INTRODUCTION

Individuals of the same species might exhibit extensive phenotypic variation, which is influenced by both genetic and environmental sources (Agrawal, 2001; Stearns, 1989). Understanding the mechanisms of variation and how traits are inherited is the key to assess the adaptive potential of a population to changing environments (Fuhrman, Larsen, Steel, Young, & Beckman, 2018). Three factors influence the trait value of any individual: (a) genetic factors that define heritable differences within an environment, (b) environmental factors that influence the
genotypes (both gene expression and alterations of the DNA sequence such as SNPs) and (c) the variation of the trait, given that genetic and environmental factors are identical (Ziv, Shuster, Siegal, & Gresham, 2017). The resulting phenotypic variation promotes population performance and thus ecological/evolutionary success in the long run (Forsman & Wennersten, 2016). Understanding how each of these factors contributes to variation in quantitative traits remains a challenge, especially because phenotypic plasticity is common, thus blurring the relationship between genetic variants and phenotype.

Phenotypic plasticity is the ability of an organism to produce multiple phenotypes from a single genotype depending on the environment (Miner, Sultan, Morgan, Padilla, & Relyea, 2005). It is also an important mechanism that helps coping with environmental perturbations (Alberto et al., 2013; Charmantier et al., 2008). Although phenotypic plasticity is advantageous in heterogeneous and/or fast changing habitats, its maintenance is associated with costs (DeWitt, Sih, & Wilson, 1998; Van Buskirk & Steiner, 2009) and sometimes becomes maladaptive (Ghalambor, McKay, Carroll, & Reznick, 2007; Langerhans & Dewitt, 2002). A wide diversity of organisms exhibits phenotypic plasticity in response to biotic and abiotic factors in their environments (DeWitt et al., 1998; Harvell, 1990; Karban & Myers, 1989; Sultan, 2000) leading to changes in behaviour, morphology, physiology and life history traits. These plastic responses can be expressed either within the lifespan of a single individual (Young, Stanton, & Christian, 2003) or across generations (Miner et al., 2005).

Examining the genetic architecture of phenotypic traits not only identifies causal mutations but also helps in understanding past and predicting future evolutionary processes of adaptation (Ronnegard et al., 2016). Genetic variation can be studied at two levels of organization: patterns of gene expression and the DNA/RNA sequence level. Associating gene expression profiles to phenotypic traits can be accomplished by constructing gene co-expression networks, which identify clusters of co-expressed genes. Often co-expressed genes within one module or gene cluster share conserved biological functions revealing their potential genetic pathways (Subramanian et al., 2005). The benefit of gene network analysis lies in the opportunity to correlate the gene expression information to biological information; by gathering insights of the biological association of genes and traits, candidate genes can be identified. For example, a study on lake whitefish used a weighted gene co-expression network analysis (WGCNA) to identify gene clusters correlating to three phenotypic traits such as trophic behaviour, trophic morphology (gill rakers) and reproduction (Filteau, Pavey, St-Cyr, & Bernatchez, 2013).

Associating the sequence level of genetic variation to phenotypic traits is accomplished in genome-wide association studies (GWAS) which make use of single nucleotide polymorphisms (SNPs). GWAS have been applied extensively in a medical context (Busch et al., 2016; Elsing et al., 2016; Kao, Leung, Chan, Yip, & Yap, 2017), and in plant and animal breeding (Atanasov, Barboza-Barquero, Tiburcio, & Alcazar, 2016; Duijvestein et al., 2010; Hayes et al., 2009; Wood et al., 2009; Zhao et al., 2011).

The genotype–environment interaction (GxE) is a common phenomenon describing how a genetic variant has different phenotypic effects in different environments (Smith & Kruglyak, 2008). For example, human individuals with sickle cell anaemia have a survival advantage in endemic areas of malaria but are at a disadvantage in areas without malaria (Ferreira et al., 2011). Recently, biologists have applied genomic data and traditional pedigree information to explain phenotypic differences in life history traits such as horn shape in Soay sheep (Johnston et al., 2013), clutch size in collared flycatchers (Husby et al., 2015; Ronnegard et al., 2016) and Glanville fritillary butterfly (Duplouy, Wong, Corander, Lehtonen, & Hanski, 2017) where life history trade-offs may be involved in promoting genetic variation at one or several loci in the species.

Daphnia plays a central role in the trophic cascade of freshwater ecosystems (Carpenter et al., 2001; Sommer et al., 2003). This small, filter-feeding crustacean has become a widely used isogenic model organism in subfields as varied as ecology, toxicology and evolutionary biology, not only because of short reproduction times and easy handling, but also due to its ability to reproduce parthenogenetically. Daphnia is an ecological link between primary producers (algae) and secondary consumers (planktivorous fish and invertebrates) (Gaede & Straile, 1998). Predation is thus (together with food quality, parasites and toxins, among others) an important selective force shaping phenotypic variability and plasticity in Daphnia (Reger, Lind, Robinson, & Beckerman, 2018). Fish are visual predators and strongly influence growth and reproductive strategies in their prey (Macháček, 1991). This selection might appear to be strongly directional towards smaller body size (e.g. Gliwicz & Boavida, 1996; Weider & Pijanowska, 1993) and skewing towards an r strategy. However, phenotypic variability remains substantial, likely because of other contrary selective forces through other predator types and frequency-dependent selection.

