

Full Paper

# Epigenetic regulation of serotype expression antagonizes transcriptome dynamics in *Paramecium tetraurelia*

Miriam Cheaib<sup>1</sup>, Azim Dehghani Amirabad<sup>1,2</sup>, Karl J. V. Nordström<sup>3</sup>, Marcel H. Schulz<sup>2</sup>, and Martin Simon<sup>1,\*</sup>

<sup>1</sup>Molecular Cell Dynamics, Centre for Human and Molecular Biology, Saarland University, Saarbrücken 66123, Germany, <sup>2</sup>Cluster of Excellence for Multimodal Computing and Interaction, Saarland University and Max Planck Institute for Informatics, Saarbrücken 66123, Germany, and <sup>3</sup>Epigenetics Department, Centre for Human and Molecular Biology, Saarland University, Saarbrücken 66123, Germany

\*To whom correspondence should be addressed. Tel. +49 681-302-4938. Fax. +49 681-302-2703.  
E-mail: martin.simon@uni-saarland.de

Edited by Dr Minoru Yoshida

Received 6 April 2015; Accepted 28 June 2015

## Abstract

Phenotypic variation of a single genotype is achieved by alterations in gene expression patterns. Regulation of such alterations depends on their time scale, where short-time adaptations differ from permanently established gene expression patterns maintained by epigenetic mechanisms. In the ciliate *Paramecium*, serotypes were described for an epigenetically controlled gene expression pattern of an individual multigene family. Paradoxically, individual serotypes can be triggered in *Paramecium* by alternating environments but are then stabilized by epigenetic mechanisms, thus raising the question to which extend their expression follows environmental stimuli. To characterize environmental adaptation in the context of epigenetically controlled serotype expression, we used RNA-seq to characterize transcriptomes of serotype pure cultures. The resulting vegetative transcriptome resource is first analysed for genes involved in the adaptive response to the altered environment. Secondly, we identified groups of genes that do not follow the adaptive response but show co-regulation with the epigenetically controlled serotype system, suggesting that their gene expression pattern becomes manifested by similar mechanisms. In our experimental set-up, serotype expression and the entire group of co-regulated genes were stable among environmental changes and only heat-shock genes altered expression of these gene groups. The data suggest that the maintenance of these gene expression patterns in a lineage represents epigenetically controlled robustness counteracting short-time adaptation processes.

**Key words:** antigenic variation, epigenetics, ciliate, telomere position effect, heat-shock

## 1. Introduction

Alterations in gene expression patterns are the regulatory basis for phenotypic plasticity: one single genome may produce quite different phenotypes, thus representing an important mechanism to adapt for alternating environmental conditions including morphological, physiological and behavioural changes.<sup>1</sup> Similarly, differentiation processes in developmental stages of multicellular organisms allow one single

genotype to differentiate into distinct tissues, representing the regulation of gene expression patterns in the meaning of phenotypic plasticity: epigenetic mechanisms have the aim to fix these gene expression patterns rather than to allow subsequent flexibility.<sup>2</sup>

Among the diverse mechanisms regulating alterations in gene expression during phenotypic variation, one can distinguish mechanisms that evolved as a fast reaction to stresses, e.g. the heat-shock response,

and adaptations to long-term changes in environments, e.g. such as temperature and food availability. The molecular regulation of the heat-shock proteins (HSPs) represents a tightly controlled transcriptional activation of molecular chaperones that assist proper folding of proteins.<sup>3</sup> Usually, this transcriptional activation of HSPs occurs in minutes and only for a few minutes, thus being a rapid, programmed and reversible response to stress.

Although it is clear that short-time stress response and long-term phenotypic plasticity are key players for survival during short-time and long-term environmental changes, a general understanding of the complex genetic and epigenetic mechanisms controlling phenotypic plasticity as a response to environmental changes is still missing.<sup>4</sup> Only few studies addressed the extent of phenotypic plasticity based on genome-wide transcriptomic profiling in response to altered environments.

Adaptation to different environmental temperatures represents one of the most important stress responses of organisms. As ectothermic species may use any environmental heat source to regulate their body temperature thus optimizing their metabolism, some organisms are poikilothermic, meaning that their body temperature is equal to the ambient environmental temperature which challenges the cellular metabolism to work properly in a huge range of temperatures. Depending on the level of temperature variation, adaptation processes have to control survival critical functions like efficient metabolic turnover. In the past, much attention was spent on the composition of membrane lipids during cold adaptation and heat-shocks,<sup>5</sup> but only few studies investigated transcriptomic changes to identify general metabolic pathways altered during temperature adaptation processes.

Paramecia are unicellular model organisms belonging to the ciliate clade. These are distributed all over the world in all climate areas. However, a single population may have to cope with drastic temperature alterations: on the one hand, these cells are so small that they cannot prevent poikilothermy by increasing body mass/volume ratio, and on the other hand, their natural ecosystem of small ponds undergoes drastic short-time temperature alterations. It is not surprising that paramecia in laboratory cultures can easily be cultivated at a broad temperature range from 4°C to 32°C. Therefore, *Paramecium* represents an excellent model to investigate mechanisms of temperature adaptation, and this ciliate has been subject of several studies investigating heat-shock response of individual genes,<sup>6–8</sup> but a genome-wide analysis is still missing.

This unicellular model organism has become even more attractive for the characterization of phenotypic plasticity as several epigenetic mechanisms have been described that allow differentiation of lineages into different cell states, which are reminiscent of developmental differentiation processes in multicellular organisms. One prominent example concerns the expression of the mating-type system,<sup>9</sup> and similarly, but not by the same mechanisms, the antigenic system was described to be epigenetically controlled a long time ago.<sup>10</sup> Recent studies indicate that the antigenic system is regulated by complex mechanisms involving small RNAs and dynamic chromatin alterations<sup>11,12</sup> (see Cheaib et al., in preparation): *Paramecium* displays mutually exclusive expression of variable surface antigen genes similar to the phenomenon of antigenic variation in pathogens.<sup>13</sup> Serotypes, meaning exclusive expression of one antigen gene of the surface antigen gene (SAg) multigene family, are initially triggered by different environmental conditions, e.g. temperature and food availability.<sup>14</sup> They are stabilized by mechanisms similar to epigenetic differentiation processes, which are under control of an RNAi mechanism likely involved in transcriptional silencing.<sup>11</sup> It was shown that a telomere position effect (TPE) is required for SAg regulation, suggesting that

spreading of telomeric heterochromatin is involved in RNAi-mediated silencing.<sup>12</sup> Differentiation into distinct serotypes therefore represents an epigenetically controlled manifestation of gene expression patterns, which interestingly can also be passed on to sexual progeny. In contrast to mating-type determination, serotype switches can occur anytime without the need for sexual recombination. Hence, they represent an important example of epigenetically controlled and heritable transcriptomic variations of a single genotype.<sup>13</sup>

To understand the processes of long-term and shocked temperature adaptations in the context of serotype expression, we analysed the transcriptome information of serotype pure cultures to see to which extent the transcriptome becomes altered during temperature adaptation and how this is related to the epigenetically controlled SAg multigene family. We furthermore analysed the subtelomeres of SAg containing chromosomes to characterize a possible heterochromatic spreading from telomeric repeats by transcriptomic analysis.

## 2. Materials and methods

### 2.1. *Paramecium* culture conditions, serotype analysis and heat-shocks

*Paramecium tetraurelia* strain 51 and d4-2 were grown in wheat grass powder (WGP, Pines International Co., Lawrence, KS, USA) infusion medium bacterized with *Klebsiella pneumoniae* supplemented with 0.8 µg/ml β-sitosterol, unless otherwise stated (serotype 51A, 51H of strain 51) and 1 : 1 diluted with Volvic® mineral water (serotype 51D, 51B of strain d4-2). To stabilize/trigger respective serotypes, different cultivation temperatures were chosen: 51A at 31°C, 51B at 24°C/6°C, 51D at 24°C and 51H at 14°C. Serotype expression was verified by immobilization with specific antibodies (1 : 100) (rabbit α-51A/-51B/-51D/-51H) in depression slides. Only serotype-pure cultures (refers to 100% immobilization) were used for further analysis. For heat-shock experiments, serotype-51D cultures were harvested, washed in Volvic® mineral water and exposed to 39°C for 20 min.<sup>8</sup> 51B cells were selected out of a 51D culture by immobilization with α-51D serum (1 : 100). Individual survived cells were transferred to fresh WGP-medium bacterized with *K. pneumoniae* after washing in Volvic® mineral water. One hundred per cent serotype 51B expressing cultures were cultivated at 24°C. They were then split, and the counterpart was cultivated at 14°C for 1 day before transfer to 6°C.

