Additive genetic variation of transcriptional regulation: metallothionein expression in the soil insect *Orchesella cincta*

D Roelofs, L Overhein, ME de Boer, TKS Janssens and NM van Straalen

Vrije Universiteit, Institute of Ecological Science, Department of Animal Ecology, De Boelelaan 1085, 1081 HV, Amsterdam, The Netherlands

Field-selected metal tolerance in *Orchesella cincta* is correlated with overexpression of the single copy cadmium (Cd) inducible *metallothionein* (*mt*). Previously, we have demonstrated large phenotypic variation in *mt* gene expression, and a higher frequency of high-expression phenotypes in a tolerant population. Here, we describe midparent–offspring regression analysis of *mt* gene expression in a laboratory culture originating from a noncontaminated natural population. Families were either not exposed ($n = 47$) or exposed to 0.5 μmol Cd per gram dry food ($n = 46$). Mean *mt* gene expressions normalized to 28S rRNA and β-actin RNA were generated using real-time RT-PCR applied to parents and offspring RNA and subjected to regression analysis. A significant heritability (h^2) for *mt* gene expression was estimated between 0.36 (β-actin normalized) and 0.46 (28S normalized) in Cd exposed families. Nontreated families did not yield a significant h^2 value. Restriction Fragment Length Polymorphism analysis of the metallothionein promoter sequence revealed eight promoter alleles that show structural variation. Three alleles show increased frequencies in families with high *mt* expression. Another gene, *croquemort* (isolated from a differential screening for 1 μmole Cd treatment) showed no h^2 of gene expression in response to 0.5 μmol Cd. This gene codes for a receptor-protein involved in recognition of apoptotic cells and may participate in the general stress response. The present data suggest that evolution of metal tolerance in *O. cincta* can occur in the field by selection for high *mt* expression due to structural changes in *mt* cis-regulation.
Heredity (2006) **96,** 85–92. doi:10.1038/sj.hdy.6800756; published online 23 November 2005

Introduction

Several recent reviews have suggested that transcriptional regulation of genes has had an important impact on the evolution of genotype–phenotype relationships (Purugganan, 2000; Wray *et al*, 2003). Thus, the regulatory sequences that flank genes are expected to be as evolutionarily important as the coding sequences themselves. Variation in these sequences may cause changes in gene expression followed by phenotypic change and fitness differences. Crawford *et al* (1999) argued that adaptive change in gene regulation has a greater probability of occurring than change in enzyme structure altering catalytic activity, because mutations in promoters are not constrained by codon usage or negative pleiotropic effects in the tertiary/quaternary structure of the protein. However, there is only limited data to support this argument. In this paper, we present evidence of the evolution of transcription regulation in a model for pollution-induced tolerance in a soil-living invertebrate.

Field-selected tolerance to heavy metals has been reported for *Orchesella cincta* (Arthropoda: Collembola,

springtails) populations occurring at metal-contaminated mining sites (Van Straalen *et al*, 1987). Populations from mining sites were characterized by a higher cadmium (Cd) excretion efficiency and a lower Cd-induced growth reduction (Posthuma, 1990; Posthuma *et al*, 1992). To determine whether *O. cincta* can genetically adapt to heavy metals, additive genetic variation for Cd excretion efficiency was estimated by Posthuma *et al* (1993) in a reference population (originating from a clean soil). They showed a significant heritability (h^2) of 0.33 in a parent–offspring analysis and a significant h^2 of 0.48 in a half-sib analysis for this trait, implying that the reference population would be able to respond to selection resulting from heavy-metal exposure.

At the molecular level, an unknown number of gene products may influence excretion efficiency of heavy metals. Among these, the small cysteine-rich protein *metallothionein* is involved in binding and sequestration of Cd in the gut (Hensbergen *et al*, 1999). Several studies have shown that metallothionein (*mt*) expression is induced upon heavy metal treatment (Stürzenbaum *et al*, 2001; Saydam *et al*, 2002). For instance, Stürzenbaum *et al* (2001) revealed that, of the two isoforms of MT present in *Lumbricus rubellus* (wMT1 and wMT2), wMT2 responds to Cd on the transcriptional as well as the translational level and is targeted to intracellular compartments that are involved in detoxification. However, no evidence was found for genetic adaptation of

L. rubellus populations occurring at contaminated sites (Sturzenbaum and Morgan, personal communication).

In the case of *Drosophila melanogaster*, Otto *et al* (1986) showed that Cd tolerance in the laboratory was accomplished by duplication of an *MtnB* allele, resulting in increased *Mtn* gene expression. Naturally-occurring amplified *Mtn* loci were identified (Maroni *et al*, 1987) and the ecological significance of these naturally amplified *Mtn* alleles was further investigated by Lange *et al* (1990). They compared frequencies of *MtnB* duplications among flies collected from metal contaminated and noncontaminated orchards at three locations in the US, but frequencies of *Mtn* duplications were not significantly higher at the contaminated sites. These results raised doubts that *Mtn* duplications and the resulting elevated *Mtn* gene expression of *Drosophila* had arisen through directional selection from metal exposure.

Based on Southern blot analysis Sterenborg and Roelofs (2003) suggested that the *mt* gene occurs as a single copy in *O. cincta*. Nevertheless, they showed that Cd-induced *mt* gene expression was 2–4-fold higher in a laboratory-raised strain originating from the metal polluted mining site in Plombières (Belgium) than in the reference population. In the absence of Cd treatment, however, the constitutive *mt* gene expression levels were not significantly different between the polluted and reference populations. In addition, there was a large variation between individual springtails in constitutive and Cd-induced *mt* expression. Individuals with high *mt* expression might be able to enhance sequestration/detoxification of the heavy metals found at increased levels at mining sites, thereby resulting in higher fitness, which could explain the evolution of tolerance at the mine site.