To date, however, only a few studies in Daphnia research have associated phenotypic traits with the individual’s genetic setup by using transcriptomic information and/or genome-wide information. Routtu et al. (2014) identified putative QTLs underlying variation in life history traits and phototactic response for Daphnia magna by analysing an F2 panel. Herrmann, Henning-Lucass, Cordellier, and Schwenk (2017) used a genotype–phenotype association approach to show the effects of temperature on fitness in D. galeata. Further, Bento et al. (2017) investigated the genetic basis of parasite resistance in D. magna by using a quantitative trait locus (QTL) approach, and Bourgeois, Roulin, Müller, and Ebert (2017) analysed the correlation between parasite susceptibility and genotype in D. magna, making use of RADseq data collected for 97 genotypes; this approach allowed the identification of two genomic regions involved in resistance to the bacterium Pasteuria ramosa. Finally, Asselman, Pfrender, Lopez, Shaw, and De Schampheletaere (2018) and Malcom, Juenger, and Leibold (2018) have explored the correlation between gene co-expression modules and phenotypic traits in Daphnia pulex.

In the present study, we associate genotypic and phenotypic data of 24 clonal lines of European D. galeata by applying two approaches: a genome-wide association (GWA) analysis and a weighted gene co-expression network analysis (WGCNA). We applied a GWA analysis to the existing life history trait data set (Tams, Lüneburg, Seddar, Detampel, & Cordellier, 2018) and the corresponding
transcriptome-based SNP data set (Herrmann, Ravindran, Schwenk, & Cordellier, 2018; Ravindran, Herrmann, & Cordellier, 2019). Further, we applied a WGCNA to understand the association between gene expression and life history traits. Finally, we estimated heritability of the measured life history traits to infer their susceptibility to selection. Based on these analyses, we addressed the following questions: (1) Which SNPs (genotype) associate to different phenotypic life history traits in the two different environments? and (2) Which gene co-expression modules are correlated to life history traits in control environment? By answering these questions, we contribute to the understanding of the genetic basis of phenotypic variation in the absence and presence of predation risk in *D. galeata*.

2 | MATERIALS AND METHODS

The SNP data set used in the present study has been described in Herrmann et al. (2018) and is available on Dryad (https://doi.org/10.5061/dryad.56jt85j - SNP data Ravindran). The life history trait data set was described in Tams, Lüneburg, et al. (2018). Variance stabilized normalized read counts for the gene co-expression network analysis are available on Dryad (SI_2 - https://doi.org/10.5061/dryad.p85m5) together with an overview of sequencing and mapping results as described in Ravindran, Herrmann, et al. (2019). Functional annotation of the *D. galeata* transcriptome has been described in Huylmans, López Ezquerra, Parsch, and Cordellier (2016) and Ravindran, Herrmann, et al. (2019). A summary of the data sets used for the present analysis is represented in Figure 1. Although data were collected at different time points (years) in different studies, the clonal lines were maintained in standardized and constant conditions in the laboratory throughout the years. To avoid confusion with terminology, we use the term ‘clonal line’ for the 24 genotypes and the term ‘genotype’ for SNPs throughout this manuscript.

2.1 | Study organism

The cladoceran *D. galeata* is a widely distributed keystone species in freshwater ecosystems. Their parthenogenic life cycle allows rearing of many genetically identical individuals from one genotype (clonal line). Despite their identical genetic makeup, individuals from one clonal line can present different phenotypes. For all data sets (i.e. genotype (SNP), phenotype and gene expression) summarized below, we used 24 clonal lines of *D. galeata* from four European lake populations (six clonal lines per population): Greifensee (Switzerland), Jordan reservoir (Czech Republic), Lake Constance (South Germany) and Müggelsee (North Germany). Clonal lines were established from dormant eggs extracted from sediment cores, which have been used and described in previous studies (Henning-Lucass, Cordellier, Streit, & Schwenk, 2016; Herrmann et al., 2018). They were maintained in laboratory cultures (18°C, 16 hr light/8 hr dark cycle, food: *Acutodesmus obliquus*, medium: Aachener Daphnia Medium (ADaM)), (Klütten, Dülmer, Engels, & Ratte, 1994).

2.2 | Phenotype data set and design of life history experiment

In *Daphnia*, inducible defences are mediated by chemicals released by the predators (such as fish), called ‘kairomones’. This chemical messenger benefits the receiver rather than being advantageous...
for the predator (sender) itself (Brown, Eisner, & Whittaker, 1970). Phenotypic data originate from the life history experiment investigating the effect of fish kairomones on D. galeata (described in detail in Tams, Lüneburg, et al., 2018) for a total of 684 experimental individuals (aim: 24 clonal lines × 2 environments × 15 replicates = 720 individuals). Prior to the experiment, each clonal line was bred in kairomone-free water (control) and in kairomone water (fish) for two subsequent generations to minimize inter-individual variances. Breeding and experimental phases were conducted at a temperature of 20°C and a 16 hr light/8 hr dark cycle in a brood chamber with a light intensity of 30% of its maximum light intensity (Rumed, Typ 3201D). Experimental individuals (F2) were female neonates of the 3rd to 5th brood. The experiment lasted for 14 days. A total of ten phenotypic traits were recorded, eight life history traits: age at first reproduction (AFR) [day of releasing offspring from brood pouch], total numbers of broods per female ('broods') together with

**Figure 2** Violin plots visualizing the data distribution of the life history trait age at first reproduction ('AFR') for each clonal line. Average values for each clonal line is indicated by a full circle. (a) control environment (control). (b) predation risk environment (fish)
brood size (‘brood1’, ‘brood2’, ‘brood3’, ‘brood4’), total numbers of neonates per female (‘offspring’), ‘survival’ [in days] (see Table S1 in Tams, Lüneburg, et al., 2018); and two morphometric traits, somatic growth rate (‘SGR’) [in μm/day] and body ‘length’ [in μm]. ‘SGR’ was calculated based on body ‘length’ of the experimental individual at the end of the experiment (day 14) and an averaged ‘neonate body length’ of each clonal line. The latter was measured independently for neonates (15 replicates per clonal line), at the beginning of the experiment (day 0). The experiment revealed a change of life history trait values when exposed to fish kairomones concordant to previous studies, for example early maturation in the presence of fish kairomones. Nevertheless, there was high intraspecific phenotypic variation of life history traits within each population as well as among the four populations; the life history trait ‘AFR’ is showcased in Figure 2.

