SUPPLEMENTARY MATERIAL

A lipocalin protein, Neural Lazarillo, is key to social interactions that promote termite soldier differentiation

Hajime Yaguchi, Shuji Shigenobu, Yoshinobu Hayashi, Satoshi Miyazaki, Kouhei Toga, Yudai Masuoka, Kiyoto Maekawa

**Supplementary Materials and Methods**

*Molecular phylogenetic analysis*

NLaz homologs in other insects were obtained from the NCBI database, and the gene copy number was confirmed by orthology data in OrthoDB [1]. NLaz homologs in *Cryptocercus punctulatus* were found from the *de novo* transcriptome data [2]; DDBJ Sequence Read Archive database accession number: DRA004598). NLaz homologs in *Blattella germanica* (accession No. PSN35678, PSN36978) [3], *Cryptotermes secundus* (Nos. XP\_023718675, XP\_023713938) [3] and *Macrotermes natalensis* (Nos. MN\_004031, MN\_003354) [4] were obtained from the genome data. The domain region in these amino acid sequences was checked using the Pfam web interface in EMBL-EBI. According to the previous phylogenetic analysis of the lipocalin family [5], we used *Drosophila melanogaster* GLaz (No. AF276506), *Schistocerca americana* axon guidance protein lazarillo (No. U15656), *Manduca sexta* insecticyanin A (No. X64715) and *Galleria mellonella* gallerin (No. L41641) as outgroups. Multiple alignments of the amino acid sequences in domain regions were performed using CLUSTAL (Gap opening penalty = 3.0, Gap extension penalty = 1.8) in MEGA version 7 [6]. We obtained estimations of tree topologies under the maximum likelihood (ML) and maximum parsimony (MP) methods. For ML, 1,000 bootstrap replicates were performed based on the most appropriate model of amino acid sequence evolution (LG+G model) determined using model selection option in MEGA7. All sites were included in the analysis, and the tree inference options were obtained by the default setting. For MP analysis, all characters were included and weighted equally, and 1,000 bootstrap replicates were performed using MEGA7. Subtree-Pruning-Regrafting algorithm with the default setting was used.

The ω [= non-synonymous (*d*N) /synonymous (*d*S) substitution rates] ratio across the following 5 monophyletic groups, (1) *Znev\_05665*, *XP\_023718675* and *MN\_004031* (termite *NLaz1*, see results), (2) termite *NLaz1* and *Comp41569\_c0\_seq1* (identified in *C. punctulatus*), (3) termite *NLaz1*, *Comp41569\_c0\_seq1* and *PSN35678* (identified in *B. germanica*), (4) *Znev\_08074*, *XP\_023713938* and *MN\_003354* (termite *NLaz2*, see results), (5) termite *NLaz2* and *PSN36978* (identified in *B. germanica*) was estimated by the PAML 4.9 package [7] using the program codeml for the branch-site model with default parameters. The likelihood ratio test was performed in the null hypothesis using parameters with ω = 1.0, fix\_ω = 1. The foreground branch indicates the branch leading to the focal clade and the background branch indicates the other branch.

*Western blotting*

The target amino acid sequences for the synthetic peptide antigen are SEVSDNTSTKPTNEK. During the immunization period, the antibody titre was evaluated by ELISA. Rabbit sera were purified using protein A. The ZnNLaz1 antibody was stored at −30°C until use. Proteins for western blotting were extracted from whole bodies, guts, and additional bodies without guts of No. 1 and No. 2 larvae (n = 5). For protein extraction, No. 1 and No. 2 larvae were collected at day 3 after their appearance in each incipient colony. Guts and bodies without guts were separated by dissection and rinsed with the phosphate-buffered saline (PBS) buffer to remove any contaminating haemolymph present. These samples were homogenized in Laemmli sample buffer (BioRad, Hercules, CA, USA) with β-mercaptoethanol (Promega Madison, WI, USA). These solutions were boiled at 95°C and centrifuged at 10,000 g for 15 min at 0°C. Supernatants were used for the sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE). Denaturing polyacrylamide gels were produced with 4.5% and 10% polyacrylamide including SDS for stacking and separating, respectively. Running buffer contained 25 mM Tris, 192 mM glycine and 0.1% SDS. Electrophoresis was performed at 150 V and 10 mA (stacking) or 150 V and 20 mA (separating). Western blotting was performed in accordance with the semi-dry transfer method [8]. Proteins in each sample were blotted onto polyvinylidene difluoride membrane in buffer containing 48 mM Tris, 39 mM glycine, and 0.0375% SDS, using a Trans-Blot SD semi-dry Transfer Cell (BioRad, Hercules, CA, USA) equipped with PowerPac HC High-Current Power Supply (BioRad) for 15 min at 25 V and 3.0 A. The ZnNLaz1 antibody was used (diluted 1:1000) in PBS with Tween-20 (PBST). The goat anti-rabbit IgG-HRP: sc-2004 (Santa Cruz) used as secondary antibody was reacted at 1:5000 in PBST. Detection of ZnNLaz1 protein was performed with TMB Liquid Substrate System with the membrane (Sigma-Aldrich, St Louis, MO, USA).

