BMC Genomics

Disentangling the aging gene expression network of termite queens

José Manuel Monroy Kuhn[1,2*], Karen Meusemann[1,3] and Judith Korb[1*]

Abstract

Background: Most insects are relatively short-lived, with a maximum lifespan of a few weeks, like the aging model organism, the fruit-fly *Drosophila melanogaster*. By contrast, the queens of many social insects (termites, ants and some bees) can live from a few years to decades. This makes social insects promising models in aging research providing insights into how a long reproductive life can be achieved. Yet, aging studies on social insect reproductives are hampered by a lack of quantitative data on age-dependent survival and time series analyses that cover the whole lifespan of such long-lived individuals. We studied aging in queens of the drywood termite *Cryptotermes secundus* by determining survival probabilities over a period of 15 years and performed transcriptome analyses for queens of known age that covered their whole lifespan.

Results: The maximum lifespan of *C. secundus* queens was 13 years, with a median maximum longevity of 11.0 years. Time course and co-expression network analyses of gene expression patterns over time indicated a non-gradual aging pattern. It was characterized by networks of genes that became differentially expressed only late in life, namely after ten years, which associates well with the median maximum lifespan for queens. These old-age gene networks reflect processes of physiological upheaval. We detected strong signs of stress, decline, defense and repair at the transcriptional level of epigenetic control as well as at the post-transcriptional level with changes in transposable element activity and the proteostasis network. The latter depicts an upregulation of protein degradation, together with protein synthesis and protein folding, processes which are often down-regulated in old animals. The simultaneous upregulation of protein synthesis and autophagy is indicative of a stress-response mediated by the transcription factor *cnc*, a homolog of human *nrf* genes.

Conclusions: Our results show non-linear senescence with a rather sudden physiological upheaval at old-age. Most importantly, they point to a re-wiring in the proteostasis network and stress as part of the aging process of social insect queens, shortly before queens die.

Keywords: RNASeq, Transcriptomes, Ageing, Social insects, Weighted gene co-expression networks, WGCNA, Time series, Termite, Lifespan, Senescence

* Correspondence: manuel.kuhn@helmholtz-muenchen.de;
judith.korb@biologie.uni
[1]Department of Evolutionary Biology & Ecology, Institute of Biology I, Albert
Ludwig University of Freiburg, Hauptstr. 1, D-79104 Freiburg (i. Brsg.),
Germany
Full list of author information is available at the end of the article

Monroy Kuhn *et al. BMC Genomics*   (2021) 22:339

Page 2 of 17

Background

Almost all animals age, but at a different pace [1]. The fruit fly *Drosophila melanogaster* lives only for around seven weeks [2], while the clam Ocean Quahog, *Arctica islandica*, can have a lifespan of more than 400 years [3], and the giant barrel sponge *Xestopongia muta* can live more than two millennia [4]. Generally, organisms with large differences in aging rates are found between widely divergent species [1, 5], which makes controlled comparisons of the underlying aging mechanisms difficult.

Classical model organisms typically have a short lifespan and can be characterized by r-life history strategies ('live fast, have many offspring and die young') as exactly these traits make them good model organisms. Social insects such as termites, ants, or the honeybee, offer promising new insights into aging because individuals with the same genetic background can differ by orders of magnitudes in lifespan. Within a social insect colony, which is generally a large family, the reproducing queen (and in termites, also the king) can reach lifespans of more than 20 years, while non-reproducing workers have a lifespan of a few months only [6–9]. However, quantitative demographic data covering the whole lifespan of queens are inherently rare (for ants: [10–12] and references therein; for termites: [13, 14]) and many reports on queen-longevity are more anecdotal. Thus, it is largely unknown for long-lived queens whether they age gradually or whether aging is a more sudden event.

During recent years, several pioneering studies, especially on the honeybee, revealed exciting new insights into the mechanisms of how queens can live so long. Generally, the TOR (target of rapamycin) and the IIS (insulin/insulin-like growth factor1 signalling) pathway have been associated with longevity in model organisms from *D. melanogaster* to mice and humans [15–17]. They are the most intensively studied aging-related pathways and they have also been associated with caste differences in social Hymenoptera (e.g.,[18–23]). Additionally, in the honeybee, juvenile hormone (JH) seems to have lost its direct gonadotropic function in adults so that queens have a high expression of *vitellogenin* (*Vg*), which encodes yolk precursors, without requiring high JH titers (e.g., [24, 25]). This result has led to the hypothesis that an uncoupling between JH and *Vg* expression might account for the long life of honeybee queens [24], as well as social insect queens more generally [26], because the life-time shortening consequences of high JH titers are absent. However, this re-wiring along the JH-*Vg* axis is not universal for all social Hymenoptera since the queens of many ant and bee species require JH for vitellogenesis (e.g. [27] and references therein). For termites, fewer studies exist but JH is required for vitellogenesis [28, 29] and a recent study revealed that no re-wiring exists along the JH-*Vg* axis [30].

Hence, other mechanisms must exist to explain the long life of termite queens. Studies of the subterranean termite *Reticulitermes speratus* implicated the involvement of a *breast cancer type 1 susceptibility* (*BRCA1*) homolog [31], which is involved in DNA repair [32], and better protection against oxidative stress by superoxide dismutases and catalases [33, 34]. The latter has also been discussed for other social Hymenoptera, including ants and the honeybee. Yet, the overall evidence of the role of oxidative stress is less clear (reviewed in [35–38]). Furthermore, regulation of the activity of transposable elements (TEs) [14] and changes in the insulin/insulin-like growth factor1 signalling (IIS) and target of rapamycin (TOR) pathways [39] have been linked with caste-specific aging differences in termites Yet, all studies on social insects suffer from a lack of time-series data to investigate molecular changes across the lifespan of long-lived queens. Like the demographic life history data, such data are inherently difficult to obtain due to the long lifespan of queens. However, they are necessary (i) to understand the aging process, (ii) to work out potential changes compared to solitary insects, and (iii) to identify the relevant age-classes for detailed studies. The latter is a completely overlooked issue but highly relevant. Differences across studies might be consequences of non-comparable age-classes between studies, if, for example, aging is a non-linear process.

We studied aging in termite queens of known age across their entire lifespan to measure at the ultimate, eco-evolutionary level age-dependent survival and at the proximate, mechanistic level age-specific changes in gene expression. For the latter, we generated head / thorax transcriptomes of queens of different ages (for an outline of the workflow, see Additional File 1, Figure S1; for rational of tissue choice, see Methods). We used field collected, newly established colonies of the wood-dwelling (i.e., one-piece nesting) termite *Cryptotermes secundus* (Hill, 1925) (Blattodea, Isoptera, Kalotermitidae) that were kept under identical conditions in the laboratory for a time span of up to 15 years. Laboratory conditions have been adjusted to *C. secundus* so that colony development (e.g. fecundity, growth rates, colony composition, molting types) does not differ compared to that of field colonies as has been shown in earlier studies [40, 41]. Keeping colonies under constant good conditions without external mortality allowed us to study intrinsic aging, disentangled from causes of extrinsic mortality such as predation, food shortage, or disease. As typical for wood-dwelling termite species, *C. secundus* colonies are founded in a piece of wood, which serves as food and shelter and which workers never leave to forage outside. Such species have a low social complexity with small colonies and totipotent workers that develop into sexuals.

Results

Survival analysis

Our survival analysis covered a time period of up to 15 years; yet none of the queens that would have had an age of 14 and 15 years was alive and the maximal age we recorded was 13 years (Additional File 2, Table S1). In fact, most queens with an expected age of ≥ 12 years were dead. Out of eight queens in this 'old-age' class, only a single queen (13 years) had survived (Fig. 1; Additional File 2, Table S1). Kaplan Meier survival analysis estimated the median longevity of the queens to be 12.0 years (SE: ± 0.54) (mean longevity: 11.1 years, SE: ± 0.66) in the laboratory after successful colony foundation (Fig. 1).

Identifying transcripts that change their expression with age: age-related DETs

To study gene expression changes over the life-time of queens, we generated transcriptomes of head / thorax from twelve queens with different chronological age since the onset of reproduction, from two until 13 years, covering the complete lifespan of *C. secundus* queens: 2, 3, 4, 5, 6, 7, 8, 9, 10 (two samples), 11, and 13 years (Additional File 2, Table S2). The queens used for gene expression analyses came from the same data set as those for the survival analysis; they were alive queens that entered the survival analysis as censored data (for more details see Methods).

A total of 169 transcripts were significantly differentially expressed (DETs) over time as revealed by Iso-MaSigPro time series analysis (Additional File 2, Table S3). According to their expression pattern, DETs were grouped into six Iso-MaSigPro clusters (hereafter, 'cluster') (Fig. 2). Cluster 1 represented 44 DETs, which were slightly expressed in young queens followed by a decline at middle ages and a strong increase when queens became older. The 32 DETs of cluster 2 characterized young queens with a declining expression with age. Clusters 3 and 5 comprised 31 and 37 DETs, respectively, that were highly expressed in middle-aged queens, while cluster 4 and cluster 6 (15 and 10 DETs) characterized old queens with no expression in young ones. Thus, in the following text, we refer to the DETs as young (cluster two), middle-aged (clusters three and five) and old DETs (clusters one, four and six). Details for all clusters are provided in Additional File 2, Table S3.

Identifying modules of co-expressed transcripts

To identify modules of co-expressed transcripts, we performed a weighted gene co-expression network analysis (WGCNA). It revealed a total of 254 modules of co-expressed transcripts. Based on eigengene values, 13 modules correlated significantly positively with age and 13 negatively (see Additional File 1, Figures S2 and S3; Additional File 3 (WGCNA module-age association, shown are eigengene values for all modules).
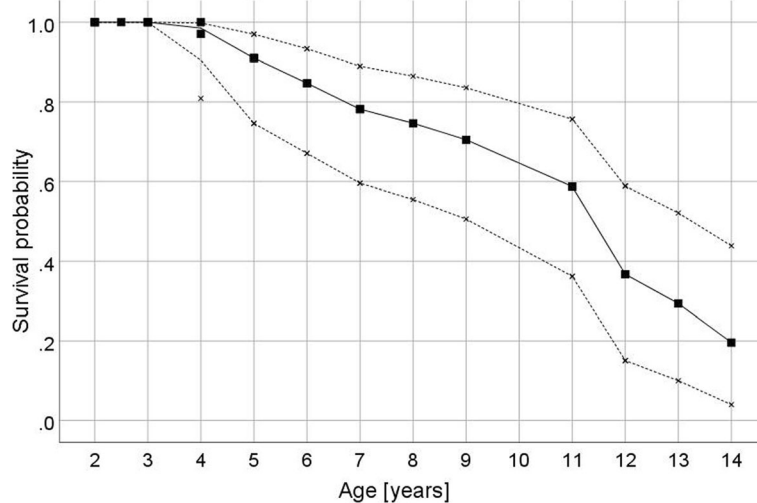
Fig. 1 Survival plot of *C. secundus* queens. Shown is the age-dependent survival probability (filled squares) of queens with 95 % confidence intervals (x) estimated with Kaplan Meier survival analyses. The median longevity of queens in the laboratory after successful colony establishment was estimated with Kaplan Meier survival analysis to be 12 years (mean longevity: 11.1 years). The maximum lifespan was 13 years. After an age of around 11 years, life expectancy declines rapidly; out of eight queens with a potential age ≥ 12 years, all had died, except one 13-year old queen. Note, the x-axis starts at an age of two because by default the queens had to survival for the first year to be included in our study

Fig. 2 Median expression profiles of DETs assigned to Iso-MaSigPro clusters. Iso-MaSigPro grouped the differentially expressed transcripts (DETs) into six clusters. DETs of cluster 1, 4 and 6 were especially highly expressed in old queens, while those of cluster 3 and 5 characterized middle-aged queens and those of cluster 2 young queens. The expression values correspond to normalized counts (see Methods). The youngest queen (age: 2 years) was taken as time step zero and each of the subsequent older queens (based on chronological age) were considered to be one time step older. One age class (time step 8; age: 10 years) consisted of two samples

Identifying transcript co-expression modules with age-related DETs

Within the age-correlated WGCNA modules, we identified age-related DETs. The negatively age-correlated module 'seashell4' had the highest number of young DETs (10 DETs). No gene ontology (GO) term was enriched for this module. The highest number of old DETs was found in the positively age-correlated modules 'cyan' (89 DETs) and 'tan' (79 DETs) (Additional File 2, Table S4 and S5). Only broad categories were enriched in the 'cyan' module (e.g., RNA metabolic process and gene expression). The 'tan' module was enriched for ribosomal and tRNA related functions (Additional File 1, Figure S4).

