1949

Investigation of the gene expression of *Lethrus apterus* in the perspective of parental behaviour

Nagyfejű csajkók (*Lethrus apterus*) génexpressziójának vizsgálata az utódgondozó viselkedés szempontjából

Thesis for the Degree of Doctor of Philosophy (PhD)

by: Nikoletta Andrea Nagy

Supervisors:
Prof. Zoltán Barta
professor

Dr. Zoltán Németh
lecturer

UNIVERSITY OF DEBRECEN
Doctoral Council of Natural Sciences and Information Technology
Juhász-Nagy Pál Doctoral School
Debrecen, 2021

Hereby I declare that I prepared this thesis within the Doctoral Council of Natural Sciences and Information Technology, **Juhász-Nagy Pál Doctoral School**, University of Debrecen in order to obtain a PhD Degree in Natural Sciences at University of Debrecen.

The results published in the thesis are not reported in any other PhD theses.

Debrecen, 2021.

…………………………………………

Nagy Nikoletta Andrea

Hereby I confirm that **Nikoletta Andrea Nagy** candidate conducted her studies with my supervision within the **Biodiversity** Doctoral Program of the Juhász-Nagy Pál Doctoral School between 2016 and 2020. The independent studies and research work of the candidate significantly contributed to the results published in the thesis.

I also declare that the results published in the thesis are not reported in any other theses.

I support the acceptance of the thesis.

Debrecen, 2021.

…………………………………………

Prof. Zoltán Barta

Hereby I confirm that **Nikoletta Andrea Nagy** candidate conducted her studies with my supervision within the **Biodiversity** Doctoral Program of the Juhász-Nagy Pál Doctoral School between 2016 and 2020. The independent studies and research work of the candidate significantly contributed to the results published in the thesis.

I also declare that the results published in the thesis are not reported in any other theses.

I support the acceptance of the thesis.

Debrecen, 2021.

…………………………………………

Dr. Zoltán Németh

Investigation of the gene expression of *Lethrus apterus* in the perspective of parental behaviour

Nagyfejű csajkók (*Lethrus apterus*) génexpressziójának vizsgálata az utódgondozó viselkedés szempontjából

Dissertation submitted in partial fulfilment of the requirements for the doctoral (PhD) degree in Biology

Written by **Nikoletta Andrea Nagy** certified Biologist

Prepared in the framework of the **Juhász-Nagy Pál Doctoral School** of the University of Debrecen
(Biodiversity programme)

Dissertation advisors:

Prof. Zoltán Barta
Dr. Zoltán Németh

The official opponents of the dissertation:
....
....

The evaluation committee:
chairperson: ....
members: ....
....
....
....

The date of the dissertation defence: ....

Table of contents

General introduction..1
  Introducing the model species...7
  Outline and objectives of the dissertation..9
Study 1: Predominant female care in the beetle *Lethrus apterus* with supposedly biparental care...11
  Objectives...11
  Results...11
  Discussion..12
Study 2: Draft genome of a biparental beetle species, *Lethrus apterus*..................14
  Introduction and objectives...14
  Materials and methods...15
    Sampling and DNA extraction...15
    DNA library preparation and sequencing...16
    Quality control and filtering..16
    Genome assembly...17
    Genome annotation...17
    Search for reproductive behaviour related genes...18
    Orthology and phylogenetic analysis..19
  Results and discussion...20
    Assembly quality and completeness..20
    Genome annotation...25
    Annotation of reproductive behaviour related genes...26
    Comparison with other coleopteran species..29
Study 3: Evaluation of potential reference genes for gene expression studies on the parental care of *Lethrus apterus*...31
  Objectives...31
  Results...32
  Discussion..32
Study 4: Inotocin, a potential modulator of reproductive behaviour in a biparental beetle, *Lethrus apterus*...34
  Introduction and objectives...34
  Materials and methods...36
    The effect of humidity on expression of inotocin and its receptor.......................36
    Field observations on the seasonal behaviour of *Lethrus apterus*........................37
    Sample collection for measuring seasonal gene expression of inotocin and its receptor...38
    RNA extraction and cDNA synthesis...39
    Gene identification and primer design..39
    Real-time quantitative PCR...40
    Statistical analysis...41
  Results...42

Structure of inotocin and its receptor..42
The effect of humidity on expression of inotocin and its receptor.................44
Seasonal variation in behaviour of *Lethrus apterus*.......................................45
Seasonal expression of inotocin..46
Seasonal expression of inotocin receptor..47
Discussion...49
General discussion..54
Summary - Key findings...59
Összefoglalás - Új tudományos eredmények..60
Contribution to the studies in the dissertation...61
Acknowledgements...62
References..63
List of publications...76
Appendix..77
Availability of beetle proteomes referenced in Study 2.....................................77
Supplementary studies on the bimodal distribution of inotocin expression levels 78
Effect of sex..80
Effect of sampling time...81
Effect of body size of individuals...82
Effect of location on RT-qPCR plates...83
Effect of RNA concentration...84
Effect of protein contamination...85
Activity and leaf collection during the breeding period.....................................86
Attachments...88

General introduction

Sociality and social behaviour occur on many different levels in extremely diverse taxa. Even behaviour that was traditionally considered as non-social (e.g. fighting, mating) have social aspects as in these processes as well, one individual's fitness depends on the other's behaviour (Hofmann et al. 2014). One of the most complex social behaviour is parental care, not only in terms of its diverse forms but in the evolutionary processes underlying these forms (Royle et al. 2016). Parental care is defined as 'any parental trait that enhances the fitness of a parent's offspring, and that is likely to have originated and/or to be currently maintained for this function' (Smiseth et al. 2012). This behaviour, however, is often costly for the parents because it decreases their possibility to invest in further offspring (Trivers 1972). Nevertheless, parental care evolved independently in as distant taxa as insects and mammals varying from provisioning of gametes to care for mature offspring (Smiseth et al. 2012). Organisms can be classified into four main groups based on the parents' investment in care:

- no care,
- female only care,
- male only care, and
- biparental care.

Considering the evolutionary origins of these types, it seems that female-only and male-only care evolved from the no care state, and biparental care originates from female care by male joining (Royle et al. 2016). Among mammals, reptiles and invertebrates, the most common form of care is maternal care, whereas paternal care dominates in fish. Among amphibian species, maternal and paternal care are equally distributed, and the most complex biparental care is mainly restricted to birds (Balshine 2012).

The origin of parental behaviour is not clear yet, nevertheless, according to Wilson (1975), four main conditions are likely contributing to the evolution of parental care: stable and structured habitats, harsh environment, specialized food resource and predation risk. For example, an interesting study was performed on a lizard (*Eutropis longicaudata*) which showed that without predator stress, no care was provided, in contrast, with high predation risk, females guarded their eggs (Pike et al. 2016) supporting the above-described hypothesis. The importance of environmental factors was further strengthened using mathematical models, revealing that interaction between ecological factors, life-history traits and costs of care is needed for inducing selection for parental care (Klug & Bonsall 2010). However, these models also pointed out that other components are interplaying in the evolution of parental care including the social environment, i.e. parent-parent, parent-offspring, and sibling interactions. In addition, further characteristics of sexes, such as mortality and stage-specific maturation, are theoretically shaping the type, i.e. maternal, paternal or biparental care, and the level of care, i.e. how much does the parent invest in the offspring (Klug et al. 2013). Consequently, one of the major aims of evolutionary biology, besides uncovering the origin of parental care, is to understand the causes of the diversity observed in parental care systems, including the variance in the care provider sex as well as the forms, levels and the duration of the activity (Clutton-Brock 1991). An important aspect of this endeavour is the investigation of the proximal, e.g. genetic and physiological mechanisms underlying the diverse forms of parental care observed in the animal kingdom (Hofmann et al. 2014).

Arthropods express the most variable forms of social behaviour, however, until recently, studies were largely restricted to understanding a highly complex grouping mechanism, namely eusociality. In eusocial species, overlapping generations live together exhibiting cooperative brood care and reproductive division of labour (Wilson 1975). Besides this decidedly intriguing form of sociality, not much attention was paid to the organisation and behaviour of non-eusocial species, often called ’other’ societies (Costa 2006). One category within the ’other’ societies was ’subsocial’ consisting of species that show any level of parental care of the eggs or the young (Wilson 1975). However, this terminology was argued to be inaccurate as it suggests that the species in question is not fully social, despite the fact that these societies can provide insights from the motivation and regulation of group formation (Costa 2018). More precisely, narrow phylogenetic comparisons can reveal the background of choice of diverse social paths between closely related species, whereas comparisons between distant taxa can answer the questions on the relationship between convergent evolution of parental behaviour and physiology (Trumbo 2012).

Within arthropods, insects can be outstanding models to investigate all aspects of parental care, including its mechanism, ecology and evolution. They usually have a relatively short life cycle and are highly manipulatable, besides, insect species exhibit diverse forms of care, from egg attendance to a well-coordinated system for offspring provision (Trumbo 2012). Moreover, both facultative, i.e. offspring do not fully depend on the care, and obligate care, i.e. offspring do not survive without care, can be found among insects (e.g. Capodeanu-Nägler et al. 2016). In a comprehensive study by Gilbert & Manica (2015), the evolution of parental care was investigated and compared between two groups of insect species: holometabolans which undergo complete metamorphosis and usually have immobile larvae; and hemimetabolans which develop through incomplete metamorphosis with mobile and self-sufficient nymphs. They found that parental care evolved in 44 families of nine orders among hemimetabolans and 51 families of four orders among holometabolans. The most characteristic form of care in both groups was found to be female only care, i.e. 87.2 % of parental hemimetabolan and 73.9 % of parental holometabolan species, respectively. In contrast, male only care was only described in hemimetabolan species (6.4 % of parental species). Besides, biparental care is much more frequent in holometabolans than in hemimetabolans (26.1 % and 6.4 %, respectively). These insect species provide excellent targets of comparative research on the evolutionary origin and the regulation of parental care (Kronauer & Libbrecht 2018).

Studies aiming to understand the regulation of parental care in insects are scarce. Nearly 20 years ago, a review was published aiming to summarise the knowledge on the hormones involved in parental care in insects (Trumbo 2002). In this review, information on the following six main groups of previously investigated care was collected: egg carrying by female cockroaches, guarding and grooming behaviour of female earwigs, direct biparental food provision and nest protection of burying beetles, egg and nymph attendance of burrower bugs, egg guarding of lace bugs and nesting and food provision of scarab beetles sometimes exhibited by both parents. In all referred studies, the main focus was on juvenile hormone. The level of this gonadotropic hormone peaks and decreases right before oviposition and remains at low levels during egg carrying in cockroaches. In addition, allatostatins and octopamine were found to inhibit juvenile hormone biosynthesis, therefore, can be potential modulators of behaviour forms related to reproduction. In earwigs, juvenile hormone and ecdysteroids are both expressed at high levels which decreases right after oviposition and remains at low levels during parental care. However, it was also demonstrated that high titre of juvenile hormone is not incompatible with parental care suggesting an indirect relationship between this hormone and the maternal care of earwigs. Juvenile hormone is present in the highest concentrations at the beginning of parental care in both sexes of burying beetles and decreases continuously during care. Interestingly, the increase of juvenile hormone occurs soon after finding a carrion that is necessary for offspring development. Experiments on inhibition of juvenile hormone in burrower bugs and lace bugs resulted in reduced offspring attendance of females suggesting a role in parental care. However, juvenile hormone directly was not investigated in scarab beetles, in spite, the volume of the secretory brain region, corpora allata was found to be bigger after insemination and decreases in size during the parental period implying a similar trend of juvenile hormone secretion in these species as described in the above-mentioned insects (Trumbo 2002).

Albeit the small number of publications on the topic of the hormonal regulation of parental care in insects, one important conclusion can be made, specifically, juvenile hormone plays a central role in all examined species. Therefore, genes and hormones related to the regulation of synthesis and secretion of this sesquiterpenoid are excellent targets of studies aiming to understand the regulation of parental behaviour in insects. Besides the above-mentioned allatostatins, ecdysteroids and octopamine, other important molecules affecting the control of juvenile hormone or being influenced by the absence or presence of this hormone are to be investigated, including allatotropins, neuropeptides, methoprene-tolerant as juvenile hormone receptor and vitellogenin (De Loof & Schoofs 2021). For instance, short neuropeptide F is a suppressor of juvenile hormone biosynthesis in the silkworm *Bombyx mori* (Kaneko & Himura 2014), and vitellogenin activation is induced by juvenile hormone in Colorado potato beetle (*Leptinotarsa decemlineata*, De Loof & de Wilde 1970).

In the past two decades, the number of studies aiming to understand the regulation of parental care in insects increased. However, from the potential models suggested by Trumbo, burying beetles, especially *Nicrophorus vespilloides* became the centre of such research. Species belonging to the *Nicrophorus* genus use small vertebrate carcasses as a food resource for their offspring. Parents directly feed the offspring with regurgitated liquid carrion, moreover, they maintain the nest and protect the offspring from predators (Scott 1998). Studying this species has several advantages, e.g. they can be easily bred and manipulated in laboratory and they have highly flexible parental care system including female only care, male only care and biparental care (Ratz & Smiseth 2018). Parker et al. (2015) performed differential gene expression analysis on five groups of individuals: biparental females, biparental males, uniparental females, uniparental males and control females and males (mated but not caring). Overall, they found 867 genes that were differentially expressed between parenting and control individuals. Within this gene set, genes that were related to food breakdown were enriched, most likely because of the regurgitation of food for the offspring. Additionally, two genes were highlighted, namely vitellogenin and takeout, to be downregulated in both parenting sexes compared to the control group suggesting a role in the regulation of parental care (Parker et al. 2015). Another study investigating the role of biogenic amins (octopamine, serotonin and dopamine) in *N. orbicollis* was published one year later (Panaitof et al. 2016). No relationship between parental care and octopamine or serotonin was found, whereas, dopamine significantly increased when offspring feeding started. Finally, Cunningham et al. (2016) examined the expression of neuropeptide F and its receptor in different stages of parental care in *N. vespilloides*. They reported lower expression levels of neuropeptide F receptor during care compared to other reproductive stages suggesting that this neuropeptide is involved in the regulation of parental care.

Another species has also attracted the interest of researchers, namely the small carpenter bee (*Ceratina calcarata*). Females collect pollen for each offspring separately in advance of their hatching. After this, mothers stay in the nest and protect the offspring from predators, furthermore, they groom and feed even the fully developed offspring. Besides, dwarf oldest daughters also forage to provide food for the other offspring (Rehan & Richards 2013). Considering the parallel of the dwarf daughters and the mothers of small carpenter bee with the workers and queens of eusocial insect societies as well as its close relatedness to honey bee (*Apis mellifera*), this species provides an outstanding opportunity to study the evolution of social behaviour modulation in social insects (Rehan et al. 2014). However, to date, not many studies were published aiming to identify genes that are involved in the parental behaviour of *C. calcarata*. Transcriptome and differential expression analysis of the mothers and dwarf oldest daughters at different time points of the reproductive period resulted in 180 differentially expressed genes associated with offspring care (Rehan et al. 2014). Among these, an odorant-binding protein was highlighted as evidence of the importance of olfaction in the regulation of parental care. A similar analysis was conducted recently, where differentially expressed genes were searched among foraging, guarding and nesting females (Shell & Rehan 2019). They found that compared to the nesting individuals, 483 and 1104 genes were upregulated in foraging and guarding females, respectively. Among the foraging and guarding associated genes, several were reported from other Hymenoptera species as well to be involved in similar behaviour, taking us one step closer to understanding the evolution of eusocial insect societies.

Unfortunately, the study of only a few species may bias evolutionary theories, therefore, it is essential to investigate as many species as possible for comparative objectives in order to make reliable conclusions on the evolution of parental care. Here, we aimed to widen our knowledge on the parental behaviour occurring in beetles and their genetic background using a species with well-developed division of labour between parents to contribute to the endeavour to understand the evolution of this complex behaviour.

Introducing the model species

One of the few insect species with biparental care is *Lethrus apterus* (Laxmann, 1770), a species belonging to the family Geotrupidae. Its distribution ranges from Hungary to the river Don where the species prefers grassland habitats (Král et al. 2013). The beetle is plain black and flightless since its elytra are fused. The body length of an adult beetle can vary between 2-4 cm and the pronotum width between 8.5-14 mm (Rosa et al. 2019). Both sexes have strong mandibles which are used to cut off leaves of diverse plants serving as food for themselves and the offspring. Sexual dimorphism can be observed as males have a pair of processes (often called tusks) on the ventral side of their mandibles which females do not bear. The size of the tusks of a male individual significantly positively correlates with its body size (Rosa et al. 2019).

L. apterus is only active during its breeding period which lasts from early March till the beginning of June depending on the weather (Merkl & Vig, 2009). In contrast to the literature from the previous century, the adults of *L. apterus* can live more than one year, spending the time between two reproductive periods underground in diapause (Kiss et al. 2020a). After emergence from the overwintering, individuals search for mates and excavate underground nests together (Schreiner 1906). The nests begin with a slant entrance channel which continues in a vertical 50-90 cm deep main channel with 6-11 chambers branching from it each of them ending in brood cells (von Lengerken 1939). According to earlier literature, a division of labour can be observed between sexes: the male collects fresh leaves cutting them with their strong mandibles, then the female forms a ball from these leaves into each brood cell and finally, lays one egg into each brood cells (Clutton-Brock 1991). In a field study, males were found to travel longer distances during the parental care period than females suggesting that active mate searching behaviour is present during the whole breeding season (Kiss et al. 2020b).

Accordingly, males protect their nests from intruders, being it either mate searching conspecific or predator species (Heymons 1915). The contests between two male individuals can last up to 30 minutes (Rosa et al. 2017) and the majority of them results in the victory of the resident male (Rosa et al. 2018). Interestingly, there is a difference between the behaviour of small and large males, i.e. large males initiate contests more often than small males. However, if a fight starts, duration does not depend on the body size of males (Rosa et al. 2018). According to a field experiment, nest guarding behaviour increased if the density of individuals or the proportion of males increased in the population which supports the hypothesis that males try to assure their paternity under such circumstances (Rosa et al. 2017).

At the end of the breeding period, when the parents finished with all of the brood cells, adult beetles dig themselves into the soil and start the overwintering period until next spring. Hatching larvae consume food stored in the brood chambers. According to the observations made by Emich (1884), the size of the imago correlates with the amount of leaf material provided for the larva. *L. apterus* have three larval instars followed by pupation from which imagos develop usually by September, but they will come up to the surface only in next spring (Merkl & Vig 2009).

Outline and objectives of the dissertation

Here I provide a brief summary of the four studies building up my dissertation.

Study 1: Life cycle and breeding behaviour of *L. apterus* were mainly described at the end of the 19th and the beginning of the 20th centuries. To obtain up-to-date knowledge on this species, we performed field observations on the behaviour of the beetles. We aimed to collect information on the division of labour between the parents during the breeding period and also to estimate the reproductive success of the individuals.

Study 2: To investigate the genetic and hormonal background of the parental behaviour of *L. apterus*, the sequence of the genome was essential since no genome was available from closely related species to use as a reference for downstream analyses, e.g. gene expression analysis. In the second study, we sequenced and assembled a draft genome of this species and performed structural and functional annotation to identify the genes present in it. In addition, we used the annotated genes to seek specifically for genes that were reported in other insects to be involved in the reproductive behaviour, including parental care to provide potential targets for future research on the regulation of parental behaviour in *L. apterus*.

Study 3: Gene expression analysis with RT-qPCR can provide a great opportunity to examine the relationship between a few selected genes and the behaviour of interest. A critical step for this is the use of reliable reference genes which can balance the potential technical errors during the process. Therefore, we used the draft genome of *L. apterus* to identify housekeeping gene sequences and investigated the stability of their expression regarding sex and season to find the best reference genes that can be used for data normalization in further gene expression analyses.

Study 4: Oxytocin/vasopressin and their orthologs are present in the whole animal kingdom from nematodes to mammals. These neuropeptides were shown to have conserved functions across widespread taxa from which the most remarkable are their roles in reproductive behaviour and water maintenance. We identified the gene sequences of inotocin, the insect ortholog of the peptide family, and its receptor in the genome of *L. apterus*. Based on the conserved roles found in diverse animal taxa, we investigated two questions: (1) whether the expression of inotocin or its receptor is affected by desiccation stress, and (2) whether the expression of the two molecules differs between sexes and during the breeding period, i.e. correlates with the reproductive and parental behaviour of the species.

Study 1: Predominant female care in the beetle *Lethrus apterus* with supposedly biparental care[1]

Objectives

As presented above, literature about the life cycle and breeding behaviour of *L. apterus* is scarce. Based on the descriptions from the beginning of the 20th century (Schreiner 1906; Heymons 1915; von Lengerken 1939), the majority of leaf collection for the offspring is performed by males whereas females are responsible for nest digging and ball formation from the collected leaves. However, no up-to-date observation-based characterization is available on the reproductive behaviour of *L. apterus*. Therefore, prior to starting work on the genetic background of parental care in this species, we collected data on its behaviour during the breeding period. For this purpose, we marked individually males and females in a natural population near Dorogháza, Hungary in 2013 during their active season. Leaf collection events performed by the marked individuals were recorded during the breeding period. In September, when the offspring were fully developed, the nests were excavated to estimate reproductive success of the observed individuals.

Results

Although the marked individuals were almost equally distributed between sexes, our results showed that the majority of recorded leaf collecting events belonged to females and only a small proportion to males. Furthermore, nests of males contained only dead larvae, whereas nests belonging to observed females had several fully developed, live offspring and only one dead larva.

---

[1] This is a summary of the article published as: Kosztolányi, A., **Nagy, N.**, Kovács, T., & Barta, Z. (2015). Predominant female care in the beetle *Lethrus apterus* with supposedly biparental care. *Entomological Science, 18*(2), 292-294. The full article is provided in the Attachment section.

Discussion

In contrast with descriptions of the parental behaviour of *L. apterus* from the literature, our observations showed that leaves presumably stored up for the offspring are mainly collected by females. Moreover, males that were collecting leaves seemed to have lower reproductive success than females, although we had small number of samples. These results are interesting because this species was described to have a specific division of labour between the parents, namely, the male collects the leaves and the female prepares them for the offspring. One explanation for our rather different results could be variation between populations since the majority of the descriptions were adapted based on the observations of Schreiner (1906) in a South Russian population. Such distance with barriers such as large rivers can result in divergence between populations, especially, in case of this flightless beetle species suggesting limited gene flow between the mentioned populations (Ito et al. 2010). Nevertheless, the same division of labour was reported by Emich (1884) in a Hungarian population of *L. apterus* which makes this explanation unlikely.

Another explanation could be that the predominant female leaf collection behaviour is specific to the examined population. *L. apterus* used to be a widespread, even pest species in Hungary until the transition of intense agriculture activities making the distribution area of this species very fragmented (Kovács et al. 2014). Nevertheless, this behaviour pattern was also observed in the Susa population in 2016, as well as in populations in Buda Hills and Debrecen (unpublished data).

In addition to these hypotheses, one should note that males which are staying at the nest could invest more energy into guarding it against intruders. Protection of the eggs increases the offspring survival thus the reproductive success of the parents (Wong et al. 2013). The guarding behaviour of males was also reported by Emich (1884) and Schreiner (1906), likewise the fact that females can also accomplish leaf collection if the male is missing. Moreover, as reported by Rosa et al. (2017), increasing proportion of males induces higher attendance to the nest. Consequently, this behaviour may be beneficial considering both offspring survival and paternity assurance (Figure 1).

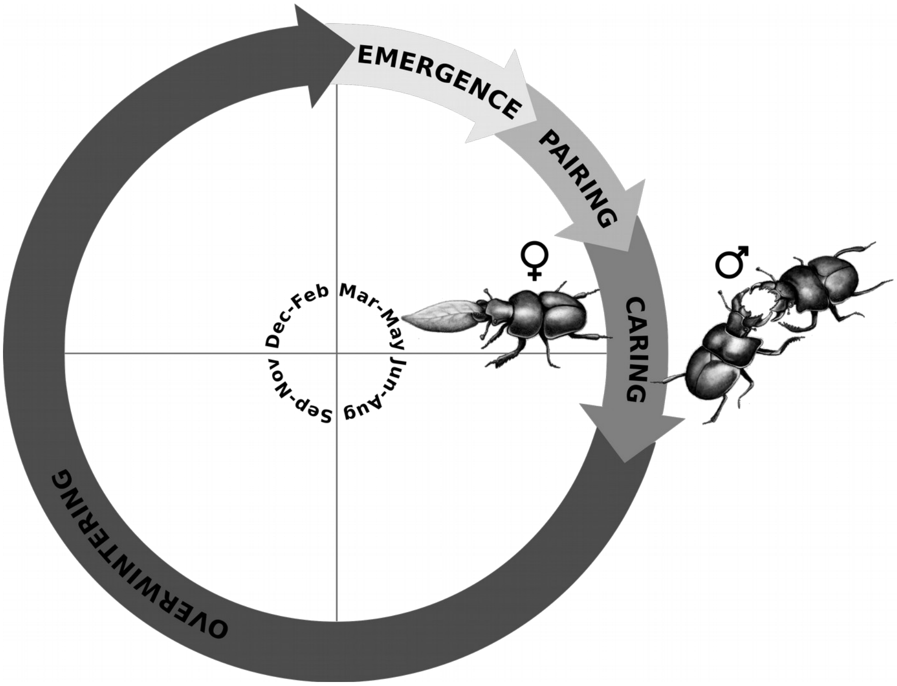


Figure 1. The breeding cycle of *L. apterus*. Beetles are only active from early March to the beginning of June. During the period of parental care, females collect leaves for the offspring while males protect the nest from intruders.

Study 2: Draft genome of a biparental beetle species, *Lethrus apterus*[2]

Nikoletta A. Nagy, Rita Rácz, Oliver Rimington, Szilárd Póliska, Pablo Orozco-terWengel, Michael W. Bruford and Zoltán Barta

Introduction and objectives

Beetles have evolved an extraordinary variety of life history strategies (McKenna & Farrel 2009), including very diverse social and reproductive behaviour, from aggregation through creating nests to biparental care (Brandmayr 1992). Parental behaviour appeared independently in 13 different insect orders, including 15 families of Coleoptera such as Scarabaeidae and Silphidae. These species are outstanding objects of sociogenomics aiming to understand the interaction between behaviour and genes regulating it, primarily under natural or naturalistic conditions (Robinson et al. 2005). However, investigating the genetic background of complex processes such as parental care in a non-model organism is challenging. One pathway to inferring the origin of sociality is via comparative analysis of genomes of many organisms with analogous parental behaviour (Cunningham 2020). However, despite the fact that beetles represent the most diverse animal order, only a handful of coleopteran genomes have been published to date, including model species like the red flour beetle (*Tribolium castaneum*), a burying beetle, *N. vespilloides* and important pests like the Colorado potato beetle (*L. decemlineata*), and the small hive beetle (*Aethina tumida*). Nevertheless, of the sequenced beetle species, only *N. vespilloides* and *Onthophagus taurus* have parental care (Gilbert & Manica 2015).

Here, we report the draft *de novo* genome of *L. apterus* which is the first published genome in the family Geotrupidae. We used the draft genome for

---

[2] Accepted as: Nagy, N. A., Rácz, R., Rimington, O., Póliska, Sz., Orozco-terWengel, P., Bruford, M. W. & Barta, Z. Draft genome of a biparental beetle species, *Lethrus apterus*. *BMC Genomics*

structural and functional annotation, with special emphasis on genes potentially related to the regulation of reproductive behaviour. Finally, we performed orthology and phylogenetic analysis using the annotated genes to achieve information about the phylogenetic position of *L. apterus* among coleopterans.