### 2.2. RNA isolation, cDNA library creation, Illumina sequencing

A total of 150,000 *P. tetraurelia* cells per sample were harvested, washed in Volvic® mineral water and incubated for 20 min at the respective temperature in non-nutrient medium to reduce bacterial occurrence in food vacuoles. Total RNA was isolated using TriReagent® (Sigma-Aldrich, Seelze, Germany). After DNase I (Invitrogen, Karlsruhe, Germany) digestion and subsequent purification with acid phenol, RNA integrity was verified on an Agilent Bioanalyzer 2100. Poly-A enrichment (NEBNext® Poly(A) mRNA Magnetic Isolation Module) out of 1 µg total RNA and library preparation (NEB-Next® Ultra™ Directional RNA Library Prep Kit for Illumina) were carried out according to manufacturer's recommendation using 11 PCR cycles of library enrichment. Multiplexed libraries were 100 nt paired-end sequenced with an average of 14 Mio reads per sample on Illumina HiSeq 2500 platform. Reads were demultiplexed with bcl2fastq (v1.8.4) and trimmed for adaptor contamination and low-quality bases with the cutadapt (v1.4.1)<sup>15</sup> wrapper trim\_galore (v0.3.3). All raw data were deposited at the European Nucleotide

Archive (ENA, <http://www.ebi.ac.uk/ena>) under study accession no. PRJEB9464.

### 2.3. Annotation of HSP70 isoforms and phylogenetic reconstruction

HSP70 candidates were identified by a proteome-wide search against the Pfam library of hidden Markov models.<sup>16</sup> Hits were extracted for PF00012 (HSP70) using an E-value threshold of  $3.6E-23$ . Amino acid sequences were aligned with ClustalW, and neighbour-joining trees were calculated with MEGA6<sup>17</sup> with 1,000 bootstrap replicates, based on the Poisson model and after complete deletion of gaps.

### 2.4. RNA-seq expression analysis

We quantified gene expression levels using the Sailfish algorithm.<sup>18</sup> Briefly, Sailfish uses transcript sequences to build a kmer index over all annotated transcript sequences. RNA-seq reads are decomposed into kmers and projected onto kmers of transcripts; then expression levels of all transcripts are deconvoluted. Finally, transcript expression levels of the same gene are summarized to give a gene expression estimate. As a normalized expression metric for plots, we used Reads per Kilobase per Million mapped reads (RPKM) reported by Sailfish. We used Sailfish version 0.6.2 with the following command lines:

```
sailfish index -t <ref_transcripts> -o <out_dir> -k 20 for building transcriptome index file and
```

```
sailfish quant -i indexfile -l 'T=PE:O =><:S=AS' -l sample_R1.fastq -2 R2.fastq -o Sample1 for strand-specific quantification of read files. We downloaded the P. tetraurelia cDNA file from the Ensembl protists database release 26 (assembly version GCA_000165425). The quantification was done for 39,635 genes.
```

After gene expression quantification, differential gene expression analysis was conducted using DESeq2 version 3<sup>19</sup> in R. Following the DESeq2 guidelines, absolute read counts per gene as obtained from Sailfish were normalized between samples to account for differences in sequencing depth. Linear models to assess DE status were fit for each gene with DESeq2 using a negative binomial distribution. After multiple testing correction,<sup>20</sup> genes with false discovery rate (FDR) <0.01 were considered as differentially expressed.

### 2.5. Transcriptome expression landscape visualization and analysis

To visualize and compare the transcriptome expression landscape for the different conditions, we applied a self-organizing map (SOM) approach using the oposSOM package.<sup>21</sup> An SOM is a machine learning technique for dimensionality reduction, which transfers a high-dimensional data matrix (expression of all genes across different samples) to a lower dimension.

The oposSOM package decomposes the data into clusters of correlated sets of genes, called metagenes, which are arranged in a fix order (here we use a lattice with dimension 20 by 20). The expression of a metagene represents the average expression of all genes in the cluster. Metagene expression values in the individual samples provide mosaic pictures visualizing condition specific over- and under-expression in terms of characteristic colour-coded textures, which allows the direct comparison of the expression of individual samples in a simple and intuitive way.

TPM scale expression levels were log-transformed as pre-processing step for training the SOM. After training, we had sample-specific transcriptome expression landscapes as well over- and under-expressed hot spots for every sample. Sample-specific transcriptome landscapes represent the expression level of the metagenes at the

respective sample while preserving the topological order of the metagenes in high dimension. Metagenes are grouped into co-regulated hot spots of similar up- or down-regulation using the following criteria: if  $\Delta e > \text{MaxEXP} * 0.9$ , where  $\Delta e$  is the expression level of the metagene under consideration and MaxEXP denotes the maximum expression of all metagenes, then the metagene is classified as over expressed. Analogously if  $\Delta e < \text{MinEXP} * 0.9$ , where MinEXP denotes the minimal expression of all metagenes, the metagene is classified as under-expressed. Metagenes with the same classification that group together in the SOM are called a regulated spot.

Hierarchical clustering (average linkage mode) of expression values was conducted using R, heatmap2 function. We used two different setups for the creation. The first was based on Euclidean distance computation, which measures the distance between all entries in the expression vectors. The second was based on 1-Pearson correlation, which measures if genes follow similar trends independent of their differences in expression values. In all cases, no scaling was used.

### 2.6. Gene ontology enrichment

The gene ontology (GO) term association file has been created by processing the parameciumDB.gff3 file, downloaded from the *ParameciumDB* database.<sup>22</sup> Differentially expressed genes (DEGs) for every pairwise condition were divided into up- and down-regulated genes. We conducted GO enrichment analysis using Ontologizer (version 2.1) software using the Parent-Child-Union method for each set of up- or down-regulated genes in each pairwise comparison.<sup>23</sup>

## 3. Results and discussion

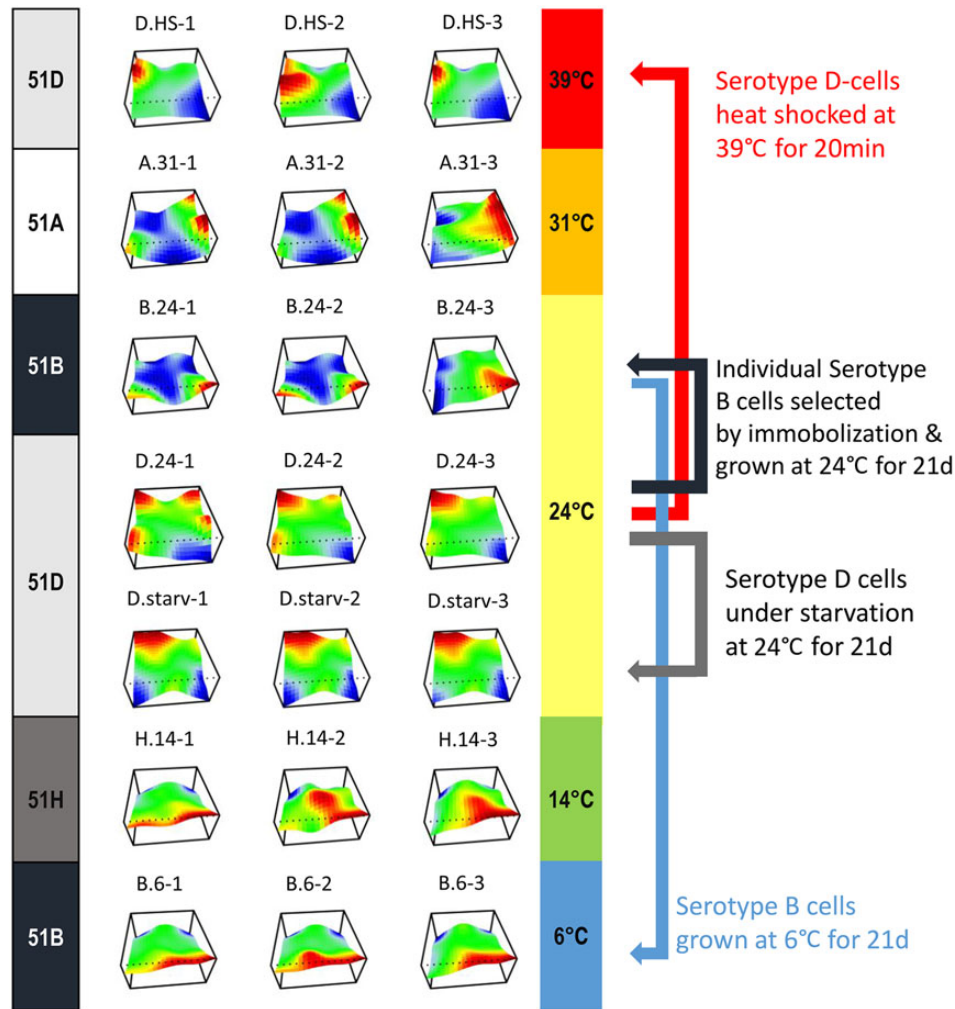
### 3.1. Experimental design

It was the aim of this study to characterize gene expression patterns following epigenetically stabilized serotype expression in *P. tetraurelia*. We therefore analysed the transcriptomes of four different serotype pure cultures (51A, 51B, 51D, 51H) and furthermore characterized transcriptome alteration during cold adaptation, starvation and heat-shocks to compare epigenetically controlled gene expression patterns to short-time adaptation processes.

