Ann Colorectal Res. 2019 June; 7(2):e93703. doi: 10.5812/acr.93703.

Published online 2019 August 17. **Research Article**

Carausius morosus (Phasmatodea) Homologues of Human Genes with Elevated Expression in the Colon

Matan Shelomi [1, *]

[1]Department of Entomology, National Taiwan University, Taipei, Taiwan

[*]*Corresponding author*: Department of Entomology, National Taiwan University, No. 27, Lane 113, Sec 4, Roosevelt Rd., Taipei, Taiwan. Tel: +886-0233665588, Email: mshelomi@ntu.edu.tw

Received 2019 May 14; **Revised** 2019 July 12; **Accepted** 2019 July 30.

Abstract

Background: Preliminary testing of novel drugs for colorectal conditions must be performed on animal models, with invertebrate models desirable for practical reasons. The insect excretory organs, the Malpighian tubules, have been cited as models for human renal disease research because they differentially express several genes homologous to those differentially expressed in human kidneys. Their role in excretion and homeostasis suggests that they could be models for human colorectal disease. The insect *Carausius morosus* (Phasmatodea) has been a model organism for decades. Regarding its potential use as a colorectal disease model, it has an advantage over other insects in that excretion in Phasmatodea is split between two organs: Malpighian tubules and the Phasmatodea-specific "appendices of the midgut".

Objectives: To find homologues of human colon genes expressed in the excretory tissues of *C. morosus* for potential use in drug testing and other experiments requiring an animal model.

Methods: Pre-existing transcriptomics data for the excretory system of the *C. morosus* were examined to find genes homologous to those known to have elevated expression in the human colon. This was done with the goal of possibly determining the excretory tissues in which they are differentially expressed.

Results: Exactly sixty transcripts from the excretory system transcriptome of *C. morosus* showed high sequence homology with human colon-specific genes, with a minimum e-value of 1e-50. Examples include solute carriers, myosin, bestrophin, carbonic anhydrase, and nitric oxide synthase. Several genes were identified with prognostic value for renal, pancreatic, endometrial, liver, skin, and urothelial cancers.

Conclusions: *C. morosus* can be used as model insect for human medical research applications, including colorectal drug testing.

Keywords: Transcriptomics, Colon, Protein, *Carausius morosus*, Model Organism

1. Background

Advances in "-omics" technologies and the ever-growing data they produce provide much novel insight into colorectal cancer and other diseases (1). Transcriptomics data in particular helps identify specific genes that are highly and/or differentially expressed in specific tissues. Once genes are identified, follow-up research can include testing drugs that affect gene expression or protein production. The integration of genomic, proteomic, and transcriptomic data enables researchers to rapidly identify potential genes of interest (2). The Human Protein Atlas is one such database of integrated human "-omics" data (3, 4). Regarding the colon, the database presently shows that of the 13496 proteins expressed in the colon (69% out of the total 19613), 166 are differentially expressed there compared to all other tissues (5).

Animal models have been used for a long time in colorectal drug discovery and testing, though the models are predominantly vertebrates such as dogs, pigs, and rodents (6). Experimenting on vertebrates is slow, expensive, and not compatible with certain ethical beliefs. Thus, alternatives to vertebrate animal research models are desirable in colorectal research as in other biomedical fields (7). Insects, which are cheaper to rear, can be experimented on with larger sample sizes at much faster rates, with insect welfare being less of a concern to most animal rights activists. While insects and humans certainly have considerable morphological and physiological differences, many of the key vital functions and their associated genes and proteins are homologous, predating our least common ancestor. Insects are thus a valid alternative to vertebrates for preclinical research (8). Insects have also benefitted from the "-omics" revolution, and a growing amount of data is available from which to search for genes of interest and run comparative genomic/proteomic assays.

It is likely that insects can be used as models for colorectal disease, with the associated benefits in terms of

cost, time, and ethics (6). The main obstacle to such practice is a lack of knowledge of insect proteins that are homologous to the human colon proteome. Compare this with the known homology between insect and human hemocytes (8) and excretory organs (9). The insect equivalents of the human nephrons are the Malpighian tubules (MpgTs), which collect nitrogenous and ionic wastes as well as certain xenobiotics, and produce a primary urine that is excreted into the hindgut and passed with food waste (10, 11). Experiments with the model insect *Drosophila melanogaster*, the fruit fly, confirm that these analogous functions of homeostasis and excretion come with homologous genes to those of the human kidney, including dozens of highly expressed *Drosophila* MpgT genes with close homology to human renal disease genes (9). While vertebrates may be more phenotypically similar to humans, genetically the insects are suitable models, as the main functions essential to life and their associated genes have been conserved across millions of years of evolution (9).