Materials and methods

Sampling and DNA extraction

In the spring of 2016, 12 individuals were collected from three neighbouring regions (Susa, Uraj and Ózd; 2 males and 2 females from each location) across Hungary. Samples were taken from the thorax of each individual which consisted of approximately 10 mg of muscle tissue. Each tissue sample was ground in a clean mortar using a pestle and liquid nitrogen and after grinding, 430 µl of extraction buffer was added. According to Gilbert et al. (2007), the buffer contained 3mM $CaCl_2$, 2% sodium dodecyl sulphate (SDS), 40 mM dithiotreitol (DTT), 250 µg/ml proteinase K, 100 mM Tris buffer pH 8 and 100 mM NaCl (final concentrations). The solution was suspended into a clean 1.5 ml microcentrifuge tube and 50 µl Ribonuclease A (50 mg/ml; VWR, Radnor, Pennsylvania, USA) was added. Suspensions were vortexed and centrifuged for 30 sec at 400 *g*. Samples were incubated overnight at 56 °C until no tissue clumps were visible. After the mixtures were cooled down at room temperature, 0.5 volume of 7.5 M ammonium-acetate was added to each sample and vortexed, and then put in a -20 °C freezer for 10-15 minutes. After this, 1 volume of chloroform:isoamylalcohol 24:1 was added to each sample and the tubes were vortexed for 2 sec. Samples were centrifuged for 5 min at 18000 *g* and the supernatant was carefully transferred to a clean tube. This step was repeated once more. Next, two volumes of ice-cold 96% ethanol were added and samples were mixed by inverting, then incubated on -20 °C overnight. Tubes were centrifuged at 18000 *g* for 10 min and the liquid phase was carefully removed from the pellets, then 500 µl 70% ethanol was added. After inverting, samples were centrifuged at 5000 *g* for 3 min, then the pellets were dried at room temperature. Finally, 50-100 µl of 10mM Tris-EDTA (pH=9.0) buffer was added to dissolve the pellets, depending on the pellet size.

DNA library preparation and sequencing

The sequencing library was prepared using 300 ng of DNA fragmented with a Bioruptor sonicator (Diagenode) with high mode 10 cycles setting (30 sec on/30 sec off). For the library preparation, the TruSeq® Nano DNA LT Sample Preparation Kit (Illumina) was used following the manufacturer's protocol. Individually barcoded libraries were diluted to 10 nM and two pools of 16 samples each were prepared for sequencing. Library preparation and paired-end sequencing of 125 bp reads on a HiSeq2500 machine were performed at the EMBL GeneCore Facility (Heidelberg).

Individual sequence coverage was on average low (~10x), therefore, the reads of the 12 individuals were combined to achieve a higher coverage that would facilitate genome reconstruction. Hereafter, we refer to this dataset of combined samples as the Susa dataset.

Quality control and filtering

The read quality of the Susa dataset was checked with FastQC (v0.11.5, Andrews 2010). The proportion of bases with a base quality of 30 or higher was above 90 % in all samples. Trimmomatic (v0.36, Bolger et al. 2014) was used to eliminate any adapter sequences from the reads, with at least 97% of the reads passing the trimming process. Next, the Susa reads were combined into two FASTQ files, one for the forward and one for the reverse sequence, and because these reads derived from different individuals, Musket (v1.1, Liu et al. 2013) was used to reduce the variation introduced by combining samples. After this correction, the quality of the dataset was checked with SGA-PreQC (v0.10.15, Simpson 2014). Results showed that 20% were PCR duplicates, high mean quality score (>Phred 35), low sequencing error rate ($<5x10^{-5}$), a peak of 51-mer distribution at count 70, and also predicted high repeat branch frequency similar to the human and oyster genomes. Additionally, GenomeScope (v1.0; Vurture et al. 2017), which works very well on short reads, was run on the 17-mer count histogram of the Susa dataset after Musket produced using Jellyfish (v2.2.6; Marçais and Kingsford 2011) to check heterozygosity and repeat region content.

Genome assembly

MEGAHIT (v1.1.1, Li et al. 2015) was used (minimum k-mer size 51, maximum k-mer size 111, k-mer step size 20) to generate the initial assembly of the Susa reads and contig length was increased through running three cycles of SSPACE (v3.0, Boetzer et al. 2011, Boetzer and Pirovano 2012) alternating with GapFiller (v1.10, Nadalin et al. 2012, Boetzer and Pirovano 2012). Another assembly of the Susa reads was run using SOAPdenovo2 (v2.04, Luo et al. 2012) with the default k-mer size. These two assemblies were then merged with gam-nsg (v1.1b, Vicedomini et al. 2013) to improve contig sizes. The final assembly was obtained by removing contaminant and duplicated contigs as reported by GenBank Contamination Screen and the potentially poor quality contigs shorter than 501 bp as short contigs can bias downstream analyses, e.g. mislead the gene prediction methods.

To assess the completeness of our assemblies, BUSCO (v4.1.2, Simão et al. 2015) was run to identify the proportion of the Endopterygota gene set that was found in the *Lethrus* genome. For this, the *T. castaneum* gene set (named ‘tribolium2012’) was used for algorithm training. The distribution of coverage and guanine-cytosine (GC) along the contigs was also analysed in R statistical environment (v3.5.2; R Core Team 2018).

Genome annotation

The distribution of repetitive sequences was analysed with RepeatModeler (version open-1.0.8, Smit & Hubley 2008-2015) and RepeatMasker (version open-4.0.7, Smit et al. 2013-2015). First, a *de novo* repeat library was built using RepeatModeler. RepeatMasker was used for repeat analysis based on RepeatModeler output and Repbase (version 20170127).

For gene prediction in the soft masked genome assembly, *ab initio* and homology-based prediction methods were combined with the BRAKER pipeline (v2.1.5; Hoff et al. 2016, 2019). *Ab initio* prediction was carried out using Augustus (v3.3.3; Stanke et al. 2006, 2008). For homology-based prediction, hints for GeneMark-EP (GeneMark-ES Suite v4.57_lic; Lomsadze et al. 2005) were generated with ProtHint (v2.4.0; Bruna et al. 2020) based on the endopterygota_db10. Augustus was then trained with these sequence hints with default settings and used them for prediction. This was followed by another round of training and prediction based on the first iteration results and GeneMark hints. Predicted coding sequences containing premature stop codons and shorter than 50 amino acids were filtered out.

Before annotation, the *ab initio* and homology-based predicted genes were merged using CD-HIT (v4.7; Fu et al. 2012) to cluster genes with the exact same amino acid sequences. The merged gene dataset was functionally annotated on the public server at usegalaxy.eu (Afgan et al. 2018) using InterProScan (v5.36-75.0; Jones et al. 2014) with Pfam (El-Gebali et al. 2019), PANTHER (Mi et al. 2019), ProDom (Servant et al. 2002), PROSITE (Sigrist et al. 2012), SMART (Letunic & Bork 2018), SUPERFAMILY (Pandurangan et al. 2019), TIGRFAM (Haft et al. 2012) and PRINTS (Attwood et al. 2012) databases. Additionally, a search with Diamond (v2.0.6.144; Buchfink et al. 2015) against Uniprot (downloaded on 6th May 2020; UniProt Consortium 2019) database with an e-value of 10^{-5} was used to further annotate the predicted genes.

Search for reproductive behaviour related genes

To identify genes potentially involved in parental behaviour, a literature search was carried out on 12th May 2020 on Web of Science using the following “Topic” keywords: (insect* AND ("reproductive behavior" OR "reproductive behaviour" OR "parental care") AND (gene OR genes)). This resulted in 96 articles from which a list of candidate genes and gene families was extracted on the basis of the genes being identified as playing a role in insect reproductive behaviour. NCBI was searched and filtered for complete and reliable protein sequences (not partial, predicted, putative, hypothetical, unknown or uncharacterised) of coleopteran species. These sequences were used as query for searching homologous proteins among the predicted genes of *L. apterus* using Diamond (in sensitive mode, with e-value set to 10-5 and query coverage to 40%). Besides, the same search was run against the proteomes of the following species: *T. castaneum* which has a well-annotated genome; *N. vespilloides* that is a model species for parental care among insects; *O. taurus*, the most closely related species with an available genome; *Asbolus verrucosus* which has a highly fragmented genome, therefore can be a scale for the accuracy of our result from the assembly of *L. apterus*.

Orthology and phylogenetic analysis

Orthofinder (v2.3.12; Emms & Kelly 2019) was used to find orthologs and species-specific genes and also create phylogenetic tree to reveal the phylogenetic relationship of *L.* to the other beetle species. For this purpose, proteomes of all available coleopteran species (*A. tumida*, *Agrilus planipennis*, *Anoplophora glabripennis*, *A. verrucosus*, *Callosobruchus maculatus*, *Dendroctonus ponderosae*, *Diabrotica virgifera*, *Ignelater luminosus*, *L. decemlineata*, *N. vespilloides*, *O. taurus*, *Photinus piralis*, *Sitophilus oryzae* and *T. castaneum*) and *Drosophila melanogaster* as an outgroup were downloaded from NCBI (accession numbers listed in Table S1). For inferring the species tree, orthogroups containing all species with only single-copy genes were used. Amino acid sequences by orthogroups were first aligned with MUSCLE (v3.8.1551; Edgar 2004) using default parameters. A concatenation-based phylogenetic tree was estimated using IQ-TREE (v1.6.9; Nguyen et al. 2015) by specifying the gene-based partitioning scheme (i.e.

allowing a different substitution model for each gene). Branch support was tested using 1000 ultrafast bootstrap replications (-bb 1000) and by a likelihood-ratio test (-alrt 1000). To further strengthen our results a coalescent based estimation of the species phylogeny was also applied as implemented in ASTRAL-III (v5.7.3; Zhang et al. 2018a). Maximum likelihood gene trees were inferred individually for each alignment also by using IQ-TREE (v1.6.9; Nguyen et al. 2015). Branch support values were assessed by 1000 ultrafast bootstrap replications (-bb 1000). Branches with lower bootstrap support than 70 were collapsed into polytomies using nw_ed from Newick Utilities (v1.6; Junier & Zdobnov 2010). Using these 37 individual gene trees as input, coalescence-based species tree was estimated with ASTRAL-III (v5.7.3; Zhang et al. 2018a). In all runs of IQ-TREE the best scoring substitution matrix was simultaneously estimated by ModelFinder Plus (-MFP; Kalyaanamoorthy et al. 2017) and the best fitting substitution model was used to reconstruct the resulting phylogenetic tree. The cophyloplot function of the R package "ape" (v5.4; Paradis et al. 2004) was used for visualisation.

Results and discussion

Assembly quality and completeness

The final assembly of *L. apterus* comprised 66,933 scaffolds with an N50 value of 8,902 bp (Table 1). The total length of the genome was estimated to be 286.93 Mbp, comparable with other beetle genomes published to date, and with the estimated size by GenomeScope (252.49 Mbp; Figure 2). The GC content of the final assembly is 31.66%, similar to other beetle genomes. GenomeScope results showed a relatively high heterozygosity rate (0.148%) and a low percentage of unique sequences (55.8%) which was probably due to the combined dataset (Figure 2). An additional peak was formed by the high frequency (0.864%) of duplicated k-mers which predicts high proportion of repeats in the genome (Vurture et al. 2017). High ratio of repeat regions together with short reads sequencing can lead to fragmented genome assemblies as repeats are often longer than the reads (Nadalin et al. 2012). Therefore, low contiguity of our assembly, even after merging assemblies generated by diverse softwares, is likely due to the high repeat sequence content to which the lack of a closely related reference genome contributed as well. Nevertheless, the *L. apterus* assembly has high gene completeness, since 93.5% of BUSCOs from the Endopterygota database were detected (Table 2).

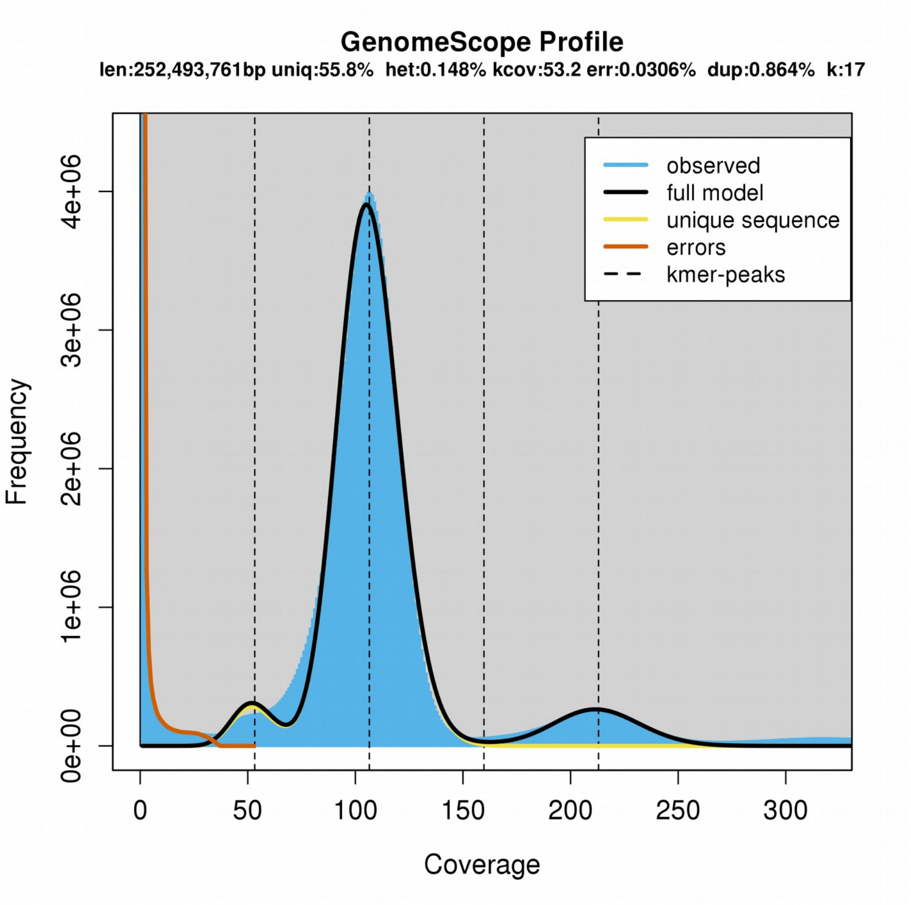


Figure 2. Genome and read characteristics produced by GenomeScope. Len: haploid genome length; uniq: overall length of unique (i.e. not repetitive) sequences; het: heterozygosity rate; kcov: mean k-mer coverage for heterozygous sequences; err: error rate of reads; dup: read duplication rate; k: k-mer length.

Table 1. Descriptive statistics of different assemblies produced during analyses.

Statistics	**MEGAHIT**	**MSG**	**SOAP**	**GAM**	**GAM_501**
Number of scaffolds	146,216	128,507	207,501	127,406	66,933
Longest scaffold (kbp)	99.34	115.32	125.01	114.98	114.98
Total length (Mbp)	306.62	307.02	230.73	307.78	286.93
N50 (bp)	7,043	8,046	5,406	8,140	8,902
GC content (%)	31.81	31.71	38.62	31.63	31.66

MEGAHIT: assembly produced by MEGAHIT; MSG: assembly produced by the MEGAHIT-SSPACE-GapFiller pipeline; SOAP: assembly produced by SOAPdenovo2; GAM: assembly produced by merging assemblies MSG and SOAP; GAM_501: the GAM assembly with contaminant and short contigs removed (for details see Materials and methods).

Table 2. Completeness of the different assemblies assessed by BUSCO.

BUSCOs	**MSG**	**(%)**	**SOAP**	**(%)**	**GAM**	**(%)**	**GAM_501**	**(%)**	**Predicted genes**	**(%)**
Complete	1,763	83.0	1,906	89.8	1,761	82.9	1,985	93.5	1,927	90.7
Single-copy	1,749	82.3	1,896	89.3	1,746	82.2	1,969	92.7	1,128	53.1
Duplicated	14	0.7	10	0.5	15	0.7	16	0.8	799	37.6
Fragmented	110	5.2	131	6.2	112	5.3	91	4.3	92	4.3
Missing	251	11.8	87	4.0	251	11.8	48	2.2	105	5.0
Total	2,124	100.0	2,124	100.0	2,124	100.0	2,124	100.0	2,124	100.0

Column headers are explained in legend of Table 1.

The distribution of contigs analysed for their coverage and GC content state-space resulted in the scaffolds separating into two groups according to their read coverage (Figure 3A). Both groups contained genes identified by the BUSCO analysis. The quotient of the coverage of the two groups was 3/4 and the proportion of males and females in the combined sample (see Materials and methods) was 1:1, suggesting that the group of scaffolds with lower coverage could reflect sequences from the sex chromosomes. This is further supported by Figure 3B showing that the lower coverage group of scaffolds are only present in males. This suggests an XY or X0 sex-determination system in *L. apterus*.

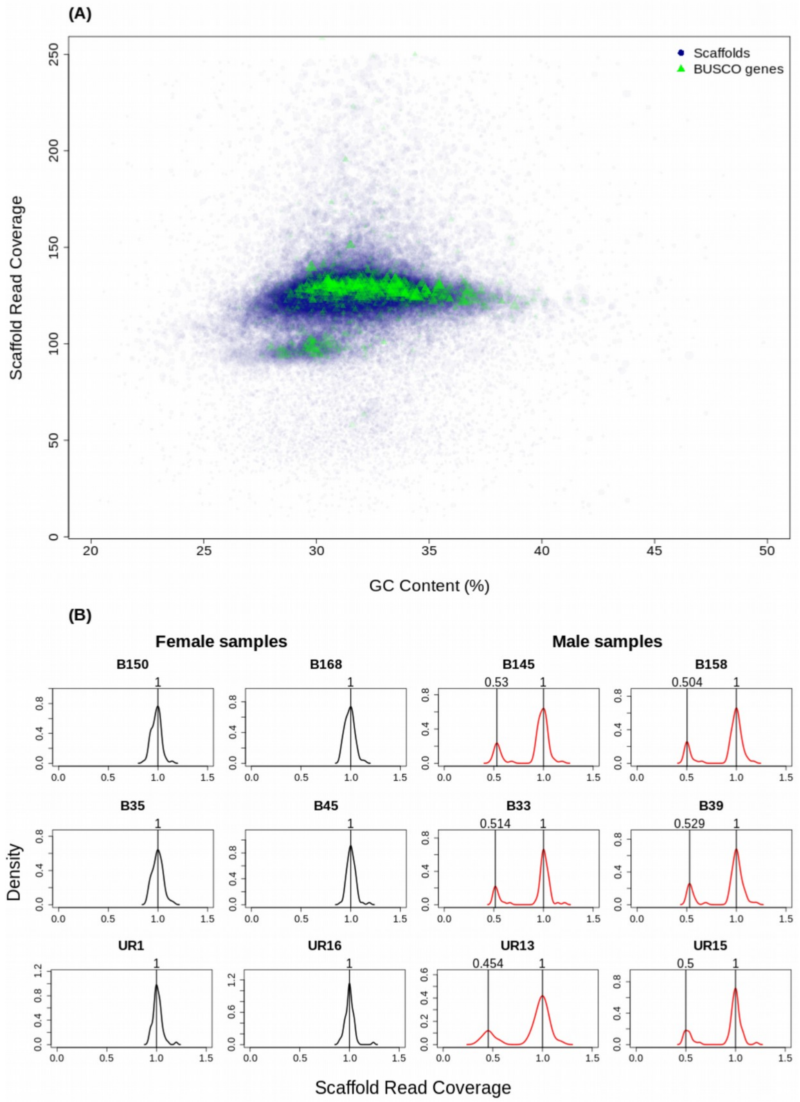


Figure 3. Read coverage distribution. (A) The distribution of scaffolds in the GC content – read coverage state space. Blue symbols mark scaffolds, size of the symbol is proportional to the length of the scaffold. Green symbols show BUSCO genes, size of the symbol is proportional to the length of the scaffold containing the gene. (B) Density plots of read coverage of scaffolds of size longer than 5 kbp. Female samples are shown with black whereas males are shown with red lines. The coverage values are rescaled so that the coverage value with maximum density is one. Numbers at the top of the plot show the coverage values of the peak densities. The relative coverage was 0.504 ± 0.03 in the lower coverage group of males.

Genome annotation

RepeatMasker was used to identify repetitive elements in the assembly of *L. apterus*. Results showed that a high proportion (36.44% of bases) of the genome contains interspersed repeats, most of which (71.46%) could not be classified as known repeats (Table 3). The most abundant repeats were A-rich sequences with low complexity. Only three of the next 10 repeat classes showed significant matches with the NCBI nt database. One matched with an inverted repeat in the pannier region of the harlequin ladybeetle (*Harmonia axyridis*), the other two had hits with uncharacterised genome regions of the mountain pine beetle (*D. ponderosae*) and the ringlet butterfly (*Aphantopus hyperantus*).

Table 3. Repetitive elements found by RepeatMasker.

Element type	Num	LO (Mbp)	PS
Total interspersed repeats	421,620	104.58	36.44%
SINEs	0	0	0.00%
LINEs	42,401	11.19	3.90%
LTR elements	2,279	1.00	0.35%
DNA elements	60,124	17.66	6.15%
Unclassified	316,816	74.73	26.04%
Small RNA	0	0	0.00%
Satellites	0	0	0.00%
Simple repeats	77,082	3.23	1.12%
Low complexity	17,892	0.88	0.31%

Num: number of elements; LO: length occupied in mega base-pairs; PS: percentage of sequence.

Ab initio gene prediction resulted in 25,385 sequences whereas homology-based prediction found 22,551. After merging and filtering the two gene sets, 34,392 remained from which 20,734 were functionally annotated by either InterProScan or Diamond using different databases. The annotated gene set had a 1,425 bp mean

CDS length, 475 amino acids mean protein length and contained 5.10 exons and 4.10 introns per gene on average.

Based on the functional annotation, the potential sex chromosome related genes coded mostly proteins necessary for cell maintenance, including housekeeping genes and mitochondrial proteins, however, some transposons and retrotransposons were also found. In addition, proteins involved in the innate immune responses, circadian rhythm and memory were also identified.

Annotation of reproductive behaviour related genes

Based on a literature search, 23 candidate genes or gene families were found in 21 research articles (Table 4). Of these, 19 were found in coleopteran species and were stored in NCBI. All 19 candidates had significant hits with the predicted genes of *L. apterus*. Compared to the other examined beetle species, *L. apterus* had lower number of hits of odorant-binding and pheromone-binding proteins. These molecules play a significant role in recognition of the signals of the environment, such as food resources or recognition of conspecifics (Fan et al. 2011). The loss of these genes may be an excellent starting point of research on the evolution of olfactory perception among dung beetles, however, we should note that the low number of hits could be caused by the fragmented genome hence further investigation would require additional data to include, such as RNA sequencing. In addition, two of the 19 candidates, namely troponin C, and octopamine receptor were found among genes located on potentially sex chromosomes and thus may serve as targets for future research on gene regulation of reproductive behaviour.

Table 4. Candidate genes and gene families involved in reproductive behaviour among coleopterans.

Gene name (reference)	**NCBI hits in Coleoptera**	**Hits in Tc**	**Hits in Nv**	**Hits in Ot**	**Hits in Av**	**Hits in La**
Fruitless (Yapici et al. 2008)	12	33	28	34	26	28
Sex peptide receptor (Yapici et al. 2008; Hanin et al. 2011)	15	29	10	14	10	7
Apolipophorin-III (Park et al. 2012; Benowitz et al. 2017)	3	1	2	8	1	3
Octopamine receptor (Cunningham et al. 2015)	73	64	57	61	49	64
Insulin receptor substrate (Woodard et al. 2014)	31	12	2	5	0	3
Krüppel homolog (Woodard et al. 2014)	2	3	4	2	1	2
Target of rapamycin (Woodard et al. 2014)	5	1	1	1	1	1
Odorant binding protein (Rehan et al. 2014; Kim et al. 2017; Wu et al. 2017)	384	55	66	54	40	20
Glucose oxidase	0	-	-	-	-	-

(Rehan et al. 2014)						
Alpha-glucosidase precursor (Rehan et al. 2014)	0	-	-	-	-	-
Troponin C (Rehan et al. 2014)	31	33	37	35	19	29
Vitellogenin (Roy-Zokan et al. 2015)	27	5	7	8	4	6
Vitellogenin receptor (Roy-Zokan et al. 2015)	7	11	12	17	11	13
Juvenile hormone acid o-methyltransferase (Wijesekera et al. 2016)	8	4	8	30	1	13
Malvolio (Mehlferber et al. 2017)	21	4	5	5	2	5
Neuropeptide F (Trumbo 2018)	5	1	2	3	1	2
Methyl geranate (Trumbo 2018)	0	-	-	-	-	-
Odorant receptor (Zhao et al. 2019; Wu et al. 2017)	418	266	51	83	259	63
Pheromone-binding protein (Senthilkumar & Srinivasan 2019)	26	36	52	40	33	13
Cryptochrome (Xu et al. 2019)	4	2	2	1	1	1

Sex peptide (Aigaki et al. 1991)	0	-	-	-	-	-
Accessory gland protein (Ram & Wolfner 2007; Dottorini et al. 2007)	2	3	6	1	2	1
Insulin-like peptide (Wigby et al. 2011)	11	7	4	13	5	4

Tc: *T. castaneum*; Nv: *N. vespilloides*; Ot: *O. taurus*; Av: *A. verrusocus*; La: *L. apterus*

Fragmented genome assembly can lead to over-prediction of paralogous genes, especially in case of gene families with a high number of similar members (Indrischek et al. 2016). The results of the candidate gene search, however, showed that the number of hits in genes of *L. apterus* and the related species were similar, suggesting that the low contiguity of the assembly did not influence the gene prediction.

Comparison with other coleopteran species

14 coleopteran proteomes available on NCBI and predicted *L. apterus* genes were used to perform comparison of orthologous genes by Orthofinder. Based on the results, 357,992 genes (95.9% of the total number of genes) were assigned to 23,528 orthogroups. All species were present in 4,754 orthogroups of which 44 included only single-copy genes. 9,618 orthogroups were species-specific from which 607 (consisting of 1,664 genes) were specific to *L. apterus*. Of the predicted genes, 436 were not assigned to any orthogroups. Finally, 42.1% of the orthogroups contained *L. apterus* genes. Our phylogenetic results are in line with relationships described in Zhang et al. 2018b. Based on the species trees reconstructed with two

independent methods, *N. vespilloides* appeared to be the sister taxa of *L. apterus+O. taurus*. Monophyly of these groups received high statistical support in all of our analyses (Figure 4). This branching not only marks the divergence of Staphylinoidea and Scarabeoidea, but also separates those three species in our dataset that have biparental care. Further studies are now needed to more precisely decipher the origin of biparental care among beetles.

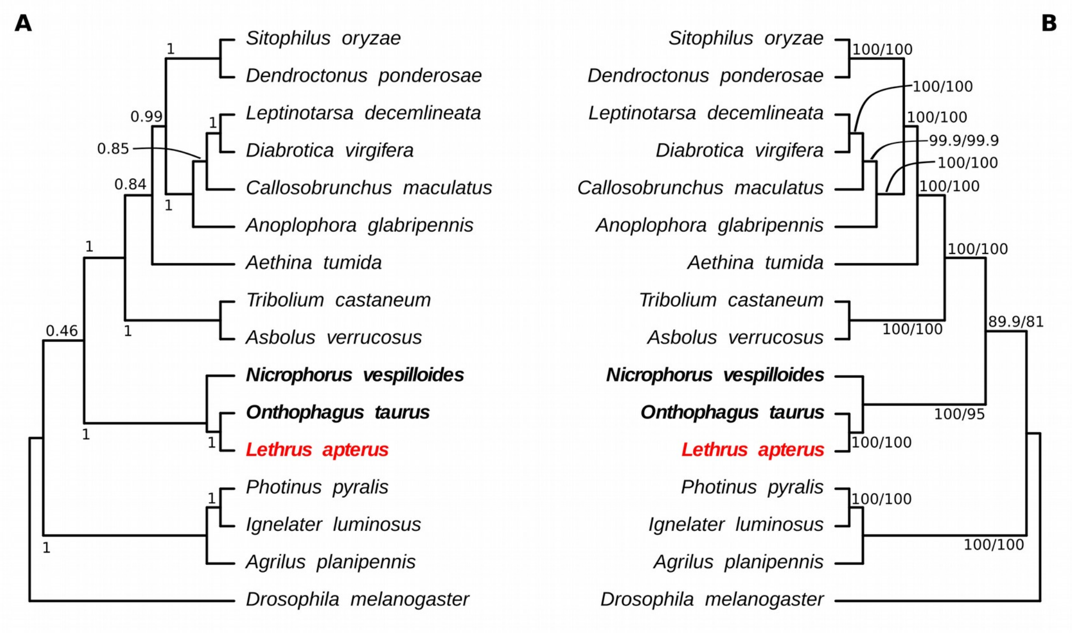


Figure 4. Phylogenetic relationships of 14 coleopteran species rooted with *D. melanogaster* as outgroup. Tree was constructed based on the 44 common single-copy protein sequences. Species with parental care are highlighted in bold and *L. apterus* is additionally highlighted in red. (A) Phylogram generated with coalescent based estimation, support values are local posterior probabilities. (B) Concatenation-based phylogenetic tree, support values are ultrafast bootstrap/aLRT.

