SI Appendix

Exaggeration and co-option of innate immunity for social defense

Mayako Kutsukake[a,1], Minoru Moriyama[a,b], Shuji Shigenobu[c], Xian-Ying Meng[a], Naruo Nikoh[d], Chiyo Noda[e], Satoru Kobayashi[f] and Takema Fukatsu[a,g,h,1]

[a]Bioproduction Research Institute, National Institute of Advanced Science and Technology (AIST), Tsukuba 305-8566, Japan; [b]Computational Bio Big Data Open Innovation Laboratory (CBBD-OIL), AIST, Tsukuba 305-8566, Japan; [c]NIBB Core Research Facilities, National Institute for Basic Biology, Okazaki 444-8585, Japan; [d]Department of Liberal Arts, the Open University of Japan, Chiba 261-8586, Japan; [e]Okazaki Institute for Integrative Bioscience, National Institute for Basic Biology, Okazaki 444-8787, Japan; [f]Life Science Center for Survival Dynamics, Tsukuba Advanced Research Alliance (TARA), University of Tsukuba, Tsukuba 305-8577, Japan; [g]Department of Biological Sciences, Graduate School of Science, University of Tokyo, Tokyo 113-0033, Japan; [h]Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba 305-8572, Japan

[1]To whom correspondence should be addressed.
E-mail: m-kutsukake@aist.go.jp or t-fukatsu@aist.go.jp

www.pnas.org/cgi/doi/10.1073/pnas.1900917116

Contents

SI Materials and Methods

Insect Materials. Galls of *Nipponaphis monzeni* were collected from *Distylium racemosum* trees at Shinkiba, Tokyo, Japan. Insects collected from the galls were subjected to experiments immediately or preserved in an ultracold freezer until use. Galls of related aphid species, *Nipponaphis distyliicola* and *Neothoracaphis yanonis*, were also collected from *D. racemosum* trees at the same locality. The pea aphid *Acyrthosiphon pisum* strain ApL, which was used for RNA-seq analysis, was collected and established in Sapporo, Hokkaido, Japan (S1).

Collection of Discharged Body Fluid from Soldier Nymphs. Upon gall repairing, each soldier nymph excretes 0.01-0.03 μl of body fluid from its cornicles. For large scale collection, the discharged body fluid was collected and pooled in distilled water. A soldier nymph was placed in a droplet of 10 μl distilled water on a dish and tapped with the tip of a dissection needle, by which the body fluid was discharged into the water pool. This procedure was repeated with up to 75 soldier nymphs, and then the body fluid sample was transferred to a plastic tube on ice using a micropipette. Each body fluid sample for SDS-PAGE or immunoblotting was prepared in 2 μl of distilled water with 15 soldier nymphs, whereas each body fluid sample for enzymatic assay was prepared in 20 μl of 0.1 M HEPES buffer (pH 7.0) with 100 soldier nymphs. For small scale collection, a soldier nymph was placed upside down on a dish and tapped with a dissection needle, and each cornicle droplet was collected with a 0.5 μl microcapillary tube (Drummond, Microcaps). After recording the volume, the body fluid was immediately pushed out into distilled water in a plastic tube on ice using a bulb dispenser. These body fluid samples were used for experiments immediately or preserved in an ultracold freezer until use.

Protein and Enzymatic Analyses. SDS-PAGE and immunoblotting were performed as described previously (S2). The antiserum against proPO of the silkworm *Bombyx mori* was provided by Dr. Tsunaki Asano (S3). The antisera against RCP and RCP-S were prepared by immunizing rabbits with synthetic peptides (CQGSGQGSYTEHEQG for RCP and CVDRNQGGFQNNKHG for RCP-S, respectively), which were designed on the basis of amino acid sequences deduced from cDNA sequences (see *SI Appendix*, Fig. S3*A*). The antiserum against no. 3 band protein, identified as a fatty acid synthase (see *SI Appendix*, Fig. S3*F*), was prepared in the same way with a synthetic peptide SNSETDDYLSRSPDY. PO activity assay was performed as described previously with some modifications (S4). Each body fluid sample was prepared from 100 soldier nymphs in 20 μl of 0.1M HEPES buffer (pH 7.0) with or without 10 mM 1-phenyl-2-thiourea. Each whole body sample was prepared from 12-15 soldier nymphs, 12-15 middle-instar aphids, or 15-30 apterous adults of *N. monzeni*; from 40 first-instar nymphs or 15 apterous adults of *N. distyliicola*; or from 30 fourth-instar nymphs of *N. yanonis*. The insects were homogenized in 100 μl of 0.1M HEPES buffer (pH 7.0) and

centrifuged at 4°C. The supernatant was combined with a reaction buffer (0.1 M HEPES buffer [pH 7.0], 5 mM $CaCl_2$, 0.2 mM L-DOPA) with 1 μg of trypsin for PO activation in the final volume of 1 ml. The enzymatic activity was measured by continuous monitoring of OD_{490} for 60 min using a photospectrometer (Beckman Coulter DU-7000) at room temperature around 22°C, wherein one unit of PO activity was defined as 0.01 increase of OD_{490} in 60 min. In case that enzymatic activity was not detectable within 60 min, the measurement was extended up to for 240 min. Total protein concentration in each sample was quantified using Bio-Rad protein assay kit (Bio Rad).

Identification and Cloning of Protein Genes. Protein bands were excised from SDS-PAGE gels. Since N-terminal amino acid sequencing of our target proteins tended to fail probably due to chemical blocking, their internal peptide sequences were analyzed. In-gel digestion was performed using V8 protease or lysyl endopeptidase. The digested peptides were electrophoresed, blotted onto polyvinylidene difluoride membranes, and subjected to N-terminal amino acid sequencing using a protein sequencer (Applied Biosystems, Procise 494HT or 491cLC). The following internal amino acid sequences were determined: three sequences (AALADDSKE, EVAVAEED and VEVLTDNQTLKPNGN) for the no. 5 protein band (PO); two sequences (HLESGHR and HQQQEFASVTFS) for the no. 6 protein band (RCP); and two sequences (DLLSNILK and ELDYQTVFDQIA) for the no. 3 protein band (fatty acid synthase). Based on the internal amino acid sequences, degenerate DNA primers were synthesized, internal gene fragments were PCR-amplified, cloned and sequenced, and their full-length cDNA sequences were determined by the 5'- and 3'-rapid amplification of cDNA ends procedure. MALDI-TOF/MS (Shimadzu, AXIMA-CFR plus) was used for direct measurement of molecular mass of RCPs in the body fluid samples. In genomic Southern hybridization, a digoxigenin-dUTP labeled DNA probe synthesized from full-length RCP cDNA was hybridized to *Eco*RI- or *Hin*d III-digested genomic DNA of *N. monzeni.*

Histological Procedures. Close morphological observations were conducted using a scanning electron microscope (JEOL, JCM-6000 NeoScopeTM). Transmission electron microscopy was performed as described previously (S5). In situ hybridization essentially followed the procedure described previously (S6). For PO and RCP, 0.2 kb gene fragments were cloned into pBluescript SK vector, and sense and antisense probes were synthesized using T7 and T3 RNA polymerases in the presence of digoxigenin-11-UTP. Paraffin tissue sections of adult insects containing soldier embryos were subjected to hybridization at 50°C overnight, and localization of bound probes was visualized using an anti-digoxigenin secondary antibody labelled with alkaline phosphatase and NBT/BCIP (Merck) as a chromogenic substrate. Immunohistochemistry was performed essentially as described previously (S7) with some modifications. Soldier nymphs were decapitated and fixed in ethanol-formalin (3:1) at 4°C overnight, embedded in Technovit 8100 resin (Kulzer), processed into 3 μm

sections and mounted on glass slides. The tissue sections were incubated with blocking solution containing 10% normal goat serum for 1 h, and then with primary antibody solution (1:5000 dilution) containing 10% normal goat serum for 3 h. After washing, the tissue sections were incubated with HistoFine Simple stain MAX-PO(R) (Nichirei Bioscience) for 30 min, washed and incubated with HistoFine Simple stain DAB solution (Nichirei Bioscience) for 20 min to detect brownish signals, and then counterstained with hematoxylin. Immunoelectron microscopy was conducted essentially as described previously (S8). Soldier nymphs were decapitated and prefixed in 4% paraformaldehyde and 0.5% glutaraldehyde, postfixed in 1% osmium tetroxide, embedded in Lowicryl HM20 resin (Polysciences), processed into 100 nm ultrathin sections, and mounted on nickel grids. The ultrathin sections were incubated with blocking solution containing 1% bovine serum albumin (BSA) for 10 min, and then with primary antibody solution (1:1500 dilution for PO or 1:300 dilution for RCP, respectively, with 1% BSA) for 90 min. After washing, the ultrathin sections were incubated with secondary antibody solution (1:50 dilution of 20 nm colloidal gold-conjugated anti-rabbit antibody) for 60 min, re-fixed with 0.5% glutaraldehyde, stained with 2% uranyl acetate and lead citrate, and observed under a transmission electron microscope. Histochemical detection of polysaccharides by periodic acid and Schiff's reagent was performed on paraffin tissue sections as described previously (S9). Lipid detection on frozen tissue sections with oil red O was conducted as described previously (S10).

Amino Acid Analysis. Amino acids were analyzed by liquid chromatography and mass spectrometry systems (LC/MS) as described previously (S11). The discharged body fluid samples and extracted hemolymph samples were dissolved in 80% methanol solution and stored at –80°C until use. Amino acids and related amines, including L-DOPA and dopamine, were purified by passing through a solid-phase extraction column (GL-Tip SDB, GL-Science) and derivatized with propyl-chloroformate (S12). Their quantification was performed using a Prominence LC system (Shimadzu) coupled with an LCQ Duo MS (Thermo Scientific) or an H-class LC system coupled with a TQ-S micro MS (Waters). The derivatized amino acids were separated using a reverse-phase C18 column (FC-ODS, 2 mm i.d. x 150 mm, Shimadzu) under a gradient condition of water containing 0.05% formic acid and 2.5 mM ammonium formate and methanol at a flow rate of 0.2 ml/min, and their protonated ions were monitored in the Electro-Spray Ionization (ESI) positive mode. Homophenylalanine was used as an internal standard.

Lipid Analysis. The global lipid profiles of *N. monzeni* secretion were analyzed using an LC/MS as described (S13). The lipid fraction was extracted with chloroform and diluted by 20 times in 2-propanol. The diluted samples were separated using a C18 column (FC-ODS, 2 mm i.d. x 150 mm, Shimadzu) in a Prominence LC system (Shimadzu) coupled with an LCQ Fleet MS (Thermo

Scientific), under a gradient condition of water:acetonitrile (4:6) containing 10 mM ammonium formate and acetonitrile:2-propanol containing 10 mM ammonium formate (1:9) at a flow rate of 0.2 ml/min. In ESI positive mode, triglycerides were detected as adduct ions [M + NH4]+. The detected triglyceride ions were further subjected to MS/MS fragment analyses to identify acyl-chain length and the number of saturation. The composition ratio of each triglyceride was estimated based on the relative peak intensity of the parent ion. The fatty acid composition of the lipid fraction was quantified by using a 2-nitrophenylhydradine derivatization method (S14) after the free fatty acids were liberated by alkaline hydrolysis. The derivatized fatty acids were analyzed using the above-mentioned LC/MS system with the same C18 column under a gradient condition of 5 mM ammonium formate and methanol at a flow rate of 0.2 ml/min. Quantification was performed based on absorbance at 390 nm, while mass information obtained by the MS was also used for fatty acid identification. Undecanoic acid (C11:0) was used as an internal control.

Production of Recombinant Proteins. Recombinant PO protein was produced using the Sf9 insect culture cells and the Bac-to-Bac Baculovirus Expression System (Thermo Fisher Scientific). The PO cDNA containing the full-length open reading frame with 6 x histidine tag at the N-terminus was ligated to the pFastBac1 vector, by which an expression plasmid was constructed. Transfection was performed according to the manufacture's protocol. A high-titer recombinant baculovirus solution was added to Sf9 cells in Sf-900 II SFM medium containing 0.25 mM $CuSO_4$, and incubated at 28°C for 68 h. Then, the cells were collected and homogenized in 10 mM Tris-HCl (pH7.5) supplemented with the protease inhibitor cocktail cOmplete Mini (Roche). For purification, the recombinant PO protein was bound to His Mag Sepharose Ni beads (GE healthcare), washed and eluted in an elution buffer (20 mM sodium phosphate [pH 7.4], 500 mM NaCl, 200 mM imidazole). Recombinant RCP protein was produced using *Eschericha coli* and the pET system. The RCP cDNA encoding a polypeptide of 354 amino acid residues, which contains 31 repeat motifs in the middle region (see Fig. 2F), was ligated to the pET15b vector, and the plasmid was transformed into the BL21(DE3)pLysS *E. coli* strain. Induction of the recombinant RCP protein was performed by an addition of 1 mM isopropyl-β-D-thiogalactopyranoside to the bacterial culture in the log growth phase. The recombinant RCP protein was purified using the GraviTrap affinity column (GE healthcare) in the elution buffer. Desalting, buffer exchange and concentration of the recombinant proteins were performed using Amicon Ultra centrifugal filter devices (Millipore). Since the recombinant RCP protein was highly adhesive to plastic tubes, a detergent-supplemented buffer (50 mM sodium phosphate buffer (pH7.0), 0.001% Tween 80) was used. The purified recombinant proteins were quantified by 2-D Quant kit (GE healthcare) using BSA as a standard protein. On native-PAGE gels, PO activity staining was performed at 4°C by incubating in 25 mM potassium phosphate buffer (pH6.3) containing an excess amount of L-DOPA and 42% 2-propanol for proPO

activation (S15).

In Vitro Clotting Assay. Concentrations of PO and RCP proteins in the discharged body fluid of soldier nymphs were evaluated densitometrically on the basis of band intensity of SDS-PAGE gels in comparison with each recombinant protein of known concentration. For in vitro clotting assay, a standard 6 μl reaction mixture consisted of 4 μl of 2.3 μg/μl recombinant PO, 0.3 μl of 9 μg/μl recombinant RCP, 0.5 μl of 40 mM tyrosine disodium in 50 mM sodium phosphate (pH7.0), and 1.2 μl of 2-propanol, while some of the components might be either removed, replaced, increased or decreased for experimental purposes (see Fig. 5). The reaction mixtures were incubated at 25°C for 60 min and photographed. For electrophoretic crosslinking assay, the reaction mixtures were added with 6 μl of 2 x SDS-PAGE lysis buffer to stop the reaction at different time intervals, boiled for 2 min, and then subjected to SDS-PAGE.