All experiments were started with serotypically pure cultures, meaning that 100% of cells expressed one single surface antigen obviously by 100% immobilization to a specific antiserum.

In this study, we analysed seven distinct conditions, three biological replicates each (Fig. 1):

- (i) Serotype 51A expressing cultures cultivated at 31°C for >6 months (A.31).
- (ii) Serotype 51D expressing cultures cultivated at 24°C for >6 months (D.24).
- (iii) Serotype 51H expressing cultures cultivated at 14°C for >6 months (H.14).
- (iv) Three aliquots of the D.24 cultures subjected to a 20 min heat-shock at 39°C (D.HS).
- (v) Three cultures (B.24) were grown from 51B expressing individuals, which were selected from the D.24 cultures by immobilization with anti-51D serum. Surviving individuals were subjected to autogamy and further cultivated at 24°C for 3 weeks.
- (vi) Three aliquots of the 51B cultures that were gradually adapted for lower temperatures (starting at 24°C, slow cool down to 14°C within 24 h, slow cool down to 6°C in 24 h) and then cultivated 3 weeks with permanent food supply at 6°C (B.6).



**Figure 1.** The experimental design and transcriptional landscape of individual cultures. The transcriptional landscape of analysed cultures is illustrated as a 3D heatmap of grouped genes. On the left side, the actual serotype is shown, meaning the surface antigen to which antiserum 100% of the cultures showed positive immobilization reaction at the time of RNA isolation. On the right side, the individual cultivation temperature is indicated (except the heat-shock cultures that were cultured at 24°C and were transferred to 39°C only for 20 min).

- (vii) Three aliquots of the D.24 cultures were subjected to autogamy and subsequently starved at 24°C for 3 weeks (D.starv).

Replicate experiments are denoted with numeric suffixes, for example, the three replicate cultures of B.6 are denoted B.6-1, B.6-2 and B.6-3. As serotypes are expressed throughout the life cycle and usually in most individuals of a population, all cultures analysed in this study represent vegetatively growing non-synchronous cultures, thus representing individuals of different cellular age.

To ensure that transcriptome differences do not result from activation of genes specifically activated during sexual recombination (autogamy), all cultures were checked for fragmented nuclei by DAPI staining prior to RNA isolation. In addition, Supplementary Fig. S1 shows gene expression profiles of different (early, intermediate, late induced) autogamy-related genes showing no significant activation in any culture.

### 3.2. Analysis of DEGs

As *P. tetraurelia* had at least three whole-genome duplications, there are a number of genes that appear in more than one highly similar

copy in the genome.<sup>24</sup> This complicates gene expression estimates as reads may map to more than one location in the genome, potentially biasing estimates of non-expressed genomic copies of a gene. We used the Sailfish algorithm,<sup>18</sup> which uses a global approach to optimize read placements on genes, to avoid false-positive expression signals of silent genes. We did a qualitative investigation by looking at the two highly similar genes SAg 51D and 51J (92.2% sequence identity) that we had previously measured with qPCR.<sup>12</sup> Whereas the 51D gene has the highest expression in D serotype cultures, the 51J gene is silent, despite the highly similar sequence, and the same pattern is observed for the Sailfish estimates and with qPCR.

After quantifying all gene expression values in the samples, we used a dimensionality reduction approach with SOMs to compare the expression landscapes between different conditions and their replicates (see Section 2). Figure 1 shows the transcriptional landscapes of all analysed cultures in 3D heatmaps. As expected, individual replicates are highly similar over the complete landscape, but may differ for smaller, individual gene groups, as seen for cultures A.31-3 and B.24-3. We decided to proceed the subsequent analysis also with these samples rather than to omit them as they represent biological fluctuations that cannot be neglected, but we are aware that some

DEGs may become masked in the subsequent analyses merging all three replicates.

At first glance, a similar but not identical landscape can be observed in the maps of the two cold-adapted experiments cultured at 6°C and 14°C (B.6 and H.14). More surprisingly, the serotype pure B.24 and D.24 cultures show entirely different patterns although cultivated under identical environmental conditions. In contrast, the heat-shocked and the starved cultures still show similarity to their origin, the D.24 samples. To describe the relationship between the individual cultures and conditions, we analysed DEGs in more detail.

Using estimated gene read counts from the Sailfish analysis, we used the DESeq2 algorithm to compute pairwise DEGs between different conditions (see Section 2) and set a cut-off for DEGs at an FDR value of 0.01. Figure 2 shows complete numbers (Fig. 2A) and MA plots (Fig. 2B) of interesting comparisons as a detailed analysis of fold changes in relation to mean expression levels (further comparisons between conditions can be found in Supplementary Fig. S2). The plots support observed differences between the B.24 and D.24 cultures with 2,202 genes being significantly down-regulated in serotype 51D cells. As single 51B cells spontaneously appeared in the 51D cultures and were selected by immobilization, these differences suggest that the altered SA<sub>g</sub>, specific for the serotype, is not the only DEG but many more genes show altered gene expression. In contrast, the other two sets of variants of the D.24 condition, heat-shock and starvation, show less differences: in the former 587 up-regulated genes total in the latter case of starvation 3,123 genes; however, most of them showing a moderate down-regulation of  $\sim\log\text{-fold change } -2$ . The cold-adapted transcriptome of the B.6 cultures shows 1,709 down- and 1,472 up-regulated genes and therefore obviously a higher degree of alteration during cold adaptation compared with starvation. However, there is still a difference between the two sets of cold cultures B.6 and the H.14, which were adapted for a much longer time. In addition, the MA plot comparing A.31 and B.24 shows many genes that are differentially regulated although the total number of significantly DEGs is lower (639 up and 1,058 down) compared with the B.24 and D.24 comparison.

### 3.3. Comparison of individual transcriptome similarity

To get an overview of the global similarity between the different transcriptomes, we used hierarchical clustering (Euclidean distance, average linkage) on the log-transformed TPM estimates. Figure 3 shows the clustering result as a heatmap, indicating transcription levels of all genes clustered according to their common expression behaviour in all analysed libraries. The dendrogram above the heatmap reflects the overall similarity between the transcriptome profiles. Excluding the D.24 and D.starved cultures, this analysis clearly clusters the individual biological replicates although few replicates (e.g. H.14-2 or D.HS-2) show divergence as indicated by comparing transcriptomic landscapes in Fig. 1. Comparing the D.24 and D.starv cultures, our analysis indicates that these still have similar transcriptomes: the D.starv cultures have been separated from the well-fed D.24 cultures and maintained without feeding bacteria for 3 weeks (Fig. 1).

As a first insight in the general comparison of the transcriptomes, the indication from the MA plots can be confirmed, because the two serotypes B.24 and D.24 that were cultured both under identical environmental conditions show huge differences and are clearly separated in the dendrogram (Fig. 3). This suggests indeed that the differences observed here are not due to the environmental temperature or other conditions, e.g. food or composition of medium, which were identical for both. In support of the latter conclusion, the two cold-adapted

cultures B.6 and H.14 are clearly separated in Fig. 3 as the overall similarity of the transcriptome reveals large differences and clusters the B.6 cultures close to the B.24 conditions, from which they were derived.

At first glance, the separation of the B.6 and H.14 transcriptomes in the dendrogram appears to be in conflict with the fact that both share the lowest number of DEGs (Fig. 2A). In comparison to the distance-based clustering in Fig. 3, Supplementary Fig. S3 shows a clustering obtained using Pearson correlation of gene expression values (see Section 2.5). Using correlation allows comparing gene expression patterns, despite their absolute differences in expression. In agreement with the low number of DEGs in the B.6/H14 comparison, both transcriptomes are clustered together in the second cluster analysis. We therefore conclude that a common mechanism of cold adaptation occurred in both the H.14 and the B.6 cultures (indicated by the low number of DEGs and clustering by Pearson correlation), but that expression of many genes in the B.6 cultures remains similar to expression in the parental B.24 cultures (see further analysis of these genes below).

The lowest number of DEGs can be seen between D.24 and its heat-shocked portions D.HS. As the MA plot in Fig. 2 shows, the heat-shock is characterized mainly by strong activation of a certain number of genes, which explains the separation of the D.HS transcriptomes from the two other conditions of the serotype D cultures at 24°C in the dendrogram (Fig. 3).

Figure 3 apparently shows that all cultures undergoing an environmental change (B.6, D.starv and D.HS) still cluster to their originating cultures. Although we did not necessarily expect a very high number of DEGs in the heat-shocked cultures, one would have expected greater differences in starvation and especially in the B.6 cultures. This suggests that these changes in the 3-week period led only to a partial rearrangement of the previously governing gene expression patterns.

