Manuscript for submission to
Molecular Ecology as Original Article.

More effective transposon regulation in fertile, long-lived termite queens than in sterile workers

Running title: Caste-specific TE expression in a termite

Frederik Post[1,†], Erich Bornberg-Bauer[1], Mireille Vasseur-Cognet[2,3,4], Mark C. Harrison[1*]

1 Institute for Evolution and Biodiversity, University of Münster, 48149 Münster, Germany
2 UMR IRD 242, UPEC, CNRS 7618, UPMC 113, INRAE 1392, Paris 7 113, Institute of Ecology and Environmental Sciences of Paris, Bondy, France
3 University of Paris-Est, Créteil, France
4 INSERM, Paris, France
† Current affiliation: Proteomics Program, Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, 2200 Copenhagen, Denmark

*Corresponding author: m.harrison@uni-muenster.de

Abstract

Transposable elements (TEs) are mobile genetic sequences, which can cause the accumulation of genomic damage in the life time of an organism. The regulation of TEs, for instance via the piRNA-pathway, is an important mechanism to protect the integrity of genomes, especially in the germ-line where mutations can be transmitted to offspring. In eusocial insects soma and germ-line are divided among worker and reproductive castes, so one may expect caste-specific differences in TE regulation to exist. To test this, we compared whole-genome levels of repeat element transcription in the fat body of female workers, kings, and five different queen stages of the higher termite, *Macrotermes natalensis*. In this species, queens can live over 20 years, maintaining near maximum reproductive output, while sterile workers only live weeks. We found a strong, positive correlation between TE expression and the expression of neighbouring genes in all castes. However, we found substantially higher TE activity in workers than in reproductives. Furthermore, TE expression did not increase with age in queens, despite a seven-fold increase in gene expression, due to a significant upregulation of the piRNA-pathway in 20-year-old queens. Our results suggest a caste- and age-specific regulation of the piRNA-pathway has evolved in higher termites that is analogous to germ-line-specific activity in individual organisms. In the fat body of these termite queens, an important metabolic tissue for maintaining their extreme longevity and reproductive output, an efficient regulation of TEs likely protects genome integrity, thus further promoting reproductive fitness even at high age.

Keywords: Transposable elements, Eusocial insects, Ageing, Reproduction, Genomic integrity, Termites

Introduction

Almost 50 % of human, more than 50 % of maize, and about 20 % of *Drosophila melanogaster* genomes consist of transposable elements (TEs) (Haberer et al., 2005, Kannan et al., 2015, McCullers and Steiniger, 2017). TEs are DNA sequences that can be divided in LTR-retrotransposons, non-LTR-retrotransposons, and DNA transposons (J.Finnegan, 1992). TEs migrate and amplify in the genome by using insertion mechanisms (Gilbert et al., 2021). While retrotransposons only need to be expressed to become potentially active, DNA transposons require transposases for excision and reintegration (Lanciano and Cristofari, 2020). TEs may reintegrate essentially anywhere in the genome, including coding regions, introns, or regulatory regions. Insertion of TEs in genes or promotor regions can lead to loss of function or loss of expression of genes (Neafsey and Hartl, 2005). This disruptive nature of active TEs classifies them as mutagenic elements. However, TEs also drive genome evolution (Feschotte et al., 2002, Schrader and

Schmitz, 2019). TEs can be important for organisms, to the point that TEs become established genes. An example is RAG1 recombinase that is essential for the variety of light chains in B- and T-cell receptors (Kapitonov and Koonin, 2015). In insects, TEs were found to be responsible for adaptations that led to an evolutionary advantage to evade predators and develop resistance to viruses (Gilbert et al., 2021).

During the life time of individuals, TE activity and associated burdens can differ depending on, e.g., the individual's age and developmental stage. Moreover, TEs can have a fundamentally different effect whether they become active in somatic or germ cells (Haig, 2016). During differentiation of germ cells, epigenetic control mechanisms are weakened and hence increased expression of TEs from usually suppressed DNA is facilitated (Hajkova et al., 2002). Furthermore, TE expression in the germ line is evolutionarily advantageous for TEs, while genomic insertions of TEs in somatic cells are not transmitted to the next generation. PIWI-interacting RNAs (piRNAs) are a species of small RNA that silence TEs and are thought to be most active in germ-line tissues (Luo and Lu, 2017, Ozata et al., 2019), although activity in somatic tissues also appears to be prevalent in arthropods (Lewis et al., 2018). Three proteins from the PIWI clade of Argonaute proteins, Piwi, Aub, and Ago3, play an important role in the synthesis and amplification of piRNAs from TE transcripts. While Piwi is involved in the primary production of piRNAs and the silencing of TE transcription in the nucleus, Aub and Ago3 combine to amplify the production of piRNAs and cleave TE transcripts in the so-called "ping-pong" cycle (Luo and Lu, 2017, Siomi et al., 2011). Although this piRNA-pathway is Dicer-independent, the regulation of small interfering RNAs (siRNAs) and microRNAs (miRNAs) with the AGO clade of Argonaute proteins (Ago1 and Ago2), on the other hand, requires Dicer activity (Ozata et al., 2019). It is broadly accepted that the piRNA-pathway predominantly occurs in germ-line tissues, in order to protect the genome integrity of offspring, although activity in somatic tissues has been described (Castañeda et al., 2011).

In some insects and most eusocial insects, sterile and reproductive individuals exist, which resembles a separation of germ line and soma, posing the question of whether TE regulation may differ between phenotypes, as previously suggested for a higher termite (Elsner et al., 2018). But, so far, no substantial differences in TE expression were found between sterile and reproductive individuals of arthropods (Kraaijeveld et al., 2012, Petersen et al., 2019, Schaack et al., 2010). However, in head transcriptomes of the highly social termite *Macrotermes bellicosus* an increase of TE expression was found for old major workers, which was accompanied by a decrease of piRNA-pathway genes. No such pattern was found in queens (Elsner et al., 2018). In that study, only specific TEs were analysed that appeared in a list of differentially expressed genes, but these findings already offer the intriguing indication that TEs may be differentially regulated between castes and with age in higher termites.

In order to investigate the existence of caste-specific TE-regulation in higher termites, we analysed transcriptional levels of genome-wide repetitive content and TE regulatory genes in the fat body of

workers, kings and queens from five different stages of maturation in *Macrotermes natalensis*, in which repetitive content of the genome has been estimated at 45.6% (Korb et al., 2015). *M. natalensis* is a higher termite, which lives in large fungus-farming colonies. King and queen of a colony can live for more than 20 years. During maturation the queen abdomen becomes hypertrophic leading to greater reproductive output in older queens (Han and Bordereau, 1982), while workers are sterile and only live for a few weeks (Séité et al., 2022). We chose to analyse the fat body since we recently showed that it is an important tissue for the extreme longevity and reproductive output of queens in this species (Séité et al., 2022). During the maturation of termite queens a metabolic reprogramming of the fat body was reported, in which transcription, cell-cycle activity and lipid and carbohydrate metabolism increased (Séité et al., 2022). Furthermore, a functioning piRNA-pathway in the fat body of flies has been discovered, which is essential for reducing TE activity and has effects on longevity and lipid metabolism (Jones et al., 2016). We present evidence for a more efficient regulation of TEs in reproductive castes than in sterile workers and a strong reduction in the proportion of TE transcripts among all mRNAs in long-lived, highly fertile queens.

Materials and Methods

Repeat annotations

Repeat annotations on the *M. natalensis* genome (Poulsen et al., 2014) were carried out in a previous study (Korb et al., 2015). For more details on the methods, we refer to that publication. Briefly, homologous repeats were identified with RepeatMasker and RepeatProteinMask v4.0.1 (Smit et al., 2015) using Repbase v17.06 (Jurka et al., 2005). *De novo* repeats were annotated with PILER v1.0 (Edgar and Myers, 2005), LTRfinder v1.05 (Xu and Wang, 2007) and RepeatModeler v1.05 (Smit et al., 2015). TE sequences were classified with RepeatClassifier (Smit et al., 2015). Homologous and de novo TE annotations were combined into one non-redundant GFF. This repeat GFF was used in the present study.