Transcriptomics and Genomics. Total RNA was extracted from insects collected from a single gall of *N. monzeni*. LGCs sampled from 50 soldier nymphs were pooled in RNAlater (TaKaRa) and subjected to RNA extraction. Embryos were dissected from adult insects in cold phosphate buffered saline containing 0.1 % Tween 20 and immediately transferred to RNAiso reagent (TaKaRa). The remaining parts of the body (adult carcasses) were also subjected to RNA preparation. Six types of samples, namely LGCs, whole bodies of soldier nymphs, whole bodies of middle-instar insects, whole bodies of adults, dissected soldier embryos, and adult carcasses, were homogenized in 1 ml of RNAiso reagent, mixed with 0.2 ml chloroform, and centrifuged. About 0.5 ml of the supernatant and an equal volume of 70 % ethanol were mixed and transferred to RNeasy column supplied in RNeasy Mini kit (QIAGEN), and subsequent procedures were performed according to the manufacture's protocol. For bacteriocytes, bacteriome from 50 soldier nymphs or 50 adult insects were dissected in cold phosphate buffered saline containing 0.1 % Tween 20, pooled in RNAiso reagent, and total RNA was extracted. For obtaining tissue-specific RNA-seq data, total RNA was extracted from hemolymph (hemocytes), fat body, gut and whole body of *A. pisum* ApL strain. Legs of about 200 adult insects were cut by a razor, and exuding hemolymph droplets were collected by 0.5 μl microcapillary glass tubes (Drummond Scientific) and immediately transferred to a plastic tube on ice containing RNAiso reagent. Fat bodies and guts were dissected from 50 and 20 adult insects, respectively, and subjected to RNA preparation. Each cDNA library was prepared from 0.5 to 1 μg of total RNA with TruSeq™ RNA Sample Preperation Kit v2 (Illumina) according to the manufacture's protocol. For RNAseq of *Buchnera* genes, rRNA was removed from total RNA using Ribo-Zero Gold rRNA Removal Kit (Epidemiology) (Illumina), and then the remaining RNA was subjected to cDNA library preparation. The quality of the libraries was inspected by an Agilent Bioanalyzer 2100 (Agilent Technologies). Paired-end or single-end sequence for each read was analyzed by Illumina HiSeq 1500/2000/2500/X

Ten. These RNA-seq data are summarized in Table S1 with accession numbers. For genome sequencing of *Buchnera* Nmo, bacteriocytes collected from 15 aphids derived from a single gall were subjected to genomic DNA extraction using QIAamp DNA Mini kit (Qiagen). The genomic DNA was sheared using Covaris S2 ultrasonication system (Covaris), and then the DNA fragments, which was about 350 bp in size, were collected by Pippin Prep (Sage Science). The DNA library was prepared from 100 ng genomic DNA with TruSeq™ Nano DNA Sample Prep Kit (Illumina). Paired-end sequence for each read was analyzed by Illumina MiSeq.

Bioinformatics. De novo assembling of raw reads of *N. monzeni* (rep1) was performed using the program Trinity program (S16). The assembled contigs were revised by replacing the incomplete short contigs of the RCP and RCP-S genes to the full-length cDNA sequences determined by molecular cloning procedures as described. Then, the reads were mapped to the revised contigs using the CLC Genomics Workbench software (Qiagen). The transcript expression levels were estimated by calculating transcripts per kilobase million (TPM) values. Genes were annotated by BLASTx searches (e-value threshold at 1E-5). Differentially expressed genes among *N. monzeni* libraries were statistically estimated by the EdgeR (S17). In this study, genes whose TPM value in the LGC library was at least twice larger than those of the other 5 libraries were regarded as "LGC-dominant genes". Gene Ontology (GO) terms enriched in LGC-dominant genes were analyzed using the GOseq application (S18). Genome and gene ID packages for fruit fly (org.Dm.eg.db) were used to link the aphid genes to the GO categories. Principle component analysis was performed using the software R 3.5.1 (S19). Cluster analysis was performed using the program Heatplus (S20). In tissue-specific RNAseq of *A. pisum*, raw reads were mapped to the sequences of the Official Gene Consensus Set ACYPI mRNA v2.1b downloaded from AphidBase (http://www.aphidbase.com/aphidbase/). Hemocyte- and fat body-dominant genes were defined as genes whose TPM value in each tissue library was at least twice larger than that of the whole body library. De novo assembly of whole genome sequence of *Buchnera* Nmo was performed using the program Velvet (S21). The sequence was manually corrected after remapping of raw reads to the assembled genome sequence. Genes were annotated using the MiGAP pipeline (S22). Gene expression levels were estimated by mapping RNAseq reads to the assembled genome sequence by using the CLC Genomics Workbench software. The accession numbers for the genomic sequences are indicated in Figure 7.

Molecular Phylogenetic and Evolutionary Analyses. Multiple alignments of amino acid sequences were generated using the program MUSCLE implemented in the software MEGA7 (S23). Model selection and maximum-likelihood phylogenies were also constructed using the MEGA7. Relative rate tests were performed using the program RRTree (S24).

SI References

S1. T. Kanbe, S. Akimoto, Allelic and genotypic diversity in long-term asexual populations of the pea aphid, *Acyrthosiphon pisum* in comparison with sexual populations. *Mol. Ecol.* **18**, 801-816 (2009).

S2. T. Fukatsu, Acetone preservation: a practical technique for molecular analysis. *Mol. Ecol.* **8**, 1935-1945 (1999).

S3. T. Asano, K. Takebuchi, Identification of the gene encoding pro-phenoloxidase A3 in the fruitfly, *Drosophila melanogaster*. *Insect Mol. Biol.* **18**, 223-232 (2009).

S4. M. S. Zufelato, A. P. Lourenco, Z. L. Simoes, J. A. Jorge, M. M. Bitondi, Phenoloxidase activity in *Apis mellifera* honey bee pupae, and ecdysteroid-dependent expression of the prophenoloxidase mRNA. *Insect Biochem. Mol. Biol.* **34**, 1257-1268 (2004).

S5. M. Kutsukake, X. Y. Meng, N. Katayama, N. Nikoh, H. Shibao, T. Fukatsu, An insect-induced novel plant phenotype for sustaining social life in a closed system. *Nat. Commun.* **3**, 1187 (2012).

S6. M. Kutsukake, H. Shibao, N. Nikoh, M. Morioka, T. Tamura, T. Hoshino, S. Ohgiya, T. Fukatsu, Venomous protease of aphid soldier for colony defense. *Proc. Natl. Acad. Sci. USA*. **101**, 11338-11343 (2004).

S7. H. Salem, E. Bauer, R. Kirsch, A. Berasategu, M. Cripps, B. Weiss, R. Koga, K. Fukumori, H. Vogel, T. Fukatsu, M. Kaltenpoth, Drastic genome reduction in an herbivore's pectinolytic symbiont. *Cell* **171**, 1520-1531 (2017).

S8. R. Amikura, K. Hanyu, M. Kashikawa, S. Kobayashi, Tudor protein is essential for the localization of mitochondrial RNAs in polar granules of *Drosophila* embryos. *Mech. Dev*. **107**, 97-104 (2001).

S9. T. Nishino, M. Tanahashi, C. P. Lin, R. Koga, T. Fukatsu, Fungal and bacterial endosymbionts of eared leafhoppers of the subfamily Ledrinae (Hemiptera: Cicadellidae). *Appl. Entomol. Zool.* **51**, 465-477 (2016).

S10. K. Uematsu, M. Kutsukake, T. Fukatsu, M. Shimada, H. Shibao, Altruistic colony defense by menopausal female insects. *Curr. Biol*. **20**, 1182-1186 (2010).

S11. H. Anbutsu, M. Moriyama, N. Nikoh, T. Hosokawa, R. Futahashi, M. Tanahashi, X. Y. Meng, T. Kuriwada, N. Mori, K. Oshima, M. Hattori, M. Fujie, N. Satoh, T. Maeda, S. Shigenobu, R. Koga, T. Fukatsu, Small genome symbiont underlies cuticle hardness in beetles. *Proc. Natl. Acad. Sci. USA* **114**, E8382-E8391 (2017).

S12. P. Uutela, R. A. Ketola, P. Piepponen, R. Kostiainen, Comparison of different amino acid derivatives and analysis of rat brain microdialysates by liquid chromatography tandem mass spectrometry. *Anal. Chim. Acta* **633**, 223-231 (2009).

S13. K. Ikeda, Y. Oike, T. Shimizu, R. Taguchi, Global analysis of triacylglycerols including oxidized molecular species by reverse-phase high resolution LC/ESI-QTOF MS/MS. *J. Chromatogr. B:*

Anal. Technol. Biomed. Life Sci. **877**, 2639–2647 (2009).

S14. H. Miwa, M. Yamamoto, Liquid chromatographic determination of free and total fatty acids in milk and milk products as their 2-nitrophenylhydrazides. *J. Chromatogr. A* **523**, 235–246 (1990).

S15. N. Asada, T. Fukumitsu, K. Fujimoto, K. Masuda, Activation of prophenoloxidase with 2-propanol and other organic compounds in *Drosophila melanogaster*. *J. Exp. Zool.* **23**, 515-520 (1993).

S16. M. G. Grabherr, B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. D. Zeng, Z. H. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B. W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, A. Regev, Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644-652 (2011).

S17. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140 (2010).

S18. M. D. Young, M. J. Wakefield, G. K. Smyth, A. Oshlack, Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol*. **11**, R14 (2010).

S19. R Core Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/ (2018).

S20. A. Ploner, Heatplus: Heatmaps with row and/or column covariates and colored clusters. R package version 2.26.0. https://github.com/alexploner/Heatplus (2015).

S21. D. R. Zerbino, E. Birney, Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res*. **18**, 821-829 (2008).

S22. H. Sugawara, A. Ohyama, H. Mori, K. Kurokawa, "Microbial Genome Annotation Pipeline (MiGAP) for diverse users". The 20th International Conference on Genome Informatics (GIW2009) Poster and Software Demonstrations (Yokohama), S001-1-2 (2009).

S23. S. Kumar, G. Stecher, K. Tamura, MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870-1874 (2016).

S24. M. Robinson-Rechavi, D. Huchon, RRTree: relative-rate tests between groups of sequences on a phylogenetic tree. *Bioinformatics* **16**, 296-297 (2000).

SI Figure Legends

Fig. S1. Structure and evolution of PO genes and proteins of *N. monzeni*. (*A*) Full-length cDNA sequence of the soldier-specific PO gene. Estimated open reading frame and amino acid sequence are also shown. Putative copper-binding histidine residues are shaded. (*B*) Molecular phylogeny of PO genes of *N. monzeni* and other insects. A maximum-likelihood phylogeny inferred from 641 unambiguously aligned amino acid sites is shown. Bootstrap probabilities no less than 50% are indicated on each node. In brackets are sequence accession numbers. (*C*) Relative-rate tests for

comparing the molecular evolutionary rates of amino acid sequences inferred from the soldier-specific and non-specific PO genes of *N. monzeni*.

Fig. S2. Gene expression levels in *N. monzeni* and *A. pisum*. (*A*-*R*) Gene expression levels in terms of TPM values in *N. monzeni*. Abbreviations: LGC, large globular cell; E, embryo; S, soldier; M, middle-instar insect; A, adult; AC, adult carcass from which ovaries containing embryos were dissected and removed. (*A*) Soldier-specific PO [LC436906]. (*B*) PO-NmA [IAEA01000001]. (*C*) PO-NmB [IAEA01000002]. (*D*) PO-NmC [IAEA01000003]. (*E*) PO-NmD [IAEA01000004]. For phylogenetic relationship of the PO genes, see *SI Appendix*, Figure S1*B*. (*F*) RCP [LC436900-LC436905]. (*G*) RCP-S [LC436897-LC436899]. For structural relationship of RCP to RCP-S, see *SI Appendix*, Figure S3*A*. (*H*) Fatty acid synthase [IAEA01000005]. (*I*) Dihydroxyphenylalanine decarboxylase (DDC) [IAEA01000006]. (*J*) Dopachrome conversion enzyme (DCE) [IAEA01000007]. For metabolic relationship of PO, DDC and DCE in the melanin synthesis pathway, see *SI Appendix*, Figure 4*A*. (*K*) Serine protease [IAEA01000008]. (*L*) Serpin [IAEA01000009]. (*M*) Phenylalanine-4-monooxigenase [IAEA01000010]. (*N*) Transaldolase [IAEA01000011]. (*O*) Phosphoenol-pyruvate carboxykinase (PEPCK) [IAEA01000012]. For metabolic relationship of phenylalanine-4-monooxygenase, transaldolase and PEPCK, see Figure 7*B*. (*P*) Transglutaminae-Nm1 [IAEA01000013]. (*Q*) Transglutaminase-Nm2 [IAEA01000014]. (*R*) Hemolectin [IAEA01000015]. (*S*-*Y*) Gene expression levels in terms of TPM values in *A. pisum*. Abbreviations: HC, hemocyte; FB, fat body; GU, gut; WB, whole body. (*S*) PO [ACYPI004484]. (*T*) PO [ACYPI001367]. (*U*) RCP-like [ACYPI086030]. (*V*) RCP-like [ACYPI005732]. (*W*) RCP-like [ACYPI081606]. (*X*) RCP-like [ACYPI088785]. (*Y*) RCP-like [ACYPI069330]. Different alphabetical characters (a-d) indicate statistically significant differences (Tukey HSD test: $P < 0.05$).

Fig. S3. Genes and proteins of RCP, RCP-S and fatty acid synthase of *N. monzeni*. (*A*) Structural features of RCP gene (top) and RCP-S gene (bottom). Homologous regions between RCP and RCP-S are indicated by dotted lines. Peptide sequences and regions used for raising antibodies are shown by bars. S and R indicate the N-terminal signal peptide region and the repeat-containing middle region, respectively. R(31-43) means, for example, the repeat region consists of 31-43 consecutive repeat units, each of which is 8 amino acid residues and classified to either of 7 sequence types. For details, see Figure 2*F*. (*B*) Southern blotting of RCP gene. (*C*) MALDI-TOF/MS analysis of RCP proteins in body fluid samples representing the gall colonies a-h (for SDS-PAGE analysis, see Fig. 2*D*). The number above each peak indicates molecular mass of the protein. (*D*) SDS-PAGE of soldier's body fluid proteins and immunoblotting against RCP-S. Asterisk indicates a non-specific signal. (*E*) Attributes and sequence accession numbers for the RCP-S alleles identified from the gall colonies i-m (see Fig. 2*E* and *G*) and additional gall colonies n and o. (*F*) SDS-PAGE of soldier's body fluid

proteins and immunoblotting against the no. 3 band protein identified as a fatty acid synthase. Three gall colonies (p-r) are analyzed. Asterisk indicates a non-specific signal.

Fig. S4. Control experimental data for detection and expression of PO and RCP. (*A*) In situ hybridization of PO gene expression using a sense probe, corresponding to Figure 3*A*. (*B*) In situ hybridization of RCP gene expression using a sense probe, corresponding to Figure 3*B*. (*C*) Immunohistochemistry of PO using a pre-immune serum, corresponding to Figure 3*C*. (*D*) Immunoelectron microscopy of PO using a pre-immune serum, corresponding to Figure 3*H*. (*E*) Immunoelectron microscopy of RCP using a pre-immune serum, corresponding to Figure 3*I*. (*F*) Activity of recombinant PO analyzed on native PAGE gels. (*G*) PO activity in soldier's body fluid analyzed on native PAGE gels. (*H*) Production of recombinant PO analyzed by SDS-PAGE and immunoblotting. (*I*) Production of recombinant RCP analyzed by SDS-PAGE and immunoblotting.