Study 3: Evaluation of potential reference genes for gene expression studies on the parental care of *Lethrus apterus*[3]

Objectives

Investigation of the hormonal background of a process in question often starts with only a few candidate genes. With the sequences of the target genes, the next step usually is to analyze the expressed gene, i.e. the RNA levels in samples, and compare them between the conditions of interest (VanGuilder et al. 2008). One of the frequently used methods for gene expression measurement is real-time quantitative polymerase chain reaction (RT-qPCR) which is an effective and sensitive approach to estimate the expression levels of a moderate number of target genes (Radonić et al. 2004). One critical step of the analysis is to normalize the expression values of the target genes in order to compensate for technical errors, e.g. pipetting errors emerging during the multi-step process (Kozera & Rapacz 2013). For this purpose, reference genes are selected of which the expression is not influenced by the examined conditions (Vandesompele et al. 2002). These reference genes are usually selected from housekeeping genes, i.e. genes that are responsible for cell maintenance processes. Accordingly, as a first step of studying specific target genes potentially involved in the regulation of parental behaviour of *L. apterus*, we aimed to find reference genes with stable expression regardless of the sex and the part of the breeding period. For this purpose, we sampled equal numbers of males and females of a natural population three times during the breeding period, and evaluated 11 candidate reference genes identified in the draft genome (GenBank accession number: GCA_018397195.1). Additionally, the optimal number of reference genes has been determined.

---

[3] This is a summary of the article published as: **Nagy, N. A.**, Németh, Z., Juhász, E., Póliska, S., Rácz, R., Kosztolányi, A., & Barta, Z. (2017). Evaluation of potential reference genes for real-time qPCR analysis in a biparental beetle, *Lethrus apterus* (Coleoptera: Geotrupidae). *PeerJ, 5*, e4047. The full article is provided in the RAttachments section.

Results

According to our results, the usage of two reference genes is required for accurate normalization for our conditions. The consensus results of the used methods showed that for head samples, ribosomal protein L7A (L7A) and ribosomal protein RP18 (RP18) are the best choices for expression data normalization, whereas, in thorax samples, the best gene pair was ribosomal protein RP4 (RP4) and L7A (Figure 5). Considering only females, RP18 and L7A are the most reliable reference genes in both head and thorax samples. In case of male head samples, ribosomal protein S8 (RPS8) and L7A should be used for normalization, whereas in thorax samples, the best two genes are L7A and elongation factor 2 (EF2). These genes were stably expressed regardless of the sampling date in the breeding season.

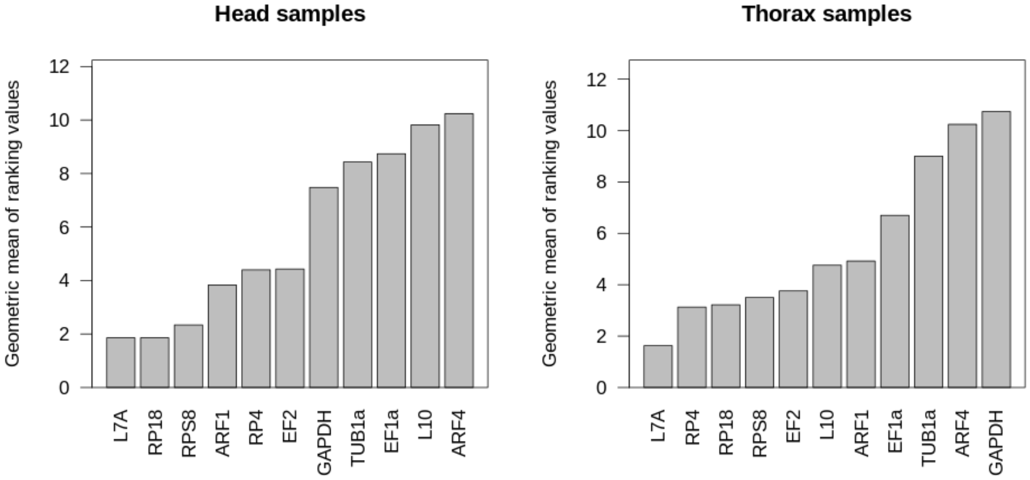


Figure 5. Ranking of the candidate reference genes based on the comprehensive results of RefFinder. Lower ranking values mean more stable expression.

Discussion

Draft genome of *L. apterus* was used to find candidate reference genes for data normalization in RT-qPCR studies. Selected genes were evaluated in male and female samples through the breeding period to find the best reference genes for gene expression studies examining the genetic and hormonal background of the parental care of this biparental species. Both for head and thorax samples, ribosomal proteins were found to be the most reliable reference genes. Ribosomal proteins are frequently used genes for normalizing data from different sexes in other coleopteran species as well, e.g. ribosomal protein L22e in the blister beetle (*Mylabris cichorii*; Wang et al. 2014), ribosomal protein L19 in the cabbage beetle *Colaphellus bowringi* (Tan et al. 2015) and ribosomal protein 49 in the Asian lady beetle (*H. axyridis*; Yang et al. 2018). On the other hand, we evaluated two commonly used reference genes, glyceraldehyde 3-phosphate dehydrogenase (GAPDH) and tubulin alpha-1 chain (TUB1a) among the least stable ones for our conditions. These results support the findings that no universal reference gene can be determined for all species, instead, housekeeping genes should be evaluated for each species and specific condition (Thellin et al. 1999).

Study 4: Inotocin, a potential modulator of reproductive behaviour in a biparental beetle, *Lethrus apterus*[4]

Nikoletta A. Nagy, Zoltán Németh, Edit Juhász, Szilárd Póliska, Rita Rácz, Johanna Kiss, András Kosztolányi, Zoltán Barta

Introduction and objectives

Oxytocin, vasopressin and their orthologs compose a highly conserved neuropeptide family as indicated by their appearance in numerous animal taxa. Orthologs were found in mammals, birds, amphibians, fishes and also in invertebrates including insects, molluscs, and even earthworms and nematodes (Beets et al., 2013). These peptides are of great interest, as not only their structures are homologous but their roles in physiology and behaviour appear to be remarkably similar even in distant species (Hanoune, 2010).

One of the ancient roles of this peptide family is related to the regulation of water balance (Hanoune, 2010). For example, vasotocin has antidiuretic functions in birds and amphibians (McCormick and Bradshaw, 2006), similarly to the role of vasopressin in mammals (Banerjee et al., 2017). An osmoregulatory role of vasotocin was also described in fish (Balment et al., 2006). Moreover, relationship between vasopressin-like peptides and water homeostasis was also found in several invertebrate species, including leeches (*Whitmania pigra* and *Erpobdella octoculata*; Fujino et al., 1999, Salzet et al., 1993) and pleated sea squirt (*Styela plicata*; Ukena et al., 2008).

Another widespread role of these peptides is their involvement in the modulation of affiliative and reproductive behaviour across distant taxa as worms, molluscs, insects and vertebrates (Donaldson and Young, 2008). For instance, oxytocin is responsible for maternal nurturing, and vasopressin affects paternal care

---

[4] Accepted as: Nagy, N. A., Németh, Z., Juhász, E., Póliska, Sz., Rácz, R., Kiss, J., Kosztolányi, A. & Barta, Z. Inotocin, a potential modulator of reproductive behaviours in a biparental beetle, *Lethrus apterus*. *Journal of Insect Physiology*

and mate guarding in prairie vole (*Microtus ochrogaster*; McGraw et al., 2010), activation of oxytocin receptor is required for stable pair formation in zebra finch (*Taeniopygia guttata*; Klatt and Goodson, 2013), and vasotocin induces egg-laying behaviour in rough-skinned newt (*Taricha granulosa*; Moore et al., 1992). Among invertebrates, lysine-conopressin-G regulates male copulatory behaviour in great pond snail (*Lymnaea stagnalis*; van Kesteren et al., 1995), annetocin plays a role in egg-laying behaviour in brandling worm (*Eisenia fetida*; Oumi et al., 1996), and arginine-conopressin-G, annetocin, as well as hirudotocin activate a sequence of reproduction-related behaviour in a medicinal leech (*Hirudo verbana*; Wagenaar et al., 2010).

The insect ortholog of the oxytocin/vasopressin peptide family, inotocin, was found in 22 out of 29 insect orders investigated using genome annotations (Liutkeviciute et al., 2016). Within the 20 previously examined coleopteran species, all but one have only one type of inotocin with the same amino acid sequence. The genome of *L. decemlineata*, however, codes two different types of this peptide, both of them differ in the 8th amino acid position from the inotocin of other coleopterans (Liutkeviciute et al., 2016).

Even though inotocin appears to be widespread across insects, its function and the specific behaviour it affects have only been investigated in a few species (reviewed in Muratspahić et al., 2020). In migratory locust (*Locusta migratoria*) and red flour beetle (*T. castaneum*), inotocin was found to be involved in controlling water balance (Proux et al., 1987, Aikins et al., 2008). Considering the conserved function in reproductive behaviour, an association between inotocin and metabolism was discovered recently in ants which is important as the reproduction and hence the related behaviour both depend on the energy availability (Liutkevičiūtė et al., 2018). A potential function of inotocin in reproductive social behaviour has also been suggested in black garden ant queens (*Lasius niger*, Chérasse and Aron, 2017), however, in this study only the expression of inotocin receptor, but not the hormone was measured. Based on the examples above and the conserved functions of the oxytocin/vasopressin system over evolutionary time, one would expect that besides the ancient diuretic function, inotocin also modulates the expression of social behaviour, however, evidence is scarce. Thus, it is an intriguing question whether inotocin may have a function in shaping social behaviour of insects similar to that of the vertebrate orthologs.

Here, we report a new form of inotocin for beetles and its receptor from *L. apterus*. Based on the knowledge so far on its function (see above), inotocin in *L. apterus* may regulate water balance and/or play a role in the control of reproductive social behaviour, probably including parental care. To uncover the role of inotocin in water management we experimentally exposed beetles from a wild population to different levels of humidity and examined the gene expression of both the preprohormone (hereafter inotocin) and the receptor. If inotocin has a function in controlling reproductive behaviour in general and parental care in particular, one would expect that the expression levels of inotocin and its receptor increase over the breeding season in concurrence with the change in the behaviour of beetles shifting from mate searching, through pair formation to parental care. To investigate this possibility, we sampled beetles over the reproductive season in the field to measure the gene expression level of inotocin and the receptor. Coinciding with this sampling we also performed field surveys to observe the reproductive behaviour of the beetles.

Materials and methods

The effect of humidity on expression of inotocin and its receptor

An experiment was performed on 24th April 2018 at Susa, Hungary (48°16'27"N, 20°15'08"E) to examine the effect of humidity on the expression of inotocin and its receptor. For this purpose, individuals were collected in the field during their daily activity, and three experimental groups were created (dried, saturated and control), each consisting of five males and five females. Individuals in the dried and saturated groups were placed for four hours in 50 ml conical centrifuge tubes which contained 20 ml indicating silica gel (Qingdao Fraken International Trading Co. Ltd, Orange indicating silica gel) which is a widely used hygroscopic agent in insect desiccation experiments (Andersen et al. 2010, Pallarés et al. 2016). For the dried group, water was completely eliminated from the silica gel by heating at 180 °C in a dry heat steriliser for two hours. In these tubes 12.7 ± 1.34 % (mean ± SD) humidity was measured with a digital hygrometer (Exo Terra PT2477 Digital Hygrometer). For the saturated group, silica gel was saturated in an experimental box in 95% humidity for six hours. In tubes filled with saturated silica gel, the humidity was 41.8 ± 3.85 % (mean ± SD). Sponge plugs were used as barriers to separate beetles from silica gel during the experiment. For the control group, samples were taken immediately, whereas sampling of individuals from experimental groups was performed after the four hours of treatment. During sample collection for the control group, relative humidity was 53.75 ± 8.02 % in the field based on four measurements.

Sample collection was approved by the Northern Hungarian Inspectorate for Environment Protection and Nature Conservation (No. 9007-8/2014) since *L. apterus* is a protected species in Hungary. Because the brain of *L. apterus* is surrounded by other tissues from which it is difficult to distinguish with the naked eye in the field, we decided, similarly to other insect species studies of expression analysis (Chérasse and Arun, 2017, Stafflinger et al., 2008), to remove all tissues from the head capsule. Each sample was immediately put in separate 1.5 ml tubes filled with 600 µl RNAlater® Stabilization Solution (Thermo Fisher Scientific, Waltham, MA, USA) in order to inhibit RNase enzyme activity. Tissue samples were taken in the field in less than five minutes after collecting individuals. Samples were stored at -20°C in the laboratory until RNA extraction.

Field observations on the seasonal behaviour of *Lethrus apterus*

To investigate the phenology and reproductive behaviour of *L. apterus*, line transect monitoring was conducted between 16 March and 23 May in 2016 in the

Susa population. During the survey, we walked on four previously selected line transects (length 285.45 ± 7.08 m, mean ± SD) covering an area of approximately 1 ha. The activity of every detected specimen was recorded as four categories (resting, travelling, leaf-carrying, fighting). In case of leaf-carrying, the taxonomic identity of the carried leaf was also determined. Surveying was usually carried out between 10 am and 6 pm depending on weather conditions and in total we had 40 survey days.

Sample collection for measuring seasonal gene expression of inotocin and its receptor

Sample collection over the reproductive period was carried out in Northern Hungary in two years. In 2015, samples were collected at Dorogháza (47°59′29″N, 19°53′36″E) on three occasions (16th April, 4th May and 28th May). In 2016, samples were collected at Susa on five occasions (18th March, 1st and 15th April, and 2nd and 17th May). On each sampling date, head samples were taken from eight males and eight females, except on 18th March 2016 when we sampled only four male and four female beetles because of the scarcity of individuals – the breeding season has just begun.

RNA extraction and cDNA synthesis

Total RNA was extracted from each sample using TRIzol Reagent (Thermo Fisher Scientific, Waltham, MA, USA) following the manufacturer's instructions. The isolated RNA was eluted in 15-30 µl RNase-free water, depending on the pellet size. Concentration and quality of the extracts were determined by NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). Genomic DNA was eliminated using RQ1 RNase-Free DNase (Promega, Madison, VI, USA) just before the reverse transcription. First-strand cDNA synthesis was performed from 1 µg DNA-free RNA using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems, Foster City, CA, USA).

Gene identification and primer design

Genes for inotocin and its receptor were identified in the draft genome of *L. apterus* (GenBank accession number: GCA_018397195.1) based on homologous sequences from *T. castaneum* using BLASTX (BLAST 2.2.31+; Camacho et al., 2009). Gene structures were determined using transcriptome data (unpublished data). For the two genes (GenBank accession numbers in Table 5), primers were designed manually using the web-based Sequence Manipulation Suite (Stothard, 2000). In addition, Multiple Primer Analyzer (Thermo Fisher Scientific, Waltham, MA, USA) was applied to check that our primers do not form any secondary structures. For normalisation, two reference genes were used, ribosomal protein L7A and L18, for which primers described in Nagy et al. (2017) were applied. Sequences of the primers are shown in Table 5.

Table 5. Primers used to measure the expression levels of the two targets and the two reference genes by RT-qPCR.

Gene	Acc. no	Primer sequence (5'-3')	AL (bp)	MT (°C)	E (%)	R^2
In	MT920922	F: ATGTTTAAAATCGTCGTTC R: CAAGCAACCGAAAGTTCCA	169	81.0	109.37	0.90
InR	MT920923	F: AGATGTCCATGAACAATACG R: CGTTATATTCCTGGTGACC	331	80.08	101.44	0.97
L7A	KY786277	F: TAGCGACTCAACTGTTCAAGG R: CCTCAATTGGATCGACGTCATGTG	224	84.8	99.54	0.95
RP18	KY786276	F: TTGTAACCACATGAACGCCTACG R: AGTTAGCTTTACGTTCACCTACTGG	186	85.2	99.75	0.96

In: inotocin; InR: inotocin receptor; L7A: ribosomal protein L7A; RP18: ribosomal protein L18; Acc. no: GenBank accession number; AL: amplicon length; MT: melting temperature; E: efficiency.

Real-time quantitative PCR

Real-time qPCR was performed on a QuantStudio 12K Flex Real-Time PCR System (Applied Biosystems, Foster City, CA, USA) using SYBR® Green PCR Master Mix (Applied Biosystems, Foster City, CA, USA) and ROX Passive Reference Dye (Affymetrix, Santa Clara, CA, USA). Amplifications were carried out under the following conditions: initial denaturation at 95 °C for 10 min followed by 40 cycles of 10 sec at 95 °C and for 1 min at the optimal annealing temperature. For melting curve analysis, temperature was raised from 65 °C to 95 °C in sequential steps of 0.05 °C for 1 sec. Three technical replicates were performed for each biological sample, and the average cycle threshold (Ct) values of triplicates were calculated. A no-template control was included in each run for each gene to exclude primer-dimers and non-specific contamination. Standard curves were created with five 5-fold dilution series from cDNA samples. Products of the qPCR were sequenced in order to check whether the correct sequences were amplified. Due to logistical reasons, inotocin expression was measured only in samples from Susa from 2016 and in samples from the humidity manipulation experiment (Susa, 2018), whereas inotocin receptor was measured in samples from Dorogháza (2015), Susa (2016) and the humidity manipulation experiment (2018).

Statistical analysis

Prior to statistical analysis, expression of inotocin and its receptor were normalised with the geometric mean of the two reference genes. As samples were collected from natural populations, no control group was determined, thus for the $2^{-\Delta\Delta Ct}$ method, sample with the highest Ct value was used as control sample for each gene (Pabinger et al., 2014).

All statistical analyses were carried out using the R statistical environment version 3.6.0 (R Core Team, 2019). Before analyses, the expression values were log-transformed. Significance of sampling dates (hereafter date) or treatment and sex and their interactions were investigated by linear models in each site separately. During model selection, non-significant interactions were excluded (all $p>=0.05$). All main effects of interest were retained in the final models and their significance were investigated with F-test. Pairwise post-hoc comparisons of the expression levels of inotocin and the receptor between sampling dates were performed using the “emmeans” package and Tukey adjustment (Lenth, 2020).

In case of inotocin measured in both the humidity manipulation experiment and the seasonal pattern analysis, expression levels split into two groups (Figure S1). To search for the reason behind the bimodal distribution, possible confounding effects were investigated (see Appendix for details). However, none of these effects explained the apparent clustering of expression levels. Therefore, a two-level factor explaining the expression groups was included in the linear models to maintain the unimodal normal distribution of errors assumption of the models. The cutoff values for the two groups were determined as the minimum of the estimated density function of expression values by two-group mixture models using the “mixsmsn” package (Prates et al., 2013).

Results

Structure of inotocin and its receptor

Both inotocin and its receptor were identified in the draft genome of *L. apterus*. Based on the comparison of mRNA sequence and the genome, no intron intermits the gene coding for the 150 amino acids long inotocin preprohormone from which the mature nine amino acids long peptide is formed by cleavage steps. Compared with the mature inotocin sequences of beetle species reported in Liutkeviciute et al. (2016), we found that valine instead of threonine is coded in position 4 of the nonapeptide of *L. apterus* (Figure 6). However, in the neurophysin part, the highly conserved 14 cysteine residues were identified similarly to other species (Stoop, 2012).

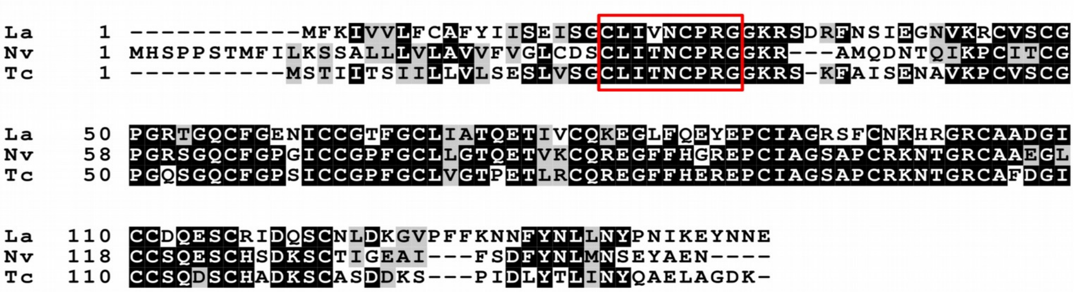
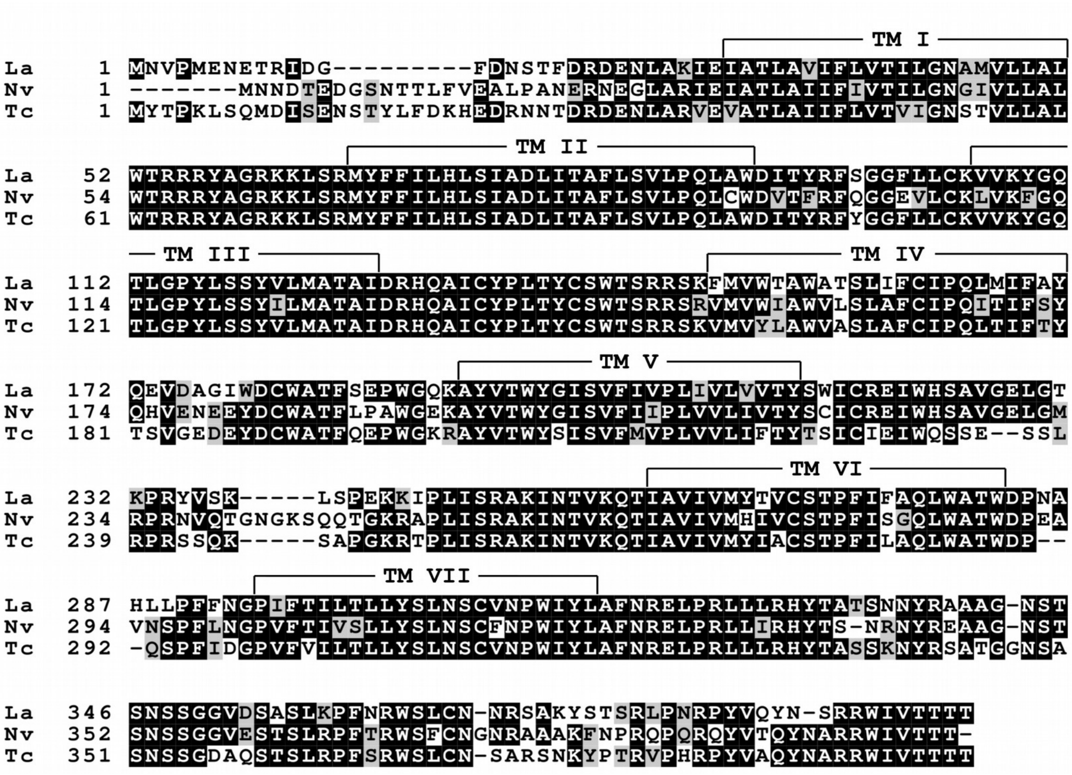


Figure 6. Multiple sequence alignment of the inotocin preprohormone of the *L. apterus* (La, GenBank accession number: MT920922), the biparental *N. vespilloides* (Nv, XP_017777933) and the model species *T. castaneum* (Tc, NP_001078831). The red box highlights the mature peptide.

In case of the receptor, the comparison of mRNA and genomic sequence revealed 5 introns in the gene. The whole protein consists of 397 amino acids and has seven transmembrane domains that are characteristic of G protein-coupled receptors (Figure 7).



```
La     1 MNVPMENETRIDG---------FDNSTFDRDENLAKIEIATLAVIFLVTILGNAMVLLAL
Nv     1 -------MNNDTEDGSNTTLFVEALPANERNEGLARIEIATLAIIFIVTILGNGIVLLAL
Tc     1 MYTPKLSQMDISENSTYLFDKHEDRNNTDRDENLARVEVATLAIIFLVTVIGNSTVLLAL

La    52 WTRRRYAGRKKLSRMYFFILHLSIADLITAFLSVLPQLAWDITYRFSGGFLLCKVVKYGQ
Nv    54 WTRRRYAGRKKLSRMYFFILHLSIADLITAFLSVLPQLCWDVTFRFQGGEVLCKLVKFGQ
Tc    61 WTRRRYAGRKKLSRMYFFILHLSIADLITAFLSVLPQLAWDITYRFYGGFLLCKVVKYGQ

La   112 TLGPYLSSYVLMATAIDRHQAICYPLTYCSWTSRRSKFMVWTAWATSLIFCIPQLMIFAY
Nv   114 TLGPYLSSYILMATAIDRHQAICYPLTYCSWTSRRSRVMVWIAWVLSLAFCIPQITIFSY
Tc   121 TLGPYLSSYVLMATAIDRHQAICYPLTYCSWTSRRSKVMVYLAWVASLAFCIPQLTIFTY

La   172 QEVDAGIWDCWATFSEPWGQKAYVTWYGISVFIVPLIVLVVTYSWICREIWHSAVGELGT
Nv   174 QHVENEEYDCWATFLPAWGEKAYVTWYGISVFIIPLVVLIVTYSCICREIWHSAVGELGM
Tc   181 TSVGEDEYDCWATFQEPWGKRAYVTWYSISVFMVPLVVLIFTYTSICIEIWQSSE--SSL

La   232 KPRYVSK-----LSPEKKIPLISRAKINTVKQTIAVIVMYTVCSTPFIFAQLWATWDPNA
Nv   234 RPRNVQTGNGKSQQTGKRAPLISRAKINTVKQTIAVIVMHIVCSTPFISGQLWATWDPEA
Tc   239 RPRSSQK-----SAPGKRTPLISRAKINTVKQTIAVIVMYIACSTPFILAQLWATWDP--

La   287 HLLPFFNGPIFTILTLLYSLNSCVNPWIYLAFNRELPRLLLRHYTATSNNYRAAAG-NST
Nv   294 VNSPFLNGPVFTIVSLLYSLNSCFNPWIYLAFNRELPRLLIRHYTS-NRNYREAAG-NST
Tc   292 -QSPFIDGPVFVILTLLYSLNSCVNPWIYLAFNRELPRLLLRHYTASSKNYRSATGGNSA

La   346 SNSSGGVDSASLKPFNRWSLCN-NRSAKYSTSRLPNRPYVQYN-SRRWIVTTTT
Nv   352 SNSSGGVESTSLRPFTRWSFCNGNRAAAKFNPRQPQRQYVTQYNARRWIVTTT-
Tc   351 SNSSGDAQSTSLRPFSRWSLCN-SARSNKYPTRVPHRPYVAQYNARRWIVTTTT
```

Figure 7. Multiple sequence alignment of the inotocin receptor of *L. apterus* (La, MT920923), *N. vespilloides* (Nv, XP_017769191) and *T. castaneum* (Tc, NP_001078830). Transmembrane regions are indicated by TMI-TMVII.

The effect of humidity on expression of inotocin and its receptor

No significant difference was found among the experimental groups either in the expression of inotocin (Figure 8A, Table 6) or in the expression of the receptor (Figure 8B, Table 6). Expression of inotocin was marginally higher in females (ß=-0.672 ± 0.37, Table 6), in contrast, receptor expression was marginally higher in males (ß=0.478 ± 0.26, Table 6).

