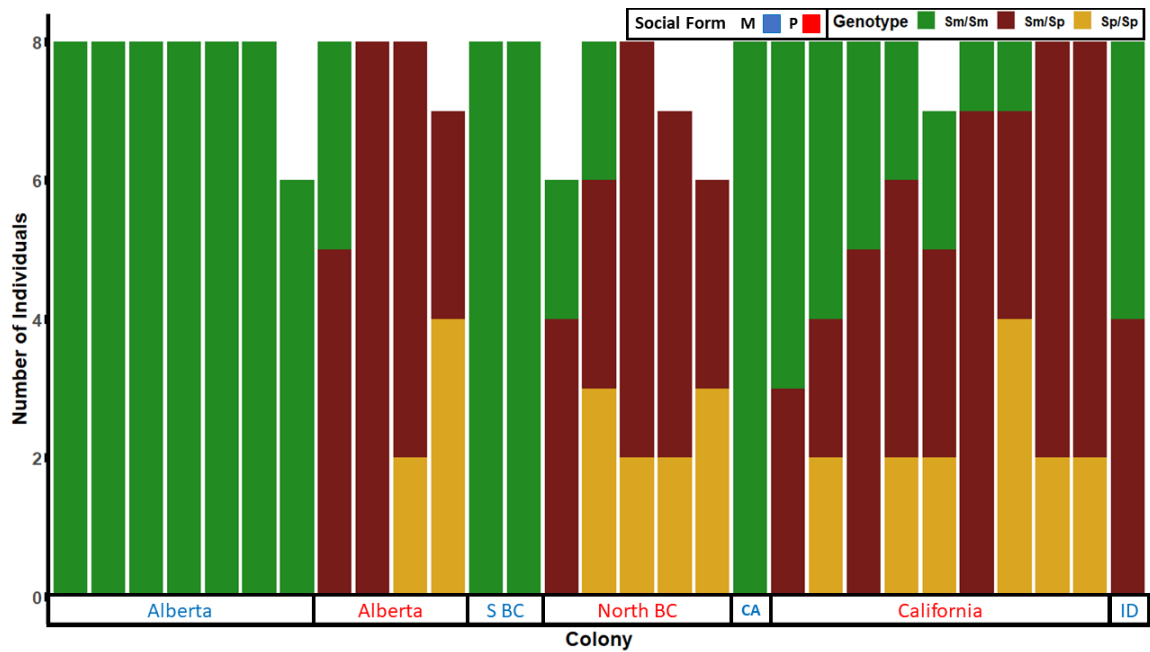
Figure II.3. Stacked bar plot displaying genotypes of samples from colonies ordered by region and haplotype frequency. Each bar represents all samples from an individual colony. A blue region name indicates that the colonies are monogyne, while a red region name indicates that colonies are polygyne. Genotype is indicated by color: green = Sm/Sm, brown = Sm/Sp, and Yellow = Sp/Sp. Genotype (green, brown, yellow) concerns each individual worker, while phenotype (red, blue) relates to the colony as a whole. Colonies with fewer than six individuals, or those deemed ambiguous, were excluded from this analysis.
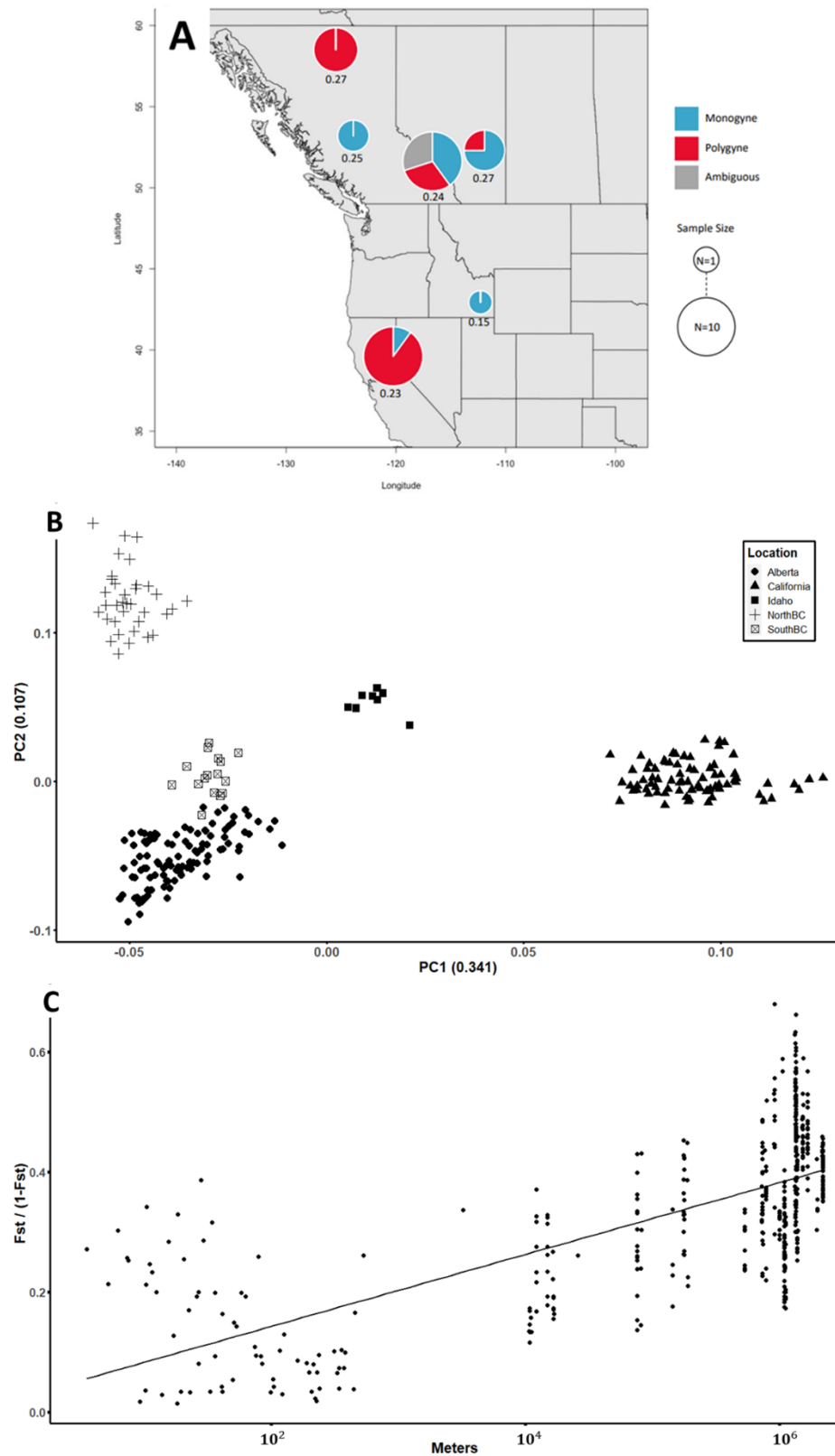
Figure II.4. Panel **A**. Location, social form, and expected heterozygosity of the six regions spanning the sample sites. Pie charts show the proportion of monogyne, polygyne, and ambiguous colonies in the population, with the pie chart size indicating sample size. Expected heterozygosity is shown below each pie chart. Panel **B**. A PCA without chromosome 3 data showing principal components 1 (PC1) and 2 (PC2) and their respective weights in parentheses. Clusters appear to be based on geographic location. Point shapes are determined by sample location; circle = Alberta, triangle = California, filled square = Idaho, cross = Northern British Columbia, open square = Southern British Columbia Panel **C**. A scatter plot displaying isolation by distance (IBD). Each point represents a pairwise comparison between two colonies - a single colony contributes to multiple points. The linear geographic distance between the two colonies (in meters) is on the x-axis, and Rousset's (1997) distance is on the y axis.