Based on the phenotypic data from the above-described experiment (hereafter ‘full phenotypic table’, see Table S1 in Tams, Lüneburg, et al., 2018), we derived two phenotypic data sets. The raw phenotype data values (Tables S1a and S2a) were normalized using quantile transformation (‘normalize.quantiles()’ function) in R. These normalized phenotypic trait values were used to calculate the mean (hereafter, ‘mean data set’, Table S1b) and statistical variance (hereafter, ‘variance data set’, Table S2b) using the functions ‘mean()’ and ‘var()’, respectively, in R, for each environment separately and were used as input for GWA analysis. For the analysis of heritability, we used the raw phenotypic data obtained from up to 15 individuals per clonal line (hereafter ‘full phenotypic table’, see Table S1 in Tams, Lüneburg, et al., 2018). An edited version of the ‘full phenotypic table’ was used as input for the WGCNA described in detail below.

2.3 Genotype data set and SNP calling

The SNP calling procedure was performed on the control environment and has been described in (Herrmann et al., 2018) along with sequencing, mapping and assembly parameters and summary statistics. Briefly, reads from an RNA-seq experiment under control conditions were aligned to the reference transcriptome published by (Huylmans et al., 2016), merged with samtools (Li et al., 2009) and realignment around indels was performed using GATK’s IndelRealigner tool (DePristo et al., 2011). The initial variant calls were made using GATK’s HaplotypeCaller. Using GenotypeGVCF tool in GATK, samples were jointly genotyped and a single vcf file was obtained. Variants were further filtered using VariantFiltration tool implemented in GATK with the following criteria: (a) clusterWindowSsize = 35; (b) Quality by depth (QD) < 2.0; (c) Fisher Strand (FS) > 30.0. The criterion QD takes into account coverage (AD) and quality of the variant (QUAL). These settings were recommended (see GATK’s ‘best practices’) when dealing with uneven coverage data typical for RNA-seq data (DePristo et al., 2011).

To use the SNP data for GWA analysis, we further filtered variants with a minor allele frequency (MAF) of 0.1 to exclude rare variants. Furthermore, indels were excluded and only biallelic sites were considered for further analysis. A total of 144,938 SNPs distributed in 11,857 transcripts (median number of SNPs per transcript: 7, ranging between 1 and 187) were used as input for GWA analysis (will be made available on Dryad).

2.4 Genotype–phenotype association analysis

The genome-wide association analysis was performed using the ‘GenABEL’ package version 1.8 (Aulchenko, Ripke, Isaacs, & Van Duijn, 2007) in R. Each SNP (genotype) was associated with the ten life history traits (phenotype) in the absence (hereafter, control environment) and presence (hereafter, fish environment) of fish kairomones. A quality check (QC) of our input genotype and phenotype data was performed using ‘check.marker()’ function in ‘GenABEL’. This QC step helps to select which SNPs and individuals should enter the GWA analysis based on call rate, minor allele frequency (MAF), value of the chi-squared test for Hardy–Weinberg equilibrium (HWE), heterozygosity and IBS (identity by state across a random sample of markers). This step is repeated until no individuals or markers are removed further. After QC, 20 clonal lines (all except G1.6, J1, J2 and M12) from the phenotype data and 114,978 SNPs passed all the criteria and were used for all further genotype–phenotype association analysis (univariate, multivariate and GxE).

‘GenABEL’ was used to perform a univariate analysis, where each SNP was tested for association to the normalized mean and variance phenotypic trait values in each environment (control and fish). To test for univariate association between a phenotypic trait and SNPs, the function ‘egscore()’ was run with 100,000 permutations. This function implements the method of Price et al. (2006, EIGENSTRAT), which makes use of principal components of the genomic kinship matrix to adjust both phenotypes and genotypes for possible stratification. The -log10 p-values were calculated automatically by ‘GenABEL’ and visualized using Manhattan plots in R.

Multivariate analysis was done using ‘MultiABEL’ package version 1.1-6 in R (Shen et al., 2017), where each SNP was tested for association to the combination of all ten life history traits in each environment. For the multivariate analysis, the function ‘MultiLoad()’ was used to test for association between a SNP and combination of all phenotypic traits and was permuted for 100,000 times.

PLINK v.1.07 (Purcell et al., 2007) was further used to test for differences in genotype–phenotype associations between the two environments. The ‘--gxe’ command was used to reveal the univariate genotype–environment interactions (GxE). For the multivariate GxE analysis, multivariate PLINK was used with the option ‘--mqlfam’. Settings were applied to correct for population stratification in the data set by permuting for 100,000 iterations within populations for the univariate GxE and 1,000 iterations within populations for multivariate GxE analysis.

Both univariate and multivariate analyses were performed for the control and fish environment and the GxE interaction separately. For the multivariate and GxE (univariate and multivariate) analysis, all p-values obtained were corrected for multiple testing using the
Bonferroni correction method in R (R Development Core Team, 2008). The \(-\log_{10} p\)-values were calculated on adjusted \(p\)-value and visualized using Manhattan plots in R. A SNP was said to be associated with a phenotypic trait if it had a \(-\log_{10} p\)-value of \(\geq 1.3\) (\(p\)-value < .05).

2.5 | Estimation of trait heritability

Thanks to the high replication of the life history data, we could precisely estimate heritability for life history traits using the function ‘marker_h2()’ in the R package ‘.ut’ (v1.3, Kruijer et al., 2015). Originally conceived for immortal lines of plants, which resemble clonal lines in Daphnia, this function uses a genetic relatedness matrix and phenotypic observations at individual level. Heritability estimates (\(H^2\)) are based on REML estimates of the genetic and residual variance and their standard errors. Most importantly and unlike the GWA analysis, this procedure was conducted on the full data set and not on summary statistics for each genotype, that is 563 data points were used (274 and 289 observations for the control and the fish conditions, respectively).