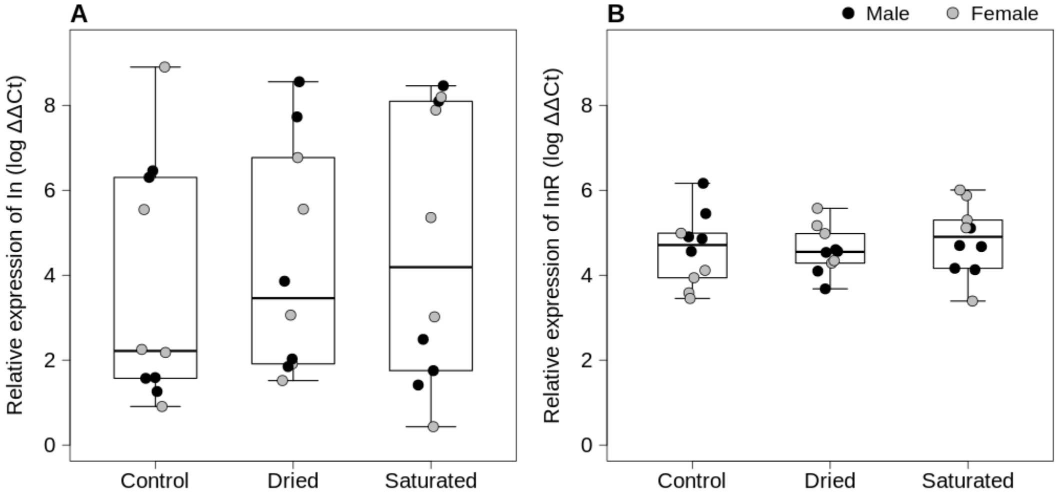


Figure 8. Effect of humidity manipulation experiment on the expression patterns of inotocin (In) and inotocin receptor (InR). Expression of inotocin (A) and its receptor (B) in the different treatment groups. Within each box, horizontal black lines denote median values; boxes extend from the 25th to the 75th percentile of each group's distribution of values; vertical extending lines denote the minimum and maximum values. Original data points are represented as dots shaded black for males and grey for females.

Table 6. Results of linear model selections on the gene expression levels of inotocin and its receptor from different years. Effect sizes of the final models (see Materials and methods) are presented as adjusted R-squared. Significant results are highlighted in bold.

Year	Gene	Effect	dfs	F value	p value
2015	Inotocin receptor ($r^2 = 0.072$)	Date	2,30	1.921	0.164
		Sex	1,30	1.822	0.187
2016	Inotocin ($r^2 = 0.897$)	Date	4,43	10.645	**<0.001**
		Sex	1,43	0.627	0.433
	Inotocin receptor ($r^2 = 0.25$)	Date	4,44	5.335	**0.001**
		Sex	1,44	0.156	0.695
2018	Inotocin ($r^2 = 0.876$)	Drying experiment	2,25	0.998	0.383
		Sex	1,25	3.227	0.085
	Inotocin receptor ($r^2 = 0.039$)	Drying experiment	2,26	0.418	0.663
		Sex	1,26	3.328	0.079

Seasonal variation in behaviour of *Lethrus apterus*

In the beginning of the active season, males were observed more frequently, by contrast, in the second half of the season females tended to be more active, i.e., they were observed above ground more frequently (Figure S8A, S8B). Increased number of leaf carrying events was observed in the second half of the breeding period of which the majority was done by females (Figure S8C, S8D). A large proportion of the collected leaves were classified as members of the Fabaceae family, especially later in the breeding season, presumably indicating the parental care period (Figure 9).

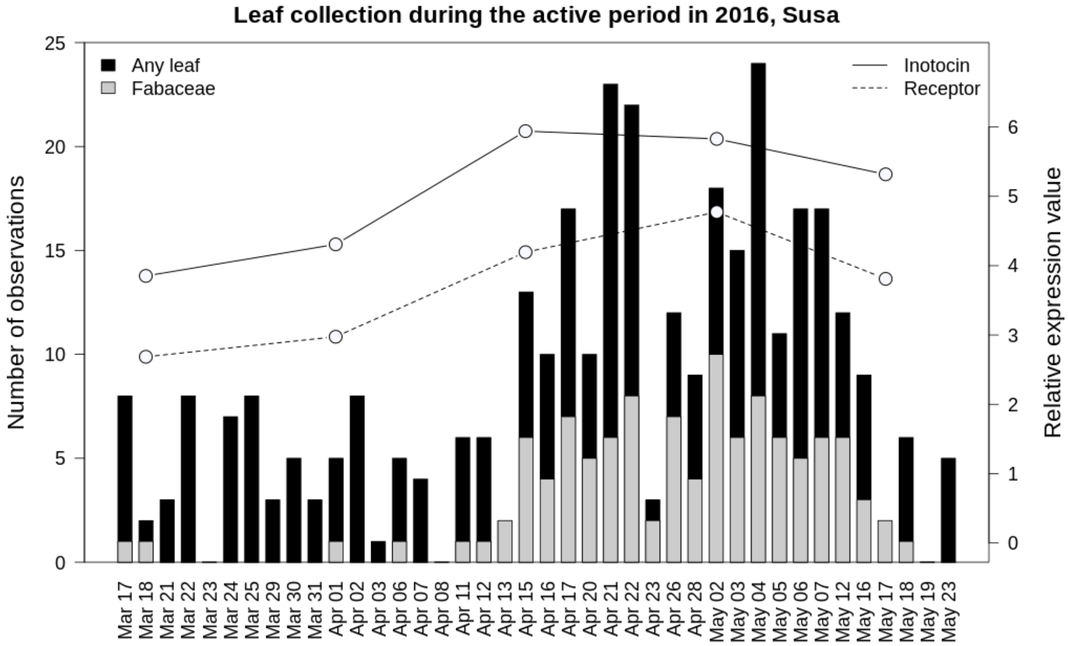


Figure 9. Number of leaf collection events of both sexes during the active period of *L. apterus* in 2016, Susa. Grey bars represent the collection of a leaf of a legume species, whereas black bars show leafs of any other taxa. Estimated marginal means of expression levels of inotocin and inotocin receptor are represented with solid and dashed lines, respectively.

Seasonal expression of inotocin

Date had a significant effect on inotocin expression (Table 6). Specifically, expression levels at the first two sampling dates differed significantly (or marginally) from the gene expression levels at the last three sampling dates (Figure 10C, Table 7) which change was in line with the increase in leaf collecting events, i.e., inotocin expression was higher during the period of parental care (Figure 9). On the other hand, no difference between sexes was detected (Table 6).

Seasonal expression of inotocin receptor

In the Dorogháza population (2015), expression of the receptor showed an increase with date but this was not significant (Figure 10A, Table 6, 7). In the Susa population (2016), on the other hand, receptor expression levels changed significantly over the season (Table 6). Based on the post-hoc comparisons, expression levels on the third date was only marginally higher than the first two sampling dates, however, a peak of receptor expression on the fourth sampling date was significantly higher than the expression on the first two sampling dates (Figure 10B, Table 7). After this peak, receptor expression tended to decrease at the end of the season. These seasonal changes in receptor expression were consistent with changes in leaf collecting behaviour, i.e. the peak of receptor expression coincided with the peak of carrying legume leaves (Figure 9). Sex had no significant effect in either population (Table 6).

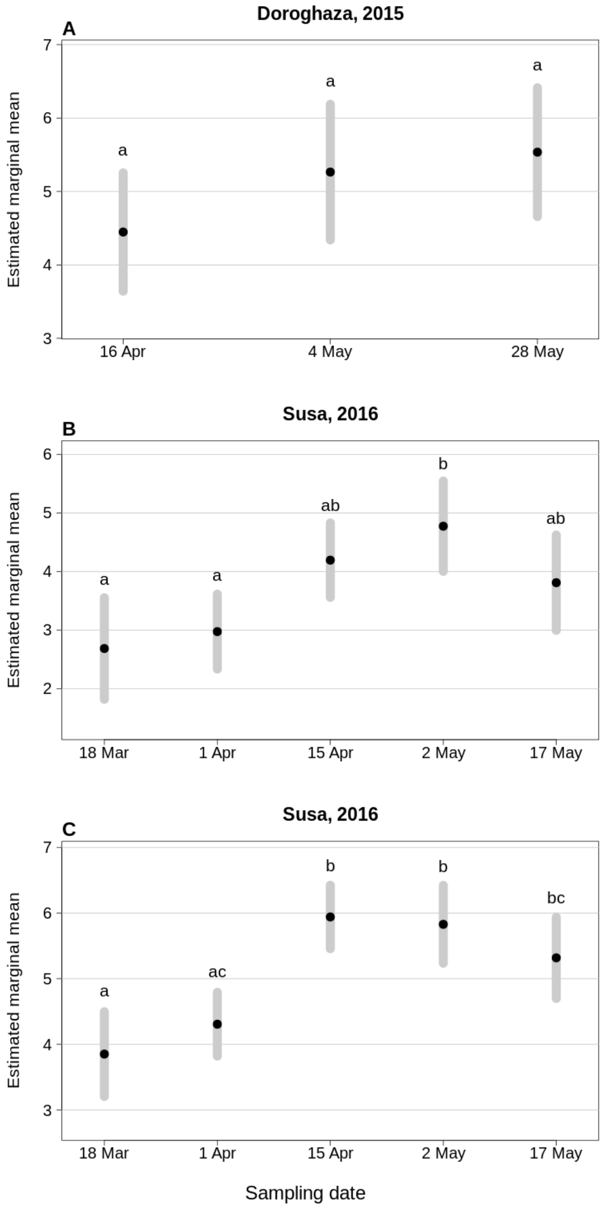


Figure 10. Expression pattern of inotocin receptor (InR) (A, B) and inotocin (In) (C) during the reproductive period of *L. apterus*. Samples were taken at five different time points in the Susa population (B, C) and three different time points in the Dorogháza population (A). Black dots represent the estimated marginal mean of expression levels, grey stripes the 95% confidence intervals for estimated marginal means. Dates with distinct letters differ significantly (Tukey-adjusted comparisons).

Table 7. Pairwise post-hoc comparisons of gene expression levels among dates in the Dorogháza and the Susa populations. Adjusted P values were calculated with the Tukey method. Results were averaged over sex, and, in case of inotocin, over the two-level factor that explains the expression groups (see Materials and methods). Significant comparisons are in bold.

Contrast		Estimate ± SE	p value
2015, Doroghaza			
Inotocin receptor (df = 30)			
16 Apr	4 May	-0.817 ± 0.599	0.3720
16 Apr	28 May	-1.088 ± 0.582	0.1654
4 May	28 May	-0.271 ± 0.622	0.9007
2016, Susa			
Inotocin (df = 43)			
18 Mar	1 Apr	-0.455 ± 0.399	0.7842
18 Mar	15 Apr	-2.089 ± 0.402	**0.0001**
18 Mar	2 May	-1.976 ± 0.439	**0.0005**
18 Mar	17 May	-1.466 ± 0.441	**0.0150**
1 Apr	15 Apr	-1.633 ± 0.346	**0.0002**
1 Apr	2 May	-1.521 ± 0.386	**0.0026**
1 Apr	17 May	-1.010 ± 0.379	0.0764
15 Apr	2 May	0.112 ± 0.368	0.9980
15 Apr	17 May	0.623 ± 0.397	0.5239
2 May	17 May	0.511 ± 0.433	0.7627
Inotocin receptor (df = 44)			
18 Mar	1 Apr	-0.289 ± 0.536	0.9827
18 Mar	15 Apr	-1.509 ± 0.534	0.0521
18 Mar	2 May	-2.088 ± 0.578	**0.0066**
18 Mar	17 May	-1.124 ± 0.591	0.3320
1 Apr	15 Apr	-1.220 ± 0.446	0.0650
1 Apr	2 May	-1.799 ± 0.493	**0.0059**
1 Apr	17 May	-0.835 ± 0.510	0.4831
15 Apr	2 May	-0.579 ± 0.494	0.7672
15 Apr	17 May	0.385 ± 0.511	0.9427
2 May	17 May	0.964 ± 0.552	0.4176

Discussion

Our findings support our expectation that inotocin may be involved in the control of reproductive social behaviour in a beetle species, *L. apterus*. Gene expression of inotocin receptor, as well as inotocin, was found to change over the course of the breeding period, i.e. expression was higher later in the season when parental behaviour dominates the activities of both sexes. In particular, changes in expression of inotocin and its receptor were in accordance with changes in legume leaf collection events. Legumes have high nitrogen content which may be necessary for fast larval development (Ohmart et al., 1985; Heisswolf et al., 2005). Examples of a similar expression pattern of oxytocin/vasopressin related peptides correlating with different forms of reproductive social behaviour can be found in vertebrates. In the biparental prairie vole, for instance, expression of the vasopressin gene was found to be higher postpartum in both males and females suggesting a role in parental behaviour (Wang et al., 2000). Furthermore, increased isotocin expression was related to affiliative behaviour toward the pair in a monogamous cichlid species, *Neolamprologus pulcher* (O'Connor et al., 2016). In case of zebra finches, with the pair formation, the expression of both mesotocin and vasotocin increased in males as well as in females (Lowrey et al., 2014). Considering the conserved functions of the oxytocin/vasopressin peptide family, higher inotocin expression after pair formation can suggest a role in affiliative or parental behaviour in *L. apterus*. Nevertheless, we cannot rule out alternative explanations for the changed expression pattern. A possible reason for this change may be the effect of the altered spectrum of collected plant species. The followings, however, make this alternative explanation less likely. First, leaves collected during the second part of the breeding season are mainly stored up as food for the larvae (Emich, 1884), hence the changed preference for plant species does not necessarily mean that the diet of the adults also changes during the breeding period. Second, legume leaf collection, which is driving the change in species composition, is largely carried out by females (Figure S7), but gene expression patterns changed with the season in both sexes. We also have to note, however, that a study found an association between oxytocin/vasopressin related peptides and metabolism in an ant (Liutkevičiūtė et al., 2018). Therefore, it is still possible that changing activity over the season influences gene expression pattern. Nevertheless, the activity pattern changes differently in sexes (Figure S8) but no sex differences were found in gene expression patterns. One could also argue that the observed expression patterns might be related to changes in immunity during the breeding period. However, there is no obvious link between immune response and season (Kiss et al., 2020a), therefore, we would not expect that seasonal changes in inotocin or inotocin receptor gene expression correlate with changes in immunity. Finally, since our investigation was performed in a natural population, we cannot exclude the possibility that seasonally changing abiotic factors, e.g., temperature influenced the gene expression patterns of inotocin and its receptor.

In the beetle *T. castaneum* higher expression levels of inotocin preprohormone and receptor were found in the head compared to other body parts by Stafflinger et al. (2008). Based on their results, Stafflinger et al. rejected the hypothesis that inotocin contributes to water balance regulation. By contrast, inotocin was found to exert an indirect effect on diuresis in the same species (Aikins et al., 2008). Furthermore, inotocin receptor has a role in desiccation resistance in an ant species, *Camponotus fellah* (Koto et al., 2019). We found, however, that the expression of inotocin and its receptor did not change as a response to low humidity stress. Although these results suggest that inotocin is not involved substantially in the regulation of water balance in the adults of this beetle species, a more robust sampling with longer experimental time and more sampling dates could provide a more reliable result to answer this question.

An interesting result of our study is that we identified a threonine-valine substitution in the fourth position of the inotocin of *L. apterus*. Threonine is the most common amino acid in the fourth position of the mature nonapeptide in arthropods, and also threonine was found in the fourth position in all other 20 beetle species studied to date (Liutkeviciute et al., 2016). This threonine-valine substitution was only identified in a highly social ant species earlier (Gruber and Muttenthaler, 2012).

Consistent bimodal distribution was found in the expression of inotocin gene in two years which was explained neither by the sex nor any of the potential confounding effects that may have arisen during the sampling or laboratory procedures (see Appendix for details). Therefore, we suggest that the source of this clustering in the expression levels may be due to some unknown biological reasons. One explanation could be that individuals of a stable pair express inotocin in higher or lower amount compared to unpaired individuals. Another possible cause can be age-related differences in reproductive history, as both first breeders and older breeders occur in the populations. For instance, in *C. fellah* ant workers the expression of both inotocin and its receptor increased with age (Koto et al., 2019). Unfortunately, information on either the reproductive state (i.e. paired or unpaired) or the age of the sampled individuals was not collected before sampling in this study. Therefore, investigating the background of this bias in inotocin gene expression in beetles observed prior to sampling could reveal intriguing biological details at both the behavioural and molecular levels.

Even though plenty of studies focusing on the oxytocin/vasopressin peptide family have been published, few have investigated the time course of expression, especially among invertebrates. Only two studies have been published to date in which receptor levels were measured over a period of time. Levoye et al. (2005) examined vasopressin-related receptor expression in a leech species, *Theromyzon tessulatum* during its reproductive period and found that the expression increased in adult individuals during the reproductive maturation until egg-laying. Chérasse and Aron (2017) measured inotocin receptor expression over the course of the lives of black garden ant queens till colony foundation. Their results showed that expression of the receptor was lower during colony foundation than before mating which suggests a relationship between reproductive state and inotocin. Exploring and comparing the changes in expression of oxytocin/vasopressin related peptides and their receptors in a wide range of species with different types of reproductive strategy during their breeding period could reveal evolutionary processes modulating the functions of these molecules (Hofmann et al., 2014).

Based on our results we provide a new perspective on the potential functions of inotocin, the insect oxytocin-like hormone. However, it is important to note that from expression data one can only infer the activity of the gene, but not the actual levels of hormones in the haemolymph. Therefore, examining inotocin at the peptide level could provide more specific results about its function in the control of parental behaviour, additionally, a repeated experiment with humidity manipulation could lead to a better understanding of the function of inotocin in water maintenance. Further comparative investigations into the regulatory roles oxytocin-like neuropeptides play in invertebrate species are important to understand the evolution of biparental care and its neuroendocrine regulation.

General discussion

One of the major goals of sociobiology is to understand the function and evolution of parental care. Care for the offspring has been investigated from several aspects in vertebrates, especially in mammals and birds (Scheiber et al. 2017). Although this highly complex behaviour is present in many different invertebrate taxa, it is still understudied among these species (Trumbo 2012). Insects are the most diverse class in the animal kingdom presenting a vast diversity of parental behaviour, yet, the interests of behaviour ecologists have mainly been focused on the organisation and function of eusocial insect societies (Smiseth et al. 2014). Many non-eusocial species, however, display various parental behaviour forms similar to those we can observe among vertebrates, e.g. food provision and nest attendance, making them outstanding targets of investigating the origin of care. Moreover, eusociality evolved likely from solitary behaviour through parental care, therefore, exploring the underlying mechanisms of the latter can contribute to a better understanding of social evolution (Rehan & Toth 2015).

Recently, a new species has been introduced which can be a valuable target of research aiming to contribute to the deeper understanding of the evolution and molecular background of parental care. *L. apterus* is a biparental beetle species presenting a well-developed division of labour between the parents. Albeit the optional biparental offspring care observed in burying beetles, no separate roles are assigned to the parents in those species (Suzuki 2011). In contrast, our results show that in *L. apterus* both parents have their determined role in the provision and protection of the offspring. This kind of organisation can only be observed among scarab beetles and the eusocial insects and biparental vertebrate species. Comparisons against the genomes of non-eusocial and eusocial insect species could reveal important changes in the evolution of processes responsible for parental role determination. Therefore, this species can be an exciting model species in social evolution for investigating the cues displayed by the parents.

An intriguing result of ours was the differences in the roles of parents compared to the literature from the previous centuries (Study 1). *L. apterus* was widely distributed and even considered as a pest in Hungary because of the damage it caused in vineyards by cutting the fresh leaves off the plants (Merkl & Vig 2009). However, this species is sensitive to disturbance thus the intense agriculture restricted the distribution of the species into smaller areas (Kovács et al. 2014). According to Rosa et al. (2017), nest attendance increases in high-density populations. Considering the habitat loss of the species since the observations of Emich (1884), it might be a possible explanation that the narrowing distribution led to increased density in remaining populations. This could induce males to invest more in protection of the nests and assurance of their paternity. As already described by Emich (1884) and Schreiner (1906), females can replace males in leaf collection labour when necessary, hence, the increased guarding behaviour of males could lead to increased leaf collecting behaviour in females.

Investigating the genetic background of such complex processes as parental care and the division of labour between parents requires an available reference genome. However, to date, no genome was published in the genus *Lethrus* or the family Geotrupidae, therefore, we sequenced, assembled and annotated the genome of *L. apterus* (Study 2). The sequence and annotation of the genome can serve as a useful basis for transcriptome research. Differential gene expression analyses during the breeding period and between sexes can lead to identification of important gene regulatory pathways in the modulation of behaviour displayed during the period of care. We have already sampled males and females at five time points through the breeding period. Analysing the differences between dates can reveal the genes regulating reproductive behaviour. Further, we can find the genes and hormones responsible for the division of roles between parents by comparing the males and females in each date. Finally, a comprehensive analysis of the differentially expressed genes between dates for each sex followed by the comparison of the resulting genes between the parents to find the shared ones can reveal the genes that are responsible for the formation of parental behaviour exhibited by both parents.

We used the predicted genes in the genome for searching genes that are potentially related to the reproductive and parental behaviour of *L.. apterus* (Study 2). Comparison of the number of found genes in *L. apterus* with other coleopteran species showed a reduction of odorant-binding and pheromone-binding proteins coded in the genome. Considering that *L. apterus* favours legume species during collection of food for the offspring (Study 4) and that the cooperation between the parents suggest a well-developed recognition system, this is an intriguing result. On the other hand, parents and offspring do not interact during the parental period, thus no pheromone-based communication is needed for expression of parental behaviour which could be in relationship with this gene reduction. Therefore, it is worth to continuing the investigation of these proteins. One of the main steps to take is to use the RNA-sequencing reads to improve the continuity of our genome. Reads of transcriptome of different developmental and reproductive stages could provide hints for genome scaffolding (Zhang et al. 2016, Song et al. 2016), furthermore, assembled transcriptomes can be used as evidence for gene prediction as it contains the expressed protein-coding genes (Hoff et al. 2019). Besides, it would also be interesting to investigate the expressed genes directly in the antennae of *L. apterus* with transcriptome analysis. For this purpose, samples could be taken at the moment of meeting of a pair or a random conspecific, during fighting between males and leaf collection by females with special attention on the choice of legume species. This sampling could reveal the main pheromones and hormones responsible for the recognition of the social and non-social environment as well.

The draft genome sequence was used to find candidate reference genes for normalization of RT-qPCR data which is a critical step of relative gene expression analyses (Study 3). Thus, finding the best reference genes made it possible to examine the gene expression changes of inotocin and inotocin receptor during the breeding season of *L. apterus* (Study 4). The gene expression of inotocin showed consistent bimodal distribution which could not be explained with any examined technical sources. This most likely means that the difference in the expression levels originated from biological causes. Considering that the samples were taken from natural populations where several environmental factors may affect the mRNA levels, we cannot give a certain explanation for this phenomenon. However, it would be of interest to investigate some hypotheses with a more robust sampling. With respect to the fact that the grouping of expression levels was detected during the whole active season, a direct relationship between parental care and the bimodal distribution seems unlikely. Possible explanations include age, reproductive history, additionally, food consumption, i.e. whether the sampled individual has fed prior to sampling and the specific activity at the moment of capture, specifically, difference between leaf collection for the parent itself or for the offspring, nest defence and intrusion, nest cleaning and arrangement and mate searching. To test these hypotheses, information on the listed characteristics should be collected before sampling. Moreover, as gene expression does not directly predict the titre of the peptide, it would be also interesting to collect haemolymph samples from the same individuals together with the sampling for gene expression analysis. Thus, we could compare if the bimodal distribution of the inotocin gene expression has an actual effect on peptide levels.

Another interesting result of ours is that the sequence of the mature inotocin of *Lethrus apterus* differs in one amino acid from sequences of other coleopteran species, specifically, valine is found in the fourth position instead of threonine. The mature peptide is a small, nine amino acids long molecule in which the two cysteines in the first and the sixth position create a disulphide bridge making a loop formation (Gruber & Muttenthaler 2012). Thus, the threonine-valine substitution affects the loop of the molecule and may have an effect on the activity of the peptide, further, on the interaction with the receptor. For instance, in case of oxytocin, the 4-valine-oxytocin (i.e. oxytocin having valine at the fourth position) had increased activity compared to other synthesised oxytocin analogs (Sawyer & Manning 1973). Moreover, a study on the specificity of inotocin receptor in *T. castaneum* revealed that several oxytocin orthologs can activate the receptor, albeit only with much higher concentration than inotocin (Stafflinger et al. 2008). These results suggest that even small changes in the sequence and the structure of the peptide can lead to significant effects on the induced processes. Therefore, it would be of interest to examine more deeply the significance of this amino acid substitution in the function of inotocin.

To date, genetic and physiological background of parental behaviour in non-social insects were investigated in only two model species, namely the burying beetle *N. vespilloides* and the small carpenter bee (*C. calcarata*). Therefore, research on other insect species is still needed for a reliable comparative evolutionary work on both the ultimate and proximate levels of parental care. *L. apterus* provide an exceptional model species to investigate the evolution and genetic background of parental care, not only for finding the genes and hormones regulating the behaviour displayed but also for identifying the environmental factors responsible for forming the roles of parents during offspring care.

Summary - Key findings

- Consistent division of labour was found between parents in *Lethrus apterus* which differed from the previously described, i.e. females collect the leaves for the offspring while males are guarding the nest from intruders.
- Draft genome of the *Lethrus apterus* has been assembled into 66,933 contigs in which 20,734 protein-coding genes were structurally and functionally annotated.
- Probable sex chromosome related contigs were found which were present in only male samples suggesting an XY or X0 sex determination system.
- Genes potentially involved in reproductive related behaviour were identified among the annotated genes of *Lethrus apterus* from which we found a reduction of odorant-binding and pheromone-binding proteins compared to other beetle species.
- Housekeeping genes were identified and evaluated as reference genes for validation of data for gene expression analysis using RT-qPCR.
- Gene expression of inotocin and its receptor is higher later in the breeding season when the parental behaviour dominates.

Összefoglalás - Új tudományos eredmények

- Következetes munkamegosztást figyeltünk meg a nagyfejű csajkó szülők között, amely különbözött az eddig leírtaktól: a nőstények gyűjtik a leveleket az utódoknak, míg a hímek őrzik a fészket a betolakodóktól.
- Összeszereltük a nagyfejű csajkó nyers genomját, amely 66.933 kontigból áll, és amelyben 20.734 protein kódoló gént találtunk strukturális és funkcionális annotáció során.
- Lehetséges ivari kromoszómákhoz tartozó kontigokat azonosítottunk, melyek csak a hímekben megtalálhatóak, amely XY vagy X0 ivarmeghatározási rendszerre utal.
- A nagyfejű csajkó annotált génjei között azonosítottunk potenciálisan szaporodási viselkedéshez köthető géneket, melyek közül a szaganyagokhoz és feromonokhoz kötődő fehérjékből kevesebbet találtunk, mint a többi bogárfajban.
- Azonosítottunk és kiértékeltünk háztartási géneket, melyek a valós-idejű kvantitatív PCR-rel történő génexpressziós analízisek során az adatok normalizálására használandók.
- Az inotocin és receptorának génexpressziója magasabb a szaporodási időszak azon szakaszában, amikor az utódgondozó viselkedések dominálnak, azonban ivari különbséget nem találtunk.

Contribution to the studies in the dissertation

Study 1: data collection; manuscript review.

Study 2: genome assembly methodological review and quality check; genome annotation, including reproductive related gene search and phylogenetic analysis; manuscript writing.

Study 3: sample collection and processing; data analysis; manuscript writing.

Study 4: sample collection and processing; data analysis; manuscript writing.

Acknowledgements

I am most grateful to my supervisors, Dr. Zoltán Barta and Dr. Zoltán Németh for supporting me both personally and scientifically, and providing me freedom for my research. I am much obligated to Dr. András Kosztolányi for his constructive comments and guidance in the statistical analyses and manuscript writing. I am thankful to Dr. Edit Juhász for her help in the development of sampling method. I owe a large debt of gratitude to Rita Rácz and Valéria Mester not only for their help and patience through the laboratory works but for their supportive friendship as well. I am grateful to Johanna Lévai-Kiss for her immense amount of help during the field work. I highly appreciate the inspiring, supportive and friendly community of all the members of the MTA-DE Behavioural Ecology Group as well as the Department of Evolutionary Zoology and Humán Biology. Furthermore, I am much obligated to Levente Laczkó for giving me scientific as well as emotional support. Naturally, special thank owed to my family for always supporting and encouraging me in my work and decisions.