Evaluation of repeat and gene expression

RNAseq data were retrieved from a recent study (Séité et al., 2022), in which, beside other analyses, gene expression was quantified for 25 samples of *M. natalensis*, including four female (minor) workers (FW), four virgin queens (Q0m), four 3-month-old queens (Q3m), three 9-month-old queens (Q9m), three 31-month-old queens (Q31m), four more than 20-year-old queens (Q20y), and three more than

20-year-old kings (K20y). The libraries were sequenced on a HiSeq 4000 Illumina sequencer, using 150bp paired-end chemistry. We refer to the original study for further details on mRNA extraction methods. These RNAseq data were mapped against the genome sequence of *M. natalensis* (Poulsen et al., 2014) (retrieved from `http://gigadb.org/dataset/100057`; accessed March 2019) with hisat2 v2.1.074 (Kim et al., 2019) at default settings. SAMtools was used to convert SAM to sorted BAM files (Li et al., 2009). Repeat expression levels were then quantified by counting reads that uniquely mapped to repeats contained in the repeat GFF using htseq-count v0.10.0 (Anders et al., 2015). Repeats, which overlapped with exons, were removed using BEDtools *intersect* (Quinlan and Hall, 2010). The distance between a repeat and a gene was determined using the BEDtools *closest* function (Quinlan and Hall, 2010). Protein coding gene expression was quantified in a similar manner by counting reads that mapped to gene regions annotated in the gene annotation GFF (Poulsen et al., 2014).

Statistical tests and visualization

Data analysis was performed in R version 3.6.1 (R Core Team, 2016) and read counts were normalised using DESeq2 (Love et al., 2014). Normalised read counts were filtered to contain only repeats with non-zero expression in at least one sample. Binning of repeats by distance to the closest gene was performed by summing the median repeat expression of each caste in 1 kb intervals in pre- and postgenic regions. If a repeat stretched over several bins, the affiliation was determined by the location of the middle of the repeat sequence. Expression of repeats that overlapped by at least 1 base pair with introns were summed in a bin called '0', corresponding to 0 kb distance to a gene. For the PCA in figure 2B, normalised expression of the top 500 repeats with the highest variance in intergenic regions (>10 kb distance to the closest gene) was used. To identify repeat elements which significantly differed in their expression between castes and queen stages, we carried out an analysis of variance (ANOVA). Multiple testing correction was applied with the Benjamini-Hochberg method. Significance was determined using a threshold of 0.05. In the heatmap displaying the results of this ANOVA, rows were clustered using Manhattan distance (2D). For the analysis of distance to genes of repeats in clusters of the heatmap, outliers with more than 1000 kb distance to genes were removed (2F). Figures were created in R version 4.1.1 (R Core Team, 2016) using ggplot2(v3.3.5), cowplot(v1.1.1), dplyr(v1.0.7), lemon(v0.4.5), viridis(v0.6.1), RColorBrewer(v1.1-2), ggokabeito(v0.1.0), tidyverse(v1.3.1), pheatmap(v1.0.12), and gridextra(v2.3) packages.

Results

Comparison of repeat expression among castes and queen stages

In our study, we conducted a transcriptomic analysis of repetitive element expression of 25 samples comprising female workers (FW), 20-year old kings (K20y), and queens from 5 stages of adult maturation (termed Q0m to Q20y) from 4 colonies of the higher termite, *M. natalensis* (Séité et al., 2022). These queen stages were sampled at colony foundation (Q0m), after 3 months (Q3m), 9 months (Q9m) and 31 months (Q31m) of colony life when queens became physiogastric, as well as mature queens which were over 20 years of age (Q20y; Fig. 1A; Séité et al. 2022). We summarised the total expression and number of uniquely expressed repeats and genes for each caste to obtain an overview of the overall transcription activity (Fig. 1). In total, 626,178 repeats were expressed across all samples. As a first striking observation, we found that the proportion of repeat expression among all transcripts differed strongly between castes. The proportion of repeat transcripts ranged from 22.6 to 25.2% in workers, from 20.5 to 23.9% in kings and from 17.4 to 21.9% in virgin queens (Q0m, Fig. 1B). The proportion of repeat transcription also varied greatly among queen stages, decreasing continuously with increasing queen age. Lowest proportions of repeat expression were found among 31-month-old (Q31m: 4.9-5.9%) and especially in 20-year-old queens (Q20y: 2.9-3.5%; Fig. 1B). We used a distance of 10 kb to a gene as the cut-off for splitting genes in two groups of either being located in genic or intergenic regions. For these genic and intergenic repeats, we analysed both the number of unique expressed repeat elements, as well as the total level of repeat transcription, measured in total numbers of reads mapping to repeat elements.

We found that the majority of repeat expression occurs in proximity to genes: a total of 82M reads mapped to 449K repeats (182.6x) within gene regions compared to only 28M reads and 178K repeats (157.3x) in intergenic regions, even though substantially more repeats (2.67 M) are annotated in intergenic than in gene regions (0.45 M). Outside of the influential range of genes, in intergenic regions, total repeat expression showed slightly different patterns than within gene bodies. However, both in intergenic and in gene regions repeat expression level was higher in workers than in reproductive castes (Fig. 1C & D, left). This difference was more pronounced in intergenic regions, where workers had about 30% more total counts than the reproductive castes, compared to around 15% more in gene regions. Otherwise, we observed little variation in repeat expression levels among reproductive samples.

We also investigated total numbers of unique repeat elements that were expressed in each caste and queen stage. The highest numbers of expressed TEs were observed for workers and Q0m, with lowest numbers in mature queens. While the number of expressed repeats in Q20y was at 75,000 in intergenic

regions, Q9m had about twice as many expressed repeats. Kings, which were of the same age as Q20y, had intermediate numbers of expressed repeats (Fig. 1D, right).

Next, we analysed total protein coding gene expression as a comparison to the repeat transcription patterns (Fig. 1E). We observed a strong increase in protein coding gene expression level from Q9m to the 20-year-old Q20y queens, which was about seven times higher than the expression of Q0m, Q3m, Q9m, kings and workers. Already at Q31m (31 months) gene expression increased threefold compared to the 9-month old Q9m queens. This is in stark contrast to the rather stable repeat expression in queens. However, comparable to the numbers of expressed repeats, the numbers of uniquely expressed genes was the lowest for Q31m and Q20y.

Analysis of expressed TE classes

We performed a PCA and an ANOVA to investigate how similarly repeats were expressed between castes and how repeat expression in queen stages developed over time (Fig. 2). The PCA revealed that repeat transcript abundance could be used to clearly separate workers from reproductive individuals, while also separating queens into three main maturation phases (Fig. 2A). Early stages (Q0m and Q3m) as well as late stages (Q31m and Q20y) of queen maturation each formed a distinct cluster. Intermediate queens (Q9m), on the other hand, clustered with mature kings. This pattern closely resembles the pattern found for gene expression in a previous study (Séité et al., 2022). However, K20y gene expression had an intermediate position between Q9m and mature queens (Q31m and Q20y). In order to control for the effect of gene expression, we repeated the PCA for intergenic repeats (> 10kb distance from genes) and still found a strong separation of workers and reproductive individuals. Clustering of kings (K20y) and queen stages was not as clear among intergenic repeat expression although the same tendencies remained (Fig. 2B).

We analysed how many different repeats were identified in each caste and queen stage grouped by their TE class LINE (class I long interspersed nuclear elements), LTR (class I long terminal repeats), DNA (class II DNA transposon), and RC (class II rolling circle) (Fig. 2C). The 'other' group contained all elements, which did not match the classes' criteria, e.g. simple repeats. For all castes and queen stages, the most numerous class was LINE, followed by other, DNA, LTR, and RC (Fig. 2C). Repeat classifications were similar among castes and queen stages, with the proportion of LINEs ranging from 52.7% in Q9m to 53.8% in workers.

