Caste development and evolution in ants: it’s all about size

Waring Trible* and Daniel J. C. Kronauer

ABSTRACT

Female ants display a wide variety of morphological castes, including workers, soldiers, ergatoid (worker-like) queens and queens. Alternative caste development within a species arises from a variable array of genetic and environmental factors. Castes themselves are also variable across species and have been repeatedly gained and lost throughout the evolutionary history of ants. Here, we propose a simple theory of caste development and evolution. We propose that female morphology varies as a function of size, such that larger individuals possess more queen-like traits. Thus, the diverse mechanisms that influence caste development are simply mechanisms that affect size in ants. Each caste-associated trait has a unique relationship with size, producing a phenotypic space that permits some combinations of worker- and queen-like traits, but not others. We propose that castes are gained and lost by modifying the regions of this phenotypic space that are realized within a species. These modifications can result from changing the size–frequency distribution of individuals within a species, or by changing the association of tissue growth and size. We hope this synthesis will help unify the literature on caste in ants, and facilitate the discovery of molecular mechanisms underlying caste development and evolution.

KEY WORDS: Body size, Evolutionary developmental biology, Formicidae, Intercaste, Phenotypic plasticity, Polyphenism

Introduction

Many insect species, such as locusts and dung beetles, display morphological polyphenisms (Simpson et al., 2011; Hartfelder and Emlen, 2012). Perhaps the most elaborate polyphenisms are observed in ants, in which female individuals can develop into a wide range of morphological castes, including workers, soldiers, ergatoid queens and queens (Fig. 1; Molet et al., 2012). Development into alternative castes is caused by diverse factors in different ant species, ranging from strictly environmental to genetic (Brian, 1979; Wilson and Hölldobler, 1990; Anderson et al., 2008; Schwander et al., 2010). Castes have been gained and lost repeatedly across the ant phylogeny, and often exhibit striking convergent and parallel evolution, exemplified by multiple independent origins of soldiers or independent losses of workers in socially parasitic species (Oster and Wilson, 1978; Bourque and Franks, 1991; Molet et al., 2012; Peeters, 2012). The phenomenon of castes in ants has attracted much attention, as it touches on subjects including epigenetics, developmental plasticity, developmental constraints and sympatric speciation (Bourque and Franks, 1991; West-Eberhard, 2003; Bonasio, 2012; Rajakumar et al., 2012; Rabeling et al., 2014). Despite many years of investigation, however, the mechanisms underlying both caste development and evolution remain poorly understood.

The ant fossil record contains multiple extinct lineages of early ants with winged queens and wingless workers, so it is likely that the most recent common ancestor of modern ants also possessed these two castes (Barden and Grimaldi, 2016). Many researchers have discussed how various groups of ants depart from this ancestral system, for example via the loss of the worker or queen caste or the evolution of novel castes from workers, queens or intercastes (worker–queen intermediates) (Emery, 1894; Wilson, 1953; Oster and Wilson, 1978; Bourque and Franks, 1991; Wheeler, 1991; Nonacs and Tobin, 1992; Urbani and Passera, 1996; Ward, 1997; Urbani, 1998; Molet et al., 2012; Peeters, 2012; Rajakumar et al., 2012). Theories of caste development have also been proposed, describing how alternative caste morphology can arise from differential growth of imaginal disks, and how alternative caste development can be induced by genetic rather than environmental factors (Wilson, 1953; Wheeler, 1991; Anderson et al., 2008; Schwander et al., 2010). These theories provide compelling partial explanations of development and evolution in certain types of caste systems, but a coherent synthesis is still lacking.

Here, we attempt to unify the literature on ant caste development and evolution with a single theoretical framework, drawing on recent advances in evolutionary developmental biology (Carroll et al., 2001; West-Eberhard, 2003; Hartfelder and Emlen, 2012; Bopp et al., 2014; Wagner, 2014; Londe et al., 2015; Xu et al., 2015). We propose that ants possess a developmental spectrum of phenotypes, with increasing body size associated with increasingly queen-like traits (Fig. 2). We define body ‘size’ as pupal mass, but it may also be estimated by adult dry mass or body length. Measurements of body size should exclude developmentally irrelevant factors that might affect mass or volume measurements, such as gut contents or developing eggs inside ovaries (Brian, 1974; Honék, 1993). In many species, the size–frequency distribution is bimodal, with one mode of fully worker-like individuals and a second mode of fully queen-like individuals (Wilson, 1953; Wheeler, 1991; Fjerdingstad and Crozier, 2006). Species gain or lose castes over evolutionary time either by modifying this size–frequency distribution or by modifying the relationship between tissue growth and size (Fig. 3). We hope that this review will clarify understanding of caste development and evolution, and provide a framework for discovering the mechanisms that generate morphological diversity in ants as well as other types of insects.