### 3.4. Cold adaptation by alterations of metabolic, DNA metabolic and translational alterations

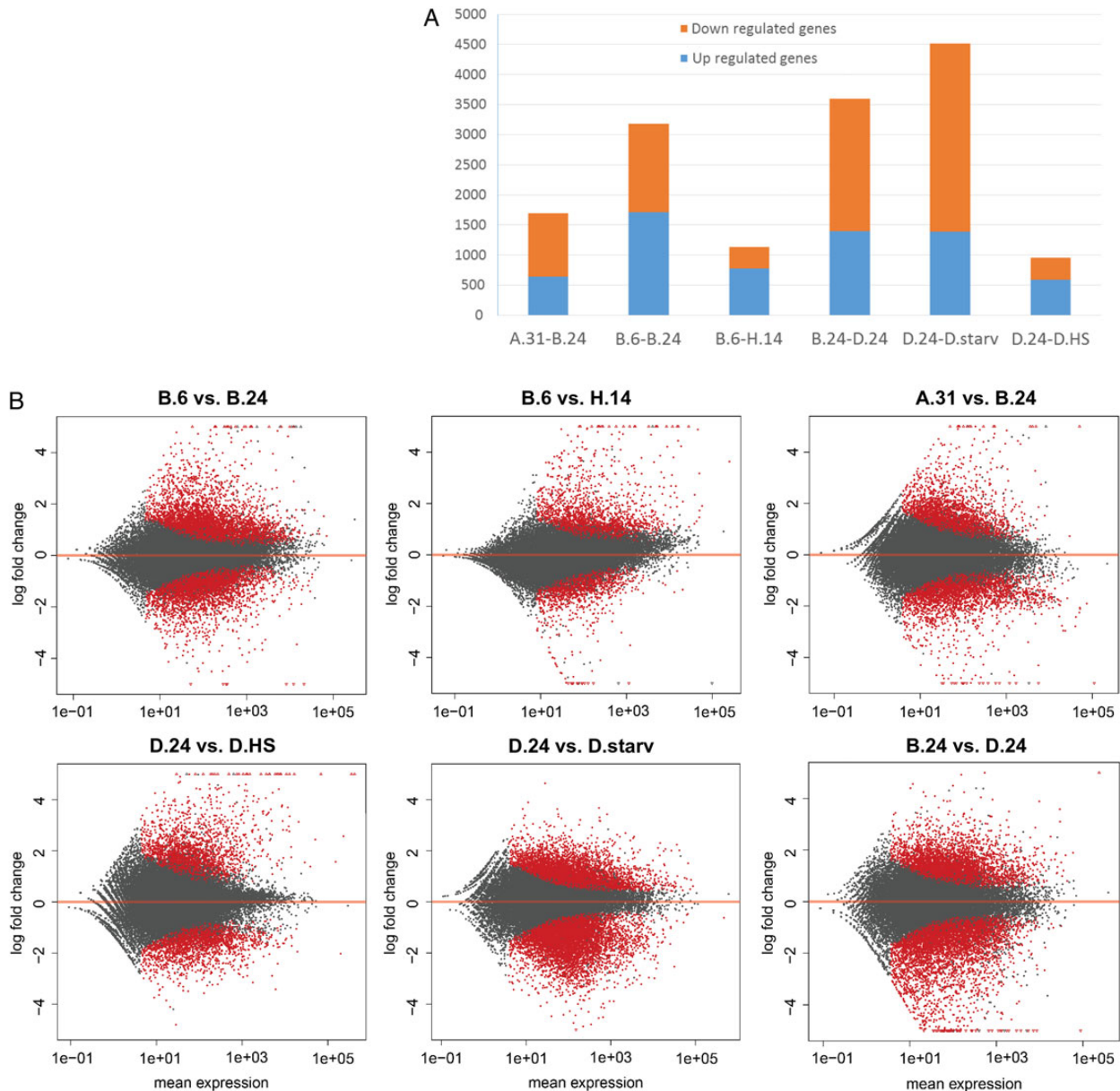
Figure 4 shows enrichment of GO terms for biological processes and individual comparisons (complete lists of differential GOs for these comparisons can be found in Supplementary File 5).

In all three warm/cold comparisons made here (B.6 vs. B.24; A.31 vs. B.6; A.31 vs. H.14), DNA metabolic GO terms including DNA biosynthesis and replication (initiation) are strongly up-regulated in the respective cold cultures. In addition, also oxidation/reduction processes are up-regulated in all cold-adapted cultures (only in the B.6 vs. B.24 comparison slightly below our set threshold). As this indicates higher energy metabolism, maybe due to the cold temperatures, this goes along with increased lipid catabolism in the B.6 cultures and with increased carbohydrate catabolism in H.14 cultures. As the representation of GO terms in Fig. 4 often reduced individual processes to parental GO terms (e.g. biological process), this indicates indeed that temperature adaptation alters many different processes, which require further investigation by subsequent experiments.

### 3.5. GO term enrichment in starvation, heat-shock and serotype comparison

The GO analysis indicates that starvation most significantly triggers microtubule-based processes and movement of cellular components, thus suggesting rearrangements of the cytoskeleton, probably because these starved cells do not undergo cell divisions. Up-regulated processes are DNA catabolism and lipid metabolism, indicating increased usage of lipids for energy metabolism.

In the heat-shocked cultures, the most significant up-regulated GO term is 'protein folding' (GO: 0006457, Supplementary File 5), which



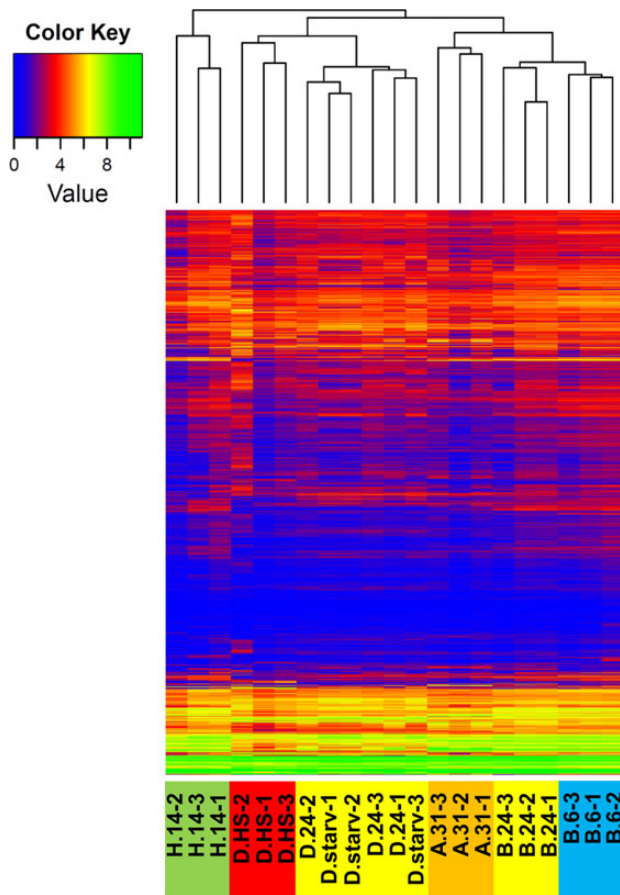
**Figure 2.** MA plots of mean differential gene expression. (A) The number of DEGs between individual comparisons ( $FDR \leq 0.01$ ). (B) MA plots of fold change expression level (y-axis) against expression level (x-axis). Each point represents a transcript; those with significant differential expression ( $FDR \leq 0.01$ ) are indicated in red. The cultures compared are indicated above each plot. A summary of all possible comparison between cultures can be found in the Supplementary Fig. 2.

involves the heat-shock response chaperones in agreement with our expectations. The GO enrichment displayed in Fig. 4 clearly shows that many other global processes are down-regulated. In the special situation of the heat-shock, these down-regulations may not be necessarily considered as controlled regulation of gene expression but also a by-product of thermal instability of individual mRNA species as discussed later.

In general, the B.24 and D.24 comparison shows that many processes are down-regulated: the most significant one is again ‘oxidoreductase activity’ (GO: 0016491), indicating that the D.24 cultures have a lower energy metabolism compared with the B.24 cells, despite identical environmental cultivation conditions and thus food supply. In this comparison, also GO terms concerning ‘chromatin assembly’ become apparent in Fig. 4. A closer look reveals that chromatin-

modifying enzymes are differentially expressed, namely genes of three histone deacetylase isoforms, chromodomain containing proteins and nuclear assembly proteins, and surprisingly the core histones themselves (Supplementary Fig. S4A). Several isoforms of histone genes H2A, H2B, H3 and H4 are down-regulated in the D.24 cultures. Supplementary Fig. S4B and C show, for histone H3, that only those isoforms are down-regulated that are constitutively expressed throughout the life cycle and that the expression of developmental specific isoforms is not affected.

