Understanding the mechanisms that allow a single genotype to generate multiple phenotypes is a major unsolved question in biology, with important implications for understanding the generation of biodiversity on Earth. This phenomenon, known as phenotypic plasticity, occurs across the tree of life: clonally-propagated plants allocate biomass differentially in response to the environment; human cells, all with the same genotype, differentiate into many tissues; and genetically similar ant larvae develop into distinct castes. Plasticity likely plays an important role in allowing organisms to exploit new niches and facilitates speciation[1]. Understanding the precise genetic mechanisms involved allows us to understand how plasticity shapes evolution, and how evolution shapes plasticity.

Ants (Family: *Formicidae*) are an excellent system in which to study phenotypic plasticity. Colonies consist of individuals with near-identical genomes yet highly differentiated morphologies and behaviors: reproductive, winged queens, and nonreproductive, wingless workers. Their plasticity is also particularly interesting because caste systems vary strikingly across the family, despite having arisen just once in the common ancestor of all ants[2]. This variation allows us to ask a broad array of interesting evolutionary questions to understand the impacts of phenotypic plasticity on molecular evolution.

Question 1: How have parallel elaborations of worker-caste polymorphism (WCP) shaped molecular evolution in the ants? In ants, workers can be monomorphic, dimorphic, or continuously polymorphic; these polymorphisms have evolved independently multiple times[3]. By comparing the genomes of species with different WCP types, we can identify 1) genes that exhibit common shifts in selective pressures, 2) genes that exhibit increased or decreased rates of sequence evolution, and/or 3) changes in gene family size across parallel evolutions of WCP.

Predictions: I expect that progressively more specialized WCPs are associated with increased positive selection, accelerated evolutionary rate, and gene family enrichment in genes related to behavior, the nervous system, and metabolism, and in transcription factors and other regulatory proteins[4]. Overall, I predict increased rates of gene family expansion associated with >1 worker caste, providing the substrate for regulatory flexibility in development. If I do not find these patterns, my results will indicate that protein-coding evolution is less important than cis-regulatory evolution in producing plastic phenotypes.

Approach: During my first year of graduate study, I **developed and tested a bioinformatic pipeline** consisting of custom Bash and Python scripts which run: Orthofinder[5], to identify orthologous genes; InterProScan[6], to assign gene functions; KinFin[7], to correlate gene family enrichment with phenotype; RERconverge[8], to correlate evolutionary rate with phenotype; and BUSTED[S][9], to detect positive selection. I have tested this pipeline with available ant genomes; **my preliminary results support the hypothesis** that more elaborate WCP is associated with gene family expansion. However, this question cannot be resolved with currently sequenced genomes, 75% of which come from monomorphic species with very low sampling from other WCP types. To overcome this limitation, I will leverage my advisor Dr. Corrie Moreau's connection to the Global Ant Genome Alliance (GAGA) and incorporate new GAGA-generated sequences from species with missing WCP types into my analysis.

Question 2: How has the evolution and elaboration of a novel worker caste polymorphism (WCP) in the genus *Cephalotes* impacted protein-coding and gene-regulatory evolution? The most recent common ancestor of all ants lived ~140-168 MYA[10]; this ancient divergence time makes studying regulatory evolution with a comparative approach difficult, as regulatory sequences (inferred from noncoding alignable regions[11]) may be quite diverged between species. Therefore, I propose using comparative genomics to describe both coding and non-coding evolution in the genus *Cephalotes*, which originated ~65 MYA[10]. The genus includes three WCP types: monomorphic species, the ancestral state; dimorphic species that have evolved a novel soldier caste characterized by unique head morphology; and polymorphic species, whose workers span a range of sizes and morphology.

Predictions: I expect increased positive selection and gene family enrichment for genes related to the nervous system, cuticle development and morphology in species with soldiers. I expect more derived plastic phenotypes to be associated with accelerated evolutionary rate in putative regulatory regions.

Approach: I will sequence the genomes of key species: *C. placidus* and *C. unimaculatis* (monomorphic); *C. opacus* (dimorphic); and *C. atratus* and *C. minimus* (polymorphic). My lab has previously sequenced a high-quality genome for *C. varians*, a second dimorphic species that will be the

reference genome for alignment. All genomes will be made publicly available as part of my data sharing plan. I will test for positive selection, gene family enrichment, and evolutionary rate variation in noncoding alignable regions[11] in association with the three WCP types.