A hierarchical continuum of phenotypes

The definition of castes has been a contentious subject (Wilson, 1953; Boven, 1970; Peeters and Crozier, 1988; Urbani, 1998). Here, we define castes as any sets of adult female ants within a species that can be morphologically distinguished (Peeters and Crozier, 1988). Following Molet et al. (2012), we primarily classify castes into four categories: workers, soldiers, ergatoid queens and queens (Fig. 1). Workers and queens differ by a suite of caste-associated traits (Molet et al., 2012). Workers are universally wingless and have reduced or absent eyes, ocelli and reproductive systems. Queens typically possess wings, large eyes and ocelli, and a well-developed...
reproductive system with a spermatheca (a sperm storage organ) and a large number of ovarioles. Ergatoid queens lack wings but have a well-developed reproductive system (Peeters, 1991). Soldiers are typically defined by allometry, with a large head size relative to body size, and they also lack wings (Wilson, 1953). These categories are useful for general discussion, but caste morphology is often continuous and will not always fit into a given classification system with discrete categories.

To our knowledge, the queen caste within an ant species is the largest and has the greatest mass, on average (see comparison of 10 subfamilies in Peeters and Ito, 2015). In many species, each caste occupies a fairly discrete mode in the size–frequency distribution, and intercastes, individuals intermediate in size between castes, are rare or absent. However, some species exhibit a continuous range of sizes from workers to queens (Molet et al., 2012; Londe et al., 2015). Fig. 2 summarizes the most detailed available data on size and phenotype for ant workers, intercastes and queens (based on studies of Myrmica rubra, Myrmiria rogeri, Pristomyrinx punctatus, Technomyrinx albipes and Temnothorax nylanderi) (Brian, 1955, 1956; Tsuji et al., 1991; Tsuji and Dobata, 2011; Okada et al., 2013). Note that Fig. 2 does not depict a developmental series, but a series of adult phenotypes at varying sizes. Comprehensive data are only available for a few species, but similar patterns are observed in many additional species and we are not aware of any exceptions to this general pattern (see also Tulloch, 1930; Hall and Smith, 1953; Bolton, 1986; Heine and Buscylinder, 1989; Ohkawara et al., 1993; Ito et al., 1994; Miyazaki et al., 2005). Intercastes possess a combination of worker- and queen-like features, illustrating that discrete workers and queens simply constitute extreme outcomes along a developmental continuum (Londe et al., 2015).

These data show that phenotypes progress from worker- to queen-like states as individuals become larger. Rather than changing abruptly at a single size threshold, different caste-associated traits switch from worker- to queen-like states at different sizes (Londe et al., 2015). For example, M. rubra intercastes can have worker-like eyes and wing morphology, but a queen-like reproductive system (Brian, 1955). As intercastes progress from slightly larger than workers to slightly smaller than queens, they first begin to develop a more queen-like reproductive system, then queen-like eyes and ocelli, and finally wings (Fig. 2). This is a hierarchical progression: within a species, individuals with queen-like ovarioles may have otherwise worker-like features, but individuals with queen-like wings also have queen-like eyes, ocelli and ovarioles. Some flexibility exists in this hierarchy: in P. punctatus and T. albipes, the smallest class of intercastes has additional ovarioles but no ocelli, while in M. rubra, the smallest class of intercastes has an additional ovariole and partially developed ocelli (Brian, 1956; Tsuji et al., 1991; Tsuji and Dobata, 2011). This flexibility is constrained, however: to our knowledge, within a species, smaller individuals on average never possess more queen-like traits than larger individuals.

The definitions of ‘caste determination’ and ‘caste differentiation’ are not consistent in the ant literature, which has contributed to confusion in the understanding of caste development. Here, we borrow from the sex development literature to provide definitions that are explicitly tied to biological processes. The available data reveal that caste-associated traits vary as a function of size: larvae that pupate at greater mass become adults with more queen-like traits. In light of this evidence, caste determination can be defined as the complex integration of many mechanisms that influence the size of larvae at pupation. The size-frequency distribution, therefore, is the outcome of caste determination across all of the larvae in a colony (Wilson, 1953; Wheeler, 1991). Caste differentiation, in contrast, is the developmental mechanism producing the association of size and phenotype. It is specifically the process by which tissues in developing larvae and

Fig. 1. Caste polymorphism in two army ants. (A) Syscia augustae has a caste system similar to the ancestral condition in ants, displaying workers (right) and queens (left) with modest differences in size and morphology. (B) Eciton burchellii has a derived caste system, displaying workers (top), subsoldiers (middle right), soldiers (middle left) and wingless ergatoid queens (bottom). Eciton burchellii exhibits extreme variation in size and morphology. Massive ergatoid queens in E. burchellii and some other army ants likely represent a secondary increase in size relative to recently evolved ergatoid queens in groups such as Monomorium salomons or Harpagoxenus subdivaevis, which are typically smaller than winged queens (Buscylinder and Winter, 1978; Bolton, 1986; Peeters, 2012). Photographs, D.J.C.K.

Fig. 2. Schematic representation of the progression of size, caste and phenotypic state in ants. Size is estimated by body mass and/or length. Phenotypic state is estimated by either the number or the presence/absence of ovarioles, eyes/ocelli and wings, and/or by measurements of abdomen, head and thorax width (Brian, 1955, 1956; Tsuji et al., 1991; Tsuji and Dobata, 2011; Okada et al., 2013). We assume that increased body mass correlates with increased body length, and that increased ovariole, eye/ocellus and wing development correlate with increased abdomen width, head width and thorax width, respectively.
pupae differentiate in a worker- or queen-like manner in association with size (Miyazaki et al., 2010).

