IDENTIFICATION AND CHARACTERIZATION OF SIGNALING MOLECULES IN NEMATODES AND BACTERIA USING LC-MS-BASED COMPARATIVE METABOLOMICS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Oishika Panda

May 2018

IDENTIFICATION AND CHARACTERIZATION OF SIGNALING MOLECULES IN NEMATODES AND BACTERIA USING LC-MS-BASED COMPARATIVE METABOLOMICS

Oishika Panda, Ph. D.

Cornell University 2018

Chemical communication forms an integral part of any organism's overall health, longevity, social interactions and various other aspects of its biology. Chemical cues produced by different organisms are products of primary and secondary metabolism, and are highly specific in their structural assembly and signaling functions. Minute differences in structures and functions of these metabolites can, for example, help an organism distinguish between pathogenic and beneficial microbes, cause the organism to extend or reduce its lifespan, signal availability or lack of sustenance, and provide snapshots of its metabolic state. As such, it is crucial to annotate structures and functions of the underlying chemical causes behind biological effects, understand how subtle structural differences cause dramatic changes in observed phenotypes, and elucidate their biosynthetic mechanisms to learn how biological information is encoded in these signaling molecules.

Ultra-high-performance liquid chromatography coupled to tandem mass spectrometry (UHPLC-MS/MS) is rapidly becoming an indispensable tool in the study of signaling molecules using comparative metabolomics. This technique far outdoes classical activity-guided fractionation methods to identify chemical cues in biological systems in terms of required time, ease of use and sensitivity. The use of both activity-guided fractionation and UHPLC-MS/MS techniques to elucidate structures of signaling molecules involved in host-pathogen interactions between nematodes and bacteria, and mitochondrion-mediated longevity in nematode are

discussed in this dissertation. Forward genetic approaches to identify biosynthetic enzymes responsible for the highly specific assembly of nematode-derived modular metabolites, and validation of their biosynthetic functions are also described. Finally, an untargeted metabolomics approach combining high resolution UHPLC-MS/MS data with genome-wide association studies to annotate the nematode metabolome and identify candidate biosynthetic genes for various metabolites is evaluated.

BIOGRAPHICAL SKETCH

Oishika Panda was born and raised in Kolkata, India. She attended Our Lady Queen of the Missions School, Kolkata, where she developed a keen interest in chemistry and biology at a very young age. She then went on to receive her Bachelors' Degree with Honors in Chemistry at St. Stephen's College, Delhi, and her Masters' Degree in Chemistry from the Indian Institute of Technology, Roorkee.

Oishika was selected to participate in the three-summer internship programs at the prestigious Jawaharlal Nehru Center for Advanced Scientific Research (JNCASR), Bengaluru, in her undergraduate days, where she was introduced to research labs in Materials, Computational and Medicinal Chemistry. While working on the synthesis and characterization of various aminoglycosides at JNCASR one summer, she became interested in the concept of mimicry of mammalian glycoproteins by synthetic compounds, and the decoration of such drugs with new side chains for better biological activity. This experience made her realize her deep interest in understanding the chemistry of small molecules in biological systems, and inspired her to pursue research in Chemical Biology.

Oishika then came to Cornell University and joined Prof. Frank C. Schroeder's research group, where she studied the biosynthesis of signaling metabolites in nematodes. She enjoyed discovering new molecules in different organisms using various techniques, and finding efficient ways to correlate large genomic and metabolomic datasets as part of her graduate research.

Dedicated to Mrs. A. Jacob, my first teacher of Chemistry, who told me when I was twelve, “Do not rely on your teacher to teach you everything. The greatest lessons you will learn will be the ones you discover yourself, because no one in the world will know them better than you.”

ACKNOWLEDGMENTS

Everything I have accomplished has only been possible because of my parents, Shailen and Shibani Panda. Their prayers, guidance and constant support have made me who I am today, and I owe them my life.

I extend my gratitude to my beautiful family away from home, the Reverend Dr. Valson Thampu, Dr. Grace Mary Valson, Drs. Sarah and Anna Valson. You have taught me to be independent and never shy away from a challenge, and for that I am eternally grateful.

I thank my guardian angel, (the late) Mrs. Rajlakshmi Haldar, my first teacher of Physics, music and brownie-baking, who would be extremely proud of me. I know she is beaming with joy from the heavens, and will always be my guiding spirit throughout my life.

Thank you, Mrs. Leena Palit, my English teacher, for being my friend, philosopher and guide throughout my life. You have taught me organization, elegance and the ability to march from strength to strength with my head held high, much beyond what is to be learned from any lesson confined within the four walls of a classroom.

I am grateful to my undergraduate instructors, Profs. SV Eswaran, (the late) KM Matthew, Vibha Sharma, Shabnam Johry, Rashmi Sachdeva, Rene Saksena, Ekta Arora, Jayanta Haldar and RK Peddinti. Thank you for teaching me to go beyond textbooks, question everything I come across and never be afraid to venture out of the boundaries of chemistry and apply my knowledge to biological systems.

Graduate school is a tough phase in the life of a scientist, and I will always be grateful to the wonderful set of friends who have helped me get through it. Thank you, Alex, for teaching me to ask the right questions, helping me answer them, picking me up and dusting me off every time I needed it. Thank you, Neelanjan, JJ, Murli, Andreas, Sarah, Sine, Jessica, Kyle, Saba, Ornella

and Eshan. I have enjoyed discussing projects with you and learning how to cope with the difficulties of graduate life from your experiences. Your burning passion for science will continue to inspire me. Thank you, Dr. Christina Mohajerani, for helping me overcome my fears and teaching me to believe in myself.

I am grateful to my PI, Prof. Frank C. Schroeder, for nurturing me not only as a scientist, but also as a team player dealing with people from different walks of life. He taught me there is no shame in saying, “I don’t know”, as long as it is followed by a resolute “but I will find out.” Thank you, Frank.

I am also grateful to Prof. S. Sylvia Lee and Prof. Hening Lin, members of my graduate committee. Thank you for your encouragement and all your help throughout my graduate career.

I thank my collaborators worldwide for educating me about so many different aspects of biology. It has been a pleasure working with you all, and I hope our paths cross again, soon.

Last but not least, thank you, Richard. You have been my rock, my constant source of strength, the sounding board for all my presentations, the shoulder to cry on, my inspiration to reach for the stars, the biggest believer in my scientific ability and strongest critic of my scientific ideas. I could not have done any of this without you.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation	Definition
2D NMR	Two-Dimensional Nuclear Magnetic Resonance
A, T, G, C, U	Adenine, Thymine, Guanine, Cytidine, Uracil
ACOX	Acyl Coenzyme A OXidase
ACS	Acyl Coenzyme A Synthetase
AMP	Adenosine MonoPhosphate
ATP	Adenosine TriPhosphate
CGC	Caenorhabditis Genetics Center
CoA	Coenzyme A
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CRTC	CREB-Regulated Transcription Coactivator
DAF	abnormal DAuer Formation
DCT	DAF-16/FOXO Controlled, germline Tumor affecting
DMSO	DiMethyl SulfOxide
DNA	DeoxyriboNucleic Acid
dqfCOSY	double quantum filtered Correlation Spectroscopy
ETC	Electron Transport Chain
GC	Gas Chromatography
GFP	Green Fluorescent Protein
GTP	Guanosine TriPhosphate
GWA	Genome-Wide Association
GWAS	Genome-Wide Association Study
HMBC	Heteronuclear Multiple Bond Correlation spectroscopy
HRMS	High Resolution Mass Spectrometry
HSQC	Heteronuclear Single Quantum Coherence spectroscopy
LC-MS	Liquid Chromatography coupled to Mass Spectrometry
LRO	Lysosome-Related Organelle
MAMP	Microbe-Associated Molecular Pattern
mQTL	metabolite Quantitative Trait Locus
NBRP	National BioResource Project
NIL	Near Isogenic Line
OAC	O-AcylTransferase
PCA	phenazine-1-carboxylic acid
PCN	phenazine-1-carboxamide
PCR	Polymerase Chain Reaction
Ppa	*Pristionchus pacificus*
PTEN	Phosphatase and TENsin homolog

Abbreviation	Definition
RNA	RiboNucleic Acid
RNAseq	RNA sequencing
ROS	Reactive Oxygen Species
SAH	S-AdenosylHomocysteine
SAM	S-AdenosylMethionine
SCP	Sterol Carrier Protein
SNP	Single Nucleotide Polymorphism
TCA	TriCarboxylic Acid
TGF-β	Transforming Growth Factor β
tRNA	transfer RNA
UHPLC	Ultra-High Performance Liquid Chromatography
UPR	Unfolded Protein Response

INTRODUCTION

Annotating the metabolome

All domains of life on earth consist of four levels of organismal regulation – the genome, transcriptome, proteome and metabolome, and all these 'omes' of life are made of molecules, large and small. These molecules generally consist of simple building blocks. DNA and RNA have different combinations of the nucleotides of A, T, G, C, U and their modifications, connected by phosphodiester bonds, giving rise to the genome and transcriptome of all organisms. Different combinations of 22 amino acids and their modifications, connected by peptide linkages, constitute the proteome. These building blocks have been extensively studied for decades, and such studies have enabled scientists to correctly predict, and eventually validate, the structures and functions of different genes and proteins. In contrast, the metabolome, or the set of small molecule metabolites produced by a given biological system, remains quite poorly understood.

Metabolites are relatively unpredictable in their structures, and more so in their functions. The sheer number of possible structures of metabolite building blocks, and the possibilities for combining these building blocks, have been intimidating obstacles in the systematic annotation of the chemical 'dark matter' of life. Annotating a few structures and functions of biogenic small molecules in any biological system only makes one aware of the expanse of chemical space still waiting to be discovered in all forms of life. New tools are being developed and technology is being advanced to allow scientists to easily explore the intricacies of chemical interactions in sub-cellular organelles, whole cells, different tissues and, of course, the whole organism. Recent technological advances in DNA sequencing, epigenetics, RNA sequencing, and metabolic profiling techniques such as 2D NMR spectroscopy and high resolution GC- and LC-MS have paved the way to integrate the different -omics in order to provide a more comprehensive understanding of complex biological systems[1].

Nematodes as model organisms

Caenorhabditis elegans, the 'chemist-friendly' free living model nematode[2], was among the first organisms to have its entire genome sequenced[3]. It bears great similarity to other organisms such as mammals in terms of major physiological pathways. Such a high level of conservation has led to extensive understanding of various signaling pathways that govern aging, social behaviors, metabolism, growth and development, diseases such as Alzheimer's, diabetes, obesity[4-16], to name but a few. The plethora of scientific literature that has emerged over the last two decades has brought to light the utility and challenges of studying *C. elegans* as a model organism to elucidate regulatory networks in the organism, which remain, at best, partly understood. *Pristionchus pacificus* has recently been established as a satellite model organism[17] to *C. elegans*. Satellite organisms are closely related to established model organisms and are especially useful for studying the evolution of the genotype-phenotype association, which can become increasingly obscured in more distant relatives[18].

Like in most model organisms, advances in the genomics and proteomics of *C. elegans* have left the annotation of its metabolome far behind. Recent investigations by the Schroeder research group (Cornell University), among others, have elucidated the structures and functions of several biogenic small molecules in *C. elegans* and *P. pacificus* that govern almost every aspect of their biology, including aging, dauer diapause, sexual attraction and chemosensation, using 2D NMR spectroscopy and LC-MS[19-35]. These studies have focused on the production of complex modular 'secondary' metabolites by the nematodes, which is not common in animals, although biosynthetic machineries for secondary metabolites are well known in bacteria, fungi and plants[33]. UHPLC-MS-based comparative metabolomics has now become an excellent tool for the high-throughput profiling[36-37] of metabolites[38-39], for both known biogenic small molecules as well as

metabolites whose structures, biosyntheses and functions are yet to be elucidated. In this dissertation, the author describes her use of the study of metabolomics in the organisms *Caenorhabditis elegans* and *Pristionchus pacificus* to understand chemical causes behind different phenotypes, the biosyntheses of such molecules, and attempts to combine genomic and metabolomic data to annotate the metabolome.

Preview of chapters

Discrimination between pathogenic and beneficial microbes is essential for host organism immunity and homeostasis. In Chapter 1, the author shows that chemosensory detection of two secondary metabolites produced by the pathogen *Pseudomonas aeruginosa* modulates a neuroendocrine signaling pathway that promotes avoidance behavior in *C. elegans*. Briefly, secondary metabolites phenazine-1-carboxamide and pyochelin activate a G-protein signaling pathway in the ASJ chemosensory neuron pair that induces expression of the neuromodulator DAF-7/TGF-β. DAF-7, in turn, via canonical TGF-β signaling in adjacent interneurons modulates aerotaxis behavior and promote avoidance of pathogenic *P. aeruginosa*[40]. This chapter highlights the utility of activity-guided fractionation techniques to elucidate metabolite interplay in a model host-pathogen system. The study provides a chemical, genetic and neuronal basis for how the behavior and physiology of a simple animal host can be modified by the microbial environment and suggest that secondary metabolites produced by microbes may provide environmental cues that contribute to pathogen recognition and host survival.

Developmental arrest as non-feeding and stress-resistant dauer larvae represents a major survival and dispersal strategy in nematodes. Detailed studies in *C. elegans* and *P. pacificus* have previously revealed that small molecule signaling in dauer formation changes rapidly in evolution resulting in extreme structural diversity of small-molecule compounds[22, 31]. In *C. elegans*, a blend

of simple ascarosides constitutes the dauer pheromone, whereas the *P. pacificus* dauer pheromone includes additional paratosides and integrates building blocks from diverse primary metabolic pathways. In Chapter 2, the author shows that genes encoding enzymes of the peroxisomal β-oxidation pathway involved in small molecule biosynthesis evolve rapidly, including gene duplications and domain switching. The thiolase *daf-22*, the most downstream factor in *C. elegans* peroxisomal β-oxidation, has duplicated in *P. pacificus*, resulting in *Ppa-daf-22.1* and *Ppa-daf-22.2*[38]. Under well-fed conditions, ascaroside biosynthesis proceeds exclusively via *Ppa-daf-22.1*. In contrast, starvation conditions induce *Ppa-daf-22.2* activity, resulting in the production of a specific subset of ascarosides. This study reveals an unexpected functional complexity of chemical communication in dauer development and evolution.

C. elegans uses simple building blocks from primary metabolism and a strategy of modular assembly to build a great diversity of signaling molecules, the ascarosides, which function as a chemical language in this model organism. In the ascarosides, the dideoxysugar ascarylose serves as a scaffold to which diverse moieties from lipid, amino acid, neurotransmitter, and nucleoside metabolism are attached. However, the mechanisms that underlie the highly specific assembly of the modular ascarosides are not understood. It is important to understand how the different modules get combined to form the complete molecules because that underlies the specificity of the biological message conveyed by each ascaroside. In Chapter 3, the author shows that the acyl-CoA synthetase ACS-7, which localizes to lysosome-related organelles, is specifically required for the attachment of different building blocks to the 4'-position of ascr#9[39]. ACS-7 is the first identified enzyme involved in the 4'-modification of ascarosides. Mutants lacking lysosome-related organelles are found to be defective in the production of all 4'-modified ascarosides, thus identifying this waste disposal system of the cell as a hotspot for the biosyntheses of signaling

molecules.

In Chapter 4, the author explores an LCMS-based comparative metabolomics technique to study the metabolome in an unbiased manner. While the above chapters focus on known metabolites such as the ascarosides, in this chapter the author ventures into the unknown to identify novel metabolites that could be potential biomarkers of aging in the metabolomes of mitochondrial mutants of *C. elegans*. These mutants of the electron transport chain exhibit peculiar lifespan extension or reduction phenotypes, and the author highlights metabolite families that correlate to this variation in longevity. Moreover, the author attempts to use MS/MS networking to study the metabolome of the mitochondria isolated from the rest of the cell.

Another untargeted comparative metabolomics study is discussed in Chapter 5, this time combining genome-wide association studies (GWAS) with metabolome annotation. The author describes efforts to explore and annotate the metabolomes of genetically diverse natural isolates of *C. elegans* whose genomes have been deep-sequenced to reveal an array of single nucleotide polymorphisms (SNPs). Correlating variation in abundances of different LC-MS 'features', which correspond to various known and unknown metabolites, with the nature, genomic location and frequency of occurrence of these SNPs can reveal candidate genes for the biosyntheses of these compounds, whose biological functions can be validated using genome-editing techniques like CRISPR/Cas9.

REFERENCES

1. Ritchie, M. D.; Holzinger, E. R.; Li, R.; Pendergrass, S. A.; Kim, D., Methods of integrating data to uncover genotype-phenotype interactions. *Nature reviews. Genetics* **2015,** *16* (2), 85-97.

2. Hulme, S. E.; Whitesides, G. M., Chemistry and the worm: *Caenorhabditis elegans* as a platform for integrating chemical and biological research. *Angew Chem Int Ed Engl* **2011,** *50* (21), 4774-807.

3. Consortium, C. e. S., Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* **1998,** *282* (5396), 2012-8.

4. O'Rourke, E. J.; Kuballa, P.; Xavier, R.; Ruvkun, G., omega-6 Polyunsaturated fatty acids extend life span through the activation of autophagy. *Genes & development* **2013,** *27* (4), 429-40.

5. Wang, M. C.; O'Rourke, E. J.; Ruvkun, G., Fat metabolism links germline stem cells and longevity in *C. elegans*. *Science* **2008,** *322* (5903), 957-60.

6. Wollam, J.; Magner, D. B.; Magomedova, L.; Rass, E.; Shen, Y.; Rottiers, V.; Habermann, B.; Cummins, C. L.; Antebi, A., A novel 3-hydroxysteroid dehydrogenase that regulates reproductive development and longevity. *PLoS Biol* **2012,** *10* (4), e1001305.

7. Shen, Y.; Wollam, J.; Magner, D.; Karalay, O.; Antebi, A., A steroid receptor-microRNA switch regulates life span in response to signals from the gonad. *Science* **2012,** *338* (6113), 1472-6.

8. Lee, S. S.; Schroeder, F. C., Steroids as central regulators of organismal development and lifespan. *PLoS Biol* **2012,** *10* (4), e1001307.

9. Fielenbach, N.; Antebi, A., C. elegans dauer formation and the molecular basis of plasticity. *Genes & development* **2008,** *22* (16), 2149-65.

10. Lucanic, M.; Held, J. M.; Vantipalli, M. C.; Klang, I. M.; Graham, J. B.; Gibson, B. W.; Lithgow, G. J.; Gill, M. S., N-acylethanolamine signalling mediates the effect of diet on lifespan in *Caenorhabditis elegans*. *Nature* **2011,** *473* (7346), 226-9.

11. Ackerman, D.; Gems, D., The mystery of *C. elegans* aging: an emerging role for fat. Distant parallels between C. elegans aging and metabolic syndrome? *BioEssays: news and reviews in molecular, cellular and developmental biology* **2012,** *34* (6), 466-71.

12. Kenyon, C. J., The genetics of ageing. *Nature* **2010,** *464* (7288), 504-12.

13. Watson, E.; MacNeil, L. T.; Arda, H. E.; Zhu, L. J.; Walhout, A. J. M., Integration of metabolic and gene regulatory networks modulates the *C. elegans* dietary response. *Cell* **2013,** *153* (1), 253-66.

14. Wittenburg, N.; Eimer, S.; Lakowski, B.; Rohrig, S.; Rudolph, C.; Baumeister, R., Presenilin is required for proper morphology and function of neurons in *C. elegans*. *Nature* **2000,** *406* (6793), 306-9.

15. Ogg, S.; Paradis, S.; Gottlieb, S.; Patterson, G. I.; Lee, L.; Tissenbaum, H. A.; Ruvkun, G., The Fork head transcription factor DAF-16 transduces insulin-like metabolic and longevity signals in *C. elegans*. *Nature* **1997,** *389* (6654), 994-9.

16. Artyukhin, A. B.; Yim, J. J.; Cheong Cheong, M.; Avery, L., Starvation-induced collective behavior in *C. elegans*. *Sci Rep* **2015,** *5*, 10647.

17. Dieterich, C.; Roeseler, W.; Sobetzko, P.; Sommer, R. J., Pristionchus.org: a genome-centric database of the nematode satellite species *Pristionchus pacificus*. *Nucleic acids research* **2007,** *35* (Database issue), D498-502.

18. Jenner, R. A.; Wills, M. A., The choice of model organisms in evo-devo. *Nature reviews. Genetics* **2007,** *8* (4), 311-9.

19. Srinivasan, J.; Kaplan, F.; Ajredini, R.; Zachariah, C.; Alborn, H. T.; Teal, P. E. A.; Malik, R. U.; Edison, A. S.; Sternberg, P. W.; Schroeder, F. C., A blend of small molecules regulates both mating and development in *Caenorhabditis elegans*. *Nature* **2008,** *454* (7208), 1115-1118.

20. Pungaliya, C.; Srinivasan, J.; Fox, B. W.; Malik, R. U.; Ludewig, A. H.; Sternberg, P. W.; Schroeder, F. C., A shortcut to identifying small molecule signals that regulate behavior and development in *Caenorhabditis elegans*. *Proc Natl Acad Sci U S A* **2009,** *106* (19), 7708-13.

21. Kaplan, F.; Srinivasan, J.; Mahanti, P.; Ajredini, R.; Durak, O.; Nimalendran, R.; Sternberg, P. W.; Teal, P. E.; Schroeder, F. C.; Edison, A. S.; Alborn, H. T., Ascaroside expression in *Caenorhabditis elegans* is strongly dependent on diet and developmental stage. *PLoS ONE* **2011,** *6* (3), e17804.

22. Bose, N.; Ogawa, A.; von Reuss, S. H.; Yim, J. J.; Ragsdale, E. J.; Sommer, R. J.; Schroeder, F. C., Complex small-molecule architectures regulate phenotypic plasticity in a nematode. *Angew Chem Int Ed Engl* **2012,** *51* (50), 12438-43.

23. Srinivasan, J.; von Reuss, S. H.; Bose, N.; Zaslaver, A.; Mahanti, P.; Ho, M. C.; O'Doherty, O. G.; Edison, A. S.; Sternberg, P. W.; Schroeder, F. C., A Modular Library of Small Molecule Signals Regulates Social Behaviors in *Caenorhabditis elegans*. *PLoS Biol* **2012,** *10* (1), e1001237.

24. Izrayelit, Y.; Srinivasan, J.; Campbell, S. L.; Jo, Y.; von Reuss, S. H.; Genoff, M. C.; Sternberg, P. W.; Schroeder, F. C., Targeted metabolomics reveals a male pheromone and sex-specific ascaroside biosynthesis in *Caenorhabditis elegans*. *ACS chemical biology* **2012,** *7* (8), 1321-5.

