BMC Evolutionary Biology

Research article

Recombinational micro-evolution of functionally different metallothionein promoter alleles from *Orchesella cincta*

Thierry KS Janssens*[1], Janine Mariën[1], Peter Cenijn[2], J Legler[2], Nico M van Straalen[1] and Dick Roelofs[1]

Address: [1]Vrije Universiteit, Institute of Ecological Sciences, Department of Animal Ecology, De Boelelaan 1085, 1081 HV Amsterdam, the Netherlands and [2]Vrije Universiteit Amsterdam, Institute for Environmental Studies (IVM), de Boelelaan 1085, 1081 HV Amsterdam, the Netherlands

Email: Thierry KS Janssens* - thierry.janssens@ecology.falw.vu.nl; Janine Mariën - janine.marien@ecology.falw.vu.nl; Peter Cenijn - peter.cenijn@ivm.vu.nl; J Legler - juliette.legler@ivm.vu.nl; Nico M van Straalen - nico.van.straalen@ecology.falw.vu.nl; Dick Roelofs - dick.roelofs@ecology.falw.vu.nl

* Corresponding author

Abstract

Background: Metallothionein (*mt*) transcription is elevated in heavy metal tolerant field populations of *Orchesella cincta* (Collembola). This suggests that natural selection acts on transcriptional regulation of *mt* in springtails at sites where cadmium (Cd) levels in soil reach toxic values This study investigates the nature and the evolutionary origin of polymorphisms in the metallothionein promoter (*pmt*) and their functional significance for *mt* expression.

Results: We sequenced approximately 1600 bp upstream the *mt* coding region by genome walking. Nine *pmt* alleles were discovered in NW-European populations. They differ in the number of some indels, consensus transcription factor binding sites and core promoter elements. Extensive recombination events between some of the alleles can be inferred from the alignment. A deviation from neutral expectations was detected in a cadmium tolerant population, pointing towards balancing selection on some promoter stretches. Luciferase constructs were made from the most abundant alleles, and responses to Cd, paraquat (oxidative stress inducer) and moulting hormone were studied in cell lines. By using paraquat we were able to dissect the effect of oxidative stress from the Cd specific effect, and extensive differences in *mt* induction levels between these two stressors were observed.

Conclusion: The *pmt* alleles evolved by a number of recombination events, and exhibited differential inducibities by Cd, paraquat and molting hormone. In a tolerant population from a metal contaminated site, promoter allele frequencies differed significantly from a reference site and nucleotide polymorphisms in some promoter stretches deviated from neutral expectations, revealing a signature of balancing selection. Our results suggest that the structural differences in the *Orchesella cincta* metallothionein promoter alleles contribute to the metallothionein -over-expresser phenotype in cadmium tolerant populations.

Background

Transcriptional regulation plays an important role in the evolution of many phenotypes, especially when the phenotype correlates with expression level of a particular key gene. Transcriptional regulation is mostly controlled at the level of transcriptional initiation, i.e. the recruitment of transcription factors which determine the stability of the RNA polymerase II holoenzyme complex and hence the frequency of transcription initiation. Due to the modularity of transcription factor binding site clusters and the lack of reading frame constraint, promoters are more evolvable than coding regions [1,2]. A few point mutations in a *cis*-regulatory region can already confer functionally different phenotypes, even when the pattern does not deviate from neutral expectations[3,4]. Some examples of variation in regulatory loci conferring adaptive phenotypes are: the *LdhB* promoter of *Fundulus heteroclitus* [4-7], the *hsp70Ba* promoter of *Drosophila melanogaster* [8,9] the chalcone synthase promoter of *Arabidopsis thaliana* [3] and the *Cyp6g1* promoter of *Drosophila melanogaster* [10]. In this paper we focus on transcriptional regulation of the *Orchesella cincta* metallothionein gene (*mt*), which is assumed to be involved in heavy metal tolerance.

Metallothioneins are low molecular weight metal-binding proteins with a high content of conserved cysteines within certain phylogenetic lineages and a lack of aromatic amino acids and histidine [11,12], although some invertebrate metallothioneins deviate from this pattern [13-15]. Metallothioneins are involved in essential metal homeostasis, metal detoxification, free radical scavenging, cell proliferation and apoptosis processes [16]. They are induced by several chemical and physical stresses [17], including free metal ions, altered redox status, oxidative stress and heat shock. The *Orchesella cincta* metallothionein (MT) [18] has a molecular weight of 7 kDa and consists of 77 amino acids. Almost all of the body burden of Cd is located in the gut epithelium and a major part of it is bound to MT [19,20]. It is suggested that the excretion of Cd from the animal occurs by a molting cycle regulated apoptotic process, by which the Cd-loaded midgut epithelium is shed [21,22]. Higher constitutive [23] and cadmium inducible [24]*mt* mRNA levels have been observed in populations from heavy metal contaminated sites, compared to populations from reference sites. Parent-offspring comparisons showed that Cd-induced expression of *mt* (h^2 = 0.48) is a heritable trait. Differences between expression level classes were linked to RFLP patterns of the *pmt* locus [25]. Although certain alleles of the *mt* coding sequence are linked to heavy metal pollution of the soil [26], it is rather unlikely that the heavy metal tolerance can be attributed to a single gene [27,28].

Inherited heavy metal tolerance has been associated with duplication events [29] and polymorphisms in the metallothionein coding sequence [26,30] The regulation of MT biosynthesis is mainly transcriptional and depends for the most part on *cis*-acting regulatory elements, such as the metal responsive element (MRE) which binds metal responsive transcription factor-1 (MTF-1), a Zn-finger protein, and the anti-oxidant responsive element (ARE) which recruits the nuclear erythroid derived related factor-2 (Nrf-2), a protein of the b-zip leucine zipper family [16,31]. However, this situation can not be generalized for all invertebrate phyla [32,33] Therefore, the induction of metallothionein can be considered as a concerted action of general, metal-specific and oxidative stress specific transcription factors.

In the past decades several studies on metallothionein promoters of invertebrates have been performed [34-42], although most mechanistic studies have been done on vertebrate model organisms, reviewed by [16,31,43]. One way to study the functionality of promoters is to fuse them to a quantitative reporter gene and analyze the induction *in vitro* in a host cell line. The comparison of allelic polymorphism in promoters by reporter assays has mainly been applied in medical biology, e.g. [44,45]. Only few studies in evolutionary ecology have compared alleles in this functional approach [5-7,46] and only one study compared the metal inducibility of metallothionein promoter alleles [47].

In the present study natural occurring allelic variation of the *Orchesella cincta* metallothionein promoter (*pmt* locus) is described. Following discovery of extensive variation in promoter sequence we formulated the following research question: "Are the *pmt* alleles observed in natural populations differentially induced by Cd, paraquat and 20-hydroxyecdysone (molting hormone) and can their induction be related to their different architecture?" Luciferase constructs were made and tested in an arthropod cell line for dose-dependent inducibilities. Paraquat is included in the experiment, because it generates reactive oxygen species in the electron transport chain [48] causing oxidative stress. This approach allows us to discriminate between the effects of Cd and oxidative stress separately. Finally, we tested if the *pmt* allele frequency distribution, based on nucleotide diversity measures, deviated from neutrality in a tolerant and sensitive population.

Results

General architecture of the pmt locus

Nine different alleles were identified in an alignment of 32 1500 bp promoter sequences (see additional files 1, 2, 3, 4, 5, 6, 7, 8, 9, 10). The consensus sequences of the respective alleles and the *O. villosa* were submitted to Genbank (DQ523588 to DQ523596, DQ641512, DQ523587

and EF106974). The general architecture of the nine promoter alleles is shown in Fig. 1, which also indicates positions of putative core promoter elements and transcription factor binding sites [16,31,49-51]. The number of the putative TFBS is summarized in Table 1 for each allele. The basal promoter consists of an initiator (Inr) consensus with an overlap of a 20-hydroxyecdysone responsive element (HERE). The *pmt*A allele contains two extra putative initiators. All the alleles, except *pmt*C, have a downstream promoter element (DPE) consensus downstream of their Inr. All *pmt* alleles contain MREs, which are all orientated in the sense direction. The proximal promoter, about 300 bp 5' from the Inr, contains most of the MREs. The alleles *pmt*A1, *pmt*A2, *pmt*B, *pmt*D1, *pmt*D2 and *pmt*BAL have five MREs in this region, named MRE-a to

MRE-e. A number of indels in this MRE-rich region make this region variable. The MRE-a was apparently lost from the *pmt*C allele by a 13 bp deletion. A 19 bp deletion 5' of the MRE-b, relative to *pmt*C, *pmt*D1, *pmt*D2 and *pmt*BAL, characterizes *pmt*A, *pmt*B, *pmt*E and *pmt*F. This deletion affects the spatial position of the MREs. The *pmt*C and *pmt*F alleles have a point mutation which disrupts the consensus MRE sequence of MRE-d and MRE-e respectively. The *pmt*D1 and *pmt*D2 alleles share a HERE between MRE-a and MRE-b, by one point mutation. An AP-1 binding site consensus was only retrieved in the forward PCR primer D1-36F, and is not further discussed. This primer was developed after genome walking resulting in a clone that was apparently *pmt*A1. All alleles, except *pmt*C share a DNA replication-related element (DRE) [50].
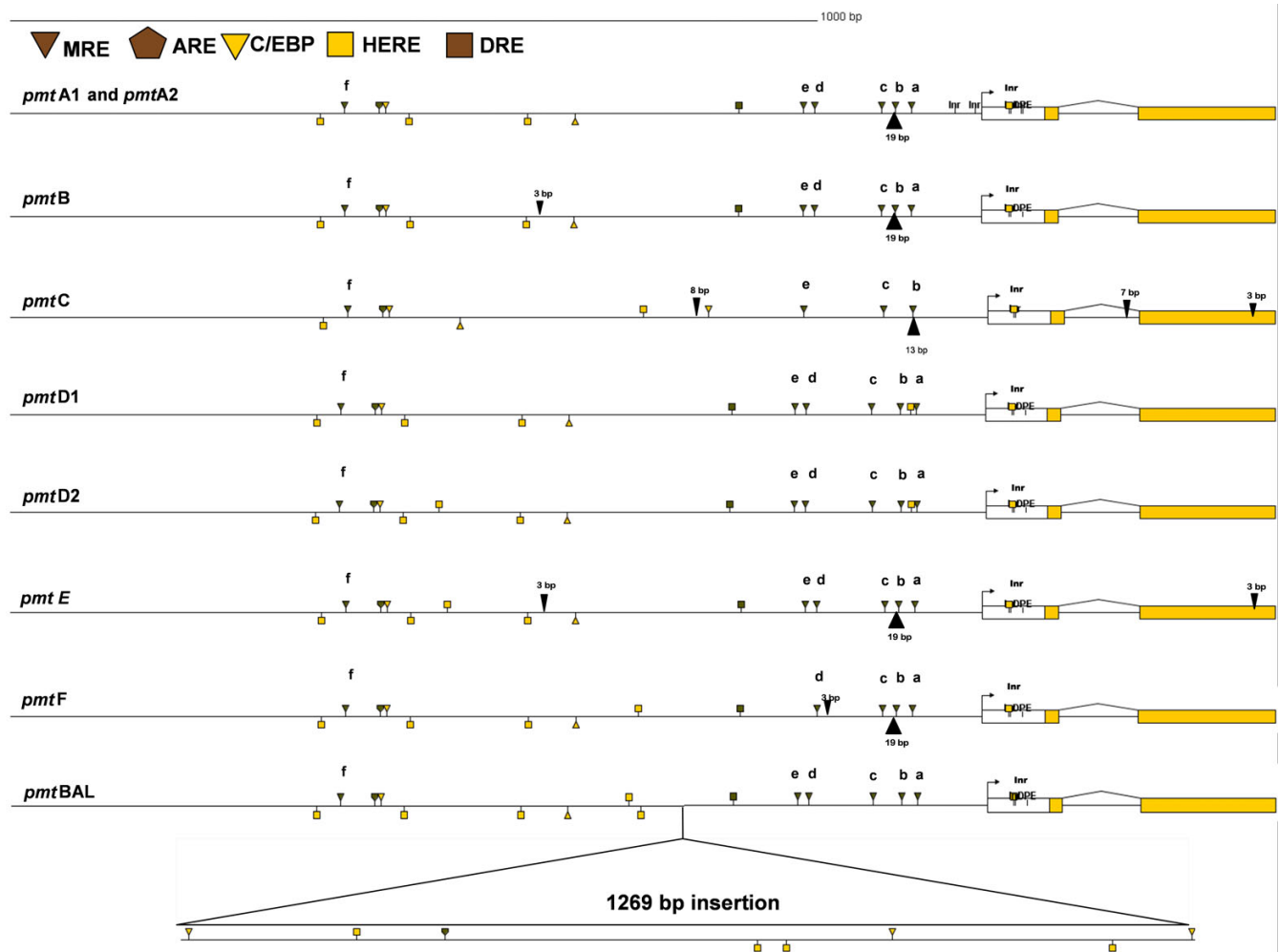
Figure 1
Architecture of the nine respective metallothionein promoter alleles (pmt). The respective putative transcription factor binding sites are represented as in the legend. Indels are indicated with black triangles. MRE, metal responsive element; ARE, anti-oxidant responsive element; DRE, DNA replication-related element; HERE, 20-hydroxyecdysone responsive element; Inr, initiator; DPE, downstream promoter element; C/EBP, CCAAT enhancer binding protein. The full sequence alignment is given in Additional File , 10.