Further wet lab work has to clarify whether the global down-regulation of the core histones in our RNA-seq data indeed indicates a genome-wide loss of nucleosome occupancy: histone gene expression was described to be complex and to occur at several levels



**Figure 3.** Relationship between transcriptomes of cultures and their replicates. Clustering heatmap showing expression levels (colour key on the left: green high and blue low expression) of the entire transcriptome per sample (bottom label). The dendrogram (top) indicates the global similarity of the transcriptome profiles using hierarchical clustering (Euclidean distance, average linkage) on the log-transformed expression values.

including post-transcriptional regulation.<sup>26</sup> Therefore, our data do not allow for conclusions that indeed histone protein levels are strongly regulated here. However, if lower nucleosome occupancy would be true in the D.24 cultures, the high number of DEGs to the B.24 cultures might also be the result of a complex re-organization of the macronuclear chromatin.

In yeast, cellular ageing was described to be accompanied by a loss of nucleosomes and a global transcriptional up-regulation,<sup>27</sup> which is different to our observations as the D.24 cultures show more down-regulated genes. As the cells of the D.24 cultures were not synchronized in cellular age, we cannot rule out an age-dependent regulation of histones, although the down-regulation is maintained in the D.starv cultures, which represent 10–20 division old F1 individuals of the D.24 cultures.

A similar result of decreased expression of the core histones was recently observed in *P. tetraurelia* by microarray analysis during silencing of an RNA-dependent RNA polymerase 3 (RDR3), which resulted in a co-expression of SAGs accompanied with decreased expression of the core histones<sup>12</sup> (M. Simon, unpublished data). As these cultures were of the same cellular age as the controls, any age-dependent influence can be ruled out in these experiments. Surprisingly, we observe a similar phenomenon of low core histone expression here in wild-type cultures of different serotypes. As *RDR3* silencing is also accompanied with drastically decreased cells divisions,<sup>11</sup> this

was also observed for D.24 cultures, which showed slightly fewer cell divisions per day compared with the B.24 cultures (data not shown). As reduced histone supply was described to extend the S phase in *Drosophila*, the delivery with core histones turns out to be a crucial component regulating the cell cycle.<sup>28</sup> Although the ciliate macronucleus divides amitotically, the prolonged cell cycle fits to our observations in the D.24 cultures and *RDR3*-silencing cultures.

In support of this conclusion, the heat-shocked derivatives of the D.24 samples show a down-regulation of the core histones (Supplementary Fig. S4). Although we did not analyse the division rate of the heat-shocked cells, the down-regulation of the histones is consistent with a recent characterization of the heat-shock in the unicellular algae *Chlamydomonas* showing decreased histone expression by proteomic analysis; in this particular case, cells show an immediate stop of cell divisions.<sup>29</sup>

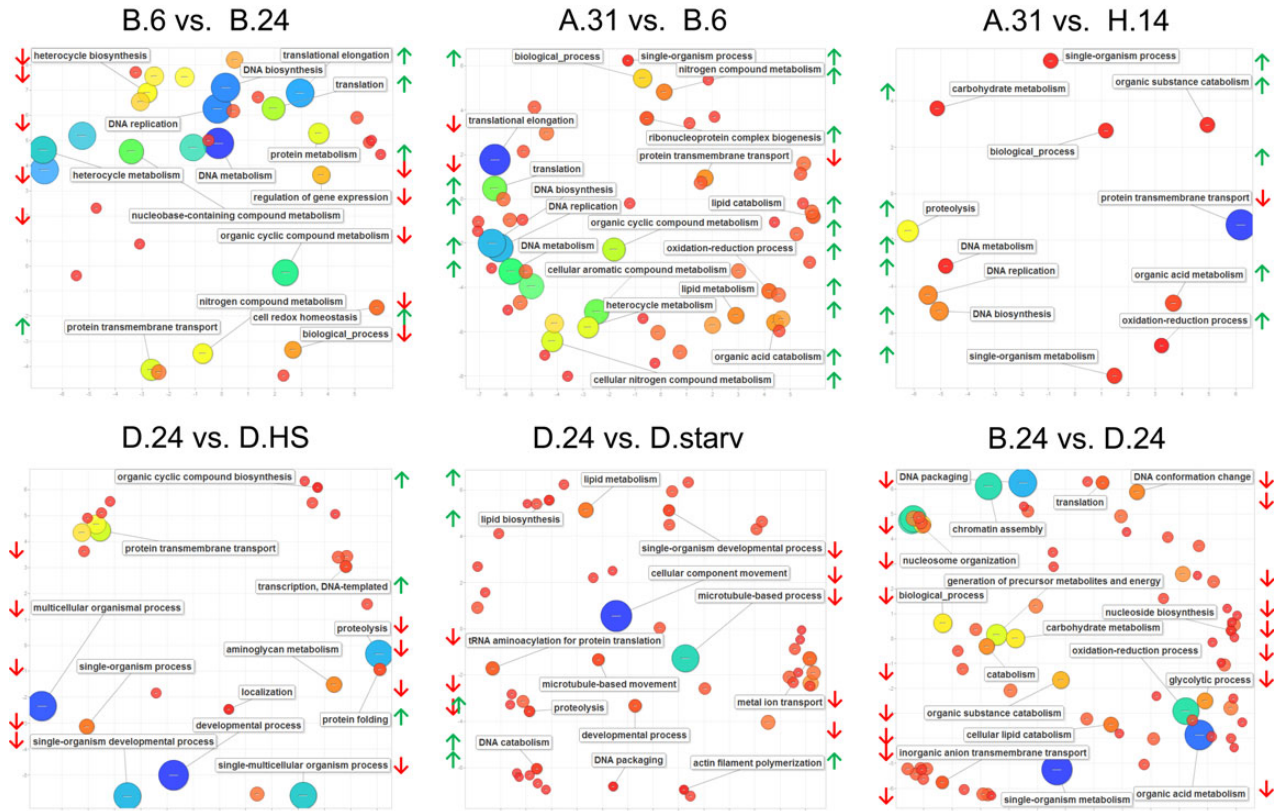
It seems tempting to speculate here that *Paramecium* controls cell division also by histone supply although the underlying regulatory mechanisms and the involvement remain unclear, especially the involvement of the *RDR3*-associated RNAi machinery. In this context, recent work in the ciliate *Stylonychia* demonstrated that the knock down of the *Stylonychia* Piwi resulted in down-regulation of individual H3 isoforms, thus providing another example for the connection of RNAi and histone expression.<sup>30</sup> Further studies have to clarify the role of histone gene expression and RNAi components as this would represent a powerful and new parameter of epigenetic influence to chromatin and genome integrity.

### 3.6. Constitutive and regulated expression of HSP70 isoforms

Ten HSP70 isoforms have been described earlier for *P. tetraurelia*,<sup>6,7</sup> and we added six more to the annotation, which were identified by a proteome-wide Pfam search (see Section 2 for details). Figure 5A shows the phylogenetic relationship between these isoforms and their putative subcellular localization. Figure 5B illustrates the TPM normalized expression levels of these HSP70 isoforms, where heat-inducible up-regulation can only be observed in the cytosolic group, in agreement with a previous analysis.<sup>7</sup> All other HSP70 isoforms show constitutive expression.

The only cytosolic HSP70 isoform that shows high expression in all cultures is HSP70Pt-01, and interestingly this isoform also shows a second expression peak in D.24 and D.starv cultures. This does not seem to be due to the cultivation temperature, as lower expression can be observed at the same cultivation temperature in B.24 cultures (Fig. 5B). Similarly, also HSP70Pt-05 shows this activation in D.24 and D.starv cultures, however to a lower degree.

Interestingly, a continuous up-regulation of HSP70 isoforms was recently also reported in *Paramecium* in response to endosymbionts, another form of external stress to the cell. The transcriptome comparison of *Paramecium bursaria* without and with endosymbiotic *Chlorella* ssp. algae revealed individual HSP70 isoforms up-regulated,<sup>31</sup> and similar findings were also reported for infections with bacterial endosymbionts in *Paramecium caudatum*.<sup>32</sup> Other studies suggest that bacterial symbiont bearing paramecia show an increased stress resistance,<sup>33</sup> and the authors concluded that permanent HSP70 activation awards increased stress tolerance. Interestingly, symbiont-induced expression of HSP70 isoforms is shown to be permanent even after the removal of bacteria by antibiotics, indicating an epigenetic manifestation of this acquired gene expression.<sup>34</sup> Here, our results support a permanent activation of an individual HSP70 isoform (HSP70Pt-01) in D.24 as well as the D.starv cultures but not in other samples.



