Gene expression is more strongly influenced by age than caste in the ant *Lasius niger*

Eric R. Lucas¹,2 | Jonathan Romiguier¹ | Laurent Keller¹

¹Department of Ecology and Evolution, Biophore, University of Lausanne, Lausanne, Switzerland
²Department of Vector Biology, Liverpool School of Tropical Medicine, Liverpool, UK

Correspondence
Eric Lucas,
Department of Vector Biology, Liverpool School of Tropical Medicine, Liverpool, UK.
Email: eric.lucas@lstmed.ac.uk
and
Laurent Keller,
Department of Ecology and Evolution, Biophore, University of Lausanne, Lausanne, Switzerland.
Email: laurent.keller@unil.ch

Funding information
Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung, Grant/Award Number: 310030_156732; FP7 People: Marie-Curie Actions, Grant/Award Number: 297898; FP7 Ideas: European Research Council, Grant/Award Number: 249375

Abstract
Phenotypic plasticity, where a single genome can give rise to different phenotypes, underlies many remarkable features of the natural world and occurs in a wide range of organisms. Understanding the transcriptional differences that underlie plastic phenotypes remains a major unsolved problem in biology. In many ants, females can develop into either queens or workers, two phenotypes with different morphology, behaviour and longevity. In comparison with workers, queens are larger, more fecund and longer lived. Here, we study gene expression differences between queens and workers in the ant *Lasius niger*. The analysis of age- and tissue-specific RNA sequencing showed that patterns of caste-biased gene expression vary considerably between ages and tissues. Expression was more tightly linked to age than caste despite the important morphological and behavioural differences between queens and workers. Our data allowed us to identify genes that are consistently biased across biological contexts. Caste-biased genes showed faster rates of molecular evolution, lower levels of DNA methylation and greater variability in expression than unbiased genes. Our results indicate that a substantial proportion of caste-biased expression is ephemeral and that taking account of age and tissue is critical to understanding the transcriptomic basis of plastic phenotypes. By contrast, the biological context of expression bias did not broadly affect methylation or the rate of evolution. The faster rate of evolution and greater variability of expression of caste-biased genes indicate that caste-biased genes evolve from loosely regulated genes that can be co-opted for caste-specific tasks because of the lax control over their expression.

KEYWORDS
molecular evolution, morph-biased expression, phenotypic plasticity, social insects

1 | INTRODUCTION

Phenotypic plasticity, where a single genome can give rise to markedly different phenotypes, underlies many remarkable features of the natural world such as context-dependent life history tactics and cooperative division of labour (Brockmann, 2001; Taborsky, Oliveira, & Brockmann, 2008; Wilson, 1971) and can allow rapid adaptation to varying environmental conditions such as temperature or altitude (Buckley, Owen, & Hofmann, 2001; Cheviron, Whitehead, & Brumfield, 2008). Plasticity is found in a wide range of species, from bacteria (Justice, Hunstad, Cegelski, & Hultgren, 2008; Kummerli, Jiricny, Clarke, West, & Griffin, 2009) to humans (Kuzawa & Bragg, 2012), and is of relevance to important topics such as adaptation to changing climatic conditions and the expansion of invasive species (Davidson, Jennions, & Nicotra, 2011; Nicotra et al., 2010). As by definition differences between plastic phenotypes can arise from a single genome, the variation in morphology and behaviour is due to differences in gene expression (Aubin-Horth & Renn, 2009). Understanding...
the transcriptional differences that underlie the striking adaptations within plastic phenotypes remains a major unsolved problem of biology, and one that is only now becoming tractable through recent advances in high-throughput whole-transcriptome sequencing (Wang, Gerstein, & Snyder, 2009).

Some of the most striking examples of phenotypic plasticity are found in social insects (ants, termites and many bees and wasps), which are characterized by their division of labour among specialized plastic phenotypic castes (Wilson, 1971). In particular, the reproductive queens and sterile workers of advanced social insects display substantial differences in morphology and behaviour despite being genetically undifferentiated in most species (Schwander, Lo, Beekman, Oldroyd, Keller, 2010; Wilson, 1971), making them excellent candidates for transcriptional studies of phenotypic plasticity. As a result, much attention has been directed towards describing the gene expression differences between queens and workers in a wide range of social insect species (Bonasio et al., 2010; Feldmeyer, Elsner, & Foitzik, 2014; Ferreira et al., 2013; Grozinger, Fan, Hoover, & Winston, 2007; Harrison, Hammond, & Mallon, 2015; Hoffmann & Goodisman, 2007; Judice et al., 2006; Ometto, Shoemaker, Ross, & Keller, 2011; Patalano et al., 2015; Well, Korb, & Rehli, 2009). Interestingly, these studies have so far revealed few consistent patterns of differential gene expression (Berens, Hunt, & Toth, 2015; Hunt & Goodisman, 2010; Morandin et al., 2016; Toth et al., 2010). A possible explanation is that different sets of genes are involved in regulating caste differences across species. Alternatively, low consistency among studies may stem from differences in methodologies and confounding factors. For example, there is substantial variation in gene expression among tissues. This has been demonstrated in mammals, where gene expression is more closely related to tissue than species (Brawand et al., 2011; Necsulea & Kaessmann, 2014), and by a recent study of regulatory genes in honey bees, where only one gene out of 495 showed consistent caste-biased expression across tissues (Johnson & Jasper, 2016). Studies that measure expression in whole bodies (e.g., Feldmeyer et al., 2014; Harrison et al., 2015; Hunt & Goodisman, 2010; Hunt, Ometto, Keller, & Goodisman, 2012; Hunt et al., 2011; Judice et al., 2006; Morandin et al., 2016; Ometto et al., 2011; Well et al., 2009) will therefore be blind to localized variation and run the risk of overgeneralizing their conclusions. Furthermore, allometric differences between morphological castes will confound whole-body comparisons as different-sized tissues will contribute to varying degrees to the total pool of sequenced transcripts. Another potentially important factor is the age of the individuals analysed. This is a particularly important problem in social insects as queens can be extremely long-lived (Keller & Genoud, 1997) and because there is a great difference in lifespan between queens and workers (Kramer & Schable, 2013). Previous studies comparing queens and workers have typically used individuals that were not controlled for age (Bonasio et al., 2010; Feldmeyer et al., 2014; Harrison et al., 2015; Hunt & Goodisman, 2010; Morandin et al., 2016; Patalano et al., 2015; Toth et al., 2010; Well et al., 2009), thus introducing a confounding factor possibly affecting the level of gene expression. Adult age can have a strong effect on gene expression in insects (Landis et al., 2004; Nipitwattanaphon, Wang, Dijkstra, & Keller, 2013; Wang et al., 2010), and developmental stage has been shown to have a stronger effect than caste on gene expression in social insects (Harrison et al., 2015; Hoffmann & Goodisman, 2007; Ometto et al., 2011; Smith et al., 2015), yet the relative importance of age and reproductive caste has yet to be investigated.