Analysis of the MT protein in a number of *O. cincta* populations throughout North-Western Europe has shown that five amino-acid mutations occur in six alleles. However, *mt* allele frequencies did not have greater variation in allele frequencies than neutral markers (Timmermans *et al*, unpublished data). Taken together, these data suggest that metal tolerance in springtails is more likely due to adaptive change in Cd-induced *mt* gene expression rather than structural variation in the gene itself.

In the present study, we investigate whether the large interindividual differences in *mt* gene expression are due to additive genetic variation. We decided to use a quantitative genetics approach, because the quantity of gene expression is the result of complex interaction between several gene products acting on the transcriptional regulatory sequences surrounding the gene of interest (Wray *et al*, 2003). Also, a microarray study of the *D. melanogaster* transcriptome suggested that a quantitative genetics approach utilizing analysis of variance-type methods is to be preferred when analyzing transcriptional h^2 (Jin *et al*, 2001). Rather than their system of microarray analysis, we used quantitative PCR, because this method can produce better quantitative estimates of gene expression.

An experimental setup was chosen similar to Posthuma *et al* (1993) to perform parent–offspring regression experiments measuring genetic variation of the *mt* molecular phenotypes in a laboratory culture of the reference population Roggebotzand (The Netherlands). Families from a laboratory strain originating from the

Roggebotzand population were either fed with noncontaminated food or with Cd amended food. To determine if Cd-induced *mt* gene expression is a heritable trait, the mean expression level of parents was compared to the mean expression level of their offspring using linear regression analysis. Furthermore, genetic variation for *mt* expression was compared to genetic variation in expression of a gene homologous to *croquemort* (*crq*), a macrophage receptor in *D. melanogaster*, mediating endocytosis of apoptotic cells (Franc *et al*, 1996, 1999). This gene was cloned from a differential screening with *O. cincta* laboratory culture animals exposed to 1 μmole Cd/g dry algae food and may be coregulated by Cd exposure (Roelofs and Mariën, unpublished data).

Materials and methods

Origin and treatment of animals

Approximately 500 *O. cincta* individuals were randomly collected from the reference site Roggebotzand (The Netherlands, Cd concentration in litter <0.005 μmol/g; Van Straalen *et al*, 1987). They were reared in a climate room (20°C, 75% humidity, LD 12:12 h) in PVC jars on a moist plaster of Paris layer in which they could feed on algae present on twigs. Laboratory-raised juveniles were allowed to grow for 2 weeks and subsequently separated from the original field population with a sieve to obtain a parent population for the crosses. At the age of 5 weeks 334 pairs of two randomly selected individuals from the parent population were placed together. Approximately, one-third (126) of the pairs generated offspring after a few days. Each clutch of approximately 40–60 eggs was immediately subdivided over four clean jars to minimize the effect of common environment during juvenile development. All animals were allowed to grow until the age of 7 weeks. Subsequently, both parents and four offspring (one from each replicate jar) of each family were transferred to a clean jar with 50 μl of clean algae paste and allowed to moult. Immediately after moulting, each animal was transferred to a new clean jar, containing either clean algae paste or paste containing 0.5 μmol Cd/g dry algae. During the exposure time of 2 days the jars were kept on a thick layer of Paris plaster soaked in water to prevent fluctuations in humidity inside the jars. After 2 days of exposure, animals were frozen in liquid nitrogen and stored at −80°C for RNA extraction.

RNA/DNA extraction and real-time RT-PCR analysis

RNA extraction was performed using the Spin Vacuum (SV) Total RNA isolation system (Promega). Animals were crushed in 250 μl SV Lysis buffer, of which 200 μl was used for RNA extraction as described by Sterenborg and Roelofs (2003). The remaining 50 μl Lysate were kept at −20°C until DNA extraction. Figure 1a shows an agarose gel loaded with total RNA samples. The 18S rRNA (lower band) and 28S rRNA bands (upper bands) can be observed indicating that the samples are not degraded by RNases. For cDNA synthesis 7 μl of total RNA (approximately 100 ng RNA/μl) was reverse transcribed with 200 units MML-V (Promega) and 25 μM random hexamers according to the manufacturer's instructions. Real-time RT-PCR was performed with SYBR green 2× mastermix (Finnzymes) and gene specific primers (Table 1) on an Opticon 1 real-time PCR