25. von Reuss, S. H.; Bose, N.; Srinivasan, J.; Yim, J. J.; Judkins, J. C.; Sternberg, P. W.; Schroeder, F. C., Comparative metabolomics reveals biogenesis of ascarosides, a modular library of small molecule signals in *C. elegans*. *J Am Chem Soc* **2012,** *134* (3), 1817–1824.

26. Choe, A.; von Reuss, S. H.; Kogan, D.; Gasser, R. B.; Platzer, E. G.; Schroeder, F. C.; Sternberg, P. W., Ascaroside signaling is widely conserved among nematodes. *Current biology : CB* **2012,** *22* (9), 772-80.

27. Artyukhin, A. B.; Yim, J. J.; Srinivasan, J.; Izrayelit, Y.; Bose, N.; von Reuss, S. H.; Jo, Y.; Jordan, J. M.; Baugh, L. R.; Cheong, M.; Sternberg, P. W.; Avery, L.; Schroeder, F. C., Succinylated octopamine ascarosides and a new pathway of biogenic amine metabolism in *Caenorhabditis elegans*. *The Journal of biological chemistry* **2013,** *288* (26), 18778-83.

28. Ludewig, A. H.; Izrayelit, Y.; Park, D.; Malik, R. U.; Zimmermann, A.; Mahanti, P.; Fox, B. W.; Bethke, A.; Doering, F.; Riddle, D. L.; Schroeder, F. C., Pheromone sensing regulates *Caenorhabditis elegans* lifespan and stress resistance via the deacetylase SIR-2.1. *Proc Natl Acad Sci U S A* **2013,** *110* (14), 5522-7.

29. Ludewig, A. H.; Schroeder, F. C., Ascaroside signaling in *C. elegans*. *WormBook* **2013**, 1-22.

30. Bose, N.; Meyer, J. M.; Yim, J. J.; Mayer, M. G.; Markov, G. V.; Ogawa, A.; Schroeder, F. C.; Sommer, R. J., Natural variation in dauer pheromone production and sensing supports intraspecific competition in nematodes. *Current biology : CB* **2014,** *24* (13), 1536-41.

31. von Reuss, S. H.; Schroeder, F. C., Combinatorial chemistry in nematodes: modular assembly of primary metabolism-derived building blocks. *Natural product reports* **2015,** *32* (7), 994-1006.

32. Manosalva, P.; Manohar, M.; von Reuss, S. H.; Chen, S.; Koch, A.; Kaplan, F.; Choe, A.; Micikas, R. J.; Wang, X.; Kogel, K.-H.; Sternberg, P. W.; Williamson, V. M.; Schroeder, F. C.; Klessig, D. F., Conserved nematode signalling molecules elicit plant defenses and pathogen resistance. **2015,** *6*, 7795.

33. Schroeder, F. C., Modular assembly of primary metabolic building blocks: a chemical language in *C. elegans*. *Chemistry & biology* **2015,** *22* (1), 7-16.

34. Yim, J. J.; Bose, N.; Meyer, J. M.; Sommer, R. J.; Schroeder, F. C., Nematode signaling molecules derived from multimodular assembly of primary metabolic building blocks. *Organic letters* **2015,** *17* (7), 1648-51.

35. Mahanti, P.; Bose, N.; Bethke, A.; Judkins, J. C.; Wollam, J.; Dumas, K. J.; Zimmerman, A. M.; Campbell, S. L.; Hu, P. J.; Antebi, A.; Schroeder, F. C., Comparative metabolomics reveals endogenous ligands of DAF-12, a nuclear hormone receptor, regulating *C. elegans* development and lifespan. *Cell metabolism* **2014,** *19* (1), 73-83.

36. Smith, C. A.; Want, E. J.; O'Maille, G.; Abagyan, R.; Siuzdak, G., XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem* **2006,** *78* (3), 779-87.

37. Aksenov, A. A.; da Silva, R.; Knight, R.; Lopes, N. P.; Dorrestein, P. C., Global chemical analysis of biology by mass spectrometry. *Nature Reviews Chemistry* **2017,** *1*, 0054.

38. Markov, G. V.; Meyer, J. M.; Panda, O.; Artyukhin, A. B.; Claassen, M.; Witte, H.; Schroeder, F. C.; Sommer, R. J., Functional Conservation and Divergence of daf-22 Paralogs in *Pristionchus pacificus* Dauer Development. *Mol Biol Evol* **2016,** *33* (10), 2506-14.

39. Panda, O.; Akagi, A. E.; Artyukhin, A. B.; Judkins, J. C.; Le, H. H.; Mahanti, P.; Cohen, S. M.; Sternberg, P. W.; Schroeder, F. C., Biosynthesis of Modular Ascarosides in *C. elegans*. *Angew Chem Int Ed Engl* **2017,** *56* (17), 4729-4733.

40. Meisel, J. D.; Panda, O.; Mahanti, P.; Schroeder, F. C.; Kim, D. H., Chemosensation of bacterial secondary metabolites modulates neuroendocrine signaling and behavior of *C. elegans*. *Cell* **2014,** *159* (2), 267-80.

CHAPTER 1: CHEMOSENSATION OF BACTERIAL SECONDARY METABOLITES MODULATES NEUROENDOCRINE SIGNALING AND BEHAVIOR OF C. ELEGANS

Reproduced in part with permission from:

"Chemosensation of bacterial secondary metabolites modulates neuroendocrine signaling and behavior of *C. elegans*"

Joshua D. Meisel, Oishika Panda, Parag Mahanti, Frank C. Schroeder, and Dennis H. Kim

Cell **2014**, *159 (2)*, 267-280. DOI: 10.1016/j.cell.2014.09.011

Introduction

The recognition of microbial pathogens and the corresponding danger they represent is essential for the survival of host organisms. Innate immune systems have evolved to respond to foreign structures derived from microbes, which help the host distinguish microbes from self. Both plants and animals possess pattern recognition receptors that serve to detect several different molecular signatures (commonly known as microbe-associated molecular patterns or MAMPs) associated with specific classes of microbes[1]. However, such molecular patterns do not necessarily help the host discriminate a microbe that is pathogenic from one that is commensal. Microbial molecular patterns such as lipopolysaccharide, flagellin etc. are found on pathogen and commensal alike. To distinguish between harmful and useful microbes, even relatively simple host organisms possess diverse sets of olfactory receptors, and therefore have the potential to detect and respond to a far greater set of relevant microbial molecules by chemosensation. Candidate receptors and specific bacterial cues that modulate host physiology and avoidance behavior have begun to be explored[2-4].

Caenorhabditis elegans is one such simple host organism that forages on decomposing

organic matter for bacterial food[5]. The presence of bacterial food affects diverse behaviors of *C. elegans* such as feeding, locomotion, thermotaxis, and aerotaxis[6-9], and differences in the species composition of the food supply can alter aspects of their physiology and behavior[10-13]. Pathogenic bacteria kill *C. elegans* and induce an aversive learning response[14] that promotes protective behavioral avoidance[2, 15-17]. Bacterial molecules including *Serratia marcescens* serrawettin and *Pseudomonas aeruginosa* quorum-sensing regulators have been implicated in the behavioral avoidance of bacterial lawns[2, 18]. As such, *C. elegans* is a useful model for dissecting the genetic and biochemical mechanisms underlying microbial discrimination in animal host-pathogen interactions.

Neuronal response in the host upon perception of pathogenic cues

Previous studies have shown that the canonical DAF-7/TGF-β signaling pathway is implicated in the avoidance behavior demonstrated by *C. elegans* in the presence of pathogenic *P. aeruginosa* as opposed to the standard laboratory diet of *Escherichia coli*[17]. Once secreted, the ligand DAF-7 binds to the TGF-β type I receptor DAF-1 and the TGF-β type II receptor DAF-4, which then act via the R-SMAD DAF-8 to antagonize the co-SMAD DAF-3[19-21] (Figure 1.1a). All *C. elegans* mutants of this pathway display enhanced susceptibility to infection by *P. aeruginosa*[22] (Figure 1.1b). Thus, the DAF-7 pathway is specifically required for avoidance of *P. aeruginosa*, functioning through the canonical DAF-3 signaling pathway to promote survival by limiting host exposure to pathogenic bacteria.

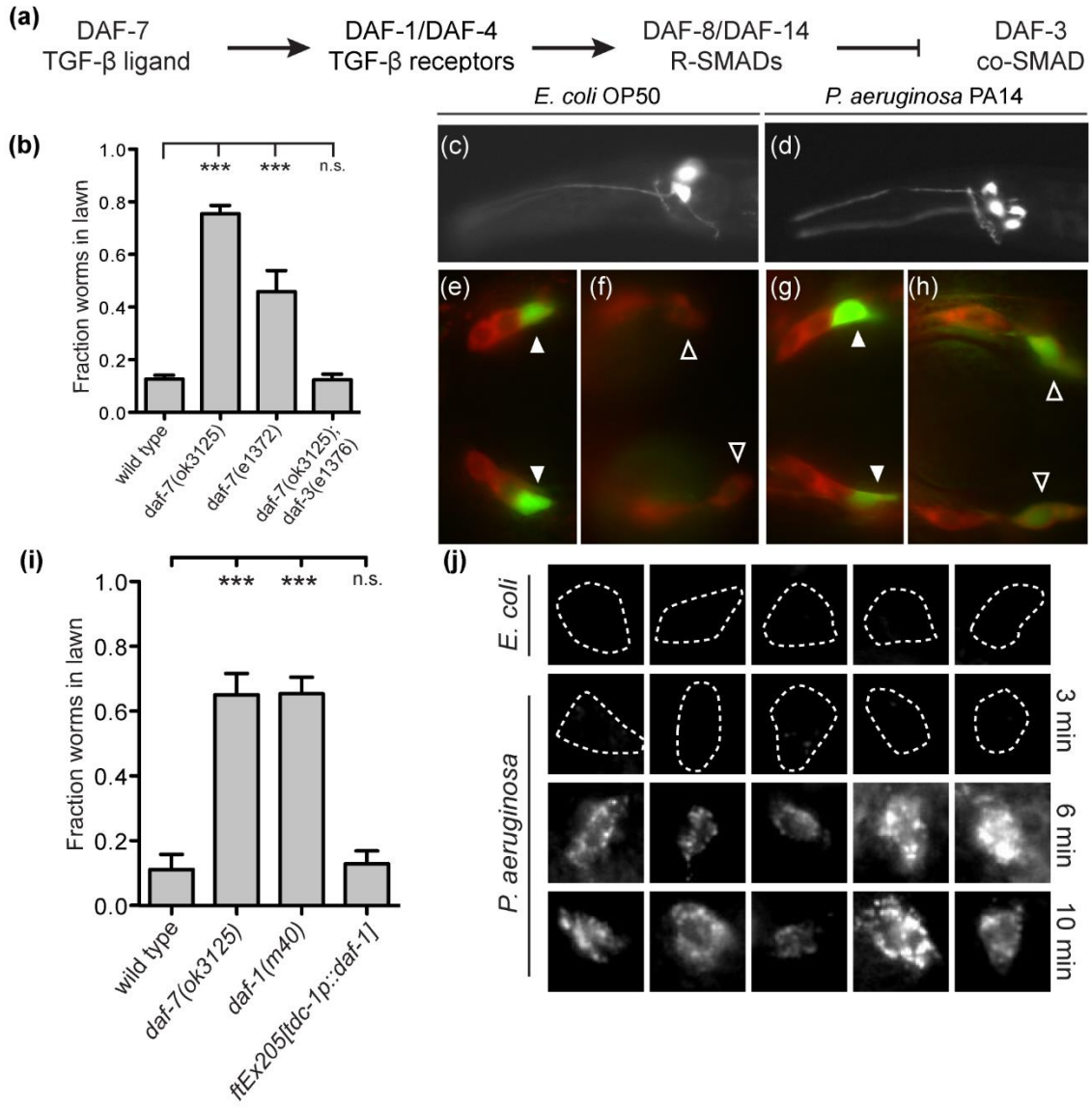

Figure 1.1. DAF-7 signaling and *P. aeruginosa* avoidance. (a) The canonical DAf-7/TGF-β signaling pathway in *C. elegans*. **(b)** Lawn occupancy of animals on *P. aeruginosa* after 15 hr, confirming that the DAF-7 pathway induces pathogen avoidance behavior in *C. elegans*. ***p < 0.001 as determined by one-way ANOVA, followed by Dunnett's multiple comparison test. ns, not significant. Values represent means of at least three independent experiments. Error bars indicate standard error. *daf-7p::gfp* expression pattern in *C. elegans* on *E. coli* **(c)** and *P. aeruginosa* **(d)**. Colocalization of *daf-7p::gfp* expression and amphid neuron-specific lipophilic dye DiI staining (red) in animals on *E. coli*, dorsal view **(e)** and ventral view **(f)**, and in animals on *P. aeruginosa*, dorsal view **(g)** and ventral view **(h)**. Filled triangles indicate ASI neurons; empty triangles indicate ASJ neurons. **(i)** Lawn occupancy of animals on *P. aeruginosa* after 15 hr. ***p < 0.001, **p < 0.01, and *p < 0.05 as determined by one-way ANOVA, followed by Dunnett's multiple comparison test. ns, not significant. Values represent means of at least three independent experiments. Error bars indicate standard deviation. *ftEx205[tdc-1p::daf-1]* indicates daf-1 expression induced under *tdc-1* promoter in *daf-1(m40)* background. **(j)** *daf-7* FISH in the ASJ neurons (colocalized with *trx-1p::gfp*) on *E. coli* and P. *aeruginosa* after various exposure times, showing rapid kinetics of this transcriptional response. Dashed lines indicate cell boundaries.

While the TGF-β ligand *daf-7* was previously shown to be expressed exclusively in the ASI pair of ciliated chemosensory neurons in *C. elegans*, Dr. Joshua D. Meisel (Kim Research Group, Massachusetts Institute of Technology) observed that exposure to *P. aeruginosa* unexpectedly induced expression of *daf-7* in the ASJ chemosensory neurons as well (Figure 1.1c-h). *daf-7* expression was not induced in the ASJ pair of neurons upon exposure to the *E. coli* strain OP50. Binding of the ASJ-produced DAF-7 ligand to the TGF-β type I receptor DAF-1 in the adjacent RIM/RIC interneurons was sufficient to induce pathogen avoidance behavior (Figure 1.1i). The observed specificity and rapid kinetics of this transcriptional response (Figure 1.1j), far faster than the kinetics of intestinal infection or aversive learning behavior[14, 23], suggested that the ASJ chemosensory neurons may be responding directly to specific *P. aeruginosa* cues.

Pathogen avoidance is promoted through the modulation of aerotaxis behavior

In the presence of bacterial food, the laboratory wild-type strain of *C. elegans* N2 does not exhibit a distinct oxygen preference, and does not avoid high atmospheric oxygen concentrations (i.e. 20% O_2). In contrast, *daf-7* mutants have altered aerotaxis behavior preferring oxygen concentrations near 8% and avoiding higher atmospheric oxygen levels[7]. The *P. aeruginosa* lawn is known to be hypoxic, and it has been shown earlier that altered oxygen preference can promote bacterial lawn avoidance in the nematodes[17, 24]. To determine whether a similar mechanism underlies the pathogen avoidance behavior promoted by *daf-7*, Dr. Joshua D. Meisel (Kim research group, Massachusetts Institute of Technology) performed a pathogen avoidance assay in an oxygen chamber at low oxygen concentrations, such that the surrounding environment would no longer induce hyperoxia avoidance. *daf-7* mutants show no avoidance defect relative to wild-type animals, supporting the hypothesis that *daf-7* mutants fail to avoid *P. aeruginosa* due to aberrant avoidance of atmospheric oxygen levels (Figure 1.2).

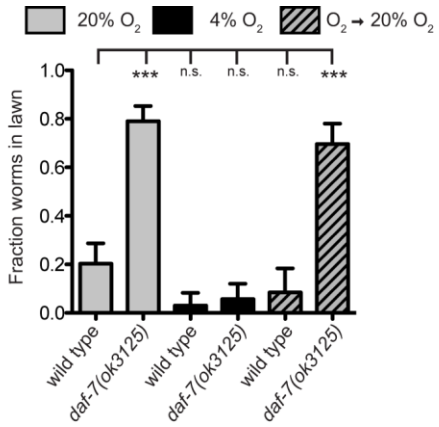

Figure 1.2. Altered aerotaxis behavior under *P. aeruginosa* influence. Altered aerotaxis behavior is reflected in the lawn occupancy of animals on *P. aeruginosa* shown here. All measurements taken after 15 hr unless otherwise noted. ***p < 0.001, **p < 0.01 as determined by oneway ANOVA followed by Tukey's Multiple Comparison Test. n.s. = not significant. Error bars indicate standard deviation.

Activity-guided fractionation leads to molecular cues produced by P. aeruginosa

The author sought to identify the specific bacterial cues inducing expression of *daf-7* in the ASJ neuron pair of *C. elegans*. Exposure of *C. elegans* to filtered *P. aeruginosa* supernatant was found to be sufficient to induce *daf-7* expression (Figure 1.3a), as evidenced by the expression of *pdaf-7::gfp* in the ASJ neurons of *C. elegans*. Testing supernatants from liquid cultures at various growth stages demonstrated that *P. aeruginosa* in stationary phase induced *daf-7* expression in ASJ neurons most dramatically (Figure 1.3a), suggesting that specific secondary metabolites produced under high cell density were being sensed by *C. elegans*. Metabolite identification was accomplished via activity-guided fractionation followed by 2D NMR spectroscopic profiling. Supernatant from large volumes of stationary phase *P. aeruginosa* strain PA14 was lyophilized and extracted to concentrate the compounds of interest. The extracts were then subjected to multiple rounds of fractionation using an automated, highly reproducible reverse phase chromatography system, and individual fractions were tested for activity in the *pdaf-7::gfp* assay. The fractions that induced *pdaf-7::gfp* expression ("active fractions") were further fractionated using an orthogonal, normal phase chromatography method to further simplify their composition,

enabling rapid identification of secondary metabolites. 1D and 2D NMR (dqfCOSY, HSQC, HMBC) spectroscopic data of the active fractions obtained thus were then analyzed. Such NMR spectroscopic analysis of metabolome fractions of relatively lowered complexity shifts focus onto testing individual identified components in bioassays, thereby reducing the need for further time-intensive purification[25].

2D NMR spectra of the active fractions revealed the presence of pantolactone[26], isovaleric acid[27], phenazine-1-carboxylic acid[28] (PCA), phenazine-1-carboxamide (PCN), pyochelin (see Appendix Figure A1) and phenyl acetic acid[29] as major components (Figure 1.3b), in addition to small quantities of monoacyl glycerides and β-hydroxy fatty acids. When synthetic and HPLC-purified samples of these compounds were tested, PCN and pyochelin displayed concentration-dependent activity in the *daf-7p::gfp* assay, (Appendix figure A2), whereas the other identified metabolites were not active in this assay. Furthermore, it was found that other *P. aeruginosa*-produced phenazines (e.g. pyocyanin and phenazine-1-carboxylic acid) and pyoverdine, another unrelated *P. aeruginosa* siderophore, do not induce *daf-7* expression in the ASJ neurons (Appendix figure A2b), indicating that general properties of these compounds are likely not the cause of their activity. The increase in *daf-7p::gfp* fluorescence in the ASJ neurons following addition of pyochelin and PCN was similar to that induced by *P. aeruginosa* supernatant (Appendix figure A2c). Interestingly, these compounds were found to have no effect on *daf-7p::gfp* fluorescence in the ASI neurons, further decoupling the responses of the ASJ and ASI neurons to *P. aeruginosa* (Appendix figure A2c).

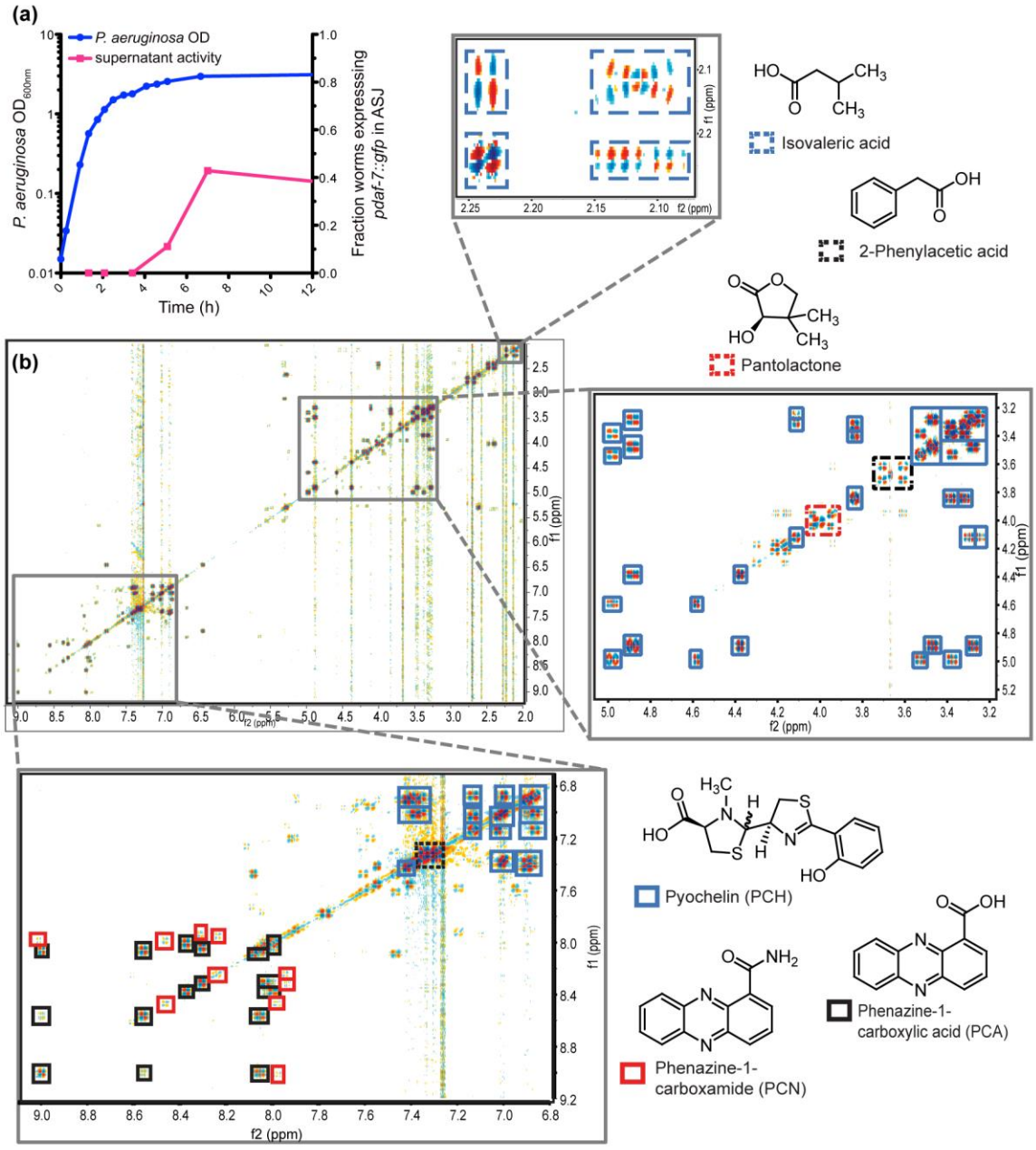

Figure 1.3. Identification of *P. aeruginosa* produced signals. (a) Growth curve of *P. aeruginosa* as measured by OD_{600nm} (blue circles) and activity of bacterial supernatant in inducing *daf-7p::gfp* expression in the ASJ neurons (magenta squares). **(b)** DQF-COSY spectrum of active metabolome fraction of *P. aeruginosa* media extract. Enlarged sections show cross-peaks of the most abundant metabolites present in this fraction.