Caste determination and differentiation position individuals within the phenotypic space of Fig. 2. Caste determination results in an individual’s size, or x-axis position, while caste differentiation results in the phenotypic states associated with that size, or y-axis position. In this space, certain combinations of phenotypes are possible (such as worker-like eyes, ocelli and wings, combined with a queen-like reproductive system), but other combinations of phenotypes are not (such as queen-like eyes, ocelli and wings, combined with a worker-like reproductive system) (Fig. 2; Londe et al., 2015). In this review, we will discuss how this simple model encompasses the available data on caste development, and how it can help explain caste evolution. In the ‘Caste determination’ and ‘Caste differentiation’ sections, below, we show how caste determination and differentiation position individuals within the phenotypic space to produce worker and queen castes in a variety of ant species. In the subsequent ‘Caste evolution’ section, we show how caste evolution occurs by modifying caste determination and differentiation to alter the phenotypes realized within a given species. For simplicity, we will focus on the ancestral worker and queen castes in the ‘Caste determination’ and ‘Caste differentiation’ sections, and limit our discussion of soldiers and ergatoid queens to the ‘Caste evolution’ section.

Caste determination

Determining adult size

Because larger size is associated with more queen-like phenotypes, understanding the mechanisms of caste determination requires understanding the mechanisms that affect size in ants. Many of these mechanisms also affect size in other insects. In holometabolous insects, growth occurs during the larval stage and adult size is dictated by the size at pupation (Nijhout and Callier, 2015). Consequently, high nutrition during larval growth in ants can induce queen development, and starvation can prevent it (Brian, 1956, 1979). However, high nutrition during larval growth does not necessarily result in large adult size, as variation in developmental timing (which can result from genetic and other factors) can cause individuals under poor nutritional conditions to prolong development and attain large size, or individuals under rich nutritional conditions to truncate development and attain small size (Brian, 1956, 1974).

Two endocrine hormones, juvenile hormone (JH) and ecdysone, are important regulators of developmental timing in insects. Pupation (and the cessation of larval growth) is triggered by ecdysone, and the action of ecdysone is inhibited by JH (Hiruma and Kaneko, 2013). Thus, treatment of larvae with JH or JH analogs leads to prolonged growth and larger size, as long as necessary nutrition is present. JH removal and/or ecdysone treatment leads to premature pupation and smaller adult size (Hiruma and Kaneko, 2013). This endocrine regulation of insect size has been demonstrated in Coleoptera and Lepidoptera, and is also supported by experiments in ants. JH treatment in Harpegnathos saltator and M. rubra, for example, prolongs larval development, allowing increased feeding (Brian, 1974; Penick et al., 2012). This in turn translates to larger size at pupation and queen development (Brian, 1974; Penick et al., 2012). In M. rubra, it has been further shown that ecdysone treatment has little effect while JH is present, but leads to premature pupation later in development after JH has been cleared (Brian, 1974). This demonstrates that nutrition and the endocrine system are important regulators of size and caste in ants.

Many factors, including genotype, maternal effects, wounding, temperature and rearing environment can influence nutrition and/or the endocrine system, affecting size and caste development. In M. rubra, larvae do not have the potential for queen development under normal conditions (Brian, 1979). However, larvae that are forced to diapause just before pupation reorganize their endocrine organs and delay pupation after growth is resumed (Brian, 1979). If sufficient nutrition is available, these totipotent larvae attain a large size and develop into queens (Brian, 1956). Many factors can prevent totipotent larvae from developing into queens, including sub-optimal rearing temperature, aggression from workers, parasite infection and small colony size (presumably leading to low nutrition) (Brian, 1953, 1973a,b, 1974, 1979). These factors appear to prevent queen development by reducing size, instead inducing worker or intercaste development (Brian, 1955, 1956).

A variety of studies corroborate the generality of Brian’s findings from M. rubra. For example, aggression prevents queen development in H. saltator, and sub-optimal rearing temperature prevents queen development in Plagiolepis pygmaea and Pheidole pallidula (Brian, 1979; Penick and Liebig, 2012). Large colony size (which is associated with high worker:larva ratios and therefore may be associated with higher nutrition) is a necessary condition for queen production in a wide variety of ant species (e.g. Brian, 1979; Tschinkel, 2006; Schwander et al., 2008; Aron et al., 2011).

The queen can also control caste development. The presence of reproductive queens is known to inhibit the production of new queens in a number of ant species (Brian, 1979). Maternal factors deposited into eggs, such as vitellogenin, ecdysone and JH, can affect size and bias larvae toward worker or queen development. In P. pallidula, overwintered queens produce eggs with a potential for queen development, possibly due to increased JH in the eggs (Passera and Suzzoni, 1979). In Pogonomyrmex rugosus, queens must be both mature and overwintered to lay eggs with a potential for queen development, possibly due to differences in ecdysone or vitellogenin levels in the eggs (Schwander et al., 2008; Libbrecht et al., 2013).