**Figure 4.** GO enrichment analysis results. Principal component analysis (PCA) scatter plots of GO representatives of biological processes generated by the REVIGO tool that summarizes the list of differential expressed GO terms ( $FDR \leq 0.02$ ) by removing redundant terms.<sup>25</sup> The distance between circles (representing individual GO terms) indicates the relationship between terms: closer distance means closer relationship. Bubble colour indicates significance of differential expression of an individual GO term (red low and blue high); the size (in  $\log_{10}$   $P$ -value) indicates the percentage of genes annotated with a term in the reference database (UniProt) and thus indicates more general terms (large) and more specific ones (small). Most significant terms are labelled in the individual plot as well as representatives for groups of terms with lower significance. Green and red lateral arrows indicate the direction of regulation (up or down).

### 3.7. Expression pattern of the surface antigen multigene family

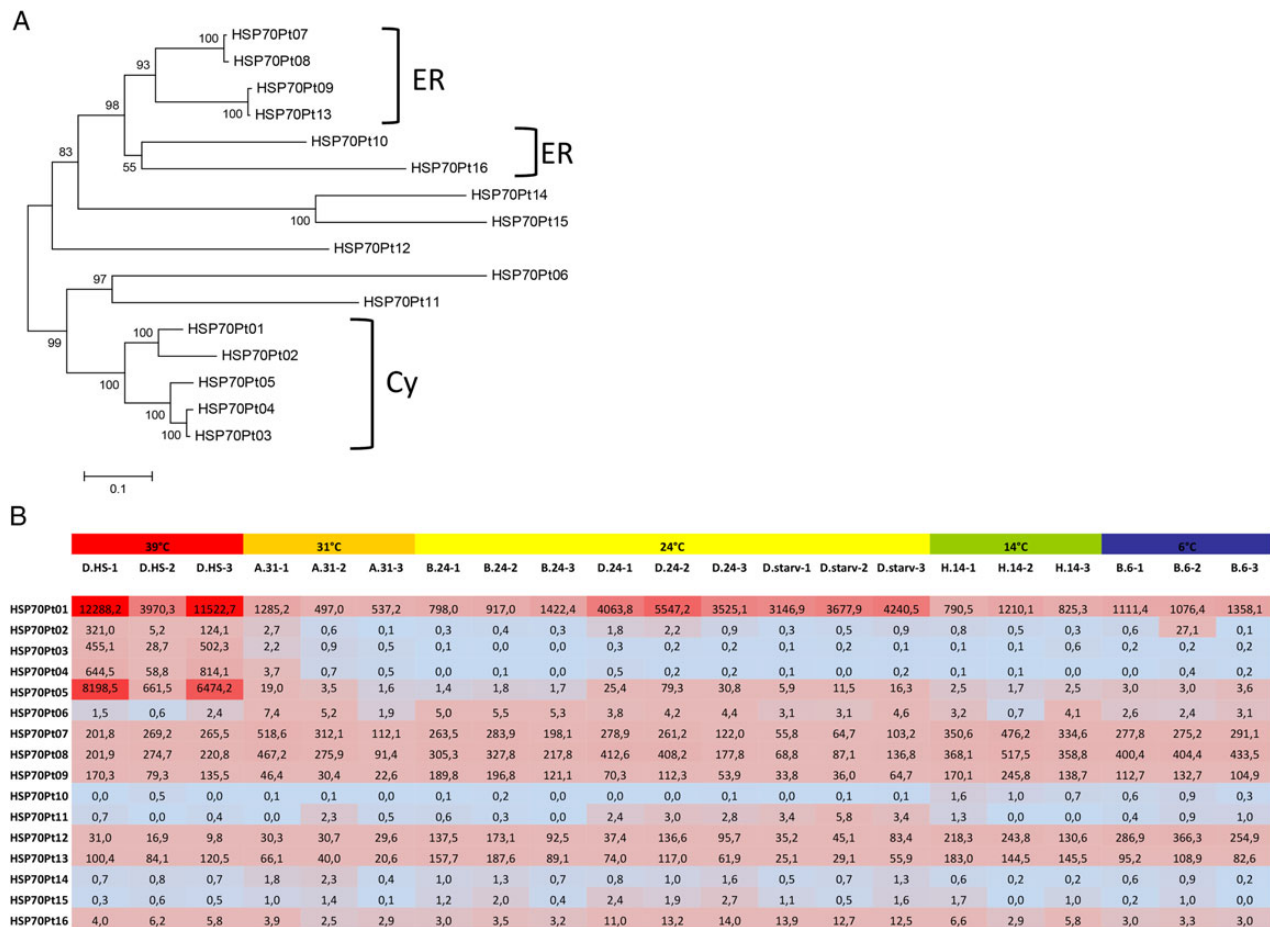
As mentioned above, serotypes represent a special kind of differentiation process manifesting a gene expression pattern of the SAg family by epigenetic mechanisms: only one gene is expressed at a time.<sup>35</sup> The most dominant environmental factor triggering serotype expression was described to be the temperature, and once induced they can be transferred to different cultivation conditions without serotype switches, thus indicating a self-stabilizing gene expression mechanism. Recent studies demonstrated that the core family consists of eight genes, some of them with a certain number of isoforms, indicated by Greek letters.<sup>12</sup> In all our experiments we started with serotype pure cultures, meaning 100% of the cells responded to specific antiserum in the immobilization reaction. Figure 6A shows the gene expression levels based on TPM normalization. The figure shows high expression levels of the individual SAg that corresponds to the antiserum with which the cells reacted. This is not surprising for the A.31, B.24, D.24 and H.14 cultures. However, the B.6 cultures, which represent an aliquot of the B.24 cultures, still retain high expression of the 51B gene. This means that the expression of the 51B gene shows stable expression, even after 3 weeks at a temperature of 6°C. Typically, *P. tetraurelia* stock 51 expresses the 51H gene at low temperatures.<sup>36</sup> Our data show that the 51H gene indeed has increased expression in the B.6 cultures, nevertheless the isolated cells still responded 100% to the 51B antiserum. This means that the actual serotype persists even

drastically temperature changes. From our experience, it is likely that some individuals of the B.6 cells will indeed shift to 51H after several weeks more of cultivation at 14°C or 6°C, but production of serotype pure cultures will require the selection of 51H expressing individuals. In such a case, serotype shifts can indeed be triggered by temperature alterations, but the selection of individual transformants is required.

Comparing the B.24 and B.6 transcriptomes in the heatmap representing the entire transcriptome, Fig. 3 indicates that many genes maintain their expression pattern, thus explaining the general high similarity. Additionally, the D.24 cultures are still highly similar to the D.starv cultures, and the serotype expression here indicates permanent expression of the 51D gene although slightly reduced (Fig. 6A). For these two examples of environmental alterations (food limitation: D.24 vs. D.starv) and extreme temperature shift (B.24 vs. B.6), our data show that the transcriptomes are indeed much more similar compared with the selected individuals of spontaneous transformants, namely the selected 51B expressing cells from the D.24 cultures that were cultivated under identical conditions.

This suggests that spontaneously occurring transcriptome alterations, which may be regulated by epigenetic mechanisms similar to the serotype system, alter gene expression in *Paramecium* to a much higher degree compared with short-time adaptation to environmental changes.





**Figure 5.** Expression level of heat-shock (HSP70) genes. (A) Neighbour-joining tree of HSP70 proteins of *Paramecium tetraurelia* (with 1,000 bootstrap replicates). The tree includes 10 previously described HSP70 genes (HSP70Pt01-10)<sup>6,7</sup> and additionally identified HSP70 isoforms (HSP70Pt11-16). Putative localization of isoforms is based on similarity analysis to HSPs in other organisms and on the classification in Krenek et al.<sup>7</sup> ER-endoplasmic reticulum, CY-cytosol. (B) TPM normalized expression values for all described HSP70 isoforms in all samples are shown, and cells are coloured according to high (red) or low (blue) expression.

### 3.8. Many genes show co-regulation with the dominant SAg and sensitivity to heat-shocks

To see whether other genes are co-regulated with the surface antigens, we defined ‘regulated spots’ in the transcriptomes, which are characteristic sets of lowly or highly expressed genes per condition (Fig. 6B). Genes that are part of a spot show not only coherent up- or down-regulation in one individual sample, but moreover they are forming a group of co-regulated genes that behave similarly among all analysed conditions and samples.