To test the ability of PCN to activate the ASJ neurons, a transgenic strain was then constructed in which the genetically encoded calcium indicator GCaMP5[30] is expressed exclusively in the ASJ neuron pair. Upon administration of PCN, but not the carrier control

DMSO, increased GCaMP5 fluorescence was observed in both the ASJ cell body and ASJ ciliated projection that is exposed to the environment (Appendix figure A2d-f). These data suggest that ASJ may be sensing the *P. aeruginosa* metabolite PCN directly and, in turn, activating *daf-7* transcription.

PCN and pyochelin are secondary metabolites produced by *P. aeruginosa* at high cell density that promote biofilm formation in soil as well as chronic infections in human lungs[31-32]. In *P. aeruginosa* production of phenazines and pyochelin are both positively regulated by GacA, a global activator of cell-density-dependent gene expression and virulence[33-34]. The *P. aeruginosa gacA* mutant was found to be deficient in inducing *daf-7p::gfp* expression in the ASJ neurons relative to wild-type bacteria (Figure 1.4a). As such, these molecules may serve as bacterial growth-stage-specific cues for *C. elegans*, alerting the host to the presence of bacteria in a "pathogenic state," and inducing a correspondingly beneficial behavioral avoidance response.

To test the specificity of the induction of *daf-7* transcription in the ASJ neuron pair of to *P. aeruginosa* as compared to other environmental microbes, *daf-7p::gfp* animals were exposed to a wide array of bacteria species and strains, covering pathogenic and non-pathogenic alpha-, beta-, and gammaproteobacteria as well as gram-positive bacterial species. Several non-*E. coli* species can induce *daf-7* expression in the ASJ neuron pair on the order of 10-fold, but that the response is 1-2 orders of magnitude greater upon exposure to *P. aeruginosa* PA14 (Figure 1.4a). This result indicated that the identity of the microbial species was the principal determinant in inducing *daf-7* expression in the ASJ neuron pair, likely due to the relatively species-specific production of *P. aeruginosa* metabolites PCN and pyochelin. The presence of low-level activity in bacterial strains other than *P. aeruginosa* are consistent with the existence of additional, unidentified bacterial determinants that also contribute to the induction of *daf-7* in the ASJ neuron pair.

Interestingly, while several other bacterial species produce phenazines, such as *Burkholderia, Brevibacterium,* and *Streptomyces,* the modifying enzyme responsible for producing phenazine-1-carboxamide, *phzH*, is not part of the canonical phenazine operon, but rather found elsewhere in the genome and limited to *P. aeruginosa* and *P. chlororaphis*[35-36]. Similarly, pyochelin production is restricted to *Pseudomonas* and *Burkholderia* species[37]. As such, these two metabolites may act as pathogen-specific cues for *C. elegans* navigation, providing a molecular readout not only of the presence of *P. aeruginosa* itself, but the presence of *P. aeruginosa* in a "pathogenic state" that is more strongly associated with virulence.

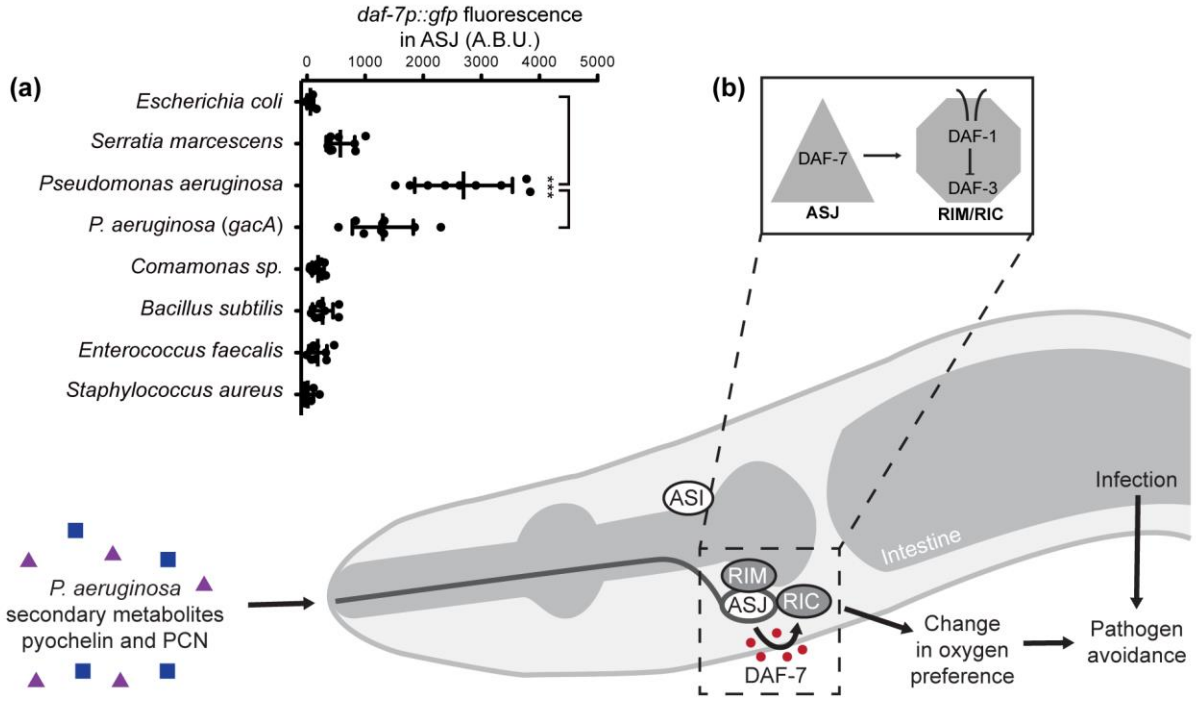

Figure 1.4. Pathogen avoidance in *C. elegans*. (a) Maximum fluorescence values of *daf-7p::gfp* in ASJ neurons after 16 hr exposure to indicated bacteria. ***p < 0.001 as determined by one-way ANOVA, followed by Tukey's multiple comparison test. Error bars indicate standard deviation. **(b)** In response to P. aeruginosa exposure, or P. aeruginosa metabolites phenazine-1-carboxamide (PCN) and pyochelin, daf-7 expression is activated in the ASJ neurons. Once secreted, DAF-7 signals to the TGF-β receptor DAF-1 the adjacent RIM/RIC interneurons. DAF-7/ TGF-β signaling acts to alter aerotaxis behavior and promote avoidance of pathogenic bacteria.

Conclusion

The author, in collaboration with Dr. Parag Mahanti (Schroeder research group, Cornell University), has demonstrated the use of the activity-guided fractionation approach in this study to enable the identification and characterization of specific compounds that elicit specific responses in biological assays, in a vast array of metabolites. While the results described here are a testimony to the usefulness of this technique, the approach relies heavily on the availability of a large amount of biomaterial to begin with, as every fractionation step leads to loss of some amount of the starting material. While this holds true for nematodes such as *C. elegans* and bacteria such as *P. aeruginosa* and *E. coli*, there are several species considered to be inherently "unculturable" under standard laboratory conditions, and as such cannot be used to generate a large amount of biomass. Molecules of interest may degrade during repeated fractionation steps, and therefore not show any activity in bioassays. In several cases, molecules of interest may also act synergistically in the natural system, making it difficult to reconstitute their activity *in vitro*.

Taken together, these results indicate that *C. elegans* respond to secondary metabolites produced by bacterial pathogens using signaling pathways in their chemosensory neurons (Figure 1.4b). In collaboration with Dr. Joshua D. Meisel (Kim research group, Massachusetts Institute of Technology), the author has shown that DAF-7 neuroendocrine signaling is necessary for the behavioral avoidance response to *P. aeruginosa*. *C. elegans* lack the antigen-specific responses of vertebrate immunity, but the utilization of chemosensory neurons and the repertoire of an estimated 1,300 GPCRs may provide a means by which a simple host organism can detect microbial pathogens. Bacterial secondary metabolism can generate a wide range of molecules that are often largely specific with regard to producer organism and regulated by environmental and growth conditions. Redox-active phenazine-1-carboxamide and the siderophore pyochelin are two

such secondary metabolites of *P. aeruginosa,* produced under conditions of high cell density and low oxygen. The decision to occupy or avoid a lawn of *P. aeruginosa* integrates multiple sensory inputs including chemosensation of bacterial compounds, chemosensation of oxygen, and mechanosensation[2, 15, 17, 38]. There has been an emerging appreciation of the profound influence that the animal microbiota can have on host organisms[39-40], and this study provides a genetic, neuronal, chemical basis for how microbes may influence host neuroendocrine physiology and behavior.

REFERENCES

1. Manosalva, P.; Manohar, M.; von Reuss, S. H.; Chen, S.; Koch, A.; Kaplan, F.; Choe, A.; Micikas, R. J.; Wang, X.; Kogel, K.-H.; Sternberg, P. W.; Williamson, V. M.; Schroeder, F. C.; Klessig, D. F., Conserved nematode signalling molecules elicit plant defenses and pathogen resistance. **2015,** *6*, 7795.

2. Pradel, E.; Zhang, Y.; Pujol, N.; Matsuyama, T.; Bargmann, C. I.; Ewbank, J. J., Detection and avoidance of a natural product from the pathogenic bacterium *Serratia marcescens* by *Caenorhabditis elegans*. *Proc Natl Acad Sci U S A* **2007,** *104* (7), 2295-300.

3. Riviere, S.; Challet, L.; Fluegge, D.; Spehr, M.; Rodriguez, I., Formyl peptide receptor-like proteins are a novel family of vomeronasal chemosensors. *Nature* **2009,** *459* (7246), 574-7.

4. Stensmyr, M. C.; Dweck, H. K.; Farhan, A.; Ibba, I.; Strutz, A.; Mukunda, L.; Linz, J.; Grabe, V.; Steck, K.; Lavista-Llanos, S.; Wicher, D.; Sachse, S.; Knaden, M.; Becher, P. G.; Seki, Y.; Hansson, B. S., A conserved dedicated olfactory circuit for detecting harmful microbes in *Drosophila*. *Cell* **2012,** *151* (6), 1345-57.

5. Felix, M. A.; Duveau, F., Population dynamics and habitat sharing of natural populations of *Caenorhabditis elegans* and *C. briggsae*. *BMC biology* **2012,** *10*, 59.

6. Avery, L.; Horvitz, H. R., Effects of starvation and neuroactive drugs on feeding in *Caenorhabditis elegans*. *The Journal of experimental zoology* **1990,** *253* (3), 263-70.

7. Chang, A. J.; Chronis, N.; Karow, D. S.; Marletta, M. A.; Bargmann, C. I., A distributed chemosensory circuit for oxygen preference in *C. elegans*. *PLoS Biol* **2006,** *4* (9), e274.

8. Hedgecock, E. M.; Russell, R. L., Normal and mutant thermotaxis in the nematode *Caenorhabditis elegans*. *Proc Natl Acad Sci U S A* **1975,** *72* (10), 4061-5.

9. Sawin, E. R.; Ranganathan, R.; Horvitz, H. R., *C. elegans* locomotory rate is modulated by the environment through a dopaminergic pathway and by experience through a serotonergic pathway. *Neuron* **2000,** *26* (3), 619-31.

10. Gusarov, I.; Gautier, L.; Smolentseva, O.; Shamovsky, I.; Eremina, S.; Mironov, A.; Nudler, E., Bacterial nitric oxide extends the lifespan of *C. elegans*. *Cell* **2013,** *152* (4), 818-30.

11. MacNeil, L. T.; Watson, E.; Arda, H. E.; Zhu, L. J.; Walhout, A. J., Diet-induced developmental acceleration independent of TOR and insulin in *C. elegans*. *Cell* **2013,** *153* (1), 240-52.

12. Shtonda, B. B.; Avery, L., Dietary choice behavior in *Caenorhabditis elegans*. *The Journal of experimental biology* **2006,** *209* (Pt 1), 89-102.

13. Watson, E.; MacNeil, L. T.; Ritter, A. D.; Yilmaz, L. S.; Rosebrock, A. P.; Caudy, A. A.; Walhout, A. J., Interspecies systems biology uncovers metabolites affecting *C. elegans* gene expression and life history traits. *Cell* **2014,** *156* (4), 759-70.

14. Zhang, Y.; Lu, H.; Bargmann, C. I., Pathogenic bacteria induce aversive olfactory learning in *Caenorhabditis elegans*. *Nature* **2005,** *438* (7065), 179-84.

15. Chang, H. C.; Paek, J.; Kim, D. H., Natural polymorphisms in *C. elegans* HECW-1 E3 ligase affect pathogen avoidance behaviour. *Nature* **2011,** *480* (7378), 525-9.

16. Pujol, N.; Link, E. M.; Liu, L. X.; Kurz, C. L.; Alloing, G.; Tan, M. W.; Ray, K. P.; Solari, R.; Johnson, C. D.; Ewbank, J. J., A reverse genetic analysis of components of the Toll signaling pathway in *Caenorhabditis elegans*. *Current biology : CB* **2001,** *11* (11), 809-21.

17. Reddy, K. C.; Andersen, E. C.; Kruglyak, L.; Kim, D. H., A polymorphism in npr-1 is a behavioral determinant of pathogen susceptibility in *C. elegans*. *Science* **2009,** *323* (5912), 382-4.

18. Beale, E.; Li, G.; Tan, M. W.; Rumbaugh, K. P., *Caenorhabditis elegans* senses bacterial autoinducers. *Applied and environmental microbiology* **2006,** *72* (7), 5135-7.

19. Estevez, M.; Attisano, L.; Wrana, J. L.; Albert, P. S.; Massague, J.; Riddle, D. L., The daf-4 gene encodes a bone morphogenetic protein receptor controlling *C. elegans* dauer larva development. *Nature* **1993,** *365* (6447), 644-9.

20. Georgi, L. L.; Albert, P. S.; Riddle, D. L., *daf-1*, a *C. elegans* gene controlling dauer larva development, encodes a novel receptor protein kinase. *Cell 61* (4), 635-645.

21. Patterson, G. I.; Koweek, A.; Wong, A.; Liu, Y.; Ruvkun, G., The DAF-3 Smad protein antagonizes TGF-beta-related receptor signaling in the *Caenorhabditis elegans* dauer pathway. *Genes & development* **1997,** *11* (20), 2679-90.

22. Meisel, J. D.; Panda, O.; Mahanti, P.; Schroeder, F. C.; Kim, D. H., Chemosensation of bacterial secondary metabolites modulates neuroendocrine signaling and behavior of *C. elegans*. *Cell* **2014,** *159* (2), 267-80.

23. Tan, M. W.; Mahajan-Miklos, S.; Ausubel, F. M., Killing of *Caenorhabditis elegans* by *Pseudomonas aeruginosa* used to model mammalian bacterial pathogenesis. *Proc Natl Acad Sci U S A* **1999,** *96* (2), 715-20.

24. Reddy, K. C.; Hunter, R. C.; Bhatla, N.; Newman, D. K.; Kim, D. H., *Caenorhabditis elegans* NPR-1-mediated behaviors are suppressed in the presence of mucoid bacteria. *Proc Natl Acad Sci U S A* **2011,** *108* (31), 12887-92.

25. Taggi, A. E.; Meinwald, J.; Schroeder, F. C., A new approach to natural products discovery exemplified by the identification of sulfated nucleosides in spider venom. *J Am Chem Soc* **2004,** *126* (33), 10364-9.

26. Nakata, K.; Gotoh, K.; Ono, K.; Futami, K.; Shiina, I., Kinetic resolution of racemic 2-hydroxy-gamma-butyrolactones by asymmetric esterification using diphenylacetic acid with pivalic anhydride and a chiral acyl-transfer catalyst. *Organic letters* **2013,** *15* (6), 1170-3.

27. Niu, D.-F.; Xiao, L.-P.; Zhang, A.-J.; Zhang, G.-R.; Tan, Q.-Y.; Lu, J.-X., Electrocatalytic carboxylation of aliphatic halides at silver cathode in acetonitrile. *Tetrahedron* **2008,** *64* (46), 10517-10520.

28. Mehnaz, S.; Saleem, R. S.; Yameen, B.; Pianet, I.; Schnakenburg, G.; Pietraszkiewicz, H.; Valeriote, F.; Josten, M.; Sahl, H. G.; Franzblau, S. G.; Gross, H., Lahorenoic acids A-C, ortho-dialkyl-substituted aromatic acids from the biocontrol strain *Pseudomonas aurantiaca* PB-St2. *Journal of natural products* **2013,** *76* (2), 135-41.

29. Korsager, S.; Taaning, R. H.; Skrydstrup, T., Effective palladium-catalyzed hydroxycarbonylation of aryl halides with substoichiometric carbon monoxide. *J Am Chem Soc* **2013,** *135* (8), 2891-4.

30. Akerboom, J.; Chen, T. W.; Wardill, T. J.; Tian, L.; Marvin, J. S.; Mutlu, S.; Calderon, N. C.; Esposti, F.; Borghuis, B. G.; Sun, X. R.; Gordus, A.; Orger, M. B.; Portugues, R.; Engert, F.; Macklin, J. J.; Filosa, A.; Aggarwal, A.; Kerr, R. A.; Takagi, R.; Kracun, S.; Shigetomi, E.; Khakh, B. S.; Baier, H.; Lagnado, L.; Wang, S. S.; Bargmann, C. I.; Kimmel, B. E.; Jayaraman, V.; Svoboda, K.; Kim, D. S.; Schreiter, E. R.; Looger, L. L., Optimization of a GCaMP calcium indicator for neural activity imaging. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **2012,** *32* (40), 13819-40.

31. Cornelis, P.; Dingemans, J., *Pseudomonas aeruginosa* adapts its iron uptake strategies in function of the type of infections. *Frontiers in cellular and infection microbiology* **2013,** *3*, 75.

32. Price-Whelan, A.; Dietrich, L. E.; Newman, D. K., Rethinking 'secondary' metabolism: physiological roles for phenazine antibiotics. *Nat Chem Biol* **2006,** *2* (2), 71-8.

33. Reimmann, C.; Beyeler, M.; Latifi, A.; Winteler, H.; Foglino, M.; Lazdunski, A.; Haas, D., The global activator GacA of *Pseudomonas aeruginosa* PAO positively controls the production of the autoinducer N-butyryl-homoserine lactone and the formation of the virulence factors pyocyanin, cyanide, and lipase. *Molecular microbiology* **1997,** *24* (2), 309-19.

34. Wei, X.; Huang, X.; Tang, L.; Wu, D.; Xu, Y., Global control of GacA in secondary metabolism, primary metabolism, secretion systems, and motility in the rhizobacterium *Pseudomonas aeruginosa* M18. *Journal of bacteriology* **2013,** *195* (15), 3387-400.

35. Chin, A. W. T. F.; Thomas-Oates, J. E.; Lugtenberg, B. J.; Bloemberg, G. V., Introduction of the phzH gene of *Pseudomonas chlororaphis* PCL1391 extends the range of biocontrol ability of phenazine-1-carboxylic acid-producing *Pseudomonas* spp. strains. *Molecular plant-microbe interactions : MPMI* **2001,** *14* (8), 1006-15.

36. Mavrodi, D. V.; Bonsall, R. F.; Delaney, S. M.; Soule, M. J.; Phillips, G.; Thomashow, L. S., Functional analysis of genes for biosynthesis of pyocyanin and phenazine-1-carboxamide from *Pseudomonas aeruginosa* PAO1. *Journal of bacteriology* **2001,** *183* (21), 6454-65.

37. Gross, H.; Loper, J. E., Genomics of secondary metabolite production by *Pseudomonas* spp. *Natural product reports* **2009,** *26* (11), 1408-46.

38. Ha, H. I.; Hendricks, M.; Shen, Y.; Gabel, C. V.; Fang-Yen, C.; Qin, Y.; Colon-Ramos, D.; Shen, K.; Samuel, A. D.; Zhang, Y., Functional organization of a neural network for aversive olfactory learning in *Caenorhabditis elegans*. *Neuron* **2010,** *68* (6), 1173-86.

39. Clemente, J. C.; Ursell, L. K.; Parfrey, L. W.; Knight, R., The impact of the gut microbiota on human health: an integrative view. *Cell* **2012,** *148* (6), 1258-70.

40. Lyte, M., Microbial endocrinology in the microbiome-gut-brain axis: how bacterial production and utilization of neurochemicals influence behavior. *PLoS pathogens* **2013,** *9* (11), e1003726.

CHAPTER 2: FUNCTIONAL CONSERVATION AND DIVERGENCE OF DAF-22 PARALOGS IN DAUER DEVELOPMENT OF *P. PACIFICUS*

Reproduced in part with permission from:

"Functional conservation and divergence of DAF-22 paralogs in dauer development of *P. pacificus*"

Gabriel V. Markov, Jan M. Meyer, Oishika Panda, Alexander B. Artyukhin, Marc Claaßen, Hanh Witte, Frank C. Schroeder, and Ralf J. Sommer

Molecular Biology and Evolution **2016**, *33 (10)*, 2506-2514. DOI: 10.1093/molbev/msw090

Introduction

The nematode species *Caenorhabditis elegans* and *Pristionchus pacificus* have become widely used models to study the chemical nature of small molecule communication and its mechanisms of action and evolution[1]. A vast modular library of small-molecule signals, the ascarosides, regulate several important behaviors in these nematodes, such as mating, aggregation and dispersal[2-6]. Ascarosides derive from the combination of the dideoxysugar ascarylose with a variety of fatty acid-like side chains. Like most small molecule pheromones, ascarosides usually function as a blend and often act in a concentration-dependent manner. These molecules are also major regulators of developmental processes, for example dauer formation. Dauer larvae are arrested, stress-resistant and long-lived alternative juveniles that are characterized by many specific morphological, physiological and behavioral adaptations[7] (Figure 2.1a). The dauer stage represents the major survival and dispersal strategy in nematodes, and many nematode species are regularly found in the wild in the dauer stage[8].