Table 1: Summary of the occurrence of the respective putative transcription factor binding sites. a: The number of MREs is given as one in the enhancer region (in the vicinity of the ARE) plus the number in the proximal promoter. b: The number of HEREs is represented as the number in the region upstream of the initiator plus the one overlapping the initiator.

	Inr	DPE	MRE [a]	ARE	C/EBP	DRE	HERE [b]
Consensus	TCAKTY [49, 84]	RGWYV [84]	TGCRCNC [31]	TGA CNNNGC [31]	CCAAT [31]	TATCGATA [50]	KNTCANTNNMM [51]
*pmt*A	3	1	1+5	1	2	1	3+1
*pmt*B	1	1	1+5	1	2	1	3+1
*pmt*C	1	0	1+3	1	3	1	2+1
*pmt*D1	1	1	1+5	1	2	1	4+1
*pmt*D2	1	1	1+5	1	2	1	5+1
*pmt*E	1	1	1+5	1	2	1	4+1
*pmt*F	1	1	1+4	1	2	1	4+1
*pmt*BAL	1	1	1+5	2	5	1	9+1

A putative enhancer, ± 850 bp upstream from the Inr (when the 1269 bp insertion of pmtBAL is not taken into account), contains another MRE (MRE-f) and an anti-oxidant responsive element (ARE). Within this enhancer and towards the proximal promoter a number of HERES and C/EBP (CCAAT enhancer binding protein binding site) binding sites are found scattered, which differ in number and position between the respective alleles. The 1269 bp indel in the pmtBAL allele is delineated at its edges (relative to the other alleles in the alignment) by two C/EBP binding sites. Furthermore this insertion contains 1 ARE, 3 HEREs and another C/EBP binding site.

Similarities among* pmt *alleles and recombinant analyses
The phi test for recombination in the Splitstree4 program [52] found statistically significant evidence for recombination (p = 1.09 × 10^{-13}). A bootstrap confidence network [52,53] based on a split decomposition analysis was developed representing the inferred recombination events in the *pmt* locus (Fig. 2). Split decomposition analysis addresses the problem of conflicting phylogenetic signals due to recombination which is not necessarily a branching or tree-like process. Parallel edges in the network represent evolutionary lineages of conflicting bifurcating trees. The parallel edges are presented in different colors in order to relate the events to results of the analysis below. The 1269 bp indel from the *pmt*BAL allele was omitted from these analyses. The evolution of the apparently ancestral alleles, *pmt*BAL, *pmt*C and *pmt*F, can be interpreted in a bifurcating pattern, whereas the more recent alleles *pmt*E, *pmt*D2, *pmt*D1, *pmt*B, pmtA2 and *pmt*A1 are consistent with a reticulate origin. The recombination analysis is presented in Fig. 3. Data below a bootscan threshold of 70% were omitted from the graphs. Breakpoints with their respective p-values from the Recco analysis, are plotted in the same graphs as a matter of convenience.

The *pmt*A1 and *pmt*A2 alleles consists of a ± 500 bp upstream block shared with *pmt*D1 and a ± 400 bp block shared with *pmt*B confirmed by very high bootstrap values. In between a slight similarity with *pmt*E was found. Recco confirmed this bootscan analysis and found a recombination breakpoint between the two dominant blocks. Allele *pmt*B contains a ± 150 bp upstream block related to *pmt*C and the ± 400 bp downstream block shared with *pmt*A as mentioned before. Both regions had recombination breakpoints at their respective 3' and 5' edges. The central part of the *pmt*B allele shares limited similarity with *pmt*D2. The *pmt*C allele has the shared ± 150 bp block with *pmt*B and a block shared with *pmt*E at the outer 3' end of the locus (the second exon of the *mt* gene), both confirmed by the Recco method, although not visible in the reticulate network. In between the two blocks bootscan analysis identified similarity with *pmt*-BAL. Numerous recombination breakpoints were detected in this region, suggesting that this allele contains a recombination hotspot. The *pmt*D1 situation is very clearcut. This allele contains the ± 500 bp block shared with *pmt*A and a ± 400 bp 3' block shared with *pmt*D2, confirmed by the Recco method. Both regions have numerous putative recombination breakpoints at their edges. The *pmt*D2 allele shares a ± 500 bp upstream block with *pmt*E and the downstream block with *pmt*D1 mentioned before, confirmed by the Recco method. In between the respective blocks a smaller ± 100 bp block similar to *pmt*B, surrounded by recombination breakpoints, was found. The *pmt*F allele has a ± 100 bp upstream block shared with *pmt*BAL immediately followed by a very small block shared with *pmt*B. Only the former block was confirmed by Recco. The *pmt*BAL situation was rather contradictory, although some recombination breakpoints provided by Recco confirmed the bootscan similarity profile with *pmt*F. The limited similarity with *pmt*C was not supported by Recco. *pmt*E consists of the clear cut upstream ± 500 bp
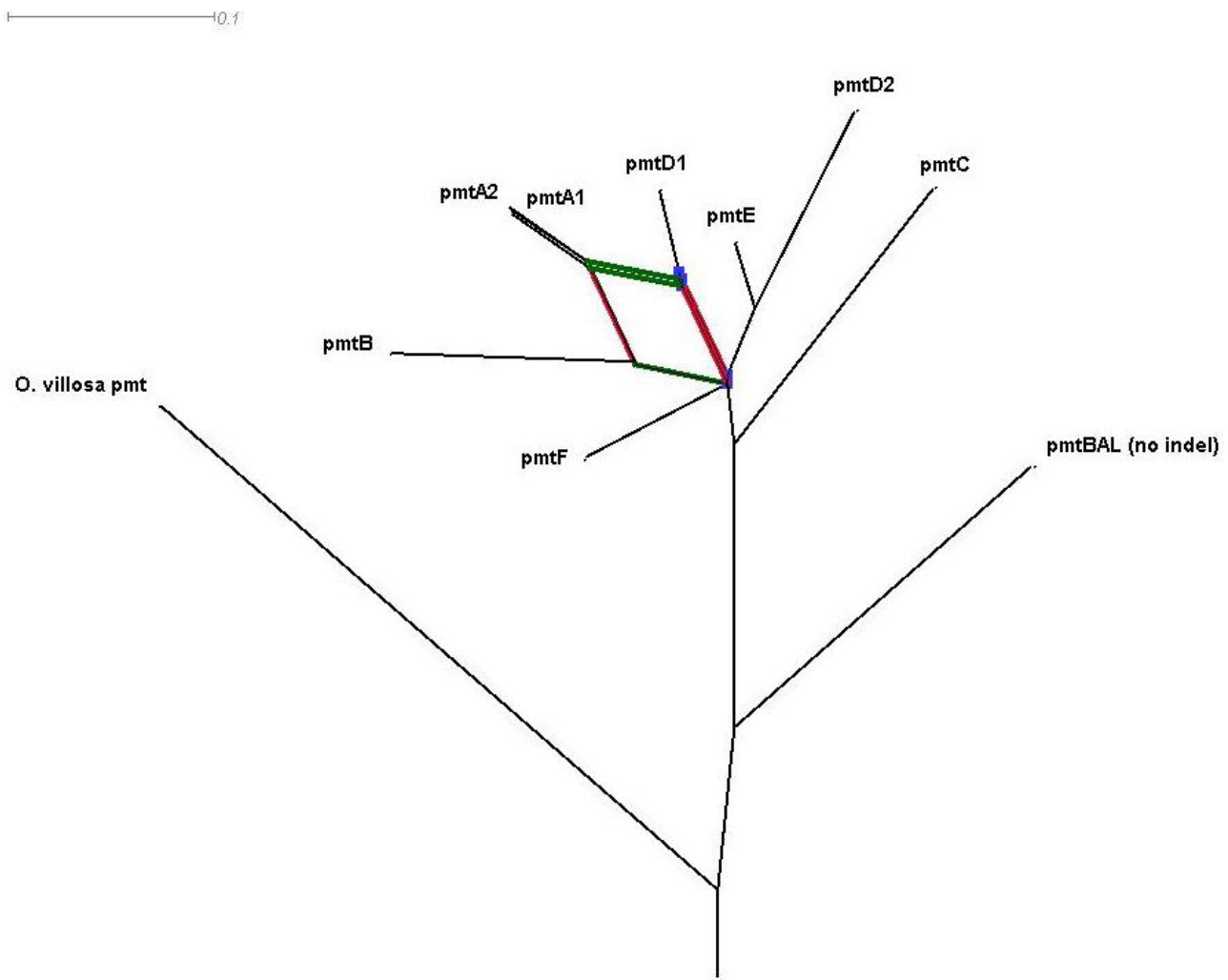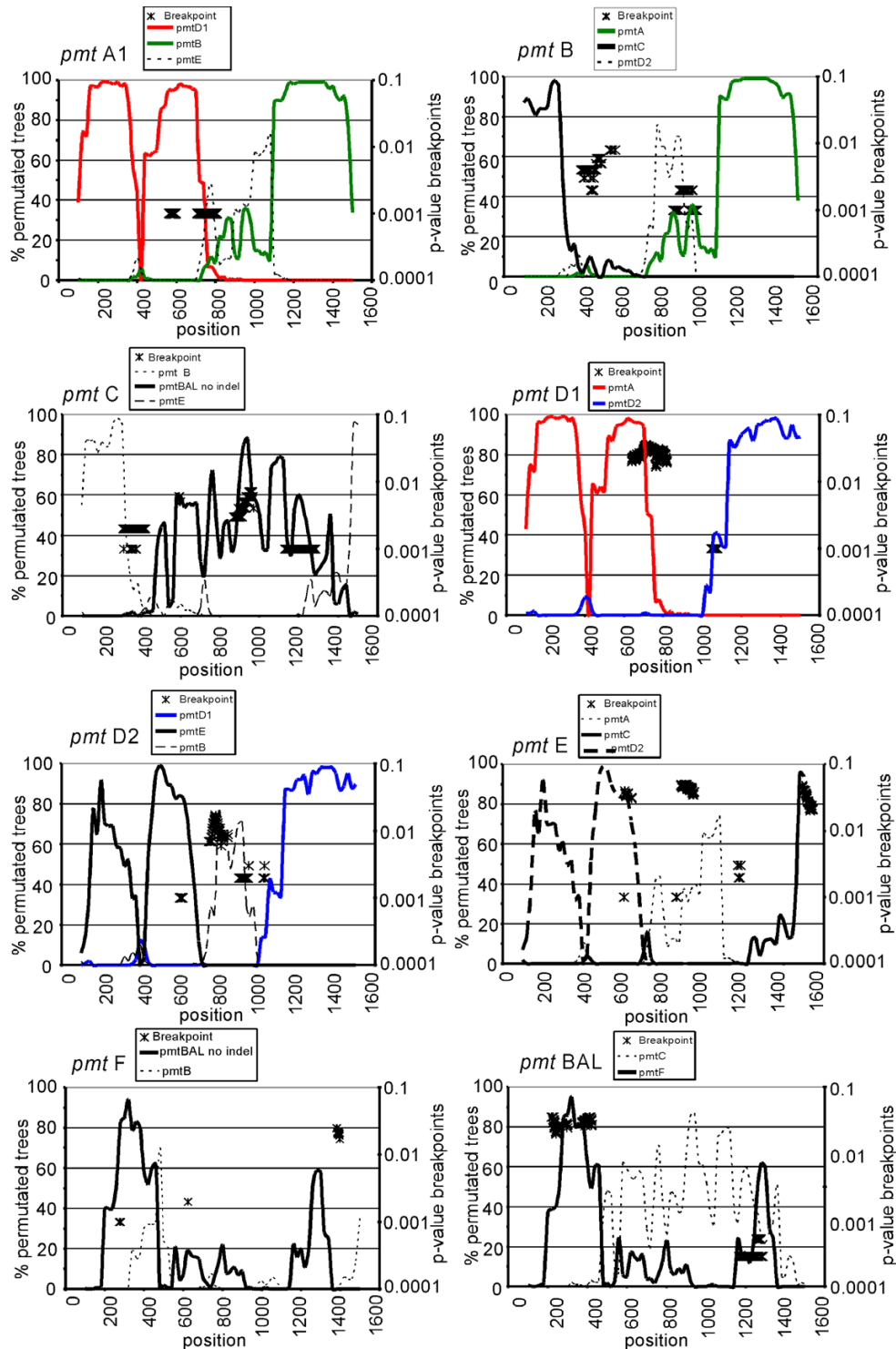
Figure 2
95% confidence reticulate network of the eight described *Orchesella cinca pmt* alleles with the *Orchesella villosa pmt* clone as an out-group, following 1000 replicate bootstraps in Splitstree v4 (uncorrected p for nucleotide substitution, NeighborNet to calculate the distance and Reticulate to calculate the splits). The colored parallel edges refer to the colors in the recombination analysis.