Fig. S5. Lipids in soldier's body fluid of *N. monzeni*. (*A*, *B*) Triglycerides in soldier's secretion before and after solidification (black and gray bars), and in the whole body of soldier nymphs (white bars). (*A*) Composition of major triglycerides. (*B*) Composition of minor triglycerides. (*C*) Fatty acid composition generated by alkaline hydrolysis of the secretion. Free fatty acids were excluded from the result. Abbreviations: C6:0, hexanoic acid; C6:2, sorbic acid, C10:0, capric acid; C12:0, lauric acid; C14:0, myristic acid; C16:0, palmitic acid; C18:0, stearic acid.

Fig. S6. Comparison of gene expression profiles of LGC with those of other tissues of *N. monzeni* and *A. pisum*. (*A*) Principal component analysis based on expression levels of 9,825 genes that exhibit 1:1 orthologous relationship between *N. monzeni* and *A. pisum*. Red shows *N. monzeni*: LGC, large globular cell; E, embryo; S, soldier; M, middle-instar insect; A, adult; AC, adult carcass from which ovaries containing embryos were dissected and removed. Blue shows adult *A. pisum*: HC, hemocyte; FB, fat body; GU, gut; WB, whole body. (*B*) Cluster analysis based on the same data set, in which a dendrogram and a heat map showing relative gene expression levels are depicted.

Fig. S7. Expression levels of genes related to tyrosine synthesis in bacteriocytes, LGCs and whole body of *N. monzeni*. (*A*) Phenylalanine-4-monooxygenase [IAEA01000010]. Note its high and bacteriocyte-specific expression. (*B*) Aspartate aminotransferase-Nm1 [IAEA01000016]. (*C*) Aspartate aminotransferase-Nm2 [IAEA01000017]. (*D*) Aspartate aminotransferase-Nm3 [IAEA01000018]. Note that expression levels of these aspartate transaminase genes are incomparably low in comparison with phenylalanine-4-monooxygenase. (*E*) Transaldolase [IAEA01000011]. Note its high and preferential expression in bacteriocytes and LGCs. (*F*) Phosphoenol-pyruvate carboxykinase (PEPCK) [IAEA01000012]. Note its high and preferential expression in LGCs. (*G*)

Enolase [IAEA01000019]. No preferential expression was observed in bacteriocytes and LGCs. BAC, bacteriocyte; WBA, whole body of adult; LGC, large globular cell; WBS, whole body of soldier. Note that LGCs are specifically found in soldier nymphs. To avoid contamination of LGCs, the bacteriocyte samples were dissected from adult insects. Asterisks indicate statistically significant differences (GLM: *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$).

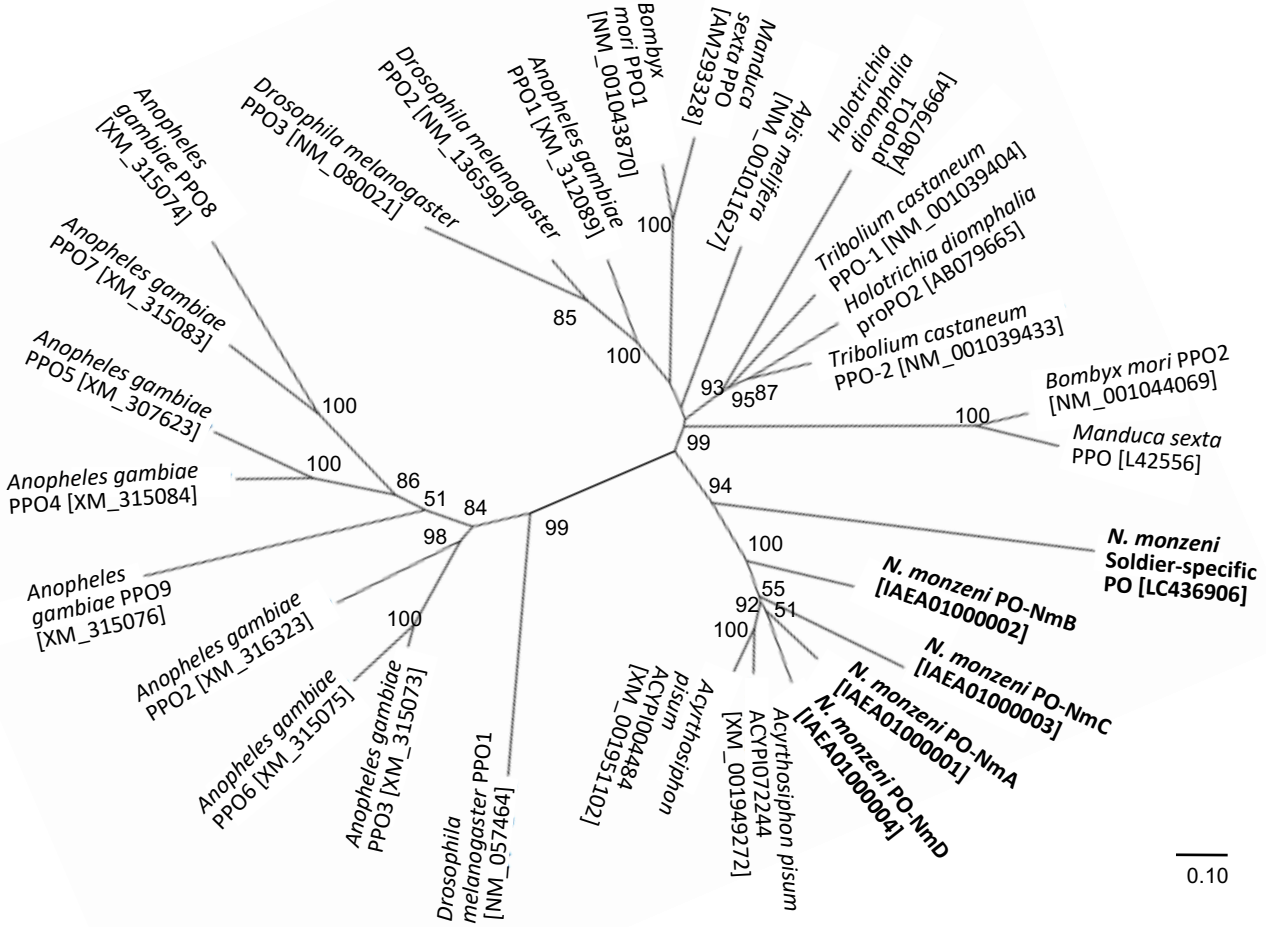
SI Movie Legends

Movie S1. Soldier nymphs of *N. monzeni* repairing their gall.

Movie S2. A soldier nymph of *N. monzeni* artificially stimulated to discharge body fluid with numerous large globular cells.

A

B

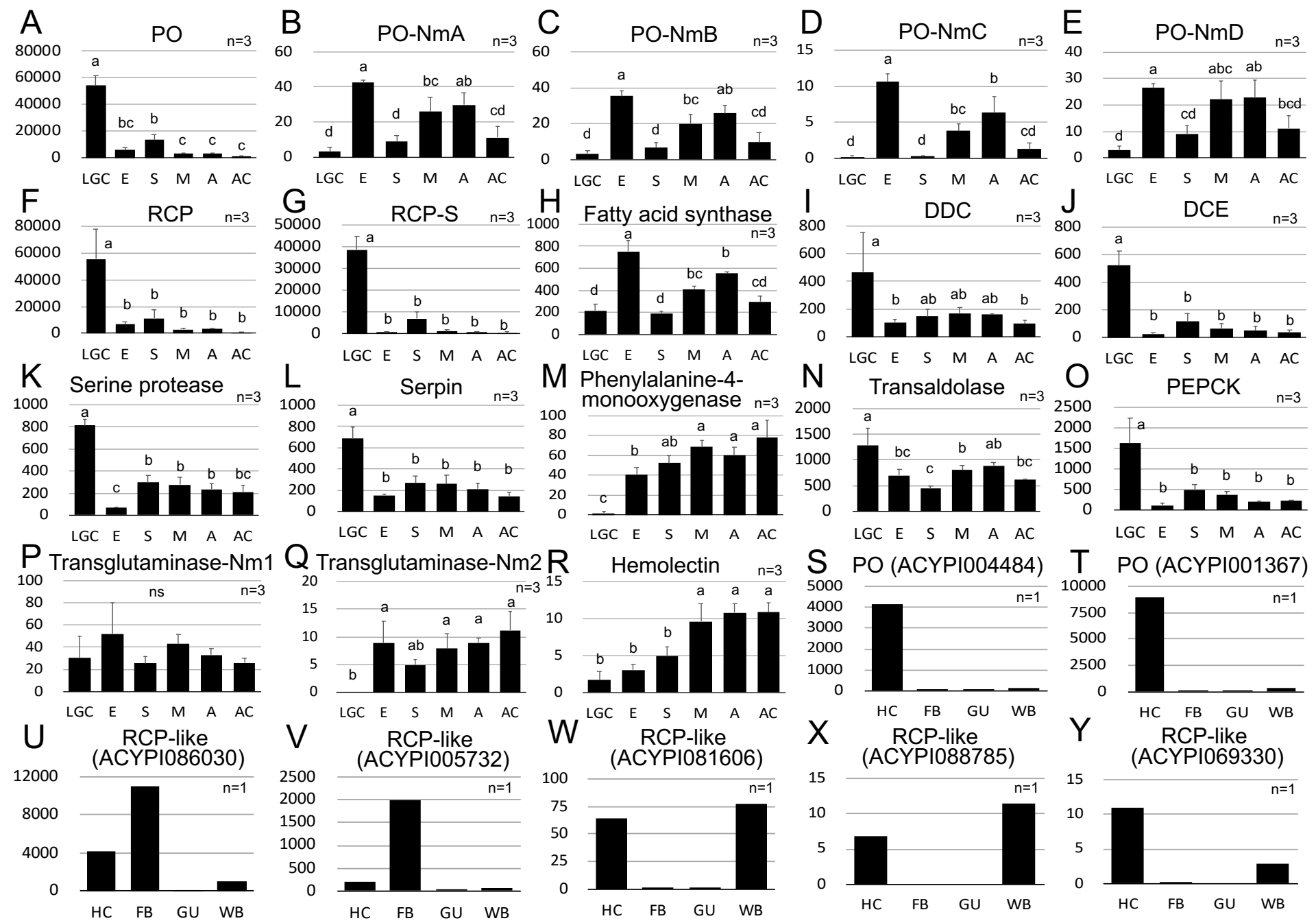


C

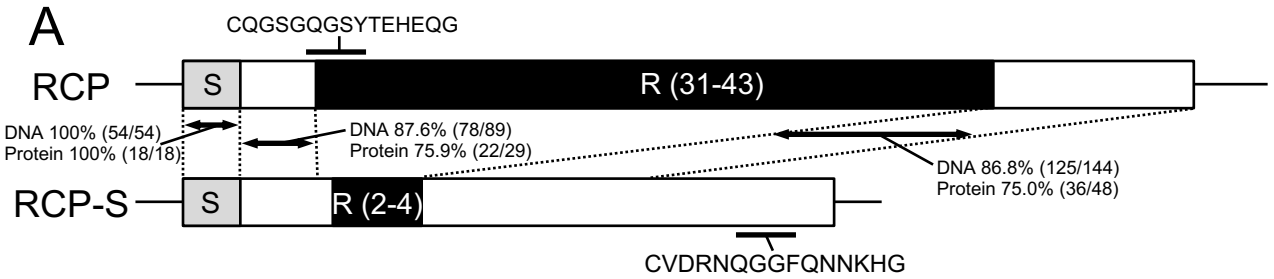
Lineage 1	Lineage 2	Outgroup[a]	K1[b]	K2[c]	K1-K2	K1/K2	*P*-value[d]
Soldier-specific PO [LC436906]	PO-NmA [IAEA01000001]	POs of bean bug [BAN20821] & whitefly [AYA71581]	0.449	0.200	0.249	2.25	1.2×10^{-7}
Soldier-specific PO [LC436906]	PO-NmB [IAEA01000002]	POs of bean bug [BAN20821] & whitefly [AYA71581]	0.437	0.218	0.219	2.00	1.3×10^{-7}
Soldier-specific PO [LC436906]	PO-NmC [IAEA01000003]	POs of bean bug [BAN20821] & whitefly [AYA71581]	0.455	0.299	0.156	1.52	0.00029
Soldier-specific PO [LC436906]	PO-NmD [IAEA01000004]	POs of bean bug [BAN20821] & whitefly [AYA71581]	0.453	0.211	0.242	2.15	1.0×10^{-7}

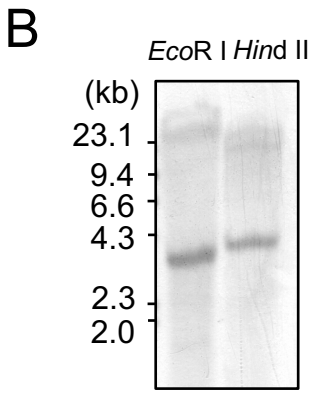
[a] Bean bug *Riptortus pedestris* and whitefly *Dialeurodes citri*.
[b] Estimated mean distance between lineage 1 and the last common ancestor of lineages 1 and 2.
[c] Estimated mean distance between lineage 2 and the last common ancestor of lineages 1 and 2.
[d] *P*-value was generated using the program RRTree (Robinson-Rechavi and Huchon, 2000).