In species with genetic caste determination, some genotypes are biased toward worker or queen development, and in extreme cases only have the potential to develop into one caste (Anderson et al., 2008; Schwander et al., 2010). Based on current evidence, it is likely that these genotypes typically influence caste by affecting size, similar to maternal effects or other caste determination factors. For example, P. rugosus larvae with queen-biased genotypes only become queens if sufficient nutrition is available. Under conditions of low nutrition they die or, rarely, become workers or intercastes, rather than becoming worker-sized queens (Anderson et al., 2008). A similar phenomenon may occur in Cataglyphis cursor, in which queen-biased eggs are produced in the early spring. In small colonies, these queen-biased eggs develop into workers or die, but in large colonies they develop into queens (presumably as a result of improved nutrition) (Aron et al., 2011). These genotypes may therefore influence caste by restricting the size range, with worker-biased genotypes producing individuals whose growth is stunted, and queen-biased genotypes producing individuals that are starvation intolerant (and therefore either attain large size or die). Both types of genotypic effects on size have been observed in Drosophila melanogaster (Kramer et al., 2008; Zhang et al., 2009).

Some of the caste determination mechanisms described above, such as genetic caste determination, can operate before an egg is even laid, while others, such as differences in hormone titers, can operate at the very end of larval development (Brian, 1974; Anderson et al., 2008; Schwander et al., 2010). Caste determination is the process that determines the ultimate size of a larva, and therefore begins to operate
Generating the size–frequency distribution

The colony-wide outcome of caste determination is the size–frequency distribution of adults. The size–frequency distribution has been employed to describe worker polymorphism, but queens are rarely included in these studies (Wilson, 1953; Wheeler, 1991; Huang and Wheeler, 2011). However, it is clear that workers and queens are typically dimorphic and queens are larger, so it follows that the size–frequency distribution is typically bimodal, with one mode for workers and one for queens, and few or no intermediates (Fjerdingstad and Crozier, 2006; Møllet et al., 2012; Peeters and Ito, 2015). Given that the extent of queen–worker dimorphism has independently increased in multiple ant lineages, it is likely that non-normal size–frequency distributions have also evolved multiple times and may be produced by different mechanisms (Fjerdingstad and Crozier, 2006; Keller et al., 2014). The size–frequency distribution arises from interactions between larval and worker traits (Links-vayer and Wade, 2005; Links-vayer, 2007; Tesse et al., 2014). Thus, the evolution of non-normal size–frequency distributions is a product of both the intrinsic development of larvae and extrinsic influences on these larvae (e.g. high nutrition of some larvae but not others), often resulting from worker behavior (Links-vayer and Wade, 2005; Links-vayer, 2007; Tesse et al., 2014).

In honeybees, which have evolved worker and queen castes similar to ants, larvae are reared individually in small or large wax cells. Larvae that are placed in large cells receive higher nutrition, resulting in a discrete queen caste. If larvae are reared outside of the colony, a full range of intermediates between workers and queens is observed (Links-vayer et al., 2011; Leimar et al., 2012). Such spatial segregation is less clear in ants, but temporal segregation is known to produce caste dimorphism: many species rear queens from cohorts of larvae that receive high nutrition. In the army ant Eciton burchellii (Fig. 1B), queens are produced together with males in special sexual broods that contain only ∼1% of the larvae of worker broods, so queen development may result from increased nutrition of female larvae in these cohorts (Schneirla, 1971). In many other species, queens are produced at the time of year when resources are most abundant (e.g. Hart and Tschinkel, 2012; Murdock and Tschinkel, 2015).

Non-normal size–frequency distributions may also arise when discrete events reprogram larval development. Often called ‘developmental switches’, it is important to note that these switches, if they even exist (see Links-vayer et al., 2011; Leimar et al., 2012), are of size (caste determination), not morphology (caste differentiation). In M. rubra, overwintered larvae have prolonged development, allowing them to attain larger size (Brian, 1956). In P. rugosus, only the first eggs laid by overwintered queens have the potential for queen development (Libbrecht et al., 2013). Overwintering is a discrete event, ubiquitous in temperate regions, so it may have been co-opted multiple times by natural selection to produce non-normal size–frequency distributions in ants.

A second promising mechanism for reprogramming development is Dyar’s rule (Dyar, 1890; O’Neal and Markin, 1975; Wheeler, 1990). Dyar’s rule states that larval growth occurs at a constant ratio between instars, i.e. the head width of larvae at the end of an instar is 1.4× the head width at the end of the previous instar (Dyar, 1890; O’Neal and Markin, 1975). This leads to exponential growth between instars, allowing larvae with small differences in absolute size at the end of the first instar to develop into larvae with large differences in absolute size at the end of the final instar (ants possess 3–5 instars) (Wheeler and Wheeler, 1976). This mechanism may amplify non-normal size–frequency distributions in some ant species. In the fire ant Solenopsis invicta, first instar larvae from incipient colonies are substantially smaller than first instar larvae from mature colonies, and give rise to tiny minim workers rather than regular workers or queens (O’Neal and Markin, 1975), suggesting that maternal effects are amplified via Dyar’s rule. Furthermore, within cohorts, the size of worker larvae at the end of the third instar is correlated with size at the end of the fourth instar (O’Neal and Markin, 1975; Wheeler, 1990). This within-cohort size variation could be due to any mechanism that produces size variation early in development. Thus, maternal effects (including queen age and overwintering), as well as early larval nutrition, may produce small size variation in early development that, through Dyar’s rule, eventually leads to large size variation at the time of pupation (O’Neal and Markin, 1975).