A centred relatedness matrix was calculated based on MAF filtered SNP data (bed file derived from tped/tfam files used in PLINK in this study) in ‘gemma’ (v0.98.1, Zhou & Stephens, 2012). Heritability was calculated separately for each environment and each phenotypic trait. Additionally, heritability was calculated on the ‘full phenotypic table’, including ‘environment’ as a covariate, similarly to the procedure outlined in Jury, Delano, and Toonen (2019). This analysis was conducted on the 20 genotypes passing all filters mentioned above (all but G1.6, J1, J2 and M12).

2.6 | Gene co-expression network analysis—Linking gene co-expression and life history traits

Gene expression data for all 24 clonal lines were available only for the control environment. The gene co-expression network analysis was based on variance stabilized read counts obtained from HTSeq data used in the R package ‘DESeq2’ (Love, Huber, & Anders, 2014) by Ravindran, Herrmann, et al. (2019) to investigate differential gene expression at population level between the four European D. galeata populations in control environment. We applied a weighted gene co-expression network analysis (WGCNA) by using the R package ‘WGCNA’ v. 1.6.1 to find putative pathways from the highly correlated genes clustered in modules (Langfelder & Horvath, 2008).

A single, signed, weighted gene co-expression network was constructed on a workstation with the R environment v.3.2.3, whereas subsequent analysis was performed in the R environment v.3.4.2 (R Development Core Team, 2008). Gene modules containing co-expressed genes were identified using the topological overlap matrices (TOM) with a soft cut-off threshold of 8 in ‘WGCNA’. Module eigengenes (ME) were calculated as the most representative gene within a module and were clustered in an eigengene dendrogram to reveal their relationships (data not shown).

2.7 | Module eigengene-trait correlation

Module eigengenes were correlated to phenotypic life history trait data in control environment. In the gene expression analysis, we had three biological replicates per clonal line, whereas we had one mean value for every life history trait measured per clonal line. The number of replicates has to match to meet the requirement of the gene co-expression analysis (equality of dimension). If we assigned the same mean trait value to all three biological replicates, potential false or inflated correlations (pseudoreplication) would have been the result. To avoid this artefact and to obtain one unique mean trait value per replicate and clonal line, the calculation of the mean trait value was based on the random selection of 75% of the available individual trait value of a life history trait (see R script ‘2_MEtraitCor.Rmd’ in Dryad, R function ‘sample_frac()’ in ‘dplyr’ package). For example, we had trait values for 15 individuals for the life history trait ‘broods’ in the clonal line J2. In the first resampling step, 75% of the individual trait values were selected to calculate a mean. The process was repeated twice to obtain three randomized mean values for this life history trait per clonal line. This step was repeated for every trait value in every clonal line. A similar approach of assigning life history trait mean values was used to perform a gene co-expression network analysis in D. pulicaria (Frisch, Becker, & Wojewodzic, 2020). Finally, the correlation of module eigengenes and the resampled life history trait mean values was calculated. This whole procedure of resampling to calculate randomized means and their correlation to module eigengenes was repeated 10,000 times to verify the robustness of ME-trait correlations. ME-trait correlations were considered as robust if occurring more than 95% of all iterations. ME-trait correlations with a correlation value of >±0.5 absolute value were considered as significant. Significant ME-trait correlations resulted in transcript sets of interest which were extracted to investigate their biological functions. In addition, hubgenes—the most interconnected gene per module—were identified and their biological functions were explored.

2.8 | Functional analysis

For every SNP/transcript associated with a life history trait in both GWA and WGCNA, we assigned Gene Ontology (GO) terms using annotations from Huylmans et al. (2016). An enrichment analysis with ‘topGO’ (Alexa & Rahnenfuhrer, 2016) was performed to investigate their biological relevance. We identified orthologs and paralogs for the transcripts associated with a phenotypic trait using orthoMCL cluster information from the reference transcriptome of D. galeata (Huylmans et al., 2016). These orthoMCL clusters are based on orthologous protein sequences from three Daphnia species (D. galeata, D. magna and D. pulex) and two insect species (Drosophila melanogaster and Nasonia vitripennis). To enhance our understanding of the ecological role of transcripts associated with life history traits, we performed a BLAST analysis on the Daphnia stressor database (Ravindran, Lüneburg, Gottschlich, Tams, & Cordellier, 2019). This enabled us to identify stressors for the candidate transcripts of interest from our GWA and WGCNA.
3 | RESULTS

3.1 | Genotype–phenotype association analysis

3.1.1 | Univariate analysis

When considering the mean values of replicates for each clonal line, we observed associations between one SNP, ‘Dgal_o2802d15519t2:1614’ and one life history trait ‘brood2’ in the control environment (Table 1). In the fish environment, the morphometric traits ‘SGR’ and ‘length’ were associated with the same SNP, namely ‘Dgal_sd13197565:1333’. Further, significant GxE interactions across the two environments were found for five life history traits, namely ‘brood2’, ‘brood3’, ‘brood4’, ‘broods’ and ‘survival’, associated with four, one, three, one and one SNPs, respectively. Among them, the SNP ‘Dgal_t1576c0t2:164’ was found to be associated with four life history traits, namely ‘brood3’, ‘brood4’, ‘broods’ and ‘survival’ in the univariate GxE analysis. No associations were found for the other phenotypic traits (Table 1 and Table S3).