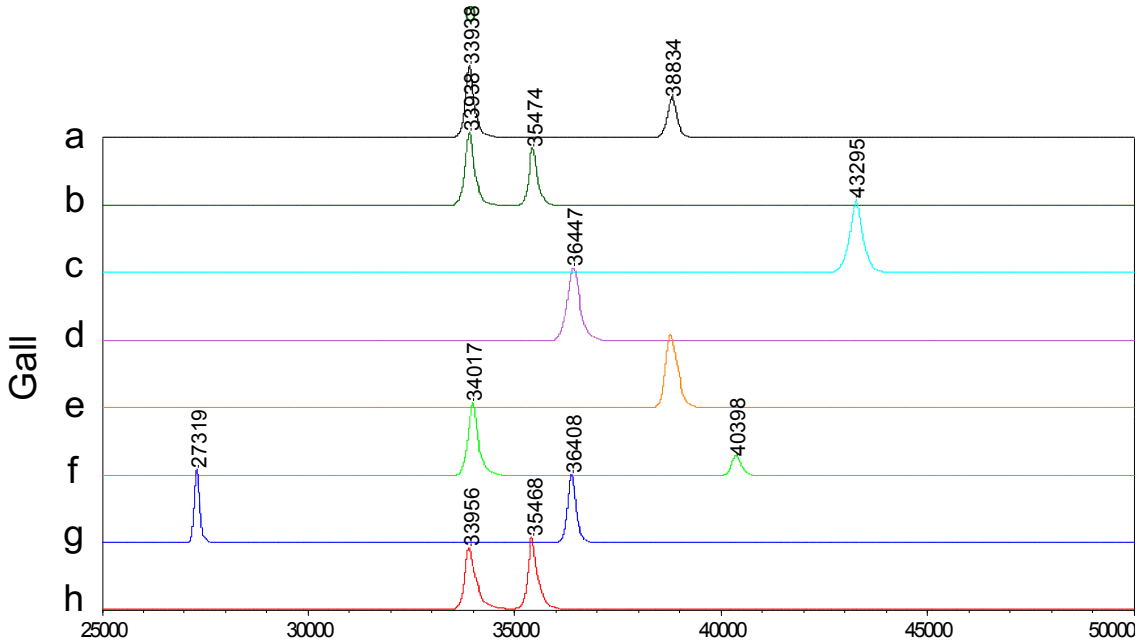
Kutsukake et al., Fig. S1

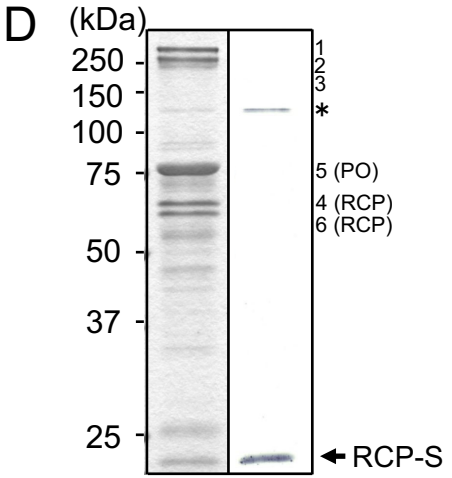


Kutsukake et al., Fig. S2



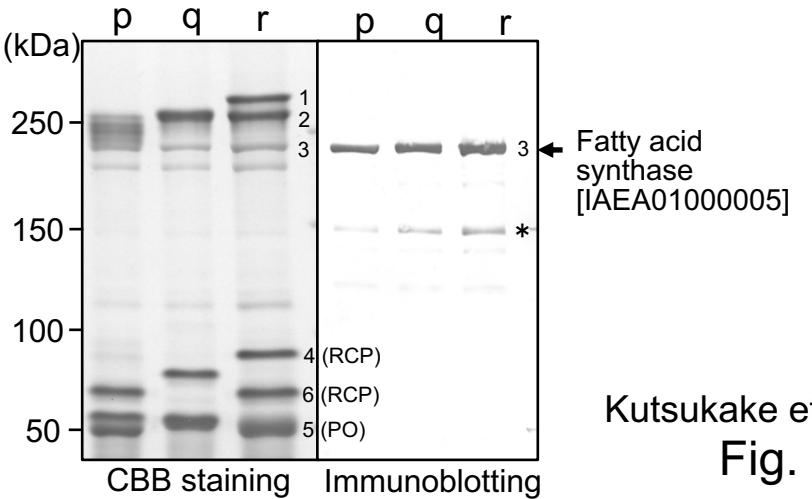






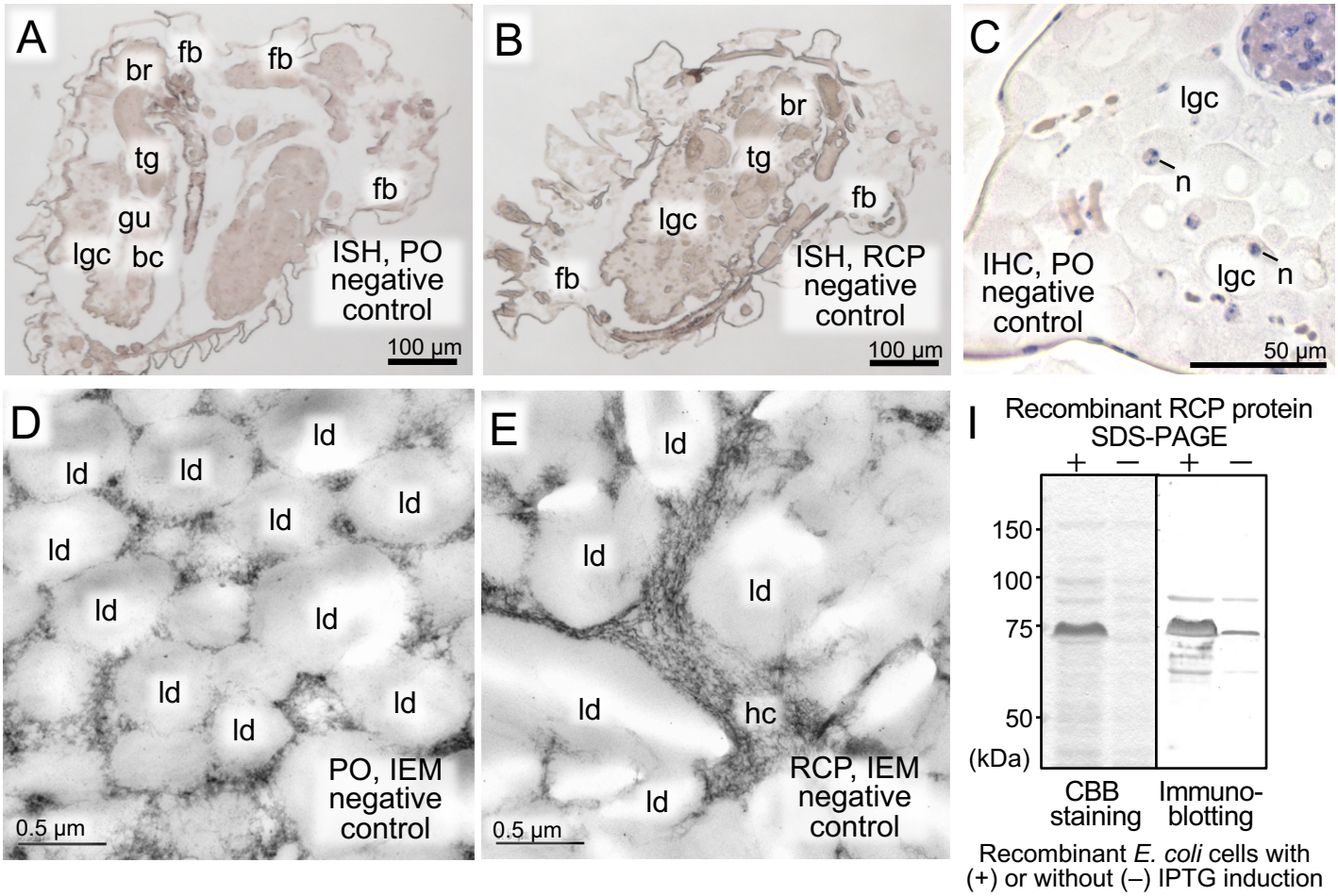
E

Galls	No. of repeats	Amino acids	Size (kDa)	Accession no.
l, m	2	217	21.0	LC436897
l, j, k, l, m, n, o	3	225	21.7	LC436898
o	4	233	22.4	LC436899

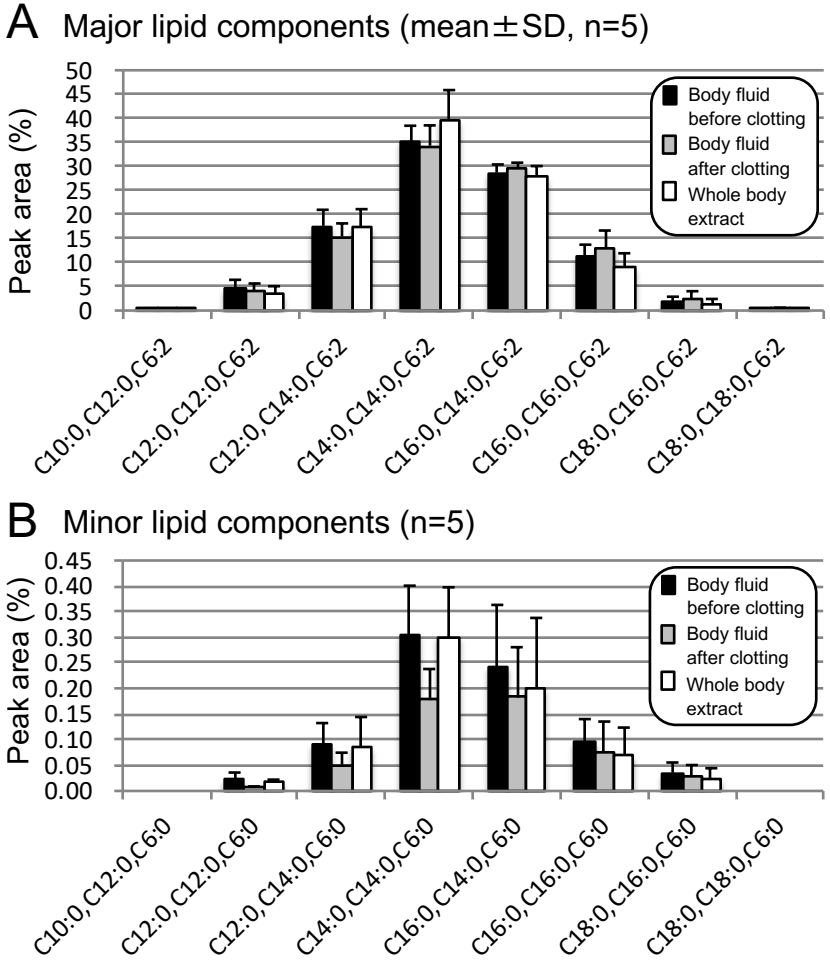


Kutsukake et al.

Fig. S3

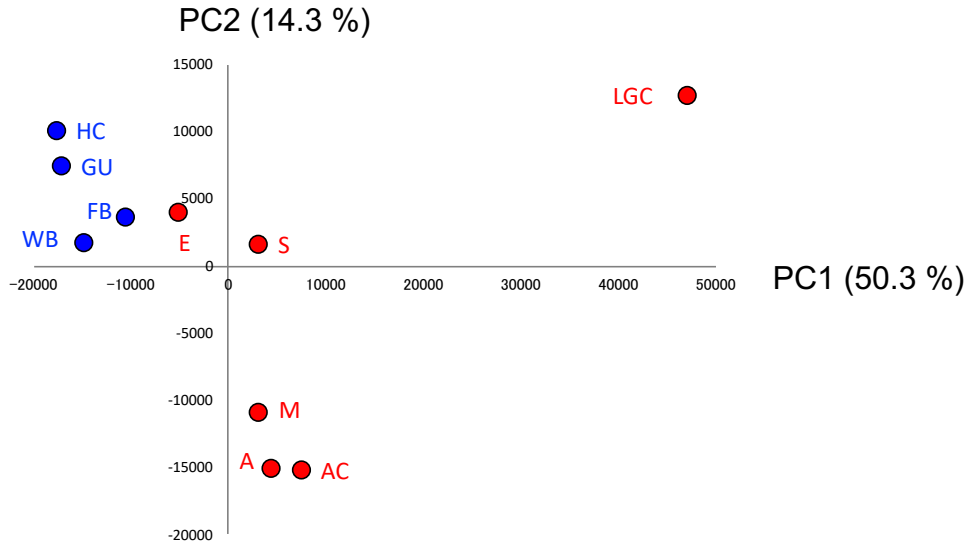


Kutsukake et al., Fig. S4

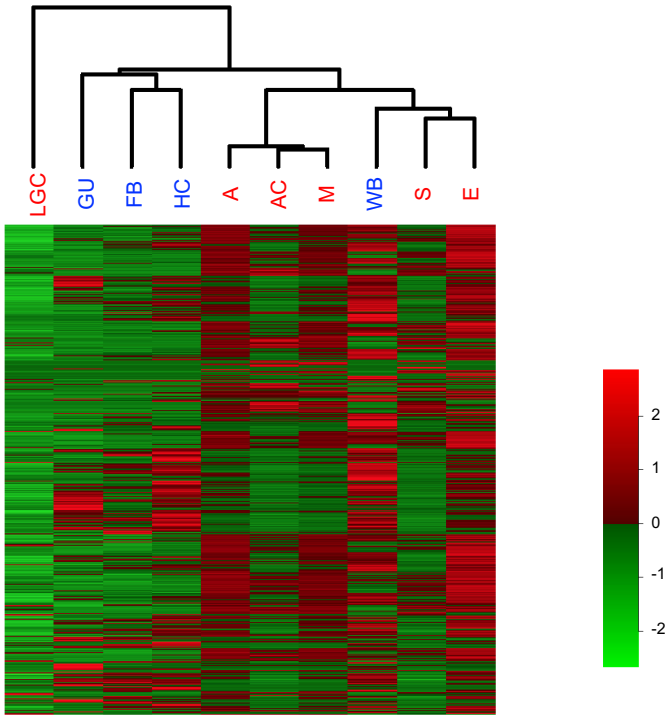


Kutsukake et al., Fig. S5

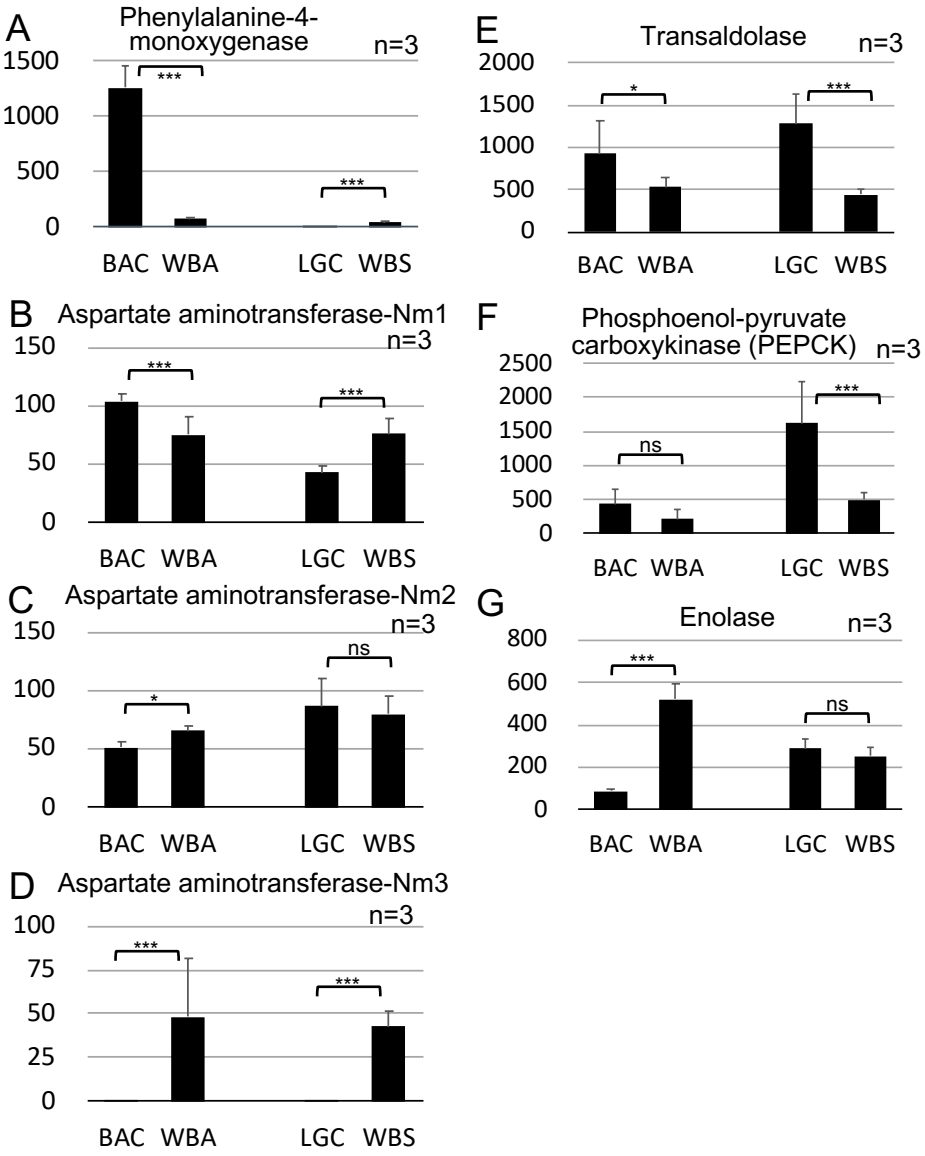
A



B



Kutsukake et al., Fig.S6



Kutsukake et al., Fig. S7

Table S1. RNA-seq data summary.