The studies presented in the dissertation were supported by the National Research, Development and Innovation Office of Hungary (NKFIH grant no. K112670 and no. K112527). Sample collection was approved by the Northern Hungarian Inspectorate for Environment Protection and Nature Conservation (No. 9007-8/2014) since *Lethrus apterus* is a protected species in Hungary. We acknowledge KIFÜ/NIIF for awarding us access to HPC resource based in Miskolc, Hungary and the Galaxy server Europe.

References

Afgan, E., Baker, D., Batut, B., Van Den Beek, M., Bouvier, D., Čech, M., et al. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Research*, *46*(W1), W537-W544.

Aigaki, T., Fleischmann, I., Chen, P. S., & Kubli, E. (1991). Ectopic expression of sex peptide alters reproductive behavior of female *D. melanogaster*. *Neuron, 7*(4), 557-563.

Aikins, M. J., Schooley, D. A., Begum, K., Detheux, M., Beeman, R. W., & Park, Y. (2008). Vasopressin-like peptide and its receptor function in an indirect diuretic signaling pathway in the red flour beetle. *Insect Biochemistry and Molecular Biology, 38*(7), 740-748.

Andersen, L. H., Kristensen, T. N., Loeschcke, V., Toft, S., & Mayntz, D. (2010). Protein and carbohydrate composition of larval food affects tolerance to thermal stress and desiccation in adult *Drosophila melanogaster*. *Journal of Insect Physiology, 56*(4), 336-340.

Andrews S. (2010) FastQC: A quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Accessed 19 Dec 2020.

Attwood, T. K., Coletta, A., Muirhead, G., Pavlopoulou, A., Philippou, P. B., Popov, I., et al. (2012). The PRINTS database: a fine-grained protein sequence annotation and analysis resource—its status in 2012. *Database*, 2012;2012:bas019.

Balment, R. J., Lu, W., Weybourne, E., & Warne, J. M. (2006). Arginine vasotocin a key hormone in fish physiology and behaviour: a review with insights from mammalian models. *General and Comparative Endocrinology, 147*(1), 9-16.

Balshine, S. (2012). Patterns of parental care in vertebrates. *The Evolution of Parental Care*, Oxford University Press, Oxford, pp. 62-80.

Banerjee, P., Joy, K. P., & Chaube, R. (2017). Structural and functional diversity of nonapeptide hormones from an evolutionary perspective: a review. *General and Comparative Endocrinology, 241*, 4-23.

Beets, I., Temmerman, L., Janssen, T., & Schoofs, L. (2013). Ancient neuromodulation by vasopressin/oxytocin-related peptides. *Worm, 2*(2), e24246.

Benowitz, K. M., McKinney, E. C., Roy-Zokan, E. M., Cunningham, C. B., & Moore, A. J. (2017). The role of lipid metabolism during parental care in two species of burying beetle (*Nicrophorus spp.*). *Animal Behaviour, 129*, 143-149.

Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D., & Pirovano, W. (2011). Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics, 27*(4), 578-579.

Boetzer, M., & Pirovano, W. (2012). Toward almost closed genomes with GapFiller. *Genome Biology, 13*(6), 1-9.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics, 30*(15), 2114-2120.

Brandmayr, P. (1992). Short review of the presocial evolution in Coleoptera. *Ethology Ecology & Evolution*, *4*(sup1), 7-16.

Brůna, T., Lomsadze, A., & Borodovsky, M. (2020). GeneMark-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins. *NAR Genomics and Bioinformatics*, *2*(2), lqaa026.

Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nature Methods, 12*(1), 59-60.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinformatics, 10*(1), 1-9.

Capodeanu-Nägler, A., Keppner, E. M., Vogel, H., Ayasse, M., Eggert, A. K., Sakaluk, S. K., et al. (2016). From facultative to obligatory parental care: interspecific variation in offspring dependency on post-hatching care in burying beetles. *Scientific Reports*, *6*(1), 1-10.

Chérasse, S., & Aron, S. (2017). Measuring inotocin receptor gene expression in chronological order in ant queens. *Hormones and Behavior, 96*, 116-121.

Clutton-Brock, T. H. (1991). *The evolution of parental care*. Princeton University Press, Princeton.

Costa, J. T. (2006). *The other insect societies*. Harvard University Press, Cambridge.

Costa, J. T. (2018). The other insect societies: overview and new directions. *Current Opinion in Insect Science*, *28*, 40-49.

Cunningham, C. B. (2020). Functional genomics of parental care of insects. *Hormones and Behavior, 122*, 104756.

Cunningham, C. B., Ji, L., Wiberg, R. A. W., Shelton, J., McKinney, E. C., Parker, D. J., et al. (2015). The genome and methylome of a beetle with complex social behavior, *Nicrophorus vespilloides* (Coleoptera: Silphidae). *Genome Biology and Evolution*, *7*(12), 3383-3396.

Cunningham, C. B., VanDenHeuvel, K., Khana, D. B., McKinney, E. C., & Moore, A. J. (2016). The role of neuropeptide F in a transition to parental care. *Biology Letters*, *12*(4), 20160158.

De Loof, A., & De Wilde, J. (1970). The relation between haemolymph proteins and vitellogenesis in the Colorado beetle, *Leptinotarsa decemlineata*. *Journal of Insect Physiology*, *16*(1), 157-169.

De Loof, A., & Schoofs, L. (2021). From One Site of Insect Juvenile Hormone Synthesis, No Identified Receptors, and a Denomination as “Status Quo Hormone” in the 1960s to Multiple, Sometimes Conflicting, Possibilities to Date. *Life: The Excitement of Biology*, *8*(4), 196-238.

Donaldson, Z. R., & Young, L. J. (2008). Oxytocin, vasopressin, and the neurogenetics of sociality. *Science, 322*(5903), 900-904.

Dottorini, T., Nicolaides, L., Ranson, H., Rogers, D. W., Crisanti, A., & Catteruccia, F. (2007). A genome-wide analysis in *Anopheles gambiae* mosquitoes reveals 46 male accessory gland genes, possible modulators of female behavior. *Proceedings of the National Academy of Sciences, 104*(41), 16215-16220.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research, 32*(5), 1792-1797.

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., et al. (2019). The Pfam protein families database in 2019. *Nucleic Acids Research, 47*(D1), D427-D432.

Emich, G. (1884). A csajkó fejlődése és átalakulása. *Rovartani Lapok, 1*, 30-33.

Emms, D. M., & Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology, 20*(1), 1-14.

Fan, J., Francis, F., Liu, Y., Chen, J. L., & Cheng, D. F. (2011). An overview of odorant-binding protein functions in insect peripheral olfactory reception. *Genetics and Molecular Research, 10*(4), 3056-3069.

Fujino, Y., Nagahama, T., Oumi, T., Ukena, K., Morishita, F., Furukawa, Y., et al. (1999). Possible functions of oxytocin/vasopressin-superfamily peptides in annelids with special reference to reproduction and osmoregulation. *Journal of Experimental Zoology, 284*(4), 401-406.

Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics, 28*(23), 3150-3152.

Gilbert, J. D., & Manica, A. (2015). The evolution of parental care in insects: a test of current hypotheses. *Evolution*, *69*(5), 1255-1270.

Gilbert, M. T. P., Moore, W., Melchior, L., & Worobey, M. (2007). DNA extraction from dry museum beetles without conferring external morphological damage. *PLoS One, 2*(3), e272.

Gruber, C. W., & Muttenthaler, M. (2012). Discovery of defense-and neuropeptides in social ants by genome-mining. *PLoS One, 7*(3), e32559.

Haft, D. H., Selengut, J. D., Richter, R. A., Harkins, D., Basu, M. K., & Beck, E. (2012). TIGRFAMs and genome properties in 2013. *Nucleic Acids Research, 41*(D1), D387-D395.

Hanin, O., Azrielli, A., Zakin, V., Applebaum, S., & Rafaeli, A. (2011). Identification and differential expression of a sex-peptide receptor in *Helicoverpa armigera*. *Insect Biochemistry and Molecular Biology, 41*(8), 537-544.

Hanoune, J. (2010). Comparative and evolutionary aspects of vasopressin. *Perspectives On Vasopressin*, Imperial College Press, London, pp. 21-38.

Heisswolf, A., Obermaier, E., & Poethke, H. J. (2005). Selection of large host plants for oviposition by a monophagous leaf beetle: nutritional quality or enemy-free space?. *Ecological Entomology, 30*(3), 299-306.

Heymons, R. (1915). Die Vielfüßler, Insekten und Spinnenkerfe. *Brehms Tierleben*, Bibliographisches Institut, Gotha.

Hoff, K. J., Lange, S., Lomsadze, A., Borodovsky, M., & Stanke, M. (2016). BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics, 32*(5), 767-769.

Hoff, K. J., Lomsadze, A., Borodovsky, M., & Stanke, M. (2019). Whole-genome annotation with BRAKER. *Gene Prediction*. Humana, New York, pp. 65-95 .

Hofmann, H. A., Beery, A. K., Blumstein, D. T., Couzin, I. D., Earley, R. L., Hayes, L. D., et al. (2014). An evolutionary framework for studying mechanisms of social behavior. *Trends in Ecology and Evolution*, *29*(10), 581-589.

Indrischek, H., Wieseke, N., Stadler, P. F., & Prohaska, S. J. (2016). The paralog-to-contig assignment problem: high quality gene models from fragmented assemblies. *Algorithms for Molecular Biology, 11*(1), 1-14.

Ito, N., Katoh, T., Kobayashi, N., & Katakura, H. (2010). Effects of straits as dispersal barriers for the flightless roving carrion beetle, *Silpha perforata* (Coleoptera, Silphidae, Silphinae). *Zoological Science, 27*(4), 313-319.

Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics, 30*(9), 1236-1240.

Junier, T., & Zdobnov, E. M. (2010). The Newick utilities: high-throughput phylogenetic tree processing in the UNIX shell. *Bioinformatics, 26*(13), 1669-1670.

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K., Von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods, 14*(6), 587-589.

Kaneko, Y., & Hiruma, K. (2014). Short neuropeptide F (sNPF) is a stage-specific suppressor for juvenile hormone biosynthesis by corpora allata, and a critical factor for the initiation of insect metamorphosis. *Developmental Biology*, *393*(2), 312-319.

Kim, I. H., Pham, V., Jablonka, W., Goodman, W. G., Ribeiro, J. M., & Andersen, J. F. (2017). A mosquito hemolymph odorant-binding protein family member specifically binds juvenile hormone. *Journal of Biological Chemistry, 292*(37), 15329-15339.

Kiss, J., Rádai, Z., Rosa, M. E., Kosztolányi, A., & Barta, Z. (2020). Seasonal changes in immune response and reproductive investment in a biparental beetle. *Journal of Insect Physiology, 121*, 104000.

Kiss, J., Németh, Z., Kosztolányi, A., & Barta, Z. (2020). Differential movement and activity patterns of sexes in a biparental beetle during the reproductive season. *Ecological Entomology*, *45*(6), 1504-1508.

Klatt, J. D., & Goodson, J. L. (2013). Oxytocin-like receptors mediate pair bonding in a socially monogamous songbird. *Proceedings of the Royal Society B: Biological Sciences, 280*(1750), 20122396.

Klug, H., & Bonsall, M. B. (2010). Life history and the evolution of parental care. *Evolution: International Journal of Organic Evolution*, *64*(3), 823-835.

Klug, H., Bonsall, M. B., & Alonzo, S. H. (2013). Sex differences in life history drive evolutionary transitions among maternal, paternal, and bi-parental care. *Ecology and Evolution*, *3*(4), 792-806.

Koto, A., Motoyama, N., Tahara, H., McGregor, S., Moriyama, M., Okabe, T., et al. (2019). Oxytocin/vasopressin-like peptide inotocin regulates cuticular hydrocarbon synthesis and water balancing in ants. *Proceedings of the National Academy of Sciences, 116*(12), 5597-5606.

Kovács, T., Merkl, O., & Rácz, R. (2014). Distribution of *Lethrus apterus* (Laxmann, 1770) in Hungary (Coleoptera: Geotrupidae). *Folia Historico-Naturalia Musei Matraensis, 38*, 67-73.

Kozera, B., & Rapacz, M. (2013). Reference genes in real-time PCR. *Journal of Applied Genetics, 54*(4), 391-406.

Král, D., Hillert, O., Drožová, D., & Šípek, P. (2013). *Lethrus* (*Lethrus*) *schneideri sp. n.*(Coleoptera, Geotrupidae) from Greece. *ZooKeys*, *339*, 93-106.

Kronauer, D. J., & Libbrecht, R. (2018). Back to the roots: the importance of using simple insect societies to understand the molecular basis of complex social life. *Current Opinion in Insect Science*, *28*, 33-39.

Lenth, R. (2020). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.4.8. https://CRAN.R-project.org/package=emmeans.

Letunic, I., & Bork, P. (2018). 20 years of the SMART protein domain annotation resource. *Nucleic Acids Research, 46*(D1), D493-D496.

Levoye, A., Mouillac, B., Riviere, G., Vieau, D., Salzet, M., & Breton, C. (2005). Cloning, expression and pharmacological characterization of a vasopressin-related receptor in an annelid, the leech *Theromyzon tessulatum*. *Journal of Endocrinology, 184*(1), 277-289.

Li, D., Liu, C. M., Luo, R., Sadakane, K., & Lam, T. W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics, 31*(10), 1674-1676.

Liutkevičiūte, Z., Gil-Mansilla, E., Eder, T., Casillas-Pérez, B., Di Giglio, M. G., Muratspahić, E., et al. (2018). Oxytocin-like signaling in ants influences metabolic gene expression and locomotor activity. *The FASEB Journal, 32*(12), 6808-6821.

Liutkeviciute, Z., Koehbach, J., Eder, T., Gil-Mansilla, E., & Gruber, C. W. (2016). Global map of oxytocin/vasopressin-like neuropeptide signalling in insects. *Scientific Reports, 6*(1), 1-9.

Liu, Y., Schröder, J., & Schmidt, B. (2013). Musket: a multistage k-mer spectrum-based error corrector for Illumina sequence data. *Bioinformatics, 29*(3), 308-315.

Lomsadze, A., Ter-Hovhannisyan, V., Chernoff, Y. O., & Borodovsky, M. (2005). Gene identification in novel eukaryotic genomes by self-training algorithm. *Nucleic Acids Research, 33*(20), 6494-6506.

Lowrey, E. M., & Tomaszycki, M. L. (2014). The formation and maintenance of social relationships increases nonapeptide mRNA in zebra finches of both sexes. *Behavioral Neuroscience, 128*(1), 61.

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., et al. (2012). SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience, 1*(1), 2047-217X.

Marçais, G., & Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics, 27*(6), 764-770.

McCormick, S. D., & Bradshaw, D. (2006). Hormonal control of salt and water balance in vertebrates. *General and Comparative Endocrinology, 147*(1), 3-8.

McGraw, L., Székely, T., & Young, L. J. (2010). Pair bonds and parental behaviour. *Social Behaviour: Genes, Ecology and Evolution.* Cambridge University Press, Cambridge, pp. 271-301.

McKenna, D. D., & Farrell, B. D. (2009). Beetles (Coleoptera). *The Timetree of Life.* Oxford University Press, Oxford, 278-289.

Mehlferber, E. C., Benowitz, K. M., Roy-Zokan, E. M., McKinney, E. C., Cunningham, C. B., & Moore, A. J. (2017). Duplication and Sub/Neofunctionalization of Malvolio, an Insect Homolog of Nramp, in the subsocial beetle *Nicrophorus vespilloides*. *G3: Genes, Genomes, Genetics, 7*(10), 3393-3403.

Merkl, O., & Vig, K. (2009). *Bogarak: a Pannon regióban*. Vas Megyei Múzeumok Igazgatósága, BKL Kiadó, Magyar Természettudományi Múzeum.

Mi, H., Muruganujan, A., Ebert, D., Huang, X., & Thomas, P. D. (2019). PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Research, 47*(D1), D419-D426.

Moore, F. L., Wood, R. E., & Boyd, S. K. (1992). Sex steroids and vasotocin interact in a female amphibian (*Taricha granulosa*) to elicit female-like egg-laying behavior or male-like courtship. *Hormones and Behavior, 26*(2), 156-166.

Muratspahić, E., Monjon, E., Duerrauer, L., Rogers, S. M., Cullen, D. A., Broeck, J. V., & Gruber, C. W. (2020). Oxytocin/vasopressin-like neuropeptide signaling in insects. *Vitamins and Hormones, 113*, 29-53.

Nadalin, F., Vezzi, F., & Policriti, A. (2012). GapFiller: a de novo assembly approach to fill the gap within paired reads. *BMC Bioinformatics, 13*(14), 1-16.

Nagy, N. A., Németh, Z., Juhász, E., Póliska, S., Rácz, R., Kosztolányi, A., & Barta, Z. (2017). Evaluation of potential reference genes for real-time qPCR analysis in a biparental beetle, *Lethrus apterus* (Coleoptera: Geotrupidae). *PeerJ*, *5*, e4047.

Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution, 32*(1), 268-274.

O'Connor, C. M., Marsh-Rollo, S. E., Aubin-Horth, N., & Balshine, S. (2016). Species-specific patterns of nonapeptide brain gene expression relative to pair-bonding behavior in grouping and non-grouping cichlids. *Hormones and Behavior, 80*, 30-38.

Ohmart, C. P., Stewart, L. G., & Thomas, J. R. (1985). Effects of food quality, particularly nitrogen concentrations, of Eucalyptus blakelyi foliage on the growth of *Paropsis atomaria* larvae (Coleoptera: Chrysomelidae). *Oecologia, 65*(4), 543-549.

Oumi, T., Ukena, K., Matsushima, O., Ikeda, T., Fujita, T., Minakata, H., et al. (1996). Annetocin, an annelid oxytocin-related peptide, induces egg-laying behavior in the earthworm, *Eisenia foetida*. *Journal of Experimental Zoology, 276*(2), 151-156.

Pabinger, S., Rödiger, S., Kriegner, A., Vierlinger, K., & Weinhäusel, A. (2014). A survey of tools for the analysis of quantitative PCR (qPCR) data. *Biomolecular Detection and Quantification, 1*(1), 23-33.

Pallarés, S., Velasco, J., Millán, A., Bilton, D. T., & Arribas, P. (2016). Aquatic insects dealing with dehydration: do desiccation resistance traits differ in species with contrasting habitat preferences?. *PeerJ, 4*, e2382.

Panaitof, S. C., Yaeger, J. D., Speer, J. P., & Renner, K. J. (2016). Biparental behavior in the burying beetle *Nicrophorus orbicollis*: a role for dopamine?. *Current Zoology*, *62*(3), 285-291.

Pandurangan, A. P., Stahlhacke, J., Oates, M. E., Smithers, B., & Gough, J. (2019). The SUPERFAMILY 2.0 database: a significant proteome update and a new webserver. *Nucleic Acids Research, 47*(D1), D490-D494.

Paradis, E., Claude, J., & Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics, 20*(2), 289-290.

Park, Y. C., Yoo, J. et al. (2012). Identification of differentially expressed genes from adult cockroach females exhibiting maternal care behavior. *Journal of Asia-Pacific Entomology, 15*(3), 343-348.

Parker, D. J., Cunningham, C. B., Walling, C. A., Stamper, C. E., Head, M. L., Roy-Zokan, E. M., et al. (2015). Transcriptomes of parents identify parenting strategies and sexual conflict in a subsocial beetle. *Nature Communications*, *6*(1), 1-12.

Pike, D. A., Clark, R. W., Manica, A., Tseng, H. Y., Hsu, J. Y., & Huang, W. S. (2016). Surf and turf: predation by egg-eating snakes has led to the evolution of parental care in a terrestrial lizard. *Scientific Reports*, *6*(1), 1-7.

Prates, M. O., Lachos, V. H., & Cabral, C. (2013). mixsmsn: Fitting finite mixture of scale mixture of skew-normal distributions. *Journal of Statistical Software, 54*(12), 1-20.

Proux, J. P., Miller, C. A., Li, J. P., Carney, R. L., Girardie, A., Delaage, M., et al. (1987). Identification of an arginine vasopressin-like diuretic hormone from *Locusta migratoria*. *Biochemical and Biophysical Research Communications, 149*(1), 180-186.

Radonić, A., Thulke, S., Mackay, I. M., Landt, O., Siegert, W., & Nitsche, A. (2004). Guideline to reference gene selection for quantitative real-time PCR. *Biochemical and Biophysical Research Communications, 313*(4), 856-862.

Ram, K. R., & Wolfner, M. F. (2007). Sustained post-mating response in *Drosophila melanogaster* requires multiple seminal fluid proteins. *PLoS Genetics, 3*(12), e238.

Ratz, T., & Smiseth, P. T. (2018). Flexible parents: joint effects of handicapping and brood size manipulation on female parental care in *Nicrophorus vespilloides*. *Journal of Evolutionary Biology, 31*(5), 646-656.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rehan, S. M., & Richards, M. H. (2013). Reproductive aggression and nestmate recognition in a subsocial bee. *Animal Behaviour*, *85*(4), 733-741.

Rehan, S. M., Berens, A. J., & Toth, A. L. (2014). At the brink of eusociality: transcriptomic correlates of worker behaviour in a small carpenter bee. *BMC Evolutionary Biology, 14*(1), 1-11.

Rehan, S. M., & Toth, A. L. (2015). Climbing the social ladder: the molecular evolution of sociality. *Trends in Ecology and Evolution, 30*(7), 426-433.

Robinson, G. E., Grozinger, C. M., & Whitfield, C. W. (2005). Sociogenomics: social life in molecular terms. *Nature Reviews Genetics, 6*(4), 257-270.

Rosa, M. E., Barta, Z., Fülöp, A., Székely, T., & Kosztolányi, A. (2017). The effects of adult sex ratio and density on parental care in *Lethrus apterus* (Coleoptera, Geotrupidae). *Animal Behaviour, 132*, 181-188.

Rosa, M. E., Barta, Z., & Kosztolányi, A. (2018). Willingness to initiate a fight but not contest behaviour depends on intruder size in *Lethrus apterus* (Geotrupidae). *Behavioural Processes*, *149*, 65-71.

Rosa, M. E., Kiss, J., Barta, Z., & Kosztolányi, A. (2019). Size-dependent investment in tusk length, testis size and sperm length in a biparental geotrupid beetle. *Journal of Zoology*, *309*(2), 106-113.

Royle, N. J., Alonzo, S. H., & Moore, A. J. (2016). Co-evolution, conflict and complexity: what have we learned about the evolution of parental care behaviours?. *Current Opinion in Behavioral Sciences*, *12*, 30-36.

Roy-Zokan, E. M., Cunningham, C. B., Hebb, L. E., McKinney, E. C., & Moore, A. J. (2015). Vitellogenin and vitellogenin receptor gene expression is associated with male and female parenting in a subsocial insect. *Proceedings of the Royal Society B: Biological Sciences, 282*(1809), 20150787.

Salzet, M., Bulet, P., Van Dorsselaer, A., & Malecha, J. (1993). Isolation, structural characterization and biological function of a lysine-conopressin in the central nervous system of the Pharyngobdellid leech *Erpobdella octoculata*. *European Journal of Biochemistry, 217*(3), 897-903.

Sawyer, W. H., & Manning, M. (1973). Synthetic analogs of oxytocin and the vasopressins. *Annual Review of Pharmacology*, *13*(1), 5-17.

Scheiber, I. B., Weiß, B. M., Kingma, S. A., & Komdeur, J. (2017). The importance of the altricial–precocial spectrum for social complexity in mammals and birds–a review. *Frontiers in Zoology, 14*(1), 1-20.

Schreiner, J. (1906). Die Lebensweise und Metamorphose des Rebenschneiders oder grossköpfigen Zwiebelhornkäfers (*Lethrus apterus* Laxm.). *Horae Societatis Entomologicae Rossicae, 37*, 197-208.

Scott, M. P. (1998). The ecology and behavior of burying beetles. *Annual review of entomology*, *43*(1), 595-618.

Senthilkumar, R., & Srinivasan, R. (2019). Sex-specific spatial and temporal gene expressions of Pheromone biosynthesis activating neuropeptide (PBAN) and binding proteins (PBP/OBP) in *Spoladea recurvalis*. *Scientific Reports, 9*(1), 1-15.

Servant, F., Bru, C., Carrere, S., Courcelle, E., Gouzy, J., Peyruc, D., et al. (2002). ProDom: automated clustering of homologous domains. *Briefings in Bioinformatics, 3*(3), 246-251.

Shell, W. A., & Rehan, S. M. (2019). Social modularity: conserved genes and regulatory elements underlie caste-antecedent behavioural states in an incipiently social bee. *Proceedings of the Royal Society B*, *286*(1916), 20191815.

Sigrist, C. J., De Castro, E., Cerutti, L., Cuche, B. A., Hulo, N., Bridge, A., et al. (2012). New and continuing developments at PROSITE. *Nucleic Acids Research, 41*(D1), D344-D347.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics, 31*(19), 3210-3212.

Simpson, J. T. (2014). Exploring genome characteristics and sequence quality without a reference. *Bioinformatics, 30*(9), 1228-1235.

Smiseth, P. T., Kölliker, M., & Royle, N. J. (2012). What is parental care?. *The Evolution of Parental Care*. Oxford University Press, Oxford, pp. 1-17.

Smiseth, P. T., Kölliker, M., & Royle, N. J. (2014). Parental care. *The Evolution of Insect Mating Systems*. Oxford University Press, Oxford, pp. 221-241.

Smit AFA, Hubley R, Green P. RepeatMasker Open-4.0. 2013 – 2015.

Smit AFA, Hubley R. RepeatModeler Open-1.0. 2008 – 2015.

Sol, D., Lapiedra, O., & González-Lagos, C. (2013). Behavioural adjustments for a life in the city. *Animal Behaviour, 85*(5), 1101-1112.

Song, L., Shankar, D. S., & Florea, L. (2016). Rascaf: improving genome assembly with RNA sequencing data. *The Plant Genome*, *9*(3), plantgenome2016-03.

Stafflinger, E., Hansen, K. K., Hauser, F., Schneider, M., Cazzamali, G., Williamson, M., et al. (2008). Cloning and identification of an oxytocin/vasopressin-like receptor and its ligand from insects. *Proceedings of the National Academy of Sciences, 105*(9), 3262-3267.

Stanke, M., Diekhans, M., Baertsch, R., & Haussler, D. (2008). Using native and syntenically mapped cDNA alignments to improve *de novo* gene finding. *Bioinformatics, 24*(5), 637-644.

Stanke, M., Schöffmann, O., Morgenstern, B., & Waack, S. (2006). Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics, 7*(1), 1-11.

Stoop, R. (2012). Neuromodulation by oxytocin and vasopressin. *Neuron, 76*(1), 142-159.

Stothard, P. (2000). The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques, 28*(6), 1102-1104.

Suzuki, S. (2011). Biparental burying beetles, *Nicrophorus quadripunctatus* (Coleoptera, Silphidae), do not show a division of labor in their carrion-burial behavior. *Elytra NS, 1*, 41-45.

Tan, Q. Q., Zhu, L., Li, Y., Liu, W., Ma, W. H., Lei, C. L., et al. (2015). A *de novo* transcriptome and valid reference genes for quantitative real-time PCR in *Colaphellus bowringi*. *PLoS One, 10*(2), e0118693.

Thellin, O., Zorzi, W., Lakaye, B., De Borman, B., Coumans, B., Hennen, G., et al. (1999). Housekeeping genes as internal standards: use and limits. *Journal of Biotechnology, 75*(2-3), 291-295.

Trivers, R. (1972). Parental investment and sexual selection. *Sexual Selection & the Descent of Man. Aldine Publishing Company,* Chicago, pp. 136-179.

Trumbo, S. T. (2002). Hormonal regulation of parental care in insects. *Hormones, Brain and Behavior*. Academic Press, Cambridge, pp. 115-139.

Trumbo, S. T. (2012). Patterns of parental care in invertebrates. *The Evolution of Parental Care*. Oxford University Press, Oxford, pp. 81-100.

Trumbo, S. T. (2018). Juvenile hormone and parental care in subsocial insects: implications for the role of juvenile hormone in the evolution of sociality. *Current Opinion in Insect Science, 28*, 13-18.