With the variance values of replicates for every clonal line (Table 1), we observed significant associations for the traits ‘brood4’ (1 SNP) and ‘broods’ (1 SNP) between genotypes and life history traits in the control environment. The trait ‘brood1’ showed association with three SNPs in the fish environment. In the GxE analysis, seven traits were associated with a SNP: ‘brood1’ (2 SNPs), ‘brood2’ (16 SNPs), ‘brood3’ (2 SNPs), ‘brood4’ (1 SNP), ‘broods’ (3 SNPs), ‘offspring’ (1 SNP) and ‘survival’ (1 SNP). One SNP among them, the SNP ‘Dgal_o14292d4485t1:1732’ was associated with all life history traits in the control environment of the ‘mean data set’ (Table 2). However, no SNPs were significantly associated with all phenotypic traits in the fish environment of ‘mean data set’. For both environments, control and fish, no significant genotype–phenotype association was found in the ‘variance data set’. The multivariate GxE analysis resulted in one SNP (Dgal_o2092t4:604) to be associated with all life history traits in the ‘variance data set’.

3.1.2 | Multivariate analysis

A multivariate testing was performed using the mean and variance trait values, where we assessed the effect of one SNP on all phenotypic traits combined, taking the interdependence of the phenotypic traits into account. Multivariate testing has been shown to be more powerful compared to univariate analysis (Galesloot, van Steen, Kliemeny, Janss, & Vermeulen, 2014). We identified one SNP (Dgal_o14292d4485t1:1732) to be significantly associated with all life history traits in the control environment of the ‘mean data set’ (Table 2). However, no SNPs were significantly associated with all phenotypic traits in the fish environment of ‘mean data set’. For both environments, control and fish, no significant genotype–phenotype association was found in the ‘variance data set’. The multivariate GxE analysis resulted in one SNP (Dgal_o2092t4:604) to be associated with all life history traits in the ‘variance data set’.

3.2 | Trait heritability estimates

We report variance (genetic and residual) and trait heritability in Table 3 for the estimation on the ‘full phenotypic table’ while setting ‘environment’ as a covariate. Further results per environment are provided in Table S4. For the traits ‘survival’ and ‘brood4’, heritability was near 0 due to an overall lack of variance in these traits. For the other traits, heritability ranged between 0.22 (±0.09) and 0.71 (±0.14). The morphometric traits ‘length’ and ‘SGR’ had the highest heritability, and life history trait ‘offspring’ was also on the upper end.

3.3 | Gene co-expression network analysis—Linking gene co-expression and life history traits

The single network construction resulted in 44 modules of co-expressed transcripts in control environment (Figure S1). Most transcripts were assigned to the module ‘turquoise’, ‘blue’, ‘brown’ and ‘yellow’. The ‘grey’ module is the largest and includes all transcripts which could not have been assigned to any module (22%; n = 7,297). For each module, the hubgene was identified. Phenotypic trait information was correlated to module eigengenes (ME) to assess the biological meaning of modules. Only one significant module-trait correlation was found

| Trait | Control: $-\log_{10}p \geq 1.3$ | Fish: $-\log_{10}p \geq 1.3$ | GxE: $-\log_{10}p \geq 1.3$
|-------|-----------------|-----------------|-----------------
|       | Mean | Variance | Mean | Variance | Mean | Variance |
| brood1 | 0 | 0 | 0 | 3 (3) | 0 | 2 (2) |
| brood2 | 1 (1) | 0 | 0 | 0 | 4 (3) | 16 (13) |
| brood3 | 0 | 0 | 0 | 0 | 1 (1) | 2 (2) |
| brood4 | 0 | 1 (1) | 0 | 0 | 3 (3) | 1 (1) |
| AFR | 0 | 0 | 0 | 0 | 0 | 0 |
| broods | 0 | 1 (1) | 0 | 0 | 1 (1) | 3 (2) |
| offspring | 0 | 0 | 0 | 0 | 0 | 1 (1) |
| survival | 0 | 0 | 0 | 0 | 1 (1) | 1 (1) |
| length | 0 | 0 | 1 (1) | 0 | 0 | 0 |
| SGR | 0 | 0 | 1 (1) | 0 | 0 | 0 |

Note: Number of significant SNP (and their transcripts indicated in brackets) associations to each life history trait in the ‘mean’ and ‘variance’ data set for the control, fish and GxE interaction.
3.4 | Functional analysis

3.4.1 | Gene Ontology analysis

Gene Ontology (GO) terms were assigned to transcripts identified in the univariate and multivariate GWA, in addition to hubgenes and transcripts of the ‘darkorange’ module which were identified in WGCNA. In total, GO terms were assigned to 39 transcripts (15 transcripts in GWA and 24 transcripts in WGCNA, Table S3) and to 15 out of the 44 hubgenes (Table S5). All these assigned GO terms were enriched for metabolic processes (Table S6). There were 18 GO terms assigned to the hubgenes and included functions for enzyme activities, binding and transport activities which are important for general metabolic processes.

3.4.2 | Comparative genomics

To gain insights into the biological relevance of our candidate transcripts identified in the GWA analysis and WGCNA, we made use of orthoMCL cluster information of the D. galeata reference transcriptome (Huylmans et al., 2016). Nine candidate transcripts identified in the ‘mean’ GWA analysis were assigned to eight orthoMCL clusters, and 20 transcripts identified in the ‘variance’ GWA analysis were assigned to 19 orthoMCL clusters (Table S3, Figure S2). Fifty-eight out of 85 transcripts identified in WGCNA were assigned to 53 orthoMCL clusters (Table S3). There was no overlap between the orthoMCL clusters identified in GWA analysis and WGCNA. One-third of the orthologs are conserved (i.e. comprised in clusters also containing insect proteins), whereas the other two-thirds are Daphnia specific. Of the Daphnia specific orthologs, 23% seem to be D. galeata specific (3).