While *Drosophila* is well studied, it might not be the best model for excretory or colorectal disease research. The stick insects, or Phasmatodea, have split their excretory functions between the MpgTs and another, Phasmatodea-specific system currently referred to as the "appendices of the midgut" for lack of a better term (12). Similar to and evolutionarily derived from MpgTs, the appendices of the midgut (AoMs) arise from the midgut rather than the hindgut, appearing to excrete different compounds (12). This division of labor has been confirmed with transcriptomics data from the historic model stick insect, *Carausius morosus* (13), demonstrating that several genes are differentially expressed in either the AoMs or the MpgTs (14). If either of these organs express genes homologous to human colon genes, then not only can *C. morosus* and other Phasmatodea be used in animal testing for colorectal pharmaceuticals, but also they may be a superior model to other insects where all excretory function is done by the same tubules, meaning non-target and target genes overlap and are harder to respectively control and test in the same experiment. In addition, Phasmatodea are known to develop pseudotumors in their gut (15, 16), making them one of the few insects with documented neoplasms and, thus, one of the few insects that could be research models for human cancers - gastrointestinal or otherwise.

2. Objectives

The goal of this study was to examine the published *C. morosus* excretory system transcriptome for transcripts homologous to the 166 genes differentially expressed in the human colon. A further goal was to see in which of the insect's excretory organ tissues they are highly and/or differentially expressed, and whether any of the human homologues are known to have favorable or unfavorable prognostic value for cancer.

3. Methods

The methods are taken from the work identifying human kidney gene homologues in *Drosophila* transcriptomes (9) (Figure 1). The *C. morosus* transcriptome comes from a previous publication (14), in which transcriptomes were made separately for the AoMs, MpgTs, and the tissue of the midgut (MG) between the origins of these tubes. The raw Sequence Read Archive (SRA) data is available in GenBank (accession number: PRJNA314295), and the assembled transcriptome of 73143 contigs is available on DRYAD (DOI: 10.5061/dryad.5rm68) (14). A nucleotide BLAST database from this transcriptome was made using an offline, terminal-based BLAST program (17).

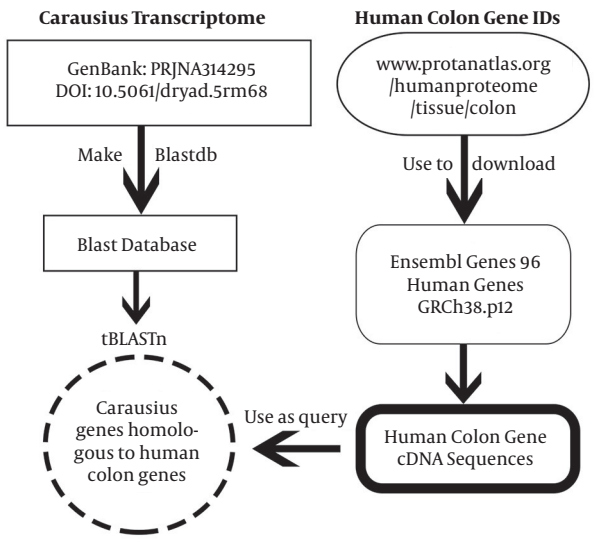

Figure 1. Flowchart of the procedure used in this study. A publicly-available *Carausius morosus* excretory tissue transcriptome was converted into a BLAST-searchable database (17). Ensemble IDs for human colon-specific genes (5) were taken from the Human Proteome Atlas (2-4) and used to download the cDNA sequences of these genes from the Ensemble Biomart (18). These cDNA sequences were used as a tBLASTn (17) query on the *C. morosus* database, and the results were all *C. morosus* excretory system transcripts homologous to human colon-specific genes.

Ensembl ID's for the 166 human colon-specific genes were obtained from the Human Protein Atlas website (2-5), before being used to download the associated unique cDNA sequences from the Ensembl Biomart dataset Ensembl Genes 96, Human Genes GRCh38.p12 (18). These were used as queries with tBLASTn (17) to examine the *C. morosus* transcriptome for homologous sequences with a con-

servative e-value cutoff of 1e-50. Data from the original *C. morosus* transcriptome publication (14) on genes that are differentially expressed in each tissue were added to the results, as well as data from the Human Protein Atlas on the cancer prognostic P values based on Kaplan-Meier analysis, where high expression of a gene is associated with favorable or unfavorable survival prognosis (19, 20). These data are combined together in Appendix 1 in Supplementary File.