Differentially expressed repeats

An ANOVA allowed us to identify 29,897 repeats which significantly differed in their expression

between castes and queen stages. We clustered these differentially expressed repeats (DERs) in a heatmap by their normalised expression levels as a Z-score (Fig. 2D). In these DERs, we found a disproportionately high proportion of worker (29.6%) and Q31m DERs (20.4%) compared to within the total number of expressed repeats (18.4% and 11.0%, respectively). Q3m and Q9m, on the other hand, were strongly under-represented in their proportions of DERs (5.2% and 9.0%) compared to within the full data set (14.6% and 16.6%). The DERs are likely a consequence of their proximity to differentially expressed genes (DEGs), since for each group of DERs, the proportion of the corresponding DEGs among the neighbouring genes ranged from 2.4 (Q9m) to 5.9 (Q3m) times higher than the corresponding proportion of DEGs in all genes (χ^2 = 79.8 - 757.5, df = 2, FDR < 4.1x10^{-19}). For example, 8.1% of all genes were significantly higher expressed in workers than in all other castes and stages but 24.9% of genes neighbouring the worker DERs are worker-specific. The repeat class distribution among these clusters of DERs (Fig. 2E) closely resembled the class distribution among all repeats (Fig. 2C). However, in Q9m the relative proportion of LINE TEs was reduced compared to the group 'other' repeats.

Additionally, we characterised the groups of DERs by the distances of the DERs to the closest gene (Fig.2 F). For Q20y, DERs were significantly closer to genes (median 2510 bp) than for all other castes and queen stages, except Q0m (Tukey post-hoc test, adjusted p-values < 1.1x10^{-6}). DERs that were specific to Q9m, on the other hand, were significantly further from genes (median 4776.5 bp) than in all other castes and queen stages (adjusted p < 1.3x10^{-4}).

To test whether DEGs were responsible for differences in observed proportions of repeats among castes and queen stages (Fig. 1B), we compared the expression levels of repeats close to genes with similar expression across samples. For this, we selected genes without significant differential expression between any combination of castes or queen-stages, but with a greater than median expression level to represent putative house-keeping genes (Fig. S2A). Expression levels of repeats close to these 195 putative house-keeping genes were similar among castes and queen stages, and in fact slightly higher in Q20y when only considering repeats within gene bodies (Fig. S2B). Interestingly, proportions of reads mapping to repeats were considerably higher within these putative house-keeping genes (32-50%; Fig. S2D) than in all genes (3-25%; Fig. 1B).

Relationship between repeat expression and gene expression

To investigate the relationship between gene proximity and repeat expression level, we summed the repeat read counts in 1 kb bins with increasing distance to a gene (Fig. 3). The bin of repeats located within gene bodies (0 kb distance), i.e. in introns, clearly had the highest total counts (range 19.2 to 19.9 log2 median counts; Fig. 3A). Repeat counts decayed with increasing distance from the closest genes, with pre- and postgenic regions showing similar declines. The repeat counts differed especially between workers

and the mature queen stages. While workers had the highest repeat counts in almost every bin, mature queens (Q31m and Q20y) mainly showed the lowest counts. We tested whether the observed pattern could be caused by mapping quality deteriorating with distance from genes due to an accumulation of assembly gaps in intergenic regions. Proportions of assembly gaps were lowest in the 1 kb bins but higher within introns (Fig. S1), so cannot explain the peak in repeat counts at 0 kb in figure 2A. As of 2 kb distance from genes, proportions of assembly gaps do not increase with distance from genes, indicating our observations represent levels of repeat transcripts rather than a systematic artefact (Fig. S1).

We performed linear regressions of repeat expression levels against the expression level of neighbouring genes per caste and queen stage. These regressions were performed in intronic regions (0 kb) and in 1 kb, 10 kb and 20 kb distances up- and downstream from genes (Fig. 3B). In general, the increase of repeat counts was lower than the increase of gene counts, resulting in slopes lower than 1. The steepest slope of repeat expression in relation to gene expression was found in introns. Q0m had the steepest slope (0.432) followed by Q9m (0.428), workers (0.404) and Q31m (0.397). Q20y had the least steep slope of all castes and queen stages (0.338 in introns). With increasing distance to the gene, the relationship between repeat and gene expression decreased strongly. While the slope at 0 kb ranged from 0.338 to 0.432, it had already decreased to 0.112 to 0.204 at 1 kb. A further strong decrease in slope was observed for repeats with 10 kb distance. At 20 kb from the nearest gene, the relationship between gene expression and repeat expression was barely detectable (0.011-0.062, only Q3m significant). During this decrease, Q0m and FW still had steeper slopes than the more mature castes.

Caste- and age-dependent expression of piRNA-pathway genes

Repeat transcript abundance may be altered by regulatory mechanisms, which include regulators of transcription, but also degradation of transcripts. Therefore, we analysed the transcript abundance of genes involved in the regulation of TEs in the piRNA pathway (Fig. 4). In *M. natalensis*, two Aub and three Ago genes have previously been identified (Elsner et al., 2018). Aub and Ago3 are the main players in the ping-pong cycle, which is responsible for TE degradation and silencing via the regulation of piRNAs (Iwasaki et al., 2015). Ago1 and Ago2, on the other hand, are likely to be involved in the regulation of other small RNAs, such as miRNAs and siRNAs, respectively (Meister, 2013). Overall, a tendency of higher transcript abundance in mature reproductives was observed for all five investigated genes. All three piRNA-pathway genes (Aub1, Aub2, Ago3) were significantly more highly expressed in Q20y than in workers and most younger queen stages. Ago 1 and Ago2, on the other hand, were highly expressed in Q31m, with Ago1 significantly lower expressed in Q20y compared to Q0m and workers. Kings only had increased levels of the Aub1 and Aub2 genes. The other genes in kings were lowly expressed and

resembled the expression in Q9m. Worker expression levels were comparable to the levels of young queens.

Discussion

Macrotermes natalensis queens can reach extremely high age (in excess of 20 years) while maintaining lifelong high reproductive output. Workers on the other hand are short-lived and sterile. Several studies suggest gene transcription plays an important role in this absence of a longevity/fecundity trade-off in termites (Elsner et al., 2018, Lin et al., 2021, Monroy Kuhn et al., 2019, 2021, Séité et al., 2022). One of these highlights the particular importance of fat body gene transcription in *M. natalensis* (Séité et al., 2022). In the present study, we were interested in understanding the importance of transposable element transcription, which can have profound mutagenic effects on genome stability thus exacerbating the effects of ageing (Sturm et al., 2015). We tested the hypothesis that queens efficiently regulate transposable element activity, thus maintaining fitness at high age, as previously suggested by Elsner et al. (2018). For this we compared transcriptomic activity of repeat elements in the fat body of 25 termite samples comprising two castes, both sexes and five stages of queen maturation of *M. natalensis*.

Repeat expression strongly related to gene expression

The majority of detected repeat element transcripts (53.3%) were LINEs, which belong to the non-LTR retrotransposons. We found a strong positive relationship between repeat element expression and gene expression in all castes and queen stages. The correlation of repeat and gene expression was especially high for repeats that resided in introns or up to 1 kb distance to a gene, but decreased strongly with distance from genes. This was further supported by high repeat expression in proximity to genes which decays strongly with increasing distance to the next gene. This finding suggests that TEs, as parasitic DNA sequences, hitch-hike actively transcribed genes. In this manner, TEs are able to utilise the transcription machinery and environment provided by a neighbouring gene.