machine (MJ Research). The cycling program was: denaturation (95°C for 10 min), 2-step amplification and quantification (92°C for 15 s, 60°C for 1 min with a single
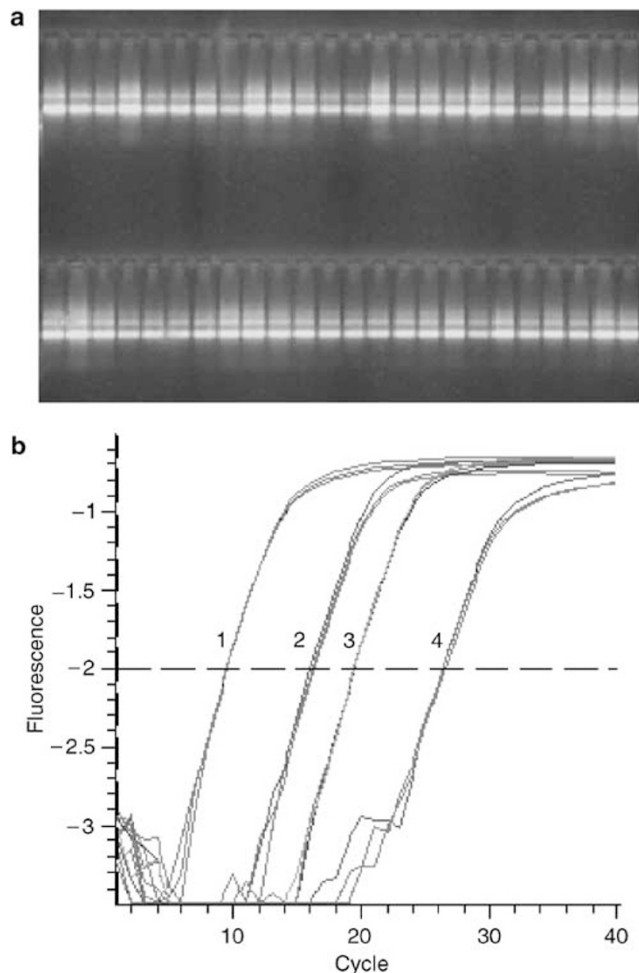

Figure 1 Integrity of total RNA samples and typical Quantitative Real-time RT-PCR profile. (a) In total, 48 total RNA samples were run on 1% agarose gel; presence of 18S rRNA (lower band) and 28S rRNA (higher band) in the samples indicate intact RNA. (b) Typical profile of a Quantitative Real-time RT-PCR from a single RNA sample, three pseudoreplicates of each gene were run. X-axis cycle number, y-axis log transformed fluorescence. Dashed line indicates Ct. (1) 28S rRNA profile, (2) β-actin profile, (3) mt profile, (4) crq profile.

fluorescence measurement), melting curve program (60–90°C with a heating rate at 0.1°C per second and fluorescence measurement every second). 28S rRNA and β-actin were chosen as internal controls for mt and crq gene quantification (see Figure 1b).

Real-time RT-PCR data were baseline-subtracted using the mean relative fluorescence from cycle 3–7 as background value. PCR efficiencies were not estimated using standard curves, but were first calculated for each individual amplification using LinRegPCR software (Ramakers et al, 2003) in the following way. Baseline-subtracted raw fluorescence unit (RFU) values were exported from the Opticon monitor software (version 1.8) to LinRegPCR (version 3.1). In the exponential phase of amplification, a linear relation exists between cycle number and log relative fluorescence. Individual PCR efficiency was estimated from the slope of the linear regression curve (Efficiency $= 10^{slope}$) and served as a quality control on the amplification.

The cycle threshold (Ct) was set at 0.01 RFUs for all assayed PCR plates and Ct values were used in the Q-Gene module to generate mean normalized expression values (MNE) of three pseudo-replicates from a cDNA synthesis (Muller et al, 2002). The mean value of all PCR efficiencies from each specific target (mt, crq, β-actin and 28S) in the first 96-well plate was taken as the set value in the Q-Gene module. Subsequent 96-well plates were checked for deviations from the set PCR efficiencies with LinRegPCR. The Q-gene module also calculates standard errors among the three pseudoreplicates and automatically highlights individuals with a standard error above 20%. These samples were re-evaluated. Table 1 summarizes sequences of primer sets, amplicon lengths and PCR efficiencies of the amplification reactions.

The lysate remaining after RNA extraction (50 μl) was used for DNA extraction by adding 200 μl of SV Lysis buffer (without β-mercaptoethanol) together with 175 μl Nuclei Lysis solution (Promega). DNA from this lysate was bound to a SV96 DNA binding plate and washed two times with 500 μl SV wash buffer. Columns were dried for 5 min under vacuum. DNA was eluted in 50 μl nuclease-free water and stored at −20°C until further use.

PCR of and RFLP analysis of mt promoter region
The following PCR primers were developed to amplify 1.6 kb promoter region based on the sequence provided

Table 1 Real-time PCR primer pairs

Gene	Amplicon length	PCR efficiency	Code	Sequence 5′–3′
mt[a]	102	2.011 ($\pm$0.039)	MT-163F	GGCAAATCGCCCACTTGTT
			MT-265R	CCTTGCAGACACAATCTGGACC
crq[b]	100	2.012 ($\pm$0.043)	SCD1C6-151F	CTTTGTTCACCCGTGGATCCT
			SCD1C6-261R	TGTGCCGCAGTGTCTATCTCA
28S[c]	94	2.027 ($\pm$0.057)	28S1-132F	TACGCCGCATTTGTTCCTCT
			28S1-22	ACCAGGACTCAAACCACAAGGAT
β-Actin[d]	90	2.005 ($\pm$0.050)	ARC4-67F	CCGTAAGGATCTGTATGCCAACA
			ARC4-157R	GGCAGTGATCTCCTTTTGCATC

PCR efficiency: mean value from LinRegPCR calculations of 24 reactions, standard deviations between brackets.
Genbank accession numbers of template sequences for primer design.
[a]AF036345 (Hensbergen et al, 1999).
[b]AY779741.
[c]AF483443 (D'Haese, 2002).
[d]AY779737.

by Janssens *et al* (unpublished data): D1-36F 5′-GGCATTCTAGTTTGAGTCAGTCTACCGG-3′, MT-265R 5′-CCTTGCAGACACAATCTGGACC-3′). PCR products of parents and offspring from Cd exposed families were cut with restriction enzymes *Cla*I, *Ssp*I, *Tau*I and *Taq*I. The restriction enzyme reactions were run on 1.5% agarose gel to identify each of the eight alleles as characterized by Janssens *et al* (unpublished data). Families were subdivided according to mt expression level into four groups: very low (LL: $<15 \times 10^{-4}$ MNE), low (L: 15–25×10^{-4} MNE), medium (M: 25–40×10^{-4} MNE), and High (H: $>40 \times 10^{-4}$ MNE). Alleles were counted for each expression group.