Ukena, K., Iwakoshi-Ukena, E., & Hikosaka, A. (2008). Unique form and osmoregulatory function of a neurohypophysial hormone in a urochordate. *Endocrinology, 149*(10), 5254-5261.

UniProt Consortium. (2019). UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research*, *47*(D1), D506-D515.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., et al. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology, 3*(7), 1-12.

VanGuilder, H. D., Vrana, K. E., & Freeman, W. M. (2008). Twenty-five years of quantitative PCR for gene expression analysis. *Biotechniques, 44*(5), 619-626.

Van Kesteren, R. E., Tensen, C. P., Smit, A. B., Van Minnen, J., Van Soest, P. F., Kits, K. S., et al. (1995). A novel G protein-coupled receptor mediating both vasopressin-and oxytocin-like functions of Lys-conopressin in *Lymnaea stagnalis*. *Neuron, 15*(4), 897-908.

Vicedomini, R., Vezzi, F., Scalabrin, S., Arvestad, L., & Policriti, A. (2013). GAM-NGS: genomic assemblies merger for next generation sequencing. *BMC Bioinformatics, 14*(7), 1-18.

von Lengerken H. (1939) *Die Brutfürsorge- und Brutpflege- instinkte der Käfer*. Akademische Verlagsgesellschaft MBH, Leipzig.

Vurture, G. W., Sedlazeck, F. J., Nattestad, M., Underwood, C. J., Fang, H., Gurtowski, J., et al. (2017). GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics, 33*(14), 2202-2204.

Wagenaar, D. A., Hamilton, M. S., Huang, T., Kristan, W. B., & French, K. A. (2010). A hormone-activated central pattern generator for courtship. *Current Biology, 20*(6), 487-495.

Wang, Y., Wang, Z. K., Huang, Y., Liao, Y. F., & Yin, Y. P. (2014). Identification of suitable reference genes for gene expression studies by qRT-PCR in the blister beetle *Mylabris cichorii*. *Journal of Insect Science, 14*(1), 94.

Wang, Z. X., Liu, Y., Young, L. J., & Insel, T. R. (2000). Hypothalamic vasopressin gene expression increases in both males and females postpartum in a biparental rodent. *Journal of Neuroendocrinology, 12*(2), 111-120.

Wigby, S., Slack, C., Grönke, S., Martinez, P., Calboli, F. C., Chapman, T., et al. (2011). Insulin signalling regulates remating in female *Drosophila*. *Proceedings of the Royal Society B: Biological Sciences, 278*(1704), 424-431.

Wijesekera, T. P., Saurabh, S., & Dauwalder, B. (2016). Juvenile hormone is required in adult males for *Drosophila* courtship. *PLoS One, 11*(3), e0151912.

Wilson, E. O. (1975). *Sociobiology: The new synthesis*. Harvard University Press.

Wong, J. W., Meunier, J., & Kölliker, M. (2013). The evolution of parental care in insects: the roles of ecology, life history and the social environment. *Ecological Entomology, 38*(2), 123-137.

Woodard, S. H., Bloch, G. M., Band, M. R., & Robinson, G. E. (2014). Molecular heterochrony and the evolution of sociality in bumblebees (*Bombus terrestris*). *Proceedings of the Royal Society B: Biological Sciences, 281*(1780), 20132419.

Wu, Z. Z., Qu, M. Q., Pu, X. H., Cui, Y., Xiao, W. Y., Zhao, H. X., et al. (2017). Transcriptome sequencing of *Tessaratoma papillosa* antennae to identify and analyze expression patterns of putative olfaction genes. *Scientific Reports, 7*(1), 1-11.

Xu, J., Gao, B., Shi, M. R., Yu, H., Huang, L. Y., Chen, P., et al. (2019). Copulation exerts significant effects on mRNA expression of cryptochrome genes in a moth. *Journal of Insect Science, 19*(2), 3.

Yang, X., Pan, H., Yuan, L., & Zhou, X. (2018). Reference gene selection for RT-qPCR analysis in *Harmonia axyridis*, a global invasive lady beetle. *Scientific Reports, 8*(1), 1-10.

Yapici, N., Kim, Y. J., Ribeiro, C., & Dickson, B. J. (2008). A receptor that mediates the post-mating switch in *Drosophila* reproductive behaviour. *Nature, 451*(7174), 33-37.

Zhang, S. V., Zhuo, L., & Hahn, M. W. (2016). AGOUTI: improving genome assembly and annotation using transcriptome data. *GigaScience*, *5*(1), s13742-016.

Zhang, C., Rabiee, M., Sayyari, E., & Mirarab, S. (2018). ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics, 19*(6), 15-30.

Zhang, S. Q., Che, L. H., Li, Y., Liang, D., Pang, H., Ślipiński, A., et al. (2018). Evolutionary history of Coleoptera revealed by extensive sampling of genes and species. *Nature Communications, 9*(1), 1-11.

Zhao, H. X., Xiao, W. Y., Ji, C. H., Ren, Q., Xia, X. S., Zhang, X. F., et al. (2019). Candidate chemosensory genes identified from the greater wax moth, *Galleria mellonella*, through a transcriptomic analysis. *Scientific Reports, 9*(1), 1-12.

List of publications

Kosztolányi, A., Nagy, N., Kovács, T., & Barta, Z. (2015). Predominant female care in the beetle *Lethrus apterus* with supposedly biparental care. *Entomological Science*, *18*(2), 292-294.

Nagy, N. A., Németh, Z., Juhász, E., Póliska, S., Rácz, R., Kosztolányi, A., & Barta, Z. (2017). Evaluation of potential reference genes for real-time qPCR analysis in a biparental beetle, *Lethrus apterus* (Coleoptera: Geotrupidae). *PeerJ*, *5*, e4047.

Appendix

Availability of beetle proteomes referenced in Study 2

Table S1. Scientific names of the species and the accession number of their proteomes used for phylogenetic analysis of *L. apterus*.

Species	**Accession number**
Aethina tumida	GCF_001937115.1
Agrilus planipennis	GCF_000699045.2
Anoplophora glabripennis	GCF_000390285.2
Asbolus verrucosus	GCA_004193795.1
Callosobruchus maculatus	GCA_900659725.1
Dendroctonus ponderosae	GCF_000355655.1
Diabrotica virgifera	GCF_003013835.1
Ignelater luminosus	GCA_011009095.1
Leptinotarsa decemlineata	GCF_000500325.1
Nicrophorus vespilloides	GCF_001412225.1
Onthophagus taurus	GCF_000648695.1
Photinus piralis	GCF_008802855.1
Sitophilus oryzae	GCF_002938485.1
Tribolium castaneum	GCF_000002335.3
Drosophila melanogaster	GCF_000001215.4

Supplementary studies on the bimodal distribution of inotocin expression levels

Gene expression levels of inotocin have bimodal distribution across our samples both from 2016 and 2018 (Fig. S1). To find out the origin of this clustering, several possible bias effects were investigated whether they correlate with the distribution of the two groups of inotocin expression: sex (Fig. S2), sampling time during the day (minutes since sunrise; Fig. S3), body size (pronotum width, only for 2016; Fig. S4), location on RT-qPCR plate (Fig. S5), concentration of the RNA samples (Fig. S6) and level of protein contamination of the samples (Fig. S7). At each sampling time, all samples were taken by one person, therefore, we did not investigate the effect of the dissecting person.

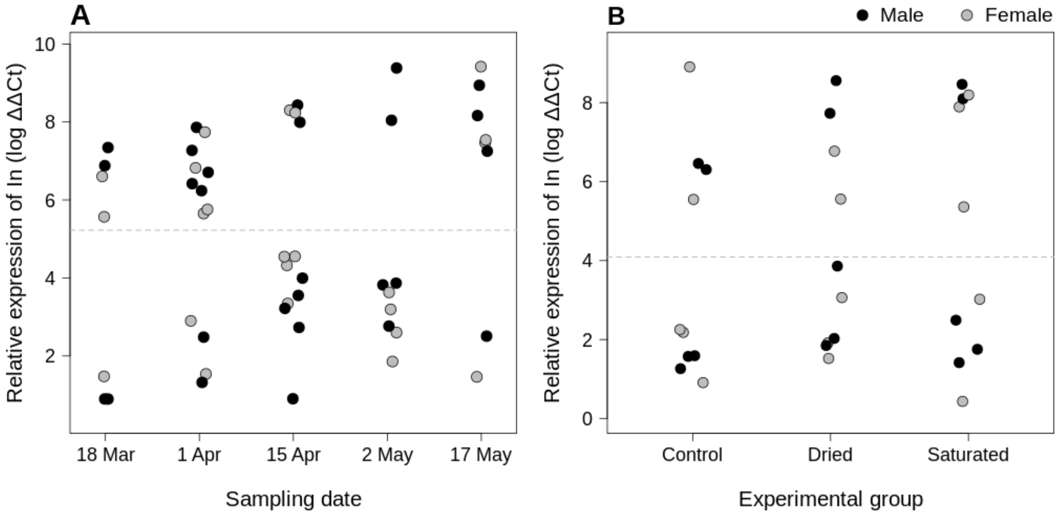


Figure S1. Expression groups of inotocin hormone during the active season in 2016 (A) and in the drying experiment in 2018 (B). Points shaded in black show males, whereas grey ones represent females. The dashed grey lines represent the minimum of the estimated density function in two-group mixture models that were used as the cutoff value for the two-level factor added to the linear models (see Material and methods).

Differences between expression groups in continuous variables (i.e. sampling time, size, concentration and protein contamination) was investigated using Wilcoxon rank sum test with continuity correction, whereas differences between groups in categorical variables (i.e. sex and location during RT-qPCR) was analysed with Fisher's Exact Test. Results of the different effects are shown separately for both years.

Effect of sex

As sexual dimorphism often can be measured in hormonal and thus gene expression levels, we tested the effect of sex, however, no difference between expression groups of sexes was found (Fig. S1; Fisher's Exact Tests, 2016: p=0.776; 2018: p=0.462).

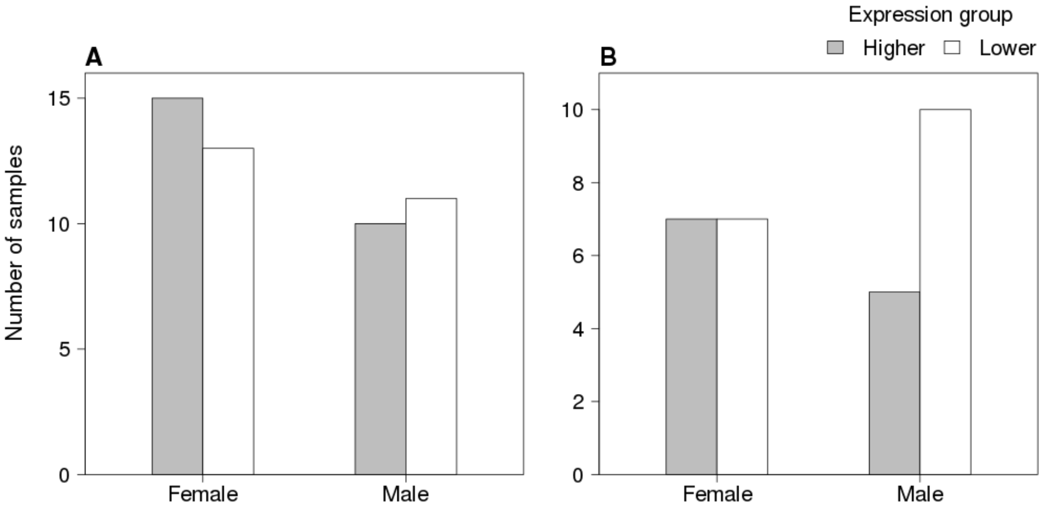


Figure S2. Samples belonging to the different expression groups of inotocin hormone in the two sexes. (A) Samples from 2016 and (B) samples from 2018.

Effect of sampling time

Hormonal and gene expression usually have a daily cycle, hence the effect of the sampling time (i.e. the time since sunrise expressed in minutes) was investigated. No difference was found between groups in sampling time (Fig. S2; Wilcoxon rank sum tests, 2016: W=350.5, p=0.317; 2018: W=84, p=0.438).

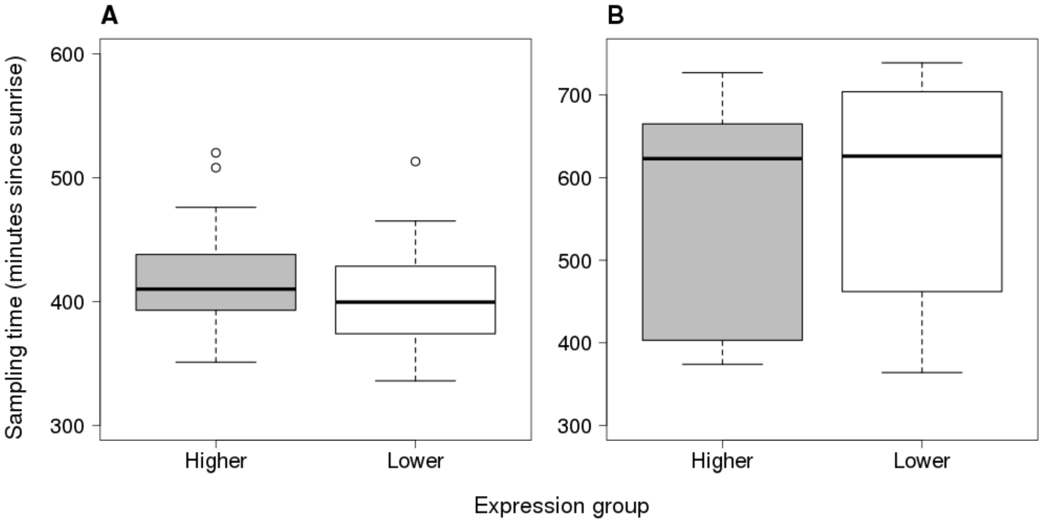


Figure S3. Sampling time during the day in the two expression groups of inotocin hormone. (A) Samples from 2016 and (B) samples from 2018.

Effect of body size of individuals

Difference between size can effect the life-history strategy of the individuals and therefore, also the hormonal levels. Individual sizes were measured as the pronotum width only in 2016 in advance of sampling. No significant difference was detected in size between groups (Fig. S3; Wilcoxon rank sum test, 2016: W=274, p=0.610).

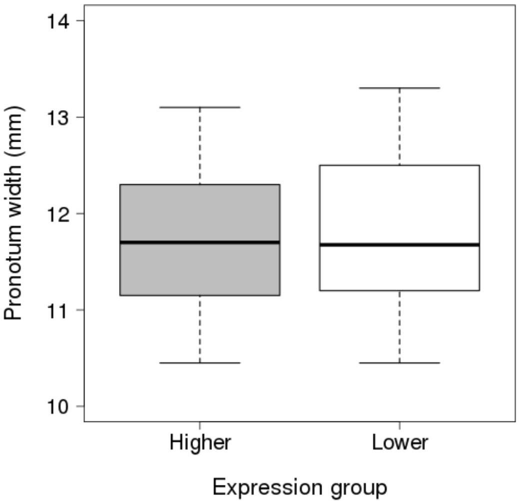


Figure S4. Individual size in the two expression groups of inotocin hormone across samples from 2016.

Effect of location on RT-qPCR plates

Technical errors during laboratory methods, e.g. different handling of samples, can cause bias in data. For example samples measured in different reactions, hence on different plates might cause variance in the results. According to the results no difference between the expression groups in RT-qPCR plate IDs (i.e. the order samples were measured) was found (Fig. S4; Fisher's Exact Tests, 2016: p=0.395; 2018: p=1.000).

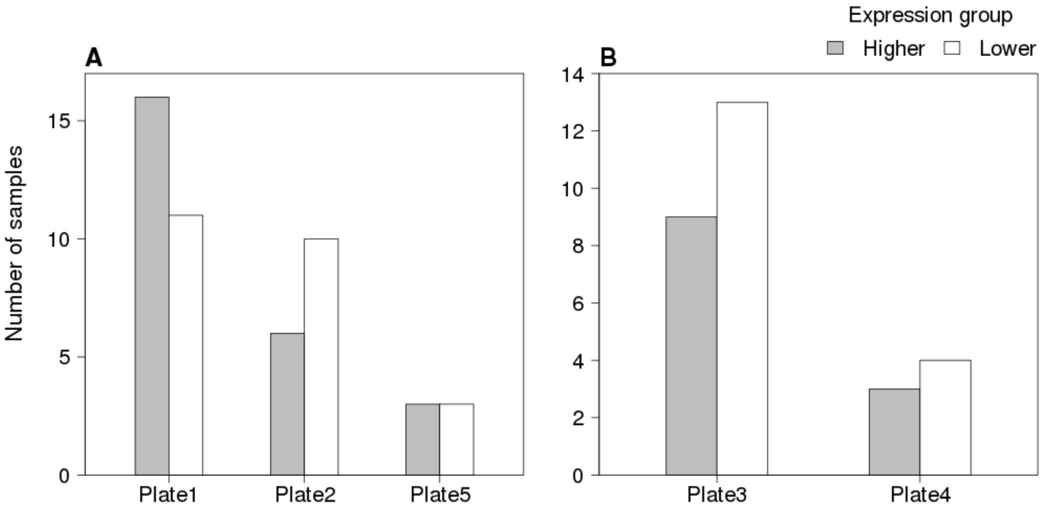


Figure S5. Location of samples during RT-qPCR in the two expression groups of inotocin hormone. (A) Samples from 2016 and (B) samples from 2018.

Effect of RNA concentration

Despite samples were diluted to the same concentration for reverse transcription, i.e., concentration differences should not cause bias in the expression levels, however, we investigated whether the starting RNA concentration differs in the groups. Results showed no significant difference (Fig. S5; Wilcoxon rank sum tests, 2016: W=350, p=0.322; 2018: W=74, p=0.223).

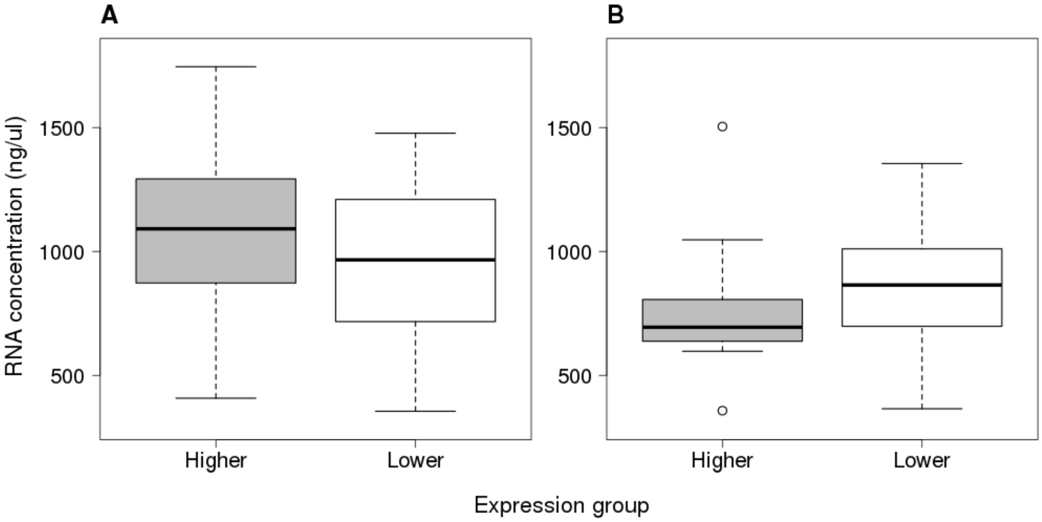


Figure S6. Starting RNA concentration in the two expression groups of inotocin hormone. (A) Samples from 2016 and (B) samples from 2018.

Effect of protein contamination

Contamination of samples by proteins may worsen the success of reverse transcription thus the quality of cDNA. Protein contamination was measured by NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). All samples had high 260/280 nm value which means they had high purity, and there was no significant difference in the contamination between the two groups (Fig. S6; Wilcoxon rank sum tests, 2016: W=217, p=0.097; 2018: W=89.5, p=0.591).

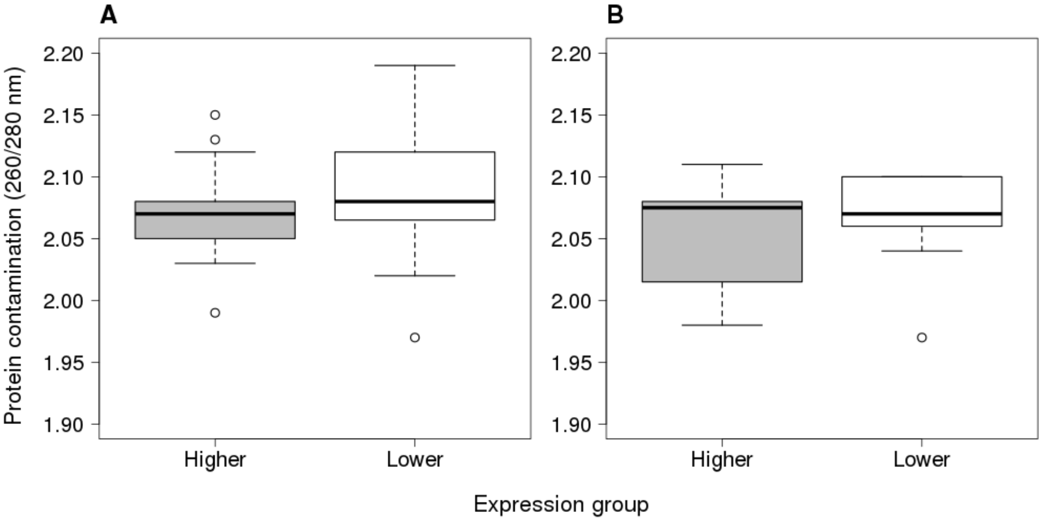


Figure S7. Protein contamination values in the two expression groups of inotocin hormone. (A) Samples from 2016 and (B) samples from 2018.

Activity and leaf collection during the breeding period

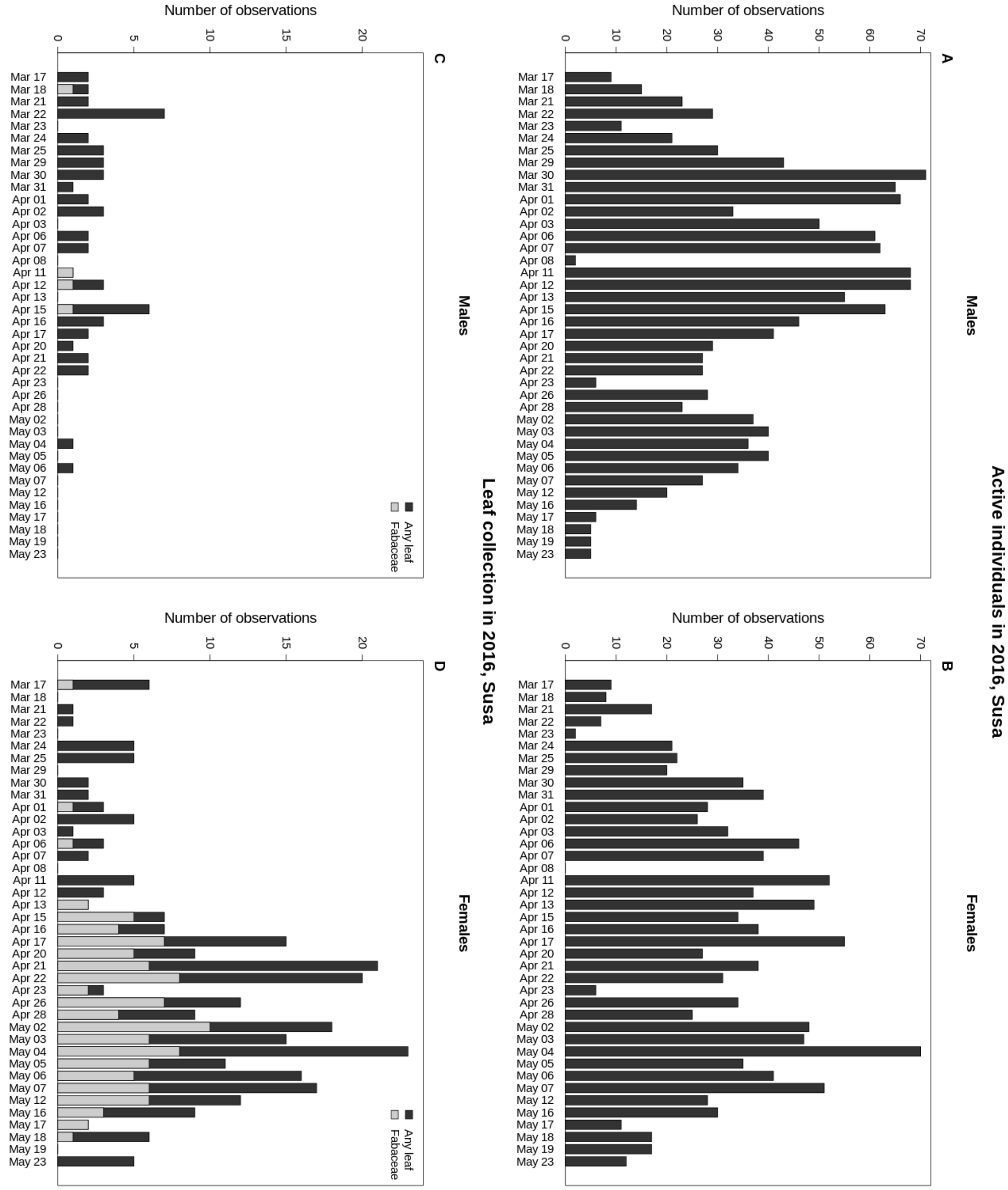


Figure S8. The seasonal variation in activity of *L. apterus*. (A) Number of observed male and (B) number of observed female individuals during the transect survey. (C) Number of leaf collection events by males and (D) females. Light grey bars represent the collection of a leaf of a legume species, whereas dark grey bars depict leafs of any taxa.

Attachments

The attachment of the published articles is approved by the journals.

1. Entomological Science - Licence number: 5024950970098
2. PeerJ - Open Access

Entomological Science

Entomological Science (2015) **18**, 292–294 doi:10.1111/ens.12123

SHORT COMMUNICATION

Predominant female care in the beetle *Lethrus apterus* with supposedly biparental care

András KOSZTOLÁNYI[1], Nikoletta NAGY[1], Tibor KOVÁCS[2] and Zoltán BARTA[1]

[1]*MTA-DE "Lendület" Behavioural Ecology Research Group, Department of Evolutionary Zoology, University of Debrecen, Debrecen, Hungary and* [2]*Mátra Museum of the Hungarian Natural History Museum, Gyöngyös, Hungary*

Abstract

Although intensive care for offspring by both parents is rare in arthropods, it occurs in some species including the beetle *Lethrus apterus*. According to previous publications, in this species the male collects leaves, which are used by the female to form balls in the underground nest burrow. These balls serve as food for the hatched offspring. Most knowledge about the behavior of this species is based on information collected more than a century ago. Therefore, we investigated above-ground breeding behavior and the status of nest burrows of this beetle in its natural habitat in Hungary. Our results suggest that contrary to previously documented cases, above-ground parental care, i.e. the collection of leaves, is done predominantly not by the males but the females. Further research is needed to understand the role of the sexes in parental care in this species and to explain the discrepancy between the previously documented cases and the results we report here.

Key words: Coleoptera, Geotrupidae, parental roles, role specialization.

Family life was long thought being ideal: in species with biparental care the female and male cooperate to raise the common offspring. In his seminal work, Trivers (1972) realized that this is not the case: family life is rife with conflicts, and there is an ongoing battle between the sexes. In the last 40 years biparental care became a model for studying cooperation and conflict between unrelated individuals (Székely *et al.* 2010). Biparental care is over-represented in some taxa; for example, most bird species have biparental care (Cockburn 2006). Care by both parents is relatively rare among invertebrates (Clutton-Brock 1991); however, it occurs in several arthropod groups (Wilson 1971; Zeh & Smith 1985; Tallamy 1999; Trumbo 2012), and it is relatively common in Scarabaeidae (Monteith & Storey 1981; Sato 1998; Hunt & House 2011).