Chemical communication systems using small molecules are highly evolvable and can

likely be modified throughout evolution to generate novel types of interactions. Indeed, comparative studies among different nematode species indicated substantial diversification of the composition of small molecules (Figure 2.1b) and the involvement of small-molecule pheromones in novel types of interactions. One organism that has been studied in greater detail in this regard is *Pristionchus pacificus. P. pacificus* was developed as a second nematode model system with a specific focus on comparative and evolutionary biology[8]. The chemical architecture of *P. pacificus* small-molecule pheromones is surprisingly diverse, containing ascarosides as well as paratosides,

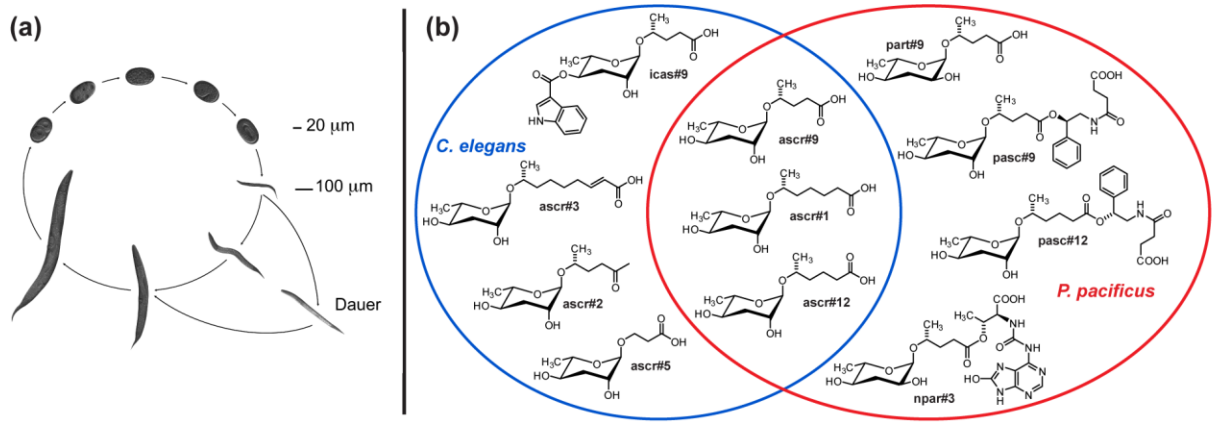

Figure 2.1. The *P. pacificus* ascarosides. (a) The life cycle of the nematode *P. pacificus*. Animals stay in the direct life cycle indefinitely if sufficient food is provided. In contrast, animals enter the dauer stage under harsh environmental conditions, such as elevated temperature, starvation, or high population density. **(b)** Selected ascarosides and ascaroside-derivatives identified from *C. elegans* and *P. pacificus*, respectively, illustrating their structural diversity and common biosynthetic origin from modular assembly of primary metabolic building blocks, including a peroxisomal β-oxidation-derived side chain. Although different compounds induce dauer in the two species, some simple short-chained ascarosides, for example, ascr#1 and

which are derived from the dideoxysugar paratose[4] (Figure 2.1a). Analysis of the ascaroside profiles of *P. pacificus*, *C. elegans* and other nematodes showed that the chemical structures of these small molecule signals are highly species-specific[4, 9].

Despite the complexity of ascaroside structures and functions, little is known about the biosynthesis of these small molecules in nematodes other than *C. elegans*. Recent studies identified and characterized the peroxisomal β-oxidation pathway to be involved in ascaroside biosynthesis in *C. elegans*[10-14]. Peroxisomal β-oxidation in *C. elegans* proceeds iteratively via a four-step

process that involves enzymes belonging to four different families (See Chapter 3 Figure 3.1c). All of them are members of multigenic families that are present across eukaryotes in varying numbers[15], raising the possibility that multiple paralogous genes from the same family contribute to ascaroside biosynthesis. Indeed, already three *acox* gene paralogs were recently shown to be involved in the synthesis of dauer-inducing ascarosides in *C. elegans*[14]. As *P. pacificus* and *C. elegans* are separated by more than 200 million years of evolution, these nematodes have largely divergent genomes and the *P. pacificus* genome was shown to encode more than 26,000 genes, around 6,000 more than *C. elegans*[16]. Interestingly, our previous work indicated that enzyme-encoding genes, particularly those that are members of multigene families, undergo rapid birth and death processes resulting in the near absence of 1:1 orthology relationships between *P. pacificus* and *C. elegans*[17]. The diversity of small molecules identified in *P. pacificus* and *C. elegans* merits an extensive study of the biosynthesis of nematode-derived small molecules by combining bioinformatic, genetic and chemical tools.

Ascaroside biosynthetic genes are non-orthologous between C. elegans and P. pacificus

To study the biosynthesis of small molecules in *P. pacificus*, Dr. Gabriel V. Markov and Dr. Jan M. Meyer (Sommer research group, Max Planck Institute for Developmental Biology) first evaluated the conservation of the putative peroxisomal β-oxidation pathway among nematodes based on orthology relationships. Analysis of the phylogenetic relationships of all closer paralogs of *C. elegans* enzymes involved in β-oxidation (*acox-1-6*, *maoc-1*, *dhs-28* and *daf-22*) revealed that none of them is conserved as a 1:1 ortholog between *P. pacificus* and *C. elegans*[18]. Further studies in *P. pacificus* ascaroside biosynthesis were then focused on the *daf-22* paralogs, which encode the most downstream acting enzymes in the *C. elegans* pathway. *Cel-daf-22* contains two thiolase domains, but lacks a SCP-2 domain that is known from thiolase genes in vertebrates,

insects and fungi[19]. Interestingly, *daf-22* is present as a single copy gene bearing a C-terminal SCP-2 domain in a diversity of nematodes including parasites of clade III (*A. suum*, *B. malayi* and *L. loa*), clade IV (*S. ratti* and *M. hapla*) and clade V (*N. americanus* and *H. contortus*) (Supplementary fig. S5). These findings suggest that the ancestral domain organization has been lost in *C. elegans* and some other nematode lineages.

In the genus *Pristionchus*, gene duplication led to two genes, *daf-22.1* and *daf-22.2*, the latter of which lost the SCP-2 domain secondarily (Appendix Figure B1a). Both paralogs differ in their chromosomal location and in their exon-intron structure (Appendix Figure B1b). *Ppa-daf-22.1* is expressed in dauer larvae, late larval stages and adults, and to a lesser extent in early larvae[20]. In contrast, *Ppa-daf-22.2* is only weakly expressed in dauers, and no expression was detectable in RNAseq experiments in other stages. Using quantitative PCR (qPCR) experiments, expression of *Ppa-daf-22.2* was observed at low levels, but *Ppa-daf-22.1* was expressed roughly 35x higher than *Ppa-daf-22.2* under well-fed conditions (Appendix Figure B2a). To study the function of the *Ppa-daf-22* genes, Dr. Gabriel V. Markov (Sommer research group, Max Planck Institute for Developmental Biology) separately inactivated the two *P. pacificus* paralogs by CRISPR-Cas9-induced deletion in the *N*-terminal part of the thiolase domain, ensuring that the catalytic activity of the encoded enzyme would be fully abolished (Appendix Figure B1b). Two mutant lines were obtained, a seven bp deletion in *Ppa-daf-22.1* (*tu489*) and a seven bp insertion in *Ppa-daf-22.2* (*tu504*), both resulting in frame shift mutations. A double mutant was generated to study the effect of knockout of both genes.

Ppa-daf-22.1 (tu489) and Ppa-daf-22.1; Ppa-daf-22.2 double mutants are deficient in ascaroside biosynthesis

To investigate the roles of the two *Ppa-daf-22* genes in ascaroside and paratoside biosynthesis in

P. pacificus, the author, in collaboration with Dr. Alexander B. Artyukhin (Schroeder research group, Cornell University), compared the metabolomes of both single mutants and *Ppa-daf-22.1; Ppa-daf-22.2* double mutants with that of wild type animals. Given that ascaroside production in *C. elegans* is known to depend on environmental conditions[11, 21], sample sets derived from two different culture protocols, in which worm liquid cultures were either fed with bacteria continuously ("*ad lib*" condition) or starved for an extended period of time prior to harvest ("starved" condition) were analyzed. Targeted LC/MS analysis of *ad lib* mixed-stage as well as synchronized cultures showed that short-chain ascaroside and paratoside biosynthesis is fully abolished in the double mutant *Ppa-daf-22.1; Ppa-daf-22.2* (Figure 2.2a, b). Analysis of the single mutants revealed similar ascaroside profiles for *Ppa-daf-22.2 (tu504)* and wild type (Figure 2.2a, b), whereas *Ppa-daf-22.1 (tu489)* was found to produce only trace amounts of a few specific ascarosides. Accumulation of large amounts of long-chain ascarosides as shunt metabolites was observed in the *Ppa-daf-22.1; Ppa-daf-22.2* double mutant and *Ppa-daf-22.1 (tu489)*. In contrast, the amounts of long-chain ascarosides in *Ppa-daf-22.2 (tu504)* were only slightly elevated compared to wild type (Figure 2.2c, d). These results indicate that in *ad lib* worms, ascaroside and paratoside biosynthesis proceeds almost exclusively via *Ppa*-DAF-22.1, whereas *Ppa*-DAF-22.2 does not significantly contribute.

In contrast, analysis of metabolome samples from starved cultures revealed that *Ppa*-DAF-22.2 can produce substantial amounts of certain ascarosides. Whereas the ascaroside profiles of starved wild type and *Ppa-daf-22.2 (tu504)* cultures remained largely unchanged from

those observed for *ad lib* cultures, starved *Ppa-daf-22.1 (tu489)* worms produced near-wild type levels of the ascaroside pasc#12, corresponding to at least 20-fold up-regulation of this compound relative to the amounts observed in *ad lib Ppa-daf-22.1 (tu489)* worms (Figure 2.2b). In addition,

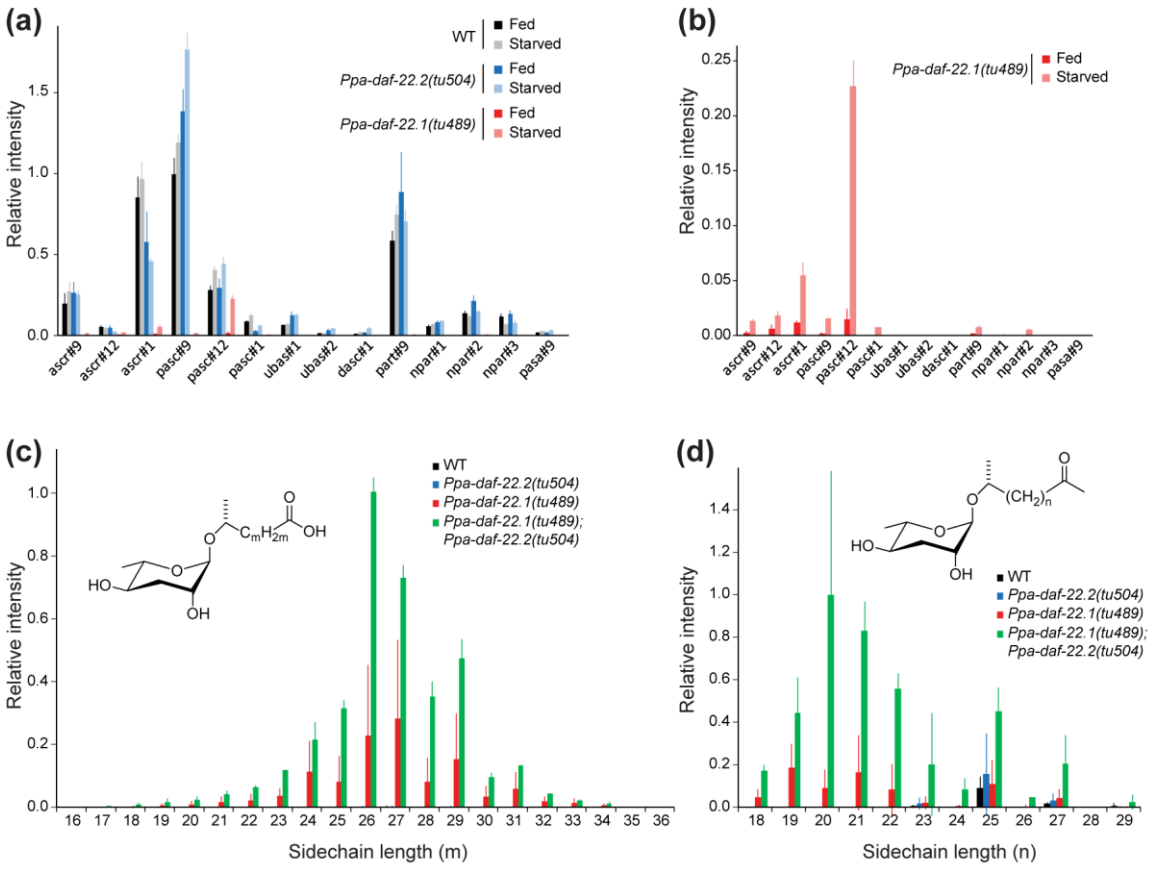

Figure 2.2. Ascaroside profiles of *Ppa-daf-22.1* and *Ppa-daf-22.2*. **(a)** Ascaroside profiles determined by LC-MS analysis from *ad lib* and starved mixed-stage liquid cultures of *P. pacificus* wild-type and *daf-22* mutants (worm media extracts, exo-metabolomes). **(b)** Enlargement of data for *Ppa-daf-22.1(tu489)* shown in **(a)** to better visualize the effect of starvation on ascaroside production in this strain. Error bars in **(a)** and **(b)** are SD, n = 3. Relative intensities in **(a, b)** are normalized to the peak area of pasc#9 in the wild-type strain in the fed condition. **(c, d)** Long-chain carboxylic acid ascarosides **(c)** and methylketone ascarosides **(d)** accumulate in worm pellets (endo-metabolome) of the *Ppa-daf-22.1(tu489)* mutant and *Ppa-daf-22.1(tu489); Ppa-daf-22.2(tu504)* double mutant. No short-chain ascarosides were found in *Ppa-daf-22.1; Ppa-daf-22.2* double mutant. Error bars in **(c, d)** are SD, n = 2. Relative intensities in **(c, d)** are normalized to the peak area of the compound with the largest peak area in each panel.

the production of several other ascarosides, including ascr#1 and ascr#12, was up-regulated in starved *Ppa-daf-22.1 (tu489)* cultures, though less markedly. The double mutant *Ppa-daf-22.1; Ppa-daf-22.2* did not produce any ascarosides under starvation conditions, indicating that

upregulation of ascaroside production in *Ppa-daf-22.1 (tu489)* in these conditions is due to *Ppa-daf-22.2* rather than some other, yet unidentified gene (Figure 2.2c, d). Notably, the chemical structures of pasc#12 and ascr#12 feature a 6-carbon carboxylic acid side chain, whereas all other major ascarosides in *P. pacificus* and *C. elegans* feature fatty acid side chains with an odd number of carbons. However, *Ppa*-DAF-22.2 does not seem to be exclusively involved in the production of ascarosides with a 6-carbon side chain, since we also observed up-regulation of ascr#1 (7-carbon side chain) and ascr#9 (5-carbon side chain) production in starved *Ppa-daf-22.1 (tu489)* cultures (Figure 2.2b).

Similar to *Cel-daf-22* worms[10], along with disappearance of short-chain ascarosides, the author observed build-up of very long-chain ascarosides and their corresponding methylketones as shunt metabolites in *Ppa-daf-22.1 (tu489)* (Figure 2.2c, d). These shunt metabolites are formed when very long-chain ascaroside precursors cannot be processed normally due to lack of thiolase enzymatic activity. This indicates that the *C. elegans* and *P. pacificus daf-22* genes catalyze similar biochemical processes.

qPCR analysis revealed no significant difference in *Ppa-daf-22.2* mRNA levels between fed and starved conditions (Appendix figure B2b). This suggests that induction of ascaroside biosynthesis in *Ppa-daf-22.1 (tu489)* mutant animals under starved conditions is regulated at the posttranscriptional level, or that starvation affects expression of other genes upstream of *daf-22* in the β-oxidation pathway. Interestingly, a reciprocal upregulation was found, resulting in a two-fold (*Ppa-daf-22.1*) and 3-4-fold (*Ppa-daf-22.2*) increase of the remaining gene (Appendix figure B2b).

Extracts from Ppa-daf-22.1 (tu489) and Ppa-daf-22.1; Ppa-daf-22.2 double mutants do not induce dauers

In *C. elegans*, knockout of *Cel-daf-22* results in loss of all dauer-inducing activity of liquid culture metabolome extracts[22]. In collaboration with the Sommer research group at the Max Planck Institute for Developmental Biology, the author found that metabolome extracts of *Ppa-daf-22.1(tu489)* and *Ppa-daf-22.1; Ppa-daf-22.2 ad lib* worms had no dauer-inducing activity in *P. pacificus* wild type animals, whereas extracts from *Ppa-daf-22.2(tu504) ad lib* worms retained most dauer-inducing activity compared to that of wild type extracts (Appendix figure B2c, d). Considering that production of short-chain ascarosides and paratosides is largely abolished in both *Ppa-daf-22.1(tu489)* and *Ppa-daf-22.1; Ppa-daf-22.2* double mutant *ad lib* cultures, these results indicate that short-chain ascarosides and paratosides are required for dauer induction in *P. pacificus*. Therefore, it appears that *Ppa*-DAF-22.1 is primarily responsible for the generation of dauer-inducing pheromones in *P. pacificus* under our *ad lib* conditions.

Conclusion

This study reveals that 1:1 orthologs of genes encoding the enzymes of peroxisomal β-oxidation are nearly absent above the genus level, with some genes showing even high birth and death rates within the genera *Caenorhabditis* and *Pristionchus*. This pattern is consistent with previous observations on the evolution of nematode detoxification-encoding enzymes[20]. Domain shuffling is observed for multiple genes in the β-oxidation pathway, causing difficulties in the assignment of enzymatic function to individual proteins. Reverse genetic approaches by CRISPR/Cas9 engineered knockout mutants of individual paralogs of the thiolase *daf*-22 have been used in this study to overcome such hurdles.

The analysis of *Ppa-daf-22* single and double mutants and the comparison of fed and

starved cultures revealed another level of regulation of peroxisomal β-oxidation. Under well-fed (*ad lib*) conditions mixed-stage or synchronized cultures of *Ppa-daf-22.1 (tu489),* like *Ppa-daf-22.1; Ppa-daf-22.2* double mutants, do not produce significant amounts of short-chain ascarosides and paratosides. In contrast, under starved conditions, *Ppa-daf-22.1 (tu489)*, mutants produce near-wild type levels of a specific subset of ascarosides, indicating that a branch of peroxisomal β-oxidation involving DAF-22.2 becomes active under starvation conditions. This activity may require posttranscriptional regulation or upstream regulatory input, as *Ppa-daf-22.2* expression is not significantly upregulated under starvation.

Taken together, these results show that orthology is a useful predictor of biosynthetic function of enzymes, but that metabolic pathways may diverge as a result of evolutionary changes that interact with other regulatory mechanisms. This partial uncoupling between the evolution of enzymes and the evolution of metabolic pathways implies that only interdisciplinary approaches of bioinformatics, genetic engineering, metabolomics, chemistry and molecular biology can reveal the full scope of evolutionary alterations and their organismic consequences. Metabolomics is maturating as a research field that can provide such information and has already provided important results in a diversity of organisms including land plants, nematodes and diatoms[23].

REFERENCES

1. Schroeder, F. C., Modular assembly of primary metabolic building blocks: a chemical language in *C. elegans*. *Chemistry & biology* **2015,** *22* (1), 7-16.

2. Srinivasan, J.; Kaplan, F.; Ajredini, R.; Zachariah, C.; Alborn, H. T.; Teal, P. E.; Malik, R. U.; Edison, A. S.; Sternberg, P. W.; Schroeder, F. C., A blend of small molecules regulates both mating and development in *Caenorhabditis elegans*. *Nature* **2008,** *454* (7208), 1115-8.

3. Pungaliya, C.; Srinivasan, J.; Fox, B. W.; Malik, R. U.; Ludewig, A. H.; Sternberg, P. W.; Schroeder, F. C., A shortcut to identifying small molecule signals that regulate behavior and development in *Caenorhabditis elegans*. *Proc Natl Acad Sci U S A* **2009,** *106* (19), 7708-13.

4. Bose, N.; Ogawa, A.; von Reuss, S. H.; Yim, J. J.; Ragsdale, E. J.; Sommer, R. J.; Schroeder, F. C., Complex small-molecule architectures regulate phenotypic plasticity in a nematode. *Angew Chem Int Ed Engl* **2012,** *51* (50), 12438-43.

5. Ludewig, A. H.; Izrayelit, Y.; Park, D.; Malik, R. U.; Zimmermann, A.; Mahanti, P.; Fox, B. W.; Bethke, A.; Doering, F.; Riddle, D. L.; Schroeder, F. C., Pheromone sensing regulates *Caenorhabditis elegans* lifespan and stress resistance via the deacetylase SIR-2.1. *Proc Natl Acad Sci U S A* **2013,** *110* (14), 5522-7.

6. Ludewig, A. H.; Schroeder, F. C., Ascaroside signaling in *C. elegans*. *WormBook* **2013**, 1-22.

7. Sommer, R. J.; Ogawa, A., Hormone signaling and phenotypic plasticity in nematode development and evolution. *Current biology : CB* **2011,** *21* (18), R758-66.

8. Sommer, R. J.; Mayer, M. G., Toward a synthesis of developmental biology with evolutionary theory and ecology. *Annual review of cell and developmental biology* **2015,** *31*, 453-471.

9. Choe, A.; von Reuss, S. H.; Kogan, D.; Gasser, R. B.; Platzer, E. G.; Schroeder, F. C.; Sternberg, P. W., Ascaroside signaling is widely conserved among nematodes. *Current biology : CB* **2012,** *22* (9), 772-80.

10. Izrayelit, Y.; Srinivasan, J.; Campbell, S. L.; Jo, Y.; von Reuss, S. H.; Genoff, M. C.; Sternberg, P. W.; Schroeder, F. C., Targeted metabolomics reveals a male pheromone and sex-specific ascaroside biosynthesis in *Caenorhabditis elegans*. *ACS chemical biology* **2012,** *7* (8), 1321-5.

11. von Reuss, S. H.; Bose, N.; Srinivasan, J.; Yim, J. J.; Judkins, J. C.; Sternberg, P. W.; Schroeder, F. C., Comparative metabolomics reveals biogenesis of ascarosides, a modular library of small molecule signals in *C. elegans*. *J Am Chem Soc* **2012,** *134* (3), 1817–1824.