The ant Lasius niger has highly morphologically differentiated queens and workers, presenting an opportunity to investigate patterns of differential expression of genes that underlie queen and worker castes. Lasius niger follows a typical claustral colony cycle, whereby new queens store large amounts of nutrients in their native colony before engaging in a single mating flight and founding an independent colony, rearing the first batch of workers alone (Hölldobler & Wilson, 1990; Keller & Passera, 1989), a process which takes a little under 2 months.

The goal of this study was to explore patterns of gene expression that underlie queen and worker castes of the ant L. niger and to determine the extent to which these differences can be confounded by age and caste. We used high-throughput RNA sequencing to compare gene expression between age-matched queens and workers in two different tissues (brains and legs) and at three different ages (1-day-old, 1-week-old and 2-month-old). The brain is a critical organ controlling behaviour and thus is likely to play an important role in the behavioural differences between queens and workers. Legs represent a tissue whose role is as similar as possible between queens and workers in order to identify differences in expression inherent to the castes rather than linked to a particular tissue-specific task. We do not claim that legs are identical in queens and workers, but they are likely to be more similar than other organs such as the thorax (which contain the queen flight muscles) or abdomen (which contain the ovaries). One-day-old and 2-month-old individuals represent contrasting life stages for both queens and workers because both castes are callows at 1 day, while workers are active in the nest and queens have founded an incipient colony by 2 months. Consistent expression differences between these two age classes would therefore provide promising candidates for consistently caste-biased signals. Because 1-day-old individuals represent a unique life-stage due to the recent transition from the pupal state, comparing the relative importance of age and caste on gene expression using only these two age groups may not provide a reliable comparison. We therefore also obtained expression data from 1-week-old individuals to more thoroughly address this question.

As well as searching for consistent patterns of caste-biased gene expression, we use our differential expression to explore the mechanistic and evolutionary context of phenotype-biased gene expression in different biological contexts. DNA methylation is a process that has received substantial attention for its potential mechanistic role in underlying gene expression variation between phenotypes. It is a process by which cytosines, typically in CpG dinucleotides, are modified by the addition of a methyl group. The role of DNA methylation is still poorly understood, but it is known to influence gene expression by modulating transcription factor binding and alternative splicing (Bonasio et al., 2012; Flores et al., 2012; Libbrecht, Oxley, Keller,
While patterns of DNA methylation across the genome vary greatly between taxa (Zemach, McDaniel, Silva, & Zipperman, 2010), methylation of coding regions is broadly conserved, and is the dominant form of methylation found in insects (Bonasio et al., 2012; Elango, Hunt, Goodisman, & Soodin, 2009; Glastad, Gokhale, Liebig, & Goodisman, 2016; Libbrecht et al., 2016; Lyko et al., 2010; Walsh et al., 2010; Wang et al., 2013; Xiang et al., 2010). Because of its potential for plasticly modifying the expression of genes, DNA methylation is of considerable interest in the field of phenotypic plasticity, and has been the focus of several studies in social insects (Bonasio et al., 2012; Elango et al., 2009; Flores et al., 2012; Foret et al., 2009; Glastad et al., 2016; Libbrecht et al., 2016; Lyko et al., 2010; Simola et al., 2013; Smith et al., 2011). Previous studies revealed that genes whose expression is variable between phenotypes, tissues or individuals tend to have low overall levels of methylation compared to genes with consistent expression (Elango et al., 2009; Foret et al., 2009; Glastad et al., 2016; Xiang et al., 2010). Whether this pattern depends on the age or tissue in which expression is measured has yet to be explored.

We investigated a proxy for historical germline methylation by calculating the frequency of CpG dinucleotides in our transcriptomic data. Because methylated cytosines have a tendency to mutate to thymine, DNA regions that are methylated in the germline tend to become depleted in CpG dinucleotides over evolutionary time (Bird, 1980). This relative depletion in CpG dinucleotides can thus be used to detect signals of historical germline methylation in regions of the genome (Elango et al., 2009; Gavery & Roberts, 2010; Simola et al., 2013; Smith et al., 2011; Walsh et al., 2010), and studies have found a good correspondence between CpG depletion and sites of active methylation (Foret et al., 2009; Simola et al., 2013).

From an evolutionary perspective, understanding phenotype-biased gene expression requires identifying the types of genes that tend to become caste-biased and understanding the sequence evolution of caste-biased genes under the selective pressures that accompany biased expression (Helenranta & Uller, 2014). Previous studies of social insect castes have mostly found that caste-biased genes have higher rates of sequence evolution than unbiased genes (Harpur et al., 2014; Hunt et al., 2010, 2012; Roux et al., 2014). Whether this is because loosely regulated genes tend to become caste-biased or because caste-biased expression results in relaxed selection remains an open question. We combined our transcriptomic data with other ant transcriptomes to estimate and compare the rate of sequence evolution between caste-biased and unbiased genes, and explore the importance of the age/tissue context of that caste-biased expression on these results.

2 | METHODS

2.1 | Sample collection

Queens and workers of Lasius niger were collected in the field and reared in controlled-climate laboratory conditions as previously described (Lucas, Privman, & Keller, 2016). Briefly, queenless colonies were established by collecting brood and adult workers from field colonies and bringing them into the laboratory, where the workers raised the brood to adulthood. Queenright colonies were established by collecting queens during their mating flight and allowing them to establish colonies in the laboratory. The 1-day-old and 2-month-old samples presented here are the same as those used in a previous study (Lucas et al., 2016). Details regarding sample collection and dissection can be found in Methods S1.