Caste differentiation

As seen in the previous section, caste determination encompasses a wide range of mechanisms that produce size variation in ants, and operates at any time until pupation, when adult body size becomes fixed. We propose that caste differentiation subsequently occurs in association with size, with tissues developing in a more queen-like manner in larger individuals. Caste differentiation, therefore, likely operates when adult tissues are formed, largely during the pre-pupal and early pupal stages.

Tissues do not necessarily respond to size directly. To our knowledge, castes are always different in size on average, but in some species the size ranges of two castes may overlap slightly (Brian, 1955, 1956, 1979; Peeters, 1991; Tsuji et al., 1991; Urbani, 1998). We propose that caste determination mechanisms cause correlated variation in size and some unknown caste differentiation factors, which in turn produce variation in tissue growth. An imperfect association of size and caste differentiation factors will produce some size overlap between castes. Such dissociation may result from an imperfect match of size and caste differentiation factors, or from differences in the time when factors affect size and tissue growth.

If the association of size and phenotype can be experimentally uncoupled, it may be possible to implicate the role of specific molecules in caste differentiation. In stag beetles, large males have disproportionately large mandibles. Analogously with ants, JH treatment of small male stag beetle larvae causes them to attain large size with large mandible growth (Gotoh et al., 2011, 2014). However, if small male stag beetle pupae are treated with JH, this results in an increase in mandible length without affecting overall body size (Gotoh et al., 2011, 2014). Thus, it seems that in stag beetles JH retains its typical effect on overall size during larval development, but during pupal development it also acts at the tissue level to influence mandible growth without affecting size. JH has been shown in ants to induce queen development, and in some cases it has been shown that this is correlated with an increase in size (Brian, 1974; Penick et al., 2012). In H. saltator, JH-treated larvae that are wounded develop into workers, rather than worker-sized queens, raising the possibility that JH primarily affects size (Penick and Liebig, 2012). However, in most studies, the size of JH-induced castes has not been reported, so it is possible that JH may also affect growth of specific tissues directly, as in stag beetles (Gotoh et al., 2014). In general, the association of size and caste in ants has not been sufficiently appreciated, and none of the caste determination factors in the ‘Caste determination’ section, above, have been tested for an effect on caste differentiation.
In developing insect larvae, cell primordia of adult tissues are sequestered in imaginal disks, which differentiate into adult tissues at the late larval/prepupal stage of development. Caste-associated phenotypes (variation in ovarioles, spermatheca, ocelli, eyes and wings) result from variation in development of these imaginal disks (Wilson, 1953; Brian, 1979; Miyazaki et al., 2010). The suite of queen-like phenotypes is similar to the ancestral state of holometabolous insects, with fully developed adult tissues, while the suite of worker-like phenotypes results primarily from reduced growth of caste-associated tissues. This type of growth variation is well known from a variety of insect species. Oariole number varies with body size across a wide range of taxa, and in D. melanogaster this is mediated by the insulin/insulin-like signaling pathway (IIS) (Honěk, 1993; Honegger et al., 2008; Tu and Tatar, 2003; Green and Extavour, 2014). IIS may also be responsible for the reduction in eye size and wing size observed in small D. melanogaster individuals (Tang et al., 2011). More extreme variation in tissue growth has been observed in other insects, where it is associated with IIS, but also JH, edysone and even the hedgehog signaling pathway (Lobbia et al., 2003; Emlen et al., 2012; Gotoh et al., 2014; Nitsu et al., 2014; Kijimoto and Moczek, 2016; Xu et al., 2015; see also the excellent discussion in Zinna et al., 2016). Any or all of these factors (and others yet to be discovered) are candidate caste differentiation factors in ants.

Caste differentiation factors can influence tissue growth at different times in development (Miyazaki et al., 2010). This effect is clearly seen in the wingless castes in the genus Pheidole, in which most species possess workers, soldiers and queens, but some species also have a super-soldier caste that is intermediate in size between soldiers and queens (Huang and Wheeler, 2011). In super-soldiers, all four wing disks are present at the time of pupation; in soldiers, only two wing disks are present; and in workers, no wing disks are present (Abouheif and Wray, 2002; Rajakumar et al., 2012). Fitting with this variation in the timing of wing disk loss, aberrant expression of the gene network underlying wing development in wingless castes occurs at more downstream parts of the pathway in larger castes (Abouheif and Wray, 2002; Rajakumar et al., 2012; Favé et al., 2015). It was proposed that selection or drift has caused wing loss to occur via different mechanisms in different wingless castes (Abouheif and Wray, 2002; Favé et al., 2015). We propose a more parsimonious and mechanistic model: the same caste differentiation factors lead to wing loss in different wingless castes, but levels of these factors vary in individuals of different sizes, changing the specific timing and pattern of wing degeneration that is observed.