12. Butcher, R. A.; Ragains, J. R.; Li, W.; Ruvkun, G.; Clardy, J.; Mak, H. Y., Biosynthesis of the *Caenorhabditis elegans* dauer pheromone. *Proc Natl Acad Sci U S A* **2009,** *106* (6), 1875-9.

13. Izrayelit, Y.; Robinette, S. L.; Bose, N.; von Reuss, S. H.; Schroeder, F. C., 2D NMR-based metabolomics uncovers interactions between conserved biochemical pathways in the model organism *Caenorhabditis elegans*. *ACS chemical biology* **2013,** *8* (2), 314-9.

14. Zhang, X.; Feng, L.; Chinta, S.; Singh, P.; Wang, Y.; Nunnery, J. K.; Butcher, R. A., Acyl-CoA oxidase complexes control the chemical message produced by *Caenorhabditis elegans*. *Proceedings of the National Academy of Sciences* **2015,** *112* (13), 3955-3960.

15. Poirier, Y.; Antonenkov, V. D.; Glumoff, T.; Hiltunen, J. K., Peroxisomal β-oxidation—A metabolic pathway with multiple functions. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **2006,** *1763* (12), 1413-1426.

16. Dieterich, C.; Clifton, S. W.; Schuster, L. N.; Chinwalla, A.; Delehaunty, K.; Dinkelacker, I.; Fulton, L.; Fulton, R.; Godfrey, J.; Minx, P.; Mitreva, M.; Roeseler, W.; Tian, H.; Witte, H.; Yang, S. P.; Wilson, R. K.; Sommer, R. J., The *Pristionchus pacificus* genome provides a unique perspective on nematode lifestyle and parasitism. *Nature genetics* **2008,** *40* (10), 1193-8.

17. Markov, G. V.; Baskaran, P.; Sommer, R. J., The same or not the same: lineage-specific gene expansions and homology relationships in multigene families in nematodes. *Journal of molecular evolution* **2015,** *80* (1), 18-36.

18. Markov, G. V.; Meyer, J. M.; Panda, O.; Artyukhin, A. B.; Claassen, M.; Witte, H.; Schroeder, F. C.; Sommer, R. J., Functional Conservation and Divergence of daf-22 Paralogs in *Pristionchus pacificus* Dauer Development. *Mol Biol Evol* **2016,** *33* (10), 2506-14.

19. Pereto, J.; Lopez-Garcia, P.; Moreira, D., Phylogenetic analysis of eukaryotic thiolases suggests multiple proteobacterial origins. *Journal of molecular evolution* **2005,** *61* (1), 65-74.

20. Baskaran, P.; Rodelsperger, C.; Prabh, N.; Serobyan, V.; Markov, G. V.; Hirsekorn, A.; Dieterich, C., Ancient gene duplications have shaped developmental stage-specific expression in *Pristionchus pacificus*. *BMC evolutionary biology* **2015,** *15*, 185.

21. Kaplan, F.; Srinivasan, J.; Mahanti, P.; Ajredini, R.; Durak, O.; Nimalendran, R.; Sternberg, P. W.; Teal, P. E.; Schroeder, F. C.; Edison, A. S.; Alborn, H. T., Ascaroside expression in *Caenorhabditis elegans* is strongly dependent on diet and developmental stage. *PLoS ONE* **2011,** *6* (3), e17804.

22. Golden, J. W.; Riddle, D. L., A gene affecting production of the *Caenorhabditis elegans* dauer-inducing pheromone. *Molecular & general genetics : MGG* **1985,** *198* (3), 534-6.

23. Kuhlisch, C.; Pohnert, G., Metabolomics in chemical ecology. *Natural product reports* **2015,** *32* (7), 937-55.

CHAPTER 3: BIOSYNTHESIS OF MODULAR ASCAROSIDES IN *C. ELEGANS*

Reproduced in part with permission from:

"Biosynthesis of modular ascarosides in *C. elegans*"

Oishika Panda, Allison E. Akagi, Alexander B. Artyukhin, Joshua C. Judkins, Henry H. Le, Parag Mahanti, Sarah M. Cohen, Paul W. Sternberg, and Frank C. Schroeder

Angewandte Chemie International Edition **2017**, *56 (17)*, 4729-4733.
DOI:10.1002/anie.201700103

Introduction

Nematodes, including the model organisms *Caenorhabditis elegans* and *Pristionchus pacificus* produce several small molecules with architectures of remarkable complexity[1]. These small molecules, called ascarosides, are involved in virtually all aspects of the life history of the nematodes, including the regulation of aging,[2] dauer development[3-5], morphology[6], mating[7] and behavior[8-12], in many cases via evolutionarily conserved signaling pathways[1, 7]. Different aspects of nematode biology are governed by different chemical structures of various members of this family of signaling molecules (Figure 3.1a, b).

In terms of chemical structure, all ascarosides consist of the dideoxy sugar ascarylose, connected to a lipid-derived side chain of varying lengths (Figure 3.1b). This central scaffold may optionally be further decorated with head groups and terminal groups derived directly from various primary metabolic pathways. Thus, ascarosides are modular libraries of signaling molecules integrating building blocks derived from conserved primary metabolic pathways, including lipid, amino acid, carbohydrate, and nucleoside metabolism[1], and as such can be

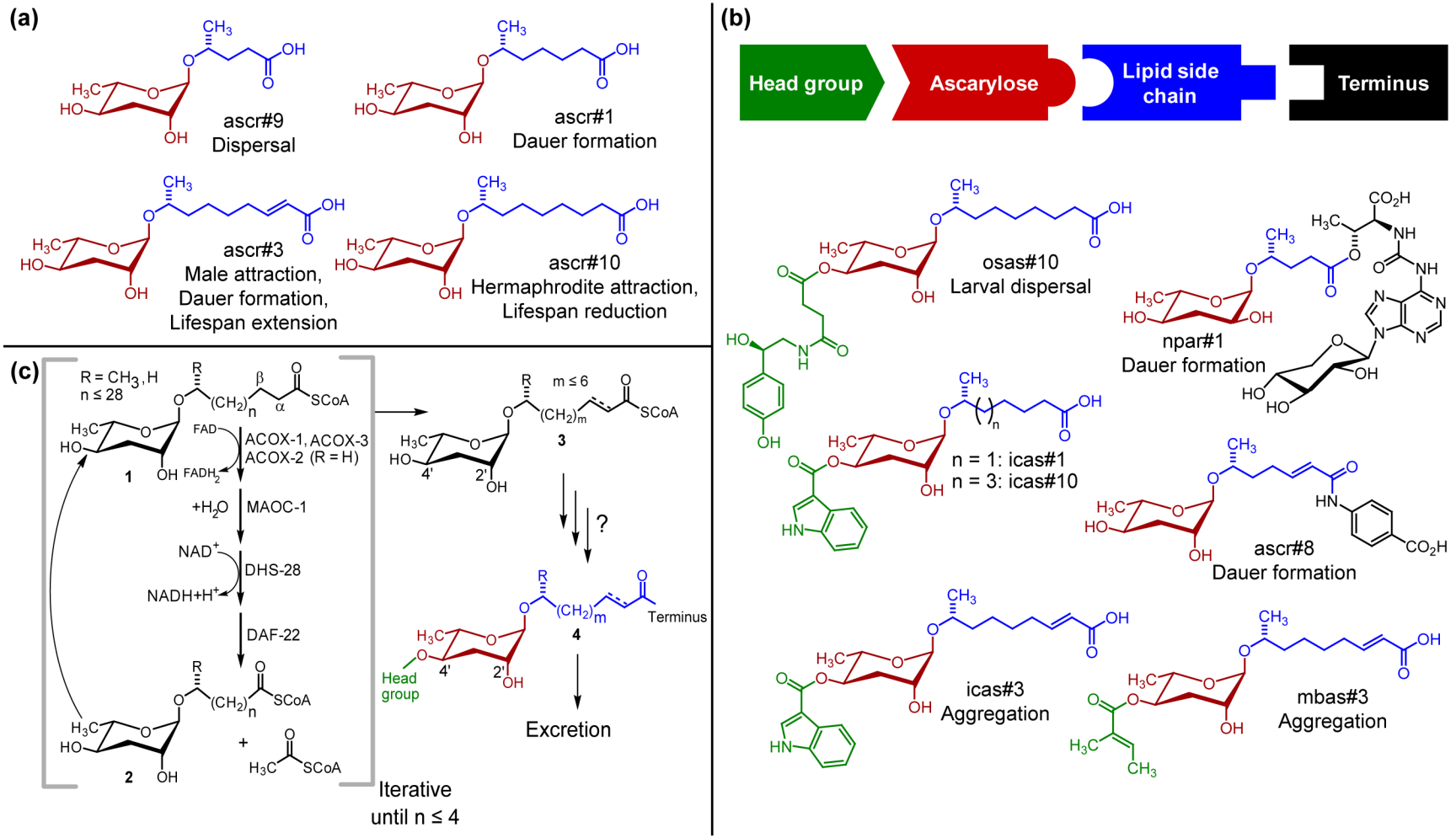

Figure 3.1. Structures and biosynthesis of ascarosides. (a) Simple ascarosides ascr#1, ascr#3, ascr#9, and ascr#10, which serve signaling functions in *C. elegans* and many other nematode species. **(b)** Modular ascarosides incorporating building blocks from folate (ascr#8), neurotransmitter (osas#10), tryptophan (icas#3, icas#10), and short-chain fatty acid (mbas#3) metabolism from *C. elegans*, and npar#1 from the satellite model organism *P. pacificus* featuring a nucleoside-derived building block. **(c)** Model for the biosynthesis of ascarosides. Iterative 4-step β-oxidation of long-chain ascarosides **(1)** produces short-chain precursors **(3)** of modular ascarosides **(4)**.

thought of as readouts for the overall metabolic state of the organism. The identification of these nematode small molecule signals demonstrated not only unexpected biosynthetic capabilities in a metazoan, but also provided a striking example of combinatorial generation of structural diversity that is reminiscent of strategies employed for the creation of synthetic combinatorial libraries in medicinal chemistry[1], such as polyketides and non-ribosomal peptides in microorganisms and plants.

Assembly of the ascarosides identified from *C. elegans* and other nematodes appears to proceed with high selectivity. For example, the likely folate-derived p-aminobenzoic acid moiety in ascr#8 is selectively attached to an unsaturated 7-carbon sidechain, although the simple ascarosides based on saturated 7-carbon and unsaturated 9-carbon sidechains, ascr#1 and ascr#3, are much more abundant[5, 13] (Figure 3.1a). Specific biological functions and highly selective assembly indicate that these modular signaling molecules are the products of dedicated biosynthetic pathways. Ultimately, the elucidation of these pathways will reveal how input from conserved primary metabolism is transduced to create signals that regulate development and behavior.

Previous efforts in elucidating ascaroside biosynthetic pathways

Efforts by the Schroeder research group and others have previously demonstrated that the fatty acid-like side chains in the ascarosides are derived from very long-chain precursors (1) via peroxisomal β-oxidation[9, 13-17] (Figure 3.1c). Other building blocks, e.g. the indole carboxy moiety in icas#3 or the octopamine moiety in osas#10, are derived from canonical primary metabolites, i.e., tryptophan or indole carboxylic acid in the case of icas#3[13] and tyrosine in the case of osas#10[8]. The author demonstrated this by conducting feeding experiments with *daf-22(ok693)* mutants of *C. elegans*, in which peroxisomal β-oxidation is disrupted leading to the absence of

medium- and short-chained ascarosides[18]. Synthetic ascr#1 and ascr#3, when fed to *daf-22(ok693)* mutants, were converted to their corresponding 4′-substituted derivatives icas#1 and icas#3, suggesting that the attachment of additional building blocks at the 4′-position occurs subsequent to and independent of side-chain β-oxidation (Figure 3.1c). Several particularly important chemical signals represent 4′-acylated ascarosides, e.g. icas#3 and icas#10, two components of the *C. elegans* aggregation pheromone[12], and osas#10, a larval dispersal signal (Figure 3.1b). However, the enzymes and cellular compartments involved in the highly selective attachment of diverse head groups at the 4′-position of the central ascarylose scaffold via ester bonds have hitherto eluded discovery.

Two-pronged approach for identifying ascaroside biosynthetic enzymes

In search of putative enzymes responsible for connecting different acyl moieties to the 4′-position of ascarylose, the author screened candidate genes based on (a) predicted enzymatic function and (b) sub-cellular localization. For example, O-acyltransferases have been shown to catalyze the formation of ester linkages in various small molecules found in nature[19]. The *C. elegans* genome includes 61 genes annotated as O-acyltransferases, and for about 30 of these likely loss-of-function mutants were available from the Caenorhabditis Genetics Center (CGC), and the National BioResource Project (NBRP) (Appendix table C1). Dr. Joshua C. Judkins, (Schroeder research group, Cornell University) reported in his PhD dissertation that comparative metabolomic analysis of liquid cultures of these mutants and wildtype (N2) controls by LC-MS revealed a promising O-acyltransferase candidate enzyme, OAC-50. The *oac-50(gk402144)* mutants produced wildtype levels of the unmodified ascr#3, but very low amounts of mbas#3 (Appendix Figure C1a). This deficiency could not be rescued by exogenous addition of synthetic tiglic acid, the proposed building block of this 4′-acylated ascaroside, suggesting that this mutant is indeed deficient in the

biosynthetic enzyme that acylates the 4′-position of ascr#3 (Appendix Figure C1b). However, the author found that the strain VC20784, which bears the *gk402144* allele of *oac-50*, could recover its mbas#3 production over several generations (Appendix Figure C1d). Untargeted comparative metabolomics using UHPLC-HRMS revealed that leucine, isoleucine and their *N*-acetyl and *N*-propionyl derivatives were also abolished in the mbas#3-deficient VC20784 cultures, and rescued concomitantly with mbas#3 in future generations (Appendix Figure C1c, d). As tiglic acid is biosynthetically derived from isoleucine[20], these observations provide evidence for metabolic rewiring of the branched-chain amino acid pathways in this strain which is corrected over several generations. Future experiments will bring to light whether this rewiring is a result of epigenetic imprinting or an inherent infection in the strain obtained from CGC.

Since shortening of the fatty acid-like side chains of the ascarosides takes place in peroxisomes, the author additionally screened several mutants of genes bearing a peroxisome targeting sequence (serine-lysine-leucine and some variations thereof, at the C-terminus of the protein)[21-22], including three putative acyl-CoA synthetases. In a pilot screen of seven genes for which deletion mutants were available, it was found that production of two 4′-acylated derivatives of ascr#9, the aggregation pheromone icas#9 and the larval dispersal signal osas#9, was abolished in *acs-7(tm6781)* mutants (Figure 3.2a, b), whereas production of the structurally related icas#3, icas#10, and osas#10 was similar or elevated compared to wildtype (Figure 3.1b, 3.2a). Abundances of all other known ascarosides, including ascr#9, were close to those found in wildtype controls (Figure 3.2a). Similarly, abundances of *N*-succinyl octopamine (5) and indole-3-carboxylic acid (7), plausible building blocks of osas#9 and icas#9, respectively, were unchanged in *acs-7* mutants (Appendix Figure C2a). The phenotype persisted even after growing the mutants for several generations, as well as after outcrossing the strain ten times to the wildtype

strain to remove any possible background mutations. Ascaroside production was largely wildtype-like in the other mutants in the pilot screen of putative peroxisomally-targeted enzymes, except for a slight reduction in the abundance of some 4′-modified ascarosides in *acs-13(ok2861)* and *acs-14(ok3391)* mutants and an increase in icas#3 in mutants of C24A3.4, an ortholog of

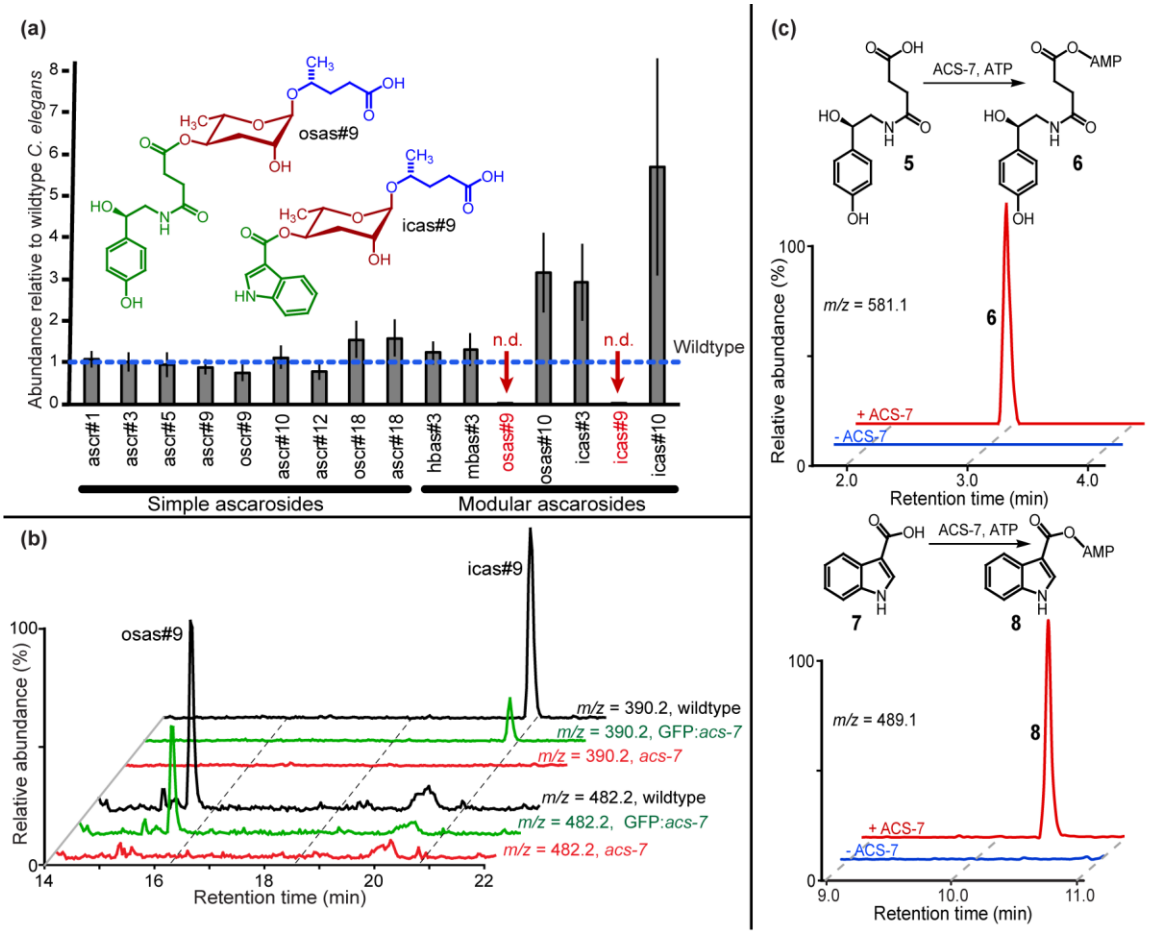

Figure 3.2. ACS-7 contributes to the biosynthesis of 4'-acylated ascarosides. (a) Abundances of ascarosides in *acs-7(tm6781)* mutants relative to wildtype *C. elegans.* Error bars represent standard deviation of three biological replicates. **(b)** HPLC-MS (ESI-) ion chromatograms for icas#9 and osas#9 in wildtype, *acs-7(tm6781)* mutants and *acs-7(tm6781)* mutants expressing *gfp::acs-7*. **(c)** HPLC-MS (ESI-) ion chromatograms showing ACS-7-dependent formation of the adenylates of *N*-succinyl octopamine and indole-3-carboxylic acid.

human α-methylacyl-CoA racemase (Appendix Figure C2b). A double mutant of *acs-13;acs-14* showed levels of 4′-modified ascarosides similar to those of the single mutants, suggesting that other redundant enzymes are involved in their biosynthetic mechanism (see Appendix Figure C2c).

Functional characterization of ACS-7

The above results indicated that ACS-7 is involved in the biosynthesis of specifically icas#9 and osas#9, which are derived from the 5-carbon side chain simple ascaroside ascr#9. *acs-7* encodes a putative acyl-CoA synthetase with no known function in *C. elegans* and ~70% sequence homology with human acyl-CoA synthetase family member 2, a mitochondrial fatty-acid CoA synthetase with preference for medium-chain substrates. To confirm the role of *acs-7* in the biosynthesis of icas#9 and osas#9, the author collaborated with Dr. Allison E. Akagi (Sternberg research group, California Institute of Technology), to express *gfp::acs-7* under the native *acs-7* promoter in the *acs-7(tm6781)* background. LC-MS analysis of the *gfp::acs-7* cultures revealed that biosynthesis of icas#9 and osas#9 was largely restored (Figure 3.2b), whereas the amounts of other ascarosides remained unchanged (Appendix Figure C2c). Additional untargeted metabolomics revealed no other significant differences between the metabolomes of *acs-7* worms and wild type. This observation confirmed that ACS-7 is involved specifically in the production of the 4′-modified ascarosides icas#9 and osas#9.

Based on sequence homology, the author hypothesized that ACS-7 may be involved in activating indole-3-carboxylic acid (7) and *N*-succinyl octopamine (5), as building blocks of icas#9 and osas#9, respectively, and subsequent 4′-acylation of ascr#9. To test this, the author collaborated with Henry H. Le (Schroeder research group, Cornell University) to purify recombinant his_6-ACS-7 from *E. coli*. Incubation of ACS-7 with indole-3-carboxylic acid, ATP, and coenzyme A rapidly produced indole-3-carboxy-AMP (8) (Km = 270±90 μM, Appendix Figure C3), whereas formation of indole carboxy-CoA was not observed. Similarly, *N*-succinyl octopamine (5) was converted into the corresponding adenylate 6. However, addition of synthetic ascr#9 or ascr#9-SCoA in these assays did not produce icas#9, osas#9, or the corresponding SCoA-derivatives. Similarly, ACS-7 did not convert ascr#9 or pentanoic acid, which had been included

as an unrelated control substrate, into either the corresponding adenylates or SCoA derivatives. These results indicate that ACS-7 contributes to the biosynthesis of icas#9 and osas#9 by activating indole carboxylic acid and *N*-succinyl octopamine, respectively; however, 4′-attachment to ascr#9 or ascr#9-SCoA may involve additional factors or require a specific cellular environment that could not be mimicked in these assays.