Repeat expression profiles differ by caste and age

Using PCA, we were able to show that repeat expression patterns in workers differ greatly from those of reproductive castes. Also among reproductives, we identified three distinct clusters of repeat expression corresponding to developmental and reproductive trajectories. Young female reproductive stages, virgin queens (Q0m) and 3-month-old queens (Q3m), shared similar repeat expression profiles, while repeat expression of 9-month-old queens (Q9m) closely resembled that of mature kings (> 20 years). Mature queen stages (Q31m: 31-month-old; and Q20y: >20 years) formed the third distinct cluster. This

repeat expression pattern closely resembles a recently reported analogous gene expression pattern in this species (Séité et al., 2022), reconfirming our findings that repeat elements rely on the expression of neighbouring genes. However, our finding that the TE expression of mature kings more closely resembled Q9m TE expression rather than being intermediate between Q9m and mature queens as was found for gene expression, indicates differing levels of TE control in kings compared to queens. In order to detect further caste- and age-specific differences in repeat element expression that were independent of gene expression, we analysed intergenic repeats with at least 10 kb distance to any gene. We still found a strong difference in worker repeat expression patterns from those of reproductive individuals. Although clustering of different queen stages was not as apparent in intergenic repeat expression, a clear trend along a maturation axis from virgins to 20-year-old mature queens was observable. We also found substantially higher expression of repeat elements in workers than in reproductive individuals, which was even more pronounced in intergenic (+30%) compared to gene bodies (+15%). These results indicate fundamental differences exist in repeat element expression patterns between sterile and reproductive castes that are only partially explained by gene transcription.

The co-expression of repeat elements with neighbouring genes likely explains the nearly 30,000 repeat elements with significant caste- and stage-specific expression (DERs). This is confirmed by an enrichment of differentially expressed genes (DEGs) among neighbouring genes of these DERs, compared to proportions of DEGs among all genes. However, we detected differences in these DERs among castes and queen stages that cannot be entirely explained by their dependence on the expression of neighbouring genes. For instance, the strength of enrichment of DEGs in these DER groups differed among castes and queen stages, ranging from almost 6-fold in 3-month-old queens (Q3m) to 2.4-fold in 9-month-old queens (Q9m). Furthermore, the proportion of LINEs within DERs was substantially reduced in Q9m queens compared to within all repeats. Additionally, the proximity of DERs to genes differed strongly. In 20-year-old queens (Q20y), DERs were closest to genes, while Q9m DERs ranged furthest from genes compared to other queen stages. In the vicinity of putative house-keeping genes, on the other hand, relative levels of repeat expression were considerably higher and varied less among castes and queen stages. This result suggests TE silencing is reduced in euchromatic regions, in which house-keeping genes more often lie, compared to the less accessible heterochromatic regions, where DEGs are more likely to reside (Ganapathi et al., 2005).

These results underline the effect of varying gene expression on repeat element expression, leading to large differences in levels of transcription as well as in the types of repeats, depending on age and fertility in these termites. Nevertheless, they suggest additional gene-independent variation in repeat element expression exists in the studied termites, indicating caste- and age-specific differences in their ability to regulate repeat elements. We therefore investigated variation in TE regulation that may explain the

observed patterns.

Increased activity of TE regulatory genes in mature queens

We quantified caste- and age-specific expression of three genes within the piRNA-pathway (Aub1, Aub2 and Ago3) to investigate differences in TE regulation. Furthermore, we added two further Ago genes (Ago1 and Ago2) to the analysis that are not involved in TE regulation. We observed a general increase in expression of the piRNA pathway in mature queens and partially in kings. All three genes were higher expressed in Q20y than in workers and younger queens, while the two Aub genes also had elevated expression in kings. The importance of Aub in regulating TEs in older insects has previously been demonstrated in *D. melanogaster* where a knock-down of Aub led to increased TE expression in 25-day-old compared to 5-day-old individuals (Yang et al., 2022). The increase in Ago3 transcription appeared to be related to reproductive maturation in queens with highest levels in Q20y, but low levels in kings. This is supported by a study of mosquito ovaries in which high expression of Ago3 was observed after laying eggs (Macias et al., 2014). The low expression of Ago3 in kings allows to conclude that kings have increased capability of producing piRNAs by Aub1 and Aub2, but are lacking Ago3-mediated amplification of piRNAs in the fat body. Further investigations of young kings could shed light on the similarities of the kings' and queens' life cycle stages or maturation phases. If kings at the age corresponding to Q9m or Q31m have similar TE activity patterns as mature kings, it will indicate that ageing has minor effects on TE activity in termites compared to reproduction. Ago1 and Ago2 are post-transcriptional regulators by orchestrating siRNA- and miRNA-mediated mRNA silencing (Tomari et al., 2007). These conventional mRNA regulators did not show the same extent of increased expression in Q20y or kings, indicating that the piRNA-pathway in particular and not silencing mechanisms in general are upregulated. In contrast to Q20y and kings, Q31m seemed to be strongly impacted by mRNA silencing mechanisms which could be a response to the threefold increase in gene expression from Q9m to Q31m in the course of several months (Fig. 1E).

Superior TE regulation in mature queens compared to workers and young queens

We previously reported a strong increase in protein and lipid synthesis in mature queens, which coincides with metabolic reprogramming of the fat body and an increase in ploidy in fat body cells, along with increased oogenesis (Séité et al., 2022). Despite this enormous upregulation in gene transcription (up to 7-fold) in 20-year-old queens compared to all other castes and queen stages, repeat transcription levels were similar among all reproductives. This is further supported by the lowest portion of RNA-seq reads

mapping to repeats in Q20y (2.9-3.5%) of all studied castes and queen stages. Repeat read proportions were also low in Q31m queens (4.9-5.9%) but otherwise ranged from 17-25%. These results suggest that repeat elements are better regulated in Q20y than other queens stages or in workers, which is supported by the upregulation of Aub1, Aub2 and Ago3, which likely efficiently silence TE transcription.

Conclusions

The role of the piRNA-pathway in protecting genome integrity in germ cells by, among others, regulating transposon activity is well established for insects (Gonzalez et al., 2015, Ma et al., 2014). Colonies of higher termites, such as *M. natalensis*, can be considered superorganisms (Boomsma and Gawne, 2018), since somatic and germ functions are divided among castes. The increased regulation of TE activity we detected in mature queens, especially compared to sterile workers, is therefore analogous to the high piRNA activity in the germline of solitary animals. An important role of the piRNA pathway has also been recognised for the fat body in flies (Jones et al., 2016). In that study, flies with depleted piRNA activity in the fat body experienced greater TE mobility and shorter lifespan. Although effects of fertility were not measured in the study, reported reductions in energetic storage metabolites are likely to have negative impacts on fertility (Jones et al., 2016). Similarly, short-lived workers in our study have low Aub and Ago3 expression and high TE activity. The fat body of queens in *M. natalensis* is an essential tissue for the production of specific proteins and lipids to sustain maximal fertility (Séité et al., 2022). The development of physogastry in termite queens, at stage Q31m, coincides with increased reproductive output (Han and Bordereau, 1982) and a massive increase in gene expression (Séité et al., 2022). Our results suggest that higher termites may have evolved mechanisms to increase the regulation of TEs in the fat body along with queen maturation. This increased activity in the piRNA-pathway likely allows the increased gene transcription which is essential for the greater reproductive activity and heightened metabolism, while also protecting the genome against ageing stress during their long life.

Acknowledgements

This study was supported by the International Human Frontier Science Program RGP0060/2018335 to M.V.-C. and E.B.-B. We are grateful to David Sillam-Dussès and Alain Robert for field assistance and the establishment of incipient colonies, and to Elias Dohmen for help with python scripts.

References

Anders, S., Pyl, P. T., and Huber, W. (2015). HTSeq—a python framework to work with high-throughput sequencing data. *Bioinformatics*, 31(2):166–169.

Boomsma, J. J. and Gawne, R. (2018). Superorganismality and caste differentiation as points of no return: how the major evolutionary transitions were lost in translation. *Biological Reviews*, 93(1):28–54.