Parent–offspring regression analysis

RNAs were pooled for the two parents and among four offspring to generate a mean value for *mt* and *crq* expression. In order to obtain a normal distribution of the data, the mean relative expression values were −log transformed and subjected to midparent–offspring regression analyses (Sokal and Rohlf, 1981) to estimate a narrow-sense h^2. The circumstances under experimental control were designed to minimize violations of the assumptions required to correctly interpret the genetic model of offspring–midparent regression analysis. A power analysis indicated that the appropriate sample size would be 50 families. The power of this design was 0.9, that is, there is a 90% chance of rejecting the null hypothesis given that the alternate hypothesis is true. Owing to mortality of parents and degradation of some RNA samples the number of families was reduced slightly to 47 in the nonexposed families and 46 in the 0.5 μmol Cd/g dry food treatment.

Statistics

A Kolmogorov–Smirnov test was performed to test if the −logMNEs of mt and crq expression show a normal distribution. A Levene's test was performed to test homogeneity of variance among parents and offspring. Regression analysis and the above-mentioned tests were performed using software package SPSS version 11.5. The significance of the h^2 estimates was evaluated by calculating 95% confidence intervals according to Falconer (1989).

A goodness of fit test (*G*-test) was performed on promoter allele frequencies to test whether the different expression groups have homogenous *mt* promoter allele frequencies (Sokal and Rohlf, 1981).

Results

PCR efficiency estimation

PCR efficiencies of 28S rRNA, *β-actin* and *mt* mRNA were 2.027 (standard deviation 0.057), 2.005 (standard deviation 0.050) and 2.011 (standard deviation 0.039) based on 24 amplifications for each gene in the same run (Table 1). These values show that primer design and template length resulted in a doubling of templates every cycle, which corresponds to maximum amplification efficiency. Subsequent PCR reactions were always tested for amplification efficiency. Amplifications in which the efficiency was less than the set values minus two times the standard deviation were reanalyzed.
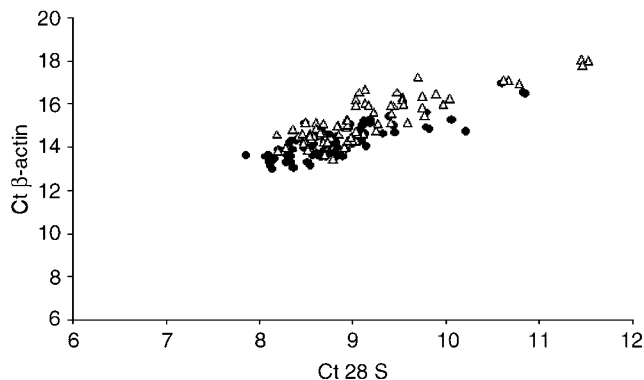

Figure 2 Correlation of Ct values (Ct) between *β-actin* (*y*-axis) and 28S (*x*-axis) for 150 RNA samples (75 random RNAs from nontreated animals: ●, 75 random RNAs from Cd exposed animals: Δ).

Table 2 Levene's test for homogeneity of variance between parents and offspring

Treatment	Internal control	Target gene	Levene statistic	df1	df2	P
−Cd	28S	*mt*	0.12	1	92	0.73
		crq	6.79	1	92	0.01[a]
	β-Actin	*mt*	1.25	1	92	0.27
		crq	7.77	1	92	0.01[a]
+Cd	28S	*mt*	0.01	1	90	0.93
		crq	5.02	1	90	0.03[a]
	β-Actin	*mt*	0.32	1	90	0.57
		crq	3.47	1	90	0.07

df = degree of freedom, *P* = significance value.
[a]Nonhomogenous distribution ($P < 0.05$).

In two samples, the *mt* amplification Ct values were observed earlier than the internal controls 28S and *β-actin*. Re-inspections of these samples on a gel, showed that either no clear ribosomal bands were visible or the lane was empty. We concluded that these samples had been accidentally degraded by RNases, and they were excluded from further analysis.

To test whether *β-actin* as internal control for normalization might have been influenced by Cd treatment, a linear regression was performed on Ct values of 28S and *β-actin* of 150 randomly chosen cDNA samples (75 Cd− samples, 75 Cd + samples). Figure 2 shows the regression plot (28S on the *x*-axis, *β-actin* on the *y*-axis). No effect is observed from the Cd exposure on the *β-actin* signal. Furthermore, from the slope of 1.1 ($R^2 = 0.85$) it can be concluded that *β-actin* slightly underestimates the total amount of RNA present in a sample, on the assumption that the 28S expression is the better estimator of total RNA input.