Site of 4'-acylated ascaroside biosynthesis

The author sought to confirm whether ACS-7 indeed localizes to peroxisomes, as predicted by its peroxisomal targeting sequence (C-terminal alanine-lysine-leucine). Surprisingly, it was found that GFP::ACS-7 is primarily expressed in punctate organelles[23] ("gut granules") in intestinal cells, rather than peroxisomes (Figures 3.3a, b). *C. elegans* intestinal cells are known to contain several different types of gut granules, including acidic lysosome-related organelles[23] (LROs). LROs play a central role in the digestion of cellular waste and the regeneration of basic building blocks, but also are suspected to contribute to the biosynthesis of diverse metabolites[24-26]. Staining with LysoTracker Red, a dye that selectively stains acidic organelles of the cell[23, 27], revealed that GFP::ACS-7 localizes specifically to the acidic LROs in the intestinal cells (Figures 3.3c, d). This suggested that biosynthesis of the modular ascarosides icas#9 and osas#9 may proceed in the acidic LROs and not in the peroxisomes.

To test whether acidic LROs are required for the biosynthesis of icas#9, osas#9, and potentially other modular ascarosides, the author, together with Dr. Parag Mahanti, Dr. Joshua C. Judkins and Dr. Alexander B. Artyukhin (Schroeder research group, Cornell University), investigated the ascaroside profiles of a several mutants that are defective in LROs or other types of gut granules. *glo-1* encodes a 24 kDa member of the Rab family of small GTPases, which is required for the biogenesis of LROs[23]. Correspondingly, *glo-1(zu437)* mutants, lacking LROs, do

not show any staining with LysoTracker Red[23]. *glo-4* encodes a guanine nucleotide exchange factor contributing to GLO-1 function[23]. Both *glo-1* and *glo-4* act downstream of *apb-3*, which encodes the worm ortholog of the β3 subunit of the mammalian adaptor protein complex AP-3[28]. It was found that biosynthesis of all known 4′-modified ascarosides is largely abolished in *glo-1(zu437)* mutants, which lack acidic LROs (Figure 3.3e), and reduced in *glo-4(ok623)* and *apb-3(ok429)* mutants, in which formation of acidic LROs is reduced, but not abolished (Appendix Figure C4).

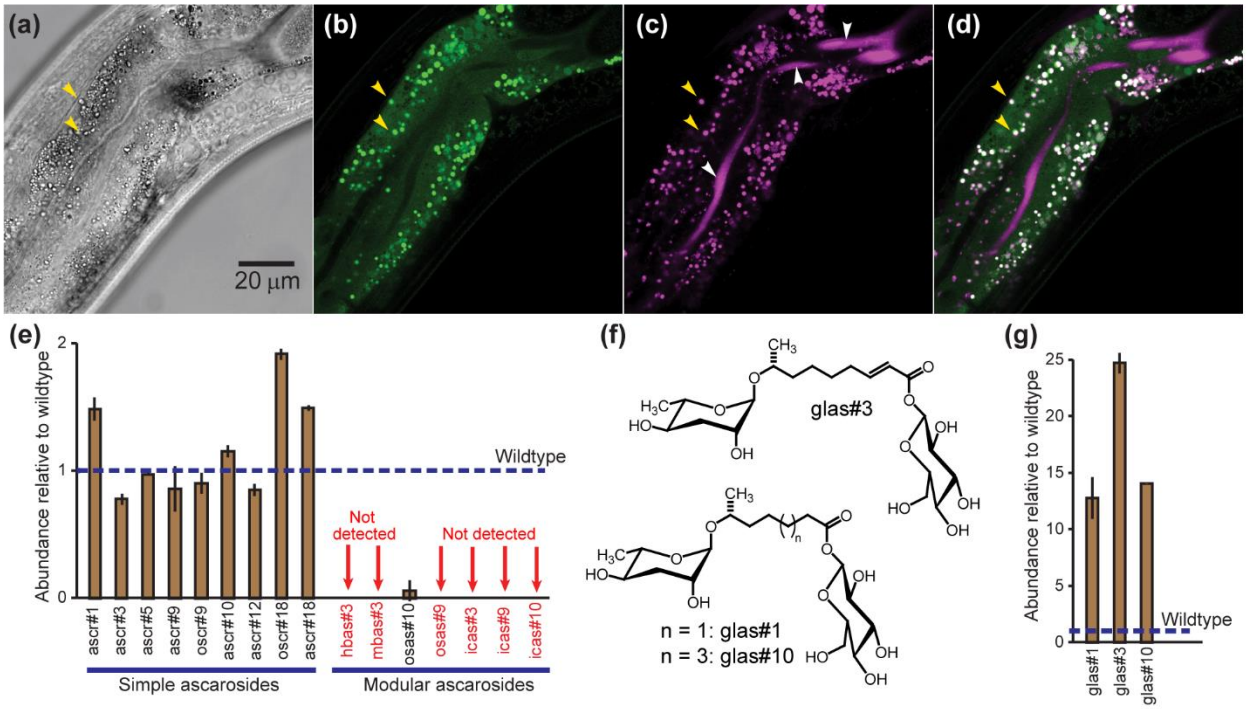

Figure 3.3. Biosynthesis of modular ascarosides in acidic LROs. (a) Bright-field image of mid section of transgenic worm carrying *acs-7p::gfp::acs-7* in *acs-7* background (young adult stage). **(b)** GFP expression is localized to punctate organelles (e.g. yellow arrows). **(c)** Staining of LROs with LysoTracker Red (white arrows: background staining of the intestinal lumen). **(d)** Overlay of images in (b) and (c). **(e)** Abundances of ascarosides in *glo-1(zu437)* mutants relative to wildtype *C. elegans*. **(f)** Structures of ascaroside glucosyl conjugates. **(g)** Abundances of glucosyl conjugates glas#1, glas#3, and glas#10 are greatly increased in *glo-1* knockout mutants. In (e) and (g), error bars represent standard deviation of two biological replicates.

The amounts of other ascarosides in these mutants were similar to those in wildtype controls, except for glycosylated derivatives glas#1, glas#3, and glas#10, which were starkly increased in *glo-1(zu437)* mutants relative to wildtype (Figure 3.3f). Glycosylation generally

serves as detoxification mechanisms in *C. elegans*[29-30], and buildup of glycosylated ascarosides in *glo-1* mutants may reflect loss of alternative waste processing pathways due to lack of acidic LROs in this mutant. Importantly, abundances of indole-3-carboxylic acid and *N*-succinyl octopamine in *glo-1* mutants were similar or only slightly reduced relative to wildtype, suggesting that abolishment of 4′-modified ascarosides in this mutant is not due to depletion of any of the building blocks (Appendix Figure C2a). In contrast, amounts of 4′ modified ascarosides in *cup-5(ar465), haf-4(ok1042)* and *haf-9(gk23)* mutants, which are defective in the formation of non-acidic gut granules, but do have acidic LROs[31], are similar or increased relative to amounts found in wildtype worms (Appendix Figure C4). These results indicated that the acidic LROs are required for the biosynthesis of all 4′-acylated ascarosides, including icas#9 and osas#9, whose biosynthesis specifically requires ACS-7 (Figure 3.4).

The acidic LROs in *C. elegans* share many characteristics with lysosomes in higher animals[23]. The lysosome, classically known as the waste disposal system of the cell, is rapidly gaining attention as a central hub for intra- and inter-cellular chemical signaling[26, 32]. Lysosomal acid lipases play critical roles under starvation conditions to generate small-molecule signals such as polyunsaturated fatty acids[33] and oleoylethanolamide[25], which in turn activate diverse downstream responses, including autophagy, a starvation-activated process by which targeted components of the cell are transferred to the lysosome for recycling[26]. The author found that in *atg-18(gk378)* mutants[34], which are deficient in autophagosome assembly, the abundances of most 4′-modified ascarosides are significantly reduced relative to wildtype (Appendix Figure C5), suggesting that autophagy contributes to lysosomal biosynthesis of modular ascaroside-based signaling molecules.

Conclusion

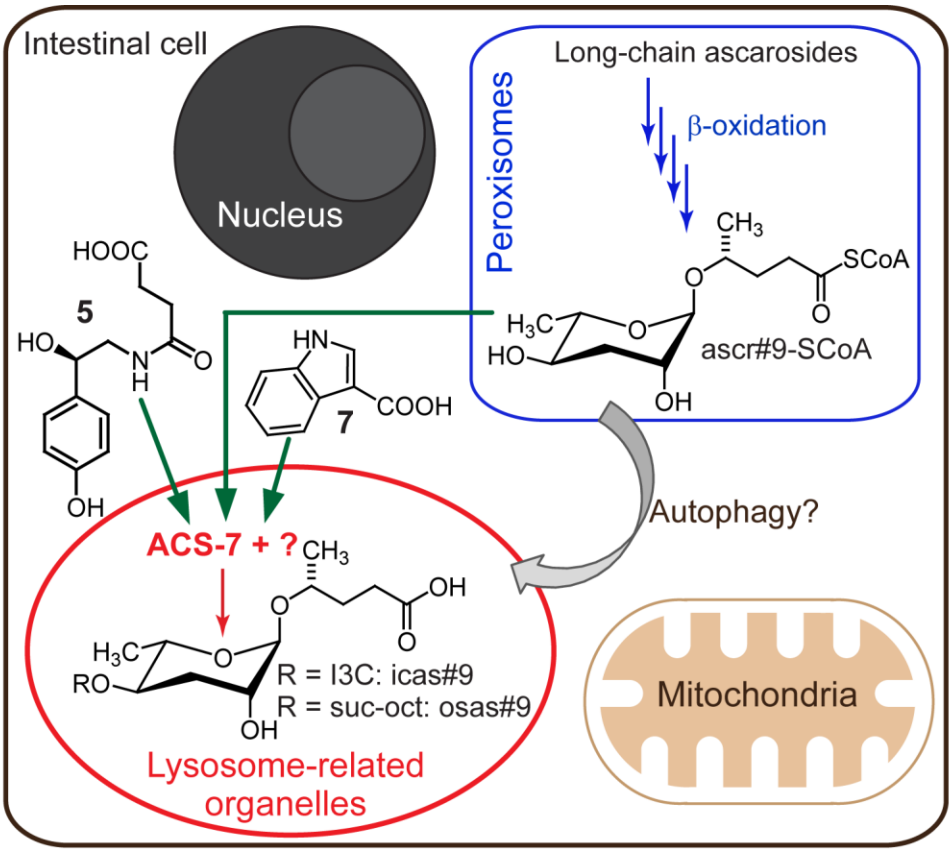

Figure 3.4. Model for modular ascaroside biosynthesis in *C. elegans*. Long-chain ascaroside precursors undergo multiple rounds of peroxisomal β-oxidation to produce SCoA esters of short-chain ascarosides such as ascr#9. Subsequently, short-chain ascarosides (or their SCoA esters) may undergo 4'-modification in acidic lysosome-related organelles (LROs). I3C: Indole-3-carboxy, suc-oct: *N*-succinyl octopamine.

Taken together, these results indicate that in *C. elegans*, homologs of genes from canonical metazoan metabolism act in different cell compartments to produce modular specialized metabolites with high specificity. The high specificity of ACS-7, being required only for the biosynthesis of 4′-modified versions of the 5-carbon side chain ascaroside ascr#9, i.e. icas#9 and osas#9, but not corresponding 7-carbon or 9-carbon side chain versions, e.g. icas#3 or osas#10 indicates that there must exist other enzymes that serve equivalent roles in the biosynthesis of 4′-modified ascarosides with longer fatty acid side chains. Identifying candidate enzymes using co-immunoprecipitation with GFP::ACS-7 and subsequent mass-spectrometry-based proteomics, or using classical yeast-two-hybrid screening approaches will revealing the interacting partners for

ACS-7. The specific localization of ACS-7 to the LROs, which serve as part of the cell's waste disposal system, and the serendipitous discovery that LROs are required for the biosynthesis of all 4′-modified ascarosides will motivate organelle-specific proteomics and metabolomics toward identifying other components in the biosynthesis of modular ascarosides. Of particular interest will be mechanisms that underlie the selection of different building blocks and that govern side chain-length specificity of the modifications, ultimately integrating metabolic state in a complex chemical message.

REFERENCES

1. von Reuss, S. H.; Schroeder, F. C., Combinatorial chemistry in nematodes: modular assembly of primary metabolism-derived building blocks. *Natural product reports* **2015,** *32* (7), 994-1006.

2. Ludewig, A. H.; Izrayelit, Y.; Park, D.; Malik, R. U.; Zimmermann, A.; Mahanti, P.; Fox, B. W.; Bethke, A.; Doering, F.; Riddle, D. L.; Schroeder, F. C., Pheromone sensing regulates *Caenorhabditis elegans* lifespan and stress resistance via the deacetylase SIR-2.1. *Proc Natl Acad Sci U S A* **2013,** *110* (14), 5522-7.

3. Butcher, R. A.; Fujita, M.; Schroeder, F. C.; Clardy, J., Small-molecule pheromones that control dauer development in *Caenorhabditis elegans*. *Nat Chem Biol* **2007,** *3* (7), 420-2.

4. Jeong, P. Y.; Jung, M.; Yim, Y. H.; Kim, H.; Park, M.; Hong, E.; Lee, W.; Kim, Y. H.; Kim, K.; Paik, Y. K., Chemical structure and biological activity of the *Caenorhabditis elegans* dauer-inducing pheromone. *Nature* **2005,** *433* (7025), 541-5.

5. Pungaliya, C.; Srinivasan, J.; Fox, B. W.; Malik, R. U.; Ludewig, A. H.; Sternberg, P. W.; Schroeder, F. C., A shortcut to identifying small molecule signals that regulate behavior and development in *Caenorhabditis elegans*. *Proc Natl Acad Sci U S A* **2009,** *106* (19), 7708-13.

6. Bose, N.; Ogawa, A.; von Reuss, S. H.; Yim, J. J.; Ragsdale, E. J.; Sommer, R. J.; Schroeder, F. C., Complex small-molecule architectures regulate phenotypic plasticity in a nematode. *Angew Chem Int Ed Engl* **2012,** *51* (50), 12438-43.

7. Ludewig, A. H.; Schroeder, F. C., Ascaroside signaling in *C. elegans*. *WormBook* **2013**, 1-22.

8. Artyukhin, A. B.; Yim, J. J.; Srinivasan, J.; Izrayelit, Y.; Bose, N.; von Reuss, S. H.; Jo, Y.; Jordan, J. M.; Baugh, L. R.; Cheong, M.; Sternberg, P. W.; Avery, L.; Schroeder, F. C., Succinylated octopamine ascarosides and a new pathway of biogenic amine metabolism in *Caenorhabditis elegans*. *The Journal of biological chemistry* **2013,** *288* (26), 18778-83.

9. Izrayelit, Y.; Srinivasan, J.; Campbell, S. L.; Jo, Y.; von Reuss, S. H.; Genoff, M. C.; Sternberg, P. W.; Schroeder, F. C., Targeted metabolomics reveals a male pheromone and sex-specific ascaroside biosynthesis in *Caenorhabditis elegans*. *ACS chemical biology* **2012,** *7* (8), 1321-5.

10. Kaplan, F.; Srinivasan, J.; Mahanti, P.; Ajredini, R.; Durak, O.; Nimalendran, R.; Sternberg, P. W.; Teal, P. E.; Schroeder, F. C.; Edison, A. S.; Alborn, H. T., Ascaroside expression in *Caenorhabditis elegans* is strongly dependent on diet and developmental stage. *PLoS ONE* **2011,** *6* (3), e17804.

11. Srinivasan, J.; Kaplan, F.; Ajredini, R.; Zachariah, C.; Alborn, H. T.; Teal, P. E. A.; Malik, R. U.; Edison, A. S.; Sternberg, P. W.; Schroeder, F. C., A blend of small molecules regulates both mating and development in *Caenorhabditis elegans*. *Nature* **2008,** *454* (7208), 1115-1118.

12. Srinivasan, J.; von Reuss, S. H.; Bose, N.; Zaslaver, A.; Mahanti, P.; Ho, M. C.; O'Doherty, O. G.; Edison, A. S.; Sternberg, P. W.; Schroeder, F. C., A Modular Library of Small Molecule Signals Regulates Social Behaviors in *Caenorhabditis elegans*. *PLoS Biol* **2012,** *10* (1), e1001237.

13. von Reuss, S. H.; Bose, N.; Srinivasan, J.; Yim, J. J.; Judkins, J. C.; Sternberg, P. W.; Schroeder, F. C., Comparative metabolomics reveals biogenesis of ascarosides, a modular library of small molecule signals in *C. elegans*. *J Am Chem Soc* **2012,** *134* (3), 1817–1824.

14. Butcher, R. A.; Ragains, J. R.; Li, W.; Ruvkun, G.; Clardy, J.; Mak, H. Y., Biosynthesis of the *Caenorhabditis elegans* dauer pheromone. *Proc Natl Acad Sci U S A* **2009,** *106* (6), 1875-9.

15. Izrayelit, Y.; Robinette, S. L.; Bose, N.; von Reuss, S. H.; Schroeder, F. C., 2D NMR-based metabolomics uncovers interactions between conserved biochemical pathways in the model organism *Caenorhabditis elegans*. *ACS chemical biology* **2013,** *8* (2), 314-9.

16. Markov, G. V.; Meyer, J. M.; Panda, O.; Artyukhin, A. B.; Claassen, M.; Witte, H.; Schroeder, F. C.; Sommer, R. J., Functional Conservation and Divergence of daf-22 Paralogs in *Pristionchus pacificus* Dauer Development. *Mol Biol Evol* **2016,** *33* (10), 2506-14.

17. Zhang, X.; Feng, L.; Chinta, S.; Singh, P.; Wang, Y.; Nunnery, J. K.; Butcher, R. A., Acyl-CoA oxidase complexes control the chemical message produced by *Caenorhabditis elegans*. *Proceedings of the National Academy of Sciences* **2015,** *112* (13), 3955-3960.

18. Panda, O.; Akagi, A. E.; Artyukhin, A. B.; Judkins, J. C.; Le, H. H.; Mahanti, P.; Cohen, S. M.; Sternberg, P. W.; Schroeder, F. C., Biosynthesis of Modular Ascarosides in *C. elegans*. *Angew Chem Int Ed Engl* **2017,** *56* (17), 4729-4733.

19. Boswell, H. D.; Dräger, B.; McLauchlan, W. R.; Portsteffen, A.; Robins, D. J.; Robins, R. J.; Walton, N. J., Specificities of the enzymes of N-alkyltropane biosynthesis in *Brugmansia* and *Datura*. *Phytochemistry* **1999,** *52* (5), 871-878.

20. Attygalle, A. B.; Wu, X.; Will, K. W., Biosynthesis of Tiglic, Ethacrylic, and 2-Methylbutyric Acids in a Carabid Beetle, *Pterostichus* (*Hypherpes*) *californicus*. *Journal of Chemical Ecology* **2007,** *33* (5), 963-970.

21. Gould, S. J.; Keller, G. A.; Hosken, N.; Wilkinson, J.; Subramani, S., A conserved tripeptide sorts proteins to peroxisomes. *The Journal of cell biology* **1989,** *108* (5), 1657-64.

22. Motley, A. M.; Hettema, E. H.; Ketting, R.; Plasterk, R.; Tabak, H. F., *Caenorhabditis elegans* has a single pathway to target matrix proteins to peroxisomes. *EMBO reports* **2000,** *1* (1), 40-6.

23. Hermann, G. J.; Schroeder, L. K.; Hieb, C. A.; Kershner, A. M.; Rabbitts, B. M.; Fonarev, P.; Grant, B. D.; Priess, J. R., Genetic Analysis of Lysosomal Trafficking in *Caenorhabditis elegans*. *Molecular Biology of the Cell* **2005,** *16* (7), 3273-3288.

24. Coburn, C.; Allman, E.; Mahanti, P.; Benedetto, A.; Cabreiro, F.; Pincus, Z.; Matthijssens, F.; Araiz, C.; Mandel, A.; Vlachos, M.; Edwards, S.-A.; Fischer, G.; Davidson, A.; Pryor, R. E.; Stevens, A.; Slack, F. J.; Tavernarakis, N.; Braeckman, B. P.; Schroeder, F. C.; Nehrke, K.; Gems, D., Anthranilate Fluorescence Marks a Calcium-Propagated Necrotic Wave That Promotes Organismal Death in *C. elegans*. *PLoS Biol* **2013,** *11* (7), e1001613.

25. Folick, A.; Oakley, H. D.; Yu, Y.; Armstrong, E. H.; Kumari, M.; Sanor, L.; Moore, D. D.; Ortlund, E. A.; Zechner, R.; Wang, M. C., Aging. Lysosomal signaling molecules regulate longevity in *Caenorhabditis elegans*. *Science* **2015,** *347* (6217), 83-6.

26. Mony, V. K.; Benjamin, S.; O'Rourke, E. J., A lysosome-centered view of nutrient homeostasis. *Autophagy* **2016,** *12* (4), 619-31.

27. Soukas, A. A.; Carr, C. E.; Ruvkun, G., Genetic regulation of *Caenorhabditis elegans* lysosome related organelle function. *PLoS genetics* **2013,** *9* (10), e1003908.

28. Boehm, M.; Bonifacino, J. S., Adaptins: the final recount. *Mol Biol Cell* **2001,** *12* (10), 2907-20.

29. Soukup, S. T.; Spanier, B.; Grunz, G.; Bunzel, D.; Daniel, H.; Kulling, S. E., Formation of phosphoglycosides in *Caenorhabditis elegans*: a novel biotransformation pathway. *PLoS One* **2012,** *7* (10), e46914.

30. Stupp, G. S.; von Reuss, S. H.; Izrayelit, Y.; Ajredini, R.; Schroeder, F. C.; Edison, A. S., Chemical detoxification of small molecules by *Caenorhabditis elegans*. *ACS chemical biology* **2013,** *8* (2), 309-13.

31. Tanji, T.; Nishikori, K.; Haga, S.; Kanno, Y.; Kobayashi, Y.; Takaya, M.; Gengyo-Ando, K.; Mitani, S.; Shiraishi, H.; Ohashi-Kobayashi, A., Characterization of HAF-4- and HAF-9-

localizing organelles as distinct organelles in *Caenorhabditis elegans* intestinal cells. *BMC cell biology* **2016,** *17*, 4.

32. Huber, L. A.; Teis, D., Lysosomal signaling in control of degradation pathways. *Current opinion in cell biology* **2016,** *39*, 8-14.

33. O'Rourke, E. J.; Kuballa, P.; Xavier, R.; Ruvkun, G., omega-6 Polyunsaturated fatty acids extend life span through the activation of autophagy. *Genes & development* **2013,** *27* (4), 429-40.

34. Zhang, H.; Chang, J. T.; Guo, B.; Hansen, M.; Jia, K.; Kovács, A. L.; Kumsta, C.; Lapierre, L. R.; Legouis, R.; Lin, L.; Lu, Q.; Meléndez, A.; O'Rourke, E. J.; Sato, K.; Sato, M.; Wang, X.; Wu, F., Guidelines for monitoring autophagy in *Caenorhabditis elegans*. *Autophagy* **2015,** *11* (1), 9-27.