Caste evolution

The last common ancestor of extant ants likely possessed morphologically distinct worker and queen castes, but many species have subsequently gained or lost castes (Molet et al., 2012; Keller et al., 2014; Barden and Grimaldi, 2016). Additionally, the degree of queen–worker dimorphism has increased in some lineages of ants (Keller et al., 2014). In the earlier section ‘A hierarchical continuum of phenotypes’, we presented a phenotypic space that permits some combinations of worker-like and queen-like phenotypes but not others (Fig. 2). The novel castes that have evolved in a wide range of ant species seem to always fit within the phenotypic space represented in Fig. 2 (Molet et al., 2012). Soldiers and ergatoid queens are phenotypically intermediate between workers and queens and, within a species, are never larger than queens or smaller than workers on average (Wilson, 1953; Peeters, 1991; Molet et al., 2012). Ergatoid queens possess queen-like ovarioles, but can otherwise range from worker- to queen-like in caste-associated features (Molet et al., 2012). Soldiers similarly range from worker- to queen-like, but do not necessarily possess ovarioles and may thus be more worker-like than ergatoid queens (Wilson and Hölldobler, 1990; Bolton and Ficken, 1994; Molet et al., 2012). In other cases, soldiers are very queen-like, with a large number of ovarioles, well-developed eyes and even wing disks (Fig. 2; Abouheif and Wray, 2002; Molet et al., 2012; Peeters et al., 2013). We propose that positions within the phenotypic space can be modified via caste determination, i.e. by changing the size–frequency distribution, and via caste differentiation, i.e. by changing the association of tissue growth and size (Fig. 3).

Caste evolution via caste determination

As with caste development, most of the attention on caste evolution has centered on caste determination. Castes can be gained via caste determination if some individuals occupy a new position in the phenotypic space (Fig. 3A). In a phylogenetic analysis, clades with a greater size difference between workers and queens were more likely to possess additional castes, supporting the notion that novel castes can evolve from worker–queen intermediates (FjerdPKG

Modified caste determination mechanisms can also lead to caste loss (Fig. 3A). In many species lacking winged queens (or lacking queens entirely), the maximum size simply seems to have been reduced (Peeters, 2012). Genetic caste determination in Harpagoxenos
Caste evolution via caste differentiation

Caste systems evolve via changes to caste differentiation when the association between size and phenotype is modified (Fig. 3B). The role of caste differentiation in caste evolution has received relatively little attention, but in theory castes can be gained or lost via changes to caste differentiation in ways analogous to caste determination (Fig. 3B). For example, it is possible that the worker caste could be lost by causing worker-sized individuals to develop queen-like morphology without changing the underlying size–frequency distribution. However, in practice, it appears that when caste differentiation is associated with caste gain or loss, changes to caste determination are typically also observed (see the following section). We will first discuss modifications to caste systems that are associated primarily with caste differentiation, which tend to involve more subtle changes to caste frequency or morphology based on current evidence.

If the association of tissue growth and size is modified, this will affect the caste system of a species without necessarily changing the size–frequency distribution. In Pheidole, the threshold size for soldier development is labile, both within and between species (Yang et al., 2004; McGlynn et al., 2012). In populations of Pheidole morrisi with a higher threshold size for soldier development, the worker size range expands to encompass larger individuals, leading to the development of more workers relative to soldiers (Yang et al., 2004). In a comparison of 26 Neotropical Pheidole species, those with larger workers also had larger and fewer soldiers (McGlynn et al., 2012). These changes in size of workers and soldiers may also change their caste morphology by modifying their position in the phenotypic space (i.e. larger soldiers may have more queen-like morphology, Fig. 2). Therefore, caste size, frequency and morphology evolve in Pheidole via changes to caste differentiation.

Modifications to caste differentiation may also include changes in the magnitude of phenotypic variation across the size range of a species. For example, in some species, both workers and queens possess eyes, while in others, queens possess eyes and workers are completely eyeless (Fig. 1; Wilson and Hölldobler, 1990; Bolton and Ficken, 1994). In still others, both workers and queens are eyeless (Wilson and Hölldobler, 1990; Bolton and Ficken, 1994). Similar variation is seen in ovariole number, ranging from modest to massive differences between workers and queens (Peeters and Ito, 2015). This may result from changes in the degree of size variation between workers and queens, or from changes in the slope of the functions of tissue growth and size. If the slope for a given trait is flat, small amounts of phenotypic variation will be observed over a given increment of size. If the slope is steep, however, large amounts of phenotypic variation will be observed over a given increment of size (Emlen and Nijhout, 2000).