2.2 | RNA extraction and sequencing

RNA was extracted from legs and brains using TRIzol with a DNA digestion step as previously described (Lucas et al., 2016) and sequenced by Illumina HiSeq 2000/2500. Further details concerning the library preparation and sequencing is provided in the Methods S1 and Data S1. The raw reads obtained from the sequencing have been deposited in the NCBI Short Read Archive (Accession nos.: SRP069113, SRP097456).

2.3 | Differential gene expression

Reads were aligned to the L. niger transcriptome assembied in (Lucas et al., 2016) using bowtie2 (Langmead & Salzberg, 2012) with default parameter values. Upon submission of the transcriptome to the NCBI Transcriptome Shotgun Assembly archive, 57 components (out of 63,718) were flagged as being a possible result of contamination. These components were therefore removed from the analysis. Counts of aligned reads were analysed using the R package edgeR (Robinson, McCarthy, & Smyth, 2010) with generalized linear modelling as previously described (Lucas et al., 2016). p-Values were adjusted using the package fdrtool to control the False Discovery Rate (FDR) at 5%. Multidimensional scaling (MDS) was performed using plotMDS in R with default parameters. An initial MDS on data from all lanes analysed separately revealed a minimal effect of lane (Fig. S1), and thus data from all lanes were pooled and analysed together. When analysing only 1-day-old and 2-month-old individuals, handling batch (samples that were handled together in the laboratory) was included as a categorical effect. As 1-week-old individuals were handled separately, batch was not included as an effect when all age groups were analysed together. Similarly, as brains and legs were handled separately, batch was not included as an effect when the two tissues were analysed together.

Sequences of interest were annotated using a blastx search (Altschul et al., 1997) against the Drosophila melanogaster nonredundant protein (nr) database (critical e-value: 0.0001). If no hits were produced, the process was repeated against the entire nr database.

2.4 | Gene ontology term enrichment analysis

Gene ontology (GO) analysis was performed using existing Biological Process GO annotations for the L. niger transcriptome (Lucas et al., 2016). All genes in the transcriptome were ranked according to their
2.5 Cluster analysis

Within each of the legs and brains data sets, we kept the top 5,000 genes with the highest variance across all samples after normalization by edgeR and used the log normalized expression values to perform cluster analyses. An initial clustering of sample combining data from brains and legs revealed a deep split between the two tissues (Fig. S9). We caution that the substantial clustering by tissue may in part be a result of the different handling complexities of legs and brains during dissection, which could have introduced confounding variation between tissues. We therefore proceeded to analyse brains and legs separately.

Samples were grouped according to similarities in their global gene expression patterns using hierarchical clustering, and genes were grouped according to similarities in their expression patterns among samples using Weighted Gene Co-expression Network Analysis (WGCNA). Both methods were implemented using the R package WGCNA (Langfelder & Horvath, 2008). WGCNA was used to divide genes into modules that showed common expression patterns, using the hybrid algorithm from the WGCNA package and a minimum module size of 30 (see Methods S1 for details). For each of the gene modules we obtained, we used a simple ANOVA to determine whether the representative eigengene of that module varied between the different age/caste combinations. For modules that showed significant variation, we used a general linear model to determine the effects of caste, age and their interaction on that module’s expression. The module’s eigengene was used as the dependent variable, while caste and age were coded as independent categorical variables.

Genes within a cluster of interest were tested for GO term enrichment using a Fisher test implemented in topGO and FDR corrections were performed with a cut-off of 0.01 using fdrtool (Klaus & Strimmer, 2015).

2.6 CpG analysis

To investigate historical germline methylation in each isogroup of the Lasius niger transcriptome, we calculated the ratio of the observed and expected number of CpG dinucleotides (CpG_{O/E}). Because methylated cytosines are highly mutable, methylated regions of the genome in the germline become depleted in CpGs over evolutionary time, and thus CpG paucity can be interpreted as a signal of historical methylation (Bird, 1980). This method has been used to estimate methylation levels in previous studies of social insect genomics (Elango et al., 2009; Simola et al., 2013; Smith et al., 2011). We calculated CpG_{O/E} as follows:

\[
\text{CpG}_{O/E} = \frac{\text{CpG}}{P_c(P_c + P_g)}
\]

where CpG is the number of observed CG dinucleotides in the sequence, \(P_c\) and \(P_g\) are the respective proportions of C and G nucleotides in the sequences and \(l\) is the length of the sequence. The expected number of CpG dinucleotides (the denominator) is therefore calculated based on the proportions of C and G nucleotides in the sequence. Unequal distribution of GC richness across the sequence will falsify this expected value. Therefore, as a control, we also calculated GpC_{O/E}, where the numerator is the observed number of GpC instead of CpG dinucleotides (Elango et al., 2009). As GpC is not considered to be a common target for methylation, GpC_{O/E} gives a measure of the neutral CpG_{O/E} expectation of in the absence of methylation.

CpG_{O/E} was calculated for four different types of sequences: “non-genic” sequences, where no open reading frame (ORF) was detected, coding DNA sequence (CDS) sequences, which constitute the coding sequences for detected ORFs, “flanking” sequences being on either side of a CDS and “intronic” sequences being located between CDS exons. For a given ORF, there may be more than one segment of flanking or intronic sequences. In these cases, all segments from a single ORF were concatenated into a single sequence before analysis.

Sequences shorter than 201 nucleotides after concatenation, or belonging to an isogroup in which several ORFs were detected in at least one transcript (fusion transcripts), were removed from the analysis. Where several isoforms existed in an isogroup, the mean CpG_{O/E} of the isoforms was calculated.

To determine whether CpG_{O/E} was associated with caste-biased expression, we used general linear models within each age/tissue combination. CpG_{O/E} was the response variable, direction of bias (queen-biased, worker-biased or unbiased) was a categorical explanatory variable and the log expression level was a continuous explanatory variable. Caste-biased genes were defined according to a threshold \(p\)-value of .05, with the direction of bias given by the sign of the coefficient estimated by EdgeR. Unbiased genes were defined as those with a \(p\)-value above the .05 threshold. Expression level was calculated as the mean Reads Per Kilobase Million across all conditions within a tissue, using normalized library sizes calculated by edgeR. The effect of removing direction of bias from the model was determined using the anova function in R. As a control, all analyses were repeated using GpC_{O/E} instead of CpG_{O/E}.