block shared with *pmt*D2 and the shared second exon with *pmt*C. On the other hand, the central part showed some above threshold similarity with *pmt*A1. Again numerous putative recombination breakpoints, at the edges of some shared blocks, were detected by Recco.

The deep trenches in the respective recombined blocks represent conserved modules, e.g. the one at position ± 400 bp is the region containing the ARE and the distant MRE.

Functional analysis
The functional significance of *pmt* variation was assessed by evaluating the effect of the different promoters on gene expression. Luciferase reporter assays of pGL3-*pmt* constructs were performed in *Drosophila* S2 cell line. No induction of the empty vector pGL3Basic neither the normalization vector *pAc5.1/V5-His/lacZ* was observed (data not shown), implying that the observed induction of the pGL3-*pmt* constructs are exclusively due to the interaction of the transcription factors from the host cells with the *mt*

Figure 3
Recombination analysis of *Orchesella cincta* metallothionein promoter alleles. Bootscanning analysis representing the percentage of permutated trees (left axis) that did coincide between the respective *pmt* alleles in a sliding window approach (200 bp width, 20 bp step size, Kimura 2-parameter for nucleotide substitution) relative to the sequence position. Only the relationships which trespass the 70% threshold of the permutated trees are presented. On the right axis the p-values of the respective breakpoints are indicated. The colours refer to the parallel edges in the reticulate network (Fig. 2).

promoters in the luciferase constructs. As a positive control for ecdysone treatment the 20-E exposure was also performed on cells transfected with the construct *pEcRhspluc* [54], a highly 20-E inducible luciferase construct (results not shown). Indeed, a high induction level was observed, comparable to Poels et al (2004) [45].

Basal luciferase expression data are presented in Fig. 4. Highly significant differences in basal expression were detected between the luciferase constructs from the different alleles, following a One Way ANOVA test, with the Tukey HSD post-hoc test. The basal expression from the *pmtCluc* construct hardly deviated from the empty vector pGL3Basic (data not shown) and was an order of magnitude lower than the basal expression values of the other constructs. The *pmt*D1luc had a higher basal expression than the *pmt*A1, *pmt*F and the *pmt*C constructs.

The dose response graphs from the Cd, paraquat and 20-E exposures and their estimated parameters are given in Figs. 5, 6, 7 and 8 respectively (see Additional File , 11 for a numerical summary of the curve fit data). It appeared that all constructs were susceptible to Cd. The RLU_{max} estimates of *pmt*D2*luc* and *pmt*F*luc* Cd exposures did not differ and were the highest observed. Exposure of the other constructs to Cd resulted in significantly different RLU_{max} values. The pmtC*luc* constructs, with the lowest RLU_{max} were the least inducible. The slope of *pmt*D2*luc* in the Cd
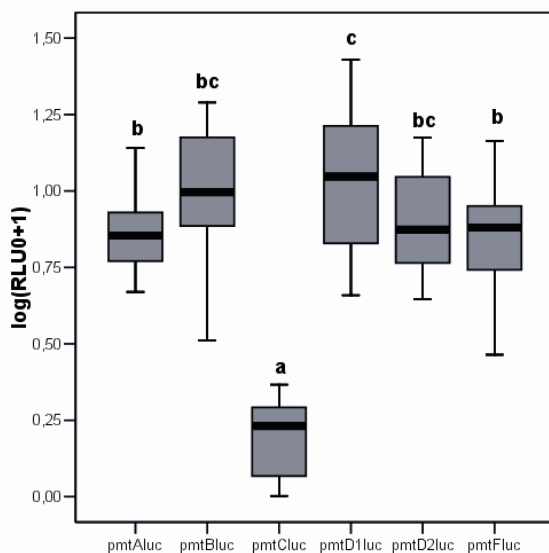
Figure 4
Boxplot representing the log transformed basal expression RLU values. One Way ANOVA; F = 71.337 and p = 0.000. A Tukey post-hoc test revealed significance groups, represented by letters.

exposure was steeper than *pmt*B*luc* and *pmt*F*luc*, implying a quicker inducibility. The most sensitive construct to induction by Cd, represented by the lowest EC_{50}, was *pmt*A1*luc*. It differed significantly from the least sensitive constructs *pmt*B*luc*, *pmt*D2*luc* and *pmt*F*luc*.

The pmtC*luc* construct did not show a significant Pearson correlation with paraquat concentration (p > 0.05) and no model fit was possible. No significant differences in the estimated parameters of the paraquat exposure were found in the responses of the inducible constructs. The 20-E exposure data revealed an inhibition of every construct. The estimated values for the parameters were abandoned, because of the wide associated 95% confidence intervals, therefore only the estimates for the RLUmin value are represented. It appeared that the pmtB*luc* construct was the least inhibited at the maximum exposure concentrations of 0.1 mM 20-E, compared to the other constructs.

At the lower range of the 20-E exposure a slight induction was observed in all constructs. We tested the RLU estimated from the inducing concentration range relative to the unexposed control in a One Way ANOVA approach (see Table 2). The *pmt*B*luc* and *pmt*D2*luc* constructs were both significantly induced at respectively 0.1 nM and 1 nM 20-E exposure.

Allele frequencies in field populations
The occurrence of the *pmt* alleles was assessed in *O. cincta* populations from a clean reference site (Roggebotzand) and a metal-polluted site (Plombières). These data are summarized in Table 3, together with information on soil metal concentrations and several indices of Cd tolerance in the two populations obtained from earlier work (Cd excretion, growth reduction and *mt* expression). The insert of the *pmt*BAL allele is omitted from this analysis.