Single-queen colonies	Species	Multiple-queen colonies
Sm/Sm ?	*Formica neoclara*	*Sp/Sp, Sp/Sm, or Sm/Sm* ?
Sm/Sm Sm	*Formica selysi*	*Sp/Sp or Sp/Sm* Sp
SB/SB SB	*Solenopsis invicta*	*SB/SB or SB/Sb* SB or Sb

Figure II.5. The genetic system underlying social organization in three ant species with the social supergene (Wang et al. 2013; Purcell et al. 2014). Offspring genotype possibilities are shown; italicized genotypes are female and non-italicized genotypes are male.
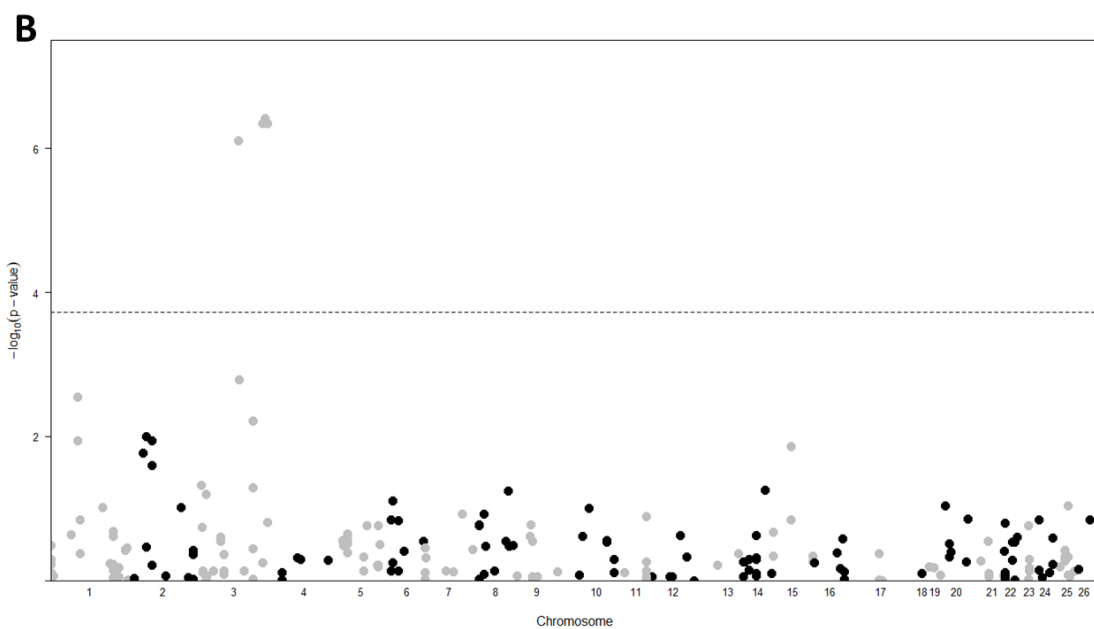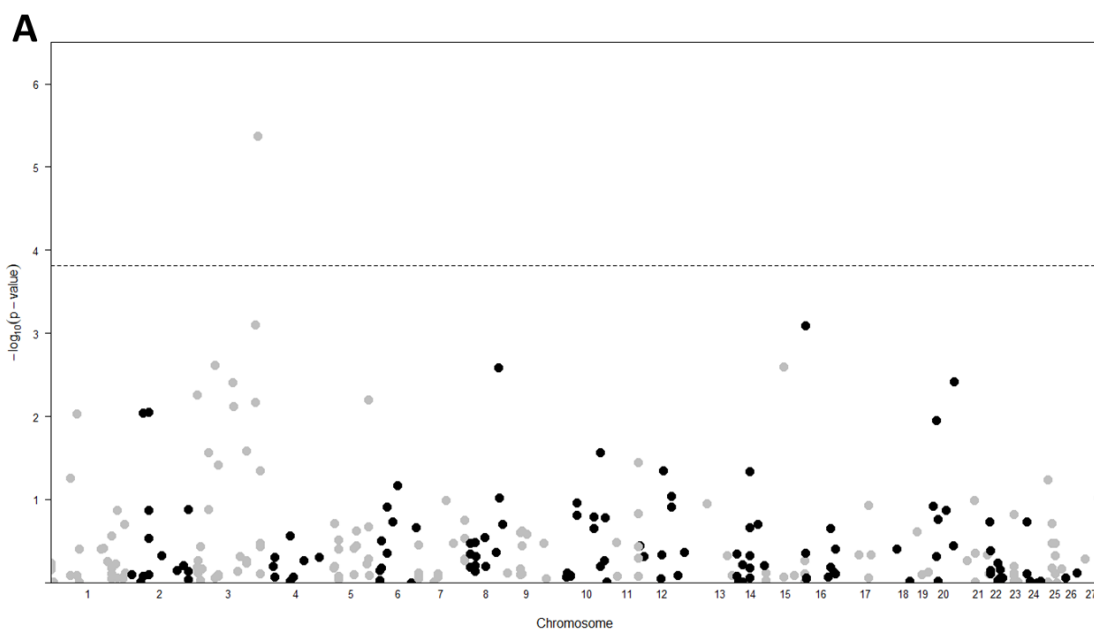
Figure II.1S. Panel **A**. Results of GEMMA analysis utilizing workers from colonies from all regions, visualized via Manhattan plot. A linear mixed model was performed, colony social form as the independent variable. Each point represents an individual SNP, with the corresponding chromosome on the x-axis and the negative logarithm of the SNP p-value on the y-axis. Only one SNP, from chromosome 3, approaches the significance level (Bonferroni corrected significance threshold: 1.52E-04; p-value: 1.63E-04). Panel **B**. Results of GEMMA analysis utilizing workers from colonies in Alberta., visualized via Manhattan plot. A linear mixed model was performed, inputted colony social form as the independent variable. Each point represents an individual SNP, with the corresponding chromosome on the x-axis and the negative logarithm of the SNP p-value on the y-axis. Five SNPs, all from chromosome 3, are above the significance threshold (Bonferroni corrected significance threshold: 1.89E-04; p-values: 7.79E-07, three at 4.43E-07, and 3.79E-07).
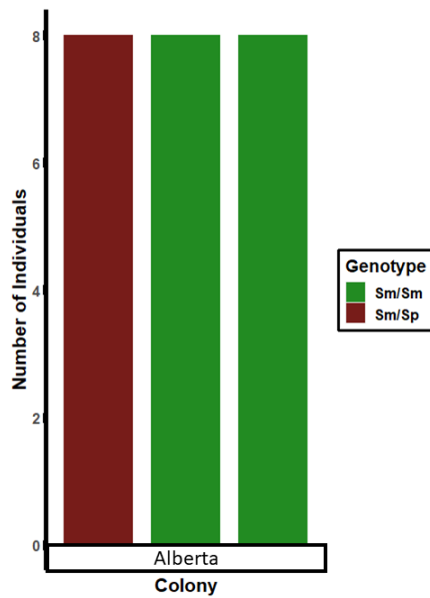
Figure II.2S. Stacked bar plot displaying genotypes of samples from colonies labeled as ambiguous in social form, ordered by region. Each bar represents all samples from an individual colony. Genotype, in relation to each individual worker, is indicated by color: green = Sm/Sm, brown = Sm/Sp, and Yellow = Sp/Sp.
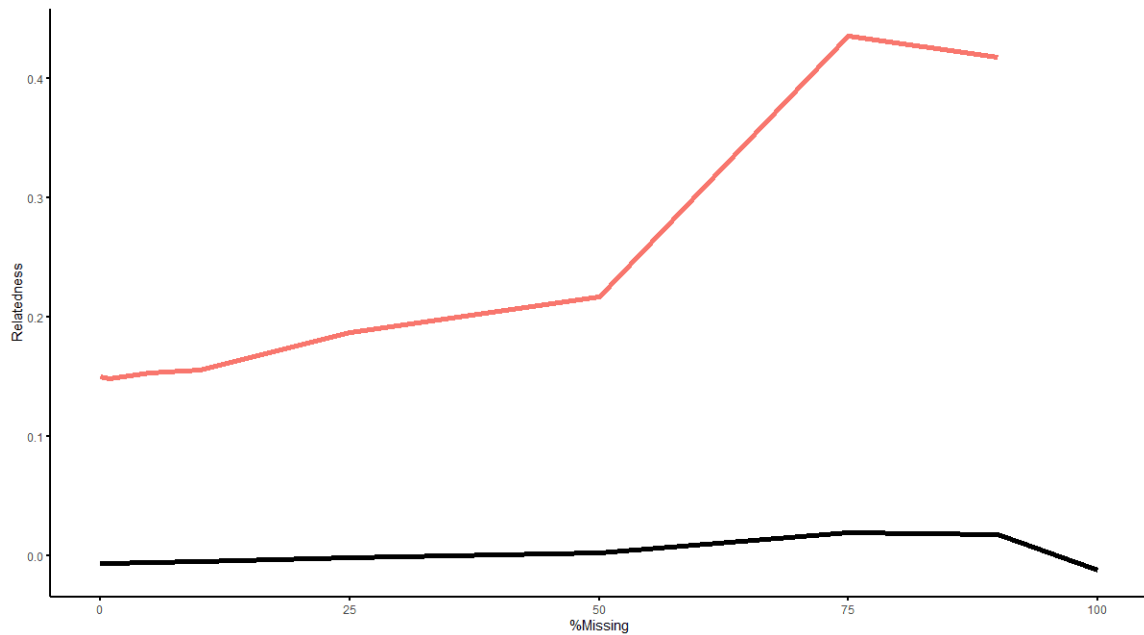
Figure II.3S. Line graph showing the effects of missingness on individual (red) and whole sample (black) relatedness values on the Huang (2015) estimator. Samples used are 30 individuals with less than 1% missingness. Missingness in one individual chosen at random (FRLC_6W8) was artificially inflated for varying increments. Notably, both individual and population pairwise relatedness values increase up to 75% missingness, with values above this showing a decrease in relatedness from the maximum.