Extracting age-related subnetworks based on age-related DETs

To generate subnetworks related to the age-related DETs, we located them in the WGCNA co-expression network. These DETs and their one- and two-step neighbors (i.e., the 'second level neighborhood') were

then extracted from the co-expression network, which resulted in 50 subnetworks of different sizes (for more details, see Methods) (Additional File 1, Figure S5). Note, DETs might be located at the boundaries of multiple WGCNA modules, which means the subnetworks obtained consist of fragments of multiple WGCNA modules. The resulting subnetworks either contained young and middle-aged DETs or old DETs, with a single exception where a middle-aged DET was in the periphery of the largest subnetwork containing old DETs. The largest subnetwork containing either young and middle-aged DETs (hereafter, young subnetwork) or old DETs (plus a single middle-age DET; hereafter old subnetwork) were further analyzed.

Young subnetwork

The largest young subnetwork comprised 164 transcripts (out of these 24 Iso-MaSigPro DETs), of which only 12 (7 %) were one-to-one orthologs to *D. melanogaster* genes (Additional File 2, Table S4). The GO enrichment

analysis of the young subnetwork showed multiple Biological Process (BP) terms related to RNA catabolism, but these GO terms were not significant after correcting for multiple testing (Additional File 2, Figure S6).

TE activity and genome instability 53 DETs (32 %) of the young subnetwork were related to TEs (Fig. 3 and Additional File 2, Table S4), comprising TEs and genes from TE defense pathways. This included one homolog of the gene *argonaute 2* (*ago2*) (two transcripts), an essential gene of the endo-siRNA pathway which silences TEs [42], and *arsenite 2* (*ars-2*), which is required for siRNA and miRNA-mediated TE silencing [43]. Additionally, we found two genes connected to DNA damage response and genome instability: *kin17* and *PIF1-like* gene.

Other signatures From well-known aging pathways, we identified (i) *inositol polyphosphate phosphatase 2* (*mipp2*) and (ii) *adenylyl cyclase 76E* (*ac76E*). The former is part of the TOR pathway and has been associated with longevity [44], and the latter is activated by the transcription factor 'Forkhead box O' (*foxo*). Additionally, we found several fecundity-related DETs. They included two transcripts of the gene *hu li tai shao* (*hts*) (one a DET of IsoMaSigPro cluster two) and one homolog of the gene *bällchen* (*ball*) (two transcripts) (one a DET of cluster five, Fig. 3).

Old subnetwork
The largest 'old subnetwork' comprised 1,098 transcripts (out of these 42 Iso-MaSigPro DETs). 521 transcripts (47 %) were identified as one-to-one orthologs of *D. melanogaster* genes (Additional File 2, Table S4). Iso-
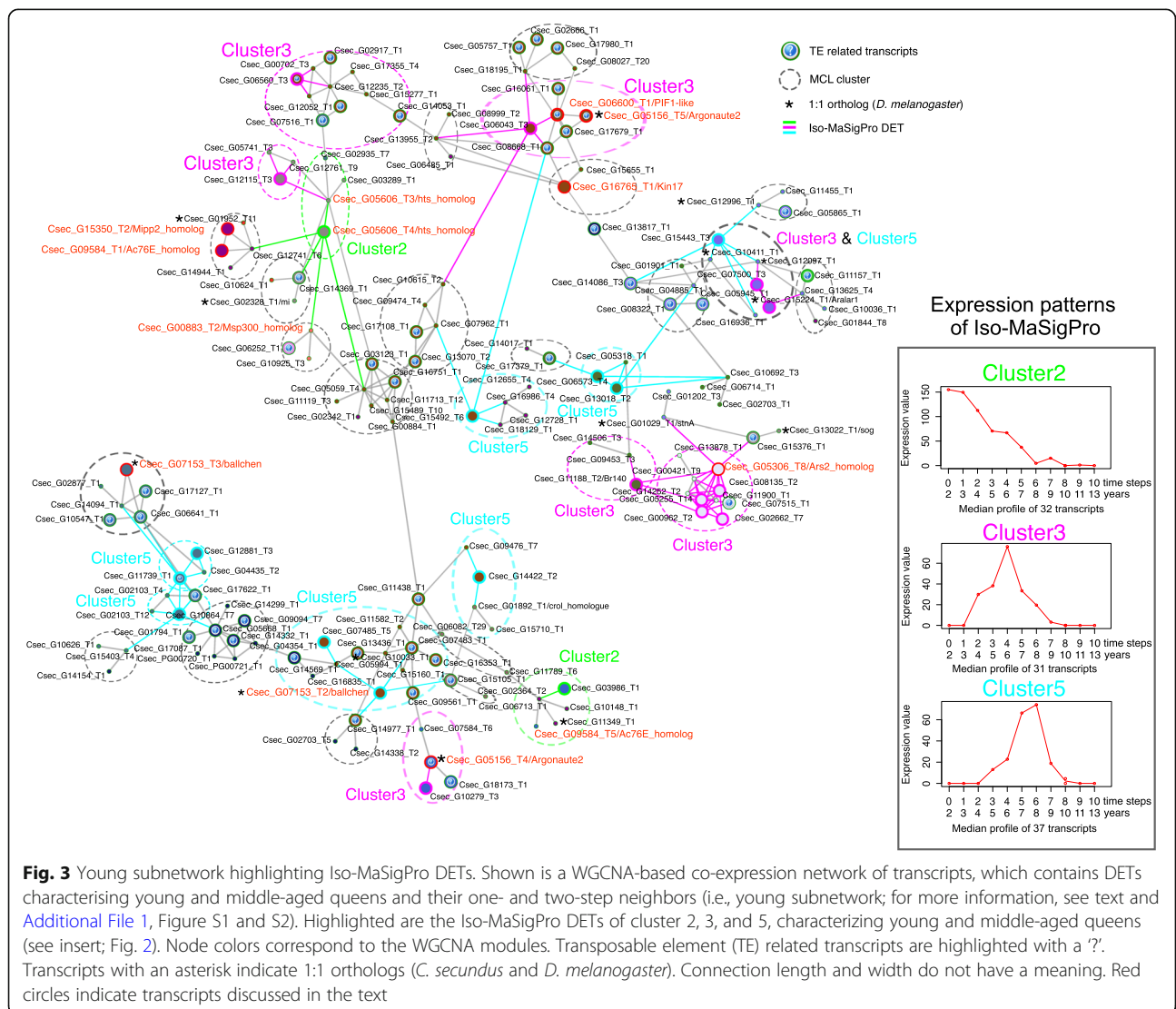

Fig. 3 Young subnetwork highlighting Iso-MaSigPro DETs. Shown is a WGCNA-based co-expression network of transcripts, which contains DETs characterising young and middle-aged queens and their one- and two-step neighbors (i.e., young subnetwork; for more information, see text and Additional File 1, Figure S1 and S2). Highlighted are the Iso-MaSigPro DETs of cluster 2, 3, and 5, characterizing young and middle-aged queens (see insert; Fig. 2). Node colors correspond to the WGCNA modules. Transposable element (TE) related transcripts are highlighted with a '?'. Transcripts with an asterisk indicate 1:1 orthologs (*C. secundus* and *D. melanogaster*). Connection length and width do not have a meaning. Red circles indicate transcripts discussed in the text

Monroy Kuhn *et al. BMC Genomics*   (2021) 22:339

Page 6 of 17

MaSigPro DETs in the old subnetwork belonged mainly to Iso-MaSigPro clusters 1 and 4. The second level neighborhoods of these DETs were connected in the network, and a GO enrichment analysis revealed multiple GO terms associated with protein-related functions, including translation, protein folding, unfolded protein binding, proteolysis involved in cellular protein catabolic process, protein targeting to ER, ribosome, and proteasome complex (Additional File 1, Figures S7 and S8). 198 transcripts of the old subnetwork (18 %) were involved in protein translation, protein folding, and protein catabolism and proteolysis [45, 46] (Figs. 4 and 5, and Additional File 2, Table S5). Additionally, 61 transcripts (∼ 6 %) were related to TEs (Additional File 2, Table S5).
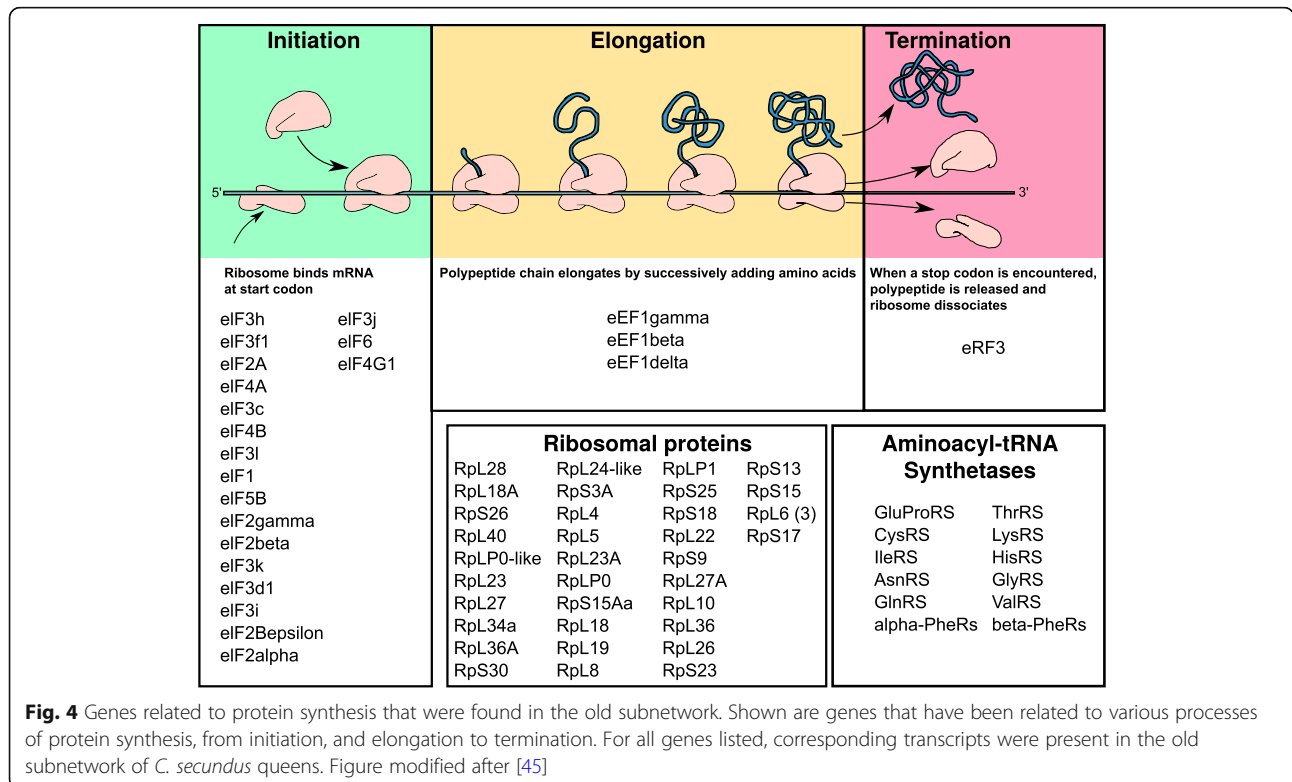
Epigenetic modifications, transcriptional regulation, and TE activity Many old subnetwork genes are involved in de/acetylation and methylation of DNA, which are important epigenetic modifications that regulate gene expression and genome stability [47–49] (Additional File 2, Table S5).
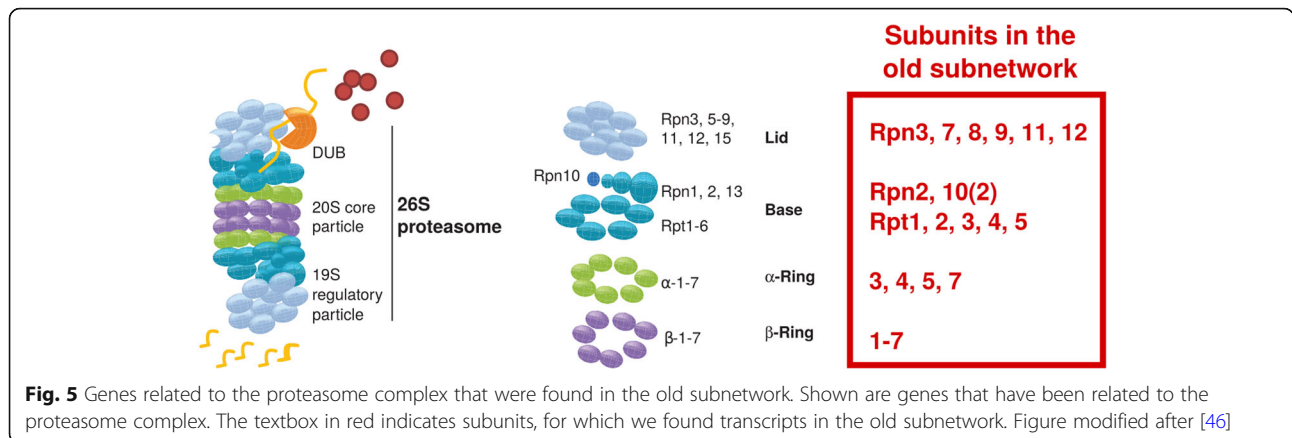
Most strikingly, two crucial histone acetylation modifying complexes, the Tip60 acetyltransferase complex and the male specific lethal (msl) complex were represented in the old subnetwork. The former included the genes *dom, ing3, mrg15, pont* and *rept*, and the latter *msl-1, msl-3* and *mof*. Genes involved in deacetylation of DNA were, for instance, *sirtuin 1* (*sirt1*), *histone deacetylase 3* (*HDAC3*), and *histone deacetylase 6* (*HDAC6*). Genes linked to epigenetic histone methylation included, for instance, *ash-1* and *lid*. Another well-represented group of genes connected to expression regulation in the old subnetwork were spliceosome components and splicing factors. Additionally, we found in the old subnetwork important transcripts related to TE silencing: *dicer-2, Hsc70-4, Hsc70-3, Hsp83, trsn, armi, Rm62, Gasz, Tudor-SN*, and *Hel25E*. Details are given in Additional File 2, Table S5.