Castañeda, J., Genzor, P., and Bortvin, A. (2011). piRNAs, transposon silencing, and germline genome integrity. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 714(1-2):95–104.

Edgar, R. C. and Myers, E. W. (2005). PILER: identification and classification of genomic repeats. *Bioinformatics*, 21(suppl_1):i152–i158.

Elsner, D., Meusemann, K., and Korb, J. (2018). Longevity and transposon defense, the case of termite reproductives. *Proceedings of the National Academy of Sciences*, 115(21):5504–5509.

Feschotte, C., Jiang, N., and Wessler, S. R. (2002). Plant transposable elements: where genetics meets genomics. *Nature Reviews Genetics*, 3(5):329–341.

Ganapathi, M., Srivastava, P., Sutar, S. K. D., Kumar, K., Dasgupta, D., Pal Singh, G., Brahmachari, V., and Brahmachari, S. K. (2005). Comparative analysis of chromatin landscape in regulatory regions of human housekeeping and tissue specific genes. *BMC bioinformatics*, 6(1):1–10.

Gilbert, C., Peccoud, J., and Cordaux, R. (2021). Transposable elements and the evolution of insects. *Annual Review of Entomology*, 66(1):355–372.

Gonzalez, J., Qi, H., Liu, N., and Lin, H. (2015). Piwi is a key regulator of both somatic and germline stem cells in the drosophila testis. *Cell Reports*, 12(1):150–161.

Haberer, G., Young, S., Bharti, A. K., Gundlach, H., Raymond, C., Fuks, G., Butler, E., Wing, R. A., Rounsley, S., Birren, B., Nusbaum, C., Mayer, K. F., and Messing, J. (2005). Structure and architecture of the maize genome. *Plant Physiology*, 139(4):1612–1624.

Haig, D. (2016). Transposable elements: Self-seekers of the germline, team-players of the soma. *BioEssays*, 38(11):1158–1166.

Hajkova, P., Erhardt, S., Lane, N., Haaf, T., El-Maarri, O., Reik, W., Walter, J., and Surani, M. (2002). Epigenetic reprogramming in mouse primordial germ cells. *Mechanisms of Development*, 117(1-2):15–23.

Han, S. H. and Bordereau, C. (1982). Origin and formation of the royal fat body of the higher termite queens. *Journal of morphology*, 173(1):17–28.

Iwasaki, Y. W., Siomi, M. C., and Siomi, H. (2015). PIWI-interacting RNA: Its biogenesis and functions. *Annual Review of Biochemistry*, 84(1):405–433.

J.Finnegan, D. (1992). Transposable elements. *Current Opinion in Genetics & Development*, 2(6):861–867.

Jones, B. C., Wood, J. G., Chang, C., Tam, A. D., Franklin, M. J., Siegel, E. R., and Helfand, S. L. (2016). A somatic pirna pathway in the *Drosophila* fat body ensures metabolic homeostasis and normal lifespan. *Nature communications*, 7(1):1–9.

Jurka, J., Kapitonov, V. V., Pavlicek, A., Klonowski, P., Kohany, O., and Walichiewicz, J. (2005). Repbase update, a database of eukaryotic repetitive elements. *Cytogenetic and genome research*, 110(1-4):462–467.

Kannan, S., Chernikova, D., Rogozin, I. B., Poliakov, E., Managadze, D., Koonin, E. V., and Milanesi, L. (2015). Transposable element insertions in long intergenic non-coding RNA genes. *Frontiers in Bioengineering and Biotechnology*, 3.

Kapitonov, V. V. and Koonin, E. V. (2015). Evolution of the RAG1-RAG2 locus: both proteins came from the same transposon. *Biology Direct*, 10(1).

Kim, D., Paggi, J. M., Park, C., Bennett, C., and Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature biotechnology*, 37(8):907–915.

Korb, J., Poulsen, M., Hu, H., Li, C., Boomsma, J. J., Zhang, G., and Liebig, J. (2015). A genomic comparison of two termites with different social complexity. *Frontiers in Genetics*, 6.

Kraaijeveld, K., Zwanenburg, B., Hubert, B., Vieira, C., Pater, S. D., van Alphen, J. J. M., den Dunnen, J. T., and de Knijff, P. (2012). Transposon proliferation in an asexual parasitoid. *Molecular Ecology*, 21(16):3898–3906.

Lanciano, S. and Cristofari, G. (2020). Measuring and interpreting transposable element expression. *Nature Reviews Genetics*, 21(12):721–736.

Lewis, S., Quarles, K., Yang, Y., Tanguy, M., Frézal, L., Smith, S., Sharma, P., Cordaux, R., Gilbert, C., Giraud, I., Collins, D., Zamore, P., Miska, E., Sarkies, P., and Jiggins, F. (2018). Pan-arthropod analysis reveals somatic pirnas as an ancestral defence against transposable elements. *Nature, Ecology and Evolution*, 2(1):174–181.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Subgroup, . G. P. D. P. (2009). The sequence alignment/map format and samtools. *Bioinformatics*, 25(16):2078–2079.

Lin, S., Werle, J., and Korb, J. (2021). Transcriptomic analyses of the termite, *Cryptotermes secundus*, reveal a gene network underlying a long lifespan and high fecundity. *Communications biology*, 4(1):1–12.

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(550).

Luo, S. and Lu, J. (2017). Silencing of transposable elements by piRNAs in *Drosophila*: an evolutionary perspective. *Genomics, Proteomics & Bioinformatics*, 15(3):164–176.

Ma, X., Wang, S., Do, T., Song, X., Inaba, M., Nishimoto, Y., ping Liu, L., Gao, Y., Mao, Y., Li, H., McDowell, W., Park, J., Malanowski, K., Peak, A., Perera, A., Li, H., Gaudenz, K., Haug, J., Yamashita, Y., Lin, H., quan Ni, J., and Xie, T. (2014). Piwi is required in multiple cell types to control germline stem cell lineage development in the *Drosophila* ovary. *PLoS ONE*, 9(3):e90267.

Macias, V., Coleman, J., Bonizzoni, M., and James, A. A. (2014). piRNA pathway gene expression in the malaria vector mosquito *Anopheles stephensi*. *Insect Molecular Biology*, 23(5):579–586.

McCullers, T. J. and Steiniger, M. (2017). Transposable elements in *Drosophila*. *Mobile Genetic Elements*, 7(3):1–18.

Meister, G. (2013). Argonaute proteins: functional insights and emerging roles. *Nature Reviews Genetics*, 14(7):447–459.

Monroy Kuhn, J. M., Meusemann, K., and Korb, J. (2019). Long live the queen, the king and the commoner? Transcript expression differences between old and young in the termite cryptotermes secundus. *PLoS One*, 14(2):e0210371.

Monroy Kuhn, J. M., Meusemann, K., and Korb, J. (2021). Disentangling the aging gene expression network of termite queens. *BMC genomics*, 22(1):1–17.

Neafsey, D. E. and Hartl, D. L. (2005). Convergent loss of an anciently duplicated, functionally divergent RH2 opsin gene in the fugu and tetraodon pufferfish lineages. *Gene*, 350(2):161–171.

Ozata, D. M., Gainetdinov, I., Zoch, A., O'Carroll, D., and Zamore, P. D. (2019). Piwi-interacting RNAs: small RNAs with big functions. *Nature Reviews Genetics*, 20(2):89–108.

Petersen, M., Armisén, D., Gibbs, R. A., Hering, L., Khila, A., Mayer, G., Richards, S., Niehuis, O., and Misof, B. (2019). Diversity and evolution of the transposable element repertoire in arthropods with particular reference to insects. *BMC Ecology and Evolution*, 19(1).