Chapter III. The Degree of Synteny Between
Polyergus* and *Formica selysi

Introduction

Social parasitism within the social Hymenoptera frequently garners scientific attention. Within the Hymenoptera, there are at least 492 socially parasitic species, and out of those at least 401 social parasites have been recognized in the ants - Formicidae (Rabeling, 2021). Studying socially parasitic ants provides insight into the conditions surrounding and the consequences following the shift from cooperation to exploitation. Within the genus *Formica*, interspecific social parasitism is common, having the highest number of social parasite species among ants (Borowiec et al., 2021). Not only are *Formica* ants diverse in terms of parasitic species, but the genus also hosts a diverse range of other socially parasitic genera. *Polyergus* is a widespread genus of obligate socially parasitic ants (Hölldobler & Wilson, 1990). These ants are entirely dependent on their hosts, as they cannot care for brood or even feed themselves (Johnson et al., 2001; King & Trager, 2007). Sister to the *Formica-Iberoformica* clade (Boroweic et al., 2021), *Polyergus* parasitizes *Formica* ants exclusively (Trager, 2013). The process begins when a newly mated *Polyergus* queen infiltrates a host *Formica* nest. The *Polyergus* queen will seek out the natal *Formica* queen, kill her, and usurp her position as head of the colony (Topoff et al., 1988; Mori et al., 1995; Johnson et al., 2001). In some instances, depending on the species, a single parasitized colony can have host *Formica* workers outnumber their *Polyergus* parasites nearly eight to one (Talbot, 1967). As *Polyergus* colonies are reliant on their *Formica* hosts, and the natal *Formica* queen is killed by the parasite, it follows that *Polyergus* colonies

need to continuously obtain new host workers. *Polyergus* maintain their numbers of host

Formica by engaging in raids, in which *Polyergus* workers locate and raid *Formica* nests of their

brood - mainly pupae (Talbot, 1967; Hasegawa & Yamaguchi, 1994). Despite their multiple

morphological and behavioral adaptations, *Polyergus* is not always successful in raiding host

colonies (Talbot, 1967; Sapp et al., 2020) or in queen establishment (Topoff et al., 1988).

Additionally, when *Polyergus* successfully raids an intended host, the host colony typically

survives (Topoff et al., 1984).