Proteostasis and oxidative stress Related to proteostasis, we detected a strong signal for protein synthesis and degradation. Regarding protein synthesis, the old subnetwork comprised many transcripts coding for initiation, elongation and termination factors, as well as many ribosomal proteins and aminoacyl-tRNA synthetases (Fig. 4; Additional File 2, Table S5). Regarding protein degradation, almost all subunits of the ubiquitin proteasome system (UPS) were present (Fig. 5), as well as autophagy genes, heat shock proteins, and the transcription factor *xbp1*. Xbp1 is involved in the 'unfolded protein response' (UPR) and the ER-associated protein degradation (ERAD) pathway [50, 51].

Additionally, *BRCA1* was also present in the old subnetwork. This gene is involved in oxidative stress response, and in the transcriptional activation of

Fig. 4 Genes related to protein synthesis that were found in the old subnetwork. Shown are genes that have been related to various processes of protein synthesis, from initiation, and elongation to termination. For all genes listed, corresponding transcripts were present in the old subnetwork of *C. secundus* queens. Figure modified after [45]

Fig. 5 Genes related to the proteasome complex that were found in the old subnetwork. Shown are genes that have been related to the proteasome complex. The textbox in red indicates subunits, for which we found transcripts in the old subnetwork. Figure modified after [46]

proteasomal genes by stabilizing the transcription factor *cnc/nrf-2* (*cap-n-collar/nuclear factor erythroid 2–related factor 2*) [52]. Other genes in the old subnetwork involved in oxidative stress response and transcriptionally activated by *nrf-2* were thioredoxin and S-glutathione transferase.

Other signatures Additionally, the old subnetwork was characterized by a signature of ecdysone biosynthesis with *ecdysone receptor* (*EcR*), *ecdysone-induced protein 75B* (*Eip75B*), *phantom* and *disembodied*. The presence of *Phosphatidylinositol 3 kinase 68D* (*Pi3K92E*) links to the IIS pathway.

Locating age-related co-expression modules in the age-related subnetworks

Finally, we inspected those WGCNA modules with a large fraction of transcripts in the young and old subnetworks and those modules that were significantly associated with age.

In the young subnetwork, WGCNA modules with a large fraction of transcripts were 'saddlebrown' and 'skyblue4', which both did not significantly correlate with age. Significantly age-correlated co-expression modules were firebrick2, indianred1 and seashell4 (Additional File 1, Figure S2). No GO terms were significantly enriched for any of these modules.

In the old subnetwork, the modules with a large fraction of transcripts were 'green' and 'paleturquoise', which did not significantly correlate with age. The old subnetwork contained transcripts of 13 significantly age-correlated co-expression modules (Additional File 1, Figure S3). The GO enrichment analysis of these modules revealed several terms involved in protein-related functions, including ribosome biogenesis, rRNA processing, protein folding, translation, unfolded protein binding, protein catabolic process, protein transport, tRNA aminoacylation for protein translation, and proteasome

core complex (Additional File 1, Figure S4, S9, S10 and Additional File 4, Table S6).

Discussion

Our study revealed a median maximum reproductive longevity of *C. secundus* queens of 12 years with a maximum lifespan of 13 years when excluding all causes of extrinsic mortality in the laboratory (Fig. 1, Additional File 2, Table S1). The small difference between the median and the maximum recorded lifespan reflects the rather sudden decline in life expectancy after an age of around 11–12 years. Of eight queens with a potential age ≥ 12 years, all had died, except one 13-year old queen. The survival curve indicates a type I survivorship with high age-specific survival probabilities until midage and a rapid decline in survival later in life, after queens successfully founded a colony and without extrinsic mortality. Including the early colony founding stages, which are characterized by very high mortalities with more than 99 % of the dispersing sexuals dying in *C. secundus* [53], this suggests a bathtube curve of mortality for queens. A high early failure period is followed by a stable failure period and a final wear-out failure period [54].

Our transcriptome study identified six clusters of transcripts that were significantly differentially expressed with age (DETs) (Fig. 2): one cluster for young queens (cluster 2), two for medium-aged queens (cluster 3 and 5) and three for old queens (cluster 1, 4, and 6). This implies that three 'molecular' life stages can be distinguished in *C. secundus* queens, with the third corresponding to old, aged queens that will probably die soon as no queen reached a lifespan beyond 13 years. The co-expression network analysis, which extracted subnetworks based on age-related DETs, resulted in two main subnetworks, a young and an old subnetwork. This implies that there are two *age-related* 'molecular' life stages, as DETs/genes of young and medium ages belonged to the same young subnetwork.

Young subnetwork

Compared to the old subnetwork, the young subnetwork contained relatives few DETs (164 vs. 1,098 in the old subnetwork) and they were characteristic for young and medium ages. This shows the similarity in expression of these two age stages. Not unexpectedly, the young subnetwork indicates an upregulation of transcripts linked with fecundity (*e.g., hts, ball*) and of the TOR pathway (*mipp2*) which has been associated with longevity [44]. The upregulation of *Ac76E* may imply that the IIS pathway is down because this gene is activated by the transcription factor foxo, which is inhibited by an upregulated IIS pathway. However, other evidence suggests that, like in other social insects (e.g. [27] and references therein), queens are characterized by an upregulated IIS pathway [30, 39]. Additionally, we detected several upregulated TEs-related transcripts associated with signs of an upregulation of the endo-siRNA pathway (e.g., *ago2, ars*), which is a transcriptional and post-transcriptional TE-defence mechanism of the soma [42, 43, 55].

In an earlier study on *C. secundus*, we compared one-year old reproducing queens with queens that were at least seven years old, though the exact age of the later was unknown [39]. We identified 193 DETs between both age classes with no signs of physiological upheaval as revealed in the current old subnetwork. This implies that the old queens used in the earlier study were not physiological old and rather classify as queens of medium age. This is in line with the gene functions identified in the earlier study [39]. Similar to the current study, we had found an IIS- and TE-related signal as well as some fecundity-related genes, the latter were higher expressed in the medium- than young-age queens.

Old subnetwork

The old subnetwork contained many more transcripts (1,098 vs. 164 in the young subnetwork). Our results imply a physiological stage of upheaval shortly before queens die. There were strong signs of decline and repair at the upstream transcriptional level of epigenetic control as well as at the post-transcriptional level of TE-activity and the proteostasis network.

Epigenetic modifications

An upregulation of genes modifying histone marks implied considerable epigenetic changes that lead to altered gene expression as is typical for aging organisms:

First, our results indicate dynamic changes of 'active' histone marks of euchromatin because many genes related to H3K4 and H3K36 de/methylation and H4K16 de/acetylation were present in the old subnetwork (Additional File 2, Table S5). For instance, the Tip60- as well as the msl-complex were well represented. Both complexes are involved in the acetylation of histones, including H4K16, which, for instance, is indicative for replicative aging in yeast [56]. The upregulation of *Sirt1* may function as an antagonist as it deacetylates H4K16. The same applies to the deacetylases HDAC6 and HDAC3, which can also deacetylase histones [49].