Poulsen, M., Hu, H., Li, C., Chen, Z., Xu, L., Otani, S., Nygaard, S., Nobre, T., Klaubauf, S., Schindler, P. M., Hauser, F., Pan, H., Yang, Z., Sonnenberg, A. S. M., de Beer, Z. W., Zhang, Y., Wingfield, M. J., Grimmelikhuijzen, C. J. P., de Vries, R. P., Korb, J., Aanen, D. K., Wang, J., Boomsma, J. J., and Zhang, G. (2014). Complementary symbiont contributions to plant decomposition in a fungus-farming termite. *Proceedings of the National Academy of Sciences*.

Quinlan, A. R. and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6):841–842.

R Core Team (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Schaack, S., Pritham, E. J., Wolf, A., and Lynch, M. (2010). DNA transposon dynamics in populations of *Daphnia pulex* with and without sex. *Proceedings of the Royal Society B: Biological Sciences*, 277(1692):2381–2387.

Schrader, L. and Schmitz, J. (2019). The impact of transposable elements in adaptive evolution. *Molecular Ecology*, 28(6):1537–1549.

Séité, S., Harrison, M. C., Sillam-Dussès, D., Lupoli, R., Dooren, T. J. M. V., Robert, A., Poissonnier, L.-A., Lemainque, A., Renault, D., Acket, S., Andrieu, M., Viscarra, J., Sul, H. S., de Beer, Z. W., Bornberg-Bauer, E., and Vasseur-Cognet, M. (2022). Lifespan prolonging mechanisms and insulin upregulation without fat accumulation in long-lived reproductives of a higher termite. *Communications Biology*, 5(1).

Siomi, M. C., Sato, K., Pezic, D., and Aravin, A. A. (2011). Piwi-interacting small RNAs: the vanguard of genome defence. *Nature reviews Molecular cell biology*, 12(4):246–258.

Smit, A., Hubley, R., and Green, P. (2013-2015). Repeatmasker open-4.0. `http://www.repeatmasker.org`.

Sturm, A., Ivics, Z., and Vellai, T. (2015). The mechanism of ageing: primary role of transposable elements in genome disintegration. *Cellular and molecular life sciences*, 72(10):1839–1847.

Tomari, Y., Du, T., and Zamore, P. D. (2007). Sorting of *Drosophila* small silencing RNAs. *Cell*, 130(2):299–308.

Xu, Z. and Wang, H. (2007). LTR FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic acids research*, 35(suppl_2):W265–W268.

Yang, N., Srivastav, S. P., Rahman, R., Ma, Q., Dayama, G., Li, S., Chinen, M., Lei, E. P., Rosbash, M., and Lau, N. C. (2022). Transposable element landscapes in aging *Drosophila*. *PLoS genetics*, 18(3):e1010024.

Data Accessibility and Benefit-Sharing Section

Data Accessibility Statement

Python and R scripts are available from a dedicated github (`github.com/MCH74/TEsInMnat/`) and the data sets necessary for performing analyses can be downloaded from dryad (`doi.org/10.5061/dryad.18931zd1`). Additional support can be requested from the corresponding author.

Benefit-Sharing Statement

Benefits from this research accrue from the sharing of our data and results on public databases as described above.

Authors' contributions

M.C.H. conceived the project, M.V.-C. provided data and support on biological interpretations. F.P. carried out all bioinformatics analyses, assisted by M.C.H. F.P. and M.C.H. wrote the first manuscript draft. All authors contributed to writing the manuscript and interpreting the findings.

Figures

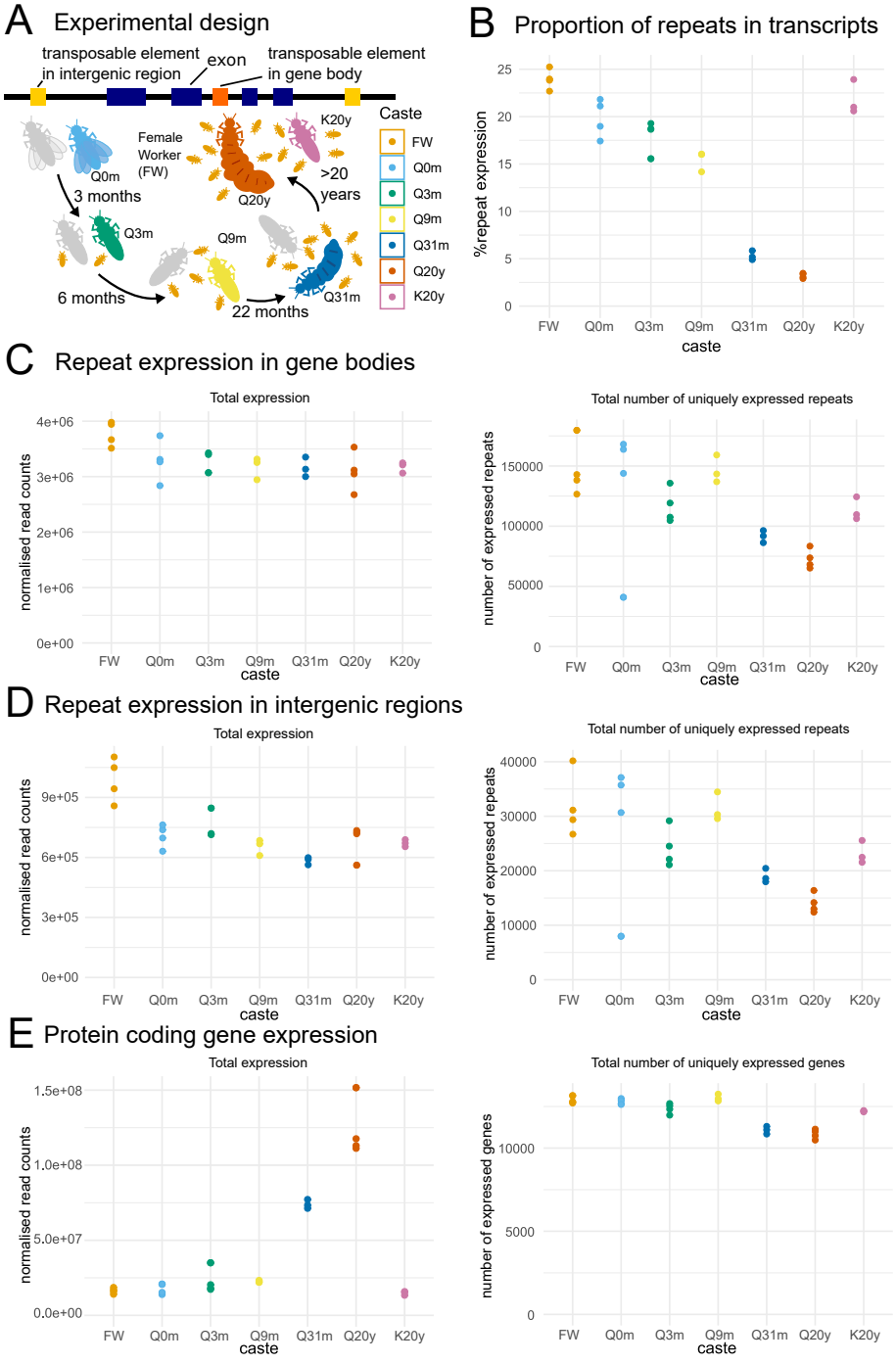

Figure 1: Repeat expression in different termite castes and queen stages. **A** Experimental design of repeat classification and sampling in the life cycle of termites. **B** Relative repeat expression among total RNA expression. **C** Total repeat expression in gene bodies (< 10 kb distance to closest gene). Left: total normalised read counts; Right: total number of unique repeats expressed in each sample. **D** Total repeat expression in intergenic regions (> 10 kb distance to closest gene). Left: total normalised read counts; Right: total number of unique repeats expressed in each sample. **E** Total protein coding gene expression. Left: total normalised read counts; Right: total number of unique repeats expressed in each sample. FW = female worker; Q0m to Q20y = queen stages; king = K20y.