Polydomy, a nest strategy where a single ant colony has multiple spatially separated

nests (Hölldobbler & Wilson, 1990), in *Formica* may be a response or a preadaptation for social

parasites such as *Polyergus*. Raids on individual nests of polydomous host colonies would have

less impact on overall colony fitness compared to monodomous colonies (Bono et al., 2007).

Polygynous and polydomous colonies can exchange brood, meaning that one raid on any

individual sub colony would not deplete brood from any one particular queen. Facultative

polygyny and polydomy arose early in the evolution of *Formica* (~30 Ma), these ancestral traits

were likely present in the ancestor of extant *Formica* (Borowiec et al., 2021). The social

supergene described in multiple *Formica* species responsible for the social form polymorphism

(monogyny and polygyny) is the oldest trans-species autosomal supergene described to date,

being conserved in a handful of Formica species separated by 30 million years of independent

evolution (Brelsford et al., 2020; Purcell et al., 2021). Moreover, *Formica* and *Iberoformica* split

from *Polyergus* around 33 million years ago (Borowiec et al., 2021), with all described *Polyergus*

species being monogynous and monodomous (Bono et al., 2007; Trager, 2013). As *Polyergus*

and *Formica* are taxa with close evolutionary and parasite-host relationships, this raises

questions on the synteny between the *Formica* social supergene and homologous regions of the *Polyergus* genome. Does the *Polyergus* genome resemble the ancestral *Formica* genome? Has chromosome 3 in *Formica* changed much since divergence, and if so, how?

Linkage mapping has a variety of uses and is viable in any system where recombination occurs, with the best study candidates producing many offspring. The most classic example is narrowing a trait of interest down to a particular genomic region; though linkage mapping is also used to assemble genomic data, assist in determining karyotype (Blande et al., 2020), and provide improved versions of previous genome sequences (Boyle et al., 2021). Additionally, the linkage maps of two species can be compared with one another through a process called comparative linkage mapping. By using comparative linkage mapping, the relative evolution of two species can be compared by identification of synteny (Boyle et al., 2021), recombination rate (Meznar et al., 2010), chromosomal duplications (Liu et al., 2017), inversions, and translocations. This can be performed between related or distant species, identifying events that have changed or that have maintained parts of the respective genomes since the initial divergence of the respective species (Tanksley et al., 1988; Whitkus et al., 1992). Comparative linkage mapping has been performed on a wide variety of organisms, including many insects. Recombination rates between two species of honey bee (Meznar et al., 2010), identification of *Papilio* butterfly mimicry loci (Obara et al., 2010), and diversity assessment in aphids (Guo et al., 2017) have all been performed utilizing linkage maps. Furthermore, linkage mapping with ants has illuminated the genetic underpinnings of sex determination in this taxon (Miyakawa & Mikheyev, 2015) as well as determine the synteny of the monogyne-associated social supergene haplotype across several *Formica* species (Brelsford et al., 2020). In this chapter, I aim to

determine the degree of synteny between *Polyergus* and *Formica*, particularly in terms of the

social supergene on the *Formica* chromosome 3 (Purcell et al., 2014; Brelsford et al., 2020). In

doing so, I aim to gain further insight into the evolution of the *Formica* supergene, particularly

the timing and conditions surrounding the suppression of recombination within the supergene

region.

Methods

Field Collection

More than 100 *Polyergus* workers were collected into 100% ethanol from a group of

workers that were actively raiding brood from a *Formica* colony in Snoqualmie Pass,

Washington. The *Polyergus* species is not yet confirmed.

DNA Extraction and Library Preparation

I extracted DNA from 96 *Polyergus* workers, all from the same colony, using the QIAGEN

DNeasy Blood and Tissue kit, following the insect tissue protocol with several modifications.

Modifications included manually grinding tissue while the 1.7 mL tubes were immersed in liquid

nitrogen, using generic spin columns (BPI-tech.com), using 70% ethanol for the second DNA

wash, and eluting the DNA in 30 μL of buffer EB. I then used a double-digest restriction site

associated DNA sequencing (ddRADseq) approach in preparation for sequencing (protocol at

Brelsford et al., 2016). I digested the DNA using enzymes MseI and EcoRI (New England Biolabs

Inc.) and ligated barcoded adapters. Next, I removed small DNA fragments using a mix of Sera-

Mag SpeedBeads™ Magnetic Carboxylate-Modified Particles (Thermo Fisher Scientific, cat.

#65152105050250) and PEG/NaCl buffer (as in Rohland and Reich, 2012). I then amplified each

sample in four separate PCR reactions, pooled all PCR products, and did a final round of small

fragment removal using the Sera-Mag bead mixture. The library was pooled with another

RADseq library (for an unrelated project) and sequenced on an Illumina HiSeq X ten by

Novogene (en.novogene.com), yielding 150 base pair paired end reads.

Bioinformatics

Following sequencing, I demultiplexed reads using the process_radtags command in

Stacks, with default parameters (Catchen et al., 2011). To merge paired end reads and remove

the adapter sequence, I used PEAR (Zhang et al., 2014). I then aligned the reads to the *Formica*

selysi reference genome (Brelsford et al., 2020) and created a de novo map (Stacks,

denovo_map.pl), each being performed independently of the other. Samples that had two

orders of magnitude less data than average were excluded from the analysis (N = 1).

I filtered genotypes using VCFtools (v0.1.16, Danecek et al., 2011) for read depth (--

minDP 7), missing data (--max-missing 0.8), and sites with a minor allele frequency less than

0.15 (--maf 0.15). Any individuals with more than 20% missing data were removed from analysis

(N = 13). Additionally, excessively homozygous and heterozygous SNPs were observed (--hardy)

and subsequently removed (--exclude-positions). This filtering process resulted in 2928 retained

loci in 83 workers.

After filtering, I performed linkage mapping using MSTmap (Wu et al., 2015). To account

for unknown phasing, I followed the method of Gadau (2009) during construction of the

MSTmap input file. Namely, artificially doubling the number of markers and assigning the

respective complement markers the opposite zygosity identifiers. I ran MSTmap with the

following parameters: p_value_cutoff (0.00001), no_map_dist (30), and missing_threshold (0.2).

This resulted in an initial number of 606 linkage groups, including genotyping errors marked as single linkage groups with only one marker (N = 499) - which were subsequently removed. Additionally, any linkage groups with less than fifteen loci were excluded from subsequent analyses (N = 64). I then compared pairs of analogous linkage groups and chose the more optimal group based on map distance, loci number, and bin number. Linkage groups with gaps of more than 50 centimorgans (cM) were split into separate linkage groups at the gaps (N = 4, with the fragments being less than fifteen loci and thus removed). Following this, I arranged the linkage groups, ordering them by cM. Using the cM and position values from the *Polyergus* and *F. selysi* reference genome physical positions, respectively, I constructed an Oxford grid (Fig III.1).