Pollution by heavy metals in the abandoned Pb-Zn mine of Plombières dates back to the Middle Ages and has proceeded until the beginning of the 20th century [55]. This is reflected by the almost 200-fold higher total Cd content of this soil compared to the clean site in Roggebotzand. This latter site is located on reclaimed land, which fell dry in 1968. A G-test for differences of allele frequencies between both populations was highly significant (p < 0.001). The frequencies of *pmt*A1 and *pmt*A2 are relatively low in the population from the mining site Plombières, while the *pmt*C, *pmt*B and *pmt*D2 alleles are more represented compared to the situation in the reference Roggebotzand population. In addition, a higher nucleotide, haplotype and nucleotide diversity was observed in the Plombières population. When calculating the Tajima's D test, no deviation from neutral expectations was observed. The Fu and Li's tests, on the other hand, did show a signif-
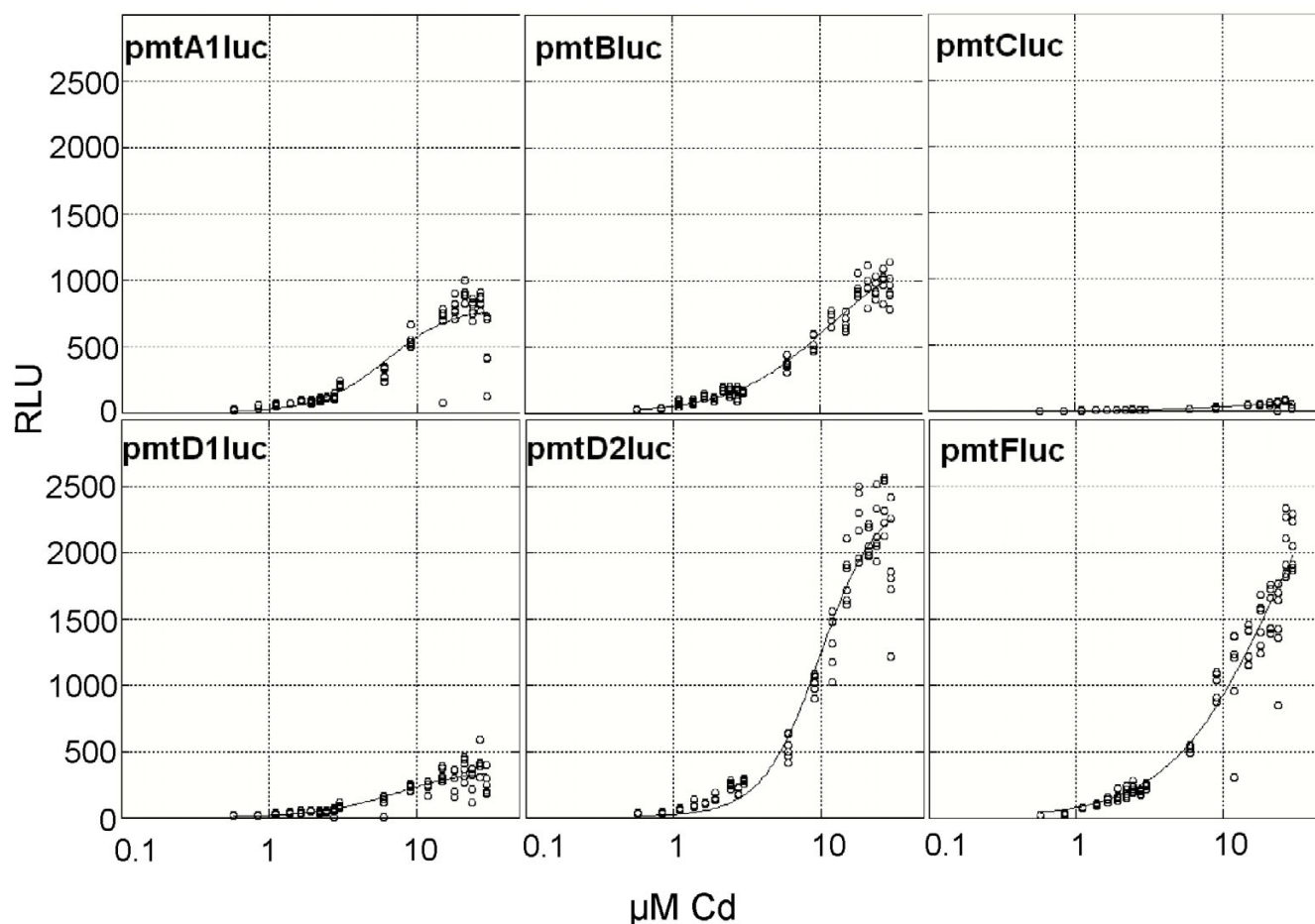
Figure 5
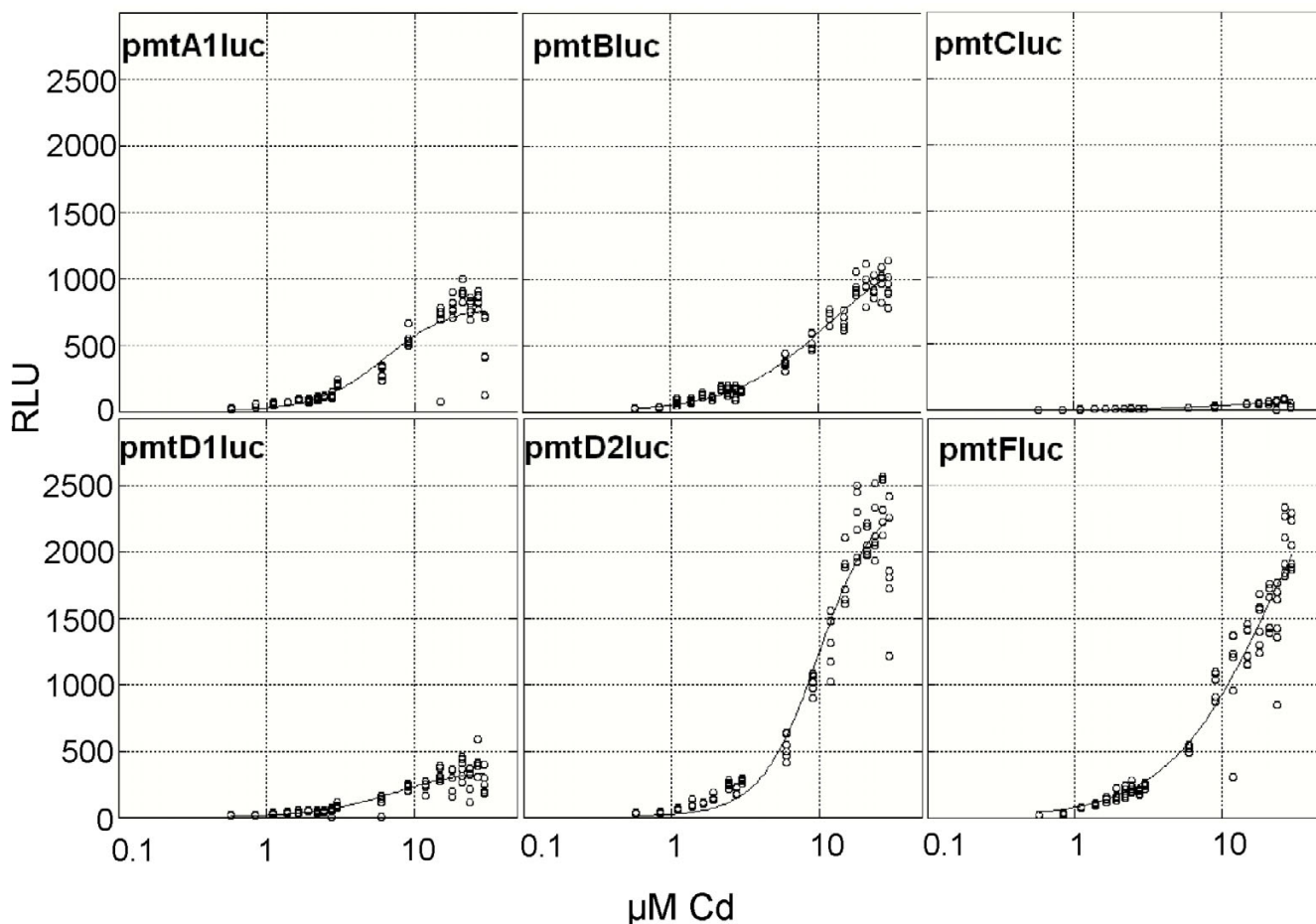Dose response relationships of the six luciferase constructs. On the Y-axis the β-galactosidase normalized relative luciferase units (RLU) are presented. On the X-axis the exposure concentrations of cadmium (Cd) are indicated.

icant departure from neutrality in the Plombières population, both with and without taking the out-group into account (D and D* respectively). Positive values of D and D* suggest the presence of an excess of intermediate frequency variants in the sample, which means that balancing selection is acting. When we calculated the Fu and Li's D in a sliding window approach, the promoter stretches on which balancing selection is taking place can be identified (see Fig 9). The spectrum of Fu and Li's D is aligned to a graph of the general architecture of the *pmt* locus. Non-significant, but negative D values are observed in the vicinity of the putative enhancer. This suggests that this stretch is under positive selection. Truly significant deviations from neutrality, with positive D values, are detected in the region between the two C/EBP binding sites, upstream from the DRE and in the vicinity of the MRE-d and MRE-e.

Discussion
From our data we have very good evidence that allelic diversity at the *pmt* locus has evolved by extensive recombination events, although we do not have knowledge about the genetic mechanism, e.g. crossing-over or gene conversion. The patterns in the splits decomposition network could somehow be linked to the results of the recombination analysis. The most important recombination blocks, which became evident in the bootscanning and Recco methods, were visible in the network. Especially the more recent alleles, *pmt*A1, *pmt*B, *pmt*D1, *pmt*D2 and *pmt*E revealed clear-cut signals of recombination among each other. Recombination between the older alleles, *pmt*BAL, *pmt*F and *pmt*C, were not detected at all in the Recco and the Splitstree method Reasons for this could be that the parental alleles were not sampled or that former recombination events could be hidden behind the muta-
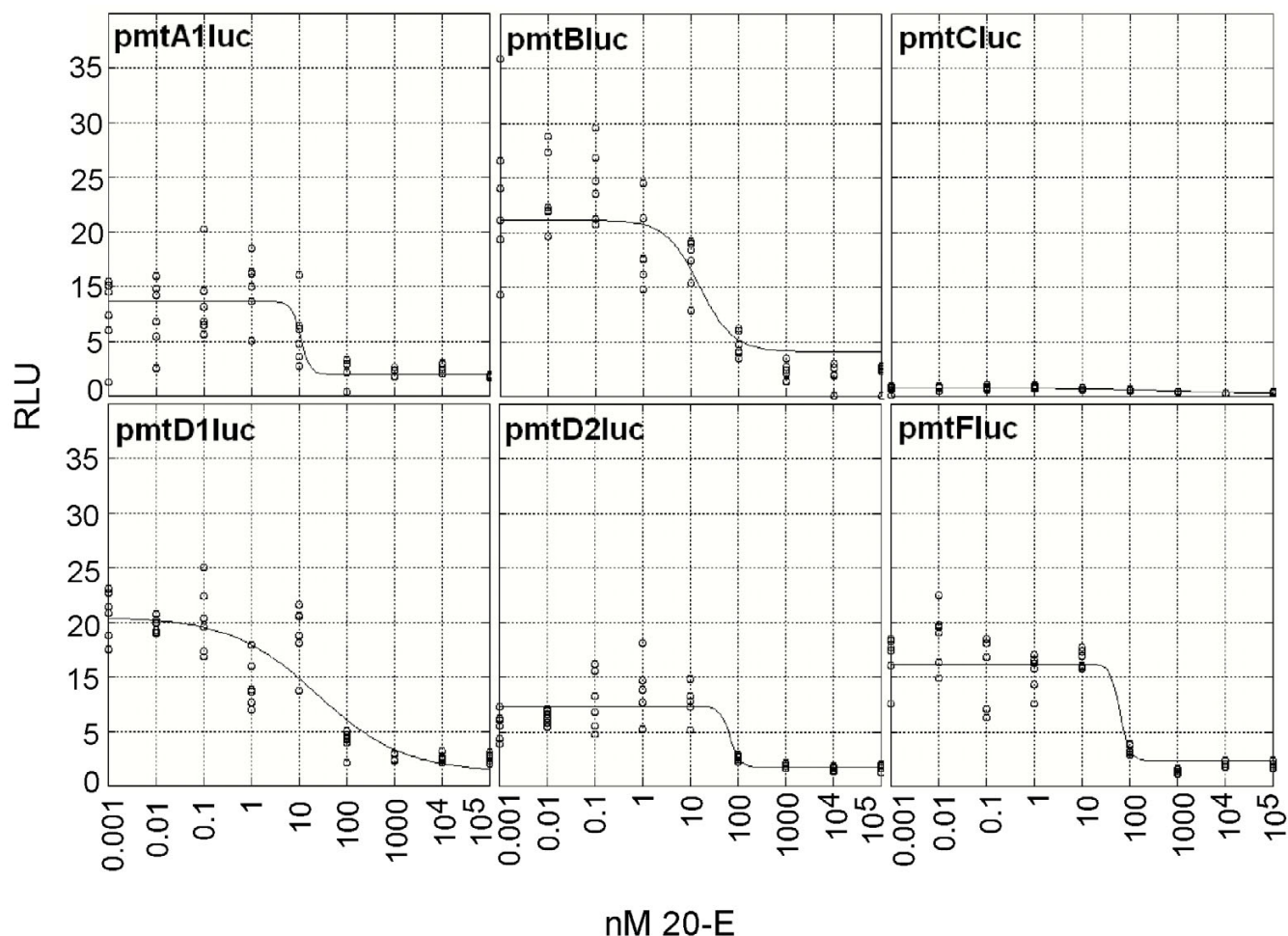
Figure 6
Dose response relationships of the six luciferase constructs. On the Y-axis the β-galactosidase normalized relative luciferase units (RLU) are presented. On the X-axis the exposure concentrations of paraquat are indicated.

tional load. Because the Recco method takes the minimization of recombination and mutation costs into account, instead of the tree-like model in the bootscanning approach, it detects recombination false positives to a lesser extent [56]. The absence of reticulations in the edges of these alleles in the 95% confidence reticulate network, points towards less conflicting bifurcations (Fig. 2). These may be caused by the accumulation of mutations following past recombination events.

The numerous recombination breakpoints at the *pmt* locus can be explained by the fact that transcriptionally active chromatin with recruited transcription factors is hypersensitive to recombination initiation [57]. Not transcription *in se* but the recruitment of recombination machinery by environmentally activated transcription factors following chromatin remodeling causes recombination events. For example, in fission yeast,

phosphorylation by the stress-activated protein kinase from the MAPK pathway increases the affinity of the transcription factor ATF1. PCR1 for a cAMP-responsive element-like (CRE-like) DNA sequence and remodels the chromatin [58]. This process and the respective CRE-like sequence are associated with a recombination hotspot in fission yeast [59]. In every *pmt* allele two CRE core sequences, CGTCA [60] were found. Because of the importance of the composition of the flanking sequences of the CRE core sequence and the large number of CREB proteins, these data are not further discussed, although it may be of importance for transcriptional regulation and initiation of recombination at the *Orchesella cincta pmt* locus.

The 1269 bp insertion in the *pmt*BAL allele, which does contain some relevant putative transcription factor binding sites, is suggested to be a possibly recombined or

Figure 7
Dose response relationships of the six luciferase constructs. On the Y-axis the β-galactosidase normalized relative luciferase units (RLU) are presented. On the X-axis the exposure concentrations of 20-hydroxyecdysone (20-E) are indicated.

duplicated region from another promoter. Similar events, where promoter regions were swapped between loci, have been described before [46].

The number of MREs in stress gene promoters in general [61] and metallothionein promoters in particular varies extensively. The same applies to their sequence and their spatial context. However, three MRE's (MRE-b, MRE_c and MRE-e) were conserved in sequence (TGCACAC) between *O. cincta* and the out-group species *O. villosa*. Previous studies revealed that the proximal MRE cluster in metal-responsive promoters is necessary [62] and MREs in the proximal promoter region need to work cooperatively for the full inductive capacity [36,43,63]. The most proximal MRE to the transcription start (MRE-a), has been shown to be most contributing to induction by Zn and Cd of the human *hmt-IIA* promoter [64] and cooperates with

more upstream MREs for the complete heavy metal induction [65].