Sample	Locality and collection date	Stage/morph, tissue	Number of pairs/reads	HiSeq	Accession number
N. monzeni replicate 1	Shinkiba, Tokyo, Japan 16 May 2012	1st instar soldier nymphs, LGCs	29,202,505	100bp paired	DRA007668
		Embryos, whole body	41,666,434	100bp paired	DRA007672
		1st instar soldier nymphs, whole body	30,624,551	100bp paired	DRA007669
		Middle instar aphids, whole body	27,668,289	100bp paired	DRA007670
		Adults, whole body	28,016,416	100bp paired	DRA007671
		Adults, carcass without ovaries and embryos	40,155,928	100bp paired	DRA007673
N. monzeni replicate 2	Shinkiba, Tokyo, Japan 16 May 2012	1st instar soldier nymphs, LGCs	6,958,119	100bp paired	DRA007674
		Embryos, whole body	8,097,348	100bp paired	DRA007678
		1st instar soldier nymphs, whole body	7,710,432	100bp paired	DRA007675
		Middle instar aphids, whole body	8,657,918	100bp paired	DRA007676
		Adults, whole body	8,457,851	100bp paired	DRA007677
		Adults, carcass without ovaries and embryos	7,729,867	100bp paired	DRA007679
N. monzeni replicate 3	Shinkiba, Tokyo, Japan 16 May 2012	1st instar soldier nymphs, LGCs	8,271,511	100bp paired	DRA007680
		Embryos, whole body	8,273,314	100bp paired	DRA007684
		1st instar soldier nymphs, whole body	8,236,393	100bp paired	DRA007681
		Middle instar aphids, whole body	7,087,768	100bp paired	DRA007682
		Adults, whole body	8,400,338	100bp paired	DRA007683
		Adults, carcass without ovaries and embryos	7,380,855	100bp paired	DRA007685
N. monzeni bacteriocytes, polyA selected	Shinkiba, Tokyo, Japan 7 May 2018	Adults, bacteriocytes replicate 1	13,050,748	150bp paired	DRA007686
		Adults, bacteriocytes replicate 2	6,075,936	150bp paired	DRA007688
		Adults, bacteriocytes replicate 3	7,206,319	150bp paired	DRA007690
		Adults, whole body replicate 1	7,668,216	150bp paired	DRA007687
		Adults, whole body replicate 2	8,349,565	150bp paired	DRA007689
		Adults, whole body replicate 3	7,850,483	150bp paired	DRA007691
N. monzeni bacteriocytes, rRNA removed	Shinkiba, Tokyo, Japan 19 & 27 May 2015	1st instar soldier nymphs, bacteriocytes replicate 1	18,489,980	66bp single	DRA007692
		1st instar soldier nymphs, bacteriocytes replicate 2	25,938,326	66bp single	DRA007693
		1st instar soldier nymphs, bacteriocytes replicate 3	21,243,725	66bp single	DRA007694
Acyrthosiphon pisum laboratory strain ApL	Sapporo, Hokkaido, Japan	Adults, hemocytes	29,652,916	100bp paired	DRA007695
		Adults, fat body cells	27,522,381	100bp paired	DRA007696
		Adults, gut	21,76,1730	100bp paired	DRA007697
		Adults, whole body	23,62,4012	100bp paired	DRA007698

Table S2. LGC-dominant genes identified by RNA-seq analysis.

	Genes (species)[a]	E-value	TPM (mean, n = 3)						Ratio (LGC/S)
			LGC	E	S	M	A	AC	
1	unknown		142222	4236	22327	20402	24746	18888	6.4
2	unknown		57625	1861	11548	1992	1824	1105	5.0
3	**RCP**		55475	6681	11104	2927	3192	578	5.0
4	**Phenoloxidase (*Drosophila melanogaster*)**	0	54357	5751	13628	3230	3237	897	4.0
5	Lipid storage droplets surface-binding protein 1 (*Drosophila melanogaster*)	6.0.E-41	46538	3510	10351	3105	2910	1259	4.5
6	**RCP-S**		38690	673	6932	1200	741	502	5.6
7	Lipid storage droplets surface-binding protein 2 (*Drosophila melanogaster*)	4.0.E-17	23537	1739	4961	1653	1427	627	4.7
8	unknown		20347	1078	5012	1704	1276	893	4.1
9	Lipase (*Xenopus laevis*)	1.0.E-37	13842	1376	3279	887	760	248	4.2
10	unknown		9879	496	1663	508	490	364	5.9
11	unknown		8387	436	2064	883	742	507	4.1
12	unknown		7540	328	1556	675	670	557	4.8
13	Esterase FE4 (*Myzus persicae*)	1.0.E-55	5236	156	1112	329	330	232	4.7
14	Dehydrogenase/reductase (*Mus musculus*)	2.0.E-51	5087	128	1264	463	326	324	4.0
15	Chitinase (*Bombyx mori*)	7.0.E-118	4835	1096	1427	1295	1202	826	3.4
16	*Kunitz-type proteinase inhibitor (Anemonia sulcata)*	1.0.E-12	4618	393	885	767	895	571	5.2
17	Uncharacterized protein (*Plasmodium falciparum*)	1.0.E-10	4007	1557	1081	1932	1847	1065	3.7
18	unknown		3431	917	880	658	876	398	3.9
19	unknown		2936	60	565	33	38	11	5.2
20	Carboxylesterase (*Mesocricetus auratus*)	2.0.E-23	2770	92	599	192	183	133	4.6
21	Malate dehydrogenase (*Caenorhabditis elegans*)	3.0.E-54	2574	796	1175	999	964	809	2.2
22	Golgi-associated plant pathogenesis-related protein (*Homo sapiens*)	1.0.E-23	2382	420	441	225	303	111	5.4
23	Megourin (*Megoura viciae*)	8.0.E-14	2243	77	454	64	63	29	4.9
24	unknown		2207	136	527	373	361	310	4.2
25	Malate dehydrogenase (*Bos taurus*)	8.0.E-75	1940	606	872	681	729	631	2.2
26	1,4-alpha-glucan-branching enzyme (*Mus musculus*)	0.0.E+00	1921	757	623	865	904	482	3.1
27	unknown		1906	428	583	659	562	325	3.3
28	Gelsolin (*Drosophila melanogaster*)	1.0.E-69	1836	366	786	725	684	609	2.3
29	Lipase (*Homo sapiens*)	7.0.E-64	1783	158	534	171	180	136	3.3
30	Gelsolin (Drosophila melanogaster)	2.0.E-133	1742	38	406	241	189	137	4.3
31	Glutamine synthetase (*Drosophila melanogaster*)	0	1727	830	836	858	811	790	2.1
32	unknown		1707	25	697	429	370	438	2.5
33	Phosphoenolpyruvate carboxykinase (*Drosophila melanogaster*)	0	1626	101	493	371	187	208	3.3
34	SPARC (*Caenorhabditis elegans*)	7.0.E-63	1561	371	518	404	439	261	3.0
35	Gelsolin (*Drosophila melanogaster*)	9.0.E-54	1441	291	571	601	499	497	2.5
36	Dehydrodolichyl diphosphate synthase (*Homo sapiens*)	3.0.E-104	1370	184	351	261	313	214	3.9
37	Cathepsin B (*Bos taurus*)	2.0.E-147	1245	390	405	472	490	317	3.1
38	unknown		1216	444	374	570	526	291	3.3
39	unknown		1204	288	401	325	331	197	3.0
40	Ornithine decarboxylase (*Rattus norvegicus*)	2.0.E-112	1160	440	547	517	530	443	2.1
41	Androgen-induced gene (*Homo sapiens*)	4.0.E-41	1075	288	240	144	188	62	4.5
42	unknown		986	262	295	440	470	297	3.3
43	FAD-dependent oxidoreductase domain-containing protein (*Xenopus laevis*)	8.0.E-117	945	111	259	218	190	128	3.6
44	2-acylglycerol O-acyltransferase (*Mus musculus*)	2.0.E-89	940	316	263	319	292	155	3.6
45	unknown		939	66	241	144	115	79	3.9

46	Myrosinase (*Brevicoryne brassicae*)	1.0.E-81	913	117	226	67	73	11	4.0
47	unknown		901	116	167	176	143	88	5.4
48	unknown		859	180	217	225	212	119	4.0
49	unknown		827	226	250	253	248	124	3.3
50	unknown		821	21	230	127	89	77	3.6
51	Serine protease (*Megabombus pennsylvanicus*)	4.0.E-80	815	72	296	275	236	212	2.8
52	Androgen-dependent TFPI-regulating protein (*Mesocricetus auratus*)	5.0.E-17	815	266	193	114	157	29	4.2
53	FAD-dependent oxidoreductase domain-containing protein (*Bos_taurus*)	3.0.E-18	772	59	182	164	115	77	4.2
54	Lambda-crystallin (Rattus norvegicus)	4.0.E-58	711	150	182	155	144	77	3.9
55	Excitatory amino acid transporter (*Ambystoma tigrinum*)	2.0.E-135	700	218	258	198	195	141	2.7
56	Serpin (*Homo sapiens*)	1.0.E-61	688	152	268	261	213	145	2.6
57	15-hydroxyprostaglandin dehydrogenase (*Rattus norvegicus*)	6.0.E-39	685	72	159	93	115	80	4.3
58	Acyl-protein thioesterase (*Homo sapiens*)	7.0.E-59	670	133	142	63	78	32	4.7
59	unknown		668	42	157	128	122	82	4.3
60	unknown		653	136	167	231	240	165	3.9
61	Uncharacterized_protein (*Acanthamoeba polyphaga mimivirus*)	9.0.E-06	642	309	127	118	141	22	5.1
62	Phytanoyl-CoA dioxygenase domain-containing protein (*Caenorhabditis elegans*)	2.0.E-78	640	105	192	245	222	129	3.3
63	unknown		561	127	232	148	134	160	2.4
64	UDP-glucuronosyltransferase (*Oryctolagus cuniculus*)	2.0.E-57	558	76	145	61	62	26	3.9
65	UDP-glucuronosyltransferase (*Homo sapiens*)	3.0.E-70	551	37	141	45	44	30	3.9
66	Acid phosphatase (*Macaca fascicularis*)	2.0.E-52	551	88	246	180	164	153	2.2
67	Tumor protein (*Homo sapiens*)	1.0.E-21	546	205	237	230	238	163	2.3
68	Alpha-tocopherol transfer protein (*Pongo abelii*)	8.0.E-18	543	30	131	66	44	33	4.2
69	Yellow-c (*Drosophila melanogaster*)	1.0.E-100	522	27	122	67	51	38	4.3
70	Short/branched chain specific acyl-CoA dehydrogenase (*Mus musculus*)	9.0.E-168	501	168	179	193	203	97	2.8
71	GMP reductase (*Homo_sapiens*)	5.0.E-180	496	141	170	246	220	142	2.9
72	N-acetyl-D-glucosamine kinase (*Mus musculus*)	2.0.E-67	482	150	186	178	216	128	2.6
73	unknown		475	14	157	14	15	6	3.0
74	Spatzle (*Drosophila melanogaster*)	7.0.E-20	474	100	135	131	143	105	3.5
75	unknown		469	171	156	141	191	142	3.0
76	Dopa decarboxylase (*Drosophila melanogaster*)	0	468	105	146	169	160	95	3.2
77	unknown		467	93	135	74	67	42	3.5
78	Xylulose kinase (*Homo sapiens*)	0	432	128	155	159	161	113	2.8
79	unknown		432	5	85	48	38	26	5.1
80	Pro-X carboxypeptidase (*Bos taurus*)	4.0.E-156	418	107	116	118	104	65	3.6
81	Innexin (*Drosophila melanogaster*)	3.0.E-91	417	101	123	108	106	69	3.4
82	Regucalcin (*Xenopus laevis*)	1.0.E-56	402	96	134	85	63	38	3.0
83	Dehydrogenase/reductase (*Gallus gallus*)	3.0.E-48	395	61	120	81	79	60	3.3
84	Riboflavin kinase (*Drosophila melanogaster*)	7.0.E-53	383	73	139	102	104	93	2.8
85	Cytochrome P450 6a14 (*Drosophila melanogaster*)	1.0.E-102	382	92	179	171	148	136	2.1
86	LIM/homeobox protein (*Gallus gallus*)	6.0.E-43	380	138	96	136	129	69	4.0
87	Lipase (*Cavia porcellus*)	1.0.E-31	380	28	173	41	29	51	2.2
88	Cathepsin B (*Bos taurus*)	1.0.E-127	377	44	112	172	159	114	3.4
89	unknown		355	33	123	96	82	85	2.9
90	Triacylglycerol lipase (*Sus scrofa*)	6.0.E-92	350	54	76	40	40	18	4.6
91	unknown		350	101	140	141	124	123	2.5
92	Delta-1-pyrroline-5-carboxylate synthase (*Mus musculus*)	0	337	101	101	95	92	49	3.3