Second, there is also evidence for changes of repressive histone mark of heterochromatin (e.g., linked to H3K9 and H3K27 acetylation) (Additional File 2, Table S5). For instance, the old age transcript ALP1 is an antagonist of HP1, the latter is involved in the maintenance of heterochromatin. HP1 generally decreases with age, and its overexpression can lead to an increased lifespan in *D. melanogaster* [57, 58].
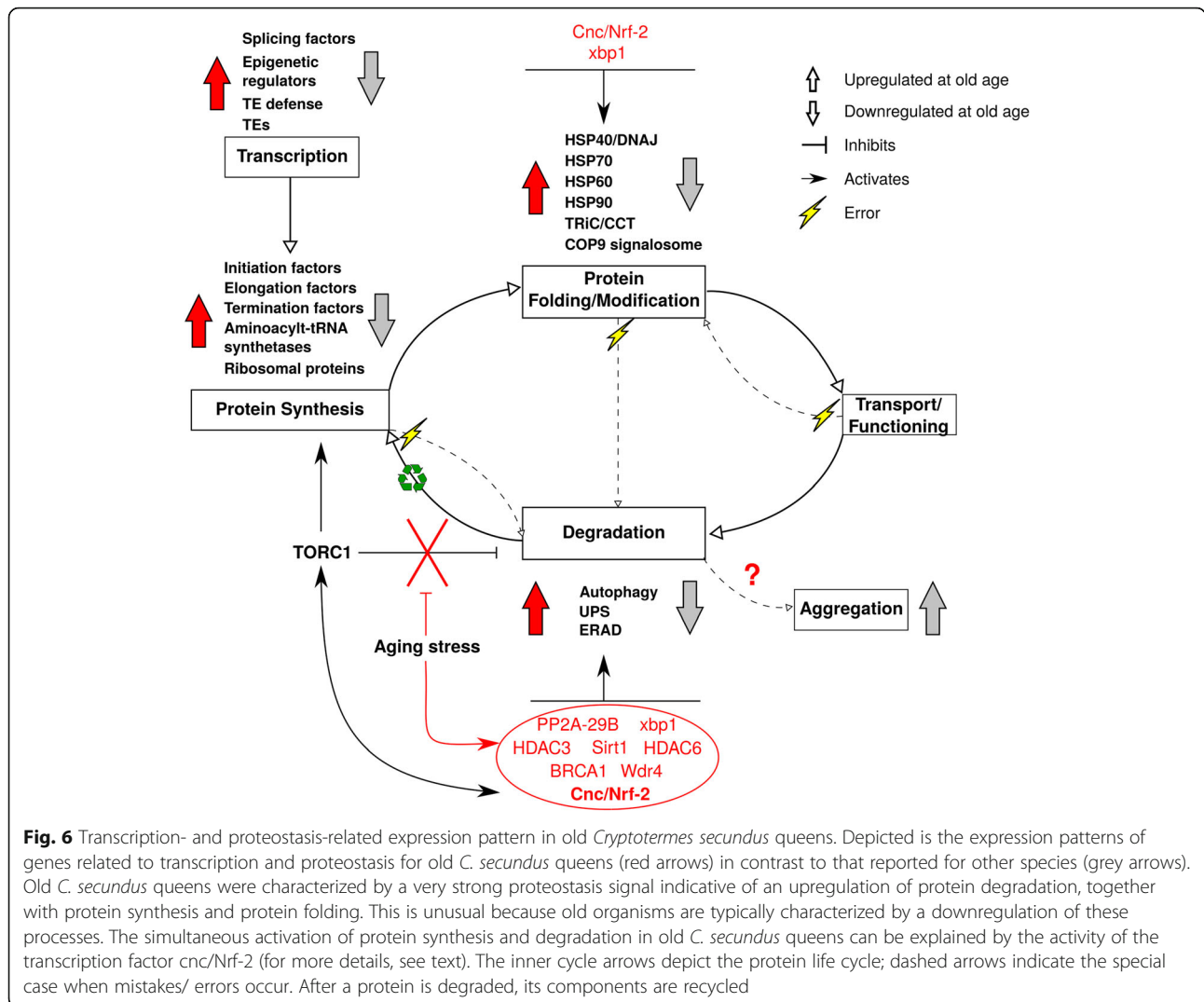
TE activity

In line with a deregulation of repressive histone mark of heterochromatin, several TE-related transcripts were also connected with the old subnetwork (Additional File 2, Table S5). TEs often are accumulated in heterochromatin, which silences their activity [59]. Yet, dysregulation of heterochromatin can allow TEs to become active and this has been associated with aging [60, 61].

Also, the upregulation of several genes from two TE-defense pathways - the endo-siRNA pathway (e.g., *dicer2, Hsc-70-4, trsn*) as well as the piRNA pathways (e.g., piRNA biosynthesis. *armi, gasz, Hel25E, Rm62*; ping-pong cycle: *Tudor-SN, qin*) - support the notion of active TEs. Both pathways silence TEs posttranscriptionally [62–64]. TE activity and the piRNA pathway have also been associated with aging and termite reproductives' longevity in another termite species [14].

Loss of proteostasis

Our results revealed a very strong proteostasis signal indicative of an upheaval in protein synthesis, protein folding and protein degradation in old queens. Many genes from the proteostasis network were detected in the old subnetwork (Additional File 2, Table S5). They indicate an upregulation of protein degradation, together with protein synthesis and protein folding (Figs. 4 and 5). This is unusual because old organisms are typically characterized by a downregulation of all these processes (Fig. 6).

Protein synthesis Many transcripts coding for ribosomal proteins and aminoacyl-tRNA synthetases were found in the old subnetwork, indicative of upregulated protein synthesis (Additional File 2, Table S5). This is further supported by a strong signal of an active TORC1 system which promotes protein synthesis [65–67] (Fig. 7). Thus, for instance, many downstream eukaryotic initiation factors (eIF4A, eIF4B, eIF4E, eIF4G) and

Fig. 6 Transcription- and proteostasis-related expression pattern in old *Cryptotermes secundus* queens. Depicted is the expression patterns of genes related to transcription and proteostasis for old *C. secundus* queens (red arrows) in contrast to that reported for other species (grey arrows). Old *C. secundus* queens were characterized by a very strong proteostasis signal indicative of an upregulation of protein degradation, together with protein synthesis and protein folding. This is unusual because old organisms are typically characterized by a downregulation of these processes. The simultaneous activation of protein synthesis and degradation in old *C. secundus* queens can be explained by the activity of the transcription factor cnc/Nrf-2 (for more details, see text). The inner cycle arrows depict the protein life cycle; dashed arrows indicate the special case when mistakes/ errors occur. After a protein is degraded, its components are recycled
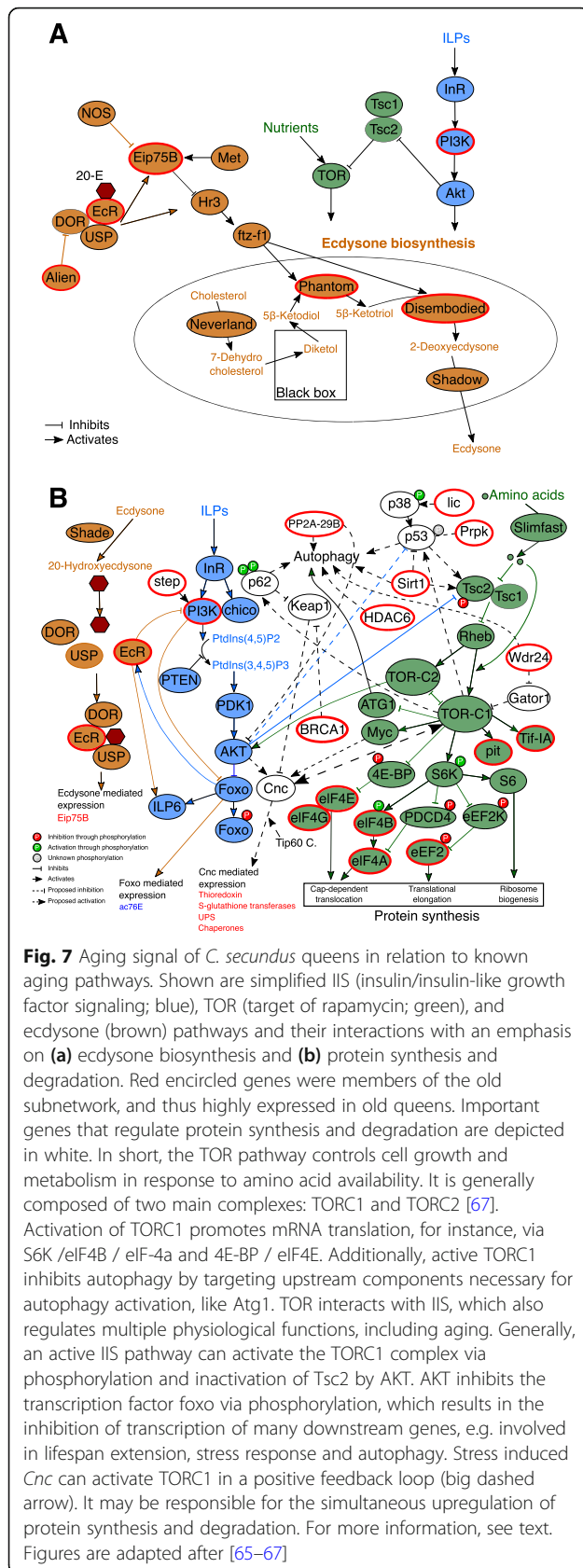
eukaryotic elongation factor (eEF2) transcripts were found in the old subnetwork (Fig. 4). They activate ribosome biogenesis, translational elongation, and cap-dependent translocation (Fig. 7).

Protein degradation Normally, an active TORC1 system is associated with a downregulation of protein degradation as it inhibits proteolytic systems [66–68] and autophagy (e.g., upregulated TORC1 inhibits ATG1, which is necessary for autophagy activation; Fig. 7). Surprisingly, however, we found strong evidence of upregulated protein degradation in the old subnetwork. Several transcripts linked to autophagy, almost all subunits of the UPS, the UPR-, and the ERAD pathway as well as heat shock proteins characterized the old subnetwork (Figs. 5 and 6; Additional File 2, Table S5).

Linking protein synthesis and degradation The simultaneous upregulation of protein synthesis and autophagy might be explained by a stress response. In *D. melanogaster* and humans, upregulated TORC1 enhances an oxidative stress response under stress conditions, controlled by the transcription factor *cnc*, a homolog of human *nrf* genes (Fig. 7). Ubiquitinated proteins and damaged mitochondria activate *cnc/ nrf-2* via p62, supported by upregulated TORC1, which then activates oxidative stress response genes [69, 70]. Additionally, *cnc* is known to activate chaperones (protein folding) and the proteasome [71], and this has been associated with lifespan expansion in *D. melanogaster* and *Caenorhabditis elegans* [72, 73].

Support for the conclusion of a stress-associated, active *cnc* transcription factor in old queens comes from several transcripts in the old subnetwork (Fig. 7): (i) BRCA1, which indirectly actives *cnc* by inhibiting *cnc*

Fig. 7 Aging signal of *C. secundus* queens in relation to known aging pathways. Shown are simplified IIS (insulin/insulin-like growth factor signaling; blue), TOR (target of rapamycin; green), and ecdysone (brown) pathways and their interactions with an emphasis on **(a)** ecdysone biosynthesis and **(b)** protein synthesis and degradation. Red encircled genes were members of the old subnetwork, and thus highly expressed in old queens. Important genes that regulate protein synthesis and degradation are depicted in white. In short, the TOR pathway controls cell growth and metabolism in response to amino acid availability. It is generally composed of two main complexes: TORC1 and TORC2 [67]. Activation of TORC1 promotes mRNA translation, for instance, via S6K /eIF4B / eIF-4a and 4E-BP / eIF4E. Additionally, active TORC1 inhibits autophagy by targeting upstream components necessary for autophagy activation, like Atg1. TOR interacts with IIS, which also regulates multiple physiological functions, including aging. Generally, an active IIS pathway can activate the TORC1 complex via phosphorylation and inactivation of Tsc2 by AKT. AKT inhibits the transcription factor foxo via phosphorylation, which results in the inhibition of transcription of many downstream genes, e.g. involved in lifespan extension, stress response and autophagy. Stress induced Cnc can activate TORC1 in a positive feedback loop (big dashed arrow). It may be responsible for the simultaneous upregulation of protein synthesis and degradation. For more information, see text. Figures are adapted after [65–67]

inhibitor *Keap1*, and (ii) the *Tip 60* complex as well as genes that are transcriptionally activated by *cnc*, such as *thioredoxin, S-Glutathione transferases* and *UPS* genes (Fig. 7).

Hence, our results imply that old *C. secundus* queens are in a stage of stress. They have mounted stress response systems mediated by *cnc*, including protein degradation and protein folding. However, it is unusual that old queens can do this. In *D. melanogaster*, only young individuals can mount this stress response [71]. The constant activation of the proteasome in these very old queens may lead to their death (note, the studied queens had reached their maximum lifespan, we never had older queens) as the proteasome's constant activation in transgenic flies was detrimental for survival [71].