**Supplementary Figure and Table Legends**

**Figure S1.** Experimental design for RNAi experiments. Incipient colonies were carefully checked for whether the No. 1 larva (= the firstly moulted 3rd instar larva in the nest) had moulted from the 2nd instar larva. At the next day (Day 1) after the No. 1 larva appearance, DDW or siRNA solutions were injected into the side of the thorax of the No. 1 larva, and the injected larva was kept in the dish with female and male reproductives (i.e. total 3 individuals in the dish). Each dish was recorded for observation of the frequencies of trophallactic behaviours from reproductives until Day 5 after the appearance (120 min per day). Then, each dish was checked every 24 hours to confirm whether the injected No. 1 larva had moulted into a presoldier or a 4th instar.

**Figure S2.** Expression levels of the *ZnNLaz1* gene (mean ± SD, biological replicate numbers = 3) in the No. 1 and No. 2 larvae at day 0 and day 3 after their appearance as measured by quantitative real-time PCR. Relative expression levels in larval heads (A) and larval thorax and abdomen (B). GeNorm [9] and Normfinder [10] software showed that *ZnRPL-13a* (stability scores: 0.591 and 0.099, respectively) was selected as the most stable house-keeping gene compared with other 5 genes (0.648 – 1.155 and 0.214 – 0.752, respectively). Expression levels were calibrated using the mean expression level of the No. 1 larvae at day 0 after their appearance as 1.0. Different letters above the columns indicate significant differences between categories (two-way ANOVA followed by Tukey’s test, p < 0.05). By statistical analyses using two-way ANOVA, interactions were detected between developmental stages (No. 1 or No. 2 larvae) and sampling periods (day 0 or day 3 after appearance) in both body parts (head: d.f. = 1, F = 6.88, p = 3.05E-02; thorax and abdomen: d.f. = 1, F = 9.35, p = 1.56E-02). The result showed a statistical difference between No. 1 and No. 2 larvae in both body parts (head: d.f. = 1, F = 7.07, p = 2.89E-02; thorax and abdomen: d.f. = 1, F = 9.04, p = 1.69E-02). Significant differences were also found between day 0 and day 3 after appearance in both body parts (head: d.f. = 1, F = 6.91, p = 3.02E-02; thorax and abdomen: d.f. = 1, F = 8.86, p = 1.77E-02).

**Figure S3.** Molecular phylogeny of insect NLaz homologs. Molecular phylogeny of NLaz inferred from the amino acid sequences in the domain region (84–146 amino acids) based on the ML method. The tree with the highest likelihood score is shown (-ln = 5771.83). All sequences except for *Blattella germanica* [3], *Cryptocercus punctulatus* [2], *Cryptotermes secundus* [3] and *Macrotermes natalensis* [4] are obtained from Orthodb7 (ID: EOG7Z6NTK). Numericals above and below branches indicate the level of bootstrap probabilities (1000 replicates) in ML and MP methods, respectively. Asterisk indicates that a node that is not supported by the MP method. Gene ID or accession number used in the NCBI database is shown in the parenthesis after species name. Thick branches (1-5) indicate those tested for significant higher evolutionary rate with the branch-site model using PAML analysis. Only in the branch leading to the clade (2) (red line), a significantly higher ω ratio was detected (*p* = 2.10E-03) (Table S10). Likelihood ratio test was performed in the null hypothesis [lnL = -1652.834262 (1), -1653.10635 (2), -1653.210735 (3), -1655.083269 (4) and -1655.08327 (5)] and alternative hypothesis [lnL = -1652.315208 (1), -1648.372234 (2), -1653.21073 (3), -1655.083269 (4) and -1655.083269 (5)], for the branch with the clade (1) to (5) shown above, respectively.

**Figure S4.** Relative expression levels of the *ZnNLaz1* gene in the No. 1 larvae (mean ± SD, biological replicate numbers = 13) 24 hours after siRNA injection. Expression levels were calibrated using the mean expression level of *gfp*-siRNA injected individuals as 1.0. GeNorm [9] and Normfinder [10] software showed that *ZnEF-1a* (stability scores: 1.139 and 0.419, respectively) was selected as the most stable house-keeping gene compared with other 2 genes (1.225 – 1.281 and 0.623 – 0.714, respectively). Statistical test was performed using the Wilcoxon rank sum test (p < 0.05).