The induction by paraquat was several orders of magnitude lower than the induction by Cd, which possibly reflects that not all the MREs are involved. In a study by [66] three different oxidative stressors elicited a twofold response of the rainbow trout MT-B promoter luciferase constructs, comparable to our study with paraquat. Beside the ARE and the MRE-a it was suggested that ARE half-sites (TGAC) are as well responsible for oxidative stress inducibility [66].

The reporter assays show that allele C is hardly inducible by Cd and oxidative stress (Fig. 5) and shows no basal transcription level above background (data not shown). This may be explained by the deletion of the MRE-a in this
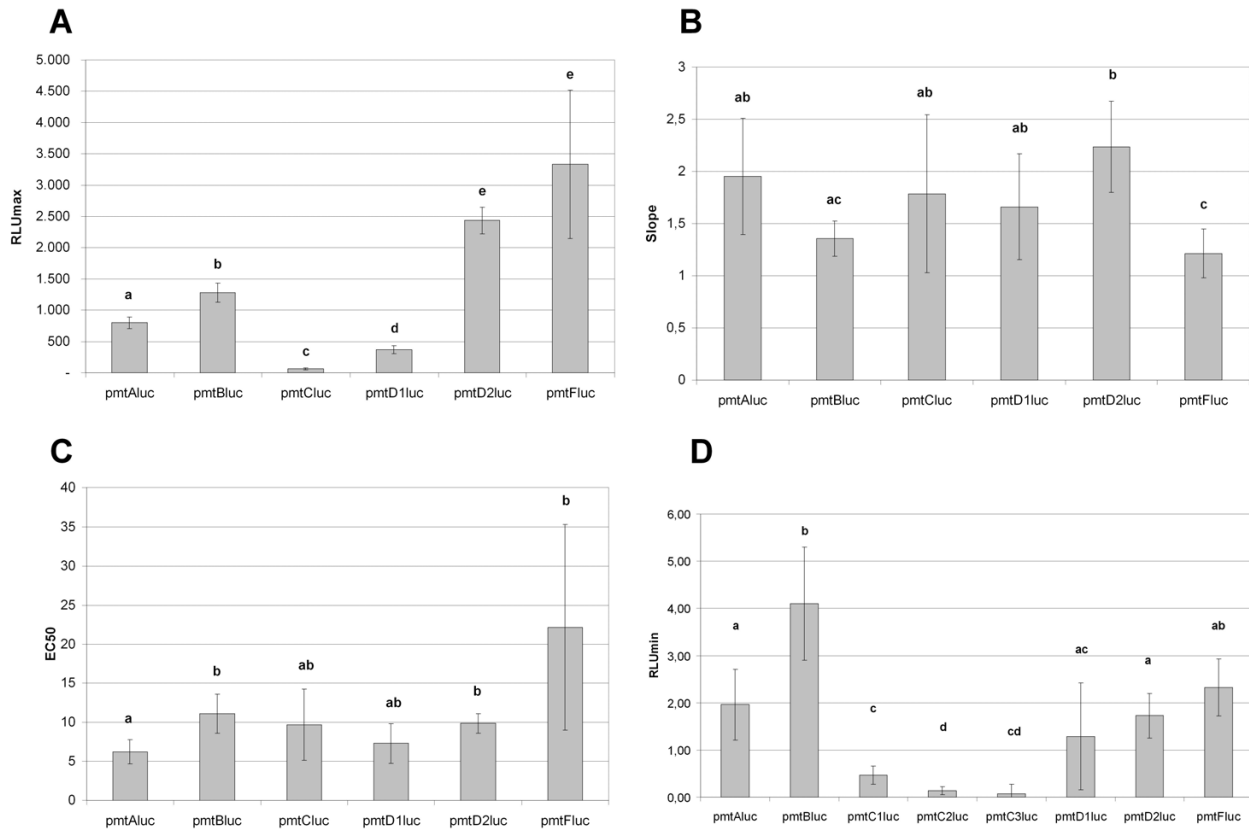
Figure 8
A: RLU_{max} estimates from the cadmium (Cd) exposure data, B: Slope estimates from the Cd exposure data, C: EC_{50} estimates from the Cd exposure data D: RLU_{min} from the 20-E exposure data.

Table 2: One Way ANOVA test comparing the measured RLU values at the control and at the putative inducing 20-E concentration. F-ratios and p-values are given.

Construct	[20-E] (nM)	F-ratio	p-value
pmtAluc	0.1	0.121	0.739
pmtBluc	0.1	10.320	**0.009**
pmtCluc	1	1.783	0.211
pmtD1luc	0.1	0.256	0.624
pmtD2luc	1	7.812	**0.019**
pmtFluc	0.01	4.598	0.058

Table 3: Background information of the sampled populations and summary of the molecular evolutionary analysis in DnaSP.

		Plombières	Roggebotzand
$[Cd]_{tot}$ soil (mg/kg) (Janssens, unpublished)		30.45 ± 14.82	0.17 ± 0.05
Allele frequencies per population (%) (Janssens, unpublished)	*pmt*A1	32.9	47.0
	*pmt*A2	7.6	15.7
	*pmt*B	24.7	10.4
	*pmt*C	11.4	2.2
	*pmt*D1	10.1	13.4
	*pmt*D2	8.9	1.5
	*pmt*E	0.6	6.0
	*pmt*F	1.9	2.9
	*pmt*BAL	1.9	0.7
	N	79	67
Sample size (number of alleles)		158 (9)	134 (9)
S (ς)		201(214)	201 (214)
H_d		0.798 ± 0.017	0.726 ± 0.032
π		0.02634 ± 0.00159	0.01491 ± 0.00162
Tajima's D		0.17187 NS	-1.3781, NS
Fu and Li's D		2.88149, p < 0.02**	-0.75998, NS
Fu and Li's D*		2.77359, p < 0.02 **	-2.13596, NS
Average constitutive *mt* expression. MNE relative to β-actin (Roelofs, unpublished)		0.52 ± 0.18	0.04 ± 0.00
Induced (1 μmole Cd/g food) *mt* expression. MNE relative to β-actin (Roelofs, unpublished)		2.63 ± 0.57	1.18 ± 0.66
Mean Cd excretion efficiency per moult (Posthuma 1993)		45%	38%

Total cadmium (Cd) content of the soil; allele frequencies of the *pmt* alleles; S, total number of variable (segregating) sites; η, total number of mutations; Hd, haplotype diversity ± SD; π, nucleotide diversity per site ± SD; Tajima's D, Fu and Li's D and Fu and Li's D*, mean normalized expression, p-value of the G-test performed on the reconstituted dataset (df = 8), MNE, mean normalized expression at mRNA level; Cd excretion efficiency; percentage of Cd-induced growth reduction.

allele as follows. During exposure to Cd and paraquat, *trans*-activation in non-mammalian metallothionein promoters can occur by interaction of an enhancer (containing MREs and/or ARE) with the MRE-a and other proximal MREs [43,63,66-69]. The induction of *pmt*C by Cd is low and completely abolished by paraquat possibly due to the incapability of *trans*-activation between the MRE-a (lacking in the C allele) and the distal enhancer, directly or in combination with CCAAT/enhancer binding proteins [70,71]. Binding sites for CCAAT/enhancer binding proteins were detected in all alleles of the *pmt* locus. Also, the MRE-a is often important in determining basal expression levels [43,64-66]. Deletion of MRE-a in *pmt* C can therefore explain the low basal expression of this allele. Finally the importance of MRE-a is reflected in the highly conserved core sequence: *O. cincta* MRE-a has the same core sequence as *Onchorhynchus mykiss*, *Strongylocentrotus purpuratus* and *Drosophila* metallothionein promoters.

There may be an alternative explanation for the very low basal expression levels of *pmt*C as well as low stress induction, related to the absence of DRE. The DRE [50] is an element found in DNA replication related genes, as well as in stress responsive genes (a.o. glutathione-S-transferase, catalase). This element coordinates the cell cycle specific expression of metallothionein. Since *pmt*C lacks this element (disruption of the consensus by a point mutation) it

may contribute to the low inducibility and basal expression.

We realize that we tested *O. cincta* derived promoters in a *Drosophila* genetic background. Thus, the observed levels may not reflect the levels *in vivo* due to absence of *O. cincta* specific transcription factors or altered binding specificity of *Drosophila* transcription factors to *O. cincta* specific transcription factor binding sites. Reporter assays are sometimes unable to generate the tissue, temporal and species specific transcriptional regulation [72-75]. For instance, aberrant expression levels have been observed when expression levels of 12 orthologous genes of human and chimpanzee were compared in cell lines from different human tissue origin [76] when compared to their *in vivo* expression levels. In contrast with that, Crawford *et al* [7] obtained consistent results using two unrelated fish cell lines (rainbow trout hepatoma cells and salmon cardiac cells) to study transcriptional activity of killifish lactate dehydrogenase-B promoters.

The general pattern of 20-E exposure is the inhibition of the metallothionein promoter. This probably occurs because of the overlapping HERE and Inr. The moderate inducibility of *pmt*B and *pmt*D2, at respectively 0.1 and 1 nM 20-E, can not be explained straightforward, although *pmt*D2 has a larger number of HEREs (5+1, See Table 1).
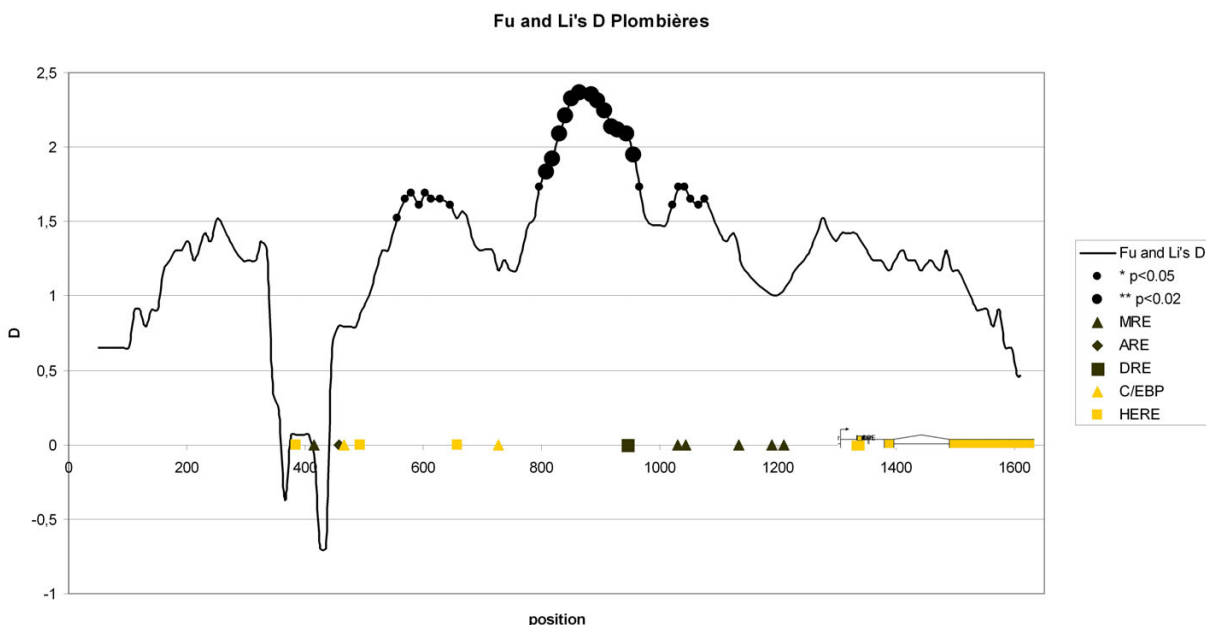
Figure 9
Sliding window analysis of the Fu and Li's D statistic on the reconstituted dataset of the Plombières population. A step size of 10 bp and a window length of 100 bp were applied. Significant deviations from neutral expectations are represented with ● and ● for p-values < 0.05 and 0.02 respectively. A cartoon of the general architecture of the *pmt* locus is provided with the putative transcription factor binding sites.