Parent–offspring analysis

Mean normalized *mt* and *crq* expression values (MNE) were estimated for 47 nonexposed families and 46 families exposed to 0.5 μmol Cd/g dry food using 28S or *β-actin* as internal controls. The variances of gene expression (−logMNE values) from parents and offspring were subjected to Levene's homogeneity test (results in Table 2). The variances for *mt* expression from

the unexposed and 0.5 μmol Cd exposed parents were not significantly different from their offspring ($P > 0.05$). However, variances between parents and offspring for *crq* gene expression were significantly different, except for *β-actin* normalized gene expression in 0.5 μmol Cd exposed families ($P = 0.07$). These observations suggest that results from the *crq* linear regression analysis need to be interpreted cautiously.

Figure 3 shows the graphs of the linear regression analysis with midparent *mt/crq* expression values as the independent character state on the *x*-axis and the mean offspring *mt/crq* expression as dependent value on the *y*-axis. Metallothionein gene expression is induced by Cd treatment, shown as a shift towards a −logMNE of zero in Figure 3. The mean value of all parent and offspring −logMNE decreases from 4.60 to 3.65 when normalized to 28S and from 2.98 to 1.94 when normalized to *β-actin*. Croquemort expression was not induced by 0.5 μmol Cd-exposure: no shifts of −logMNEs were observed when nonexposed families were compared to Cd-exposed families. Additive genetic variation for *mt* gene expression is demonstrated for 0.5 μmol-exposed families. An
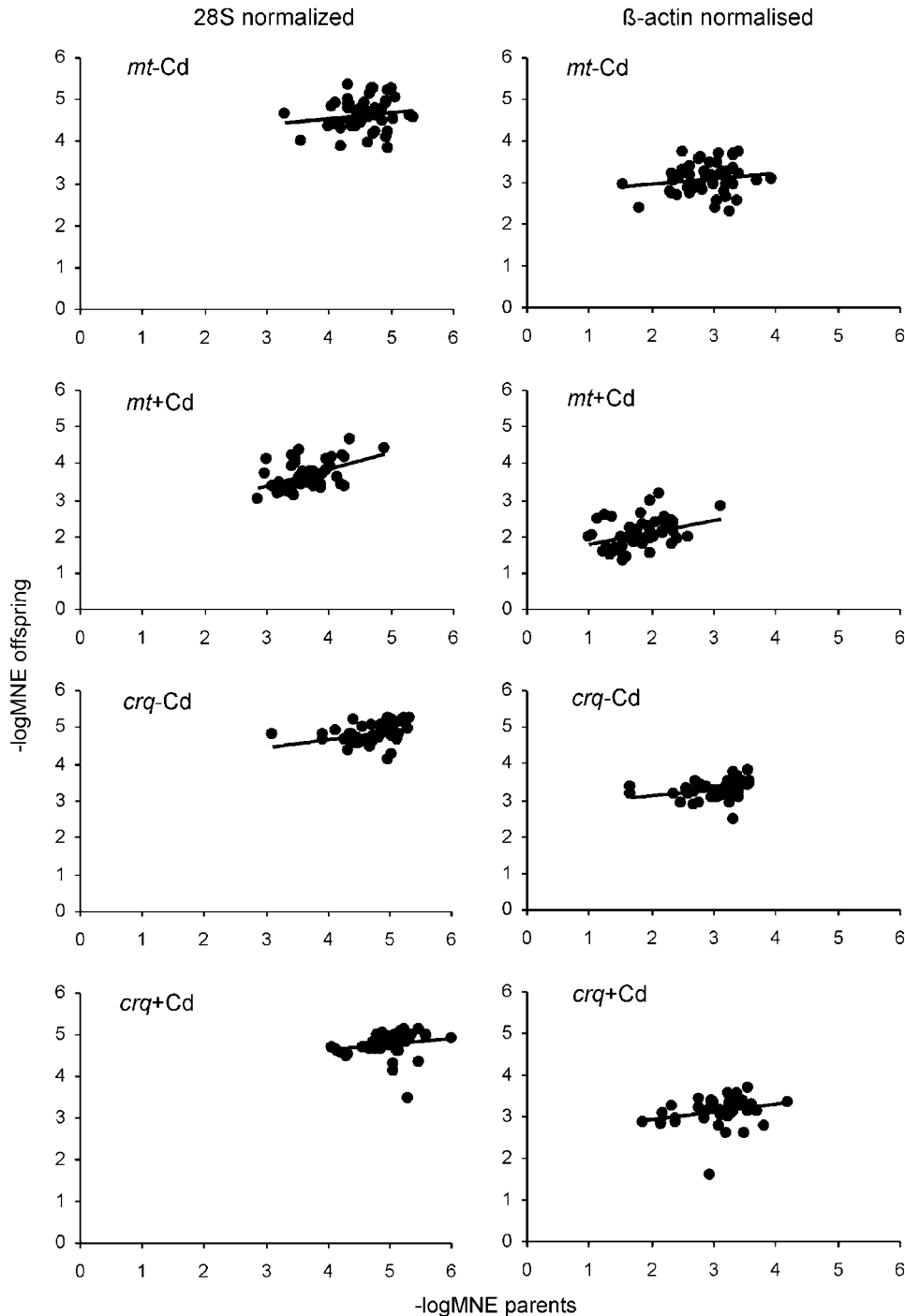

Figure 3 Parent–offspring regression of *metallothionein* (*mt*) and *croquemort* (*crq*) expression. −Cd no cadmium exposed food, + Cd 0.5 μmol cadmium/g dry food.