2.7 Rate of evolution

To investigate the rates of evolution of L. niger coding sequences, we compared our transcriptome assembly to those of Lasius neglectus and Lasius niger (assembly available from Morandin et al., 2016) while using Formica cunicularia as an out-group (assembly available from J. Romiguier, J. Rolland, J. Morandin, L. Keller, in preparation). Coding-sequence orthology was defined using OrthoMCL (Chen, Mackey, Stoeckert, & Roos, 2006). Sequences were
aligned using MACSE (Runwex, Harispe, Delsuc, & Douzery, 2011). Alignments were then carefully cleaned following the cleaning procedure of the Selectome database (Moretti et al., 2014), see details in http://selectome.unil.ch/cgi-bin/methods.cgi that uses a combination of MAHALIGN (Gouveia-Oliveira, Sackett, & Pedersen, 2007), PAGAN (Loptynoja, Vilella, & Goldman, 2012), MCCoffee (Wallace, O’Sullivan, Higgins, & Notredame, 2006), GUIDANCE (Penn, Privman, Landan, Grauf, & Pupko, 2010) and TRIMAL (Capella-Gutierrez, Silla-Martinez, & Gabaldon, 2009). We only considered the 1-1 ortholog gene groups across the four species (n = 2,736) to calculate dN (nonsynonymous rates of evolution), dS (synonymous rates of evolution) and dN/dS ratios with codeml in the PAML package (Yang, 2007). We used a pairwise approach (runmode = -2) and computed dN/dS ratios between L. niger and both L. neglectus and L. turcicus. Both comparisons provided similar results and we present comparisons with L. turcicus in the result section (results with L. neglectus available in Data S2).

Points with extreme dN or dS values (dN = 1, dS = 1, dN/dS > 5) or with too few substitutions in a sequence (N < 100, S < 100) were considered unreliable and removed from the analysis. For the analysis of dN/dS, we also excluded points where dS = 0, as these make the calculation of dN/dS meaningless. dN/dS, dN and dS were compared between unbiased and caste-biased or age-biased genes using Wilcoxon signed rank tests. Caste bias was determined in the same way between unbiased and caste-biased or age-biased genes using Wilcoxon signed rank tests.

Using the same set of alignments, we also performed an analysis of positive selection detection using the branch-site tests implemented in PAML/codeml (Yang, 2007) and following the same procedure used for the Selectome database (Moretti et al., 2014, see details in http://selectome.unil.ch/cgi-bin/methods.cgi).

2.8 Expression variability

Variability of expression within groups was measured as the dispersion of the negative binomial distribution model applied to the read counts. The dispersion is a measure of the proportion of the variance in the data that can be attributed to variation between biological replicates as opposed to random error in the sampling process (Robinson et al., 2010). Dispersion was calculated in edgeR using tagwise estimates.

3 RESULTS

3.1 Effects of age and caste on gene expression

Multidimensional scaling on the combined data of all ages, castes and tissues was able to separate samples according to age and tissue, but not caste (Figure 1; the same data colour-coded by only age or caste are presented in Fig. S2). The first MDS dimension separated samples according to tissue, while the second dimension separated 1-day-old individuals from older individuals. The third and fourth dimensions separated 1-week-old and 2-month-old individuals. Even up to the fifth and sixth dimension, castes were not separated by the MDS. Only when fewer sample types were included did the MDS successfully separate caste. For example, when tissues were analysed separately and only 1-day-old and 2-month-old individuals were included, caste was separated on the second MDS dimension in legs, and to some extent in 1-day-old brains (Fig. S3).

Expression was highly correlated between queens and workers of the same age in both brains (Fig. S4; Table S1) and legs (Fig. S5; Table S1). Expression was also highly correlated between individuals of different age and the same caste in both brains (Fig. S4; Table S1) and legs (Fig. S5; Table S1). The values of r2 were higher between queens and workers than between samples of the same caste but different age (Figs S4 and S5), indicating that age has a stronger influence on expression than caste. To confirm this, we calculated residuals around the mean expression value for each gene. These residuals were positively correlated between workers and queens of the same age, but negatively correlated between individuals of different age but the same caste (Figs S6 and S7, Table S1), confirming that expression is more consistent within age groups than within castes.

In brains, variability in gene expression within individuals was significantly lower in 2-month-old individuals than 1-day-old individuals.

![Figure 1](https://example.com/figure1.png) Multidimensional scaling (MDS) of normalized expression data, performed using the MDSplot function in edgeR with default parameters.
(\(W = 2,411,217,924, n_1 = 60,651, n_2 = 60,651, p < .0001\), Fig. S8).

In legs, variability in expression was also significantly lower in 2-month-old than 1-day-old individuals, although the magnitude of the difference was very small (\(W = 1,358,236,562, n_1 = 52,347, n_2 = 52,347, p = .015\), Fig. S8).

### 3.2 Caste-biased genes

Of 63,661 isogroups (henceforth called genes for simplicity), 178 genes and showed a significant interaction between age and caste in brains. Of the genes that did not show a significant interaction, 2,007 were significantly differentially expressed between queens and workers, and 14,027 were significantly differentially expressed between 1-day-old and 2-month-old individuals. In legs, 5,051 genes showed a significant interaction between age and caste. Of the genes that did not show a significant interaction, 7,123 were significantly differentially expressed between queens and workers, and 25,352 were significantly differentially expressed between 1-day-old and 2-month-old individuals. A possible explanation for the lower number of significantly caste-biased genes in brains than in legs is that there was a greater dispersion of expression values in brains than in legs (\(W = 8.4 \times 10^8, n_1 = 52,347, n_2 = 60,651, p < .0001\)). This higher dispersion could represent higher intrinsic variation in expression between individuals in this tissue, or may be an artefact of the dissection procedure, which was more complex for brains than for legs and therefore could have increased the variance between replicates.

Considering each age/tissue context separately, 5,792 (9.1%) genes showed caste-biased expression in the legs of 1-day-old individuals, 10,400 (16.3%) in the legs of 2-month-old individuals, 1,384 (2.2%) in the brains of 1-day-old individuals and 486 (0.763%) in the brains of 2-month-old individuals. Thirteen genes were consistently queen-biased and eight genes were consistently worker-biased in all four age/tissue contexts (Figure 2, Data S3). Annotations for these consistently biased genes are listed in Data S3. The genes that were consistently queen-biased include an Elongation Factor 1-alpha, which has been linked with extended longevity in Drosophila (Shepherd, Waldorf, Hug, & Gehring, 1989), and the immunity-related gene Serpin 27A.