93	L-threonine dehydratase catabolic TdcB (*Staphylococcus aureus*)	3.0.E-42	333	36	118	68	72	93	2.8
94	unknown		328	111	92	108	130	130	3.6
95	PRELI domain-containing_protein (*Mus musculus)*	4.0.E-13	325	46	78	63	74	47	4.2
96	Nogo-B receptor (*Danio rerio*)	7.0.E-15	317	88	79	63	70	44	4.0
97	GMP reductase (*Homo sapiens*)	4.0.E-24	312	85	89	143	126	83	3.5
98	Myrosinase (*Brevicoryne brassicae*)	3.0.E-69	308	38	66	15	23	7	4.7
99	Cytochrome b5-related protein (*Drosophila melanogaster*)	2.0.E-74	304	62	131	107	124	101	2.3
100	unknown		304	27	60	23	18	4	5.0
101	2-acylglycerol O-acyltransferase (*Mus musculus)*	3.0.E-12	304	91	73	86	94	48	4.2
102	*Spondin (Mus musculus)*	1.0.E-92	302	56	101	132	129	94	3.0
103	Glucose-6-phosphate 1-dehydrogenase (*Homo sapiens*)	0	299	122	115	89	107	78	2.6
104	unknown		295	66	84	95	118	85	3.5
105	UDP-glucuronosyltransferase (*Homo sapiens*)	8.0.E-80	294	43	112	113	102	92	2.6
106	unknown		289	125	112	140	137	90	2.6
107	Serpin (*Homo sapiens*)	3.0.E-62	289	79	139	112	113	105	2.1
108	Solute carrier organic anion transporter (*Mus musculus*)	4.0.E-79	289	94	110	143	113	70	2.6
109	Lysosome membrane protein (*Rattus norvegicus*)	2.0.E-52	288	100	87	93	92	66	3.3
110	Angiotensin-converting enzyme (*Gallus gallus*)	0	286	65	81	70	81	39	3.5
111	Bombyxin (*Samia cynthia*)	2.0.E-07	278	15	58	36	32	24	4.8
112	unknown		270	31	68	23	22	8	4.0
113	Lambda-crystallin (*Mus musculus*)	7.0.E-10	264	50	65	65	48	28	4.0
114	Dehydrogenase/reductase (*Mus musculus*)	1.0.E-64	259	48	63	26	23	11	4.1
115	Methylthioribose-1-phosphate isomerase (*Aedes aegypti*)	7.0.E-174	248	112	110	99	103	75	2.3
116	Sialin (*Homo sapiens*)	2.0.E-99	248	102	106	123	118	109	2.3
117	unknown		246	30	102	40	39	26	2.4
118	Neuronal acetylcholine receptor (*Rattus norvegicus*)	6.0.E-27	244	51	52	51	51	22	4.7
119	unknown		235	42	73	85	65	46	3.2
120	Vitellogenin (*Oscheius brevesophaga*)	2.0.E-24	233	35	83	77	73	46	2.8
121	Glycogen synthase (*Drosophila melanogaster*)	0	233	60	69	85	85	53	3.4
122	Trehalose transporter (*Culex quinquefasciatus*)	2.0.E-52	230	21	78	83	69	62	2.9
123	unknown		218	17	37	9	11	3	5.9
124	Myrosinase (*Brevicoryne brassicae*)	2.0.E-60	214	24	50	12	15	4	4.3
125	UDP-glucuronosyltransferase (*Rattus norvegicus*)	3.0.E-55	211	18	58	42	40	22	3.7
126	unknown		208	32	47	26	31	14	4.4
127	unknown		207	41	86	78	77	64	2.4
128	Glycine cleavage system H protein (*Drosophila melanogaster*)	4.0.E-50	202	77	77	86	93	75	2.6
129	Retinaldehyde-binding protein (*Homo sapiens*)	9.0.E-16	202	6	42	19	11	9	4.8
130	MFS-type transporter (*Mus musculus*)	4.0.E-68	191	30	79	67	73	63	2.4
131	Nicotinate phosphoribosyltransferase (*Drosophila melanogaster*)	0	189	69	82	66	78	55	2.3
132	unknown		186	41	43	52	43	34	4.4
133	UDP-glucuronosyltransferase (*Oryctolagus cuniculus*)	2.0.E-72	185	70	64	67	61	47	2.9
134	unknown		183	32	43	47	33	20	4.3
135	unknown		176	70	66	52	52	44	2.7
136	Scavenger receptor (*Rattus norvegicus*)	5.0.E-66	174	56	76	76	79	47	2.3

137	unknown		173	36	60	60	55	51	2.9
138	unknown		172	60	66	62	65	62	2.6
139	Gonadotropin-releasing hormone II receptor (*Clarias gariepinus*)	1.0.E-65	166	45	62	73	76	46	2.7
140	Chondroitin sulfate synthase (*Mus musculus*)	0	165	23	55	17	19	11	3.0
141	15-hydroxyprostaglandin dehydrogenase (*Homo sapiens*)	1.0.E-31	158	44	47	24	34	17	3.4
142	Angiotensin-converting enzyme (*Mus musculus*)	6.0.E-15	156	34	44	29	40	19	3.5
143	Adenine phosphoribosyltransferase (*Drosophila pseudoobscura*)	4.0.E-46	156	54	74	64	67	60	2.1
144	Cathepsin D (*Mus musculus*)	2.0.E-45	156	44	54	57	74	44	2.9
145	Alpha-N-acetylgalactosaminidase (*Gallus gallus*)	3.0.E-133	150	60	62	56	68	45	2.4
146	UDP-glucuronosyltransferase (*Mus musculus*)	2.0.E-63	150	61	48	52	69	41	3.1
147	Ornithine transporter (*Homo sapiens*)	4.0.E-38	148	21	41	37	29	24	3.6
148	Kynurenine-oxoglutarate transaminase (*Homo sapiens*)	2.0.E-152	146	57	57	57	64	39	2.6
149	Box A-binding factor (*Drosophila melanogaster*)	1.0.E-25	143	69	56	62	62	27	2.6
150	unknown		142	70	32	33	36	10	4.4
151	Cationic amino acid transporter (*Gallus gallus*)	4.0.E-124	141	32	35	46	37	23	4.0
152	Trehalose transporter (*Polypedilum vanderplanki*)	5.0.E-51	140	23	51	60	58	52	2.8
153	Thiol reductase (*Homo sapiens*)	6.0.E-13	138	38	51	48	47	26	2.7
154	Hydroxypyruvate isomerase (*Danio rerio*)	2.0.E-75	138	54	53	66	67	46	2.6
155	Lipoamide acyltransferase component of branched-chain alpha-keto acid dehydrogenase (*Mus musculus*)	4.0.E-102	136	49	42	56	52	33	3.2
156	unknown		133	23	36	35	39	26	3.7
157	unknown		133	32	34	31	43	20	3.9
158	unknown		130	8	45	5	5	1	2.9
159	unknown		129	5	36	7	6	4	3.6
160	Peroxidase (*Drosophila melanogaster*)	2.0.E-64	128	40	47	15	23	6	2.7
161	Inorganic phosphate cotransporter (*Drosophila ananassae*)	4.0.E-98	128	54	49	46	50	26	2.6
162	MFS-type transporter (*Homo sapiens*)	6.0.E-76	126	34	38	51	45	28	3.3
163	unknown		125	19	34	34	34	20	3.6
164	Cysteine dioxygenase (*Danio rerio*)	4.0.E-73	123	38	38	53	58	24	3.3
165	Clotting factor (*Tachypleus tridentatus*)	2.0.E-31	122	36	49	54	51	35	2.5
166	Glutamate transporter (*Bos taurus*)	2.0.E-113	122	34	37	38	38	20	3.3
167	Beta-galactosidase (*Canis familiaris*)	2.0.E-162	120	26	39	32	28	18	3.1
168	unknown		114	35	43	38	42	38	2.7
169	Serine carboxypeptidase (*Apis mellifera*)	2.0.E-146	112	56	43	50	54	28	2.6
170	unknown		112	41	30	31	45	20	3.8
171	Thiamine transporter (*Homo sapiens*)	2.0.E-78	112	42	41	31	44	29	2.8
172	unknown		110	5	26	6	5	6	4.3
173	unknown		110	30	25	26	24	11	4.4
174	Esterase (*Myzus persicae*)	5.0.E-76	106	32	34	25	27	19	3.2
175	Glucosylceramidase (*Pan troglodytes*)	5.0.E-135	106	29	30	39	33	21	3.6
176	Wnt (*Xenopus laevis*)	2.0.E-73	104	15	25	10	12	5	4.2
177	UDP-glucuronosyltransferase (*Rattus norvegicus*)	2.0.E-58	101	14	29	34	33	20	3.5
178	fuseless (*Drosophila melanogaster*)		99	13	45	40	38	33	2.2
179	Elastase inhibitor (*Mus musculus*)	2.0.E-59	98	19	35	27	23	16	2.8
180	unknown		97	22	25	24	24	17	3.9
181	Thymidine phosphorylase (*Rattus norvegicus*)	1.0.E-96	97	28	25	37	33	19	3.9
182	UDP-glucuronosyltransferase (*Canis familiaris*)	6.0.E-72	93	25	35	22	22	20	2.7
183	Dehydrogenase/reductase (*Gallus gallus*)	4.0.E-42	92	24	31	22	20	17	3.0

184	Dehydrogenase/reductase (*Gallus gallus*)	8.0.E-47	92	17	26	21	17	8	3.5
185	unknown		92	10	31	15	15	10	3.0
186	unknown		91	1	15	3	5	5	6.0
187	Cation transport regulator-like protein (*Danio rerio*)	3.0.E-44	90	13	28	24	23	24	3.2
188	ATP-binding cassette sub-family (*Mus musculus*)	7.0.E-133	88	12	20	17	15	11	4.4
189	unknown		86	35	32	34	35	17	2.7
190	Pyrroline-5-carboxylate reductase (*Pongo abelii*)	2.0.E-66	85	28	27	38	33	18	3.2
191	Acyl-CoA Delta desaturase (*Trichoplusia ni*)	2.0.E-100	84	12	26	14	14	4	3.2
192	Yellow-c (*Drosophila melanogaster*)	2.00E-73	82	7	24	14	12	12	3.3
193	UDP-glucuronosyltransferase (*Cavia porcellus*)	3.0.E-70	80	34	23	22	28	14	3.5
194	Ion transport peptide (*Schistocerca gregaria*)	2.0.E-41	80	18	26	28	29	31	3.0
195	unknown		79	6	23	22	12	5	3.4
196	Retinol dehydrogenase (*Homo sapiens*)	7.0.E-75	79	28	24	29	27	20	3.2
197	unknown		78	11	25	20	26	22	3.2
198	unknown		78	29	30	23	23	19	2.6
199	Carboxypeptidase (*Rickettsia conorii*)	1.0.E-34	78	8	33	20	22	17	2.4
200	Trehalose transporter (Drosophila willistoni)	4.0.E-42	77	16	34	22	23	14	2.3
201	unknown		76	18	16	15	16	6	4.8
202	Kynurenine/alpha-aminoadipate aminotransferase (*Bos taurus*)	3.0.E-95	76	25	21	23	20	12	3.5
203	unknown		75	8	22	6	7	4	3.4
204	ATP-binding cassette sub-family (*Homo sapiens*)	1.0.E-157	74	37	21	18	24	6	3.5
205	Glycine dehydrogenase (*Homo sapiens*)	0	74	11	27	18	18	12	2.7
206	*Alpha-mannosidase (Macaca fascicularis)*	0	72	17	28	28	25	19	2.5
207	Cysteine dioxygenase (*Danio rerio*)	3.0.E-73	71	22	20	33	32	26	3.5
208	unknown		69	10	20	20	18	15	3.5
209	Potassium channel protein (*Drosophila melanogaster*)	3.0.E-76	69	20	31	28	27	21	2.2
210	unknown		68	4	25	8	4	3	2.8
211	UDP-glucuronosyltransferase (*Rattus norvegicus*)	1.0.E-50	67	27	28	29	29	23	2.4
212	Potassium channel (*Gallus gallus*)	8.0.E-117	66	14	21	16	15	11	3.1
213	Alpha-N-acetylgalactosaminidase (*Gallus gallus*)	3.0.E-133	65	12	17	12	9	6	3.8
214	unknown		64	23	18	16	20	8	3.5
215	Lipase (*Drosophila melanogaster*)	9.0.E-90	63	20	25	20	22	15	2.5
216	D-alanine-D-alanine ligase (*Rickettsia bellii*)	3.0.E-176	63	12	28	18	18	11	2.3
217	unknown		62	5	19	19	17	18	3.3
218	Glycoprotein 3-alpha-L-fucosyltransferase (*Drosophila melanogaster*)	3.0.E-101	62	4	22	11	8	8	2.8
219	Metalloproteinase (*Mus musculus*)	1.0.E-87	61	11	22	16	13	8	2.7
220	Clotting factor (*Tachypleus tridentatus*)	7.0.E-23	60	7	19	26	22	17	3.2
221	unknown		57	7	20	23	17	18	2.8
222	Acetyl-coenzyme A transporter (*Rattus norvegicus*)	6.0.E-128	57	4	18	14	12	9	3.2
223	unknown		56	25	21	21	21	8	2.6
224	unknown		56	16	27	19	17	9	2.1
225	Acyl-CoA Delta desaturase (*Trichoplusia ni*)	1.0.E-72	53	2	10	2	2	0	5.2
226	unknown		52	13	14	13	9	6	3.7
227	unknown		51	2	16	3	4	3	3.3
228	*Allatostatin-A receptor (Bombyx mori)*	1.0.E-89	51	15	24	22	19	17	2.2
229	BDNF/NT-3 growth factors receptor (*Gallus gallus*)	9.0.E-83	51	16	19	16	15	8	2.7
230	Cytochrome P450 6a2 (*Drosophila melanogaster*)	4.0.E-72	51	16	20	22	21	23	2.6
231	CG32112 (*Drosophila melanogaster*)	5.0.E-148	50	20	25	24	24	22	2.1