Oxidative stress in other social insects

There has been a debate about the role of oxidative stress to explain the long lifespan of social Hymenoptera queens, yet the evidence is inconclusive (e.g., reviewed by [35–38]). For instance, markers of oxidative stress in honey bee workers' brains do not increase with age, although they live shorter than queens [74]. In the ant *Lasius niger*, both workers and males show higher activity of the antioxidant superoxide dismutase than queens, but both live shorter than queens [75]. These studies have shown that higher expression of oxidative stress response genes does not necessarily correlate with longer lifespans. A recent comparative analysis including one termite, one ant and several bee species, also did not obtain consistent results of an association between oxidative stress and longevity [38]. For the termite *Reticulitermes speratus*, it has been suggested that queens are better protected against oxidative stress as qRT-PCRs studies showed a higher expression of the antioxidants catalases and peroxiredoxins in queens compared to workers [33], while kings were characterized by a high expression of *BRCA1* (in the fat body) compared to workers [31]. Unfortunately, the age of the studied reproductives is not known. It would be of interest to study expression of these genes with age, as this would contribute to a better understanding of the aging process, but such studies are rare.

Conclusions

Our results imply that *C. secundus* queens do not age gradually. At old age, there is a physiological stage of upheaval, characterized by signs of stress (activity of TEs, active *crc*), defence (piRNA pathways) and repair (protein degradation and synthesis) before the animals die.

This sudden decline is in line with our life history records for *C. secundus* (Fig. 1), which match survival curves for the ant *Cardiocondyla obscurior* [11]. Most of the few other studies that surveyed survival of social

insect colonies and queens did not cover the final stage of death, although some were long-term studies [10, 12]. These studies recorded a low, steady rate of mortality, which is consistent with the stable failure period of a bathtube curve of mortality, before a short, final wear-out failure period starts [54]. These non-linear aging curves stress the importance of using queens of known age for aging studies as processes revealed for middle-aged versus old queens are expected to differ considerably.

Methods

Figure S1 (Additional File 1) provides a schematized workflow of the analyses described in the following sections.

Collection and colony maintenance

From 2002 until 2016, *C. secundus* colonies were collected from mangroves near Palmerston - Channel Island (12°30' S, 131°00' E; Northern Territory, Australia) when they were less than one year old [40]. Colonies of an age of less than one year can be unambiguously identified by the size and slightly lighter sclerotization of the founding reproductives (primary reproductives), the presence of fewer than 20 workers and short tunnel systems of a few centimeters. All collected colonies were transferred to *Pinus radiata* wood blocks and transported to the laboratory in Germany, where they were kept under standardized conditions in a climate room with a temperature of 28 °C, 70 % humidity and a 12 h day/night rhythm. Under these conditions, colonies develop like in the field (see [40, 41]). Only colonies that survived the transfer from the field to the laboratory in Germany and the re-establishment in the new wood block were used for survival analysis.

Survival analysis

The survival of primary queens (and kings) was determined by their presence /absence in a colony.

Founding (primary) queens can be identified by their dark brown color, compound eyes and wing abscission scars. If the primary reproductives had died and the colony was still alive, they had been replaced by neotenic replacement reproductives that lack these traits.

Survival census of colonies was restricted to the times when colonies had to be transferred to new wood blocks to provide consistently abundant food conditions (low food conditions change colony trajectories and interfere with normal development, [40]). We restrained from splitting wood blocks more often as splitting and colony transfer is stressful for colonies. Thus, censuses were restricted to around every third year. To ensure that time of death of the queen – and hence her age - was

accurately determined (i.e., that they have died just within a few months prior to splitting), we only included colonies / queens, which clearly had newly developed neotenic replacement reproductives. Newly developed neotenic replacement reproductives can be unambiguously identified, for instance, by their very slight sclerotization.

We did not have a single cohort that we followed through time, as it is impossible to collect sufficient sample sizes of colonies with a known age of one year in the field. Hence, we used all one –year old colonies that we had collected over time to maximize sample size and which we had left undisturbed in the laboratory (except for necessary translocations to new wood blocks). These colonies were split in either 2017 or in 2018, which were the endpoints of this study. As sample size was still relatively small in 2017, we used another set of colonies that we split in 2018 to obtain more reliable survival analysis. Queens that were alive in 2017, respectively 2018, entered the survival analysis as censored data; representative samples of these queens were used for the transcriptome study. In total, we had over 100 one-year old, established colonies that entered this experiment. Out of these colonies, 41 colonies / queens could be used as the age of death could be reliably determined. Thus, the sample size was 41 queens (Additional File 2, Table S1).

The median longevity of queens was determined using Kaplan Meier survival analysis in SPSS 23 [76] with alive queens entering the data set as censored data. Confidence intervals were based on a transformation of intervals for the log-minus-log survival function according to recommendations by Hosmer et al. [77]. Overall, we had surviving primary queens from one year up to a maximum of 13 years; the three expected 14- and 15-year old queens were all dead (Additional File 2, Table S1).

Transcriptome study
Rational of tissue choice

We choose brain and thorax (including legs) to obtain a brain and endocrine signature, including the corpora allata and some hemolymph and fat body from the thorax. Recent work on *C. secundus* has pointed out that transcriptomes of the head and (pro)thorax conveys important information of queens and life history traits [30].

RNA extraction and sequencing

RNA was extracted from twelve queens with different chronological ages since onset of reproduction from two years until 13 years: 2, 3, 4, 5, 6, 7, 8, 9, 10 (two samples), 11, and 13 years. In colonies older than 13 years, the original queen was always replaced by a neotenic replacement queen.

Monroy Kuhn *et al. BMC Genomics*   (2021) 22:339

Page 12 of 17

An in-house protocol was followed for RNA extraction (see [39]). Individuals were placed on ice and the gut was removed and discarded. The head together with the thorax were used for RNA extraction. Samples were transferred into peqGOLD Tri Fast™ (PEQLAB) and homogenized in a Tissue Lyser II (QIAGEN). Chloroform was used for protein precipitation. From the aqueous phase, RNA was precipitated using Ambion isopropyl alcohol and then washed with 75 % ethanol. Obtained pellets were solved in nuclease-free water. DNA was subsequently digested using the DNase I Amplification Grade kit (Sigma Aldrich, Cat. No. AMPD1). We performed an RNA Integrity Number Analysis (RIN Analysis) measuring the RNA concentration with the Agilent RNA 6000 Nano Kit using an Agilent 2100 Bioanalyzer (Agilent Technologies) for quality control. Samples with total RNA were sent on dry ice to Beijing Genomics Institute (BGI) Tech Solutions (HONGKONG) Co. and then to the BGI-Shenzhen (PR China) for sequencing. BGI performed the preparation of the cDNA libraries according to their internal and proprietary standard operating procedure. The cDNA libraries were paired-end sequenced (not-strand specific) on Illumina HiSeq 2500 and 4000 platforms (100 base pairs read length and about 4 Giga bases per sample). Index sequences from the machine reads were demultiplexed and a quality-check and filtering of raw reads was done using the package soapuke (-n 0.05 -l 20 -q 0.2 -p 1 -i -Q 2 -G --seqType 1 and -A 0.5, http://soap.genomics.org.cn/).

Processing of RNASeq raw reads

FastQC (v0.11.4) [78] was used to evaluate the quality of the cleaned raw reads. To obtain a transcript count table, the cleaned raw reads were pseudo-aligned with Kallisto (default settings, v0.43.0) [79] against a *C. secundus* transcriptome obtained from a draft version of the *C. secundus* genome (with estimated gene and transcript models, see [39, 80]). The counts estimated by Kallisto were normalized using DESeq2 (v1.18.1, count/size factor) [81] in R (v3.4.4) [82].

Time course analysis to identify age-associated differentially expressed transcripts (DETs)

The normalized counts were used as input for the R package Iso-MaSigPro (v1.50.0) [83] to test for differentially expressed transcripts (DETs) across time. Iso-MaSigPro is designed for the analysis of multiple time-course transcriptome data. It implements negative binomial generalized linear models (GLMs) [83, 84]. Significantly differentially expressed transcripts (FDR corrected p-value set to 0.05) were clustered using the clustering algorithm mclust in Iso-MaSigPro and the k.mclust option set as 'true'. With this option, mclust computes

an optimal number of clusters (k) [83]. We obtained six Iso-MaSigPro clusters (Additional File 2, Table S2).

Weighted gene co-expression Network analysis (WGCNA) to identify networks of co-expressed transcripts

To obtain networks of co-expressed transcripts that were categorized as modules, we performed a Weighted Gene Co-expression Network Analysis (WGCNA). The counts obtained with Kallisto (v0.43.0) [79] were transformed using variance stabilizing transformation (vst) as implemented in DESeq2 (v1.18.1) [81]. The vst transformed counts were used to perform a co-expression network analysis with the R package WGCNA (v1.63) [85]. For more details on the methodology, see [85–87]. In short, (Additional File 1, Figure S1, workflow, right side), a similarity matrix was built by calculating Pearson correlations between the expression values of all pairs of transcripts. Using the similarity matrix, a signed weighted adjacency matrix was obtained as described by the formula:

$$a_{ij} = \left(\frac{1}{2} \left(1 + \mathrm{cor}_{ij} \right) \right)^{\beta}$$

Where cor_{ij} is the Pearson correlation between the expression pattern of transcript 'i' and transcript 'j' (the similarity value). The value of β was chosen based on the soft-thresholding approach [85]. With this value of β, we obtained a weighted network with an approximate scale-free topology ($\beta = 14$, scale-free topology $R^2 = 0.84$). In a signed weighted adjacency matrix negative and small positive correlations get negligibly small adjacency values shifting the focus on strong positive correlations. Seeing the adjacency matrix as a network, the nodes correspond to the transcripts and the connections between nodes correspond to the adjacency values (transformed correlation coefficients). A topological overlap matrix (TOM), which in addition to the adjacency matrix considers topological similarity (shared neighbors reinforce the connection strength between two nodes), was constructed using the adjacency matrix [88]. To define transcript modules, a hierarchical clustering tree was constructed using the dissimilarity measure (1-TOM). Transcript modules were defined by cutting the branches of the tree using the Dynamic Hybrid Tree Cut algorithm [89] and the minimum module size was set to 30 transcripts. Transcript modules with similar expression profiles were merged by hierarchical clustering of the eigengene correlation values. Briefly, a hierarchical clustering tree was created with the eigengene dissimilarity measure (1-correlation coefficient of eigengenes) and a tree height cut of 0.20 was used (corresponds to an eigengene cor ≥ 0.80). Eigengenes were calculated with the function moduleEigengenes (default settings) [85]. A module

Monroy Kuhn *et al. BMC Genomics*   (2021) 22:339

Page 13 of 17

eigengene corresponds to the module's first principal component and can be seen as a weighted average expression profile [85]. To find significantly associated modules with age, correlations between age and eigengenes of the merged modules were calculated. Each module was named after a color by WGCNA.

The adjacency matrix of the WGCNA was visualized using Cytoscape (v3.7.1) [90], only including pairs of nodes with a $cor_{ij} \geq 0.90$. Each module's color corresponds with the respective module name (e.g., saddle-brown color for the saddlebrown module).

To identify co-expression modules containing age-related DETs, we looked for age-related DETs from the Iso-MaSigPro analysis in the WGCNA modules. Those modules that were significantly correlated with age and held the highest number of Iso-MaSigPro DETs were further inspected.

Identifying/Extracting age-related subnetworks based on age-related DETs

To identify age-related subnetworks within the co-expression network, we combined the results of the Iso-MaSigPro analysis with those from the WGCNA and extracted subnetworks that were based on age-related Iso-MaSigPro DETs. Therefore, we extracted 1st and 2nd neighbors of DETs based on the WGCNA co-expression network (i.e., the visual representation of the adjacency matrix). To do this, we used the 'First neighbors' function of Cytoscape. We selected an age-related DET from Iso-MaSigPro as transcript of interest. By calling the function, the neighboring transcripts were selected, which were then extracted to form a subnetwork. By calling the function twice, one obtains the one- and two-step neighbors (called 'second level neighborhood') of the transcript of interest. This was done for each DET identified in IsoMaSigPro.