CHAPTER 4: UNTARGETED COMPARATIVE METABOLOMICS REVEALS POTENTIAL AGING BIOMARKERS IN MITOCHONDRIAL MUTANTS OF *C. ELEGANS*

Introduction

Of the numerous biochemical processes taking place in the mitochondria, classically regarded as that "powerhouse of the cell", perhaps the best studied is the mitochondrial electron transport chain (ETC). The mitochondrial ETC consists of five protein complexes, which, via series of redox reactions, transfer electrons from donor molecules to acceptor molecules, thereby creating an electrochemical proton gradient that drives the synthesis of adenosine triphosphate (ATP), the "energy currency" of the cell. Figure 4.1 shows a schematic for the redox reactions and electrochemical gradients in the ETC[1].

Given this critical role of the mitochondria in any eukaryotic cell, perturbations to the ETC are usually very damaging to the organism. However, recent studies have shown that mild to moderate ETC dysfunction can, quite counter-intuitively, increase lifespan of organisms ranging from yeast to mice[2-7], whereas a more severe ETC dysfunction usually (but not always) shortens

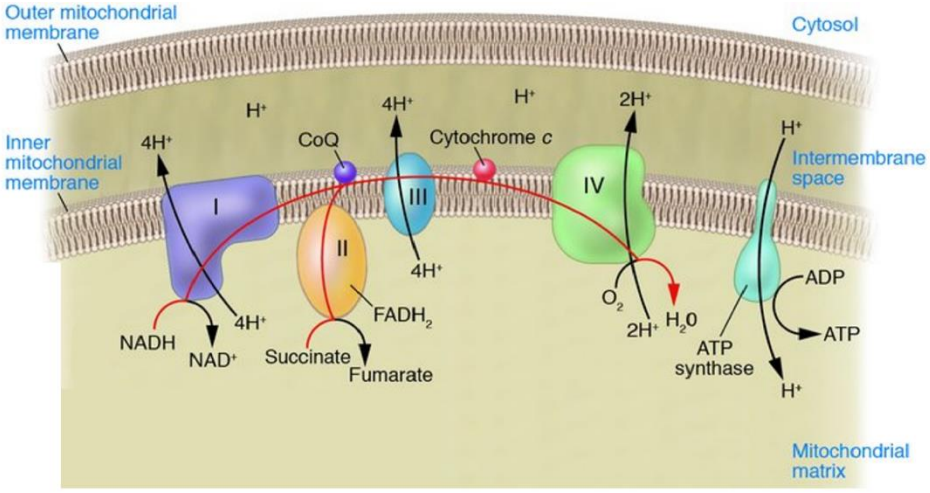

Figure 4.1 (adapted from Reference 1)**. The mitochondrial electron transport chain.** Complexes I and II are the "loading zones" for products of sugar, lipid, protein catabolism which are chemically transformed here by redox couples NADH/NAD+ and FADH/FAD, thereby releasing electrons. These electrons are shuttled to Complex III via ubiquinone (CoQ) and then to Complex IV, where it encounters molecular oxygen and reduces it to water. These complexes act as proton pumps, shuttling protons out of the matrix to create an electrochemical gradient across the intermembrane space. The dissipation of this electrochemical gradient in Complex V results in the production of ATP.

lifespan[8-9]. Several mutants of different complexes of the ETC have been identified in *C. elegans*, which exhibit an increase or decrease in longevity. *isp-1(qm150)* harbors a point mutation in the Rieske-like iron-sulfur protein, a key sub-unit of the ETC Complex III, and shows a lifespan extension of up to 50% compared to wild type *C. elegans*[10]. *nuo-6(qm200)* harbors a point mutation in the NADH-ubiquinone oxidoreductase of the ETC Complex I, and yet lives longer than wild type[11]. In both of these mutants, activity of Complex III and Complex I, respectively, is severely reduced[11]. In contrast, the *mev-1(kn1)* mutant, which bears a point mutation in the succinate dehydrogenase subunit of Complex II, exhibits significantly reduced lifespan than wild type[12].

The Lee Research Group, Cornell University, discovered that when two transcription factors, encoded by *cep-1*[13] and *ceh-23*[14-15], were inactivated in these long-lived and short-lived mitochondrial mutants of *C. elegans*, the longevity phenotype in both cases was rescued. CEP-1 is the only *C. elegans* ortholog of mammalian p53, a transcription factor involved in DNA repair, tumor suppression, cell cycle regulation, and apoptosis. The p53 functions of apoptosis and cell cycle regulation are conserved in *C. elegans*[16-18]. *ceh-23* encodes a homeodomain protein in *C. elegans*, and is required for the functions of the AIY interneuron[19]. Recent studies have shown that *ceh-23* and *cep-1* act downstream of AMP kinase signaling and the CRTC-1 transcriptional cofactor to mediate longevity and promote stress resistance[14].

Despite their intriguing lifespan phenotypes, not much is known about the metabolic profiles of the mitochondrial (Mit) mutants of *C. elegans*. A few studies have shown somewhat elevated levels of α-ketoacids and α-hydroxyacids in the long-lived Mit mutants, relative to the short-lived ones, and have been regarded as a putative pro-longevity signal in the long-lived Mit mutants[20-21]. The metabolomes of the short-lived Mit mutants appear to be enriched in amino acids and TCA cycle intermediates[20]. It has been shown that reduced mitochondrial ETC in *C. elegans*

neurons activates the mitochondrial unfolded protein response (UPR^{mt}) in intestinal cells[22], and that this activation is essential for the extension of lifespan in the ETC mutant nematodes. Durieux *et al*. proposed that a pro-longevity cue ('mitokine') originating from the neurons is responsible for signaling cell-non-autonomously to the intestinal cells to induce UPR^{mt}. However, the identity of the cell-non-autonomous signal that mediates this communication between neurons and intestine is yet to be discovered.

In this study, the author used an untargeted metabolomics approach, based on high resolution UHPLC-MS, to compare the *endo*- and *exo*-metabolomes of the long-lived and short-lived Mit mutants to those of wild type *C. elegans*. Several metabolites were found to be differentially regulated among these mutants. Some of these lifespan-correlated (or anti-correlated) metabolites were restored to wild type levels in the double mutants of the Mit mutants with *cep-1(gk138)* and *ceh-23(ms23)*, similar to their longevity phenotype. The Mit mutants and double mutants used in this study provide an excellent platform to look for small molecule biomarkers of aging as well as investigate conserved regulatory pathways governing aging.

Untargeted comparative metabolomics approach to study Mit mutant metabolomes

The author, along with Patrick Loi and Joseph Matragrano (Schroeder Research Group, Cornell University), grew at least three independent biological replicates of *isp-1*, *nuo-6*, *mev-1* and N2, as well as their double mutants with *cep-1* and *ceh-23*; i.e., *isp-1;cep-1*, *isp-1;ceh-23, nuo-6;cep-1*, *nuo-6;ceh-23*, *mev-1;cep-1* and *mev-1;ceh-23*. The nematodes were pelleted out and the worm pellet (WP) and media supernatant (WM) were separately processed and analyzed using a high resolution UHPLC-MS. The author made use of the 'centwave' centroid mode peak picking algorithm from the Bioconductor package XCMS in R[23], with a few modifications made by Dr. Susan Strickler (Mueller Research Group, Cornell University) and Dr. Maximilian Helf (Schroeder

Research Group, Cornell University), to automate the detection and quantification of thousands of features from the raw data obtained from the UHPLC-MS. The UHPLC-MS data were collected in the profile MS mode, based on instrument specifications. For profile MS data, one usually makes use of the 'matched filter' peak picking algorithm for XCMS. While this reduces chances of picking noisy peaks and therefore reduces the number of false positives, the author found that several real features of low intensity can be dismissed as noise by this algorithm, so one runs the risk of not getting comprehensive coverage of differential features in the samples of interest. However, the 'centwave' algorithm picks several features that are in fact noise, or labels the tail ends of a relatively large peak as separate features. As such, one needs to spend a lot of time manually filtering through the noise in the massive output list to arrive at a small list of actually differential features.

Using Microsoft Excel, each detected feature (a particular m/z value eluting at a specific retention time) was normalized to the sum total of all features detected in that sample. This normalization is performed to eliminate any variables such as sample preparation differences, biomass differences etc., which may be introduced as sample-to-sample variation. The strongest differences are the ones which follow the same trend (and differ only in absolute values), irrespective of the normalization method or the lack therof. The normalized features in each of the mutant samples were then compared to their corresponding normalized features in the wild type (N2) samples, that were grown in parallel with that mutant replicate. Only features from this analysis that were found to be at least two-fold up- or downregulated in every replicate of the different Mit mutants relative to N2 were further analyzed. The author then subjected these features to additional manual filtering of false positives and different isotopes and/or adducts corresponding to the same feature, using the "EIC Reader" code in R written by Dr. Maximilian Helf (see

Appendix D for codes). This was followed by manual verification of the differential intensities of these shortlisted features from the raw data files quantified using Thermo XCalibur Qual Browser viewing software. In the following sections, the author will discuss some classes of metabolites whose abundances correlate with the lifespan phenotype of the Mit mutants.

Phosphohexose derivatives

Several compounds were found to be upregulated in the short-lived *mev-1* mutant, but downregulated in the long-lived *isp-1* and *nuo-6* mutants. In the WP samples, which represent the *endo*-metabolome of the nematodes, the author discovered a family of phosphohexose derivatives that follow this trend (Figure 4.2). MS/MS analysis in the ESI- ionization mode demonstrated that all these compounds show fragments at m/z 79.96688 (PO_3H^-) and 96.96962 ($PO_4H_2^-$), which are characteristic of compounds containing phosphate groups. Further, these compounds show fragments at m/z 223.00131 ($C_6H_8PO_7^-$) and/or 241.01188 ($C_6H_{10}O_8P^-$), which are indicative of a phosphohexose moiety (Appendix Figures D1-D3). It is not possible to determine the position of the phosphate attachment to the hexose group solely based on MS/MS structure assignment, and NMR spectroscopy will be required to unambiguously determine the complete structures of these compounds. Like the ascarosides introduced in Chapter 2, these nematode-derived metabolites also appear to be modular in nature, with inputs from conserved primary metabolic pathways (amino acid metabolism, carbohydrate metabolism etc.) coming together to form the complete structures like pieces of a jigsaw.

At the metabolomic level, the double mutants with *cep-1(gk138)* and *ceh-23(ms23)* appear to rescue the upregulated phosphohexose derivatives 'metabotype' in the short-lived *mev-1(kn1),* to varying degrees (Figure 4.3). This suggests that the metabolic contribution of CEP-1 and CEH-23 to the lifespan phenotype is largely similar in *mev-1*. In the long-lived *isp-1(qm150)* and *nuo-*

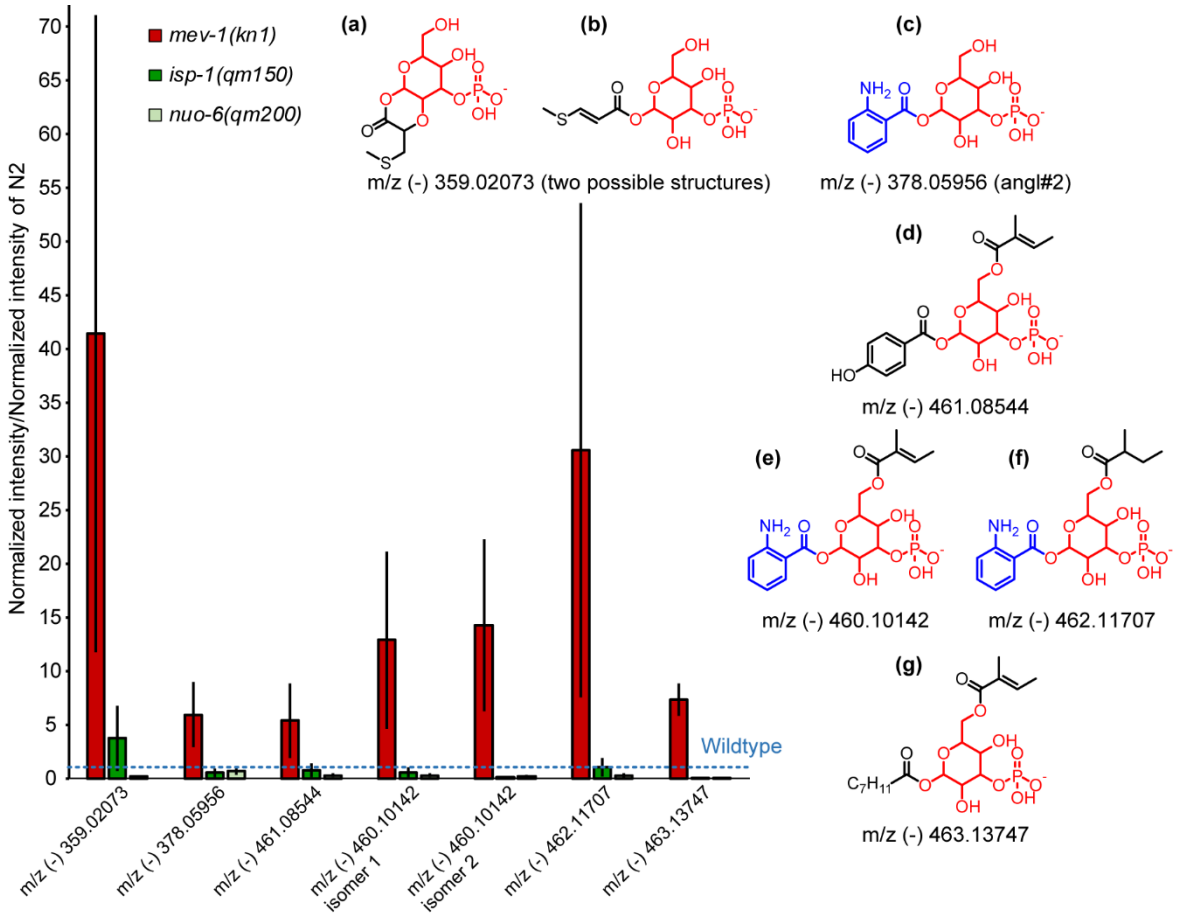

Figure 4.2. Phosphohexose derivatives in Mit mutants. (a) and **(b)** cannot be distinguished based on MS/MS fragments alone. **(c)** is angl#2, a metabolite which contributes to the 'death fluorescence' in *C. elegans*. **(d)** has 5 isomers, all of which show a similar trend in abundance in the endometabolome samples. **(e)** and **(f)** have 2 isomers each, all of which show a similar trend in abundance in the endometabolome samples. **(g)** requires further structural confirmation by NMR spectroscopy. Positions of attachment to of the different modules to the hexose ring, identity of the hexose moiety (angl#2 contains glucose), and positions of the heteroatoms on the aromatic rings remain to be confirmed by NMR spectroscopy. Normalized intensities in ESI negative ionization mode are shown here. Error bars = s. e. of at least 3 independent biological replicates.

6(qm200) mutants, the double mutants with *cep-1(gk138)* and *ceh-23(ms23)* appear to rescue the metabotype for most, but not all of these metabolites. This suggests that even in the same metabolic pathways that may contribute to the lifespan phenotype, CEP-1 and CEH-23 may have different modes of action in long-lived mutants compared to the short-lived *mev-1*.

Angl#2 (Figure 4.2c) has been described as a component of the death fluorescence

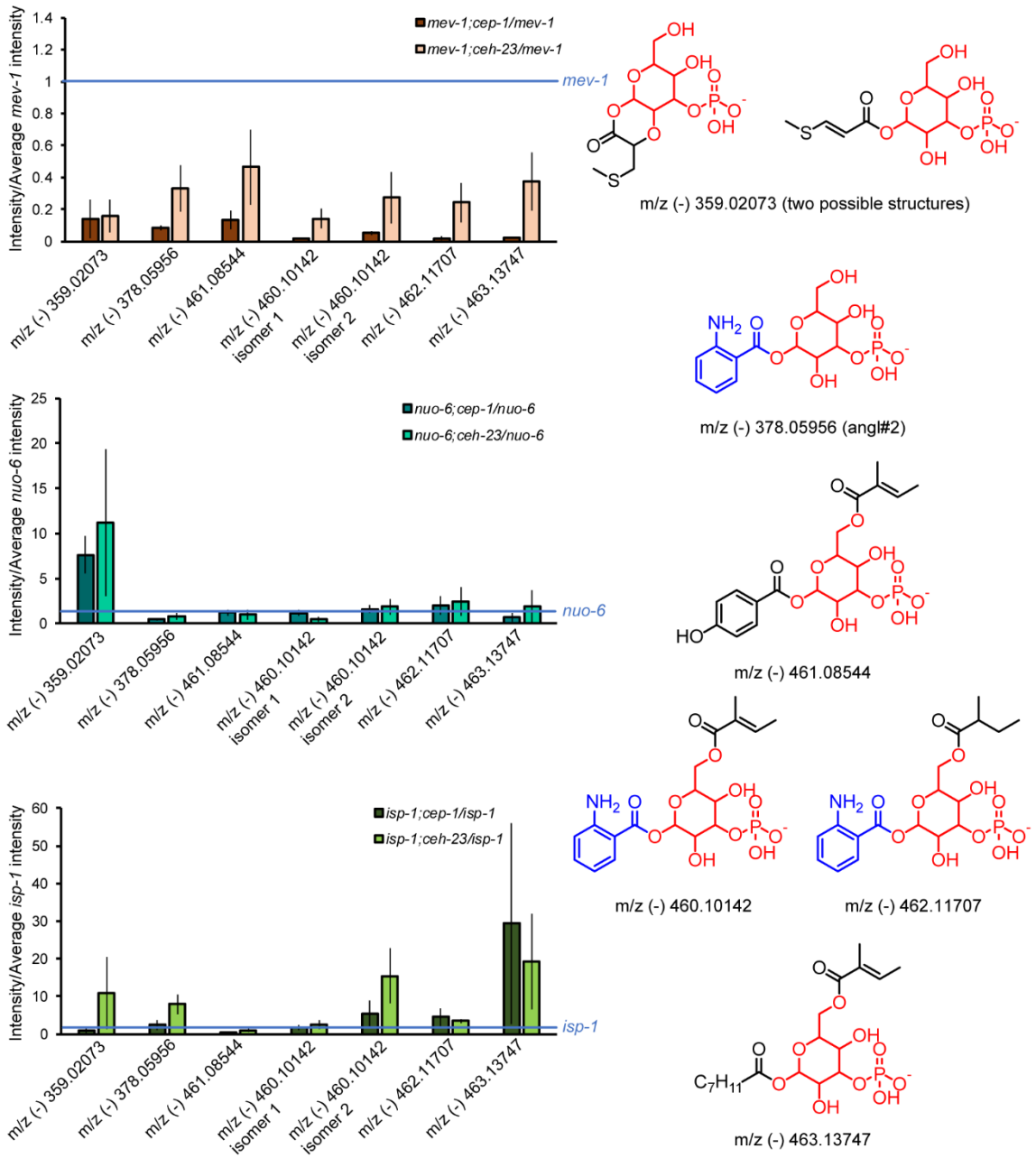

Figure 4.3. Phosphohexose derivatives in the double mutants. At the metabolomic level, the double mutants of *cep-1(gk138)* and *ceh-23(ms23)* both appear to rescue the metabolite abundance phenotype for all differentially-produced phosphohexose derivatives in the short-lived *mev-1(kn1)* mutant, and for some of these compounds in the long-lived *isp-1(qm150)* and *nuo-6(qm200)* mutants, albeit to varying degrees.Normalized intensities in ESI negative ionization mode are shown here. Error bars = s.e. of at least three independent biological replicates.

chemical cocktail originating the lysosome-related organelles in *C. elegans*, which causes a burst of intense blue fluorescence at the time of organismal death[24]. By analogy with angl#2 and similar compounds[24], one can speculate that the hexose (most likely glucose) in these phosphohexose

derivatives, is phosphorylated at position 3. The accumulation of angl#2 and similar anthranilate derivatives in the short-lived *mev-1* mutant suggests a possible signaling role for these metabolites in governing lifespan. Alternatively, the phosphohexose moiety could also be thought of as a scavenger group to rapidly collect products of primary metabolic pathways and remove them from the cell to prevent toxicity at otherwise higher concentrations[25]. In the long-lived mutants, decreased metabolic rate[11] possibly results in a 'cleaner' cell, which causes much slower accumulation of these metabolites, thereby reducing the need for phosphohexose conjugation to purge them from the cells.

Nucleic acid derivatives

Comparative metabolomic analysis of the WM samples from the Mit mutants revealed a feature at m/z 134.04722 (ESI-) that was upregulated in the short-lived *mev-1* mutant, but downregulated in both the long-lived *isp-1* and *nuo-6* mutants. The m/z value corresponds to the molecular formula $C_5H_4N_5^-$, which may suggest adenine. However, the retention time for this feature did not match with that of a synthetic standard of adenine. Upon closer inspection, this feature was found to be a fragment of m/z 307.11603 ($C_{12}H_{15}N_6O_4^-$), which followed the same trend in abundance as the adenine fragment. The search for an m/z of 293.10038 ($C_{11}H_{13}N_6O_4^-$), which is one methylene (-CH_2) group less than m/z 307.11603, revealed a compound with similar MS/MS fragmentation pattern as $C_{12}H_{15}N_6O_4^-$. This compound was found to be downregulated in all long-lived and short-lived mutants relative to N2, although its absolute amounts were higher in *mev-1* compared to both *isp-1* and *nuo-6*. MS/MS fragmentation and 2D NMR spectroscopy revealed these compounds to be ethyl and methyl esters of threonylcarbamoyladenine (Appendix Figures D4 and D8), and were assigned the SMIDs (www.smid-db.org) tade#2 and tade#1, respectively (Figure 4.4). Two closely-eluting isomers

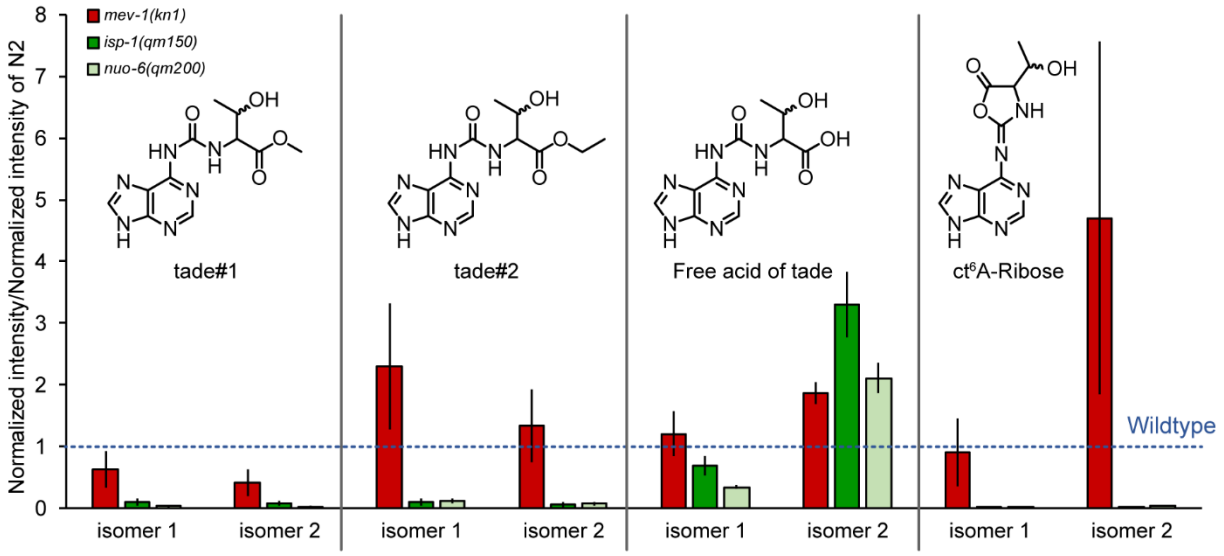

Figure 4.4. The tade family of metabolites in the Mit mutant excretome. The SMID '**tade**' is assigned based on the **t**hreonylcarbamoyl**ade**nine moiety in this family of nucleoside derivatives. Normalized intensities in ESI negative ionization mode are shown here (except for ct^{6}A-Ribose, which ionizes in ESI positive mode. All isomers follow a similar trend in abundances in worm pellet samples. Error bars = s.e. of at least 3 biological replicates.

were found for each of these features, possibly corresponding to stereo- or positional isomers, although comparison with synthetic standards and NMR spectroscopy will be required for confirmation. The double mutants appear to rescue the metabolite abundance phenotype for most of these compounds in all the Mit mutants, to varying degrees (Figure 4.5).