The spatial and sequence context could be the reason for their inducibility. In the firebrat *Thermobia domestica* 20-E equivalents peak to 5 μM during apolysis [77] falling back to the basal concentration of 80 nM, indicating that the order of magnitude of our exposure range was well within the physiologically relevant range. These data suggest that the metallothionein expression is switched off during ecdysone-induced apoptosis of the gut epithelium, when the onset of new cuticle formation is set.

Our small-scale population genetic comparison between populations from the clean site in Roggebotzand and the polluted one in Plombières reflected a positive sign of balancing selection at Plombières, because of the significant positive Fu and Li's D values, higher nucleotide diversity per site, and higher haplotype diversity in the latter population. A study by Timmermans *et al.* [26] revealed selection on certain alleles of the *O. cincta mt* coding sequence by heavy metal content in the soil. However, no signatures of any selection were detected in the amino acid sequence (d_n/d_s, Fisher's exact test) or in the nucleotide composition (Tajima's D). The population at Plombières is characterized by relatively high frequencies of *pmt*D2 and *pmt*B,

which is in accordance with their greater inducibility compared to the most common *pmt*A1 allele. This suggests a fitness advantage to phenotypes with high metallothionein expression at polluted sites [23]. On the other hand, the Plombières population also has a relatively high frequency of the less responsive *pmt*C allele, This is consistent with the signature of balancing selection detected by Fu and Li's D values and can be understood if there are not only advantages in terms of metal tolerance but also fitness costs associated with high *mt* expression. Further work is necessary to elucidate the precise relationship between fitness and *pmt* genotype.

Conclusion
The *Orchesella cincta* metallothionein promoter contains a high degree of polymorphism, reflected in the different number and spacing of consensus transcription factor binding sites involved in relevant regulatory processes, such as heavy metal, oxidative stress induction and the regulation by the molting cycle.

In general, the evolution of transcriptional regulation can occur by stabilizing selection by which erosive mutations

in TFBS and the emergence of new ones are compensating each other. Evolution due to transcriptional regulation has been reported in macro-evolutionary processes, e.g. the even-skipped-2 enhancer of *Drosophila* species [78-80]. Alternatively, micro-evolutionary processes rather occur by point mutations in (putative) TFBS [3,6,7]. Here, we provide evidence of extensive recombination, reshuffling the nucleotide variation of insertion, deletions and point mutations in the micro-evolution of a regulatory locus.

Since induction of metallothionein is suggested to be associated with a metal-tolerant phenotype [24,25] the difference in inducibility between the *pmt* alleles provides a scope for natural selection in field populations. The deviations from neutral expectations in the cadmium tolerant population from Plombières support this suggestion. Future approaches will be the detection of DNA-binding activity of selected transcription factors, measure the *in vivo* transcription by real-time Q-PCR, and screen field populations for their respective *pmt* allele frequencies.

Methods

DNA purification
DNA of individual animals was purified by the SV genomic DNA purification system (Promega Corporation). The maxiprep used for the genome walking procedure was done by a modified CTAB extraction method [81] on 100 adult individuals.

Genome Walking and PCR
Different aliquots of 10 µg of genomic DNA of *Orchesella cincta* (laboratory culture) and *Orchesella villosa* (Belcaro, Italy) were digested overnight with 5 to 10 U different blunt cutting restriction enzymes, *Dra*I, *Eco*RV, *Hinc*II, *Hpa*I, *Sca*I, *Sma*I and *Stu*I. These digests were cleaned up by phenol chloroform extraction, followed by ethanol precipitation and consequently ligated overnight at 4°C to the adapter originating from the hybridized adapters 1 and 1A (see Additional File , 12).

Based on the sequences resulting from the Universal fast walking method [82] on the metallothionein promoter, (Mariën and Roelofs, unpubl.), a nested reverse primer pair, (R34 and R130, see Additional File , 12) in the proximal promoter was designed. A nested gradient PCR approach was conducted with above mentioned primers to each of the differentially digested aliquots. The same approach was done in two steps for the *O. villosa* promoter with respectively the primer pairs (Ovimt210R and Ovimt233R) (Spinsanti, unpubl.) and the resulting primer pair (Ovipmt and Ovipmtnested).

In the resulting sequences for *O. cincta* and *O. villosa*, the respective primers, D1-36F and OvipmtfarF were developed on the 5' end. The metallothionein promoter was amplified by PCR ($T_{ann} = 55°C$) with the primer combination D1-36F and MT-265R [25] and OvipmtfarF and Ovimt233R for *O. cincta* and *O. villosa* respectively. The reverse primer hybridizes in the second exon of the *mt* gene, and consequently allows the alignment of the previously described *Orchesella cincta mt* sequences (Timmermans *et al.* in press) [25]) to the resulting *Orchesella cincta pmt* sequences.

DNA-Sequencing
32 *pmt* fragments, originating from various populations (see Additional File , 13), with previously known proximal promoter SSCP genotype or with deviating RFLP patterns (Janssens, unpublished) were amplified with the D1-36F an MT265R primer pair. Resulting PCR fragments were ligated in the pGEM-T vector (Promega Corporation) and cloned in JM109 or XL1-Blue competent cells by respectively heat-shock or electroporation procedures. Plasmid purification was executed with the SV Minipep System (Promega Corporation). The clones were sequenced with Big Dye V 1.1 (ABI) on a ABI 3100 capillary sequencer, analyzed with Vector NTI Software 10.0.1 (Invitrogen) and aligned to the *mt* alleles previously described by Timmermans et al. (pers. comm.). The alignments served to make a consensus for every allele, and are provided as Additional Files 2, 3, 4, 5, 6, 7, 8, 9. One single *O. villosa pmt* sequence (Zelzate, Belgium) was cloned and sequenced to serve as out-group in phylogenetic analysis.

Transcription Factor Binding Site Analysis
Because of the lack of functional studies on this promoter, the descriptive work of the observed sequence variation was restricted to the identification of consensus binding sites available in the literature by using the program Genepalette 1.2 [83]. The core promoter structure was analyzed by using consensus sequences from the literature [49,83,84]. Consensus binding sites for transcription factors known from metallothionein induction [16,31] cell cycle regulated and stress genes [50] and molting cycle regulated transcription processes [51] were included in the analysis.

Luciferase reporter assay
From six of the *O. cincta pmt* alleles a luciferase reporter construct was made by PCR on the respective minipreps with the primers, D136-F*Kpn*I and MT-73R*Xho*I, containing restriction sites for these respective enzymes. The clones on which these luciferase reporter constructs were made are summarized in Additional File , 13 section. The fragments were ligated in *Kpn*I and *Xho*I double-digested pGL3Basic luciferase vector (Promega Corporation) and

cloned in XL1-Blue by electroporation. As an internal control for transfection efficiency and number of cells, a β-galactosidase reporter plasmid with a constituve actin promoter of *Drosophila melanogaster* (pAc5.1/V5-His/lacZ (Invitrogen), was used. Purification of transfection grade and endotoxin-free plasmid was done with the Nucleobond PC 500 EF kit (Macherey-Nagel)

Drosophila S2 cells (Gibco) were grown at 28°C in Schneiders *Drosophila* Medium (revised) (Gibco) containing 15% fetal calf serum (Gibco). On the first day cells were plated out in 96 wells plates at approximately 25000 cells/150 μl and grown for 24 hours at 28°C. The second day every well was transfected using a calcium phosphate precipitation method [85] with 0.72 μg of DNA in 15 μl. The DNA used for transfection was a 1/1 ratio of the respective luciferase constructs with the pAc5.1/V5-His/lacZ. Following overnight incubation the medium was removed and replaced by spiked exposure media. Exposure ranges for Cd, paraquat and 20-hydroxyecdysone (20-E) were 0–30 μM, 0–1800 μM and 0–100 μM respectively. Every exposure was executed in six wells. The basal expression was measured in hexaplicate in four independent experiments.

Thousand fold stock solutions of the respective exposure concentrations were filter sterilized (0.2 μm) and stored at 4°C. Cadmium and paraquat were dissolved in water whereas 20-E was dissolved in DMSO. The 20-E stock in DMSO was not filter sterilized. Schneider's *Drosophila* Medium (revised) (Gibco), containing 5% fetal calf serum (Gibco), was spiked with the respective stock solution to achieve the required final concentrations.

The exposure with Cd and 20-E was performed during 24 hours and the paraquat exposure was limited to 6 hours due to cytotoxicity. After the exposure period the cells were lysed in 100 μl lysis buffer containing (25 mM Tris, 2 mM dithiothreitol, 2 mM trans-1,2-diaminocyclohexane-N,N,N9,N9-tetraacetic acid monohydrate, 10% glycerol, and 1% Triton® X-100 [Sigma-Aldrich, Steinheim, Germany] in demineralized water, pH 7.8 overall buffer).

A 50 μl aliquot was used to measure luciferase activity using 100 μl glowmix, (20 mM tricine, 1.07 mM C4H2Mg5O14, 2.6 mM MgSO4, 0.1 mM ethylenedi-amine-N,N,N9,N9-tetraacetic acid, 33.3 mM dithiothreitol, 0.27 mM coenzyme A, 0.46 mM luciferine, and 0.53 mM adenosine-59-triphosphate in demineralized water), on a LucyII luminometer (Anthos Labtec instruments). Beta-galactosidase activity was measured in a 50 μl lysis aliquot, diluted in 100 μl lysis buffer and 90 μl of ONPG-mix. The ONPG mix consisted of 24 μl of 10xZ buffer (60 mM Na2HPO4, 40 mM NaH2PO4, 10 mM KCl and 2 mM MgSO4), 66 μl water, 0.1 μl β-mercaptoethanol and 0.16

mg 2-nitrophenyl β-D-galactopyranoside (ONPG). The reactions were incubated during one hour at 28°C and subsequently the absorbance at 420 nm was measured on a Spectramax 340pc (Molecular Devices) spectrophotometer.

Firstly, luciferase signals in each well were normalized with an internal luciferase standard, in order to standardize the readings between different plates. Secondly, the normalizations for the number of cells, and the transfection efficiency, were done by dividing the latter values by A_{420} values from the ONPG measurement. The basal expression values of the constructs and the pGL3Basic vector were measured in four replicate experiments, six wells per experiment.

Data analysis

DNA sequences were processed and aligned in Vector NTI version 10.0.1 (Invitrogen). Recombination events, which took place at this locus, make the construction of a bifurcating tree less relevant. Therefore, the Splitstree4 software [52] was used to construct a reticulate network to represent evolutionary relationships between the respective alleles. The assumptions under which the network was constructed are: uncorrected P for nucleotide substitution, NeighborNet to calculate the distances and the reticulate method to treat the splits. Gaps, constant and non-parsimonous positions were omitted from the analysis. Bootstrap analysis was performed on one thousand replicates and subsequently a 95% confidence network was constructed.

Recombination sites were detected by using two methods. The first method we used relies on the bootscan method implemented in the program Simplot [86]. It constructs replicate trees in a sliding window approach. The general accepted threshold level for the detection of a recombination is a clustering in 70% of the permutated trees in a certain window. Thousand replicate neighbor-joining trees were made by using the following parameters, window size 200 bp, step size 20 bp and the Kimura 2-parameter as a model to estimate nucleotide substitution. Alternatively, a non-phylogenetic method, applying dynamic programming that minimizes the mutation and recombination cost between sequences was used. This method is implemented in the software Recco [56]. The parameter α, representing the ratio of mutation cost to recombination cost was set to 0.2, the methods to calculate mutation and recombination costs were respectively Hamming and Delta Dirac.