The obtained subnetworks were clearly separated in those containing young Iso-MaSigPro DETs (young subnetworks) and those containing old Iso-MaSigPro DETs (old subnetworks). The largest subnetwork obtained for each group was used for further analysis paying attention to both transcript identity and WGCNA module content. Thus, for instance, we looked for WGCNA modules that had been identified to be age-related within the global WGCNA in these subnetworks.

The AutoAnnotate Cytoscape plug-in (v1.3) [91] was used to annotate the subnetworks using the clustering algorithm 'Markov Cluster' (MCL) [92] to define and visualize sub-clusters, and the labeling algorithm 'Adjacent Words' to label the sub-clusters. The Cytoscape plug-in BiNGO (v3.0.3) [93] was used for GO enrichment analysis. The p-values of the GO enrichment analysis were adjusted for multiple testing using the FDR

approach [94]. Subnetworks were graphically processed with Inkscape (v0.91, www.inkscape.org).

Transcript (functional) annotation

Nucleotide and protein sequences were obtained from the draft version of *C. secundus* genome [39, 80]. For annotation, the translated transcripts were searched against the Pfam database (Pfam A, release 30) [95] with the software *hmmscan* (option --cut_ga, HMMer v.3.1b2) [96] and against the InterPro database with the software InterProScan (v5.17-56.0) [97]. Additionally, we did a BLASTX search (NCBI BLAST suite v. 2.3.0) [98] with an e-value of $1e^{-05}$ as cut-off against the protein coding sequences of the termite *Zootermopsis nevadensis* (official gene set version 2.2) [99]. To further assist the annotation, we inferred a set of clusters of orthologous sequence groups (COGs) from the official gene sets at the amino acid level of *C. secundus* (draft version) and *D. melanogaster*, and a BLASTP search of *C. secundus* sequences against the protein coding sequences (longest isoforms only) of *D. melanogaster* with a threshold e-value of $1e^{-05}$.

To detect possible TEs, transcripts were searched against the Dfam database (v2.0) [100] using nhmmer [101]. A transcript was considered TE related if there was a hit against the Dfam database and the other annotation sources (Pfam, Interpro and BLAST) were not pointing in the direction of a known gene.

Gene identification and construction of gene trees

In addition to the functional annotation, we inferred phylogenetic trees for selected transcripts (Supplementary Archive 1, DRYAD, doi available upon acceptance, see below for a private reviewer link). Following the procedure described in [39], we retrieved protein coding sequences of the respective cluster of orthologous sequence groups (COGs) from OrthoDB v.9.1 [102] for the following species: *D. melanogaster* (DMEL), *Apis mellifera* (AMEL), *Cardiocondyla obscurior* (COBS), *Polistes canadensis* (PCAN), *Tribolium castaneum* (TCAS), *Z. nevadensis* (ZNEV) and *Blattella germanica* (BGER). COGs were identified using text search by searching for the gene name or IDs of *D. melanogaster*. In case Selenocysteine (U) was included in sequences, "U" was replaced by "X" to avoid problems in downstream analyses since many programs cannot handle Selenocysteine properly. Protein sequences of COGs of the above-listed species were separately aligned with MAFFT (v.7.294b) applying the G-INS-i algorithm and otherwise default options [103]. For each multiple sequence alignment, a profile hidden Markov model (pHMM) was built with hmmbuild (HMMER v.3.1b2) [96]. Then the pHMM was used to search (hmmsearch) for corresponding protein coding sequences in *C.*

Monroy Kuhn *et al. BMC Genomics*   (2021) 22:339

Page 14 of 17

secundus and *Macrotermes natalensis* to identify orthologous candidate sequences for each COG in both species. For gene (transcript) tree inference, we only kept sequences with a threshold e-value of $\leq 1e^{-40}$. In addition, we annotated all candidate sequences identified in *C. secundus* and *M. natalensis* against the Pfam database (Pfam A, release 30) using hmmscan (HMMER v.3.1b2).

To infer phylogenetic gene trees, we merged for each COG the COGs (amino-acid level) of the seven species listed above with the putatively orthologous amino-acid sequences of *C. secundus* and *M. natalensis*. We generated multiple sequence alignments for a total of 29 genes of interest applying MAFFT (G-INS-i, see above). Ambiguously aligned sequence sections were identified with Aliscore (v. 2 [104, 105]; settings: -r: 10,000,000 and otherwise defaults) and removed with Alicut (v. 2.3, https://www.zfmk.de/en/research/research-centres-and-groups/utilities); masked alignments are provided as Supplementary Archive S1 (deposited at DRYAD, doi available upon acceptance, see below for a private reviewer link). Phylogenetic trees were inferred with IQ-TREE (1.7-beta12 [106]) for each gene. The best model was selected with the implemented ModelFinder [107] from all available nuclear models implemented in IQ-TREE plus the two protein mixture models LG4M and LG4X [108] based on the Bayesian Information Criterion (BIC). We applied default settings for rates and the number of rate categories. Statistical support was inferred from 2,000 non-parametric bootstrap replicates. Unrooted trees with the bootstrap support mapped were visualized with Seaview (v4.5.4 [109]) and provided in Newick Format with Supplementary Archive S1 at DRYAD (doi available upon acceptance, see below for a private reviewer link).

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-021-07649-4.

Additional file 1: Figure S1. Schematic workflow. Details are explained in the Methods. **Figure S2.** WGCNA modules of co-expressed transcripts that negatively correlate with age. Modules are named after colors by WGCNA. Eigengenes of all these modules showed a negative correlation with age while * indicates that the age-correlation was significant. Modules marked with $ were found in the young subnetwork. **Figure S3.** WGCNA modules of co-expressed transcripts that positively correlate with age. Modules are named after colors by WGCNA. Eigengenes of all these modules showed a positive correlation with age; * indicates that the age-correlation was significant.Additionally, modules marked with $ were found in the old subnetwork. **Figure S4.** GO enrichment for the WGCN Amodule 'tan' that positively correlated with age and which had many old-age DETs. Details are shown for BP (Biological Process), which revealed an enrichment of transcripts for ribosomal and tRNA related functions. **Figure S5.** Young and old transcript subnetworks corresponding to the second level neighborhood of Iso-MaSigPro DETs. Age-related DETs were located in the WGCNA co-expression network and these DETs and their one- and two-step neighbors (*i.e.,* the 'second

level neighborhood') were then extracted from the co-expression network to provide the shown networks. **Figure S6.** BiNGO GO enrichment(Biological Process) for the young subnetwork. No terms were significantly enriched after correcting for multiple testing (FDR). **Figure S7.** BiNGO GO enrichment (Biological Process) for the old subnetwork. Colored nodes are GO terms that were significantly enriched after correcting for multiple testing (FDR). **FigureS8.** BiNGO GO enrichment of (Molecular Function; Cellular Component) for the old subnetwork. Colored nodes are GO terms that were significantly enriched after correcting for multiple testing (FDR). **Figure S9.** BiNGO GO enrichment (Biological Process; MolecularFunction; Cellular Component) for the 'green' WGCNA module, which is part of the old subnetwork. Colored nodes are GO terms that were significantly enriched after correcting for multiple testing (FDR). **Figure S10.** BiNGO GO enrichment (Biological Process, Molecular Function; Cellular Component) for the 'paleturquoise' WGCNA module, which is part of the old subnetwork. Nodes in color are GO terms significantly enriched after correcting for multiple testing (FDR).

Additional file 2: Table S1. Data for survival analysis. **Table S2.** Sample information of samples included in this study. **Table S3.** Differentially expressed transcripts of the Iso-MaSigPro analysis, separately for cluster 1-6. **Table S4.** Transcripts in the young subnetwork (SNW). Transcripts in the young subnetwork (SNW) classified into major categories; TE-related transcripts of the young subnetwork (SNW). **Table S5.** Transcripts in the old subnetwork (SNW). Transcripts in the old subnetwork (SNW) classified into major categories; TE-related transcripts of the old subnetwork (SNW).

Additional file 3. WGCNA module-age associations. Listed are the eigengenes and the correlation coefficients with respect to age (**p**-values in parenthesis).

Additional file 4: Supplementary Table S6. GO terms for enriched differentially expressed transcripts (DETs) included in BiNGO module green; BiNGO module paleturquoise; BiNGO module tan; BiNGO module Thistle2; BiNGO module snow;BiNGO module cyan; BiNGO module deeppink1; BiNGO module navajowhite; BiNGO module blue2_NS; BiNGO module cornflowerblue_NS; BiNGO module pink3_NS; BiNGO module steelblue4_NS.

SAMN17305303 (https://www.ncbi.nlm.nih.gov/sra/17305303),
SAMN17305304 (https://www.ncbi.nlm.nih.gov/sra/17305304),
SAMN17305305 (https://www.ncbi.nlm.nih.gov/sra/17305305),
SAMN17305306 (https://www.ncbi.nlm.nih.gov/sra/17305306),
SAMN17305307 (https://www.ncbi.nlm.nih.gov/sra/17305307),
SAMN17305308 (https://www.ncbi.nlm.nih.gov/sra/17305308).
see also: Additional File 2: Table S2). Additional supplementary data are deposited on the Dryad digital repository DRYAD (https://doi.org/10.5061/dryad.nvx0k6dr6).

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Department of Evolutionary Biology & Ecology, Institute of Biology I, Albert Ludwig University of Freiburg, Hauptstr. 1, D-79104 Freiburg (i. Brsg.), Germany. [2]Computational Discovery Research, Institute for Diabetes and Obesity, Helmholtz Zentrum München, Ingolstaedter Landstr. 1, D-85764 Neuherberg, Germany. [3]Australian National Insect Collection, CSIRO National Research Collections Australia, Clunies Ross Street, Acton, ACT 2601 Canberra, Australia.