Finally, caste differentiation can change by modifying the interaction of tissues during development. Queens are larger than other castes in overall size, but smaller castes can still be larger than queens in some body dimensions. In particular, soldiers often possess heads or mandibles larger than those of both queens and workers (Fig. 1B; Wilson, 1953; Jaffé et al., 2007; Molet et al., 2014). In dung beetles, artificial selection for increased horn size also leads to a reduction in eye size, indicating that developing tissues may compete for resources (Emlen, 1996; Nijhout and Emlen, 1998). Queens invest large amounts of energy in traits such as wing and ovariole growth, so it is possible that individuals just below the size threshold for queen-like trait development could have more resources available for growth of other organs. This may occur in M. rogeri, where the largest intercastes (which lack wings) have longer legs than the smallest queens (which bear wings) (Londe et al., 2015). Such competition may also explain the overgrowth of the head or mandibles in soldiers relative to queens (Fig. 1); the total growth must be a positive function of size, but the growth of a single tissue could be greater in smaller individuals. Developing tissues have many potential interactions, particularly in the context of the dramatic morphological divergence between workers and queens. In addition to the caste-associated traits discussed here, any
morphological trait, or even traits such as behavior and lifespan, may exhibit size-associated variation and evolve according to our model.

Caste determination and differentiation often change together
In many cases, caste evolution occurs via changes to both caste determination and caste differentiation. In *Mystrium* red’, the reproductive caste has shifted from queens to small intermorphs (Molet et al., 2007). The loss of the queen caste is due to caste determination: the large individuals that develop into queens in the closely related *M. rogeri* are absent in *M. red’* (Fig. 3A; Molet et al., 2007). In addition, changes to caste differentiation allowed the gain of a novel caste: the size range of workers is restricted in *M. rogeri* (Fig. 3B; Molet et al., 2007). Note that, while these intermorphs form a reproductive caste smaller than the worker caste, they do not appear to be more queen-like than workers in terms of caste-associated morphology (i.e. ovariole number, eyes/ocelli and wings).

Similar changes are also observed in social parasites. Queens in social parasites have repeatedly evolved reduced size, producing the smallest ant queens (Nonacs and Tobin, 1992; Bourke and Franks, 1991; Aron et al., 1999). Social parasites often also have a worker caste that is reduced in size or absent. For example, *Plagiolepis xene* is a workerless social parasite closely related to the non-parasitic *Plagiolepis pygmaea*. The large individuals that become queens in *P. pygmaea* are absent in *P. xene*, and smaller individuals that would develop into workers in *P. pygmaea* instead develop into queens in *P. xene* (Aron et al., 1999). It has been proposed that the loss of the worker caste results from selection lowering the threshold size for queen development (Nonacs and Tobin, 1992; Aron et al., 1999; Linksvayer et al., 2013). However, it is also possible that the social context of social parasites could simultaneously impose selection for a reduction of queen size and loss of the worker caste.

This confounding factor is partially mitigated in *Myrmoxenus*, in which queens in most species obligately found colonies parasitically, but the worker caste is retained to conduct slave raids (Heinze et al., 2015). *Myrmoxenus* queens are reduced in size relative to queens in closely related non-parasitic species, demonstrating that parasitic colony foundation is associated with a reduction in queen size even when workers are present (Heinze and Foitzik, 2009). *Myrmoxenus alderzi* is a workerless species closely related to *Myrmoxenus ravouxi*, and it shows an additional reduction in queen size (phylogeny from Heinze et al., 2015; queen measures from Douwes et al., 1988). Thus, it appears that the social context of parasitic queens may lead to a reduction in body size, but a further reduction in size occurs when the worker caste is lost. This sequence of caste evolution occurs via a change to caste differentiation, i.e. the reduction of the size of queens to displace or eliminate the worker caste, and caste determination, i.e. the loss of the large individuals that developed into queens in the ancestral species. Caste evolution in this case can reduce the degree of queen–worker dimorphism (Fig. 3B), and the same processes may also provide a mechanism by which queen–worker dimorphism can be increased.

Conclusions
The developmental data reviewed in ‘A hierarchical continuum of phenotypes’, ‘Caste determination’ and ‘Caste differentiation’ closely fit with the evolutionary data reviewed in ‘Caste evolution’, implying that caste determination and differentiation position individuals within the phenotypic space described in Fig. 2, and caste evolution occurs via modifications to caste determination and caste differentiation (Fig. 3; Wilson, 1953; Oster and Wilson, 1978; Molet et al., 2012; Rajakumar et al., 2012; Favé et al., 2015). This theory constitutes an hourglass model, in which a large number of evolutionarily labile caste determination factors influence a conserved set of caste differentiation factors, which then produce diverse and evolutionarily labile caste morphologies (Fig. 4; Akhshabi and Dovrolis, 2011; Bopp et al., 2014; Wagner, 2014). This is reminiscent of sex development in insects, in which sex determination and sex morphology are highly variable, but sex differentiation results from conserved roles of the genes *doublesex and transformer* (Fig. 4; Akhshabi and Dovrolis, 2011; Bopp et al., 2014; Wagner, 2014). Caste development and sex development share many similarities, with variation in environmental (e.g. temperature and nutrition) or genetic (e.g. sex chromosomes and genetic caste determination) factors leading to development of alternative phenotypic states (Brian, 1979; Anderson et al., 2008; Schwander et al., 2010; Bopp et al., 2014; Klein et al., 2016). However, caste development is more frequently mediated by environmental factors, whereas sex development is more frequently mediated by genetic factors.