The population genetic data were achieved by RFLP analysis of the amplified *pmt* fragments by PCR (Janssens unpublished). A reconstituted dataset was made by pasting the consensus for every allele the number of times it

was observed in the sample, according to Timmmermans *et al.*, in press These datasets were aligned with each other and an alignment with the out-group *Orchesella villosa pmt* sequence was provided. Molecular diversity indices and deviations from the neutral theory (Tajima's D, Fu and Li's D and D*) were calculated using DnaSP v4.10 [87].

The basal expression RLU data were log(x+1) transformed to approach normality of the data, A One Way ANOVA with $p < 0.05$ was executed. The LSD post-hoc test was performed to discriminate significance groups.

Initial curve-fitting was done in Kaleidagraph v 3.5 for a rough estimate of the parameters, and the curve plotting. Fine tuning of the curve fitting was done in SPSS v 12.0.1.

The dose-responses of the Cd and paraquat exposures on the RLU of every construct were compared by estimating the RLU_{max}, EC_{50} and the slope by fitting a curve with the following formula, from which RLU_{max} by summing RLU_0 (the average RLU at unexposed conditions) and RLU_e.

$$RLU = RLU_0 + \frac{RLU_{est}}{1 + e^{(-slope.\ln([Cd]/EC_{50}))}}$$

The 20-E exposure data were fitted with the following formula and RLU_{min}, slope and EC_{50} were estimated.

$$RLU = RLU_{est} + \frac{RLU_0}{1 + e^{(slope.\ln([Cd]/EC_{50}))}}$$

The RLU values from 20-E exposure concentrations at which a putative induction was observed were tested by a One Way ANOVA ($p < 0.05$) to test for induction.

Authors' contributions

TJ performed the genome walking, cloning, sequencing, reporter assays, the data analysis and drafted the manuscript. JM and DR performed the initial universal fast walking and assisted in the field and lab work. JL and PC supported the cell culture and luciferase reporter assays and evaluated the manuscript. DR and NvS assisted in outlining the study, directing research and revision of the manuscript.

Additional material

Additional file 1
Alignment of the primary sequences. The primary sequences were aligned with the consensus sequences of the alleles identified in the mt *coding sequence [26]*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2148-7-88-S1.msf]

Additional file 2
Alignment per allele. These alignments show the variation between the clones from which the consensus sequence for every allele was extracted.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2148-7-88-S2.msf]

Additional file 3
Alignment per allele. These alignments show the variation between the clones from which the consensus sequence for every allele was extracted.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2148-7-88-S3.msf]

Additional file 4
Alignment per allele. These alignments show the variation between the clones from which the consensus sequence for every allele was extracted.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2148-7-88-S4.msf]

Additional file 5
Alignment per allele. These alignments show the variation between the clones from which the consensus sequence for every allele was extracted.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2148-7-88-S5.msf]

Additional file 6
Alignment per allele. These alignments show the variation between the clones from which the consensus sequence for every allele was extracted.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2148-7-88-S6.msf]

Additional file 7
Alignment per allele. These alignments show the variation between the clones from which the consensus sequence for every allele was extracted.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2148-7-88-S7.msf]

Additional file 8
Alignment per allele. These alignments show the variation between the clones from which the consensus sequence for every allele was extracted.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2148-7-88-S8.msf]

Additional file 9
Alignment per allele. These alignments show the variation between the clones from which the consensus sequence for every allele was extracted.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2148-7-88-S9.msf]

Additional file 10

Alignment of the pmt *alleles with* Orchesella villosa pmt *as an outgroup. These data were used in the phylogenetic and recombinational analysis.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2148-7-88-S10.msf]

Additional file 11

Curve fit estimates of the luciferase reporter assay. Overview of the curve fit estimates for the exposure of every luciferase construct to Cd, paraquat and 20-E.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2148-7-88-S11.doc]

Additional file 12

Primer Table. In this table the name and sequence of the used primers are given.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2148-7-88-S12.doc]

Additional file 13

Sequence strategy Table. In this table the number and origin of the clones sequenced per allele are given.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2148-7-88-S13.DOC]

Acknowledgements

References

1. Wray GA, Hahn MW, Abouheif E, Balhoff JP, Pizer M, Rockman MV, Romano LA: **The evolution of transcriptional regulation in eukaryotes.** *Molecular Biology and Evolution* 2003, **20(9):**1377-1419.
2. Arnosti DN: **Analysis and function of transcriptional regulatory elements: Insights from Drosophila.** *Annual Review of Entomology* 2003, **48:**579-602.
3. de Meaux J, Goebel U, Pop A, Mitchell-Olds T: **Allele-specific assay reveals functional variation in the chalcone synthase promoter of** *Arabidiopsis thaliana* **that is compatible with neutral evolution.** *Plant Cell* 2005, **17(3):**676-690.
4. Schulte PM: **Environmental adaptations as windows on molecular evolution.** *Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology* 2001, **128(3):**597-611.
5. Schulte PM, Gómez-Chiarri M, Powers DA: **Structural and functional differences in the promoter and 5' flanking region of Ldh-B within and between populations of the teleost Fundulus heteroclitus.** *Genetics* 1997, **145(3):**759-769.
6. Schulte PM, Glemet HC, Fiebig AA, Powers DA: **Adaptive variation in lactate dehydrogenase-B gene expression: Role of a stress-responsive regulatory element.** *Proceedings of the National Academy of Sciences of the United States of America* 2000, **97(12):**6597-6602.
7. Crawford DL, Segal JA, Barnett JL: **Evolutionary analysis of TATA-less proximal promoter function.** *Mol Biol Evol* 1999, **16(2):**194-207.
8. Lerman DN, Michalak P, Helin AB, Bettencourt BR, Feder ME: **Modification of heat-shock gene expression in Drosophila melanogaster populations via transposable elements.** *Molecular Biology and Evolution* 2003, **20(1):**135-144.
9. Michalak P, Minkov I, Helin A, Lerman DN, Bettencourt BR, Feder ME, Korol AB, Nevo E: **Genetic evidence for adaptation-driven incipient speciation of Drosophila melanogaster along a microclimatic contrast in "Evolution Canyon," Israel.** *Proceedings of the National Academy of Sciences of the United States of America* 2001, **98(23):**13195-13200.
10. Daborn PJ, Yen JL, Bogwitz MR, Le Goff G, Feil E, Jeffers S, Tijet N, Perry T, Heckel D, Batterham P, Feyereisen R, Wilson TG, ffrench-Constant RH: **A single P450 allele associated with insecticide resistance in Drosophila.** *Science* 2002, **297(5590):**2253-2256.
11. Dallinger R: **Metallothionein research in terrestrial invertebrates: synopsis and perspectives.** *Comp Biochem Physiol C Pharmacol Toxicol Endocrinol* 1996, **113(2):**125-133.
12. Coyle P, Philcox JC, Carey LC, Rofe AM: **Metallothionein: The multipurpose protein.** *Cellular and Molecular Life Sciences* 2002, **59(4):**627-647.
13. Stürzenbaum SR P. Kille & A.J. Morgan: **The identification, cloning and characterization of earthworm metallothionein.** *FEBS Letters* 1998, **431:**437-442.
14. Imagawa M, Onozawa T, Okumura K, Osada S, Nishihara T, Kondo M: **Characterization of Metallothionein Cdnas Induced by Cadmium in the Nematode Caenorhabditis-Elegans.** *Biochemical Journal* 1990, **268(1):**237-240.
15. Freedman JH, Slice LW, Dixon D, Fire A, Rubin CS: **The Novel Metallothionein Genes of Caenorhabditis-Elegans - Structural Organization and Inducible, Cell-Specific Expression.** *Journal of Biological Chemistry* 1993, **268(4):**2554-2564.
16. Miles AT, Hawksworth GM, Beattie JH, Rodilla V: **Induction, regulation, degradation, and biological significance of mammalian metallothioneins.** *Critical Reviews in Biochemistry and Molecular Biology* 2000, **35(1):**35-70.
17. Kägi JHR: **Overview of Metallothionein.** *Methods in Enzymology* 1991, **205:**613-626.
18. Hensbergen PJ, Donker MH, van Velzen MJM, Roelofs D, van der Schors RC, Hunziker PE, van Straalen NM: **Primary structure of a cadmium-induced metallothionein from the insect Orchesella cincta (Collembola).** *Eur J Biochem* 1999, **259(197-203):**.
19. Hensbergen PJ, van Velzen MJ, Nugroho RA, Donker MH, van Straalen NM: **Metallothionein-bound cadmium in the gut of the insect Orchesella cincta (Collembola) in relation to dietary cadmium exposure.** *Comp Biochem Physiol C Toxicol Pharmacol* 2000, **125(1):**17-24.
20. Sterenborg I, Vork NA, Verkade SK, Van Gestel CAM, Van Straalen NM: **Dietary zinc reduces uptake but not metallothionein binding and elimination of cadmium in the springtail, Orchesella cincta.** *Environmental Toxicology and Chemistry* 2003, **22(5):**1167-1171.
21. Posthuma L, Hogervorst RF, Joosse ENG, Van Straalen NM: **Genetic variation and covariation for characteristics associated with cadmium tolerance in natural populations of the springtail Orchesella cincta (L.).** *Evolution* 1993, **47(2):**619-631.
22. Posthuma L, Hogervorst RF, van Straalen NM: **Adaptation to soil pollution by cadmium excretion in natural populations of Orchesella cincta (L.) (Collembola).** *Arch Environ Contam Toxicol* 1992, **22:**146-156.
23. Timmermans M, Ellers J, Roelofs D, Van Straalen NM: **Metallothionein mRNA expression and cadmium tolerance in metal-stressed and reference populations of the springtail Orchesella cincta.** *Ecotoxicology* 2005, **14(7):**727-739.
24. Sterenborg I, Roelofs D: **Field-selected cadmium tolerance in the springtail Orchesella cincta is correlated with increased metallothionein mRNA expression.** *Insect Biochemistry and Molecular Biology* 2003, **33(7):**741-747.
25. Roelofs D, Overhein L, de Boer ME, Janssens TKS, van Straalen NM: **Additive genetic variation of transcriptional regulation: metallothionein expression in the soil insect** *Orchesella cincta.* *Heredity* 2006, **96(1):**85-92.
26. Timmermans MJTN, Ellers J, van Straalen NM: **Allelic diversity of metallothionein in Orchesella cincta (L.): traces of natural selection by environmental pollution.** *Heredity* 2007, **98(5):**311-319.