4. Results

A total of 60 *C. morosus* excretory system transcripts were found to be homologous to the 166 human genes differentially expressed in the colon, with some of these insect transcripts homologous to more than one human gene (Appendix 1 in Supplementary File). Of these, 43 were "differentially expressed" in one of the three *C. morosus* tissues according to the Cufflinks (http://cole-trapnell-lab.github.io/cufflinks) analysis from Shelomi et al. (14). Three were also "highly expressed", defined as expression levels greater than 10 times the mean (14). Nine of these genes have unfavorable prognostic P values in humans for renal, pancreatic, liver, and skin cancers, as well as favorable P values for renal, endometrial, and urothelial cancers. None of the human genes linked to colorectal cancer had insect homologues above the inclusion threshold.

Examples of genes differentially expressed in the human colon with homologues in insect excretory tissue include ATPase phospholipid transporter 10B (ATP10B, with six homologous *C. morosus* genes), bestrophin (BEST2 and BEST4, with one *C. morosus* homologue), carbonic anhydrase (CA1, CA2, and CA7, with one *C. morosus* homologue), hepatocyte nuclear factor 4 alpha (HNF4A, with four *C. morosus* homologues), monoacylglycerol O-acyltransferase 3 (MOGAT2 and MOGAT3, with one *C. morosus* homologue), myosin 1A (MYO1A, with at least 11 *C. morosus* homologues), nitric oxide synthase 2 (NOS_2, with two *C. morosus* homologues), solute carrier family 17 member 4 (SLC17A14, homologous to four *C. morosus* inorganic phosphate cotransporter genes), solute carrier family 26 members 2 and 3 (SLC26A2 and SLC26A3, homologous to three *C. morosus* solute carrier family 26 member 10-like isoforms), UDP glucuronosyltransferase family 2 member B17 (UGT2B17, with at least 11 *C. morosus* homologues), and Villin 1 (VIL1, with two *C. morosus* homologues). Notable human colon-specific genes with no homologues identified in the *C. morosus* excretory organs include aquaporin, carcinoembryonic antigen related cell adhesion molecule 5, chloride channels, claudin, insulin, intelectin, galectin, mucin, neurexophilin, otopetrin, resistin, and tetraspanin.

5. Discussion

The finding of 60 *C. morosus* excretory organ genes homologous to human colon disease genes, among which 43 are differentially expressed and 3 are highly expressed, is a comparable result to that of Wang et al.'s assay seeking renal function homologue genes in *Drosophila* MpgTs (9). This paper's e-value cutoff of 1e-50 is conservative: if larger e-values such as the default e-10 were tolerated, even more homologous genes would have been found. Note too that the query in this study was limited to genes for colon-specific proteins rather than all the 13496 proteins expressed in the colon, reducing the likelihood of finding a match in the insect transcriptome. For example, *C. morosus* excretory organs are known to express aquaporins, chloride channels, and otopetrin (14), even though none were identified in this study. Some of these insect genes may be homologous to those expressed in multiple human tissues including the colon, though none were homologous with the 166 colon-specific genes. Likewise, the *C. morosus* transcriptome was produced solely with excretory tubule and gut tissues, not with the whole body as was done in Wang et al.'s *Drosophila* assay (9). This means that colon gene homologues may exist in non-excretory *C. morosus* tissues. Even with these limitations, a considerably number of Phasmatodea gene homologues to human colon-specific genes and human cancer genes were found. Disappointingly, no homologues were found to colorectal or gastric cancer genes.

5.1. Conclusions

Carausius morosus and other stick insects can be used as model insects for human medical research, including drug testing for renal and colorectal disease. Further experimental data in required on the compounds that affect the primary urines of the AoMs compared with the MpgTs, the genes that are involved in Phasmatodean pseudotumor production, and the genes that are expressed in the *C. morosus* hindgut itself. One these details are revealed, the use of stick insects as model insects for colorectal drug testing may become a reality.

Supplementary Material

Supplementary material(s) is available here [To read supplementary materials, please refer to the journal website and open PDF/HTML].

Footnotes

Conflict of Interests: The authors have no financial or other conflicts of interest to declare.

Funding/Support: Matan Shelomi is funded in part by the Ministry of Science and Technology of Taiwan, grant no. 106-2311-B-002-002-MY3.

Ethical Approval: There are no applicable ethical considerations to consider.