Results

Linkage Mapping

The linkage map of *Polyergus* made with *F. selysi* as a reference had 606 linkage groups, though the majority of these were single markers not aligning to any particular linkage group (n=499). Excluding fewer than fifteen markers (N = 64) resulted in 43 linkage groups. The expected karyotype of *Polyergus* is 26 or 27, depending on species (Lorite & Palomeque, 2010), so this linkage map is slightly over split. Linkage groups range in size from 30cM to 355cM.

Oxford Grid

Using *Polyergus* and *F. selysi* genetic data, I constructed an Oxford grid. This comparison clearly shows regions of synteny between the *Polyergus* and *F. selysi* genomes on all *F. selysi* chromosomes (27 boxes along the x-axis, Fig. III.1). Some loci in the *Polyergus* map align to

smaller contigs in the *F. selysi* genome (larger rectangle on the right side of the x-axis, Fig. III.1).

Through this comparison, we can see clear instances of inversions between the *Polyergus* and *F. selysi* genomes, on *F. selysi* chromosomes 13, 14, 17, and 22. Notably, the social supergene in *Formica*, on chromosome 3, is split into six blocks of synteny that are distributed across six different linkage groups in *Polyergus*. We cannot yet determine whether these regions were unlinked in the ancestor of the *Polyergus*, *Iberoformica*, and *Formica* clade and underwent a chromosomal fusion in a more recent *Formica* ancestor, or whether this chromosome is split up in the *Polyergus* lineage. These six syntenic blocks span from 0.3 to 1.7Mb, 2.6 to 3.4 Mb, 3.5 to 5.2Mb, multiple within 5.5Mb, 6.4 to 11.5 Mb, and from 12.9 to 13.8Mb on chromosome 3 in *F. selysi*.

Discussion

Polyergus has large regions of synteny with *Formica selysi* across all chromosomes. However, not all genetic elements are found within the same chromosomes between the two species. The linkage map for *Polyergus* constructed using *F. selysi* as a reference resulted in 43 distinct linkage groups used for the construction of the Oxford grid. The number of linkage groups in *Polyergus* does not correspond with the number of chromosomes in this genus. *Formica*, *Polyergus*, and *Cataglyphis* all have karyotype numbers ranging from 26 to 27 (Lorite & Palomeque, 2010). Despite supergene elements potentially preceding the divergence of these two species (Brelsford et al., 2020; Borowiec et al., 2021), *Polyergus* does not have all genetic elements of the *F. selysi* chromosome 3 colocalized within the same linkage group (Fig. III.1). However, there is one region of synteny on Polyergus linkage group three, spanning from 6.4 to 11.5 Mb on the *F. selysi* reference chromosome 3. The remaining *F. selysi* chromosome 3

elements are scattered among six additional *Polyergus* linkage groups, ranging from one marker

to 1.7 Mb. These findings provide a preliminary look into the synteny of elements of the *Formica*

chromosome 3 within a non-*Formica* species. These results pave the way for future directions in

terms of timing and synteny of the *Formica* supergene.

Future directions should observe the patterns in other related species and outgroups.

Candidates include *Cataglyphis* spp., *Iberoformica*, and other related *Formica* spp. In doing so,

we may gain further insight into the evolution of the social supergene by capturing snapshots of

it through evolutionarily divergent lineages. In addition to this, gene content between the

groups should also be analyzed. Many candidate genes on chromosome 3 in *Formica* are

conserved across species within the genus (Brelsford et al., 2020). Identifying and searching for

candidate genes across genera may elucidate genes necessary for the evolution of the social

supergene or changes in gene content and function within the social supergene.

Comparative linkage mapping is a useful tool to observe chromosomal trends between

species of the course of divergent evolution. In performing comparative linkage mapping with *F.

selysi* and Polyergus, I contribute to our understanding of the trends in *Formica* social supergene

evolution as well as the synteny with its outgroup *Polyergus*.

References

Blande, D., Smolander, O.-P., Ahola, V., Rastas, P., Tanskanen, J., Kammonen, J. I., Oostra, V., Pellegrini, L., Ikonen, S., Dallas, T., DiLeo, M. F., Duplouy, A., Cem Duru, I., Halimaa, P., Kahilainen, A., Kuwar, S. S., Kärenlampi, S. O., Lafuente, E., Luo, S., … Saastamoinen, M. (2020). Improved chromosome level genome assembly of the Glanville Fritillary Butterfly (*Melitaea cinxia*) based on SMRT sequencing and Linkage Map. *BioRxiv*. https://doi.org/10.1101/2020.11.03.364950

Bono, J. M., Blatrix, R., Antolin, M. F., & Herbers, J. M. (2007). Pirate ants (*Polyergus breviceps*) and sympatric hosts (*Formica occulta and Formica sp. cf. argentea*): Host specificity and Coevolutionary Dynamics. *Biological Journal of the Linnean Society*, *91*(4), 565–572. https://doi.org/10.1111/j.1095-8312.2007.00817.x

Borowiec, M. L., Cover, S. P., & Rabeling, C. (2021). The evolution of social parasitism in *Formica* ants revealed by a global phylogeny. *Proceedings of the National Academy of Sciences*, *118*(38). https://doi.org/10.1073/pnas.2026029118

Boyle, J. H., Rastas, P. M., Huang, X., Garner, A. G., Vythilingam, I., & Armbruster, P. A. (2021). A Linkage-Based Genome Assembly for the Mosquito *Aedes albopictus* and Identification of Chromosomal Regions Affecting Diapause. *Insects*, *12*(2), 167. https://doi.org/10.3390/insects12020167

Brelsford, A., Purcell, J., Avril, A., Tran Van, P., Zhang, J., Brütsch, T., Sundström, L., Helanterä, H., & Chapuisat, M. (2020). An ancient and eroded social supergene is widespread across *Formica* ants. *Current Biology*, *30*(2). https://doi.org/10.1016/j.cub.2019.11.032

Gadau, J. (2009). Phase-unknown linkage mapping in ants. *Cold Spring Harbor Protocols*, *2009*(7). https://doi.org/10.1101/pdb.prot5251

Guo, J., Wang, Z., & Francis, F. (2017). Use of molecular markers for entomological diversity assessment and their application in population study of aphids. *Faunistic Entomology*, *70*, 49–62. https://doi.org/10.25518/2030-6318.3647

Hallar BL, Krieger MJ, Ross KG. Potential cause of lethality of an allele implicated in social
    evolution in fire ants. Genetica. 2007;131(1):69–79.