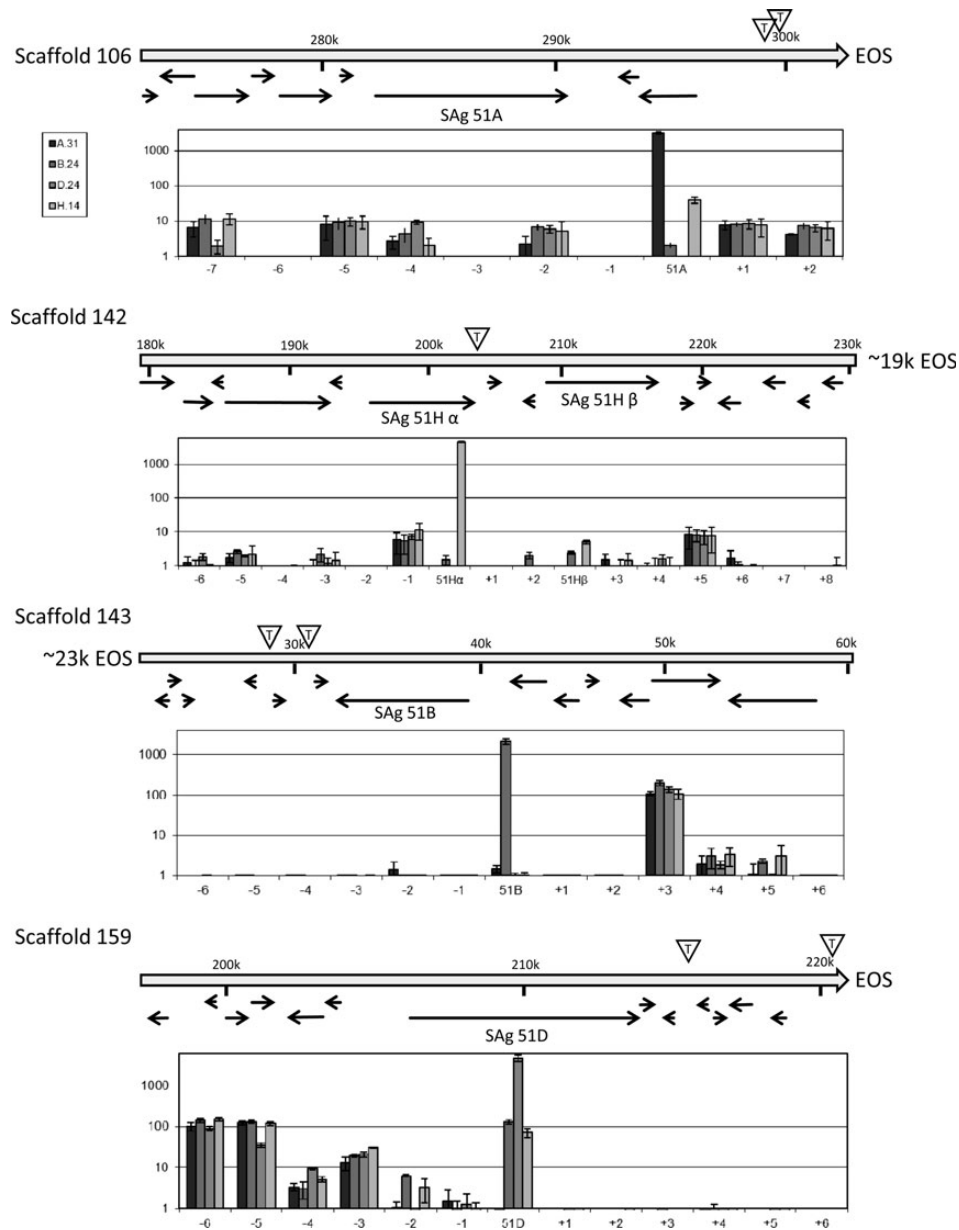
These data show that a huge number of genes are indeed co-regulated with individual antigens. These groups contain a large number of genes: on average between the replicates 2,468 genes in A.31, 1,180 in B.24 and 1,860 in D.24, thus representing a high percentage of the genome. In case of the B.24/D.24 comparison, the visualization of the regulated spots almost shows a mirror image of one another, thus explaining the clear separation of both states in the dendrogram (Fig. 3); it is still surprising that this difference is not due to environmental alterations but more likely due to spontaneous alterations. The short-time derivatives of these cultures (B.6 and D.starv) still show these spots (open arrow), although the pattern seems slightly different (Fig. 6B), thus indicating that a huge number of genes follows the expression behaviour of the SAg family and shows stability to environmental alterations within 3 weeks.

Surprisingly, only the heat-shocked cultures (D.HS) alter serotype expression and, moreover, show a loss of the highly expressed spot of regulated genes, which was characteristic for all 51D expressing cultures. This means that a heat-shock of 20 min alters the expression of genes, which we discussed to be co-regulated with the epigenetically controlled SAg family.

Comparing the transcript level of the cytosolic HSP70 isoforms per replicate to the expression levels of SAGs in Fig. 6C, the 51D mRNA is drastically reduced after 20 min heat-shock. This is apparent in the D.HS-1 and -3 cultures, and only to a lesser extent in the D.24-2 culture (only ~50% reduction). Further, comparing this with the HSP expression data, an inverse correlation becomes apparent: culture D.HS-2 also shows the lowest activation of the cytosolic HSP70 isoforms, thus indicating that this culture was only undergoing a moderate heat-shock but shows the lowest reduction of SAG 51D transcripts.

As a consequence, SAg expression levels seem to be sensitive for heat-shocks which is supported by studies in the close relative *Tetrahymena thermophila*, where mRNA of variable surface antigens (Ser genes) was shown to be sensitive to heat-shocks; however, not necessarily because of physical instability but by active degradation by a *de novo* synthesized protein factor.<sup>37,38</sup> This finding fits also very well to reports that a sudden change in temperature is much more efficient to induce serotype shifts compared with gradual changes.<sup>10</sup>





**Figure 7.** Genomic map and expression of SAG and neighbouring genes in subtelomeric regions. Genomic maps of subtelomeric regions of scaffold 106 (SAG 51A), scaffold 142 (SAG 51H), scaffold 143 (SAG 51B) and scaffold 159 (51D) indicate individual ORFs by arrows (data from *ParameciumDB*<sup>22</sup> and Baranasic et al.<sup>12</sup>). The position in the scaffold is indicated above in kbp (k). Expression level of the SAG and nearby genes (numbered by  $-X$  upstream and  $+X$  downstream relative to the SAG) are shown by TPM normalized data (logarithmic scale) in the graph below the genomic maps, respectively. Internal telomeric sites representing chromosome heterogeneity are indicated by the triangle containing the letter 'T'; EOS, end of scaffold.

### 3.10. Conclusions

Cell differentiation is mainly discussed in context of the ability of stem cells to differentiate into different tissues, i.e. an epigenetic mechanism manifests an individual gene expression pattern of the genome. Are these mechanisms for regulating and manifesting gene expression the only difference to phenotypic variation of unicellular eukaryotes?

*Paramecium* was demonstrated in the past to differentiate its vegetative genome into different phenotypes with one prominent example in serotypes being regulated by epigenetically controlled heritable gene expression patterns whose regulation involves small RNA-induced chromatin regulation<sup>12,13</sup> (Cheaib et al., in preparation). Interpretation of our transcriptome resource here indicates that not only the

surface antigen multigene family but also many more genes follow the expression pattern of SAGs and show persistence while transcriptome adaptation to environmental changes. This suggests that gene expression patterns of gene groups can be manifested by epigenetic mechanisms and consequently that also unicellular species can differentiate into distinct epigenetically controlled phenotypes. However, such differentiation should be interpreted as a kind of transcriptomic robustness and needs to be distinguished from regular adaptation to environmental changes representing short-time flexibility for adaptation.

It seems clear that our study can only estimate the extent of epigenetic differentiation by the above-discussed mechanisms, and the high number of genes co-regulated with the SAGs justifies future approaches

to clarify the epigenetic mechanisms controlling and maintaining gene expression. Of special interest here is the heat-shock, as the data indicate that this might represent a trigger for chromatin remodelling, not only to activate individual HSP genes but for remodelling many more chromosomal loci.

Our data suggest that serotypes cannot be restricted to the activation of a single gene as we identified a certain number of co-regulated genes, thus indicating that SAg expression is representative for individual gene expression profiles. *Paramecium* serotypes can be triggered by the environment or occur spontaneously (e.g. 51B cells in 51D cultures). It seems tempting to speculate that the co-regulated genes can be interpreted as a pre-installed genome program running to achieve a complex predetermined phenotype. We therefore need to clarify the epigenetic mechanisms allowing gene silencing, activation and maintenance of expression, to understand the epigenetic network that acts to coordinate the co-regulation of so many genes. Here, small RNA pathways, as indicated for the serotype system, would be an attractive hypothesis.

## Acknowledgements

The entire work was carried out in the context of the CNRS supported research group 'Paramecium Genome Dynamics and Evolution' and the European COST action BM1102 Ciliates as model systems to study genome evolution, mechanisms of non-Mendelian inheritance and their roles in environmental adaptation. We thank Emanuel Staudt for assistance with serotype pure cultures, Simone Marker for correcting HSP70 gene models, Gilles Gasparoni for kind support in coordination of sequencing, and Martina Schrällhammer for critical comments to the manuscript.