Question 3: What genetic mechanisms underlie worker differentiation in a strongly dimorphic ant? Larval development is the process that translates a single embryonic genome into the differentiated adult's morphology and behavior. To understand how different plastic phenotypes evolve, we must understand the genetic basis of caste differentiation. Most work on ant phenotypic plasticity compares queens and workers, so findings center on genes underlying reproduction. Worker subtypes are less well studied, though experiments in late-stage *Pheidole* larvae and adult *Camponotus* ants implicate juvenile hormone (JH) as a regulator of worker differentiation[12,13]. However, early genetic determinants are unknown, and caste fate cannot be predicted until late in development. I will address this knowledge gap by identifying gene expression patterns that predict worker caste fate in dimorphic *Cephalotes varians*.

Predictions: I expect that the expression patterns of known genes like JH synthetase[12], *vestigial*[14], and *foraging*[15], will predict caste fate. I also expect to identify many novel genes and gene modules that are predictive of caste fate. Lack of evident developmental trajectory is unlikely; at some time point in development, morphological differences must arise by differential gene expression. However, if we do not find early predictive genes, this would suggest social control of caste fate by provision of external substances by nurse ants, providing a new avenue for further research.

Approach: Previous gene expression studies have been limited by coarse temporal sampling and use of whole-tissue RNAseq. To address this limitation, I will collect single-cell RNA sequencing data from larval, pupal, and newly eclosed adult ant brains and imaginal discs (5 animals/time point $\times$ tissue[16,17]) throughout development. I will use pseudotemporal trajectory analysis to reconstruct the developmental trajectory of each caste[18]. This is a **novel approach** that has not been used in a nonmodel insect. To accomplish this project, I will work with the Cornell Genomics Innovation Hub, and a key collaborator and dissertation committee member, Dr. Leslie Babonis, who is exploring similar methods in cnidarians.

Intellectual merit: My proposed research improves our understanding of the evolution of and mechanisms producing phenotypic plasticity, contributing to **Understanding the Rules of Life, one of NSF's 10 Big Ideas**. My bioinformatic pipeline establishes an **efficient and, crucially, reproducible workflow** that other researchers can use for comparative genomics studies in any system. Questions 2 and 3 also **develop the *Cephalotes* genus as a novel system** for work on plasticity and development, furthering research and collaboration on fundamental questions in biology and myrmecology.

Broader impacts: As part of Question 1 and 2, I am developing a **web-based teaching tool** for high school and undergraduate biology educators that incorporates both introductory computational skills and molecular evolution content. **Deliverables** include a website, hosting a tutorial for my comparative genomics pipeline and lessons covering evolutionary biology content, and a Docker container in which students can use the pipeline for their own projects. Activities map to the **Next Generation Science Standards** for high school students. I am working with **Arti Jewett, the Ithaca HS biology department leader**, and **Dr. Michelle Smith, a discipline-based education researcher** in my department as I continue to develop both the content and associated learning assessments to benchmark efficacy.

To increase STEM accessibility, I will create and supervise independent study projects related to Question 3 for students from New College of Florida (NCF), through the college's **Alum Fellows program**. NCF is an 800 student, undergraduate liberal arts college which, due to size, has limited infrastructure to support advanced genomics research. This will create an opportunity for undergraduates to receive training in animal rearing, genetics/genomics, and bioinformatics research.

---

[1]West-Eberhard, PNAS 2005.[2]Hölldobler & Wilson, The Ants 1990. [3]Blanchard & Moreau, Evolution 2016. [4]Favreau et al., Curr Opin Insect Sci. 2018. [5]Emms & Kelly, Gen Biol. 2019. [6]Jones et al., Bioinf. 2014. [7]Laetsch & Blaxter, G3 2017. [8]Kowalczyk et al., Bioinf. 2019. [9]Wisotsky et al., Mol Bio. 2020. [10]Moreau et al., Science 2006. [11]Rubin et al., Philos Trans R B 2019. [12]Rajakumar et al., Science 2012. [13]Glastad et al., Mol Cell 2020. [14]Rajakumar et al., Nature 2018. [15]Ingram et al., BMC Ecol. 2011 [16]Hammond et al., Immunity 2019 [17]Jevitt et al., PLoS Biol. 2020. [18]Lescroart et al., Science 2018.