### 3.3 Cluster analysis

We used hierarchical clustering to sort individuals into clusters according to their expression profiles. Individuals of the same age clustered together, with caste-based clusters appearing only within the age-based cluster (Figure 3), confirming our finding from the global expression analysis that age has a stronger effect than caste on gene expression.

Grouping genes into modules (clusters) with similar expression patterns provided further evidence that age affects patterns of gene expression more strongly than caste. In brains, 38 gene modules were obtained (Fig. S10). These modules were dominated by the effect of age, with only a minor influence of caste. Fourteen modules showed an effect of age with no effect of caste, 24 were significantly affected by both age and caste or showed an interaction between age and caste, but no module showed an effect of only caste with no effect of age (Table S2). There were two modules (Module 1 and Module 21) that showed a stronger significant effect of caste than age with no interaction. Module 1 was not enriched for any GO terms, but Module 21 was significantly enriched for five GO terms, all of which were related to metabolism (significant GO terms: lipid metabolic process, cellular lipid metabolic process, hormone metabolic process, oxidation-reduction process, isoprenoid metabolic process).

In legs, 32 gene modules were obtained (Fig. S11). These modules were again dominated by the effect of age, with only a minor influence of caste. Ten modules showed a significant effect of age but not caste, 22 were significantly affected by both age and caste or showed an interaction between age and caste, but no module showed an effect of only caste with no effect of age (Table S2). One module (Module 1) showed a stronger significant effect of age than caste with no interaction between them, but this module was not enriched for any GO terms.

### 3.4 Gene ontology terms enriched for genes with caste-biased expression

The lists of all significantly enriched GO terms can be found in Data S4. The effect of caste on gene expression differed markedly between ages and tissues for a range of biological processes. In
several cases, the same GO term was queen-biased in one context, but worker-biased in another context. For example, GO terms linked to cellular respiration were enriched for queen-biased genes in the legs of 1-day-old individuals, but enriched for worker-biased genes in the legs and brains of 2-month-old individuals. Similarly, GO terms linked to organ morphogenesis and development were enriched for queen-biased genes in 1-day-old brains, but enriched for worker-biased genes in 1-day-old legs.

No GO term was significantly biased in the same direction in all four age/tissue combinations, but some GO terms were consistently biased in the same direction within either a tissue or age (Data S4). For example, terms linked to organic acid metabolism and biosynthesis were enriched for queen-biased genes in the legs of both 1-day-old and 2-month-old individuals, while GO terms linked to cell–cell signalling were enriched for worker-biased genes in the brains of 2-month-old individuals, as well as the legs of both 1-day-old and 2-month-old individuals.

3.5 | CpG analysis

Coding regions contained the strongest signal of historical germline methylation, having the lowest mean value of CpG_{O/E} both in absolute terms and relative to GpC_{O/E}. CpG_{O/E} was highest in nongenic sequences, followed by flanking, then intronic and finally coding sequences (nongenic vs. flanking: \( t = 2.84, df = 13422, p = .004 \); flanking vs intronic: \( t = 13.13, df = 3340.7, p < .0001 \); intronic coding: \( t = 19.72, df = 2791.1, p < .0001 \); Figure 4). We controlled for factors that could confound the value of CpG_{O/E}, such as uneven distribution of GC content across a sequence, by comparing CpG_{O/E} with GpC_{O/E} (Elango et al., 2009). In nongenic, flanking and intronic sequences, values of CpG_{O/E} were significantly greater than GpC_{O/E} (nongenic: mean difference = 0.25, \( t = 99.5, df = 48246, p < .0001 \); flanking: mean difference = 0.22, \( t = 42.6, df = 8632, p < .0001 \); intronic: mean difference = 0.08, \( t = 6.1, df = 2488, p < .0001 \); Figure 4). In the coding sequences, while the difference between CpG_{O/E} and GpC_{O/E} was significant due to the large sample size, the magnitude of this difference was very small (coding: mean difference = 0.02, \( t = 4.8, df = 12255, p < .0001 \); Figure 4), further supporting the conclusion that CpG dinucleotides are rarest in coding sequences.

Genes that were either worker- or queen-biased had higher CpG_{O/E} (but not GpC_{O/E}) than unbiased genes (Figures 5 and 6, Data S5), suggesting that historical germline CpG methylation may primarily occur in unbiased genes. These differences were mostly highly significant, with the exception of genes that were queen-biased in 1-day-old legs, suggesting that the association between caste bias and methylation may be weaker in this group of genes. Interestingly, there was one isolated case in which the direction of bias was reversed. In nongenic sequences whose expression was queen-biased in 2-month-old brains, CpG_{O/E} was significantly lower than in unbiased genes.
As with caste-biased genes, age-biased genes had higher CpG
(Obs/Exp) than unbiased genes (Figs S12 and S13, Data S5),
again supporting the idea that historic germline CpG methylation pri-
arily occurs in genes with stable, consistent expression. The exccep-
tion was in the intronic regions of genes biased towards 2-month-
old individuals, where there was no significant difference in CpG
(Obs/Exp) with unbiased genes.

3.6 | Rates of molecular evolution

3.6.1 | Caste-biased genes

dN/dS was higher in caste-biased genes than unbiased genes (signifi-
cant in four of eight comparisons, with nonsignificant trends in the
same direction in the other four comparisons), but there was no sig-
nificant difference in dN/dS between queen-biased and worker-
biased genes (Data S6, Figure 7). When considering dN and dS sepa-
 separately, dN gave the same results as dN/dS analyses, being higher in
caste-biased genes than unbiased genes, but there were no significant
differences in dS between caste-biased and unbiased genes (Data S6,
Figure 8).

3.6.2 | Age-biased genes

Genes biased towards 1-day-old individuals in brains had higher dN/
dS than unbiased genes (workers: \(W = 205.556, p = .003\); queens:
\(W = 228.474, p = .009\)), but this was not the case for genes biased
towards 2-month-old individuals, or for any caste-biased genes in
legs (Data S6). This indicates that age-biased genes tend to have
higher rates of molecular evolution than average, although this is less
pronounced than in caste-biased genes. When considering dN and
dS separately, dN mirrored the results obtained for dN/dS, but there
were no significant differences in dS between caste-biased and unbi-
ased genes (Data S6).