Table 3 95% confidence intervals for heritability estimates (h^2) compared to 0 for metallothionein and *crq* gene expression from the reference site Roggebotzand

Exposure	Internal control	Target gene	h^2	95% confidence interval for h^2 t = 0.43	
				Lower	Upper
−Cd	28S	*mt*	0.13	−0.16	0.42
		crq	0.22	−0.07	0.51
	β-Actin	*mt*	0.14	−0.15	0.43
		crq	0.17	−0.12	0.46
+Cd	28S	*mt*	0.46*	0.16	0.76
		crq	0.13	−0.17	0.43
	β-Actin	*mt*	0.36*	0.07	0.67
		crq	0.17	−0.13	0.47

t = intraclass correlation coefficient.
*Significant heritability.

Table 4 Allele frequencies (absolute numbers) of the 1.6 kb proximal *mt* promoter

Allele	*mt* Expression group			
	LL	L	M	H
A	10	11	6	5
B[a]	0	1	1	5
C1	2	0	2	0
C2	2	0	2	0
C3	4	6	3	4
D1[a]	2	2	4	6
D2[a]	0	2	2	4
F	5	3	1	2

Families were subdivided according to *mt* expression level: very low (LL: $<15 \times 10^{-4}$ MNE), low (L: $15-25 \times 10^{-4}$ MNE), medium (M: $25-40 \times 10^{-4}$ MNE), and High (H: $>40 \times 10^{-4}$ MNE).
[a]Increased in frequency in families with high *mt* expression (H).

h^2 of 0.46 was estimated with 28S normalized *mt* expression and 0.36 with β-actin normalized data (Table 3). Furthermore, an h^2 of 0.22 was found for *crq* expression in nonexposed families.

We tested the hypothesis that the h^2 is significantly different from zero by calculating 95% confidence intervals according to Falconer (1989). To correct the standard error of h^2 for phenotypic correlation between members of each family the intraclass correlation coefficient (t) was estimated from a full sib analysis based on a previous experiment from 45 crosses exposed to 1 μmol Cd/gram dry food (data not shown). This experiment was setup similarly as described here except that families were raised in a single jar instead of being separated in four jars per family. The full sib analysis yielded a t of 0.43 (De Boer and Roelofs, unpublished data) for *mt* gene expression, which was used to correct for correlation within families in this study. This t-value should be considered an overestimate, because the common environment effect in our case was minimized by raising offspring in separate jars. Table 3 shows the 95% confidence intervals ($h^2 \pm 1.96 \times$ standard error, SE) for each h^2 estimate for $t = 0.43$. For $t = 0.43$ a SE of 0.15 was calculated for unexposed families ($n = 47$) and 0.16 for exposed families ($n = 46$). Both 28S and β-actin normalized h^2 estimates for Cd-induced *mt* gene expression are significant at $t = 0.43$. Linear regression of 28S normalized *crq* expression in nonexposed families ($h^2 = 0.22$) shows a significant slope of 0.22 ($P < 0.01$). However, when corrected for phenotypic correlation within families, this h^2 does not significantly deviate from zero (the lower 95% confidence limit is −0.07, Table 3). Finally, we tested significance of Cd-induced *mt* expression in the hypothetical case that the phenotypic correlation of the offspring is maximal ($t = 1$), but this did not change our conclusion that the h^2 estimate for Cd-induced *mt* expression is significantly different from zero.

RFLP analysis of promoter alleles

Janssens *et al* (unpublished data) identified eight *mt* promoter alleles designated as A, B, C1, C2, C3, D1, D2 and F. These alleles differ in the position and number of metal responsive elements and other potential transcription factor binding sites as well as single nucleotide polymorphisms and indels. Each allele could be identi-

fied by the digestion patterns of *Cla*I, *Ssp*I, *Tau*I and *Taq*I. The families were arbitrarily subdivided in the four groups based on *mt* expression level. Each group contained between 11 and 12 families. Table 4 shows the promoter allele frequencies within each group. Allele A is the most abundant one occurring in high frequency in all groups, which is in concordance with the data of Janssens *et al* (unpublished). To test whether the expression groups show homogenous allele frequency distribution a G-test was performed. Since the calculated G_H value of 34.84 exceeded the critical value of $X^2_{0.05[21]} = 32.67$, there is evidence that the promoter allele frequencies are not homogenous among the expression groups. Promoter alleles B, D1 and D2 have a high frequency in the group of high *mt* expression families (H) as compared to the allele frequencies in the other groups (M, L and LL).

Discussion

The present study demonstrated significant additive genetic variation for Cd-induced *mt* gene expression in a collembolan species. This implies that the reference population under study can respond to selection driven by heavy-metal exposure. Families with high *mt* expression differ significantly in *mt* promoter allele distribution, which suggests that the additive genetic variation may partly be explained by structural changes in *cis*-regulation. Metallothionein gene expression under nonexposed conditions does not seem to show a significant additive genetic component, which would imply that selection can act on the enhancement of expression, not on the constitutive level of expression.