232	unknown		50	14	18	12	19	14	2.7
233	Wos2 (*Schizosaccharomyces pombe*)	9.0.E-19	50	2	16	5	3	1	3.1
234	Beta-glucosidase (*Oryza sativa*)	4.0.E-10	49	6	6	0	2	0	8.4
235	unknown		49	6	13	19	24	13	3.7
236	unknown		48	11	15	12	10	4	3.2
237	unknown		47	8	10	18	22	14	5.0
238	Beta-galactosidase (*Mus musculus*)	1.0.E-10	45	7	10	10	10	9	4.5
239	unknown		43	8	14	16	18	12	3.0
240	Serine/threonine-protein phosphatase 6 regulatory ankyrin repeat subunit A (*Homo sapiens*)	3.0.E-91	42	12	16	5	6	2	2.6
241	unknown		41	1	14	4	2	1	2.8
242	UDP-glucuronosyltransferase (*Rattus norvegicus*)	1.0.E-54	41	6	12	10	9	8	3.3
243	unknown		38	7	17	13	15	14	2.2
244	unknown		38	2	7	5	6	4	5.4
245	TELO2-interacting protein (*Mus musculus)*	2.0.E-12	37	2	11	7	8	6	3.2
246	LIM homeobox transcription factor (*Mesocricetus auratus*)	2.0.E-14	36	4	7	2	4	1	5.6
247	unknown		35	12	12	14	11	10	3.0
248	unknown		34	6	10	12	15	14	3.6
249	Not2 (*Xenopus laevis*)	2.0.E-24	34	3	8	6	6	5	4.5
250	unknown		34	9	7	9	12	4	5.1
251	UDP-glucuronosyltransferase (*Rattus norvegicus*)	3.0.E-56	33	5	9	15	14	10	3.8
252	CG7120 (*Drosophila melanogaster*)	2.0.E-107	33	16	12	11	14	7	2.7
253	Serine/threonine-protein kinase (*Drosophila melanogaster)*	8.0.E-56	33	7	13	10	9	5	2.5
254	TELO2-interacting protein (*Homo sapiens*)	1.0.E-86	31	11	11	10	12	9	2.8
255	unknown		31	5	6	8	5	7	5.1
256	Myrosinase (*Brevicoryne brassicae*)	4.0.E-49	29	7	7	3	4	1	4.1
257	Acyl-coenzyme A thioesterase (*Mus musculus*)	3.0.E-31	28	3	5	1	2	1	5.1
258	unknown		28	5	13	4	9	4	2.1
259	unknown		28	14	6	5	7	3	4.7
260	unknown		27	11	10	8	9	6	2.6
261	unknown		27	6	4	4	5	3	6.4
262	Acyl-CoA Delta desaturase (*Trichoplusia ni)*	8.0.E-102	27	2	3	1	1	0	8.1
263	Myrosinase (*Brevicoryne brassicae*)	2.0.E-16	25	3	4	1	4	0	6.0
264	unknown		24	1	4	1	1	0	5.5
265	unknown		24	6	12	8	10	7	2.0
266	PRELI domain-containing protein (*Mus musculus*)	7.0.E-11	22	5	5	6	8	3	4.4
267	unknown		22	2	5	2	1	1	4.8
268	2-hydroxyacylsphingosine 1-beta-galactosyltransferase (*Rattus norvegicus*)	3.0.E-15	22	6	7	4	4	1	3.1
269	Glycerol-3-phosphate acyltransferase (*Rattus norvegicus)*	4.0.E-38	21	3	7	3	5	5	3.2
270	RING finger and transmembrane domain-containing protein (*Mus musculus*)	2.0.E-42	21	2	8	6	6	4	2.6
271	Krueppel (*Mus musculus*)	1.0.E-37	20	4	4	3	4	2	4.6
272	CG6761 (*Drosophila melanogaster*)	3.0E-101	20	1	7	2	2	1	2.7
273	Peroxidase (*Drosophila melanogaster*)	5.0.E-70	20	1	4	2	2	1	4.6
274	Amnionless (*Mus musculus*)	1.0.E-27	20	4	5	6	5	3	3.8
275	unknown		20	9	6	5	5	3	3.1
276	unknown		19	3	6	5	6	6	3.0
277	unknown		18	4	4	6	6	5	4.2
278	Ninjurin (*Mus musculus*)	4.0.E-09	18	7	7	5	5	2	2.8
279	UDP-glucuronosyltransferase (*Oryctolagus cuniculus*)	6.0.E-52	18	5	7	8	8	6	2.4
280	unknown		18	3	6	6	5	3	2.8

281	UDP-glucuronosyltransferase (*Oryctolagus cuniculus*)	1.0.E-42	17	4	5	2	3	1	3.4
282	unknown		17	1	3	2	1	0	5.5
283	Cubilin (*Canis familiaris*)	1.0.E-29	16	6	6	5	5	2	2.9
284	unknown		16	5	6	6	7	4	2.5
285	unknown		15	2	3	3	3	2	5.2
286	Stearoyl-CoA desaturase (*Bos taurus*)	1.0.E-103	15	0	2	0	0	0	6.3
287	unknown		15	7	5	2	7	2	2.9
288	unknown		15	2	6	3	3	2	2.6
289	unknown		13	4	2	1	2	1	7.0
290	Ras-related protein (*Oryctolagus cuniculus*)	9.0.E-65	11	0	4	4	4	3	3.0
291	unknown		11	0	3	1	1	0	4.4
292	Cytochrome P450 6a13 (*Drosophila melanogaster*)	4.0.E-102	11	3	3	2	2	1	4.2
293	Retrovirus-related Pol polyprotein (*Drosophila melanogaster*)	2.0.E-11	11	2	3	2	3	1	3.9
294	unknown		11	4	4	3	5	1	2.9
295	unknown		9	0	2	0	0	0	4.9
296	Retrovirus-related Pol polyprotein (*Drosophila melanogaster)*	4.0.E-06	9	2	3	1	2	3	2.7
297	unknown		8	1	3	1	1	1	2.4
298	unknown		8	3	2	2	2	1	4.3
299	unknown		8	0	3	1	3	2	2.6
300	CG17322 (*Drosophila melanogaster)*	2.0E-53	7	1	2	3	2	1	3.7
301	unknown		7	1	1	2	2	3	8.3
302	Beta-carotene 15,15' monooxygenase (*Gallus gallus*)	2.0.E-57	7	1	2	1	1	1	2.8
303	unknown		6	1	3	1	2	2	2.4
304	unknown		5	2	1	0	1	1	6.4
305	Stearoyl-CoA desaturase (*Bos taurus*)	1.0.E-26	5	0	0	0	0	0	-
306	Retinoid isomerohydrolase (*Rattus norvegicus*)	3.0.E-07	4	0	0	0	0	0	54.6
307	Lactase-phlorizin hydrolase (*Rattus norvegicus*)	3.0.E-30	4	1	1	0	1	0	2.6
308	unknown		3	1	1	1	0	0	4.8
309	unknown		3	0	1	1	0	0	4.6
310	PiggyBac transposable element-derived protein (*Homo sapiens*)	3.0.E-06	3	0	0	0	0	0	22.0
311	unknown		2	1	1	1	1	0	3.7
312	PiggyBac transposable element-derived protein (*Homo sapiens*)	3.0.E-26	2	0	1	1	0	0	3.0
313	unknown		2	0	0	0	0	0	10.2
314	unknown		2	0	0	0	0	0	6.7
315	unknown		2	0	1	1	0	0	3.1

[a]Colors highlight the following gene categories: red, PO and related genes; blue, RCP and related genes; green, lipid-related genes; orange, sugar-related genes.

Table S3. Gene ontology categories enriched in LGC-dominant genes.

GO ID	Term	FDR	Ontology
GO:0044281	small molecule metabolic process	4.1E-06	Biological process
GO:0016042	lipid catabolic process	4.2E-05	Biological process
GO:1901605	alpha-amino acid metabolic process	1.0E-03	Biological process
GO:0055114	oxidation-reduction process	1.7E-03	Biological process
GO:0006082	organic acid metabolic process	1.7E-03	Biological process
GO:0005975	carbohydrate metabolic process	5.1E-03	Biological process
GO:0009071	serine family amino acid catabolic process	1.9E-02	Biological process
GO:0044242	cellular lipid catabolic process	2.7E-02	Biological process
GO:0009110	vitamin biosynthetic process	3.0E-02	Biological process
GO:0006629	lipid metabolic process	3.3E-02	Biological process
GO:0009225	nucleotide-sugar metabolic process	4.5E-02	Biological process
GO:0003824	catalytic activity	6.0E-09	Molecular function
GO:0016757	transferase activity, transferring glycosyl groups	3.9E-05	Molecular function
GO:0016758	transferase activity, transferring hexosyl groups	3.1E-04	Molecular function
GO:0004553	hydrolase activity, hydrolyzing O-glycosyl compounds	4.3E-04	Molecular function
GO:0016491	oxidoreductase activity	4.3E-04	Molecular function
GO:0016798	hydrolase activity, acting on glycosyl bonds	7.0E-04	Molecular function
GO:0052689	carboxylic ester hydrolase activity	2.2E-02	Molecular function

Table S4. Highly-expressed genes in *A. pisum* tissues identified by RNA-seq analysis.

	AphidBase[a]	Function[a]	HC[b]	WB[b]	Ratio (HC/WB) [b]
1	ACYPI002789	Collagen	15078	858	17.6
2	**ACYPI001367**	**Phenoloxidase**	8904	341	26.1
3	ACYPI004656	Collagen	8473	468	18.1
4	ACYPI009912	Actin	6753	2340	2.9
5	ACYPI001359	SPARC	6269	353	17.8
6	ACYPI083140	Uncharacterized protein	5652	1483	3.8
7	ACYPI003478	Hemocytin	4411	39	113.2
8	ACYPI086030	Uncharacterized protein	4222	1025	4.1
9	**ACYPI004484**	**Phenoloxidase**	4151	140	29.6
10	ACYPI009414	Uncharacterized protein	3676	1173	3.1
11	ACYPI073884	Glutamine synthetase	3288	1506	2.2
12	ACYPI004672	Omega amidase	3162	775	4.1
13	ACYPI001736	Uncharacterized protein	2970	120	24.7
14	ACYPI001864	Uncharacterized protein	2489	390	6.4
15	ACYPI082655	Aspartic protease	2380	208	11.4
16	ACYPI001483	Echinoderm microtubule-associated protein	2369	169	14.0
17	ACYPI003572	Myophilin	2359	418	5.6
18	ACYPI008158	Gelsolin	2253	577	3.9
19	ACYPI008989	GILT-like protein	2137	656	3.3
20	ACYPI001365	Chitinase	2003	855	2.3
21	ACYPI081126	Uncharacterized protein	1973	610	3.2
22	ACYPI009721	Histone	1860	196	9.5
23	ACYPI006974	Cathepsin L	1823	538	3.4
24	ACYPI087208	Uncharacterized protein	1820	314	5.8
25	ACYPI56425	Uncharacterized protein	1792	68	26.4
26	ACYPI003244	Papilin	1730	77	22.5
27	ACYPI28737	60S ribosomal protein	1556	38	41.0
28	ACYPI082627	Uncharacterized protein	1530	342	4.5
29	ACYPI008487	Uncharacterized protein	1495	31	48.4
30	ACYPI081449	Laminin	1487	120	12.4
31	ACYPI008657	Glutathione S-transferase	1478	467	3.2
32	ACYPI34442	Uncharacterized protein	1412	514	2.7
33	ACYPI086042	Uncharacterized protein	1410	453	3.1
34	ACYPI001378	Amino acid transporter	1376	477	2.9
35	ACYPI009011	Fatty acid-binding protein	1361	365	3.7
36	ACYPI001776	Laminin	1328	143	9.3
37	ACYPI009530	Integral membrane protein	1277	445	2.9
38	ACYPI004906	Leucine-rich repeats and immunoglobulin-like domains protein 1	1273	109	11.7
39	ACYPI001993	Serine protease	1247	39	32.0
40	ACYPI50923	Fibulin	1230	21	59.5
41	ACYPI28373	Uncharacterized protein	1217	65	18.8
42	ACYPI007421	Uncharacterized protein	1217	28	43.0
43	ACYPI064152	Collagen	1152	72	16.0
44	ACYPI088246	Uncharacterized protein	1134	221	5.1
45	ACYPI004891	Peptidylprolyl isomerase B	1120	551	2.0
46	ACYPI008883	Glutamate transporter	1120	13	87.0
47	ACYPI089272	Uncharacterized protein	1083	163	6.6
48	ACYPI001380	Acetylcholine receptor	1036	28	36.8
49	ACYPI001274	Excitatory amino acid transporter	1006	258	3.9
50	ACYPI010019	Laminin	998	99	10.1

	AphidBase[a]	Function[a]	FB[b]	WB[b]	Ratio (FB/WB)[b]
1	ACYPI004672	Omega-amidase	12792	775	16.5
2	ACYPI003223	Odorant-binding protein	12684	1385	9.2
3	ACYPI000294	Uncharacterized protein	12123	657	18.4
4	ACYPI005249	Uncharacterized protein	11040	2369	4.7
5	**ACYPI086030**	**Uncharacterized protein**	10968	1025	10.7
6	ACYPI081676	Uncharacterized protein	8148	347	23.5
7	ACYPI009769	Glyceraldehyde-3-phosphate dehydrogenase	7399	3539	2.1
8	ACYPI009030	Uncharacterized protein	7323	1285	5.7
9	ACYPI060629	Uncharacterized protein	7134	371	19.3
10	ACYPI083140	Uncharacterized protein	7125	1483	4.8
11	ACYPI004796	Uncharacterized protein	6393	1593	4.0
12	ACYPI009414	Uncharacterized protein	5731	1173	4.9
13	ACYPI087473	Aldehyde dehydrogenase	5550	339	16.4
14	ACYPI066751	Uncharacterized protein	5526	974	5.7
15	ACYPI000057	Xylulose reductase	5397	909	5.9
16	ACYPI005806	Enolase	5247	1128	4.7
17	ACYPI007027	Fructose-bisphosphate aldolase	5159	1430	3.6
18	ACYPI073884	Glutamine synthetase	4442	1506	2.9
19	ACYPI081559	Uncharacterized protein	4343	1749	2.5
20	ACYPI080203	Uncharacterized protein	4026	146	27.6
21	ACYPI001461	Glutamine synthetase	3807	1154	3.3
22	ACYPI061541	Uncharacterized protein	3325	143	23.2
23	ACYPI083147	Uncharacterized protein	3257	485	6.7
24	ACYPI53781	Uncharacterized protein	3091	1537	2.0
25	ACYPI073004	Uncharacterized protein	3035	273	11.1
26	ACYPI006784	L-threonine dehydratase catabolic TdcB	2880	448	6.4
27	ACYPI008389	Uncharacterized protein	2872	1089	2.6
28	ACYPI007294	Uncharacterized protein	2753	139	19.8
29	ACYPI085343	Uncharacterized protein	2747	114	24.1
30	ACYPI002175	Lysozyme	2663	1303	2.0
31	ACYPI007254	Uncharacterized protein	2654	407	6.5
32	ACYPI002789	collagen	2653	858	3.1
33	ACYPI008106	Ornithine decarboxylase	2649	934	2.8
34	ACYPI001365	Chitinase	2642	855	3.1
35	ACYPI085855	Uncharacterized protein	2576	66	38.9
36	ACYPI008158	Gelsolin	2575	577	4.5
37	ACYPI006984	Spermidine synthase	2322	704	3.3
38	ACYPI005016	Serpin	2154	346	6.2
39	ACYPI008569	Apolipophorin	2152	399	5.4
40	ACYPI005113	Cytochrome P450 4g15	2128	867	2.5
41	**ACYPI005732**	**Uncharacterized protein**	2004	63	31.7
42	ACYPI069547	UTP-glucose-1-phosphate uridylyltransferase	1950	650	3.0
43	ACYPI004689	Uncharacterized protein	1912	319	6.0
44	ACYPI007739	Uncharacterized protein	1906	536	3.6
45	ACYPI010135	Hexokinase	1879	766	2.5
46	ACYPI000376	Cysteine proteinase	1869	418	4.5
47	ACYPI001378	Amino acid transporter	1858	477	3.9
48	ACYPI000007	Adenosylmethionine decarboxylase	1842	604	3.1
49	ACYPI004546	Glutamate–cysteine ligase	1766	231	7.6
50	ACYPI008065	Fatty acid synthase	1765	454	3.9

[a]Colors highlight the following gene categories: red, PO genes; blue, RCP-related genes; green, lipid-related genes; orange, sugar-related genes.
[b]HC, hemocyte; FB, fat body; WB, whole body.