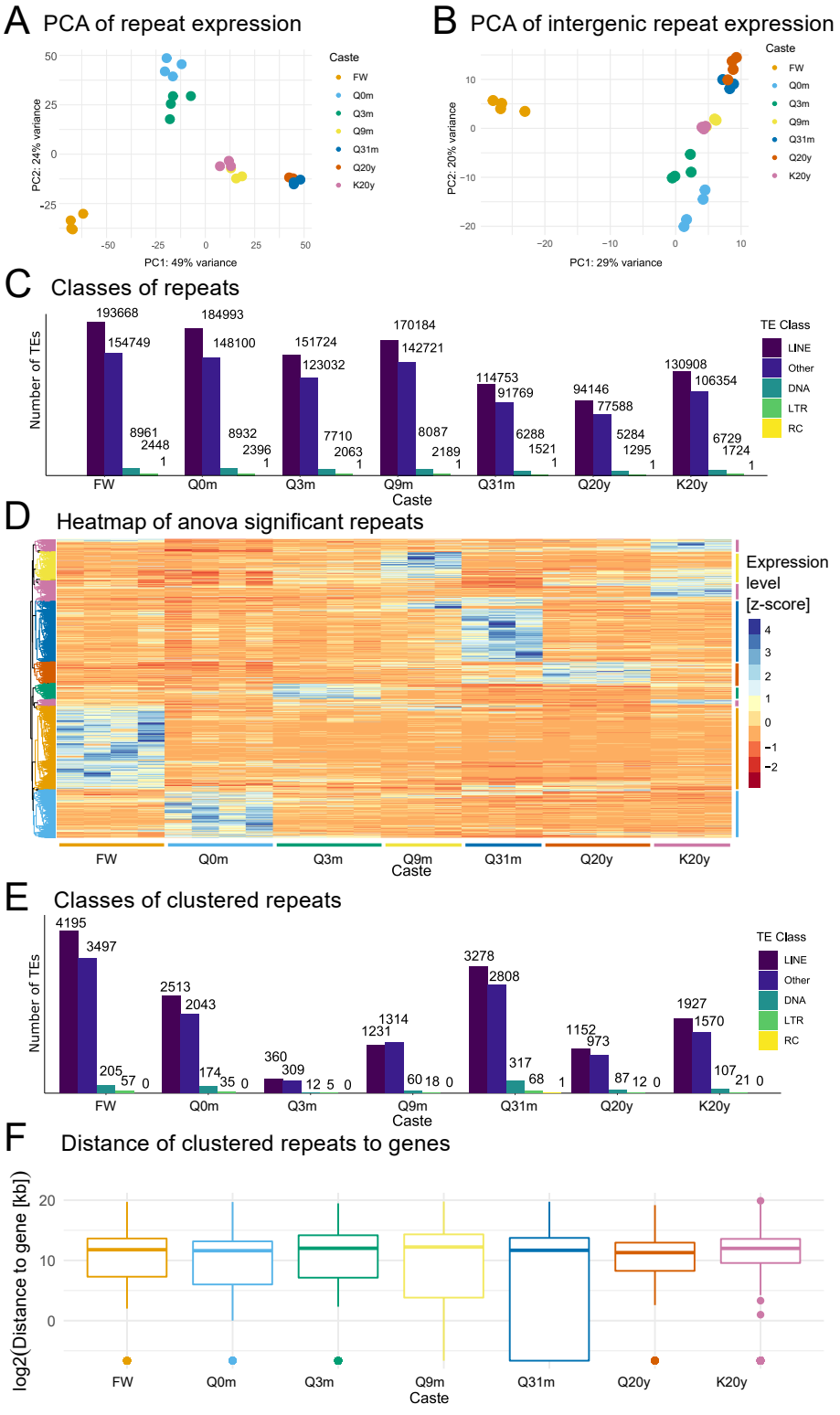

Figure 2: Differential repeat expression in castes of *M. natalensis*. **A** Principal component analysis (PCA) of repeat expression. **B** Principal component analysis (PCA) of intergenic repeat expression (>10 kb distance to the closest gene). **C** Number of expressed repeats by class. **D** Heatmap of repeats that significantly differ in their expression between castes (ANOVA). Columns show expression of these significant repeats in each individual, rows are clustered by expression similarity to caste- and stage-specific clusters using the Manhattan method. **E** Number of expressed repeats by classes of heatmap clusters. **F** Distance of differentially expressed repeats from heatmap clusters to the closest gene.
FW = female worker; Q0m to Q20y = queen stages; king = K20y.

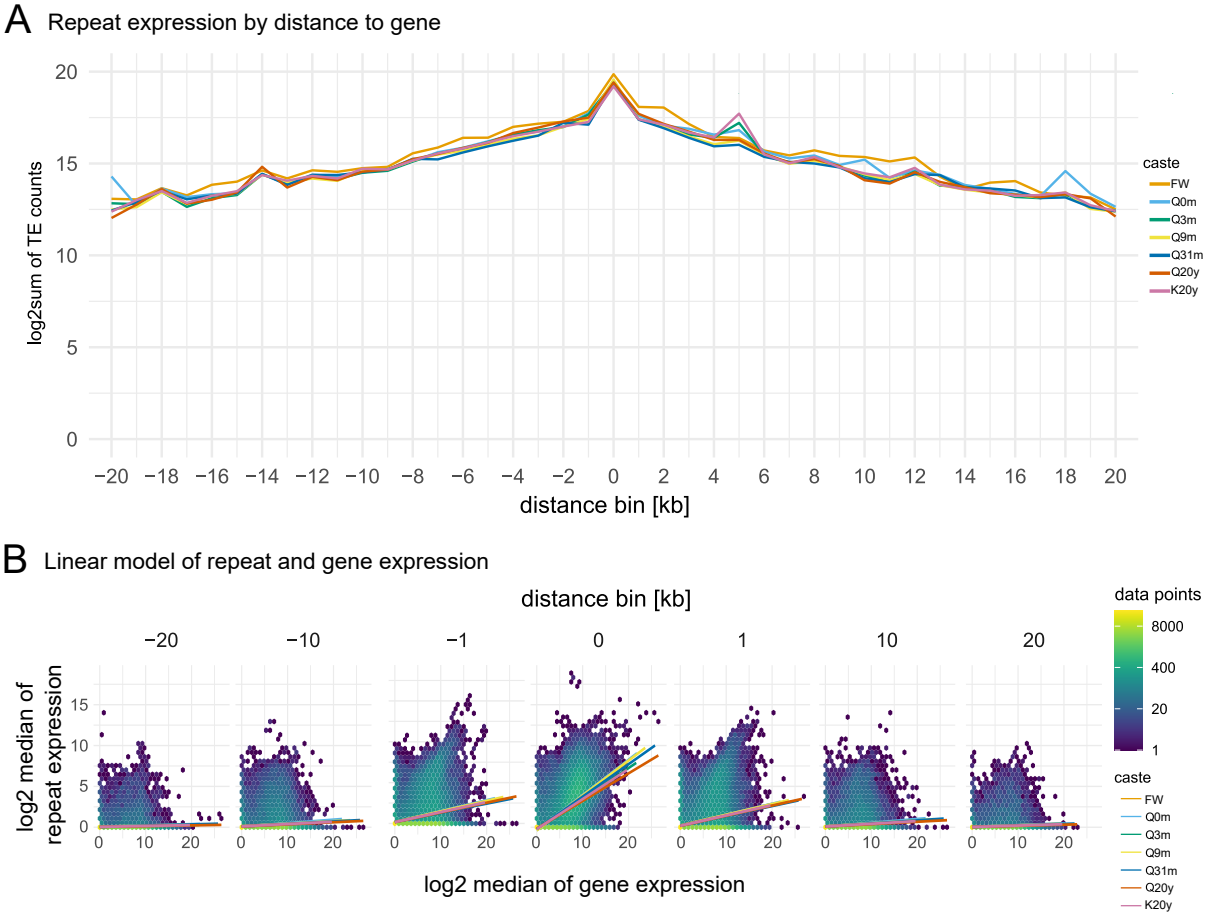

Figure 3: Analysis of repeat expression relative to the distance to the next gene. **A** Total repeat expression of repeats overlapping with genes or in 1kb bins up to 20 kb into pre- and postgenic regions. **B** Linear model of repeat and gene expression between castes of binned repeats within introns (0), or 1 kb bins with a distance of 1 kb, 10 kb and 20 kb pre- and postgenic. Data points were summed in hexagons and the density visualised by color scaling.
FW = female worker; Q0m to Q20y = queen stages; king = K20y.