3.6.3 | Signature of positive selection

A branch-site test revealed nine genes with a significant signal of
positive selection on the Lasius niger branch of the phylogeny (Data
S8). Annotation of these sequences using blastx revealed that one of
these genes is a Cytochrome P450 (CYP) reductase. Expansion of
CYPs and sequence evolution in this family has been suggested to
be linked with the success of L. niger in urban environments
(Konorov et al., 2017).

4 | DISCUSSION

4.1 | Effects of age and caste on gene expression

Much attention has focused on trying to identify genes or genetic
toolkits that are consistently involved in caste differences across
social insects, yet studies in different species have so far struggled
to identify consistent patterns (Berens et al., 2015; Hunt & Goodis-
man, 2010; Morandin et al., 2016; Toth et al., 2010). Our results
demonstrate that there is only limited consistency within adults of a
single species in the biological processes that are linked to caste dif-
fences. At the gene level, caste-biased patterns were more consis-
tent with age in brains than in legs. In brains, just over 2,000 genes
showed a direct global effect of caste, while only 178 genes showed
a significant interaction between age and caste. In legs, over 7,000
genes showed a direct global effect of caste, with just over a further
5,000 showing a significant interaction. One possible explanation for
this difference between brains and legs is that transcriptional differ-
ences between castes are more stable with age in brains than legs.
Another possible explanation is that the higher variance and lower
sample size (five replicates in brains compared to six in legs) reduced
the power to detect interactions between age and caste in brains.
Thus, the proportion of genes with no age/caste interaction in brains
may have been underestimated.

We found that age has a stronger effect than caste on patterns of
gene expression. Expression within an age group was more tightly
correlated than within a caste, more genes were differentially
expressed with age than with caste, and samples clustered together
by age rather than caste according to both MDS (Figure 1) and hier-
archical clustering (Figure 3) analyses. The variation that we identi-
fied with age and tissue may help explain the lack of consistent
patterns among studies of caste-biased gene expression. Uncon-
trolled variation in age and allometry between castes within a study
could lead to incorrect identification of caste-biased genes. Similarly,
variation in age and tissue between studies could cause different

![Relative abundance of CpG and GpC dinucleotides](image-url)

FIGURE 4 Observed/Expected CpG and GpC (CpGO/E and
GpCO/E, respectively) in nongenic, coding, flanking and intronic
sequences [Colour figure can be viewed at wileyonlinelibrary.com]
snapshots of expression bias to be captured, leading to the appearance of inconsistencies. A more detailed approach that recognizes the dynamic nature of expression biases is needed if truly general conclusions are to be made on the gene expression patterns that underlie caste differences in social organisms.

Our results are consistent with other studies in social insects that have shown that developmental stage has a stronger effect than caste on gene expression (Harrison et al., 2015; Hoffman & Goodman, 2007; Ometto et al., 2011; Smith et al., 2015). Also, in the fire ant Solenopsis invicta, age was found to have a stronger effect than queen phenotype in a dimorphic system where queens can either be large and head a colony singly or smaller and part of a multiqueen colony (Nipitwattanaphon et al., 2013). By contrast, behavioural caste has been found to have a stronger effect than age on gene expression in honeybees (Whitfield, Cziko, & Robinson, 2003). This interesting result may be due to the fact that caste is typically linked with maturation in honeybee workers. To obtain age-controlled castes, the honeybee workers needed to be manipulated into accelerating or slowing down their caste transition, which may have also affected their rate of physiological maturation.

4.2 Caste-biased genes

Of the genes that were consistently differentially expressed across all four age/tissue contexts, two stand out as particularly interesting. One gene was annotated as Elongation Factor 1alpha48D

FIGURE 5 Barplot showing CpGO/E and GpCO/E as a function of caste-biased expression in brains
Elongation Factor 1α (Ef1α) is involved in protein synthesis, and increased expression of this gene in *Drosophila* is associated with extended lifespan, possibly by compensating the decline in protein synthesis that accompanies ageing (Shepherd et al., 1989). Like many advanced social insects, queens of *Lasius niger* live substantially longer than either workers or solitary insects (Holldobler & Wilson, 1990; Keller & Genoud, 1997). Our results raise the possibility that increased expression of Ef1α could be implicated in the extended longevity of queens.

The serine protease inhibitor Serpin 27A (Spn27A) was also consistently more highly expressed in queens than workers. Spn27A negatively regulates the encapsulation immune response by inhibiting the conversion of pro-phenoloxidase (PPO) to phenoloxidase (PO) (De Gregorio et al., 2002; Nappi, Frey, & Carton, 2005). One possible conclusion is that increased expression of Spn27A in queens is indicative of a reduced immune response compared to workers, perhaps reflecting the fact that queens live sheltered within their colony (Wilson, 1971). However, it seems curious that low immune response should be achieved by active suppression of immune capability. Furthermore, this conclusion contrasts with findings in honeybees, where queens have been reported to have higher levels of PO activity than workers (Schmid, Brockmann, Pirk, Stanley, & Tautz, 2008). A second possibility is therefore that queens have higher levels of PPO and PO than workers, associated with their greater longevity, and require high Spn27A expression to keep PO levels in check when there is no active immune threat.

**FIGURE 6** Barplot showing CpGO/E and GpCO/E as a function of caste-biased expression in legs.
Gene ontology terms enriched for genes with caste-biased expression

The lack of consistency in expression patterns between age/tissue contexts was particularly manifested at the level of GO terms. No GO term showed a consistent pattern of caste-biased gene enrichment across the four age/tissue contexts analysed in this study, and some GO terms were caste-biased in opposite directions in different age/tissue contexts. Enrichment of a GO term does not necessarily signal an up-regulation of the related biological process, as genes assigned a GO term may up- or down-regulate that process. Enrichment of a GO term in opposite directions in different contexts may therefore not necessarily indicate that the biological process is also biased in different directions, but it does at least indicate that the process is being modulated differently in the two contexts.