Hasegawa, E., & Yamaguchi, T. (1994). Raiding behavior of the Japanese slave-making ant
    Polyergus samurai. *Insectes Sociaux*, *41*(3), 279–289.
    https://doi.org/10.1007/bf01242299

Hölldobler Bert, & Wilson, E. O. (1990). *The ants*. The Belknap Press of Harvard University Press.

Johnson, C., Vander Meer, R., & Lavine, B. (2001). Changes in the Cuticular Hydrocarbon Profile
    of the Slave-Maker Ant Queen, Polyergus breviceps Emery, After Killing a Formica Host
    Queen (Hymenoptera: Formicidae). *Journal of Chemical Ecology*, *27*, 1787–1804.
    https://doi.org/https://doi.org/10.1023/A:1010456608626

Joron, M., Frezal, L., Jones, R. T., Chamberlain, N. L., Lee, S. F., Haag, C. R., Whibley, A., Becuwe,
    M., Baxter, S. W., Ferguson, L., Wilkinson, P. A., Salazar, C., Davidson, C., Clark, R., Quail,
    M. A., Beasley, H., Glithero, R., Lloyd, C., Sims, S., … ffrench-Constant, R. H. (2011).
    Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly
    mimicry. *Nature*, *477*(7363), 203–206. https://doi.org/10.1038/nature10341

King, J. R., & Trager, J. C. (2007). Natural history of the slave making ant, *Polyergus lucidus*, sensu
    lato in northern Florida and its three *Formica pallidefulva* group hosts. *Journal of Insect
    Science*, *7*(42), 1–14. https://doi.org/10.1673/031.007.4201

Kunte, K., Zhang, W., Tenger-Trolander, A., Palmer, D. H., Martin, A., Reed, R. D., Mullen, S. P., &
    Kronforst, M. R. (2014). *doublesex* is a mimicry supergene. *Nature*, *507*(7491), 229–232.
    https://doi.org/10.1038/nature13112

Liu, H., Fu, B., Pang, M., Feng, X., Yu, X., & Tong, J. (2017). A High-Density Genetic Linkage Map
    and QTL Fine Mapping for Body Weight in Crucian Carp (*Carassius auratus*) Using 2b-
    RAD Sequencing. *G3 Genes|Genomes|Genetics*, *7*(8), 2473–2487.
    https://doi.org/10.1534/g3.117.041376

Lorite, P., & Palomeque, T. (2010). Karyotype evolution in ants (Hymenoptera: Formicidae), with a review of the known ant chromosome numbers. *Myrmecological News*, *13*, 89–102.

Meznar, E. R., Gadau, J., Koeniger, N., & Rueppell, O. (2010). Comparative Linkage Mapping Suggests a High Recombination Rate in all Honeybees. *Journal of Heredity*, *101*(Supplement 1). https://doi.org/10.1093/jhered/esq002

Miyakawa, M. O., & Mikheyev, A. S. (2015). QTL Mapping of Sex Determination Loci Supports an Ancient Pathway in Ants and Honey Bees. *PLOS Genetics*, *11*(11). https://doi.org/10.1371/journal.pgen.1005656

Mori, A., D'Ettorre, P., & Le Moli, F. (1995). Host nest usurpation and Colony Foundation in the European Amazon Ant, *Polyergus rufescens* Latr. (Hymenoptera: Formicidae). *Insectes Sociaux*, *42*(3), 279–286. https://doi.org/10.1007/bf01240422

Obara, O., Furukawa, M., & Hori, H. (2010). Genetic linkage mapping of the mimetic gene H in *Papilio* butterfly: Development of EST-based interspecific PCR length variations. *Insect DNA Research Newsletter*, *12*, 8–10.

Purcell, J., Brelsford, A., Wurm, Y., Perrin, N., & Chapuisat, M. (2014). Convergent genetic architecture underlies social organization in ants. *Current Biology*, *24*(22), 2728–2732. https://doi.org/10.1016/j.cub.2014.09.071

Rabeling, C. (2021). Social parasitism. *Encyclopedia of Social Insects*, 836–858. https://doi.org/10.1007/978-3-030-28102-1_175

Sapp, J. R., Yost, J., & Lyon, B. E. (2020). The socially parasitic ant *Polyergus mexicanus* has host-associated genetic population structure and related neighbouring colonies. *Molecular Ecology*, *29*(11), 2050–2062. https://doi.org/10.1111/mec.15468

Schwander, T., Libbrecht, R., & Keller, L. (2014). Supergenes and complex phenotypes. *Current Biology*, *24*(7). https://doi.org/10.1016/j.cub.2014.01.056

Talbot, M. (1967). Slave-raids of the ant *Polyergus lucidus* Mayr. Psyche: A Journal of
Entomology, 74(4), 299–313. https://doi.org/10.1155/1967/24519

Tanksley, S. D., Bernatzky, R., Lapitan, N. L., & Prince, J. P. (1988). Conservation of gene
repertoire but not gene order in pepper and tomato. *Proceedings of the National
Academy of Sciences*, *85*(17), 6419–6423. https://doi.org/10.1073/pnas.85.17.6419

Topoff, H., Cover, S., Greenberg, L., Goodloe, L., & Sherman, P. (1988). Colony founding by
Queens of the obligatory slave-making ant, *Polyergus breviceps*: The role of the Dufour's
gland. *Ethology*, *78*(3), 209–218. https://doi.org/10.1111/j.1439-0310.1988.tb00231.x

Topoff, H., LaMon, B., Goodloe, L., & Goldstein, M. (1984). Social and orientation behavior of
Polyergus breviceps during slave-making raids. *Behavioral Ecology and Sociobiology*,
15(4), 273–279. https://doi.org/10.1007/bf00292989

Trager, J. C. (2013). Global revision of the dulotic ant genus *Polyergus* (Hymenoptera:
Formicidae, Formicinae, formicini). *Zootaxa*, *3722*(4), 501.
https://doi.org/10.11646/zootaxa.3722.4.5

Tuttle, E. M., Bergland, A. O., Korody, M. L., Brewer, M. S., Newhouse, D. J., Minx, P., Stager, M.,
Betuel, A., Cheviron, Z. A., Warren, W. C., Gonser, R. A., & Balakrishnan, C. N. (2016).
Divergence and functional degradation of a sex chromosome-like supergene. *Current
Biology*, *26*(3), 344–350. https://doi.org/10.1016/j.cub.2015.11.069

Wang, J., Wurm, Y., Nipitwattanaphon, M., Riba-Grognuz, O., Huang, Y.-C., Shoemaker, D. W., &
Keller, L. (2013). A Y-like social chromosome causes alternative colony organization in
fire ants. *Nature*, *493*(7434), 664–668. https://doi.org/10.1038/nature11832