The use of a quantitative genetics model applied to the expression profile of an individual gene is justified by the argument that it takes the action of many genes to determine the expression of any single gene. The transcription initiation complex contains more than eight subunits all encoded by separate genes. Furthermore, RNA polymerase II is required for RNA synthesis, and finally, specific transcription factors modulate transcription. For instance, *mt* expression is coactivated by metal transcription factor-1, which initiates a response to heavy metals (Zhang *et al*, 2001). It will probably take a minimum of 10 gene products to determine the expression profile of a gene, which may easily cause a normal

distribution of gene expression phenotypes. A Kolmogorov–Smirnov test was used to test if our data differed from a normal distribution (Sokal and Rohlf, 1981). Neither the 28S-normalized nor the β-actin-normalized mt expression were significantly different from a normal distribution (28S normalized mt $D = 0.0941$, $P > 0.2$; β-actin normalized mt $D = 0.1288$, $P > 0.2$) implying that gene expression is a quantitative trait. Metallothionein expression variation seems to be gradual and not discrete, which suggests that no single gene has a major effect on mt gene regulation. At present, it is difficult to verify which genes are involved in mt gene regulation and to what extent they influence mt expression. Genome-wide transcription profiling of the Cd response on $O. \, cincta$ cDNA microarrays may elucidate to which extent the candidate genes are involved in mt gene regulation.

The Cd concentration used in this study is comparable to that found at contaminated field sites. Van Straalen et al (1987) measured 0.525 μmole Cd/g dry wt in litter samples of the Plombières mining site and 1.63 μmole Cd/g dry wt at the Stolberg mining site, which was recently confirmed for the Plombières mining site by Lock, Janssens, and Janssen (2003). Sterenborg and Roelofs (2003) demonstrated that metal tolerant springtails from the Plombières mining site showed a significant Cd enhanced expression of mt mRNA, up to four-fold that of the reference population. Combined with the data presented here, this observation suggests that the Cd elevated mt gene expression assayed in the Plombières animals is due to adaptive evolution of transcriptional mt regulation to Cd contamination encountered in the field.

Previously, selection was inferred for populations of $O. \, cincta$ living in metal contaminated areas based on Cd excretion efficiency (EE), growth rate (RGR), age at maturity and maintainance of growth during Cd exposure (Van Straalen et al, 1987; Posthuma, 1990; Posthuma et al, 1992). Genetic variation was indeed determined for EE, RGR and another characteristic associated with Cd tolerance, moulting frequency (Posthuma et al, 1993). The h^2 estimates for EE are comparable to those estimated for mt gene expression in the Roggebotzand population, and the question can be raised, does this imply that mt expression is mechanistically related to excretion efficiency? Excretion efficiency is a function of Cd retention in the gut epithelium. This tissue is renewed at each moulting cycle, a process including apoptosis and excretion of the cells. As a result, the excreted gut pellet is loaded with Cd. This makes it very likely that up-regulated metallothionein protein enhances the capacity for Cd excretion through renewal of the gut epithelium at moult.

As Cd excretion in springtails is tightly connected to periodical apoptosis of gut epithelial cells during moulting, we also investigated a gene associated with apoptosis. $Croquemort$ gene expression did not respond to 0.5 μM Cd-exposure, even though this expressed sequence tag was isolated from a differential screening for genes responding to 1 μmole Cd/g dry food. Additional real-time RT-PCR data showed an increase of crq expression upon exposure to higher concentrations (≥ 1 μmole Cd/g dry food, Roelofs, unpublished data). The observation that crq expression is not influenced at 0.5 μmol Cd/g dry food treatment may either suggest that crq gene expression is less sensitive to Cd exposure or that its response to Cd is the result of the alternative oxidative stress response that can be triggered by Cd as well (Haq, Mahoney, and Koropatnick, 2003). This pathway may only become transcriptionally significant when toxic levels of Cd treatment start inducing reactive oxygen species.

There are only few examples demonstrating evolution through transcriptional regulation occurring in the wild. For instance, $in \, vivo$ transcription rates of the $lactate$ $dehydrogenase$-B gene (Ldh-B) were two-fold greater in livers from Northern $versus$ Southern populations of the common killifish $Fundulus \, heteroclitus$. This resulted in a two-fold greater LDH-B$_4$ enzyme concentration, which compensates for the lower Northern temperatures (Crawford and Powers, 1992). Analysis of the Ldh-B proximal promoter sequence revealed that sequence variation in this region is biologically important, and can be explained by directional selection on the Ldh-B proximal promoter. In Northern populations of $F. \, heteroclitus$ this selection appears to have led to an increase in the enzyme expression (Crawford, Segal, and Barnett, 1999). These authors elegantly showed, in footprinting assays, that fixed mutations in the SP1 transcription-binding site of Northern populations caused a more efficient binding of this transcription factor in the Ldh-B proximal promoter. This results both $in \, vitro$ and $in \, vivo$ in overexpression of Ldh-B in Northern populations. They argue that adaptive change in gene expression has a greater probability of occurring than change in enzyme structure altering catalytic activity, because mutations in promoters are not constrained by codon usage or amino-acid replacements in the tertiary/quaternary structure of the protein. Recently, Janssens et al (unpublished data) isolated 1.6 kb of the $O. \, cincta$ proximal mt promoter and showed extensive natural variation in this region; eight alleles were identified in Benelux populations. Their preliminary data suggest that genetic variation in mt expression may be due to changes in cis-regulatory elements: variation in the amount and physical spacing of several transcription factor binding sites is associated with variation in Cd enhanced mt expression. For instance, the C alleles (Table 4) lack 1–2 metal responsive elements as compared to the other alleles and are 100 times less inducible in mt promoter luciferase reporter assays in the $D. \, melanogaster$ S2 cell line (data not shown). We observed that the total number of C alleles (C1, C2 and C3) is low within the high mt expression group (group H in Table 4), when compared to the other expression groups. In particular, the frequencies of alleles B, D1 and D2 are increased in the H group. It can be hypothesized that frequencies of these alleles are elevated in populations from contaminated field sites and, if selection is particularly strong, they may not be in Hardy–Weinberg equilibrium. Currently, we are screening populations in contaminated as well as clean sites to test this hypothesis.