A full list of significantly enriched GO terms can be found in Data S4. Here, we discuss broad GO categories of interest that are consistent across more than one age/tissue context. GO terms linked with organ morphogenesis and development were enriched for caste-biased genes in both the brains and legs of 1-day-old individuals. A likely explanation for this finding is that individuals of this age still undergo important changes in their morphology and physiology (Keller & Passera, 1989; Passera & Keller, 1992). Caste-biased expression of genes related to development could therefore reflect the different juvenile developmental processes between the two castes.

The enrichment of caste-biased genes in GO terms linked to cellular respiration may reflect the different energy requirements of queens and workers. Interestingly, the bias was significant in opposite directions in 1-day-old and 2-month-old individuals. The bias was towards queens in the legs of 1-day-old individuals, when queens are accumulating resources for their mating flight. In contrast, the bias was towards workers in legs and brains of 2-month-old individuals, when the legs and brains of queens are largely at rest as they focus on egg-laying. The expression differences between queens and workers are thus changeable and may track the changing requirements of these castes with age.

Gene ontology terms linked to metabolism and biosynthesis were enriched for queen-biased genes in the legs of both 1-day-old and 2-month-old individuals. Metabolism GO terms were also enriched in a WGCNA module strongly associated with caste in brains. These differences may reflect the higher nutritional intake of queens, as well as their need to accumulate lipid and protein reserves for the purposes of reproduction (Keller & Passera, 1989). Metabolism stands out as one of the few processes that are consistently found to be associated with caste differentiation across studies and taxa. Genes involved in metabolism have previously been found to be differentially expressed between castes in both larvae and adults of other social insects taxa (wasps: Hoffman & Goodisman, 2007; Toth et al., 2010; honeybees: Barchuk et al., 2007; Cristina et al., 2006; Grozinger et al., 2007; ants: Bonasio et al., 2010; Patalano et al., 2015) and have consistently been found to be queen-biased (Smith, Toth, Suarez, & Robinson, 2008; see Berens et al., 2015 for an exception in wasp larvae). Whether these queen-biases in metabolism are simply a result of their different diets, or whether they represent a causative factor underlying caste differentiation, remains to be explored.
4.4 Variability in gene expression

Variability in gene expression between biological replicates was higher in 1-day-old individuals than 2-month-old individuals. A possible explanation for this is that expression in very young individuals changes rapidly. One-day-old individuals in our study were collected between 0 and 24 hr post-eclosion from the pupal cocoon, potentially leaving variation in the developmental state at which the individuals were collected. Even if all individuals were collected at the same time post-eclosion, they may not be released from their pupae by their fellow workers at exactly the same moment, leading to further variation in developmental stage. Interestingly, the effect was much stronger in brains than in legs, suggesting that gene expression in brains is undergoing greater changes than in legs at this early stage in adult life. An alternative explanation could be that gene expression is less tightly regulated in 1-day-old individuals, perhaps because gene expression needs more than a few hours of adulthood before stabilizing to optimal levels.

4.5 CpG analysis

Coding regions exhibited a higher signal of historical germline methylation than intronic, flanking and nongenic sequences, matching the general observation in hymenopterans that methylation is concentrated in exons (Bonasio et al., 2012; Elango et al., 2009; Glastad et al., 2016; Libbrecht et al., 2016; Lyko et al., 2010; Walsh et al., 2010; Wang et al., 2013; Xiang et al., 2010). In both coding and noncoding region, the signal of historic germline methylation was stronger in unbiased than either caste-biased or age-biased genes. This is consistent with differences in methylation level between caste-biased and unbiased genes in honeybees, ants and termites (Elango et al., 2009; Glastad et al., 2016; Patalano et al., 2015), and extends these results to age-biased genes, supporting the notion that methylation in invertebrates occurs mainly in housekeeping genes, perhaps serving to stabilize expression (Bonasio et al., 2012; Foret et al., 2009; Lyko et al., 2010; Roberts & Gavery, 2012; Simola et al., 2013). The association between caste-biased expression and methylation signal was consistent across nearly all age/tissue contexts. Only queen-biased nongenic sequences in a single context showed a stronger signal of methylation in caste-biased than unbiased genes. Studies investigating the association between historical germline methylation and caste bias are therefore less exposed to the confounding factors of age and tissue than studies that directly explore the expression differences that underlie caste differentiation.

An important assumption that underlies CpG\textsubscript{O/E} analysis is that highly methylated regions will become depleted in CpG dinucleotides due to the hypermutability of methylated cytosines (Bird, 1980). An alternative interpretation is that selection preserves regions with high CpG density if it is important for these regions to be potentially methylated (Elango et al., 2009), which would reverse the conclusion to be drawn from our results. However, our conclusions that methylation is concentrated in gene bodies and in caste-biased genes are supported by other studies that investigated levels of active methylation (Bonasio et al., 2012; Libbrecht et al., 2016).

CpG depletion is likely to be particularly strong in regions that are consistently methylated across cell types and developmental stages, as this will include the germline, where mutations must occur for CpG depletion to result. Regions that are plasticly methylated in specific nongermine contexts will experience only somatic mutations that will not propagate and leave their mark in the genome. CpG\textsubscript{O/E} analysis is therefore unable to detect genes that are differentially methylated between castes. Whether consistent caste-biased methylation exists at all has even been called into question (Libbrecht et al., 2016). Rather than making a claim about the caste-biased methylation status, we report that the CpG density in the transcriptome suggests that the global hypermethylation of caste-biased genes commonly reported in social insects also occurs in L. niger, and, importantly, that this difference is consistent with respect to the age/tissue context in which that caste bias exists.