The beetle *Lethrus apterus* (Laxmann, 1770) (Coleoptera: Geotrupidae) is a biparental species with Eastern European and Anatolian distribution. The western edge of its distribution is in Hungary (Merkl & Vig 2009). The sexes are dimorphic, as males are larger and have two tusks (ventral mandibular processes). According to the literature the male defends the nest burrow from intruders and pulls cut leaves into the nest underground, whereas the female forms balls from the leaves in the brood chambers where the eggs were laid previously (von Lengerken 1939; Wilson 1971; Clutton-Brock 1991).

We investigated the breeding behavior of *L. apterus* in its natural habitat in Northern Hungary, and here we report on the observed division of labor between sexes in the above-ground leaf collecting activity and on the status of the brood cells in the nest burrows.

We studied *L. apterus* in Northern Hungary in the Mátra Mountains at Dorogháza village (47°59′29″N, 19°53′36″E), where more than 5000 individuals can be found on the grass-covered slopes of hills. The study was carried out between 15 April and 17 June and between 12 and 27 September 2013. Beetles emerged early April and their above-ground activity had mostly ceased by mid June. Data on leaf collecting behavior of individually marked individuals were collected. Between 15 April and 7 June, 89 individuals (44 females and 45 males) were individually marked with paint marker (Marabu Brilliant, Bietigheim-Bissingen, Germany or

---

Correspondence: András Kosztolányi, MTA-DE "Lendület" Behavioural Ecology Research Group, Department of Evolutionary Zoology, University of Debrecen, H-4032 Debrecen, Egyetem tér 1, Hungary.
Email: andras.kosztolanyi@gmail.com

Received 12 March 2014; accepted 27 June 2014.

Marvy Uchida DecoColor, Tokyo, Japan) after their capture either at the entrance of their nest or if encountered above ground. Between 1 May and 12 June we spent over 100 h in the field, and recorded all occasions when an individually marked individual was observed carrying a leaf. The collected leaves can be used to form food balls for the offspring but can serve also as food for the parents themselves. Therefore, to ensure that the leaves were collected also for creating food balls, some nests were filled with plaster of Paris and dug out to check the status of the food cells underground (on 23 May and between 12 and 27 September 2013). The entrances of nests investigated in September had been filled with a 10–15 cm long plastic tube in June after the above-ground activity of beetles ceased to prevent the entrance from collapsing due to grazing stock. A large nail was also hammered in the soil at the entrance, and a metal detector (Silver Star 3; F. Chrenkó, Szigetszentmiklós, Hungary) was used to find the exact position of nests in September.

The distribution of number of marked males and females either observed or not observed to collect leaf material was analyzed in 2 × 2 contingency tables with Chi-squared and Fisher's exact tests in R v3.0.3 (R Core Team 2014).

From the 89 marked individuals 21 were observed at least once to carry leaf material. While the number of marked individuals was approximately equally distributed between the sexes (44 females and 45 males), the distribution of leaf-collecting individuals was highly biased toward females: 18 females *vs* three males (Fig. 1; Chi-squared test, $\chi^2_1 = 12.63$, $P < 0.001$). This highly unequal distribution holds also if only the individuals that presumably had already mated (i.e. both a female and a male were marked at the same nest) are considered: from 16 females and males caught at 16 nests, eight females and zero males were observed to carry leaf material (Fisher's exact test, $P = 0.002$).

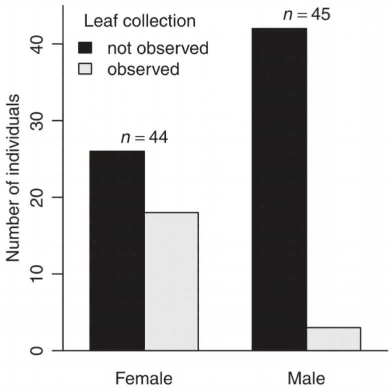


Figure 1 Number of individually marked females and males observed (light gray) or not observed (dark gray) to collect leaf material in a *Lethrus apterus* population in Northern Hungary.

The status of the nest burrow of two females observed to collect leaf material was checked on 23 May. These two nests contained three and four brood cells, respectively, filled with leaves and also an egg was found at one of the brood cells. In September when the offspring are already fully developed we checked the nest of 13 females and two males observed to collect leaf material in May and June. All 15 nests contained brood cells, and there were 6.5 ± 4.56 (mean ± SD) brood cells in the nests. The nests of the two males contained no developed offspring, only one and two dead larvae. However, 12 nests of the 13 females contained 3.9 ± 2.84 developed offspring, and one nest contained only one dead larva.

We found that in a population of *L. apterus* in Northern Hungary during the breeding season there was a remarkable sex difference in leaf-collecting activity of individually marked individuals: predominantly the females were responsible for carrying the leaves to the nest. Because at least 12 from the 21 leaf-collecting individuals reproduced successfully, we can conclude that these leaves were collected at least partly to serve as food for the offspring.

Our findings are surprising in a way given that this species has been characterized as a good example for role specialization between the parents with males collecting the leaves for the offspring (e.g. von Lengerken 1939; Wilson 1971; Clutton-Brock 1991). An obvious explanation for this discrepancy can be that reviews on the parental behavior of this species (von Lengerken 1939; Wilson 1971; Clutton-Brock 1991) are heavily based on the detailed observations of Schreiner (1906) in South Russia at that time. One may argue that the same species has different behavioral patterns in a population located more than 1000 km from Schreiner's (1906) original observations. This argument seems to be especially reasonable because this flightless species has presumably low dispersal ability, and intensive speciation was observed in other, closely related *Lethrus* species on the Balkan peninsula (Král & Hillert 2013). However, this explanation is unlikely because Emich (1884) observed *L. apterus* in Hungary, and reported the same parental care behavior as Schreiner (1906). Although Schreiner (1906) disagrees with Emich's (1884) findings at several points, he describes the division of labor between the sexes in leaf collection and ball forming the same way as Emich (1884). Thus, we can conclude that parental behavior was similar in South Russia and Hungary more than a century ago; at both places males heavily took part in above-ground leaf collection.

Another explanation for our results may be that this behavior is characteristic for our study population only. Until recently *L. apterus* was a common species in Hungary; however, because of intense agricultural activities in the last century, most of its habitats were destroyed and now it can be found only in fragmented populations (Merkl & Vig 2009). Altered circumstances may lead to quick changes in behavior (cf. the recent behavioral changes in many bird species in relation to urbanization; Sol *et al.* 2013). However, it seems that this behavioral pattern is not restricted to the population investigated in this study, as undocumented observations suggest that *L. apterus* in the Buda Hills in Hungary located about 100 km away the focal population has similarly female-biased leaf collecting role specialization (S. Bérces, pers. comm., 2013).

At the moment, it is unclear what roles the *L. apterus* males have in parental care, if they have any, and whether males participate in underground food ball preparation. Nests of two leaf-collecting males were checked for offspring and these nests contained brood cells and dead larvae, but unfortunately we had no information on female behavior at these nests. Thus, it is unknown what proportion of food balls was collected by the males in these cases. The nests of 15 leaf-collecting females were excavated (two in May and 13 in September), and at eight of these nests the male was also marked. That is, a male was present at least for a part of the breeding period; however, these males were not observed to collect leaf material. Thus, it is unknown whether they took part in above-ground parental care.

In summary, we showed that in a beetle species with presumed role specialization during parental care, the division of labor between the sexes is different from the accessible documented cases. Further investigations of the parental role of the sexes and especially the exploration of underground behavior in this beetle are needed to solve this enigma.

ACKNOWLEDGMENTS

We are grateful to Tamás Varga for allowing us to conduct fieldwork on his property. Permission (KTVF: 5866-4/2013) for the fieldwork was provided by the National Inspectorate for Environment, Nature and Water of Hungary.

REFERENCES

Clutton-Brock TH (1991) *The Evolution of Parental Care*. Princeton University Press, Princeton, NJ.

Cockburn A (2006) Prevalence of different modes of parental care in birds. *Proceedings of the Royal Society of London Series B, Biological Sciences* **273**, 1375–1383.

Emich G (1884) Die Metamorphose des *Lethrus apterus*. *Mathematische und Naturwissenschaftliche Berichte aus Ungarn* **2**, 184–188.

Hunt J, House C (2011) The evolution of parental care in the Onthophagine dung beetles. In: Simmons LW, Ridsdill-Smith TJ (eds) *Ecology and Evolution of Dung Beetles*, pp 153–176. Wiley-Blackwell, Chichester.

Král D, Hillert O (2013) Three new *Lethrus* species close to *L. raymondi* (Coleoptera: Geotrupidae) from the Balkan Peninsula. *Acta Entomologica Musei Nationalis Pragae* **53**, 219–244.

von Lengerken H (1939) *Die Brutfürsorge- und Brutpflege-instinkte der Käfer*. Akademische Verlagsgesellschaft MBH, Leipzig.

Merkl O, Vig K (2009) *[Beetles in the Pannonian Region]* . Vas Megyei Múzeumok Igazgatósága, BKL Kiadó, Magyar Természettudományi Múzeum, Szombathely. (In Hungarian.)

Monteith GB, Storey RI (1981) The biology of *Cephalodesmius*, a genus of dung beetles which synthesizes "dung" from plant material (Coleoptera: Scarabaeidae: Scarabaeinae). *Memoirs of the Queensland Museum* **20**, 253–277.

R Core Team (2014) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna. [Cited 7 Jan 2014.] Available from URL: http://www.R-project.org/

Sato H (1998) Male participation in nest building in the dung beetle *Scarabaeus catenatus* (Coleoptera: Scarabaeidae): mating effort versus paternal effort. *Journal of Insect Behavior* **11**, 833–843.

Schreiner J (1906) Die Lebensweise und Metamorphose des Rebenschneiders oder grossköpfigen Zwiebelhornkäfers (*Lethrus apterus* Laxm.). *Horae Societatis Entomologicae Rossicae* **37**, 197–208.

Sol D, Lapiedra O, González-Lagos C (2013) Behavioural adjustments for a life in the city. *Animal Behaviour* **85**, 1101–1112.

Székely T, Moore AJ, Komdeur J (eds) (2010) *Social Behaviour: Genes, Ecology and Evolution*. Cambridge University Press, Cambridge.

Tallamy DW (1999) Child care among the insects. *Scientific American* **280**, 72–77.

Trivers RL (1972) Parental investment and sexual selection. In: Campbell B (ed.) *Sexual Selection and the Descent of Man*, pp 136–179. Aldine Press, Chicago, IL.

Trumbo ST (2012) Patterns of parental care in invertebrates. In: Royle NJ, Smiseth PT, Kölliker M (eds) *The Evolution of Parental Care*, pp 81–100. Oxford University Press, Oxford.

Wilson EO (1971) *The Insect Societies*. Belknap Press of Harvard University Press, Cambridge, MA.

Zeh DW, Smith RL (1985) Paternal investment by terrestrial arthropods. *American Zoologist* **25**, 785–805.

Evaluation of potential reference genes for real-time qPCR analysis in a biparental beetle, *Lethrus apterus* (Coleoptera: Geotrupidae)

Nikoletta A. Nagy[1,2], Zoltán Németh[1,2], Edit Juhász[1], Szilárd Póliska[3], Rita Rácz[1,2], András Kosztolányi[2,4] and Zoltán Barta[1,2]

[1] Department of Evolutionary Zoology, University of Debrecen, Debrecen, Hungary
[2] MTA-DE Behavioural Ecology Research Group, University of Debrecen, Debrecen, Hungary
[3] Genomic Medicine and Bioinformatic Core Facility, Department of Biochemistry and Molecular Biology, Faculty of Medicine, University of Debrecen, Debrecen, Hungary
[4] Department of Ecology, University of Veterinary Medicine Budapest, Budapest, Hungary

Submitted 29 July 2017
Accepted 26 October 2017
Published 28 November 2017

Corresponding author
Nikoletta A. Nagy,
nnolett@gmail.com

Academic editor
Ilaria Negri

Additional Information and Declarations can be found on page 13

DOI **10.7717/peerj.4047**

ABSTRACT

Hormones play an important role in the regulation of physiological, developmental and behavioural processes. Many of these mechanisms in insects, however, are still not well understood. One way to investigate hormonal regulation is to analyse gene expression patterns of hormones and their receptors by real-time quantitative polymerase chain reaction (RT-qPCR). This method, however, requires stably expressed reference genes for normalisation. In the present study, we evaluated 11 candidate housekeeping genes as reference genes in samples of *Lethrus apterus*, an earth-boring beetle with biparental care, collected from a natural population. For identifying the most stable genes we used the following computational methods: geNorm, NormFinder, BestKeeper, comparative delta Ct method and RefFinder. Based on our results, the two body regions sampled (head and thorax) differ in which genes are most stably expressed. We identified two candidate reference genes for each region investigated: ribosomal protein L7A and RP18 in samples extracted from the head, and ribosomal protein L7A and RP4 extracted from the muscles of the thorax. Additionally, L7A and RP18 appear to be the best reference genes for normalisation in all samples irrespective of body region. These reference genes can be used to study the hormonal regulation of reproduction and parental care in *Lethrus apterus* in the future.

Subjects Entomology, Genetics, Molecular Biology
Keywords Insect, Parental care, Housekeeping gene

INTRODUCTION

Hormonal regulation in insects generates great interest among entomologists but hormones have only been studied in detail in a few species (*Gullan & Cranston, 2014*). Insect hormones of particular interest include juvenile hormones, ecdysteroids and neuropeptides. These molecules regulate a vast number of physiological and developmental processes as well as behaviours (*Gäde, Hoffmann & Spring, 1997*). Studying these hormones used to be difficult considering their small amount and the occasional instability (*Gullan & Cranston, 2014*).

How to cite this article Nagy et al. (2017), Evaluation of potential reference genes for real-time qPCR analysis in a biparental beetle, *Lethrus apterus* (Coleoptera: Geotrupidae). **PeerJ 5:e4047; DOI 10.7717/peerj.4047**

The technical revolution of molecular biology and genetics, however, made it attainable to discover the details of genetic and hormonal regulation in insects (*Raikhel, Brown & Belles, 2005*). Some of the processes controlled by hormones mentioned above, such as ecdysis (*Mykles et al., 2013*), are already well described. Nevertheless, there are many other interesting physiological and behavioural mechanisms, like parental care, the hormonal regulation of which are not well understood (*Panaitof et al., 2016*). One way to increase our understanding of hormonal regulation is to identify patterns of gene expression associated with the hormones in question (*Champagne & Curley, 2012*).

Real-time quantitative polymerase chain reaction (RT-qPCR) is a commonly used method for analysing gene expression as it is a sensitive, fast and reproducible method; moreover, it requires only a minimal amount of RNA (*Radonić et al., 2004*). With this method, gene expression levels can be measured simultaneously in several different samples for a limited number of genes. Gene expression analyses with RT-qPCR, however, require some kind of normalisation in order to control the variation caused by stochastic processes occurring during the analytic procedure (*Vandesompele et al., 2002*). This normalisation is usually achieved by taking into account the expression level of so-called reference genes (*VanGuilder, Vrana & Freeman, 2008*). These genes are usually selected from housekeeping genes, which produce proteins vital for maintaining fundamental cell functions, like ribosomal or cytoskeletal proteins. Therefore, the expression levels of these reference genes are thought to be relatively stable. Thus, comparing the expression level of genes of interest with the expression levels of the reference genes, we can eliminate the differences caused by the different amount and quality of starting material. With this method we are able to control for differences occurring due to technical errors during sample preparation as well (e.g., RNA isolation and cDNA synthesis, *Radonić et al., 2004*). Nonetheless, expression levels of the housekeeping genes may also vary considerably under certain circumstances because they can be involved in processes other than maintenance functions of the cell, e.g., apoptosis (*Nicholls, Li & Liu, 2012*), cytokinesis (*D'Souza-Schorey & Chavrier, 2006*) and development (*Zhou et al., 2015*). Therefore, a given housekeeping gene cannot automatically serve as reference gene, and normalisation with unstable reference genes can lead to erroneous quantification results and conclusions (*Thellin et al., 1999*). Consequently, reference genes must be carefully selected so that their expression levels are similar between the different samples and should not be influenced significantly by different experimental conditions (*VanGuilder, Vrana & Freeman, 2008*). According to *Vandesompele et al. (2002)*, the combination of two or more reference genes is highly recommended for normalisation to obtain more accurate results. In case of multiple reference genes, it is advised to use the geometric mean for normalisation since it better controls for extreme values and the possible differences between expression levels of the different genes (*Vandesompele et al., 2002*).

Lethrus apterus (Laxmann, 1770) (Coleoptera: Geotrupidae) is an earth-boring beetle that has biparental care during which the parents provision food for their offspring in advance their hatching (*Kosztolányi et al., 2015*). This kind of parental care is a complex and relatively rare trait among insects (*Smiseth, Kölliker & Royle, 2012*) and makes this beetle an outstanding model species for studying the hormonal background of parental care. In order to do so, however, stably expressed reference genes have to be identified.

In recent years, numerous studies aimed to identify stable reference genes in insects (*Lord et al., 2010*; *Ponton et al., 2011*; *Bansal et al., 2012*; *Li et al., 2013*; *Pan et al., 2015*; *Yu et al., 2016*). Nevertheless, there is a lack of reference gene studies that use individuals from natural populations. Our objective in this study was to examine the expression stability of several housekeeping genes in *Lethrus apterus* across different times of the breeding period in a natural population in order to identify the most stable reference gene(s). With the right combination of reference genes, the expression levels of hormone regulating genes involved in parental care in *Lethrus apterus* can be examined accurately in the future. Based on the literature (*Shi et al., 2013*; *Liang et al., 2014*; *Zhu et al., 2014*; *Yang et al., 2015*), we probed eleven housekeeping genes.

MATERIALS AND METHODS

Sample collection

Samples were collected near Dorogháza, northern Hungary (47°59′29″N, 19°53′36″E) on 16th April, 4th May and 28th May in 2015, which dates corresponded to the beginning, middle and end of the breeding season of *Lethrus apterus*, respectively. Sample collection was approved by the Nothern Hungarian Inspectorate for Environment Protection and Nature Conservation (No. 9007-8/2014). The first sampling date represents the period of mate choice, while the second and third samplings were done during the period when parents were collecting leaves for the offspring. On each sampling dates, head and thorax samples were collected from eight males and eight females. All tissues were removed from the head capsule and muscle samples were taken from the thorax. Samples were collected in the field in less than five minutes after euthanizing the individuals. Each head and thorax sample was put immediately into separate eppendorf tubes which already contained 600 µl RNA*later*® Stabilization Solution (Thermo Fisher Scientific, Waltham, MA, USA), then stored at −20 °C in the laboratory in order to inhibit RNase enzyme activity until RNA extraction.

RNA extraction and cDNA synthesis

Total RNA was isolated from each samples using TRIzol® Reagent (Thermo Fisher Scientific, Waltham, MA, USA) following the manufacturer's instructions. The extracted RNA was eluted in 15–30 µl RNase-free water, depending on the pellet size. Yield of RNA was quantified by NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). To eliminate genomic DNA, samples were treated with RQ1 RNase-Free DNase (Promega, Madison, WI, USA) just before the reverse transcription. First strand cDNA was synthesized from 1 µg DNA-free RNA using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems, Foster City, CA, USA).

Reference gene selection and primer design

Using a draft genome of *Lethrus apterus* (unpublished data) eleven reference genes, which were already described as stable reference genes in other arthropods, were selected

Table 1 The list of the candidate housekeeping genes with their biological functions.

Gene	Symbol used	Function	Reference
glyceraldehyde 3-phosphate dehydrogenase	GAPDH	glycolytic enzyme	*Liang et al. (2014)*
tubulin alpha-1 chain	TUB1a	cytoskeletal structural protein	*Liang et al. (2014)*
elongation factor 1-alpha	EF1a	protein synthesis	*Liang et al. (2014)*
elongation factor 2	EF2	protein synthesis	*Zhu et al. (2014)*
ADP-ribosylation factor-like protein 1	ARF1	GTP-binding protein	*Shi et al. (2013)*
ADP-ribosylation factor 4	ARF4	GTP-binding protein	*Shi et al. (2013)*
ribosomal protein S8	RPS8	structural constituent of ribosome	*Yang et al. (2015)*
ribosomal protein L4	RP4	structural constituent of ribosome	*Shi et al. (2013)*
ribosomal protein L7A	L7A	structural constituent of ribosome	*Zhu et al. (2014)*
ribosomal protein L10	L10	structural constituent of ribosome	*Zhu et al. (2014)*
ribosomal protein L18	RP18	structural constituent of ribosome	*Shi et al. (2013)*

(Table 1). We manually designed primers (Table 2) using the web-based Sequence Manipulation Suite (*Stothard, 2000*) and Multiple Primer Analyzer (Thermo Fisher Scientific, Waltham, MA, USA) in order to avoid the forming of possible secondary structures of the primers. To check the specificity of primer pairs and to determine optimal annealing temperature, PCR reactions were performed in 10 µl volumes containing the following components: 10x buffer, 2 mM $MgCl_2$, 0.2 mM dNTP, 0.02 U/µL Taq DNA polymerase enzymes (DreamTaq Green, Thermo Fisher Scientific, Waltham, MA, USA), 0.2 µM forward and 0.2 µM reverse primer and 0.1 µg cDNA. PCR conditions were optimized by determining the optimal annealing temperature using temperature gradient ranging from 54 °C to 62 °C for primer binding. In this study, we used ABI Veriti® 96-Well Thermal Cycler (Applied Biosystems, Foster City, CA, USA). Cycling conditions consisted of a denaturing step at 95 °C for 2 min followed by 40 cycles at 95 °C for 30 sec, at a temperature gradient (54 °C, 56 °C, 58 °C, 60 °C or 62 °C) for 30 s and at 72 °C for 90 s, and finally at 72 °C for 10 min. PCR amplicons were run on 1% agarose gel stained with GelRed™ (Biotium, Fremont, CA, USA).

Real-time quantitative PCR

RT-qPCR was performed on a QuantStudio 12K Flex Real-Time PCR System (Applied Biosystems, Foster City, CA, USA) using SYBR® Green PCR Master Mix (Applied Biosystems, Foster City, CA, USA) and ROX Passive Reference Dye (Affymetrix, Santa Clara, CA, USA). Amplifications were carried out under the following conditions: initial denaturation at 95 °C for 10 min followed by 40 cycles of 10 sec at 95 °C and for 1 min at the optimal annealing temperature. This was followed by a melting curve analysis in which the temperature raised from 65 °C to 95 °C in sequential steps of 0.05 °C for 1 s. Three technical replicates were performed for each biological sample, and the average cycle threshold (Ct) values of triplicates were calculated. Furthermore, no-template control was done in order to check whether primer-dimers or contamination with amplified PCR product were detectable. Five 5-fold serial dilution was made from

Table 2 **The primers used to measure gene expression levels for the candidate reference genes by RT-qPCR.**

Gene	GenBank accession number	Primer sequence (5′to 3′)[a]	Amplicon length (bp)	Tm (°C)[b]	*E* (%)[c]	R^2[d]
GAPDH	KY786279	F: GCCATTCCAGTAAGTTTTCCATTGAG	157	85.0	100.75	0.91
		R: GCTGTTACTGCTACACAAAAGAC				
TUB1a	KY786273	F: CAGACTGCACGTTGGACTTTAGC	172	83.6	100.04	0.96
		R: TACAGAGGAGATGTTGTCCCCAAG				
EF1a	KY786281	F: AAACCTTTGCGTCTTCCACTACAGG	184	81.7	99.83	0.94
		R: CTTCAGTTGTAAGACCAACAGGTG				
EF2	KY786280	F: GATGAGAAATCCACATGTCCAG	244	82.0	102.00	0.86
		R: CGACTCCCTAGTATCAAAGG				
ARF1	KY786283	F: GTATGACAGTAGCTGAAGTTC	141	81.4	112.70	0.84
		R: CTGTTTTGTAAAGCATTGGC				
ARF4	KY786282	F: TAGTACGGACGGTCAAGTC	197	89.1	105.91	0.81
		R: GTAGACCGTCACCTGTTATGGC				
RPS8	KY786274	F: CATTATGTACGTACGAGAGGAGGCAACG	200	84.0	99.96	0.91
		R: TCTAAAGGGAGTAGCGTCGATAACG				
RP4	KY786275	F: TAATGGACCACGACGCTGTATGC	248	84.5	100.33	0.92
		R: CGTACCAGCTTTAGTAATGAGCAAGG				
L7A	KY786277	F: TAGCGACTCAACTGTTCAAGG	224	84.8	99.54	0.95
		R: CCTCAATTGGATCGACGTCATGTG				
L10	KY786278	F: CGTAGAGCCTCGATAACTTGG	210	84.7	99.33	0.94
		R: TCATGTGCTGGAGCTGATAGG				
RP18	KY786276	F: TTGTAACCACATGAACGCCTACG	186	85.2	99.75	0.96
		R: AGTTAGCTTTACGTTCACCTACTGG				

Notes.

[a] F, forward primer; R, reverse primer.

[b] melting temperature.

[c] real-time qPCR efficiency (calculated by the standard curve method).

[d] regression coefficient (calculated from the regression line of the standard curve).

cDNA samples to create a standard curve, and the amplification efficiency was determined for each candidate gene. The efficiency (E) values were calculated according to the equation: $E = (10^{(-1/\mathrm{slope})} - 1) \times 100$, where slope is the slope of the standard curve (*Radonić et al., 2004*).

Statistical analysis of raw Ct values

In order to examine the differences between sample groups, random intercept mixed-effects models were used with sample id as a random factor for each gene. Significance of fixed terms was investigated by likelihood ratio tests. For likelihood ratio tests models were fitted using Maximum Likelihood estimation. The analyses were carried out using "lme4" (*Bates et al., 2014*) and "car" (*Fox & Weisberg, 2011*) packages in the R statistical environment version 3.3.2 (*R Core Team, 2016*).

Determination of reference gene expression stability

In order to determine the expression stability of the selected reference genes, we used the following methods: geNorm (*Vandesompele et al., 2002*), NormFinder (*Andersen, Jensen & Ørntoft, 2004*), BestKeeper (*Pfaffl et al., 2004*), delta Ct method (*Silver et al., 2006*) and RefFinder (*Xie et al., 2012*). For the analyses with the geNorm and NormFinder procedures, the average Ct values were transformed to relative quantities by dividing sample values by the lowest average Ct value. For calculations by BestKeeper, delta Ct method and RefFinder, the untransformed average Ct values were used. All calculations, except the ones done by the web-based RefFinder, were carried out in R with "NormqPCR" package (*Perkins et al., 2012*).

geNorm calculates the expression stability value M by assessing the mean pairwise expression ratio for each candidate gene against all the other candidates (*Vandesompele et al., 2002*). The basic assumption of this method is that the expression ratio between two reference genes is identical across the samples. The lower the M value the more stable the expression of the candidate reference gene. Stepwise exclusion of the genes with the highest M value results in the selection of the two most stably expressed reference genes in the tested samples both sharing the same M value. *Vandesompele et al. (2002)* also suggest not to accept candidate genes as stably expressed reference genes with M value higher than 1.5. Moreover, the procedure determines the normalisation factor by taking the geometric mean of the expression levels from the most stable genes and then additively recalculating with each of the next most stable gene. The pairwise variation, $V_{n/n+1}$ between two sequential normalisation factors is then calculated in order to determine the effect of each newly added gene to the normalisation factor. The optimum number of genes is the lowest number of genes with $V_{n/n+1}$ less than 0.15 (*Vandesompele et al., 2002*).

NormFinder determines the stability of the candidate reference genes by measuring the intra- and intergroup variation between user specified groups (e.g., male and female groups or treated and control groups) first. Stability values for each candidate gene are then calculated by adding the two sources of variation. The lowest stability value means the most stable expression (*Andersen, Jensen & Ørntoft, 2004*).

BestKeeper calculates, for each candidate reference gene across the samples, the geometric mean, the arithmetic mean, the minimal and the maximal Ct values, in addition to the average absolute deviation from the arithmetic mean. Genes with the lowest average absolute deviation can be considered as stably expressed reference genes. BestKeeper Index is calculated as the geometric mean of the Ct values of the candidate reference genes. Inter-gene relations are estimated by performing pairwise correlation analyses of all possible reference gene pairs. Furthermore, correlation between the expression level of each candidate gene and the BestKeeper Index is calculated, describing the relation between the index and the contributing genes by the Pearson correlation coefficient, coefficient of determination and the corresponding p-value (*Pfaffl et al., 2004*).