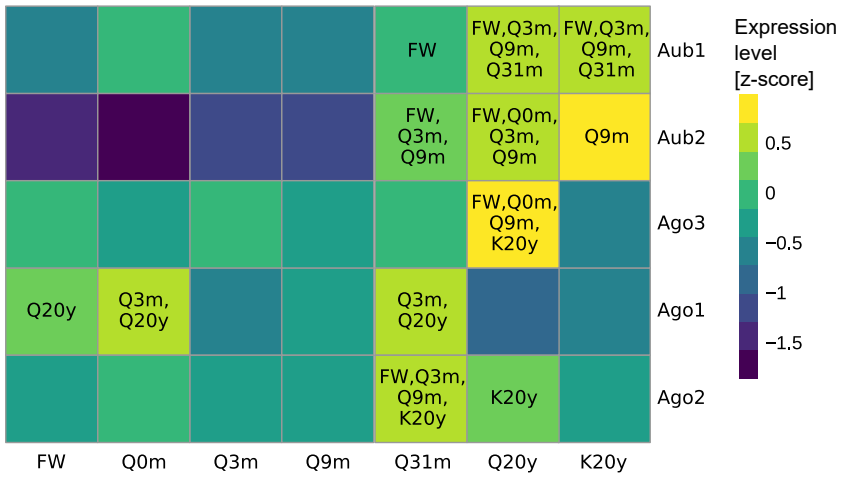

Figure 4: Expression level (as z-score) of piRNA-pathway genes by caste and queen stage. Colours signify relative intensity of expression. Significant differences in expression are denoted by inserted labels, that show the caste or queen stage, in which expression is lower - e.g. Aub1 is significantly higher expressed in Q31m vs. FW.
FW = female worker; Q0m to Q20y = queen stages; king = K20y

Supporting information

Supplementary figures

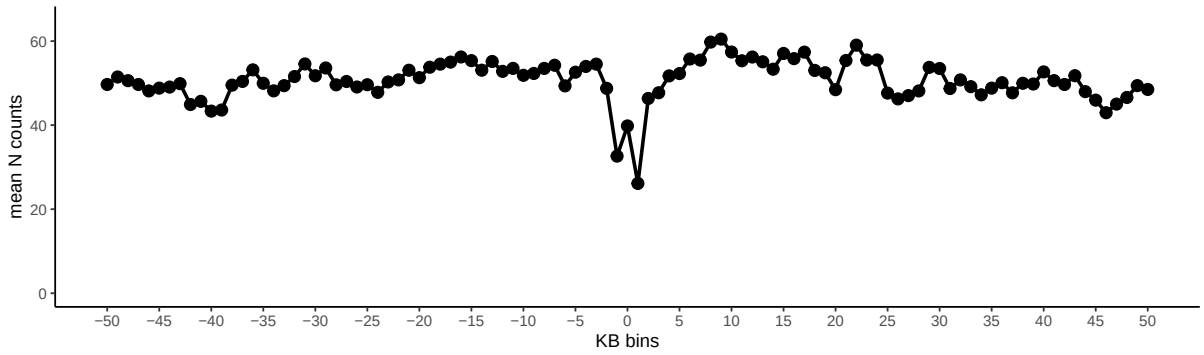

Figure S1: Mean proportions of gaps (Ns) in the *M. natalensis* assembly in varying distances (1kb bins) from genes.

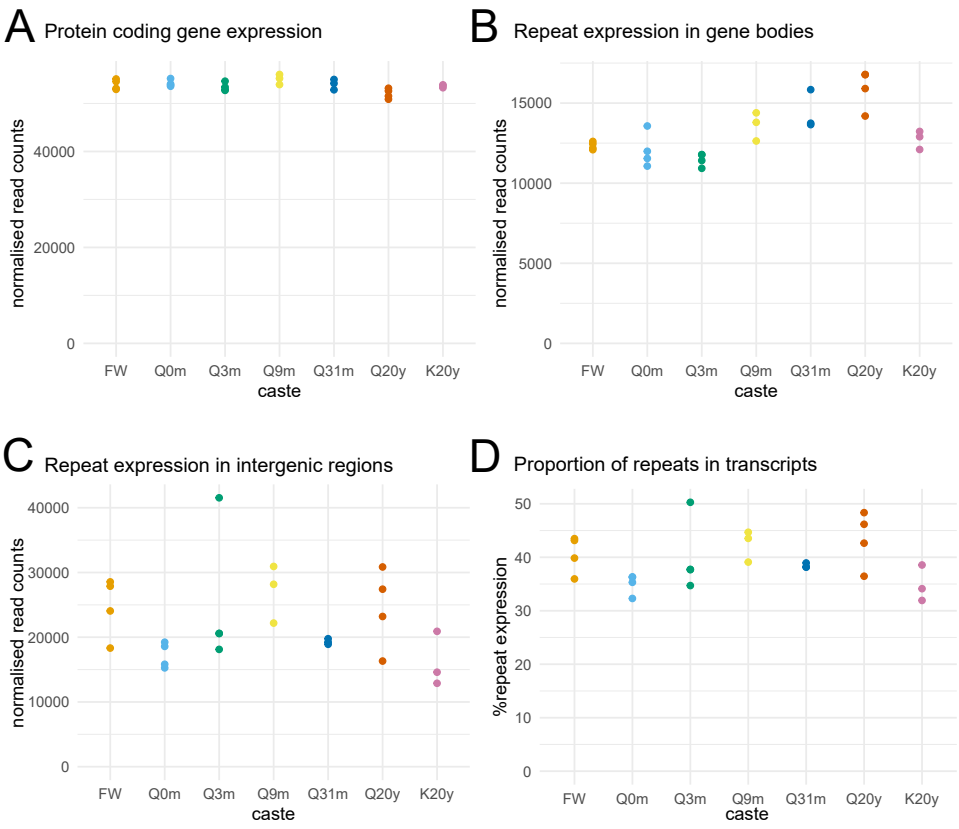

Figure S2: Expression levels of house-keeping genes and neighbouring repeats.

Supplementary tables

Table S1: RNAseq sample information, including sample ID, colony ID and mapping rates.

Sample ID	Caste	Colony ID	Mapping rate
AAAC	FW	3	0.701
AAAE	FW	6	0.700
AAAF	FW	7	0.759
AABE	FW	5	0.758
AAAG	Q0m	3	0.806
AAAH	Q0m	5	0.718
AAAI	Q0m	6	0.799
AAAJ	Q0m	7	0.773
AAAK	Q3m	3	0.744
AAAL	Q3m	5	0.737
AAAM	Q3m	6	0.778
AAAN	Q3m	7	0.541
AAAP	Q9m	7	0.770
AACH	Q9m	Blue	0.723
AACI	Q9m	Green	0.701
AACE	Q31m	3	0.794
AACF	Q31m	7	0.808
AACG	Q31m	6	0.831
AAAQ	Q20y	3	0.803
AAAR	Q20y	5	0.806
AAAS	Q20y	6	0.727
AAAT	Q20y	7	0.798
AAAU	K20y	5	0.781
AAAV	K20y	6	0.786
AAAW	K20y	7	0.780