Identifying which stressor causes differential expression of a transcript is another way of looking into the functional aspects of a transcript (Figure 3). To this aim, we compared the sequences of transcripts identified in the GWA and WGCNA to those available in the Daphnia stressor database. For 18 transcripts (7 from GWA and 11 from WGCNA), the database contained a Daphnia transcript whose expression was influenced by stressors in previous studies (Table S3). These stressors are mostly abiotic factors such as salinity (2), temperature (2) and phosphorus (1), among others; and only two are biotic factors: sex-biased (6) and fish kairomones (1). The transcript ‘Dgal_o3698t3’ with the stressor-based hit for fish kairomones (Daphnia pulex transcript hx AUG25s185g21111) is grouped to the orthogroup ‘ORTHO_ALL324’ which contains protein-coding transcripts of D. galeata (10), D. pulex (3), D. magna (2) and Drosophila melanogaster (2) (Table S3). However, no GO annotation exists for that specific transcript or any of the other D. galeata transcripts clustered in the orthogroup ‘ORTHO_ALL324’.

4 | DISCUSSION

4.1 | Trait heritability

The question of whether Daphnia are capable of adaptation to predation risk relies critically on whether traits under strong selective pressure are heritable, with individual variation upon which natural selection can act. Growth and reproduction strategies leading to a higher individual fitness are only relevant on the long term if they can be transmitted to descendants. In this study, we show that many of the traits are heritable to a certain extent. Put simply, residual variance is quite substantial and heritability is low to moderate in the traits related to reproduction, with the exception of total number of ‘offspring’. On the contrary, traits related to growth are highly heritable, and this is relevant in the context of size-dependent predation pressure.

4.2 | Genotype—phenotype association analysis

We found no direct evidence of genotype—phenotype associations for most of the phenotypic traits except for ‘brood2’, ‘brood4’ and
broods in the control environment and 'length', 'SGR' and 'brood1' in the fish environment. These genotype–phenotype associations for the above-mentioned traits are promising candidates for true associations. To verify these associations, subsequent investigations should test the candidate SNPs for their biological relevance for the corresponding phenotypic trait experimentally, for example via a RNA knock-out or silencing of the respective transcript/gene. Alternatively, a broader sampling strategy could be applied to establish a diverse clone panel, and their genotypes and phenotypes recorded to explicitly test the association identified here. It is to be noted, however, that trait recording might cause a bias in 'brood4' as the experiment was stopped at 14 days, and the value for this trait was set to zero if the individual had not produced a fourth brood. Our results for this trait are therefore likely an artefact.

The observed significant GxE interaction effect for the trait 'offspring' in the 'variance data set' could be best explained by the significant associations of SNPs to 'brood1', 'brood2', 'brood3', 'brood4' and 'broods' since these life history traits are not independent. The trait 'offspring' describes the total number of offspring per female and thus includes the total number of offspring for the first, second, third and fourth brood ('brood1' to 'brood4'), whereas the trait 'broods' describes the total number of broods. This correlation of life history traits might have biased the statistical genotype–phenotype association and led to a false-positive association in the GxE analysis.

By leveraging the high replication in phenotypic trait recording and considering the variance of traits instead of mean values, we did include a proxy for phenotypic plasticity. Predator-induced plastic changes in Daphnia are a textbook example of phenotypic plasticity (Tollrian, 1995), and the data gathered by Tams, Lüneburg, et al. (2018) and used here demonstrate this. Reger et al. (2018) showed how predation drives local adaptation of phenotypic plasticity in D. pulex. Here, we could identify significant associations between the genotype of D. galeata clonal lines and the variance for reproduction-related phenotypic traits (e.g. 'broods', 'offspring'). Although the same precautions as above apply for 'brood 4' for interpreting these results, plasticity for 'brood1' and 'brood2' sizes might have a genetic basis and this warrants further investigation.

The overall number of 131 identified genotype–phenotype associations for ten phenotypic traits in our analysis is concordant to previous genotype–phenotype associations in different organisms. Two genomic regions with about 30 out of 7,034 SNPs were associated with parasite resistance in D. magna (Bourgeois et al., 2017). One GWAS found 139 genotype–phenotype associations to one...
morphometric trait, wing shape, in *Drosophila melanogaster* (Pitchers et al., 2019). Further, GWAS revealed a total of 410 genotype–phenotype associations to 30 phenotypic traits in a wild tree population (McKown et al., 2014), whereas one out of 50,000 SNPs was significantly associated with the trait clutch size in the collared flycatcher (Husby et al., 2015).

Several explanations for the lack of association between the genotype and phenotypic traits are possible: low statistical power, low genetic variation, polygenic basis (many loci of small effect), missing/untargeted genomic regions and epigenetic effects, among others. The statistical power of our study is limited because of the relatively small sample size. GWAS is often conducted on human populations with many hundreds of individuals, and in a breeding context (i.e. for plants) with a few hundred accesses (e.g. Atwell et al., 2010; van Heerwaarden, Hufford, & Ross-Ibarra, 2012; Zhao et al., 2011). It is all the more remarkable that we uncovered a few associations, but the lack of association found here should not lead to the conclusion that the phenotypic traits have indeed no genetic basis.

Low genetic variation can be dismissed in our case because substantial genetic variability at the sequence level among the studied *D. galeata* populations was reported in Ravindran, Herrmann, et al. (2019). To date, the genome-wide variation has not been assessed systematically in *D. galeata* populations and is the subject of ongoing investigations. *Daphnia*’s mode of sexual reproduction, dispersal or resting stages and clonal growth all contribute to a high diversity among populations; it is especially true for species such as *D. galeata* which inhabit large lakes and establish permanent populations monopolizing resources (De Meester, Vanoverbeke, Kilsdonk, & Heerwaarden, 2012; Hufford, & Ross-Ibarra, 2012; Zhao et al., 2011). It is all the more remarkable that we uncovered a few associations, but the lack of association found here should not lead to the conclusion that the phenotypic traits have indeed no genetic basis.