The delta Ct method compares relative expression of pairs of candidate genes within each sample in order to identify the stably expressed housekeeping genes. If the ΔCt value of the two genes fluctuates when analysed in different samples, it means that one or both genes are variably expressed. If the ΔCt value remains constant, both genes are stably expressed among the samples (*Silver et al., 2006*).

Each procedure mentioned above uses different algorithms to calculate an expression stability value which represents the suitability of the candidate genes as reference genes, therefore the ranking of the examined genes according to the methods may vary. The web-based tool RefFinder (*Xie, 2012*) was used in order to combine our results and rank the candidate genes. This user-friendly program integrates the four methods mentioned above. Using the ranking from each program, it assigns an appropriate weight to an individual gene and calculates the geometric mean of their weights for the overall ranking. The lowest rank indicates the most stably expressed gene (*Xie et al., 2012*).

For each analysis, except for NormFinder, seven sample groups were used: all samples irrespective of body part or sex; head samples irrespective of sex; male head samples; female head samples; thorax samples irrespective of sex; male thorax samples; female thorax samples. The calculation by NormFinder requires subgroup specification, therefore, body regions were set as subgroups for the analysis of all samples. In order to investigate the effect of sexes, male and female subgroups were specified for the analysis of head and thorax sample groups separately. In this way, three sample groups, each divided into two subgroups, were analysed by NormFinder.

RESULTS

Transcriptional profiling of candidate reference genes

Before the evaluation of expression stability of the eleven candidate genes, specificity of each primer pair was checked on 1% agarose gel which showed single products with the expected sizes. Moreover, gene-specific amplification was confirmed by single melting curve peaks. These results indicate that no primer-dimers or nonspecific amplification products were formed. Additionally, no fluorescent signals were detected in the negative control during the RT-qPCR. Each amplicons were sequenced and annotated to the sequences from which the primer design was based in order to check that the correct genes were amplified. The sequences are available in File S1. The efficiency of the eleven candidate genes ranged from 99.33 to 112.70%. The efficiency values and other basic information of the RT-qPCR required based on the guideline of *Bustin et al. (2009)* are included in Table 2.

Raw Ct values ranged from 11.66 (TUB1a) to 30.12 (ARF4) (Fig. 1). The mean and standard deviation (SD) of the Ct values across all samples were calculated for each gene (Table 3). Since the mean Ct values ranged between 15 and 30 for all the candidate reference genes, all of them were analysed further (*Kozera & Rapacz, 2013*). ARF1 had the least variable expression level with the lowest SD value (SD = 1.85), while ARF4 had the most variable expression level (SD = 3.04). Low average Ct values indicate high expression level in TUB1a and EF2 ($Ct_{mean} = 15.11$), on the other hand, high Ct values of ARF1 ($Ct_{mean} = 20.77$) indicated low expression.

Based on the likelihood ratio tests, sex had no significant effect on the expression level of the candidate genes. However, significant effect of body region was found in case of six genes: GAPDH, EF1a, ARF1, ARF4, RP4 and L10. The interaction of sex and bodypart had no significant effect on the expression level of the candidates, except for RPS8 (Table 4).

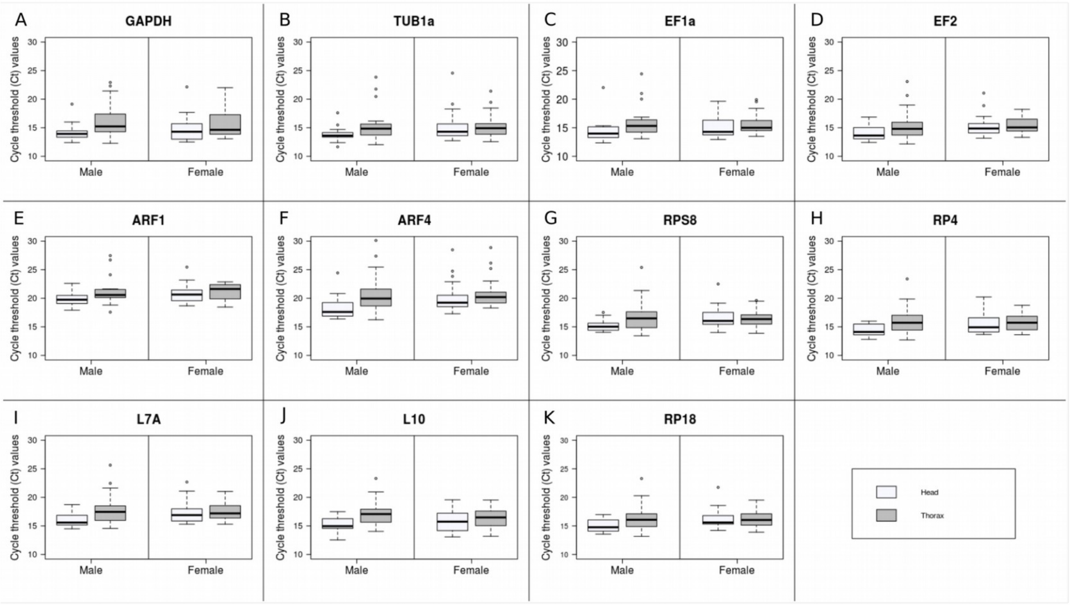


Figure 1 Expression profiles of the 11 candidate reference genes.

Full-size DOI: 10.7717/peerj.4047/fig-1

Table 3 Mean, standard deviation and coefficient of variation for the Ct values of 11 candidate reference genes, calculated across all samples.

Genes	Mean	SD	CV
GAPDH	15.27	2.62	0.17157826
TUB1a	15.11	2.64	0.17471873
EF1a	15.44	2.34	0.1515544
EF2	15.11	2.05	0.13567174
ARF1	20.77	1.85	0.08907078
ARF4	20.22	3.04	0.15034619
RPS8	16.38	2.12	0.12942613
RP4	15.52	1.94	0.125
L7A	17.27	2.1	0.12159815
L10	16.15	1.97	0.12198142
RP18	16.02	1.93	0.12047441

Expression stability of candidate reference genes

Based on geNorm analysis for all samples, eight candidate genes had an M value below the threshold of 1.5 (Table S1). The results show that the lowest *M* value was 0.390 for RPS8 and L7A. Among the head samples irrespective of sex, all of the tested genes except

Table 4 **Results of likelihood ratio tests on the effects of body region, sex and their interaction on the expression levels of the eleven candidate reference genes.**

Gene	Sex		Bodypart		Sex*Bodypart	
	χ^2	*P*	χ^2	*P*	χ^2	*P*
GAPDH	0.019	0.889	**5.668**	**0.017**	0.965	0.326
TUB1a	0.642	0.423	1.960	0.162	1.995	0.158
EF1a	0.002	0.968	**4.096**	**0.043**	0.857	0.355
EF2	1.216	0.270	2.271	0.132	2.030	0.154
ARF1	0.476	0.490	**4.147**	**0.042**	2.459	0.117
ARF4	1.417	0.234	**4.885**	**0.027**	1.289	0.256
RPS8	0.586	0.444	2.307	0.129	**4.007**	**0.045**
RP4	0.356	0.551	**4.272**	**0.039**	2.384	0.127
L7A	0.556	0.456	3.087	0.079	2.901	0.089
L10	0.246	0.620	**8.509**	**0.004**	3.319	0.068
RP18	1.028	0.311	2.336	0.126	2.896	0.089

Notes.

Significant effects are highlighted in bold.

L10 had an M value below 1.5, and RPS8 and RP18 were co-ranked as the most stable genes from the candidates ($M = 0.304$). Furthermore, the same two genes had the lowest M value considering male and female head samples separately ($M = 0.264$ for females and $M = 0.346$ for males). In case of the thorax samples irrespective of sex, eight genes had an M value below the threshold. RPS8 and L7A were the most stable candidate gene pair with an M value of 0.358. In thorax samples collected from females, RPS8 and RP18 were the most stable genes as well with an M value of 0.222. However, in thorax samples of males, RPS8 and L7A were ranked as the best reference gene pair ($M = 0.288$).

According to NormFinder, L7A was the most stable gene when calculating with all samples divided into groups of head and thorax samples (Table S2). The second and third genes were RP4 and RPS8, indicating that these are also worth considering as reference genes. In the case of specifying males and females as subgroups within head and thorax samples, L7A was found again to be the most stably expressed gene among the candidate ones. In both head and thorax samples, L7A was followed by similar ranking order: EF2, RP4, RP18 and RPS8 as second, third, fourth and fifth genes, respectively.

Based on BestKeeper, across all samples ARF1 had the lowest mean absolute deviation (MAD) value; however, L7A had the highest correlation r value (Table S3). In the group of head samples irrespective of sex, ARF1 had the lowest MAD value, while among the thorax samples irrespective of sex, L10 was the most stable according to the MAD value. This was surprising as the other programs ranked this gene consistently as one of the least stable genes. On the other hand, in both head and thorax samples, L7A had the highest r value. In head samples of males RPS8 (MAD = 0.827), and of females ARF1 (MAD = 1.122) were the most stable candidate genes. In both male and female head samples, L7A had the highest correlation r value. Considering male thorax samples L10 had the lowest MAD value (MAD = 1.576), while in female thorax samples EF2 was ranked as the most stable

Table 5 **Stability ranking of the eleven candidate reference genes in the different sample groups as calculated by RefFinder.**

Rank	All samples	Head samples			Thorax samples		
		All head samples	Male head samples	Female head samples	All thorax samples	Male thorax samples	Female thorax samples
1	L7A	L7A	RPS8	RP18	L7A	L7A	RP18
2	RP18	RP18	L7A	L7A	RP4	EF2	L7A
3	RPS8	RPS8	RP18	RPS8	RP18	RPS8	EF2
4	EF2	ARF1	EF2	EF2	RPS8	RP4	RPS8
5	ARF1	RP4	TUB1a	ARF1	EF2	L10	RP4
6	RP4	EF2	RP4	RP4	L10	EF1a	L10
7	EF1a	GAPDH	ARF1	EF1a	ARF1	RP18	ARF1
8	TUB1a	TUB1a	GAPDH	GAPDH	EF1a	ARF1	EF1a
9	L10	EF1a	L10	ARF4	TUB1a	ARF4	TUB1a
10	GAPDH	L10	ARF4	L10	ARF4	TUB1a	GAPDH
11	ARF4	ARF4	EF1a	TUB1a	GAPDH	GAPDH	ARF4

with MAD value 1.236. L7A had the highest r value in female thorax samples, however, in male thorax samples EF2 had the highest correlation r value.

According to the delta Ct method, L7A was the most stable gene among the candidates overall with the stability value always below 1.0 (Table S4).

Finally, the candidate genes were evaluated by RefFinder to combine the results of individual methods (Table 5). Using all samples irrespective of body region and sex, and separately the head and thorax samples irrespective of sex, L7A was ranked first, as the most stably expressed gene among the candidate reference genes. In head samples, RP18 was co-ranked with L7A as the most stable reference genes. In thorax samples, RP4 was ranked on the second place. In female head and thorax samples, RP18 was the most stably expressed gene of the candidates. Considering head samples of males, RPS8 was ranked on the first place, while in thorax samples of males, L7A was ranked as the best reference gene. L7A was ranked on the second place in all subgroups, with the exception of male thorax samples, where EF2 was the second best reference gene according to RefFinder.

Optimal number of reference genes

To determine the minimal number of genes necessary for normalisation, the V-value was computed by geNorm. The results demonstrated that across all samples $V_{2/3}$ was the first V-value lower than the cut-off value of 0.15 (Fig. 2). Considering separately the head and thorax samples, $V_{2/3}$ was again lower than 0.15. Separate analyses of female and male samples within head and thorax groups showed that $V_{2/3}$ was also the first value below the threshold in all cases (results not shown). Therefore, two stably expressed reference genes are sufficient for normalisation in any case of sample classification.

DISCUSSION

RT-qPCR is a widely used method for measuring gene expression levels due to its relatively low cost, high accuracy and sensitivity. A critical step of this method is data

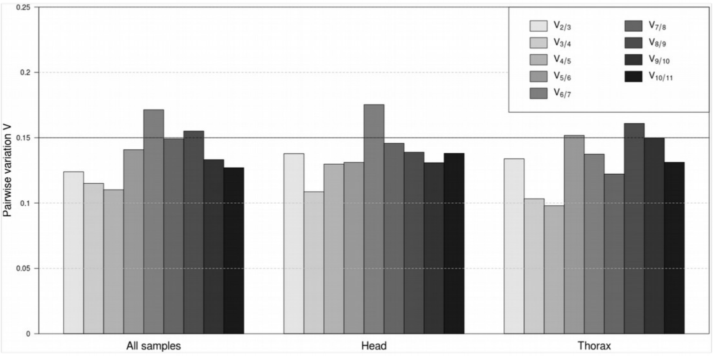


Figure 2 Pairwise variation analyses by geNorm to determine the optimal number of reference genes for accurate normalization. Pairwise variation for all samples together, as well as separately for head and thorax samples. The lowest number of genes with $V_{n/n+1}$ less than 0.15 means the optimum number of genes.

Full-size DOI: 10.7717/peerj.4047/fig-2

normalisation which requires careful selection of reference genes for the given experimental or environmental conditions. With these stably expressed genes, technical errors and variance resulting from the method can be moderated (*Udvardi, Czechowski & Scheible, 2008*). Several studies have examined the stability of reference genes in various insect species in the past decade and these studies suggest that no universally stable reference gene can be found that is applicable for all species, tissue types and experimental conditions. Hence, it is necessary to identify the most suitable reference genes for the specific circumstances in a given study for a given species (*Zhu et al., 2014*).

In the present study, variation in expression levels of eleven housekeeping genes were evaluated across a span of 1.5 months covering most of the breeding period of the biparental beetle *Lethrus apterus*. To date, no study investigated the possible reference genes either in this species, or in the family of Geotrupidae. We analyzed the expression stability of the candidate reference genes by four frequently used programs: geNorm, NormFinder, BestKeeper and comparative delta Ct method. The outcomes of these programs can vary because of the differences in the algorithms. Therefore, the combined use of them ensures more reliable results. For this purpose, RefFinder, a freely available web-based tool was used to calculate a comprehensive ranking value for each candidate gene.

According to the comprehensive ranking by RefFinder, the most stably expressed reference gene was L7A across all samples, irrespective of body region and sex. Based on the results of geNorm analysis, two reference genes are sufficient for normalisation in gene expression analysis in *Lethrus apterus* during the breeding period. For accurate normalisation, we recommend the use of L7A and RP18 in head samples irrespective of sex. When considering the sexes separately, RPS8 and L7A should be used for head samples of males, and RP18 and L7A for females. In thorax samples irrespective of sex, L7A and

RP4 are the best reference genes. In case of thorax samples, L7A and EF2 are recommended for normalisation in males, RP18 and L7A in females.

Consistent with our results, ribosomal proteins are reported to be the best reference genes in many insect species. In a study by *Zhu et al. (2014)*, ribosomal protein L7A was ranked as one of the best reference genes in *Spodoptera exigua* in different tissues, specific larval physiological stages and male individuals. Studies of other coleopterans gave similar results: RP4 and RP18 were the best reference genes in *Leptinotarsa decemlineata* (*Shi et al., 2013*), RPS3 (ribosomal protein S3), RPL13a (ribosomal protein 13a) and RPS18 (ribosomal protein S18) were suitable reference genes for *Tribolium castaneum* (*Lord et al., 2010*; *Sang et al., 2015*), and RPL22e (ribosomal protein 22e) was one of the best reference genes in *Mylabris cichorii* both in males and females (*Wang et al., 2014*). In other species, e.g., in *Drosophila melanogaste r* Rpl32 (ribosomal protein L32) was a suitable reference gene in individuals on different diets (*Ponton et al., 2011*), and in *Aphis craccivora,* RPS8, RPL14 (ribosomal protein L14), and RPL11 (ribosomal protein L11) were the three most stable housekeeping genes across different developmental stages and temperature conditions (*Yang et al., 2015*).

Interestingly, two frequently used reference genes, GAPDH and TUB1a were ranked as less stable genes in this study, beside ARF4, with stability values above the threshold values of all the programs used. L10 was also found to be an unstable candidate gene in all but the geNorm analysis. These results correspond with the findings of *Thellin et al. (1999)*, i.e., housekeeping genes should be evaluated as reference genes across the given experimental conditions in the given species. Based on our results, we recommend to avoid the use of these last four genes for normalisation in studies investigating gene expression patterns during the reproductive period in this species.

CONCLUSION

By evaluating the stability of eleven candidate housekeeping genes in samples collected during the breeding period of free-living *Lethrus apterus*, we conclude that two of them provide sufficient reference for normalising target gene expression. In head samples, these two genes appear to be L7A and RP18, whereas in thorax samples L7A and RP4 should be used. In both thorax and head samples of females, RP18 and L7A are the best choices for normalisation. Based on our results, in head samples of males, RPS8 and L7A, while in thorax samples of males, L7A and EF2 are recommended to use. These results provide reliable reference genes that are suitable normalizers for further RT-qPCR investigations on the hormonal regulation in *Lethrus apterus*.

ACKNOWLEDGEMENTS

We are grateful to Tamás Varga for allowing us to conduct fieldwork on his property and Adrien Fónagy for helping us in the development of the autopsy.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

The study was financed by NKFIH Grants (No. K112670 and No. K112527), and partially supported by the European Union and the European Social Fund through project EFOP-3.6.1-16-2016-00022, András Kosztolányi was supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences and Zoltán Németh was supported by an NKFIH Fellowship (No. PD121013). There was no additional external funding received for this study. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
NKFIH: K112670, K112527, PD 121013.
Hungarian Academy of Sciences.
European Union and the European Social Fund: EFOP-3.6.1-16-2016-00022.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Nikoletta A. Nagy conceived and designed the experiments, performed the experiments, analyzed the data, wrote the paper, prepared figures and/or tables, reviewed drafts of the paper.
- Zoltán Németh and András Kosztolányi conceived and designed the experiments, analyzed the data, wrote the paper, reviewed drafts of the paper.
- Edit Juhász, Szilárd Póliska and Rita Rácz conceived and designed the experiments, performed the experiments, wrote the paper, reviewed drafts of the paper.
- Zoltán Barta conceived and designed the experiments, analyzed the data, contributed reagents/materials/analysis tools, wrote the paper, reviewed drafts of the paper.

Field Study Permissions

The following information was supplied relating to field study approvals (i.e., approving body and any reference numbers):

Sample collection was approved by the Northern Hungarian Inspectorate for Environment Protection and Nature Conservation.

DNA Deposition

The following information was supplied regarding the deposition of DNA sequences:

The housekeeping gene sequences described here are available via GenBank accession numbers KY786273 to KY786283.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.4047#supplemental-information.

REFERENCES

Andersen CL, Jensen JL, Ørntoft TF. 2004. Normalization of real "—time quantitative reverse transcription"—PCR data: a model "—based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Research* **64(15)**:5245–5250.

Bansal R, Mamidala P, Mian RMA, Mittapalli O, Michel AP. 2012. Validation of reference genes for gene expression studies in *Aphis glycines* (Hemiptera: Aphididae). *Journal of Economic Entomology* **105**:1432–1438 DOI 10.1603/EC12095.

Bates D, Mächler M, Bolker B, Walker S. 2014. Fitting linear mixed-effects models using lme4. ArXiv preprint. arXiv:1406.5823.

Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, Mueller R, Tolan T, Pfaffl MW, Shipley GL, Vandesompele J, Wittwer CT. 2009. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clinical Chemistry* **55**:611–622 DOI 10.1373/clinchem.2008.112797.

Champagne FA, Curley JP. 2012. Genetics and epigenetics of parental care. In: Royle NJ, Smiseth PT, Kölliker M, eds. *The evolution of parental care*. Oxford: Oxford University Press, 304–326.

D'Souza-Schorey C, Chavrier P. 2006. ARF proteins: roles in membrane traffic and beyond. *Nature Reviews Molecular Cell Biology* **7**:347–358 DOI 10.1038/nrm1910.

Fox J, Weisberg S. 2011. *An R companion to applied regression*. Thousand Oaks: Sage Publications.

Gäde G, Hoffmann K-H, Spring JH. 1997. Hormonal regulation in insects: facts, gaps, and future directions. *Physiological Reviews* **77**:963–1032.

Gullan PJ, Cranston PS. 2014. *The insects: an outline of entomology*. Oxford: John Wiley & Sons, Ltd.

Kosztolányi A, Nagy N, Kovács T, Barta Z. 2015. Predominant female care in the beetle *Lethrus apterus* with supposedly biparental care. *Entomological Science* **18**:292–294 DOI 10.1111/ens.12123.

Kozera B, Rapacz M. 2013. Reference genes in real-time PCR. *Journal of Applied Genetics* **54**:391–406 DOI 10.1007/s13353-013-0173-x.

Li R, Xie W, Wang S, Wu Q, Yang N, Yang X, Pan H, Zhou X, Bai L, Xu B, Zhou X, Zhang Y. 2013. Reference gene selection for qRT-PCR analysis in the sweet-potato whitefly, *Bemisia tabaci* (Hemiptera: Aleyrodidae). *PLOS ONE* **8**:1–8 DOI 10.1371/journal.pone.0053006.

Liang P, Guo Y, Zhou X, Gao X. 2014. Expression profiling in *Bemisia tabaci* under insecticide treatment: indicating the necessity for custom reference gene selection. *PLOS ONE* **9**:1–8 DOI 10.1371/journal.pone.0087514.

Lord JC, Hartzer K, Toutges M, Oppert B. 2010. Evaluation of quantitative PCR reference genes for gene expression studies in *Tribolium castaneum* after fungal challenge. *Journal of Microbiological Methods* **80**:219–221 DOI 10.1016/j.mimet.2009.12.007.

Mykles DL, Adams ME, Gäde G, Lange AB, Marco HG, Orchard I. 2013. Neuropeptide action in insects and crustaceans. *Physiological and Biochemical Zoology* **83**:836–846 DOI 10.1086/648470.Neuropeptide.

Nicholls C, Li H, Liu JP. 2012. GAPDH: a common enzyme with uncommon functions. *Clinical and Experimental Pharmacology and Physiology* **39**:674–679 DOI 10.1111/j.1440-1681.2011.05599.x.

Pan H, Yang X, Siegfried BD, Zhou X. 2015. A comprehensive selection of reference genes for RT-qPCR analysis in a predatory lady beetle, *Hippodamia convergens* (Coleoptera: Coccinellidae). *PLOS ONE* **10**:1–15 DOI 10.1371/journal.pone.0125868.

Panaitof SC, Yaeger JDW, Speer JP, Renner KJ. 2016. Biparental behavior in the burying beetle *Nicrophorus orbicollis*: a role for dopamine? *Current Zoology* **62**:285–291 DOI 10.1093/cz/zow032.

Perkins JR, Dawes JM, McMahon SB, Bennett DL, Orengo C, Kohl M. 2012. ReadqPCR and NormqPCR: R packages for the reading, quality checking and normalisation of RT-qPCR quantification cycle (Cq) data. *BMC Genomics* **13**:296 DOI 10.1186/1471-2164-13-296.

Pfaffl MW, Tichopad A, Prgomet C, Neuvians TP. 2004. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper–Excel-based tool using pair-wise correlations. *Biotechnology Letters* **26**:509–515 DOI 10.1023/B:BILE.0000019559.84305.47.

Ponton F, Chapuis MP, Pernice M, Sword GA, Simpson SJ. 2011. Evaluation of potential reference genes for reverse transcription-qPCR studies of physiological responses in *Drosophila melanogaster*. *Journal of Insect Physiology* **57**:840–850 DOI 10.1016/j.jinsphys.2011.03.014.

Radonić A, Thulke S, Mackay IM, Landt O, Siegert W, Nitsche A. 2004. Guideline to reference gene selection for quantitative real-time PCR. *Biochemical and Biophysical Research Communications* **313**:856–862 DOI 10.1016/j.bbrc.2003.11.177.

Raikhel AS, Brown MR, Belles X. 2005. Hormonal control of reproductive processes. In: Gilbert LI, Iatrou K, Gill SS, eds. *Comprehensive molecular insect science*. Oxford: Elsevier, 433–491.

R Core Team. 2016. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. *Available at https://www.R-project.org/*.

Sang W, He L, Wang X-P, Zhu-Salzman K, Lei C-L. 2015. Evaluation of reference genes for RT-qPCR in *Tribolium castaneum* (Coleoptera: Tenebrionidae) under UVB Stress. *Environmental Entomology* **44**:418–425 DOI 10.1093/ee/nvv010.

Shi X-Q, Guo W-C, Wan P-J, Zhou L-T, Ren X-L, Ahmat T, Fu K-Y, Li G-Q. 2013. Validation of reference genes for expression analysis by quantitative real-time PCR in *Leptinotarsa decemlineata* (Say). *BMC Research Notes* **6**:93 DOI 10.1186/1756-0500-6-93.

Silver N, Best S, Jiang J, Thein SL. 2006. Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR. *BMC Molecular Biology* **7**:33 DOI 10.1186/1471-2199-7-33.

Smiseth PT, Kölliker M, Royle NJ. 2012. What is parental care? In: Royle NJ, Smiseth PT, Kölliker M, eds. *The evolution of parental care.* Oxford: Oxford University Press, 1–20.

Stothard P. 2000. The Sequence manipulation suite: javascript programs for analyzing and formatting protein and DNA sequences. *BioTechniques* **28**:1102.

Thellin O, Zorzi W, Lakaye B, De Borman B, Coumans B, Hennen G, Grisar T, Igout A, Heinen E. 1999. Housekeeping genes as internal standards: use and limits. *Journal of Biotechnology* **75**:291–295 DOI 10.1016/S0168-1656(99)00163-7.

Udvardi MK, Czechowski T, Scheible W-R. 2008. Eleven golden rules of quantitative RT-PCR. *The Plant Cell* **20**:1736–1737 DOI 10.1105/tpc.108.061143.

Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F. 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology* **3**:1–11 DOI 10.1186/gb-2002-3-7-research0034.

VanGuilder HD, Vrana KE, Freeman WM. 2008. Twenty-five years of quantitative PCR for gene expression analysis. *BioTechniques* **44**:619–626 DOI 10.2144/000112776.

Wang Y, Wang Z, Huang Y, Liao Y, Yin Y. 2014. Identification of suitable reference genes for gene expression studies by qRT-PCR in the blister beetle *Mylabris cichorii*. *Journal of Insect Science* **14**:1–10 DOI 10.1093/jis/14.1.94.

Xie F. 2012. RefFinder. *Available at http://leonxie.esy.es/RefFinder/* (accessed on 23 February 2017).

Xie F, Xiao P, Chen D, Xu L, Zhang B. 2012. miRDeepFinder: a miRNA analysis tool for deep sequencing of plant small RNAs. *Plant Molecular Biology* **80**:75–84 DOI 10.1007/s11103-012-9885-2.

Yang C, Pan H, Liu Y, Zhou X. 2015. Temperature and development impacts on housekeeping gene expression in cowpea aphid, *Aphis craccivora* (Hemiptera: Aphidiae). *PLOS ONE* **10**:1–10 DOI 10.1371/journal.pone.0130593.

Yu S-H, Yang P, Sun T, Qi Q, Wang X-Q, Xu D-L, Chen X-M. 2016. Identification and evaluation of reference genes in the Chinese white wax scale insect *Ericerus pela*. *SpringerPlus* **5**:791–798 DOI 10.1186/s40064-016-2548-z.

Zhou X, Liao WJ, Liao JM, Liao P, Lu H. 2015. Ribosomal proteins: functions beyond the ribosome. *Journal of Molecular Cell Biology* **7**:92–104 DOI 10.1093/jmcb/mjv014.

Zhu X, Yuan M, Shakeel M, Zhang Y, Wang S, Wang X, Zhan S, Kang T, Li J. 2014. Selection and evaluation of reference genes for expression analysis using qRT-PCR in the beet armyworm *Spodoptera exigua* (Hübner) (Lepidoptera: Noctuidae). *PLOS ONE* **9**:1–14 DOI 10.1371/journal.pone.0084730.