Table S5. Expression levels of *Buchnera* Nmo genes identified by RNA-seq analysis.

	Location	Gene/protein (putative)	TPM (mean, n=3)
1	Chr	cspE	624,620
2	Chr	yba3	133,742
3	Chr	dapD	19,957
4	Chr	groEL	18,605
5	pLeu	ibpA	14,458
6	Chr	dnaK	10,393
7	Chr	rpsU	8,899
8	Chr	groES	8,463
9	Chr	csrA	7,711
10	Chr	fis	4,767
11	Chr	ahpC	4,125
12	Chr	ydhD	4,051
13	Chr	dnaJ	4,001
14	Chr	yba2	3,852
15	Chr	gapA	3,146
16	Chr	tufB	2,662
17	Chr	rho	2,561
18	Chr	ptsH	2,031
19	Chr	rpsJ	1,798
20	Chr	hypothetical protein BUCNMO_041	1,757
21	Chr	rpmF	1,749
22	Chr	dapF	1,729
23	Chr	rpsI	1,603
24	Chr	yciA	1,567
25	Chr	**aroE**	1,531
26	Chr	metE	1,435
27	Chr	metK	1,419
28	Chr	dut	1,374
29	Chr	**aroC**	1,364
30	Chr	hslU	1,329
31	Chr	bacA	1,328
32	Chr	argH	1,229
33	Chr	tsf	1,227
34	Chr	ssb	1,210
35	Chr	pfkA	1,194
36	Chr	tldD	1,174
37	Chr	tktB	1,152
38	Chr	**eno**	1,121
39	Chr	rpiA	1,076
40	Chr	rplC	1,034
41	Chr	trxA	1,010
42	pLeu	repA1	993
43	Chr	himD	921
44	Chr	mt	905
45	Chr	rplB	864
46	Chr	ptsG	822
47	Chr	htrA	814
48	Chr	ribB	809
49	Chr	ppa	796
50	Chr	norM	789

	Location	Gene/protein (putative)	TPM (mean, n=3)
51	Chr	yabI	745
52	Chr	rplL	741
53	Chr	gyrB	724
54	Chr	rplM	702
55	Chr	tadA	692
56	Chr	glyQ	682
57	Chr	nuoB	680
58	Chr	rpsP	665
59	Chr	rpsL	660
60	Chr	orn	659
61	Chr	hisD	649
62	Chr	argA	629
63	Chr	rpsO	597
64	Chr	adk	592
65	Chr	atpE	590
66	Chr	rpsK	577
67	Chr	gloB	568
68	Chr	nusG	568
69	Chr	tilS	565
70	Chr	rplT	563
71	Chr	flgI	553
72	pTrp	trpG	548
73	Chr	rnc	543
74	Chr	infC	531
75	Chr	sohB	514
76	Chr	hflB	512
77	Chr	**aroK**	511
78	Chr	yccK	510
79	Chr	map	510
80	Chr	sodA	503
81	Chr	atpD	497
82	Chr	yoaE	493
83	Chr	argC	485
84	Chr	ychF	483
85	Chr	pth	480
86	Chr	yrbA	476
87	Chr	ribA	467
88	Chr	rpmH	466
89	Chr	argD	444
90	Chr	yhgI	443
91	Chr	dapA	440
92	pTrp	trpE	438
93	Chr	crr	432
94	Chr	truA	431
95	Chr	gshA	429
96	Chr	atpB	427
97	Chr	nuoCD	424
98	Chr	gnd	420
99	Chr	yebD	415
100	Chr	purA	415
101	Chr	yjgF	412
102	Chr	rpoD	398
103	Chr	deaD	390

	Location	Gene/protein (putative)	TPM (mean, n=3)
104	Chr	rplD	385
105	Chr	hypothetical protein BUCNMO_154	383
106	Chr	rplJ	382
107	Chr	rpsM	381
108	Chr	thrC	374
109	Chr	nuoF	373
110	Chr	cyoE	370
111	Chr	htpX	370
112	Chr	rnhA	368
113	Chr	lon	354
114	Chr	rplA	354
115	Chr	lgt	350
116	Chr	rplK	349
117	Chr	rpsA	347
118	Chr	flgG	347
119	Chr	rpmA	346
120	Chr	yhbZ	345
121	Chr	ung	344
122	Chr	flhB	336
123	Chr	yhgN	333
124	Chr	purH	333
125	Chr	rsmC	333
126	Chr	infB	331
127	Chr	rpe	324
128	Chr	asnS	323
129	Chr	metF	322
130	Chr	greA	319
131	Chr	murE	318
132	Chr	rpsG	317
133	Chr	trpS	316
134	Chr	smrB	315
135	Chr	glyA	315
136	Chr	ilvC	305
137	Chr	miaA	304
138	Chr	prsA	298
139	Chr	rply	298
140	Chr	ilvI	294
141	Chr	rluD	293
142	Chr	rpoB	292
143	Chr	secA	284
144	Chr	pmbA	278
145	Chr	ygjD	278
146	Chr	mrcB	276
147	Chr	yb1688	273
148	Chr	flgC	273
149	Chr	pykA	270
150	Chr	rpoC	270
151	Chr	fliE	270
152	Chr	iscU	268
153	Chr	folD	265
154	pLeu	leuA	263
155	Chr	rpsN	263
156	Chr	secY	260

	Location	Gene/protein (putative)	TPM (mean, n=3)
157	Chr	rpmE	259
158	Chr	fabI	251
159	Chr	nif3-like protein BUCNMO_237	250
160	Chr	hslV	248
161	pLeu	yqhA	247
162	pLeu	leuD	247
163	Chr	lipA	246
164	Chr	sbcB	245
165	Chr	rpoA	245
166	Chr	gidA	242
167	Chr	murA	239
168	Chr	**aroB**	236
169	Chr	fusA	236
170	Chr	rpmJ	236
171	Chr	apaH	235
172	Chr	smpB	229
173	Chr	nuoI	228
174	Chr	gshB	227
175	Chr	pcnB	227
176	Chr	clpP	226
177	Chr	trmD	226
178	Chr	atpA	225
179	Chr	rnhB	225
180	Chr	flgH	224
181	Chr	cysS	222
182	Chr	rplS	221
183	Chr	acpS	220
184	Chr	hisS	219
185	Chr	rplW	219
186	Chr	bioA	218
187	Chr	lepA	217
188	Chr	hisG	216
189	Chr	nuoH	215
190	Chr	rpsE	215
191	Chr	pgi	214
192	Chr	thrB	214
193	Chr	ribH	213
194	Chr	rluC	212
195	Chr	asd	210
196	Chr	grpE	210
197	Chr	fabB	207
198	Chr	yabC	206
199	Chr	mviN	204
200	Chr	rpsD	203
201	Chr	**aroD(Q)**	202
202	Chr	rpsT	200
203	Chr	ybaB	200
204	Chr	aceE	198
205	Chr	rpmB	197
206	Chr	fba	196
207	Chr	rplU	196
208	Chr	rpsR	195
209	Chr	argE	194

	Location	Gene/protein (putative)	TPM (mean, n=3)
210	Chr	rpsB	193
211	Chr	flgA	191
212	Chr	**aroH**	190
213	Chr	cyoB	188
214	Chr	rpmI	187
215	Chr	rpsF	187
216	Chr	metG	187
217	Chr	tgt	185
218	Chr	rpsC	184
219	Chr	dnaQ	184
220	Chr	ydic	183
221	Chr	yajR	181
222	Chr	frr	180
223	Chr	grpE1	180
224	Chr	cmk	180
225	Chr	**aroA**	178
226	Chr	bioD	176
227	Chr	ilvD	176
228	Chr	nfo	176
229	Chr	rnpA	174
230	Chr	rplE	173
231	Chr	trmU	171
232	Chr	lig	171
233	Chr	fliQ	170
234	Chr	dnaG	169
235	Chr	rplQ	169
236	Chr	atpF	169
237	Chr	htpG	169
238	Chr	cca	168
239	Chr	rpsS	168
240	Chr	dapE	167
241	Chr	gmk	167
242	Chr	secE	166
243	Chr	trpB	165
244	Chr	flgD	164
245	pLeu	leuC	164
246	Chr	rplP	164
247	Chr	rpsH	163
248	Chr	rplF	163
249	Chr	bioH	161
250	Chr	ychE	161
251	Chr	trxB	158
252	Chr	dnaT	158
253	Chr	yciL	158
254	Chr	argB	155
255	Chr	ptsI	154
256	Chr	thrS	150
257	Chr	iscS	146
258	Chr	cyoA	146
259	Chr	phrB	146
260	Chr	yheM	145
261	Chr	hemK	145
262	Chr	lepB	145

	Location	Gene/protein (putative)	TPM (mean, n=3)
263	Chr	argF	144
264	Chr	dnaB	143
265	pLeu	leuB	143
266	Chr	rplN	142
267	Chr	yleA	141
268	Chr	tmk	141
269	Chr	gpmA	141
270	Chr	rnb	140
271	Chr	pnp	140
272	Chr	rplV	139
273	Chr	nuoG	139
274	Chr	typA	137
275	Chr	ybhE	136
276	Chr	folC	135
277	Chr	rplR	134
278	Chr	flgJ	134
279	Chr	**talA**	134
280	Chr	ileS	133
281	Chr	ftsY	133
282	Chr	nuoA	131
283	Chr	pepA	131
284	Chr	lpdA	130
285	Chr	dnaC	129
286	Chr	flhA	129
287	Chr	cyoD	129
288	Chr	murD	127
289	Chr	prfA	125
290	Chr	dnaX	125
291	Chr	ybeY	125
292	Chr	nuoE	125
293	Chr	trpC	124
294	Chr	rpmG	123
295	Chr	mutY	121
296	Chr	fpr	118
297	Chr	fliO	117
298	Chr	rplO	116
299	Chr	pgk	116
300	Chr	serC	116
301	Chr	nuoM	115
302	Chr	ftsJ	113
303	Chr	gyrA	113
304	Chr	sucA	111
305	Chr	lipB	109
306	Chr	serS	107
307	Chr	rimM	105
308	Chr	ackA	105
309	Chr	yhhF	104
310	Chr	dnaE	104
311	Chr	thrA	104
312	Chr	valS	103
313	Chr	atpH	103
314	Chr	uppS	103
315	Chr	yeaZ	102

	Location	Gene/protein (putative)	TPM (mean, n=3)
316	Chr	hisF	102
317	Chr	glnS	102
318	Chr	yfgK	101
319	Chr	yihA	101
320	Chr	dksA	101
321	Chr	yqgF	99
322	Chr	era	99
323	Chr	gltX	99
324	Chr	proS	98
325	Chr	nrdA	98
326	Chr	yidC	98
327	Chr	flgF	97
328	Chr	rep	97
329	Chr	rpoH	96
330	Chr	purB	95
331	Chr	fliH	94
332	Chr	argG	93
333	Chr	fldA	93
334	Chr	nuoJ	92
335	Chr	carA	92
336	Chr	aceF	90
337	Chr	trpA	89
338	Chr	prfB	89
339	Chr	hisB	88
340	Chr	rpsQ	88
341	Chr	nuoL	88
342	Chr	ychA	87
343	Chr	murF	86
344	Chr	nuoK	86
345	Chr	aspS	85
346	Chr	hisI	85
347	Chr	flgB	85
348	Chr	ribD1	85
349	Chr	fmt	84
350	Chr	carB	82
351	Chr	yggH	82
352	Chr	yhhP	81
353	Chr	polA	81
354	Chr	fliR	81
355	Chr	yggX	80
356	Chr	def	80
357	Chr	zwf	80
358	Chr	murC	79
359	Chr	pta	78
360	Chr	leuS	76
361	Chr	ygfZ	75
362	Chr	ilvH	74
363	Chr	fliI	72
364	Chr	rplI	70
365	Chr	recD	70
366	Chr	alaS	69
367	Chr	fabD	68
368	Chr	fabG	68

	Location	Gene/protein (putative)	TPM (mean, n=3)
369	Chr	dsbA	68
370	Chr	argS	68
371	Chr	rpmD	67
372	Chr	ksgA	67
373	Chr	yggB	67
374	Chr	pheT	67
375	Chr	dapB	67
376	Chr	**pheA**	67
377	Chr	fabZ	67
378	Chr	bioB	66
379	Chr	rplX	66
380	Chr	fliN	66
381	Chr	ribE	65
382	Chr	tpiA	64
383	Chr	ybeX	63
384	Chr	nusA	63
385	Chr	atpC	62
386	Chr	atpG	62
387	Chr	mraY	61
388	Chr	clpX	61
389	Chr	dnaN	60
390	Chr	**hisC**	59
391	Chr	ffh	59
392	Chr	pheS	58
393	Chr	murG	58
394	Chr	yceA	57
395	Chr	nusB	57
396	Chr	fig-like protein BUCNMO_223	57
397	Chr	sucB	56
398	Chr	cyoC	56
399	Chr	efp	55
400	Chr	yheN	53
401	Chr	fdx	53
402	Chr	secG	51
403	Chr	nth	51
404	Chr	infA	50
405	Chr	hisA	48
406	Chr	hisH	48
407	Chr	nrdB	48
408	Chr	glyS	46
409	Chr	lspA	46
410	Chr	yggS	45
411	Chr	recB	45
412	Chr	hypothetical protein BUCNMO_168	44
413	Chr	yidD	44
414	Chr	trpD	43
415	Chr	lysA	42
416	Chr	lysS	42
417	Chr	holB	41
418	Chr	yciC	38
419	Chr	yggW	37
420	Chr	ycfH	37
421	Chr	fliM	37

	Location	Gene/protein (putative)	TPM (mean, n=3)
422	Chr	yheL	35
423	Chr	rpmC	35
424	Chr	yedA	34
425	Chr	nuoN	34
426	Chr	thdF	34
427	Chr	fliF	32
428	Chr	hscB	27
429	Chr	fliG	27
430	Chr	holA	25
431	Chr	yfaE	24
432	Chr	hypothetical protein BUCNMO_280	24
433	Chr	ycfF	23
434	Chr	flij	21
435	Chr	rbfA	19
436	Chr	himA	19
437	Chr	hypothetical protein BUCNMO_060	17
438	Chr	acpP	15
439	Chr	hypothetical protein BUCNMO_140	13

[a] Colors highlight the following gene categories: red, shikimate pathway genes leading to synthesis of phenylalanine; blue, enzyme genes to generate starting substrates for the tyrosine synthesis, erythrose-4-phosphate and phosphoenolpyruvate.