## Supplementary data

Supplementary data are available at [www.dnaresearch.oxfordjournals.org](http://www.dnaresearch.oxfordjournals.org).

## Funding

This study was supported by the starter grant by the Saarland University (61-cl/Anschub 2014/bew-Simon) and by grant SI 1397-2 by the Deutsche Forschungsgemeinschaft (DFG) to M.S. Funding to pay the Open Access publication charges for this article was provided by Saarland University.

## References

- Kelly, S.A., Panhuis, T.M. and Stoehr, A.M. 2012, Phenotypic plasticity: molecular mechanisms and adaptive significance, *Compr. Physiol.*, **2**, 1417–39.
- Wutz, A. 2013, Epigenetic regulation of stem cells: the role of chromatin in cell differentiation, *Adv. Exp. Med. Biol.*, **786**, 307–28.
- Vabulas, R.M., Raychaudhuri, S., Hayer-Hartl, M. and Hartl, F.U. 2010, Protein folding in the cytoplasm and the heat shock response, *Cold Spring Harb. Perspect. Biol.*, **2**, a004390.
- Fusco, G. and Minelli, A. 2010, Phenotypic plasticity in development and evolution: facts and concepts, *Philos. T. Roy. Soc. B.*, **365**, 547–56.
- Guschina, I.A. and Harwood, J.L. 2006, Lipids and lipid metabolism in eukaryotic algae, *Prog. Lipid Res.*, **45**, 160–86.
- Budin, K. and Philippe, H. 1998, New insights into the phylogeny of eukaryotes based on ciliate Hsp70 sequences, *Mol. Biol. Evol.*, **15**, 943–56.
- Krenek, S., Schlegel, M. and Berendonk, T.U. 2013, Convergent evolution of heat-inducibility during subfunctionalization of the Hsp70 gene family, *BMC Evol. Biol.*, **13**, 49.
- Cheuib, M. and Simon, M. 2013, Dynamic chromatin remodelling of ciliate macronuclear DNA as determined by an optimized chromatin immunoprecipitation (ChIP) method for *Paramecium tetraurelia*, *Appl. Microbiol. Biotechnol.*, **97**, 2661–70.
- Singh, D.P., Saudemont, B., Guglielmi, G., et al. 2014, Genome-defence small RNAs exapted for epigenetic mating-type inheritance, *Nature*, **509**, 447–52.
- Beale, G.H. 1957, The antigen system of *Paramecium aurelia*, *Int. Rev. Cytol.*, **6**, 1–23.
- Marker, S., Le Mouél, A., Meyer, E. and Simon, M. 2010, Distinct RNA-dependent RNA polymerases are required for RNAi triggered by double-stranded RNA versus truncated transgenes in *Paramecium tetraurelia*, *Nucleic Acids Res.*, **38**, 4092–107.
- Baranasic, D., Oppermann, T., Cheuib, M., Cullum, J., Schmidt, H. and Simon, M. 2014, Genomic characterization of variable surface antigens reveals a telomere position effect as a prerequisite for RNA interference-mediated silencing in *Paramecium tetraurelia*, *mBio.*, **5**, e01328–14.
- Simon, M. and Plattner, H. 2014, Unicellular eukaryotes as models in cell and molecular biology: critical appraisal of their past and future value, *Int. Rev. Cell Mol. Biol.*, **309**, 141–98.
- Beale, G.H. 1954, *The genetics of Paramecium Aurelia*, Cambridge University Press: New York.
- Martin, M. 2011, Cutadapt removes adapter sequences from high-throughput sequencing reads, *EMBnet J.*, **17**, 10–2.
- Punta, M., Coghill, P.C., Eberhardt, R.Y., et al. 2012, The Pfam protein families database, *Nucleic Acids Res.*, **40**, D290–301.
- Tamura, K., Stecher, G., Peterson, D., Filipowski, A. and Kumar, S. 2013, MEGA6: molecular evolutionary genetics analysis version 6.0., *Mol. Biol. Evol.*, **30**, 2725–9.
- Patro, R., Mount, S.M. and Kingsford, C. 2014, Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms, *Nat. Biotechnol.*, **32**, 462–4.
- Love, M.I., Huber, W. and Anders, S. 2014, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, *Genome Biol.*, **15**, 550.
- Benjamini, Y., Hochberg, Y. and Roy, J. 1995, Controlling the false discovery rate: a practical and powerful approach to multiple testing, *J. R. Stat. Soc. Ser. B. (Methodological)*, **57**, 289–300.
- Wirth, H., Loeffler, M., Bergen, M.V. and Binder, H. 2011, Expression cartography of human tissues using self organizing maps, *BMC Bioinformatics*, **12**, 306.
- Arnaiz, O. and Sperling, L. 2011, *ParameciumDB* in 2011: new tools and new data for functional and comparative genomics of the model ciliate *Paramecium tetraurelia*, *Nucleic Acids Res.*, **39**, D632–6.
- Bauer, S., Grossmann, S., Vingron, M. and Robinson, P.N. 2008, Ontologizer 2.0—a multifunctional tool for GO term enrichment analysis and data exploration, *Bioinformatics*, **24**, 1650–1.
- Aury, J.M., Jaillon, O., Duret, L., et al. 2006, Global trends of whole genome duplications revealed by the ciliate *Paramecium tetraurelia*, *Nature*, **444**, 171–8.
- Supek, F., Bošnjak, M., Škunca, N. and Šmuc, T. 2011, REVIGO summarizes and visualizes long lists of Gene Ontology terms, *PLoS ONE*, **6**, e21800.
- Ratray, A.M.J. and Müller, B. 2012, The control of histone gene expression, *Biochem. Soc. Trans.*, **40**, 880–5.
- Hu, Z., Chen, K., Yia, Z., et al. 2013, Nucleosome loss leads to global transcriptional up-regulation and genomic instability during yeast aging, *Genes Dev.*, **28**, 396–408.
- Günesdogan, U., Jäckle, H. and Herzig, A. 2014, Histone supply regulates S phase timing and cell cycle progression, *eLife*, **3**, e02443.
- Hemme, D., Veyel, D., Mühlhaus, T., et al. 2014, Systems-wide analysis of acclimation responses to long-term heat stress and recovery in the photosynthetic model organism *Chlamydomonas reinhardtii*, *Plant Cell*, **26**, 4270–97.
- Forcob, S., Bulic, A., Jönsson, F., et al. 2014, Differential expression of histone H3 genes and selective association of the variant H3.7 with a specific sequence class in *Stylonychia* macronuclear development, *Epigenetics Chromatin*, **7**, 1–15.
- Kodama, Y., Suzuki, H., Dohra, H., et al. 2014, Comparison of gene expression of *Paramecium bursaria* with and without *Chlorella variabilis* symbionts, *BMC Genomics.*, **15**, 183.
- Hori, M. and Fujishima, M. 2003, The endosymbiotic bacterium *Holospira obtusa* enhances heat-shock gene expression of the host *Paramecium caudatum*, *J. Euk. Microbiol.*, **50**, 293–8.

33. Fujishima, M., Kawai, M. and Yamamoto, R. 2005, *Paramecium caudatum* acquires heat-shock resistance in ciliary movement by infection with the endonuclear symbiotic bacterium *Holospora obtusa*, *FEMS Microbiol. Lett.*, **243**, 101–5.
34. Hori, M., Fujii, K. and Fujishima, M. 2008, Micronucleus-specific bacterium *Holospora elegans* irreversibly enhances stress gene expression of the host *Paramecium caudatum*, *J. Euk. Microbiol.*, **55**, 515–21.
35. Simon, M.C. and Schmidt, H.J. 2007, Antigenic variation in ciliates: antigen structure, function, expression, *J. Eukaryot. Microbiol.*, **54**, 1–7.
36. Godiska, R. 1987, Structure and sequence of the H surface protein gene of *Paramecium* and comparison with related genes, *Mol. Gen. Genet.*, **208**, 529–36.
37. Love, H.D. Jr, Allen-Nash, A., Zhao, Q.A. and Bannon, G.A. 1988, mRNA stability plays a major role in regulating the temperature-specific expression of a *Tetrahymena thermophila* surface protein, *Mol. Cell. Biol.*, **8**, 427–32.
38. McMillan, P.J., Stanley, J.S. and Bannon, G.A. 1995, Evidence for the requirement of protein synthesis and protein kinase activity in the temperature regulated stability of a *Tetrahymena* surface protein mRNA, *Nucleic Acids Res.*, **23**, 942–8.
39. Merrick, C.J. and Duraisingh, M.T. 2006, Heterochromatin-mediated control of virulence gene expression, *Mol. Microbiol.*, **62**, 612–20.
40. Duret, L., Cohen, J., Jubin, C., et al. 2008, Analysis of sequence variability in the macronuclear DNA of *Paramecium tetraurelia*: a somatic view of the germline, *Genome Res.*, **18**, 585–96.