27. Roelofs D, Mariën J, van Straalen NM: **Differential gene expression profiles associated with heavy metal tolerance in the soil insect *Orchesella cincta*.** *Insect Biochemistry and Molecular Biology* 2007, **37(4):**287-295.
28. Morgan AJ, Kille P, Sturzenbaum SR: **Microevolution and ecotoxicology of metals in invertebrates.** *Environmental Science & Technology* 2007, **41(4):**1085-1096.
29. Maroni G, Wise J, Young JE, Otto E: **Metallothionein gene duplications and metal tolerance in natural populations of Drosophila melanogaster.** *Genetics* 1987, **117:**739-744.
30. Tanguy A, Boutet I, Bonhomme F, Boudry P, Moraga D: **Polymorphism of metallothionein genes in the Pacific oyster Crassostrea gigas as a biomarker of response to metal exposure.** *Biomarkers* 2002, **7(6):**439-450.
31. Haq F, Mahoney M, Koropatnick J: **Signaling events for metallothionein induction.** *Mutation Research* 2003, **533:**211-226.
32. Moilanen LH, Fukushige T, Freedman JH: **Regulation of metallothionein gene transcription - Identification of upstream regulatory elements and transcription factors responsible for cell-specific expression of the metallothionein genes from Caenorhabditis elegans.** *Journal of Biological Chemistry* 1999, **274(42):**29655-29665.
33. Sturzenbaum SR, Georgiev O, Morgan AJ, Kille P: **Cadmium detoxification in earthworms: From genes to cells.** *Environmental Science & Technology* 2004, **38(23):**6283-6289.
34. Bonneton F, Theodore L, Silar P, Maroni G, Wegnez M: **Response of Drosophila metallothionein promoters to metallic, heat shock and oxidative stresses.** *FEBS Letters* 1996, **380:**33-38.
35. Cserjesi P, Fairley P, Brandhorst BP: **Functional analysis of the promoter of a sea urchin metallothionein gene.** *Biochem Cell Biol* 1992, **70(10-11):**1142-1150.
36. Bai G, Stuebing EW, Parker HR, Harlow P, Nemer M: **Combinatorial regulation by promoter and intron 1 regions of the metallothionein SpMTA gene in the sea urchin embryo.** *Mol Cell Biol* 1993, **13(2):**993-1001.
37. Schmitt-Wrede HP, Koewius H, Tschuschke S, Greven H, Wunderlich F: **Genomic organization of the cadmium-inducible tandem repeat 25-kDa metallothionein of the oligochaete worm *Enchytraeus buchholzi*.** *Biochimica Et Biophysica Acta-Gene Structure and Expression* 2004, **1680(1):**24-33.
38. Bunch TA, Grinblat Y, Goldstein LSB: **Characterization and use of the Drosophila metallothionein promoter in cultured Drosophila melanogaster cells.** *Nucleic Acids Research* 1988, **16(3):**1043-1061.
39. Yin H, Smith M, Glass J: **Stable expression of C/EBP alpha in prostate cancer cells down-regulates metallothionein and increases zinc-induced toxicity.** *Prostate* 2005, **62(3):**209-216.
40. Tanguy A, Moraga D: **Cloning and characterization of a gene coding for a novel metallothionein in the Pacific oyster Crassostrea gigas (CgMT2): a case of adaptive response to metal-induced stress?** *Gene* 2001, **273(1):**123-130.
41. Otto E, Allen JM, Young JE, Palmiter RD, Maroni G: **A DNA segment controlling metal-regulated expression of the Drosophila melanogaster metallothionein gene Mtn.** *Mol Cell Biol* 1987, **7(5):**1710-1715.
42. Boldrin F, Santovito G, Gaertig J, Wloga D, Cassidy-Hanley D, Clark TG, Piccinni E: **Metallothionein gene from Tetrahymena thermophila with a copper-inducible-repressible promoter.** *Eukaryotic Cell* 2006, **5(2):**422-425.
43. Samson SLA, Gedamu L: **Molecular analyses of metallothionein gene regulation.** *Progress in Nucleic Acid Research* 1998, **59:**257-288.
44. Fang YY, Kashkarov U, Anders MW, Board PG: **Polymorphisms in the human glutathione transferase zeta promoter.** *Pharmacogenet Genomics* 2006, **16(5):**307-313.
45. Hoogendoorn B, Coleman SL, Guy CA, Smith K, Bowen T, Buckland PR, O'Donovan MC: **Functional analysis of human promoter polymorphisms.** *Human Molecular Genetics* 2003, **12(18):**2249-2254.
46. Yuasa N, Nakagawa Y, Hayakawa M, Iimura Y: **Two alleles of the sulfite resistance genes are differentially regulated in *Saccharomyces cerevisiae*.** *Bioscience Biotechnology and Biochemistry* 2005, **69(8):**1584-1588.
47. Wu MT, Demple B, Bennett RAO, Christiani DC, Fan R, Hu H: **Individual variability in the zinc inducibility of metallothionein-IIA mRNA in human lymphocytes.** *Journal of Toxicology and Environmental Health-Part A* 2000, **61(7):**553-567.
48. Suntres ZE: **Role of antioxidants in paraquat toxicity.** *Toxicology* 2002, **180(1):**65-77.
49. Cherbas C: **The arthropod initiator: The capsite consensus plays an important role in transcription.** *Insect Biochemistry and Molecular Biology* 1993, **23(1):**81-90.
50. Matsukage A F. Hirose, Y. Hayashi, K. Hamada & M. Yamaguchi: **The DRE sequence TATCGATA, a putative promoter-activating element for Drosophila melanogaster cell-proliferation-related genes.** *Gene* 1995, **166:**233-236.
51. Luo Y, Amin J, Voellmy R: **Ecdysterone receptor is a sequence-specific transcription factor involved in the developmental regulation of heat shock genes.** *Mol Cell Biol* 1991, **11(7):**3660-3675.
52. Huson DH, Bryant D: **Application of phylogenetic networks in evolutionary studies.** *Molecular Biology and Evolution* 2006, **23(2):**254-267.
53. Huson DH, Kloepper TH: **Computing recombination networks from binary sequences.** *Bioinformatics* 2005, **21(suppl ii):**ii159-ii165.
54. Poels J, Martinez A, Suner MM, De Loof A, Dunbar SJ, Vanden Broeck J: **Functional and comparative analysis of two distinct ecdysteroid-responsive gene expression constructs in Drosophila S2 cells.** *Insect Biochemistry and Molecular Biology* 2004, **34(5):**451-458.
55. Cappuyns V, Swennen R, Vandamme A, Niclaes M: **Environmental impact of the former Pb-Zn mining and smelting in East Belgium.** *Journal of Geochemical Exploration* 2006, **88(1-3):**6-9.
56. Maydt J, Lengauer T: **Recco: recombination analysis using cost optimization.** *Bioinformatics* 2006, **22(9):**1064-1071.
57. Koren A, Ben-Aroya S, Kupiec M: **Control of meiotic recombination initiation: a role for the environment?** *Current Genetics* 2002, **42(3):**129-139.
58. Mizuno K, Hasemi T, Ubukata T, Yamada T, Lehmann E, Kohli J, Watanabe Y, Iino Y, Yamamoto M, Fox ME, Smith GR, Murofushi H, Shibata T, Ohta K: **Counteracting regulation of chromatin remodeling at a fission yeast cAMP responsive element-related recombination hotspot by stress-activated protein kinase, cAMP-dependent kinase and meiosis regulators.** *Genetics* 2001, **159(4):**1467-1478.
59. Fox ME, Yamada T, Ohta K, Smith GR: **A family of cAMP-response-element-related DNA sequences with meiotic recombination hotspot activity in *Schizosaccharomyces pombe*.** *Genetics* 2000, **156(1):**59-68.
60. Smolik SM, Rose RE, Goodman RH: **A cyclic Amp-responsive element-binding transcriptional activator in Drosophila melanogaster, dCREB-A, is a member of the leucine zipper family.** *Mol Cell Biol* 1992, **12(9):**4123-4131.
61. Lichtlen P, Wang Y, Belser T, Georgiev O, Certa U, Sack R, Schaffner W: **Target gene search for the metal-responsive transcription factor MTF-1.** *Nucleic Acids Research* 2001, **29(7):**1514-1523.
62. Ren HW, Xu MX, He PF, Muto N, Itoh N, Tanaka K, Xing J, Chu MM: **Cloning of crucian carp (*Carassius cuvieri*) metallothionein-II gene and characterization of its gene promoter region.** *Biochemical and Biophysical Research Communications* 2006, **342(4):**1297-1304.
63. Scudiero R, Carginale V, Capasso C, Riggio M, Filosa S, Parisi E: **Structural and functional analysis of metal regulatory elements in the promoter region of genes encoding metallothionein isoforms in the Antarctic fish *Chionodraco hamatus* (icefish).** *Gene* 2001, **274:**199-208.
64. Koizumi S, Suzuki K, Ogra Y, Yamada H, Otsuka F: **Transcriptional activity and regulatory protein binding of metal-responsive elements of the human metallothionein-IIA gene.** *European Journal of Biochemistry* 1999, **259(3):**635-642.
65. Samson SLA, Gedamu L: **Metal-responsive elements of the rainbow trout metallothionein-B gene-function for basal and metal-induced activity.** *Journal of Biological Chemistry* 1995, **270(12):**6864-6871.
66. Samson SLA, Paramchuk WJ, Gedamu L: **The rainbow trout metallothionein-B gene promoter: contributions of distal promoter elements to metal and oxidant regulation.** *Biochimica et Biophysica Acta* 2001, **1517:**202-211.
67. Olsson PE, Kille P: **Functional comparison of the metal-regulated transcriptional regions of metallothionein genes from cadmium-sensitive and tolerant fish species.** *Biochimica et Biophysica Acta* 1997, **1350:**325-334.

68. Chen WY, John JAC, Lin CH, Lin HF, Wu SC, Lin CH, Chang CY: **Expression of metallothionein gene during embryonic and early larval development in zebrafish.** *Aquatic Toxicology* 2004, **69(3):**215-227.
69. Olsson PE, Kling P, Erkell LJ, Kille P: **Structural and functional analysis of the rainbow trout (Oncorhyncus mykiss) metallothionein-a gene.** *European Journal of Biochemistry* 1995, **230(1):**344-349.
70. Aniskovitch LP, Jacob ST: **Purification and characterization of a rat liver protein that recognizes CCAAT-homologous sequences of the metallothionein promoter and trans-activates this promoter.** *Archives of Biochemistry and Biophysics* 1997, **341(2):**337-346.
71. Aniskovitch LP, Jacob ST: **Distinct rat proteins can recognize CCAAT-homologous sequences of the metallothionein promoter and trans-activate this promoter.** *Oncogene* 1998, **16(11):**1475-1486.
72. Babiss LE, Herbst RS, Bennett AL, Darnell JE: **Factors That Interact with the Rat Albumin Promoter Are Present Both in Hepatocytes and Other Cell-Types.** *Genes & Development* 1987, **1(3):**256-267.
73. Clayton DF, Harrelson AL, Darnell JE: **Dependence of liver-specific transcription on tissue organization.** *Mol Cell Biol* 1985, **5(10):**2623-2632.
74. Li XX, Liao WSL: **Expression of Rat Serum Amyloid-A1 Gene Involves Both C/Ebp-Like and Nfkb-Like Transcription Factors.** *Journal of Biological Chemistry* 1991, **266(23):**15192-15201.
75. Bancroft JD, McDowell SA, Degen SJF: **The Human Prothrombin Gene - Transcriptional Regulation in Hepg2 Cells.** *Biochemistry* 1992, **31(49):**12469-12476.
76. Heissig F, Krause J, Bryk J, Khaitovich P, Enard W, Paabo S: **Functional analysis of human and chimpanzee promoters.** *Genome Biology* 2005, **6(7):**.
77. Delapaz AR, Delbecque JP, Bitsch J, Delachambre J: **Ecdysteroids in the hemolymph and the ovaries of the firebrat Thermobia domestica (Packard) (Insecta, Thysanura) Correlations with integumental and ovarian cycles.** *Journal of Insect Physiology* 1983, **29(4):**323-329.
78. Ludwig MZ, Bergman C, Patel NH, Kreitman M: **Evidence for stabilizing selection in a eukaryotic enhancer element.** *Nature* 2000, **403(6769):**564-567.
79. Ludwig MZ, Palsson A, Alekseeva E, Bergman CM, Nathan J, Kreitman M: **Functional evolution of a cis-regulatory module.** *Plos Biology* 2005, **3(4):**588-598.
80. Ludwig MZ: **Functional evolution of noncoding DNA.** *Current Opinion in Genetics & Development* 2002, **12:**634-639.
81. van der Wurff AWG, Chan YL, van Straalen NM, Schouten J: **TE-AFLP: combining rapidity and robustness in DNA fingerprinting.** *Nucl Acids Res* 2000, **28(24):**e105.
82. Myrick KV, Gelbart WM: **Universal fast walking for direct and versatile determination of flanking sequence.** *Gene* 2002, **284:**125-131.
83. Rebeiz MJWP: **GenePalette: a universal software tool for genome sequence visualization and analysis.** *Developmental Biology* 2004, **271(2):**431-438.
84. Smale ST, Kadonaga JT: **The RNA polymerase II core promoter.** *Annual Review of Biochemistry* 2003, **72:**449-479.
85. Sambrook J, Fritsch EF, Maniatis T: **Molecular cloning: a laboratory manual.** Cold Spring Harbor. , Cold Spring Harbor Press.; 1989.
86. Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS, Novak NG, Ingersoll R, Sheppard HW, Ray SC: **Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination.** *Journal of Virology* 1999, **73(1):**152-160.
87. Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R: **DnaSP, DNA polymorphism analyses by the coalescent and other methods.** *Bioinformatics* 2003, **19(18):**2496-2497.