4.6 Rates of molecular evolution

Our finding that dN/dS is higher in caste-biased genes than unbiased genes is consistent with results in other species, where genes differentially expressed between phenotypic morphs such as castes or sexes have been found to have higher proportions of nonsynonymous substitutions than unbiased genes (Ellegren & Parsch, 2007; Harpur et al., 2014; Hunt et al., 2010, 2011, 2012; Mikeyev & Linksvayer, 2015; Purandare, Bickel, Jaquery, Rispe, Brisson, 2014; Roux et al., 2014; Wang, Werren, & Clark, 2015; Whittle & Johannesson, 2013), although exceptions to this have been noted (Ferreira et al., 2013; Meisel, 2011; Smith et al., 2015). To explain the higher rate of molecular evolution in morph-biased genes, it has been proposed that morph bias may evolve mostly through drift in loosely regulated genes, rather than through directional selection (Helanter & Uller, 2014). As genes that are not tightly regulated are likely to be under weak purifying selection, morph-biased genes tend to accumulate slightly deleterious nonsynonymous substitutions at a higher rate than average. Under this scenario, morph-biased genes are expected to show high variability in expression within, as well as between morphs. Consistent with this view, we found that gene expression was more variable between individuals in caste-biased genes than unbiased genes, a result that has also been found in the fire ant S. invicta (Hunt et al., 2012). An alternative explanation for the greater variability in expression in caste-biased genes is that expression of these genes may be more situational than that of unbiased genes, and thus vary more strongly with the environmental factors that inevitable fluctuate between biological replicates. However, our conclusions are supported by results from S. invicta, where orthologs of caste-biased genes have been found to have higher rates of molecular evolution even in solitary wasps, supporting the idea that rapid sequence evolution pre-dated the emergence of caste-biased expression (Hunt et al., 2011). In contrast to these results indicating loose regulation of caste-biased genes, caste-biased genes have been shown to undergo higher rates of positive selection.
than unbiased genes in both ants and honeybees (Harpur et al., 2014; Roux et al., 2014). The emerging picture from our study and others is therefore that caste-biased genes evolve from a class of loosely regulated genes that are able to become caste-biased because their functions are not critical to the organism. These genes are then co-opted for specific caste-biased tasks, leading to positive selection on some genes or some sites within genes, while other genes/sites remain under loose regulation and continue to drift at a higher rate than unbiased genes.

The difference in $dN/dS$ between caste-biased and unbiased genes was not always significant. For example, for genes that were caste-biased in 1-day-old individuals, there was no significant difference in $dN/dS$ with unbiased genes. However, the direction of the difference in $dN/dS$ was the same as in the other age/tissue contexts and was close to significance in some cases. There was therefore no evidence that the effect of caste bias on the rate of molecular evolution is qualitatively dependent on the context in which the bias exists.

It has been suggested that the strength of selection should be weaker on worker-biased genes than on queen-biased genes because workers only accrue fitness indirectly by rearing the queen’s brood, leading to faster sequence evolution in worker-biased genes than queen-biased genes (Linksvayer & Wade, 2009). We found no support for this hypothesis, as rates of sequence evolution were the same in queen-biased and worker-biased genes. In other ants and in honeybees, rates of sequence evolution have even been found to be higher in queen-biased genes than worker-biased genes, thus contradicting theoretical predictions (Hunt et al., 2010, 2011). An intriguing possibility is that these results, combined with our own, indicate that colony success is more tightly dependent on optimal worker behaviour than optimal queen behaviour. This increased importance in worker performance would act to counterbalance the effect of worker-biased genes having only indirect effects on fitness.

We found that genes with age-biased expression showed a higher rate of sequence evolution than genes that were not age-biased, albeit only when the bias was towards 1-day-old individuals in brains. This may indicate that the evolution of age-biased genes has followed a similar process to that of phenotype-biased genes, and that the evolution of differentially expressed genes in general perhaps follows a consistent pattern. The correlation between $dN/dS$ and age-biased expression was weaker than with caste-biased expression, indicating that the strength of selection on a gene is more loosely associated with age bias than caste bias. This could be because caste-biased genes are exposed to reduced purifying selection in the caste where they show reduced expression. In contrast, the strength of selection on an age-biased gene is associated with the expected change in inclusive fitness after an individual reaches the age at which that gene is differentially expressed (Lee, 2003). Thus, as long as most of an individual’s expected inclusive fitness accumulates after 2 months of age there should be little difference in the strength of selection on genes biased towards 1-day-old or 2-month-old individuals. This is certainly the case in queens of L. niger, which do not begin egg-laying until they are 2 months old. It may also be true in workers of this species, as their expected lifespan may be substantially more than 2 months (Dussutour & Simpson, 2012; Kramer, Schaible, & Scheuerlein, 2016).

In conclusion, this study revealed that phenotype-biased expression patterns vary considerably with both age and tissue. We found that even age can have a stronger effect on gene expression than phenotype in a system with strongly morphologically and behaviourally differentiated plastic phenotypes. Using controlled comparisons across different ages and tissues, we were able to identify genes with consistent caste bias across biological contexts, including genes linked with longevity and immunity. In contrast to the strongly context-dependent patterns of caste-biased expression, the associations between caste-biased expression and either signals of methylation or the rate of molecular evolution were not qualitatively affected by the age or tissue in which the bias was measured. Future transcriptomic studies of phenotypic plasticity should be mindful of age and tissue in their comparisons, both using age-matched and tissue-matched comparisons and by including more than one age and tissue in their analysis. In this way, the interplay between biological context and expression can be considered, and general overarching associations between gene expression and plastic phenotypes can be obtained.

ACKNOWLEDGEMENTS

We thank Charlotte Soneson, Oksana Ribc and Eyal Privman for helpful advice on bioinformatic analysis and three anonymous reviewers for useful comments on the manuscript. For performing dissections of brain tissues, we thank Catherine Berney. For advice and guidance on ant rearing, we thank Romain Libbrecht. We thank Claire Morandin for providing the transcriptomes of Lasius neglectus and Lasius turicensis. This work was supported by a Seventh Framework Programme Marie Curie Intra European Fellowship (297898) to Eric Lucas, as well as by as well as by a Swiss National Science Foundation grant (310030_156732) and an advanced European Research Council grant (249375) to Laurent Keller.

DATA ACCESSIBILITY

Raw sequence reads for this study were deposited in the NCBI Short Read Archive: Accession nos. SRP069113, SRP097456.

AUTHORS’ CONTRIBUTIONS

E.L. designed the study, performed the experiments, analysed the gene expression and CpG data and wrote the manuscript. J.R. produced and analysed the $dN/dS$ data and wrote the manuscript. L.K. designed the study and wrote the manuscript. All authors read and approved the final manuscript.

REFERENCES


Davidson, A. M., Jennions, M., & Nicotra, A. B. (2011). Do invasive spe-


SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Lucas ER, Romiguier J, Keller L.
Gene expression is more strongly influenced by age than caste in the ant Lasius niger. Mol Ecol. 2017;26:5058–5073.
https://doi.org/10.1111/mec.14256