Whitkus, R., Doebley, J., & Lee, M. (1992). Comparative genome mapping of sorghum and maize.
Genetics, *132*(4), 1119–1130. https://doi.org/10.1093/genetics/132.4.1119
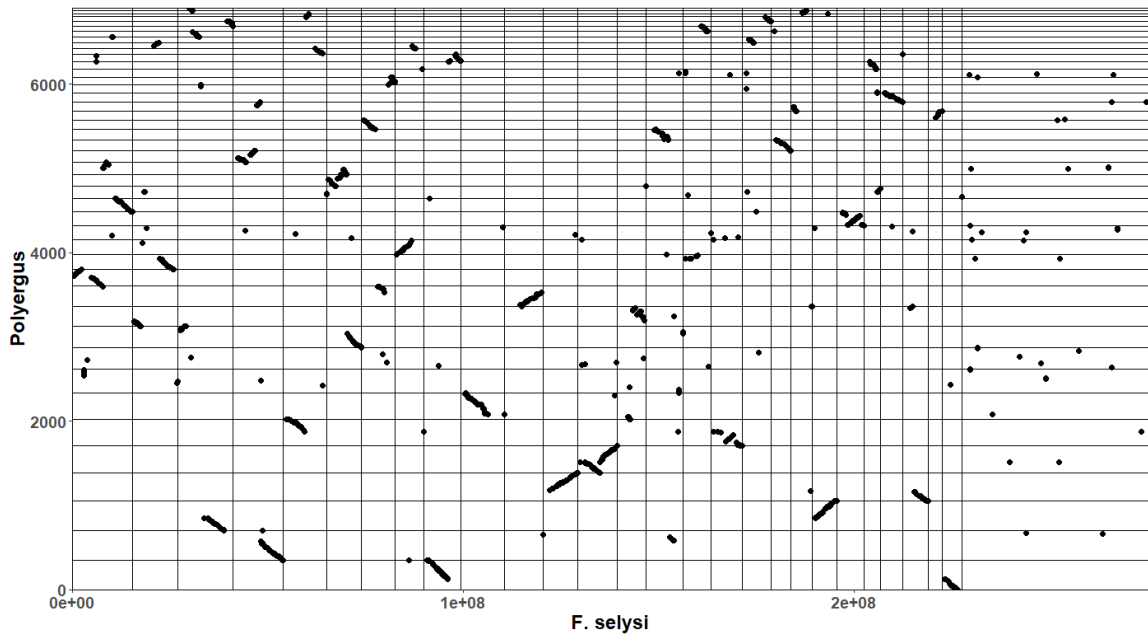
Figures

Figure III.1. An Oxford grid displaying *Polyergus* (y-axis, in centimorgans) and *Formica selysi* (x-axis, megabases) marker data. *F. selysi* data are in chromosomal position (base-pairs) and ordered by chromosome (1-27). The first chromosome is closest to the origin. Vertical lines show where one chromosome begins and the next begins. The last rectangles on the right contain contigs. Polyergus data are in linkage groups, ordered by size (centimorgans, cM). The largest linkage group is closest to the origin. Horizontal lines separate the respective linkage groups (N = 43). Here we can see signatures of synteny, chromosomal inversions, and possible translocations. These genetic themes are illustrated by continuous lines within chromosomes or linkage groups, a change in slope on those lines, and segments of lines on different linkage groups, respectively.

Chapter IV: Conclusion

Formica neoclara is a socially polymorphic ant widely distributed across western North America. Many socially polymorphic *Formica* species have been shown to harbor the social supergene (Purcell et al., 2014; 2021; Brelsford et al., 2020). Since the initial discoveries of social supergenes in *Solenopsis invicta* (Wang et al., 2013) and *Formica selysi* (Purcell et al., 2014) further studies have revealed that congenerics within these respective genera also possess these supergenes (Brelsford et al., 2020; Yan et al., 2020). Such explorations may help us gain further insight into chromosomal and thus genomic evolution (Jay et al., 2021), as well as polygenic traits such as behavior. *F. neoclara* not only provides a means of better understanding the *Formica* supergene and supergenes as a whole, but also grants us insight into the colony and population structure of this widely distributed, yet little-studied species.

Aside from a few studies focusing on the life history (Wheeler and Wheeler, 1963) and agricultural implications (Capinera & Roltsch, 1981) of *F. neoclara*, little was known about it prior to this study. Using genetic data, we were able to answer the three main research questions of the *F. neoclara* study. Formica neoclara is socially polymorphic in queen number, with the polymorphism being under the same genetic control as in previously described Formica species (Purcell et al., 2014; 2021; Brelsford et al., 2021). Additionally, *F. neoclara* populations appear to have a long history of isolation from one another. Within *F. neoclara,* the *Formica* social supergene shows some degree of variation compared to previously described *Formica* species. Unlike species such as *F. selysi* (Purcell et al., 2014), *F. neoclara* polygyne colonies tolerate workers that are homozygous for the monogyne associated haplotype. Additionally, we do not see evidence of any of the transmission control mechanisms seen in other ants

68

possessing supergenes (Avril et al., 2020; Wang et al., 2013; Keller & Ross, 1998). These

differences in genotypic distribution and lack of obvious control mechanisms illustrate that

despite sharing a genus-wide supergene, there can be a degree of variation in phenotype

between species.

The *Formica* social supergene has been observed in all genetically studied *Formica* to

date (Purcell et al., 2014; 2021; Brelsford et al., 2020). Despite variations between species, the

supergene maintains its core function of determining social form - variation in queen number.

As there are supergenes that precede speciation events, it follows that the genetic elements

that compose supergenes in a particular species are present in other related species and genera.

Polyergus is an ant genus sister to the *Formica-Iberoformica* clade, separated by ~33 Ma of

evolution (Borowiec et al., 2021). Comparative linkage mapping is useful for identifying synteny

between species and observing the chromosomal evolution since divergence time and can shed

light on the evolutionary relationships between and changes within taxa (Liu et al., 2017; Boyle

et al., 2021). To date, no *Polyergus* species have been found to exhibit alternative social

organization in the same way that *Formica* does in regard to queen number (Purcell et al., 2014;

Trager, 2013). However, this does not preclude *Polyergus* from experiencing a large degree of

synteny with *Formica* in regard to chromosome 3. By constructing an Oxford grid using genetic

data, I show that *Polyergus* displays both a degree of chromosomal synteny with *F. selysi*, as well

as translocations and inversions. Chromosome 3 in particular displays both synteny and

translocation, having a collinear relationship on the Oxford grid while being on at least seven

distinct *Polyergus* linkage groups. Several questions stand to be answered, such as what the

gene content is on chromosome 3 for the respective taxa and why the chromosomal regions

changed over the course of evolution between these two taxa.