Acknowledgements

References

Crawford DL, Powers DA (1992). Evolutionary adaptation to different thermal environments via transcriptional regulation. *Mol Biol Evol* **9**: 806–813.

Crawford DL, Segal JA, Barnett JL (1999). Evolutionary analysis of TATA-less proximal promoter function. *Mol Biol Evol* **16**: 194–207.

D'Haese CA (2002). Were the first springtails semi-aquatic? A phylogenetic approach by means of 28S rDNA and optimization alignment. *Proc R Soc Ser B* **269**: 1143–1151.

Falconer DS (1989). *Introduction to Quantitative Genetics*. Longman Scientific & Technical: New York.

Franc NC, Dimarcq JL, Lagueux M, Hoffmann J, Ezekowitz RAB (1996). Croquemort, a novel *Drosophila* hemocyte/macrophage receptor that recognizes apoptotic cells. *Immunity* **4**: 431–443.

Franc NC, Heitzler P, Ezekowitz RAB, White K (1999). Requirement for croquemort in phagocytosis of apoptotic cells in *Drosophila*. *Science* **284**: 1991–1994.

Haq F, Mahoney M, Koropatnick J (2003). Signaling events for metallothionein induction. *Mutat Res* **533**: 211–226.

Hensbergen PJ, Donker MH, Van Velzen MJM, Roelofs D, Van der Schors RC, Hunziker PE *et al* (1999). Primary structure of a cadmium-induced metallothionein from the insect *Orchesella cincta* (Collembola). *Eur J Biochem* **259**: 197–203.

Jin W, Riley RM, Wolfinger RD, White KP, Passador-Gurgel G, Gibson G (2001). The contributions of sex, genotype and age to transcriptional variance in *Drosophila melanogaster*. *Nat Genet* **29**: 389–395.

Lange BW, Langley CH, Stephan W (1990). Molecular evolution of *Drosophila* metallothionein genes. *Genetics* **126**: 921–932.

Lock K, Janssens F, Janssen CR (2003). Effects of metal contamination on the ectivity and diversity of springtails in an ancient Pb-Zn mining area in Plombières, Belgium. *Eur J Soil Biol* **39**: 25–29.

Maroni G, Wise J, Young JE, Otto E (1987). Metallothionein gene duplications and metal tolerance in natural populations of *Drosophila melanogaster*. *Genetics* **1178**: 739–744.

Muller PY, Janovjak H, Miserez AR, Dobbie Z (2002). Processing of gene expression data generated by quantitative real-time RT PCR (vol 32, pg 1378, 2002). *Biotechniques* **33**: 514.

Otto E, Young JE, Maroni G (1986). Structure and expression of a tandem duplication of the *Drosophila* metallothionein gene. *Proc Natl Acad Sci USA* **83**: 6025–6029.

Posthuma L (1990). Genetic differentiation between populations of *Orchesella cincta* (Collembola) from heavy metal contaminated sites. *J Appl Ecol* **27**: 609–622.

Posthuma L, Hogervorst RF, Joosse ENG, Van Straalen NM (1993). Genetic variation and covariation for characteristics associated with cadmium tolerance in natural populations of the springtail *Orchesella cincta*. *Evolution* **47**: 619–631.

Posthuma L, Hogervorst RF, Van Straalen NM (1992). Adaptation to soil pollution by cadmium excretion in natural populations of *Orchesella cincta* (L.) (Collembola). *Arch Environ Cont Tox* **22**: 146–156.

Purugganan MD (2000). The molecular population genetics of regulatory genes. *Mol Ecol* **9**: 1451–1461.

Ramakers C, Ruijter JM, Deprez RHL, Moorman AFM (2003). Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neurosci Let* **339**: 62–66.

Saydam N, Adams TK, Steiner F, Schaffner W, Freedman JH (2002). Regulation of metallothionein transcription by the metal-responsive transcription factor MTF-1 – Identification of signal transduction cascades that control metal-inducible transcription. *J Biol Chem* **277**: 20438–20445.

Sokal RR, Rohlf FJ (1981). *Biometry*. W.H. Freeman and company: New York.

Sterenborg I, Roelofs D (2003). Field-selected cadmium tolerance in the springtail *Orchesella cincta* is correlated with increased metallothionein mRNA expression. *Insect Biochem Mol Biol* **33**: 741–747.

Stürzenbaum SR, Winters C, Galay M, Morgan AJ, Kille P (2001). Metal ion trafficking in earthworms – Identification of a cadmium-specific metallothionein. *J Biol Chem* **276**: 34013–34018.

Van Straalen NM, Burghouts TBA, Doornhof MJ, Groot GM, Janssen MPM, Joosse ENG *et al* (1987). Efficiency of lead and cadmium excretion in populations of *Orchesella cincta* (Collembola) from various contaminated forest soils. *J Appl Ecol* **24**: 953–968.

Wray GA, Hahn MW, Abouheif E, Balhoff JP, Pizer M, Rockman MV *et al* (2003). The evolution of transcriptional regulation in eukaryotes. *Mol Biol Evol* **20**: 1377–1419.

Zhang B, Egli D, Georgiev O, Schaffner W (2001). The *Drosophila* homolog of mammalian zinc finger factor MTF-1 activates transcription in response to heavy metals. *Mol Cell Biol* **21**: 4505–4514.