GWAS are most powerful for phenotypic traits caused by one or few loci with strong effects with a minor sensitivity to environmental variation, which is rarely the case. The study by Routtu et al. (2014) on *D. magna* exposed to the presence and absence of fish kairomones reported that 8.2% of genetic variation was explained by the significant QTLs detected for phototactic behaviour and it could stem from multiple genes with small effects. Similarly, another study by De Meester (1991) in *D. magna* reported that the genetic variation underlying phototactic trait is largely additive in nature. Thus, the phenotypic traits might reflect a pattern of additive, polygenic inheritance. The relative high heritability we estimated for the morphometric traits ‘length’ and ‘SGR’ seems at odds with the fact that almost no association was found through the GWA. A polygenic basis for these traits is therefore likely and would warrants further investigation.

We studied genotype–phenotype associations at the transcriptome level, thereby introducing the limitation of looking only at genotype–phenotype associations in coding regions. Our study design did not allow testing for the role of noncoding regions since genomic data of *D. galeata* are not yet available. No significant associations between the phenotype and the transcriptome-based SNPs could also imply that the genetic basis of phenotypic variation is not predominantly on the coding sequences, but rather on the noncoding regions of the genome. Such associations between phenotypic traits and polymorphism in noncoding regions have been reported by McKown et al. (2014) in black cottonwood (*Populus trichocarpa*), where 152 out of 275 identified associated polymorphisms were in noncoding regions. Noncoding regions include essential regions for the complex mechanism of gene expression regulation, such as transcription factors and promoters which are spatially located in the close vicinity of a gene, whereas gene regulatory elements such as enhancers, locus control regions or insulators can be located several kilobases away from the expressed gene of interest (Babu, Janga, de Santiago, & Pombo, 2008).

Last but not least, future investigations should include as well the epigenetic level (Tak & Farnham, 2015), which we did not address in the present study. Only a small portion of heritability is explained by associated genetic factors (Trerotola, Relli, Simeone, & Alberti, 2015). Underlying factors and mechanisms such as different epigenetic modifications add to variation and plasticity without changing DNA sequences (Trerotola et al., 2015). The epigenetic level of inheritance in *Daphnia* is of increasing interest (reviewed by Harris, Bartlett, & Lloyd, 2012; Wojewodzic & Beaton, 2017). Although most of the existing and thus reviewed studies focus on *D. magna* in an ecotoxicological context, little to none is known if and how *D. galeata* passes on epigenetic modifications in general and under predation risk specifically.

### 4.3 Functional analysis

The functional analysis revealed GO enrichment for enzymatic activities and metabolic processes (17 in WGCNA and eight in GWA; Table S6). Since these functions are generally important for the survival of an organism, no specific conclusions can be drawn in the context of our study. However, it is to be noted that only 34% of the *D. galeata* reference transcriptome has GO annotations. Thus, we cannot exclude that a potential bias in the gene set enrichment analysis exists due to lack of additional GO terms. In general, the gene set enrichment analysis emphasizes the need of further functional annotation for the existing *Daphnia* genomes to improve biological valid conclusions.

When looking closer to the functional analysis of candidates from WGCNA, we can highlight hubgenes of modules with biological functions that are involved in predator-induced responses, according to the predation risk experiments in Tams, Nickell, Ehring, and Cordellier (2018). Growth-related GO terms, such as ‘chitin binding’ and ‘chitin metabolic process’, were identified for the ‘turquoise’ hubgene (3,017 transcripts), whereas the GO term ‘structural constituent of cuticle’ was identified for the hubgenes ‘paleturquoise’ (222 transcripts) and ‘darkgreen’ (371 transcripts). Predator-induced responses in *Daphnia* include changes of body size as well as morphological modifications (e.g. Laforsch & Tollrian, 2004; Tollrian, 1995). For example, a smaller body size in the presence of fish kairomones was observed in *D. galeata* from Greifensee, in contrast to a larger
body size in the presence of Chaoborus kairomones, an invertebrate predator (Wolinska, Löffler, & Spaak, 2007).

Digestion-related GO terms for ‘cysteine-type peptidase activity’ were found for the hubgenes of the ‘royalblue’ (405 transcripts) and ‘lightcyan1’ (70 transcripts) modules. Peptidases are major digestive proteases in the gut of Daphnia (von Elert et al., 2004) and are important for the energy allocation for Daphnia growth (Schwarzenberger, Küster, & Von Elert, 2012). Thus, the identified gene co-expression modules with hubgenes annotated to relevant GO terms are likely important for predator-induced responses in Daphnia galeata (Tams, Nickel, et al., 2018). However, this needs to be verified, for example by differential gene expression and gene co-expression analysis of all 24 clonal lines exposed to fish kairomones.

In the end, our functional analysis allowed two straightforward conclusions. First, an integrative approach is beneficial to identify putative candidate transcripts/genes. Our integrative approach resulted in a candidate transcript list of overall 131 transcripts (14 from ‘mean’ GWA, 32 from ‘variance’ GWA and 85 from WGCNA) being involved in phenotypic variation of life history traits in D. galeata (Table S3). Second, genomes still lack functional annotations thus hindering the interpretation of biological relevant transcripts. As is the case in many other nonmodel species, only about one-third of the candidate transcripts had a GO annotation, limiting our conclusions.

Depending on the genes or transcripts of interest, further research can be conducted on these candidates whose expression was shown to respond to environmental stressors in previous studies (Figure 3). Furthermore, expansion of the Daphnia stressor database (Ravindran, Lüneburg, et al., 2019) as well as newly available data may help researchers to identify biologically relevant transcripts and to infer stress responses in other Daphnia and related species.