References

1. Jones OR, Scheuerlein A, Salguero-Gómez R, Camarda CG, Schaible R, Casper BB et al. Diversity of ageing across the tree of life. Nature. 2014;505: 169–73. doi:https://doi.org/10.1038/nature12789.
2. Linford NJ, Bilgir C, Ro J, Pletcher SD. Measurement of lifespan in *Drosophila melanogaster*. J Vis Exp. 2013;71. doi:https://doi.org/10.3791/50068.
3. Ungvari Z, Ridgway I, Philipp EER, Campbell CM, McQuary P, Chow T, et al. Extreme longevity is associated with increased resistance to oxidative stress in *Arctica islandica*, the longest-living non-colonial animal. J Gerontol A Biol Sci Med Sci. 2011;66:741–50. doi: https://doi.org/10.1093/gerona/glr044.
4. McMurray SE, Blum JE, Pawlik JR 2008 Redwood of the reef: growth and age of the giant barrel sponge *Xestospongia muta* in the Forida Keys. Marine Biol. 2008;155:159–171. doi:https://doi.org/10.1007/s00227-008-1014-z.
5. Rose MR. Evolutionary biology of aging. Oxford: Oxford University Press; 1991. ill.; 25 cm.
6. Keller L, Genoud M. Extraordinary lifespans in ants: a test of evolutionary theories of ageing. Nature. 1997;389:958–60. doi:https://doi.org/10.1038/40130.
7. Keller L. Queen lifespan and colony characteristics in ants and termites. Insectes Soc. 1998;45:235–46. doi:https://doi.org/10.1007/s000400050084.
8. Toth AL, Sumner S, Jeanne RL. Patterns of longevity across a sociality gradient in vespid wasps. Curr Opin insect Sci. 2016;16:28–35. doi:https://doi.org/10.1016/j.cois.2016.05.006.
9. Korb J, Thorne B. Sociality in Termites. In: Rubenstein DR, Abbot P, editors. Comparative Social Evolution. Cambridge: Cambridge University Press; 2017. p. 124–53. doi: https://doi.org/10.1017/9781107338319.006.
10. Cole BJ The ecological setting of social evolution: the demography of ant populations. In: Gadau J, Fewell J, editors. Organization of Insect Societies. Cambridge: Harvard Univ Press; 2009. p. 74–104.
11. Heinze J, Schrempf A. Terminal investment: individual reproduction of ant queens increases with age. PLoS One. 2012;7:e35201. doi:https://doi.org/10.1371/journal.pone.0035201.
12. Tschinkel WR Lifespan, age, size-specific mortality and dispersion of colonies of the Florida harvester ant, *Pogonomyrmex badius*. Insect Soc. 2017;64:285–296. doi:https://doi.org/10.1007/s00040-017-0544-0
13. Thorne BL, Breisch NL, Haverty MI Longevity of kings and queens and first time of production of fertile progeny in dampwood termite (Isoptera;

14. Elsner D, Meusemann K., Korb J. Longevity and transposon defense, the case of termite reproductives. Proc Natl Acad Sci USA. 2018;115:5504–9. doi: https://doi.org/10.1073/pnas.1804046115
15. Evans DS, Kapahi P, Hsueh W-C, Kockel L. TOR signaling never gets old: aging, longevity and TORC1 activity. Ageing Res Rev. 2011;10:225–37. doi: https://doi.org/10.1016/j.arr.2010.04.001.
16. Partridge L, Alic N, Bjedov I, Piper MD. Ageing in *Drosophila*: the role of the insulin/Igf and TOR signalling network. Exp Gerontol. 2011;46:376–81. doi: https://doi.org/10.1016/j.exger.2010.09.003.
17. Antikainen H, Driscoll M, Haspel G, Dobrowolski R. TOR-mediated regulation of metabolism in aging. Aging Cell. 2017;16:1219–33. doi: https://doi.org/10.1111/acel.12689.
18. Libbrecht R, Corona M, Wende F, Azevedo DO, Serrão, JR, Keller L. Interplay between insulin signaling, juvenile hormone, and vitellogenin regulates maternal effects on polyphenism in ants. Proc Natl Acad Sci U S A. 2013; 110:11050–55. doi: https://doi.org/10.1073/pnas.1221781110.
19. Mutti NS, Dolezal AG, Wolschin F, Mutti JS, Gill KS, Amdam GV. IRS and TOR nutrient-signaling pathways act via uvenile hormone to influence honey bee caste fate. J Experim Biol. 2011;214:3977–84. https://doi.org/10.1242/jeb.061499.
20. Amdam GV, Norberg K, Fondrk MK, Page RE Jr. Reproductive ground plan may mediate colony-level selection effects on individual foraging behavior in honey bees. Proc Natl Acad Sci U S A. 2004;101:11350–55. doi: https://doi.org/10.1073/pnas.0403073101.
21. Warner MR, Qiu L, Holmes MJ, Mikheyev AS, Linksvayer TA. Convergent eusocial evolution is based on a shared reproductive groundplan plus lineage-specific plastic genes. Nat. Commun. 2019;10:2651. doi: https://doi.org/10.1038/s41467-019-10546-w.
22. Chandra V, Fetter-Pruneda I, Oxley PR, Ritger A, McKenzie S, Libbrecht R, Kronauer DJC.Social regulation of insulin signaling and the evolution of eusociality in ants. Science 2018;361:398–402. doi: https://doi.org/10.1126/science.aar5723.
23. Korb J, Meusemann K, Aumer D, Bernadou A, Elsner D, Feldmeyer et al. Comparative transcriptomic analysis of the mechanisms underpinning ageing and fecundity in social insects. Phil Trans R Soc B. 2021;376: 20190728. doi:https://doi.org/10.1098/rstb.2019.0728.
24. Corona M, Velarde RA, Remolina S, Moran-Lauter A, Wang Y, Hughes KA, et al. Vitellogenin, juvenile hormone, insulin signaling, and queen honey bee longevity. Proc Natl Acad Sci U S A. 2007;104:7128–33. doi:https://doi.org/10.1073/pnas.0701909104.
25. Rascon B, Mutti NS, Tolfsen C, Amdam GV. Honey bee life history plasticity: Development, behavior, and aging. In: Flatt T, Heyland A, editors. Mechanisms of Life History Evolution: The Genetics and Physiology of Life History Traits and Trade-Offs. New York: Oxford Univ Press; 2011. p. 253–66.
26. Rodrigues MA, Flatt T. Endocrine uncoupling of the trade-off between reproduction and somatic maintenance in eusocial insects. Curr Opin Insect Sci. 2016;16:1–8.
27. Weitekamp CA, Libbrecht R, Keller L. Genetics and evolution of social behavior in insects. Annu Rev Genet. 2017;51:219–39. https://doi.org/10.1146/annurev-genet-120116-024515.
28. Maekawa K, Ishitani K, Gotoh H, CornetteR, Miura T. Juvenile hormone titre and vitellogenin gene expression related to ovarian development in primary reproductives compared with nymphs and nymphoid reproductives of the termite Reticulitermes speratus. Physiol Entomol. 2010;35:52–8. doi: https://doi.org/10.1111/j.1365-3032.2009.00711.x.
29. Korb J. Juvenile hormone: a central regulator of termite caste polyphenism. In: Zayed A, Kent C, editors. Genomics, Physiology and Behaviour of Social Insects. Academic Press; 2015. p. 131 – 61. doi:https://doi.org/10.1016/bs.aiip.2014.12.004.
30. Lin S, Werle J, Korb J Transcriptomic analyses of the termite, *Cryptotermes secundus*, reveal a gene network underlying a long lifespan and high fecundity. Comm Biol. accepted
31. Tasaki E, Mitaka Y, Nozaki T, Kobayashi K, Matsuura K, Iuchi Y. High expression of the breast cancer susceptibility gene BRCA1 in long-lived termite kings. Aging (Albany NY). 2018;10:2668–83. doi:https://doi.org/10.18632/aging.101578.
32. Wu J, Lu L-Y, Yu X. The role of BRCA1 in DNA damage response. Protein Cell. 2010;1:117–23. doi:https://doi.org/10.1007/s13238-010-0010-5.
33. Tasaki E, Kobayashi K, Matsuura K, Iuchi Y. An efficient antioxidant system in a long-lived termite queen. PLoS One. 2017;12:e0167412. doi:https://doi.org/10.1371/journal.pone.0167412.

Monroy Kuhn *et al. BMC Genomics* (2021) 22:339

Page 16 of 17

34. Tasaki E, Kobayashi K, Matsuura K, Iuchi Y. Long-lived termite queens exhibit high Cu/Zn-superoxide dismutase activity. Oxid Med Cell Longev. 2018; 2018:5127251. doi:https://doi.org/10.1155/2018/5127251.

35. Münch D, Amdam G V, Wolschin F. Ageing in a eusocial insect: molecular and physiological characteristics of life span plasticity in the honey bee. Funct Ecol. 2008;22:407–21. doi:https://doi.org/10.1111/j.1365-2435.2008.01419.x.

36. Lucas ER, Keller L. Ageing and somatic maintenance in social insects. Curr Opin Insect Sci. 2014;5:31–6. doi:https://doi.org/10.1016/j.cois.2014.09.009.

37. de Verges J, Nehring V. A critical look at proximate causes of social insect senescence: damage accumulation or hyperfunction? Curr Opin Insect Sci. 2016;16:69–75. doi:https://doi.org/10.1016/j.cois.2016.05.003.

38. Kramer BH, Nehring V, Buttstedt A, Heinze J, Korb J, Libbrecht R et al. Oxidative stress and senescence in social insects: a significant but inconsistent link. Phil Trans R Soc B. 2021;76:20190732. doi:https://doi.org/10.1098/rstb.2019.0732.

39. Monroy Kuhn JM, Meusemann K, Korb J. Long live the queen, the king and the commoner? Transcript expression differences between old and young in the termite Cryptotermes secundus. PLoS One. 2019;14:e0210371. doi: https://doi.org/10.1371/journal.pone.0210371.

40. Korb J, Lenz M. Reproductive decision-making in the termite, Cryptotermes secundus (Kalotermitidae), under variable food conditions. Behav Ecol. 2004; 15:390–5. doi: https://doi.org/10.1093/beheco/arh033.

41. Korb J, Katrantzis S. Influence of environmental conditions on the expression of the sexual dispersal phenotype in a lower termite: implications for the evolution of workers in termites. Evol Dev. 2004;6:342–52. doi: https://doi.org/10.1111/j.1525-142X.2004.04042.x.

42. Kim K, Lee YS, Carthew RW. Conversion of pre-RISC to holo-RISC by Ago2 during assembly of RNAi complexes. RNA. 2007;13:22–9. doi:https://doi.org/10.1261/rna.283207.

43. Sabin LR, Zhou R, Gruber JJ, Lukinova N, Bambina S, Berman A, et al. Ars2 regulates both miRNA- and siRNA- dependent silencing and suppresses RNA virus infection in Drosophila. Cell. 2009;138:340–51. doi: https://doi.org/10.1016/j.cell.2009.04.045.

44. Ivanov DK, Escott-Price V, Ziehm M, Magwire MM, Mackay TFC, Partridge L, et al. Longevity GWAS Using the Drosophila Genetic Reference Panel. J Gerontol A Biol Sci Med Sci. 2015;70:1470–8. doi:https://doi.org/10.1093/gerona/glv047.

45. Cooper, GM. Protein Synthesis, Processing, and Regulation (Chap. 7). In: The Cell: A Molecular Approach. 2nd. Sunderland: Sinauer Associates; 2000.

46. Wong E, Cuervo AM. Integration of clearance mechanisms: the proteasome and autophagy. Cold Spring Harb Perspect Biol. 2010;2:a006734. doi: https://doi.org/10.1101/cshperspect.a006734.

47. Putiri EL, Robertson KD. Epigenetic mechanisms and genome stability. Clin Epigenetics. 2010;2:299–314. doi:https://doi.org/10.1007/s13148-010-0017-z.

48. Gozani O, Shi Y. In: Workman JL, Abmayr SM, editors. Histone Methylation in Chromatin Signaling BT - Fundamentals of Chromatin. New York: Springer; 2014. p. 213–56. https://doi.org/10.1007/978-1-4614-8624-4_5.

49. Steunou A-L, Rossetto D, Côté J. Regulating Chromatin by Histone Acetylation BT - Fundamentals of Chromatin. In: Workman JL, Abmayr SM, editors. New York, NY: Springer New York; 2014. p. 147–212. doi:https://doi.org/10.1007/978-1-4614-8624-4_4.

50. Plongthongkum N, Kullawong N, Panyim S, Tirasophon W. Ire1 regulated XBP1 mRNA splicing is essential for the unfolded protein response (UPR) in Drosophila melanogaster. Biochem Biophys Res Commun. 2007;354:789–94.

51. Remondelli P, Renna M. The Endoplasmic Reticulum Unfolded Protein Response in Neurodegenerative Disorders and Its Potential Therapeutic Significance. Front Mol Neurosci. 2017;10:187. doi:https://doi.org/10.3389/fnmol.2017.00187.

52. Gorrini C, Baniasadi PS, Harris IS, Silvester J, Inoue S, Snow B, et al. BRCA1 interacts with Nrf2 to regulate antioxidant signaling and cell survival. J Exp Med. 2013;210:1529–44. doi: https://doi.org/10.1084/jem.20121337.

53. Korb J, Schneider K. Does kin structure explain the occurrence of workers in the lower termite Cryptotermes secundus? Evol Ecol. 2007;21:817–28.