References

1. Wood LD, Parsons DW, Jones S, Lin J, Sjoblom T, Leary RJ, et al. The genomic landscapes of human breast and colorectal cancers. *Science*. 2007;**318**(5853):1108–13. doi: 10.1126/science.1145720. [PubMed: 17932254].
2. Fagerberg L, Hallstrom BM, Oksvold P, Kampf C, Djureinovic D, Odeberg J, et al. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol Cell Proteomics*. 2014;**13**(2):397–406. doi: 10.1074/mcp.M113.035600. [PubMed: 24309898]. [PubMed Central: PMC3916642].
3. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. Tissue-based map of the human proteome. *Science*. 2015;**347**(6220):1260419. doi: 10.1126/science.1260419. [PubMed: 25613900].
4. Yu NY, Hallstrom BM, Fagerberg L, Ponten F, Kawaji H, Carninci P, et al. Complementing tissue characterization by integrating transcriptome profiling from the Human Protein Atlas and from the FANTOM5 consortium. *Nucleic Acids Res*. 2015;**43**(14):6787–98. doi: 10.1093/nar/gkv608. [PubMed: 26117540]. [PubMed Central: PMC4538815].
5. Gremel G, Wanders A, Cedernaes J, Fagerberg L, Hallstrom B, Edlund K, et al. The human gastrointestinal tract-specific transcriptome and proteome as defined by RNA sequencing and antibody-based profiling. *J Gastroenterol*. 2015;**50**(1):46–57. doi: 10.1007/s00535-014-0958-7. [PubMed: 24789573].
6. Yang L, Chu JS, Fix JA. Colon-specific drug delivery: New approaches and in vitro/in vivo evaluation. *Int J Pharm*. 2002;**235**(1-2):1–15. [PubMed: 11879735].
7. Mayer EA, Bradesi S, Chang L, Spiegel BM, Bueller JA, Naliboff BD. Functional GI disorders: From animal models to drug development. *Gut*. 2008;**57**(3):384–404. doi: 10.1136/gut.2006.101675. [PubMed: 17965064]. [PubMed Central: PMC4130737].
8. Berger J. Preclinical testing on insects predicts human haematotoxic potentials. *Lab Anim*. 2009;**43**(4):328–32. doi: 10.1258/la.2008.007162. [PubMed: 19505933].
9. Wang J, Kean L, Yang J, Allan AK, Davies SA, Herzyk P, et al. Function-informed transcriptome analysis of Drosophila renal tubule. *Genome Biol*. 2004;**5**(9):R69. doi: 10.1186/gb-2004-5-9-r69. [PubMed: 15345053]. [PubMed Central: PMC522876].
10. Dow JA. Insights into the Malpighian tubule from functional genomics. *J Exp Biol*. 2009;**212**(Pt 3):435–45. doi: 10.1242/jeb.024224. [PubMed: 19151219].
11. Dow JA, Davies SA. The Malpighian tubule: Rapid insights from post-genomic biology. *J Insect Physiol*. 2006;**52**(4):365–78. doi: 10.1016/j.jinsphys.2005.10.007. [PubMed: 16310213].
12. Shelomi M, Kimsey LS. Vital staining of the stick insect digestive system identifies appendices of the midgut as novel system of excretion. *J Morphol*. 2014;**275**(6):623–33. doi: 10.1002/jmor.20243. [PubMed: 24338977].
13. Ramsay JA. The excretory system of the stick insect dixippus morosus (Orthoptera, Phasmidae). *J Exp Biol*. 1955;**32**:183–99.
14. Shelomi M. De novo transcriptome analysis of the excretory tubules of Carausius morosus (Phasmatodea) and possible functions of the midgut 'appendices'. *PLoS One*. 2017;**12**(4). e0174984. doi: 10.1371/journal.pone.0174984. [PubMed: 28384348]. [PubMed Central: PMC5383107].
15. Hoffmann P, Holtmann M, Dorn A. Degenerative and regenerative processes involved in midgut pseudotumor formation in the stick insect (Carausius morosus). *J Morphol*. 2009;**270**(12):1454–74. doi: 10.1002/jmor.10770. [PubMed: 19603413].
16. Holtmann M, Dorn A. Midgut pseudotumors and the maintenance of tissue homeostasis: Studies on aging and manipulated stick insects. *J Morphol*. 2009;**270**(2):227–40. doi: 10.1002/jmor.10685. [PubMed: 19034916].
17. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res*. 1997;**25**(17):3389–402. doi: 10.1093/nar/25.17.3389. [PubMed: 9254694]. [PubMed Central: PMC146917].
18. Kinsella RJ, Kahari A, Haider S, Zamora J, Proctor G, Spudich G, et al. Ensembl BioMarts: A hub for data retrieval across taxonomic space. *Database (Oxford)*. 2011;**2011**:bar030. doi: 10.1093/database/bar030. [PubMed: 21785142]. [PubMed Central: PMC3170168].
19. Thul PJ, Lindskog C. The human protein atlas: A spatial map of the human proteome. *Protein Sci*. 2018;**27**(1):233–44. doi: 10.1002/pro.3307. [PubMed: 28940711]. [PubMed Central: PMC5734309].
20. Uhlen M, Zhang C, Lee S, Sjostedt E, Fagerberg L, Bidkhori G, et al. A pathology atlas of the human cancer transcriptome. *Science*. 2017;**357**(6352). doi: 10.1126/science.aan2507. [PubMed: 28818916].