**Figure S5.** Western blotting analysis of ZnNLaz1 protein. The proteins extracted from whole bodies of the No. 1 and No. 2 larvae were used for the analysis (one individual was used for each lane) (A). In addition to the band observed in both lanes, a specific band was observed in the left lane (shown in arrow). The proteins extracted from guts (left) and other body parts (right) were used for the analysis (five individuals were used for each lane) (B). The No. 1-specific band was observed only in the lane with proteins extracted from guts of the No. 1 larvae.

**Table S1.** Primer sequences used in this study.

**Table S2.** The significantly differentially expressed 248 genes identified by Cufflinks between Day 0 and Day 3 in the head of the No. 1 larva (sorted by q value).

**Table S3.** The significantly differentially expressed 190 genes identified by Cufflinks between Day 0 and Day 3 in the thorax and abdomen of the No. 1 larva (sorted by q value).

**Table S4.** The significantly differentially expressed 110 genes identified by Cufflinks between the No. 1 and No. 2 larva in the head at Day 3 (sorted by q value).

**Table S5.** The significantly differentially expressed 129 genes identified by Cufflinks between the No. 1 and No. 2 larva in the thorax and abdomen at Day 3 (sorted by q value).

**Table S6.** The significantly differentially expressed 253 genes identified by Cufflinks between Day 0 and Day 3 in the head of the No. 2 larva (sorted by q value).

**Table S7.** The significantly differentially expressed 330 genes identified by Cufflinks between Day 0 and Day 3 in the thorax and abdomen of the No. 2 larva (sorted by q value).

**Table S8.** The significantly differentially expressed 52 genes identified by Cufflinks between the No. 1 and No. 2 larva in the head at Day 0 (sorted by q value).

**Table S9.** The significantly differentially expressed 53 genes identified by Cufflinks between the No. 1 and No. 2 larva in the thorax and abdomen at Day 0 (sorted by q value).

**Table S10.** The ω ratio in the alternative hypothesis predicted by the branch-site model using PAML analysis in the branch with the clade (1).

**Table S11.** The ω ratio in the alternative hypothesis predicted by the branch-site model using PAML analysis in the branch with the clade (2).

**Table S12.** The ω ratio in the alternative hypothesis predicted by the branch-site model using PAML analysis in the branch with the clade (4).

**Table S13.** The ω ratio in the alternative hypothesis predicted by the branch-site model using PAML analysis in the branch with the clade (3).

**Table S14.** The ω ratio in the alternative hypothesis predicted by the branch-site model using PAML analysis in the branch with the clade (5).

**References**

[1] Kriventseva EV, Tegenfeldt F, Petty TJ, Waterhouse RM, Simão FA, Pozdnyakov IA, Ioannidis P, Zdobnov EM. 2014 OrthoDB v8: update of the hierarchical catalog of orthologs and the underlying free software. *Nucleic Acid Res.* **43**, D250–D256. (doi:10.1093/nar/gku1220)

[2] Hayashi Y, Maekawa K, Nalepa CA, Miura T, Shigenobu S. 2017 Transcriptome sequencing and estimation of DNA methylation level in the subsocial wood-feeding cockroach *Cryptocercus punctulatus* (Blattodea: Cryptocercidae). *Appl. Entomol. Zool.* **52**, 643–651. (doi:10.1007/s13355-017-0519-7)

[3] Harrison *et al.* 2018 Hemimetabolous genomes reveal molecular basis of termite eusociality. *Nature Ecol. Evol.* **2**, 557–566. (doi:10.1038/s41559-017-0459-1)

[4] Poulsen M *et al.* 2014 Complementary symbiont contributions to plant decomposition in a fungus-farming termite. *Proc. Natl. Acad. Sci. USA* **111**, 14500-14505. (doi:10.1073/pnas.1319718111)

[5] Gutiérrez G, Ganfornina MD, Sánchez D. 2000 Evolution of the lipocalin family as inferred from a protein sequence phylogeny. *Biochim. Biophys. Acta-Protein Struct. Mol. Enzym.* **1482**, 35–45. (doi:10.1016/S0167-4838(00)00151-5)

[6] Kumar S, Stecher G, Tamura K. 2016 MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874. (doi:10.1093/molbev/msw054)

[7] Yang Z. 2007 PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591. (doi:10.1093/molbev/msm088)

[8] Kurien BT, Scofield RH. 2006 Western blotting. *Methods* **38**, 283–293. (doi:10.1016/j.ymeth.2005.11.007)

[9] Vandesompele J, Preter KD, Pattyn F, Poppe B, Roy NV, Paepe AD, Speleman F. 2002 Accurate normalization of real-time quantitative RT-PCR data by genometric averaging of multiple internal control genes. *Genome Biol.* **3**, research0034. (doi:10.1186/gb-2002-3-7-research0034)

[10] 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 Res.* **64**, 5245–5250. (doi:10.1158/0008-5472.CAN-04-0496)