54. Lienig J, Bruemmer H. Fundamentals of Electronic Systems Design. Berlin: Springer International Publishing; 2017. https://doi.org/10.1007/978-3-319-55840-0.

55. Piatek MJ, Werner A. Endogenous siRNAs: regulators of internal affairs. Biochem Soc Trans. 2014;42:1174–9. doi:https://doi.org/10.1042/BST20140068.

56. Dang W, Steffen KK, Perry R, Dorsey JA, Johnson FB, Shilatifard A, et al. Histone H4 lysine 16 acetylation regulates cellular lifespan. Nature. 2009;459:802–7. doi: https://doi.org/10.1038/nature08085.

57. Wood JG, Hillenmeyer S, Lawrence C, Chang C, Hosier S, Lightfoot W, et al. Chromatin remodeling in the aging genome of Drosophila. Aging Cell. 2010;9:971–8. doi: https://doi.org/10.1111/j.1474-9726.2010.00624.x.

58. Larson K, Yan S-J, Tsurumi A, Liu J, Zhou J, Gaur K, et al. Heterochromatin formation promotes longevity and represses ribosomal RNA synthesis. PLoS Genet. 2012;8:e1002473. doi: https://doi.org/10.1371/journal.pgen.1002473.

59. Wood JG, Helfand SL. Chromatin structure and transposable elements in organismal aging. Front Genet. 2013;4:274. doi: https://doi.org/10.3389/fgene.2013.00274.

60. Chen H, Zheng X, Xiao D, Zheng Y. Age-associated de‐repression of retrotransposons in the Drosophila fat body, its potential cause and consequence. Aging Cell. 2016;15:542–52. doi:https://doi.org/10.1111/acel.12465.

61. Wood JG, Jones BC, Jiang N, Chang C, Hosier S, Wickremesinghe P, et al. Chromatin-modifying genetic interventions suppress age-associated transposable element activation and extend life span in Drosophila. Proc Natl Acad Sci U S A. 2016;113:11277–82. doi: https://doi.org/10.1073/pnas.1604621113.

62. Liu Y, Ye X, Jiang F, Liang C, Chen D, Peng J, et al. C3PO, an endoribonuclease that promotes RNAi by facilitating RISC activation. Science. 2009;325:750–3. doi: https://doi.org/10.1126/science.1176325.

63. Kandasamy SK, Zhu L, Fukunaga R. The C-terminal dsRNA-binding domain of Drosophila Dicer-2 is crucial for efficient and high-fidelity production of siRNA and loading of siRNA to Argonaute2. RNA. 2017;23:1139–53. doi: https://doi.org/10.1261/rna.059915.116.

64. Tsuboyama K, Tadakuma H, Tomari Y. Conformational activation of Argonaute by distinct yet coordinated actions of the Hsp70 and Hsp90 chaperone systems. Mol Cell. 2018;70:722–729.e4. doi: https://doi.org/10.1016/j.molcel.2018.04.010.

65. Liu Y, Beyer A, Aebersold R. On the dependency of cellular protein levels on mRNA abundance. Cell. 2016;165:535–50. doi: https://doi.org/10.1016/j.cell.2016.03.014.

66. Cao JQ, Tong WS, Yu HY, Tobe SS, Bendena WG, Hui JHL. Chapter Three - The Role of MicroRNAs in Drosophila Regulation of Insulin-Like Peptides and Ecdysteroid Signalling: Where Are We Now? In: Verlinden HBT-A in IP, editor. Insect Epigenetics. Academic Press; 2017. p. 55–85. doi: https://doi.org/10.1016/bs.aiip.2017.02.002.

67. Laplante M, Sabatini DM. mTOR signaling at a glance. J Cell Sci. 2009;122 Pt 20:3589–94. doi: https://doi.org/10.1242/jcs.051011.

68. Zhao J, Zhai B, Gygi SP, Goldberg AL. mTOR inhibition activates overall protein degradation by the ubiquitin proteasome system as well as by autophagy. Proc Natl Acad Sci U S A. 2015;112:15790–7. doi: https://doi.org/10.1073/pnas.1521919112.

69. Ichimura Y, Waguri S, Sou Y-S, Kageyama S, Hasegawa J, Ishimura R, et al. Phosphorylation of p62 activates the Keap1-Nrf2 pathway during selective autophagy. Mol Cell. 2013;51:618–31. doi: https://doi.org/10.1016/j.molcel.2013.08.003.

70. Aramburu J, Ortells MC, Tejedor S, Buxade M, Lopez-Rodriguez C. Transcriptional regulation of the stress response by mTOR. Sci Signal. 2014;7:re2. doi: https://doi.org/10.1126/scisignal.2005326.

71. Tsakiri EN, Sykiotis GP, Papassideri IS, Terpos E, Dimopoulos MA, Gorgoulis VG, et al. Proteasome dysfunction in Drosophila signals to an Nrf2-dependent regulatory circuit aiming to restore proteostasis and prevent premature aging. Aging Cell. 2013;12:802–13. doi: https://doi.org/10.1111/acel.12111.

72. Sykiotis GP, Bohmann D. Keap1/Nrf2 signaling regulates oxidative stress tolerance and lifespan in Drosophila. Dev Cell. 2008;14:76–85. doi: https://doi.org/10.1016/j.devcel.2007.12.002.

73. Tullet JMA, Hertweck M, An JH, Baker J, Hwang JY, Liu S, et al. Direct inhibition of the longevity-promoting factor SKN-1 by insulin-like signaling in C. elegans. Cell. 2008;132:1025–38. doi: https://doi.org/10.1016/j.cell.2008.01.030.

74. Tolfsen CC, Baker N, Kreibisch C, Amdam GV. Flight restriction prevents associative learning deficits but no changes in brain protein-adduct formation during honeybee ageing. J Exp Biol. 2011;214:1322–32.

75. Parker JD, Parker KM, Sohal BH, Sohal RS, Keller L. Decreased expression of Cu-Zn superoxide dismutase 1 in ants with extreme lifespan. Proc Natl Acad Sci U S A. 2004;101:3486–9. doi:https://doi.org/10.1073/pnas.0400222101.

76. IBM Corp. Released 2015 IBM SPSS Statistics for Windows, Version 23.0. 2015. Armonk: IBM Corp.

77. Hosmer DW, Lemeshow S, May S. Applied Survival Analysis 2nd. 2008. Hoboken: Wiley.

78. Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010. http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

79. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nat Biotech. 2016;34:525–7. doi:https://doi.org/10.1038/nbt.3519.

80. Harrison MC, Jongepier E, Robertson HM, Arning N, Bitard-Feildel T, Chao H, et al. Hemimetabolous genomes reveal molecular basis of termite eusociality. Nat Ecol Evol. 2018;2:557–66. doi: https://doi.org/10.1038/s41559-017-0459-1.

81. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550. doi: https://doi.org/10.1186/s13059-014-0550-8.

82. R Core Team. 2018 R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing, Vienna). https://www.R-project.org/.

83. Nueda MJ, Martorell-Marugan J, Marti C, Tarazona S, Conesa A. Identification and visualization of differential isoform expression in RNA-seq time series. Bioinformatics. 2018;34:524–6. doi: https://doi.org/10.1093/bioinformatics/btx578.

84. Nueda MJ, Tarazona S, Conesa A. Next maSigPro: updating maSigPro bioconductor package for RNA-seq time series. Bioinformatics. 2014;30:2598–602. doi:https://doi.org/10.1093/bioinformatics/btu333.

85. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559. doi:https://doi.org/10.1186/1471-2105-9-559.

86. Zhang B, Horvath S. A general framework for weighted gene coexpression network analysis. Stat Appl Genet Mol Biol. 2005;4:17. doi:https://doi.org/10.2202/1544-6115.1128.

87. Horvath S, Dong J. Geometric interpretation of gene coexpression network analysis. PLoS Comput Biol. 2008;4:e1000117. doi:https://doi.org/10.1371/journal.pcbi.1000117.

88. Yip AM, Horvath S. Gene network interconnectedness and the generalized topological overlap measure. BMC Bioinformatics. 2007;8:22. doi:https://doi.org/10.1186/1471-2105-8-22.

89. Langfelder P, Zhang B, Horvath S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. Bioinformatics. 2008;24:719–20. doi: https://doi.org/10.1093/bioinformatics/btm563.

90. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13:2498–504. doi: https://doi.org/10.1101/gr.1239303.

91. Kucera M, Isserlin R, Arkhangorodsky A, Bader GD. AutoAnnotate: A Cytoscape app for summarizing networks with semantic annotations. F1000Research. 2016;5:1717. doi:https://doi.org/10.12688/f1000research.9090.1.

92. Enright AJ, Van Dongen S, Ouzounis CA. An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res. 2002;30:1575–84.

93. Maere S, Heymans K, Kuiper M. BiNGO: a Cytoscape plugin to assess overrepresentation of Gene Ontology categories in Biological Networks. Bioinformatics. 2005;21:3448–9. doi: https://doi.org/10.1093/bioinformatics/bti551.

94. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Ser B. 1995;57:289–300. http://www.jstor.org/stable/2346101.

95. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 2016;44:D279-85. doi: https://doi.org/10.1093/nar/gkv1344.

96. Eddy SR. Accelerated Profile HMM Searches. PLoS Comput Biol. 2011;7:e1002195. doi: https://doi.org/10.1371/journal.pcbi.1002195.

97. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014;30:1236–40. doi: https://doi.org/10.1093/bioinformatics/btu031.

98. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215:403–10. doi: https://doi.org/10.1016/S0022-2836(05)80360-2.

99. Terrapon N, Li C, Robertson HM, Ji L, Meng X, Booth W, et al. Molecular traces of alternative social organization in a termite genome. Nat Commun. 2014;5:3636. doi: https://doi.org/10.1038/ncomms4636.

100. Hubley R, Finn RD, Clements J, Eddy SR, Jones TA, Bao W, et al. The Dfam database of repetitive DNA families. Nucleic Acids Res. 2016;44:D81-9. doi: https://doi.org/10.1093/nar/gkv1272.

101. Wheeler TJ, Eddy SR. nhmmer: DNA homology search with profile HMMs. Bioinformatics. 2013;29:2487–9. doi: https://doi.org/10.1093/bioinformatics/btt403.

102. Zdobnov EM, Tegenfeldt F, Kuznetsov D, Waterhouse RM, Simao FA, Ioannidis P, et al. OrthoDB v9.1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. Nucleic Acids Res. 2017;45:D744-9. doi: https://doi.org/10.1093/nar/gkw1119.

103. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30:772–80. doi:https://doi.org/10.1093/molbev/mst010-

104. Misof B, Misof K. A Monte Carlo approach successfully identifies randomness in multiple sequence alignments: a more objective means of data exclusion. Syst Biol. 2009;58:21–34. doi: https://doi.org/10.1093/sysbio/syp006.

105. Kück P, Meusemann K, Dambach J, Thormann B, von Reumont BM, Wägele JW et al. Parametric and non-parametric masking of randomness in sequence alignments can be improved and leads to better resolved trees. Front Zool. 2010;7:10. doi: https://doi.org/10.1186/1742-9994-7-10.

106. Nguyen L-T, Schmidt HA, Von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32:268–74. doi: https://doi.org/10.1093/molbev/msu300.

107. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14:587–9. doi: https://doi.org/10.1038/nmeth.4285.

108. Le SQ, Dang CC, Gascuel O. Modeling protein evolution with several amino acid replacement matrices depending on site rates. Mol Biol Evol. 2012;29:2921–36. doi: https://doi.org/10.1093/molbev/mss112

109. Gouy M, Guindon S, Gascuel O. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol. 2010;27:221–4. doi:https://doi.org/10.1093/molbev/msp259.

Publisher's Note