Since the initial supergene concept was described by Fisher in 1930, science has

progressively increased the number of described supergenes within species. Congenerics to

those with supergenes are also being investigated, illustrating that supergenes evolution can

precede speciation events. This is especially true for the social supergenes of the *Solenopsis* and

Formica ant genera (Wang et al., 2013; Brelsford et al., 2020; Yan et al., 2020; Purcell et al.,

2021). *F. neoclara* is yet another example of a species possessing a genus-wide supergene

(Brelsford et al., 2020; Purcell et al., 2021). Future directions should focus not only on other

Formica species potentially harboring this supergene but also the gene content and function of

said supergene. While the *Solenopsis* supergene was initially described as a single odorant

binding protein locus (Keller & Ross, 1998; Wang et al., 2013), any one function of any element

within the *Formica* supergene is yet to be articulated (Purcell et al., 2021). Not only is there

opportunity for investigating supergenes within ants, but for all social insects. Wang et al. (2013)

with their work on the *Solenopsis* supergene proposed that supergenes may play a role in social

organization in other social insects. As supergenes are widespread, presenting as balanced

polymorphisms in the populations of many species, the idea that supergenes may be common in

social insects holds plausibility. Additionally, linkage mapping grants us an opportunity to look

through the passage of time by comparing maps from species of interest to related outgroups.

Future directions will focus on comparing maps within the *Formica* genus as well as outgroups

such as *Iberoformica* and *Cataglyphis*, further illuminating the events surrounding the

evolutionary history of the *Formica* supergene. In particular, highly conserved regions of

chromosome 3 between groups of monophyletic taxa should be analyzed for gene content.

Through our continued understanding of supergenes within various species, we not only

broaden our understanding of polymorphic and polygenic traits, but also the evolution of

genomic elements within and between species.

References

Avril, A., Purcell, J., Béniguel, S., & Chapuisat, M. (2020). Maternal effect killing by a supergene controlling ant social organization. *Proceedings of the National Academy of Sciences*, *117*(29), 17130–17134. https://doi.org/10.1073/pnas.2003282117

Borowiec, M. L., Cover, S. P., & Rabeling, C. (2021). The evolution of social parasitism in *Formica* ants revealed by a global phylogeny. *Proceedings of the National Academy of Sciences*, *118*(38). https://doi.org/10.1073/pnas.2026029118

Boyle, J. H., Rastas, P. M., Huang, X., Garner, A. G., Vythilingam, I., & Armbruster, P. A. (2021). A linkage-based genome assembly for the mosquito *Aedes albopictus* and identification of chromosomal regions affecting diapause. *Insects*, *12*(2), 167. https://doi.org/10.3390/insects12020167

Brelsford, A., Purcell, J., Avril, A., Tran Van, P., Zhang, J., Brütsch, T., Sundström, L., Helanterä, H., & Chapuisat, M. (2020). An ancient and eroded social supergene is widespread across *Formica* ants. *Current Biology*, *30*(2). https://doi.org/10.1016/j.cub.2019.11.032

Capinera, J. L., & Roltsch, W. J. (1981). The Predatory Ant Formica neoclara: Effect on Within-Field Bean Aphid Distribution, and Activity in Relation to Thermal Conditions. *Journal of the Kansas Entomological Society*, *54*(3), 578–586.

Fisher, R. A. (1930). *The genetical theory of natural selection*. Oxford University Press, Oxford.

Jay, P., Chouteau, M., Whibley, A., Bastide, H., Parrinello, H., Llaurens, V., & Joron, M. (2021). Mutation load at a mimicry supergene sheds new light on the evolution of inversion polymorphisms. *Nature Genetics*, *53*(3), 288–293. https://doi.org/10.1038/s41588-020-00771-1

Keller, L., & Ross, K. G. (1998). Selfish genes: A green beard in the red fire ant. *Nature*, *394*(6693), 573–575. https://doi.org/10.1038/29064

Liu, H., Fu, B., Pang, M., Feng, X., Yu, X., & Tong, J. (2017). A High-Density Genetic Linkage Map and QTL Fine Mapping for Body Weight in Crucian Carp (*Carassius auratus*) Using 2b-RAD Sequencing. *G3 Genes|Genomes|Genetics*, *7*(8), 2473–2487. https://doi.org/10.1534/g3.117.041376

Purcell, J., Brelsford, A., Wurm, Y., Perrin, N., & Chapuisat, M. (2014). Convergent genetic architecture underlies social organization in ants. *Current Biology*, *24*(22), 2728–2732. https://doi.org/10.1016/j.cub.2014.09.071

Purcell, J., Lagunas-Robles, G., Rabeling, C., Borowiec, M. L., & Brelsford, A. (2021). The maintenance of polymorphism in an ancient social supergene. *Molecular Ecology*. https://doi.org/10.1111/mec.16196

Schwander, T., Libbrecht, R., & Keller, L. (2014). Supergenes and complex phenotypes. *Current Biology*, *24*(7). https://doi.org/10.1016/j.cub.2014.01.056

Trager, J. C. (2013). Global revision of the dulotic ant genus *Polyergus* (Hymenoptera: Formicidae, Formicinae, formicini). *Zootaxa*, *3722*(4), 501. https://doi.org/10.11646/zootaxa.3722.4.5

Tuttle, E. M., Bergland, A. O., Korody, M. L., Brewer, M. S., Newhouse, D. J., Minx, P., Stager, M., Betuel, A., Cheviron, Z. A., Warren, W. C., Gonser, R. A., & Balakrishnan, C. N. (2016). Divergence and functional degradation of a sex chromosome-like supergene. *Current Biology*, *26*(3), 344–350. https://doi.org/10.1016/j.cub.2015.11.069

Wang, J., Wurm, Y., Nipitwattanaphon, M., Riba-Grognuz, O., Huang, Y.-C., Shoemaker, D. W., & Keller, L. (2013). A Y-like social chromosome causes alternative colony organization in fire ants. *Nature*, *493*(7434), 664–668. https://doi.org/10.1038/nature11832

Wheeler, G. C., & Wheeler, J. (1963). *The ants of North Dakota*. University of North Dakota Press.

Yan, Z., Martin, S. H., Gotzek, D., Arsenault, S. V., Duchen, P., Helleu, Q., Riba-Grognuz, O., Hunt, B. G., Salamin, N., Shoemaker, D. W., Ross, K. G., & Keller, L. (2020). Evolution of a supergene that regulates a trans-species social polymorphism. *Nature Ecology & Evolution*, *4*(2), 240–249. https://doi.org/10.1038/s41559-019-1081-1

Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. Nature Genetics. 2012;44(7):821–4.