Polyphenisms have evolved in many types of insects, some of which appear to be size based (e.g. in hymenopterans, dung beetles and stag beetles), while others are independent of size (e.g. in aphids, butterflies, locusts and planthoppers) (Simpson et al., 2011; Hartfelder and Emlen, 2012). We propose that female polyphenisms in ants are size based, such that differences in morphology are always associated with differences in size, on average, and caste-associated traits undergo more queen-like development in larger individuals. Wheeler (1991) and Urbani (1998) proposed that ant castes do not necessarily differ in size. Others have claimed that microgyne such as those in *Aphaenogaster tennesseensis* and the *Formica microgyna* species group can fully overlap in size with the worker caste (Wilson, 1953). The evidence we present in Fig. 2 comes from species in three subfamilies that span the ant phylogeny,
and we have reviewed evidence from many additional taxa, so it is likely that the patterns we have described here are general and ancestral in ants (Fig. 2). We therefore expect that these microevolutions are in fact larger than workers in terms of pupal mass, even though the size differences may be very small. However, it is also possible that the association of caste and size has been secondarily modified in some species, or that the association of size and phenotype that we propose here is not a general pattern in ants, in which case our model would be falsified.

It is worth mentioning some additional consequences of a size-based polyphenism in ants. First, ovariole number varies with size ancestrally in insects, so once other queen-like traits (e.g. eye and wing development) became reduced in small individuals, ovariole number became reduced as well (Honěk, 1993). Therefore, the polyphenism in flight apparatus was necessarily associated with some degree of reproductive division of labor in the early stages of ant evolution, which may have facilitated the transition to eusociality. Second, as has been noted by other authors, size-based caste determination can be built into simple models that produce the elegant patterns of sociogenesis, sexual maturity and seasonal reproduction observed in many species (Tschoinkel, 2006). Let us assume that increased worker number causes colonies to produce larger larvae (possibly as a result of improved nutritional conditions of larvae and/or queens), and those larvae above a threshold size develop into queens. Small colonies therefore produce small larvae that exclusively develop into workers. These workers contribute to colony function by nursing, foraging, etc., which allows the colony to produce both more and larger workers, which then contribute to further colony growth. This positive feedback cycle is broken when larvae become large enough to develop into queens. At this point, worker and queen production reach an equilibrium; any deviation in the production of workers or queens causes changes in future worker or queen production that bring the colony back to the appropriate level (Tschoinkel, 2006). If larval size is also associated with additional factors like queen age or season, this simple model can produce many of the complex life cycles observed throughout the ants. Intriguingly, if larval size were not correlated with worker number, this could cause colonies to collapse by continuing to produce new queens even if worker numbers decline. Such a hypothetical scenario might be quite similar to what occurs in socially parasitic lineages, in which some individuals become genetically determined to develop into queens at a lower body size threshold (Nonacs and Tobin, 1992; Aron et al., 1999; Linksvayer et al., 2013).

Finally, we would like to discuss the concept of allometry. Wilson (1953), Wheeler (1991) and others have argued that castes in ants can be defined as sets of individuals that differ in their allometric coefficients (i.e. changing slope in the regression of two trait measures). The association of body size and caste-associated phenotypes is non-linear, so it follows that individuals differing in size will often differ in allometry as well, even if these different groups do not have unique functional roles in the colony (Fig. 2; Londe et al., 2015). Conversely, individuals in other regions of the phenotypic space may differ greatly in size and their functional roles in the colony, but not in allometry (Fig. 2; Londe et al., 2015). Ants possess a phenotypic space ranging from worker-like to queen-like, and show some flexibility in the regions of this space that are occupied in any given species. In some cases it may be useful to define these regions as separate castes, while in other cases it may be more useful to consider the occupied space as a continuum.

The theory presented here provides a framework for interpreting data about caste development and evolution in ants and potentially other social insects, and we hope that it will facilitate a unified research program moving forward. In particular, we hope to highlight the critical importance of including measurements of size in any study of caste development and evolution, which is necessary to decide whether a given factor affects caste determination or differentiation. Central attributes of caste determination and differentiation remain poorly understood. We know a host of caste determination factors that affect size, but the combination of larval and worker traits that give rise to discontinuous size-frequency distributions in the vast majority of ant species remain mysterious. Caste differentiation is even less well explored. It is clear that variation in caste-associated traits is associated with size and results from variation in imaginal disk development, but how this variation arises and how it evolves remains unclear. These two issues are critical features of ant biology, and further inquiry will yield far-reaching insights in the years to come.

Acknowledgements
We thank the members of the Laboratory of Social Evolution and Behavior and many of our colleagues for helpful comments on this theory—you are too many to name. We also thank The Company of Biologists for organizing the special issue Evolution of Social Behavior.

Competing interests
The authors declare no competing or financial interests.

Author contributions
Conceptualization, W.T.; Investigation, W.T.; Writing, W.T. and D.J.C.K.; Supervision, D.J.C.K.

Funding
This work was supported by the David Rockefeller Graduate Program at The Rockefeller University.

References