5 | CONCLUSIONS AND OUTLOOK

In this study, we explored the association of phenotype, genotype and environment in European D. galeata emphasizing the complexity of their interactions. The present study helped us to identify a total of 131 candidate transcripts associated with phenotypic traits. Hence, we are now one step closer to understand the genetic basis of phenotypic variation related to predation risk in Daphnia. The present study also brought to light some shortcomings. First, an appropriate GWA approach is missing to account for the clonal nature of Daphnia. We would have gained more information and statistical power by using the complete phenotypic data set of individuals (n = ~700) rather than phenotypic means per clonal line (n = 24). Second, although we found genotype–phenotype associations at the transcriptome level, we cannot exclude the role of noncoding regions in shaping phenotypic variation. Once genomic information is available, investigating the role of noncoding regions can help to understand the interplay of genotype, phenotype and environment. Further, whole genome sequence would allow localizing the candidates identified in this study and finding out whether neighbouring regions contain loci of interest that might explain the association.

Third, more complete functional annotation information for Daphnia would help to identify biologically meaningful transcripts. Finally, to better understand the influence of predation risk here simulated by the presence of fish kairomones on Daphnia life history traits, gene expression profiles under predation risk are needed for all 24 clonal lines in the fish environment (the present study only uses gene expression data obtained in the control environment independently). These gene expression profiles would allow the application of a differential gene co-expression network analysis between the two gene co-expression networks (control versus fish kairomones), further revealing biologically significant pathways and hence candidate transcripts. Overall, the identification of biologically significant transcripts being involved in predator-induced responses in Daphnia provides a valuable source for further investigations of the environment-dependent genotype–phenotype relationships in Daphnia.

ACKNOWLEDGMENTS

Open access funding enabled and organized by Projekt DEAL.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/jeb.13699.

DATA AVAILABILITY STATEMENT

Raw SNP data are available in Dryad Digital Repository: https://doi.org/10.5061/dryad.56l85. Raw life history trait data ('full phenotypic table') are available at the following url: 10.7717/peerj.5746/supp-1. The following data files will be uploaded in a Dryad repository and the accession number provided upon publication: MAF filtered SNP data used as input for performing GWA in PLINK (.tped file) and for calculating a relatedness matrix, variance stabilized normalized read counts used as input for WGCNA.

ORCID

Suda Parimala Ravindran https://orcid.org/0000-0002-6745-4977
Verena Tams https://orcid.org/0000-0002-9539-2603
Mathilde Cordellier https://orcid.org/0000-0001-7376-4560

REFERENCES

Agrawal, A. A. (2001). Phenotypic plasticity in the interactions and evolution of species. Science, 294, 321–326.
Alberto, F. J., Aitken, S. N., Alia, R., Gonzalez-Martinez, S. C., Hanninen, H., Kremer, A., … Savolainen, O. (2013). Potential for evolutionary responses to climate change - evidence from tree populations. Global Change Biology, 19, 1645–1661.
Alexa, A., & Rahnenfuhrer, J. (2016). tapGO: Enrichment analysis for gene ontology.
Asselmans, J., Pfreuder, M. E., Lopez, J. A., Shaw, J. R., & De Schampheelaere, K. (2018). Gene co-expression networks drive and predict reproductive effects in Daphnia in response to environmental disturbances. Environmental Science & Technology, 52, 317–326.
Atanasov, K. E., Barboza-Barquero, L., Tiberio, A. F., & Alcazar, R. (2016). Genome wide association mapping for the tolerance to the polyamine oxidase inhibitor guazatine in Arabidopsis thaliana. Frontiers in Plant Science, 7, 401.
Weider, L. J., & Pijanowska, J. (1993). Plasticity of Daphnia life histories in response to chemical cues from predators. Oikos, 385–392. https://doi.org/10.2307/3545351

Wojewodzic, M. W., & Beaton, M. J. (2017) The future of environmental epigenetics: insights using the clonal water flea model. Advances in Insect Physiology, 53, 287–312.

Wolinska, J., Löffler, A., & Spaak, P. (2007). Taxon-specific reaction norms to predator cues in a hybrid Daphnia complex. Freshwater Biology, 52, 1198–1209. https://doi.org/10.1111/j.1365-2427.2007.01757.x

Wood, S. H., Ke, X., Nuttall, T., McEwan, N., Ollier, W. E., & Carter, S. D. (2009). Genome-wide association analysis of canine atopic dermatitis and identification of disease related SNPs. Immunogenetics, 61, 765–772. https://doi.org/10.1007/s00251-009-0402-y

Young, T. P., Stanton, M. L., & Christian, C. E. (2003). Effects of natural and simulated herbivory on spine lengths of Acacia drepanolobium in Kenya. Oikos, 101, 171–179. https://doi.org/10.1034/j.1600-0706.2003.12067.x

Zhao, K., Tung, C. W., Eizenga, G. C., Wright, M. H., Ali, M. L., Price, A. H., ... McCouch, S. R. (2011). Genome-wide association mapping reveals a rich genetic architecture of complex traits in Oryza sativa. Nature Communications, 2, 467. https://doi.org/10.1038/ncomms1467

Ziv, N., Shuster, B. M., Siegal, M. L., & Gresham, D. (2017). Resolving the complex genetic basis of phenotypic variation and variability of cellular growth. Genetics, 206(3), 1645–1657. https://doi.org/10.1534/genetics.116.195180

How to cite this article: Ravindran SP, Tams V, Cordellier M. Transcriptome-wide genotype–phenotype associations in Daphnia in a predation risk environment. J Evol Biol. 2021;34:879–892. https://doi.org/10.1111/jeb.13699

SUPPORTING INFORMATION
Additional supporting information may be found online in the Supporting Information section.