The threonylcarbamoyladenine moiety in tade#1 and tade#2 is likely derived from threonylcarbamoyladenosine (t^{6}A). t^{6}A is a highly conserved tRNA modification found at position 37 of the anticodon loop of nearly all ANN decoding tRNAs[26], and is believed to prevent the collapse of the anticodon loop, thereby improving translational fidelity. In *Pristionchus pacificus*, the ribose of t^{6}A possibly gets converted to xylose, and paratoside conjugation occurs to form the ascaroside npar#1[27] (Figure 4.6). npar#1 is not produced by *C. elegans*, and the ascr#3 conjugates of t^{6}A in *C. elegans*, nuclas#33 and nuclas#34 (Figure 4.6, Artyukhin *et al.*, *manuscript in revision*) are not differential among the Mit mutants. The *N*-ribosylated versions of tade#1 and tade#2 were

also not differentially produced in the Mit

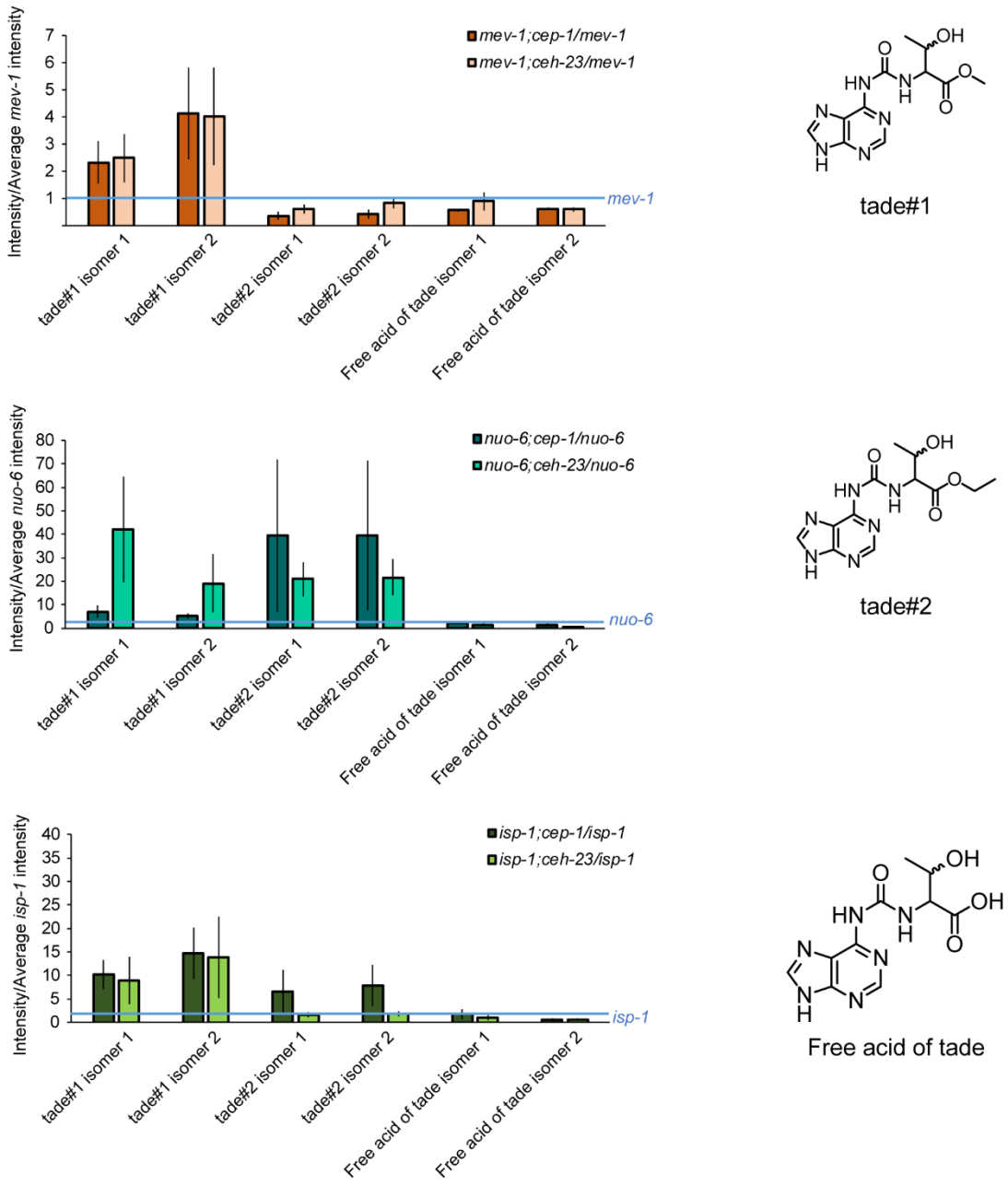

Figure 4.5. Nucleoside derivatives in the double mutants. At the metabolomic level, the double mutants of *cep-1(gk138)* and *ceh-23(ms23)* both appear to rescue the metabolite abundance phenotype for most of these derivatives in the short-lived *mev-1(kn1)* mutant, as well as in the long-lived *isp-1(qm150)* and *nuo-6(qm200)* mutants, albeit to varying degrees. Normalized intensities in ESI negative ionization mode are shown here. Error bars = s.e. of at least three independent biological replicates.

mutants. Interestingly, the free acid of tade#1 and tade#2 was not found to be differential in this system. It is possible that the ethyl and methyl esters of t^6A are artifacts of extraction that occurred

during sample preparation using methanol and ethanol as extraction solvents, and the actual differential metabolite is derived from the cyclic t^6A (ct^6A) derivative[28] (Figure 4.4). A pair of features corresponding to the mass of the ct^6A derivative (m/z 263.08871, $C_{10}H_{11}N_6O_3^+$, annotated as ct^6A-Ribose isomers 1 and 2) was found to be somewhat upregulated in *mev-1* and severely downregulated in both *isp-1* and *nuo-6* (Figure 4.4). Co-injection experiments with synthetic standards should be done to confirm the identities of these features. Extraction of the lyophilized WM samples with deuterated solvents, and corresponding shifts in the m/z values of these features should also be done to confirm whether tade#1 and tade#2 are indeed artifacts of sample preparation.

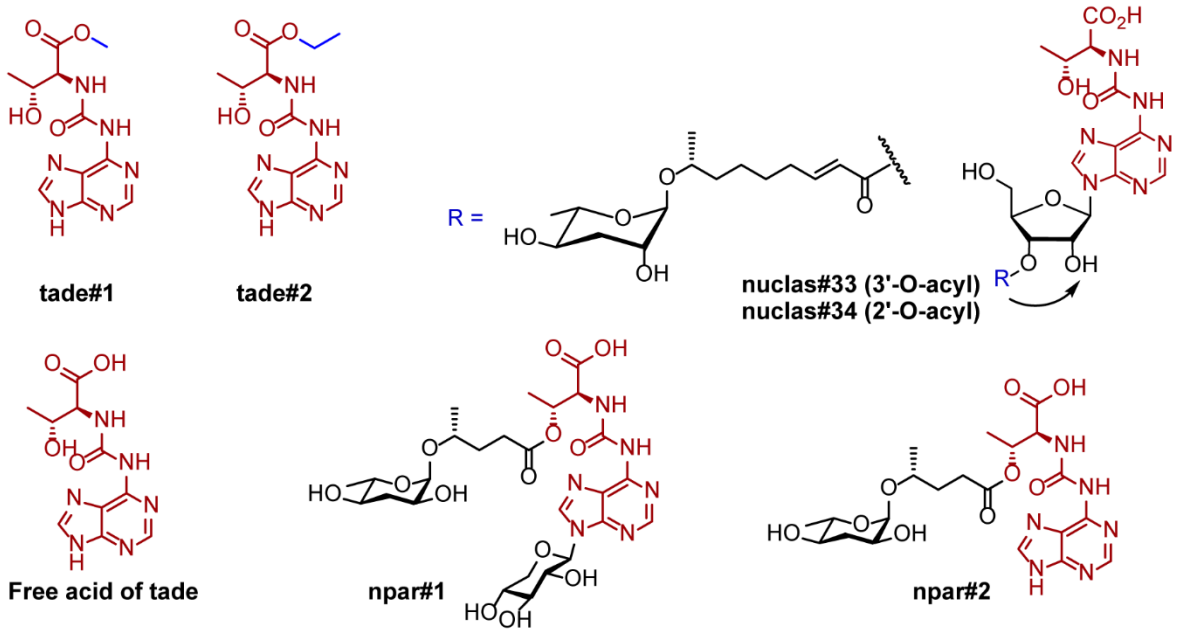

Figure 4.6. Structures of the tade family of compounds and known ascarosides in *P. pacificus* and *C. elegans* that contain the threonylcarbamoyladenine moiety (in brown).

Molecular networking in sub-cellular metabolomics

In Chapter 3, the author suggested the use of organelle-specific metabolomics to identify different components of biochemical pathways that may be involved in various aspects of nematode biology. The untargeted comparative metabolomics approach to study various organisms can be

extended to study metabolites in a tissue-specific or even sub-cellular level. To assess the utility of this approach in the sub-cellular level, the author collaborated with Dr. Yong Yu (Wang research group, Baylor College of Medicine). Dr. Yu isolated mitochondria from three independent biological replicates of wildtype, *isp-1(qm150)* and *nuo-6(qm200)* mutants. The author then analyzed the metabolome of the isolated mitochondria ('*mito*') and compared it to the metabolome of the remainder of the worm bodies ('*sup*') using high resolution UHPLC-MS. The phosphohexose derivatives described earlier appeared to be present only in the N2 *sup*, and absent the *mito* (Appendix Figure D9). These compounds were either not detected or severely downregulated in the *sup* of the *isp-1* and *nuo-6* mutants, which is consistent with the observations from the worms grown by the author in liquid culture. The tade family of compounds was found to be absent in the Mit mutants *mito* and *sup*, and very low in abundance in the N2 *sup*, possibly due to a lower number of worms used by Dr. Yu (~30,000 as opposed to over 1,000,000 in liquid culture used by the author).

Spectra for these samples were analyzed using the Bioconductor package XCMS (see Appendix D). Of the approximately 55,000 features that were detected, the author focused on the features that were enriched or depleted two-fold or more in each replicate of the *mito* relative to the *sup*. High-resolution MS/MS spectra were acquired for the features that were enriched or depleted in the mitochondria. These spectral files were then uploaded to the online workflow of Global Natural Products Social Molecular Networking (GNPS, University of California, San Diego)[29]. A molecular network was created by GNPS and analyzed using Cytoscape 3.5.1 (Figure 4.7). This networking technique is based on the degree of similarity of MS/MS fragments of different features. It allows one to visualize features that likely bear structural similarity, as suggested by their similar fragmentation patterns. In this figure, which is indicative of the chemical

space in the mitochondria as well as the rest of the worm body, the orange nodes represent features that are enriched in *mito*, and the blue nodes represent features that are

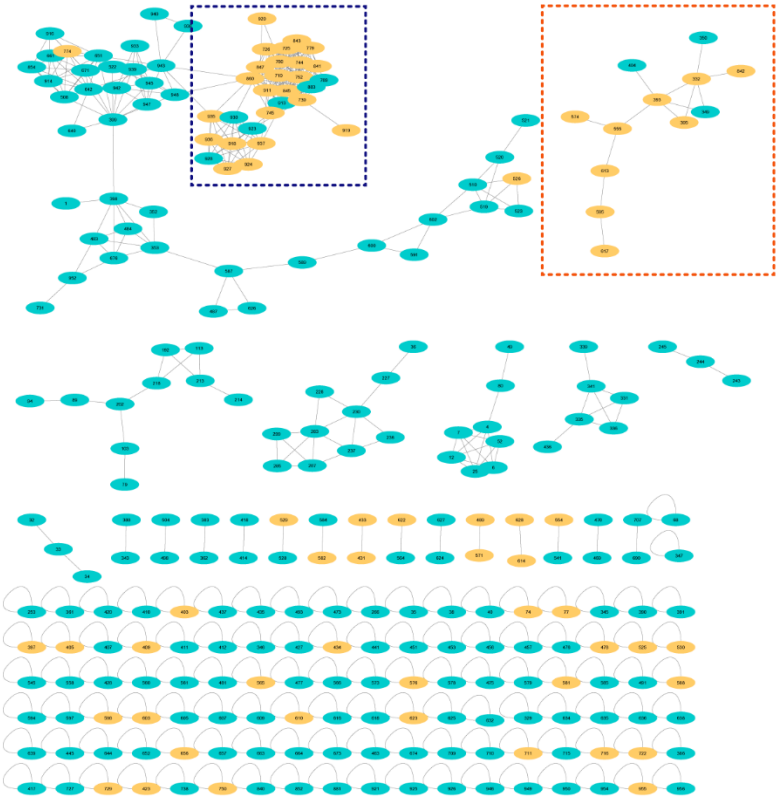

Figure 4.7. MS/MS network for the mitochondrial metabolome. Molecular network showing features in the ESI negative ionization mode, corresponding to metabolites that are enriched in the mitochondria (orange ellipses) relative to the rest of the worm body (blue ellipses). Several sturcturally related features form clusters e.g. features highlighted by the orange dotted box correspond to several *N*-terminal glutamyl dipeptides, features highlighted by the blue dotted box correspond to various phospholipids. Features that do not share common fragments with other features form self-loops at the bottom of the network.

enriched in *sup*. Some features tend to cluster well together in this chemical space, which can increase ease of elucidation of their structures as well as help understand biochemical pathways they may be part of. For example, the clustered nodes in the orange dotted box were found to be glutamate and dipeptides containing an *N*-terminal glutamyl moiety, such as Glu-Met, Glu-Phe, Glu-Val, Glu-Leu and Glu-Ile (Appendix Figure D5). The glutamate dipeptides likely have a γ-glutamyl moiety, as evidenced by a strong MS/MS fragment of m/z (-) 128.03532[30], although comparison with synthetic standards will be required to confirm their structures. Clustered nodes in the blue dotted box correspond to various phosphatidylethanolamines (Appendix Figure D6-

D7). The structures of these compounds were assigned based on the analysis of the MS/MS fragmentation data. The relatively low fold change limit of two-fold was deliberately chosen, as this increases the likelihood of a cluster to only include nodes enriched in the *mito* or enriched in the *sup*, relative to the likelihood of a cluster to feature nodes from both the *sup* and the *mito*. The author did not find any features that were consistently differential (*i.e.* over two-fold enriched or depleted in all three replicates) between Mit mutant *mito* samples and N2 *mito* samples.

The caveat to using this method for pathway analysis is that the networking algorithm uses only those features that were specified in the inclusion list for MS/MS fragmentation, which in turn were compiled following a manual analysis of fold change. If certain features are undetected or incorrectly integrated in the XCMS output they are not considered to be significantly enriched in the mitochondria in the manual analysis step, their corresponding features would not be included for MS/MS fragmentation and therefore be absent from the molecular network. Another potential problem with this method is that the molecular network is based on similarities of individual mass spectra, which is calculated using a cosine similarity score for the spectra being compared[31]. The cosine similarity score is proportional to the intensity of each m/z fragment in the spectrum[32]. For some features containing a few easily ionizable fragments (such as those corresponding to phosphate groups, which ionize very well in ESI negative mode), the cosine similarity score may be biased by the high intensities of these fragments, and the contribution of the lower intensity fragments to the similarity score may be low. This could result in incorrectly connected nodes in the network, for features that bear very little structural similarity to each other. Despite these disadvantages, MS/MS networking can help one visualize enriched metabolites and their corresponding biochemical pathways occurring in various organelles of the cell.

Discussion

The data presented in this chapter suggest that the Mit mutants of *C. elegans* could be a powerful system to study metabolites whose production correlates with lifespan, and could therefore be potential aging biomarkers. The author described two families of compounds, the phosphohexose derivatives in the *endo*-metabolome and the nucleoside derivatives in the *exo*-metabolome, that are upregulated in the short-lived mutants and downregulated in long-lived mutants. It is interesting to note that mutating the transcription factors *cep-1* and *ceh-23* rescues the metabotype of the Mit mutants, in addition to their lifespan phenotype.

mev-1(kn1) is known to have increased superoxide anion production[33], which may account for its shortened lifespan. This is in agreement with Harman's free radical theory of aging[34], which states that faster metabolism (and hence faster rate of living) in an organism, resulting in increased production of free radicals and reactive oxygen species, leads to shorter lifespan. Most of the different modules in the phosphohexose derivatives appear to be inputs from conserved primary metabolic pathways, so it is possible that *mev-1(kn1)* has an increased metabolic rate which leads to rapid production of these metabolites. The long-lived *isp-1(qm150)* and *nuo-6(qm200)* are known to have decreased metabolic rates[11], which could explain lower accumulation of primary metabolites and hence lower abundances of the phosphohexose derivatives.

Mitophagy, or the selective degradation of mitochondria by autophagy, also plays a role in the mitochondrion-mediated lifespan extension of *C. elegans*[35-36]. It has been shown that knocking down the DAF-16-controlled BNIP3 homolog DCT-1 and the PTEN-induced kinase PINK-1, two critical conserved components of mitophagy in *C. elegans*, shortens the lifespan of long-lived *isp-1(qm150)*[37]. This suggests that efficient clearing out of damaged mitochondria by mitophagy is essential for the extended lifespan of the organism. Although no significant difference in overall

autophagy was observed between *isp-1(qm150)*, *nuo-6(qm200)* and N2[11], it remains to be seen whether mitophagy specifically and autophagy in general is reduced in *mev-1(kn1)*. If this is indeed the case, it would support the hypothesis that sequestration of these primary metabolites by the phosphohexose moiety occurs to prime them for degradation in the lysosomes and related organelles. In the absence of efficient autophagy in *mev-1(kn1)*, these derivatives are found to be in the cytoplasm as opposed to being enclosed in lysosomes for degradation. Indeed, these compounds are much more abundant in the *sup* of N2 than those of the long-lived Mit mutants, and are altogether absent in the isolated mitochondria. It would be interesting to see whether these compounds are also abundant in the worm pellets of the acidic lysosome-related organelle-deficient *glo-1(zu437)* mutants, as well as in double mutants of *glo-1* and the Mit mutants. One could also look for the possible differential regulation of these compounds in other mitochondrial mutants such as the long-lived *clk-1* and short-lived *gas-1*.

t^6A has long been studied as a highly conserved tRNA modification occurring at position 37 of the anticodon loop of tRNAs decoding ANN codons[26]. It can be speculated that the ribose of t^6A gets converted to xylose, and conjugation of this molecule to a paratoside leads to npar#1 formation in *P. pacificus*[27]. In *C. elegans*, nuclas#33 and nuclas#34 are formed when t^6A gets conjugated to ascr#3 (Artyukhin *et al*., *manuscript in revision*). In this chapter, the author describes a third possible fate of t^6A in nematodes *i.e.*, the formation of tade#1 and tade#2, although the data suggest that de-ribosylated ct^6A may be the actual differential metabolite. It has recently been proposed that a diffusible factor or 'mitokine', whose identity remains to be determined, is released from one tissue in response to mitochondrial stress to relay longevity cues to other tissues[22]. As cleavage of tRNA is a conserved response to oxidative stress in eukaryotes[38], it is possible that the tade metabolites, which are likely derived from tRNA, pre-tRNA or tRNA fragments, are

components of such a cell-non-autonomous signal. The phosphohexose derivatives, which are downregulated in the long-lived mutants, could also be involved in the biosynthesis of this pro-longevity cue. Nematode aging assays with synthetic standards of all these compounds could reveal whether these metabolites influence aging by themselves, and their mechanisms of action could be elucidated using biochemical pathway mutants in similar assays. One can also look for the presence of these or similar compounds in other organisms that show mitochondrion-mediated lifespan extension, such as yeast, insects, mice etc. Experiments with synthetic standards of these compounds could then also be carried out in other these organisms, particularly disease models of mitochondrial dysfunction.

The utility of MS/MS networking has been highlighted in this chapter to study the metabolome of subcellular components. Based on their MS/MS fragments, several features cluster together and could potentially indicate participating metabolites in conserved primary metabolic pathways which occur in a specific organelle. Biomembranes, such as those of the mitochondria, are composed of phospholipids, which cluster together in the network shown in Figure 4.5. γ-glutamyl amino acids also form a cluster in the network, indicating their enrichment in the mitochondria. These compound classes show similar enrichment trends in the *mito* and *sup* of the long-lived *isp-1* and *nuo-6*, and no significant differences were seen between mutant and wildtype samples (both isolated *mito* and mixed stage liquid cultures). For effective transport across membranes, amino acids are converted to their γ-glutamyl derivatives by a γ-glutamyltranspeptidase enzyme that utilizes glutathione as a substrate in the Meister cycle[39]. Once inside the mitochondria, these amino acids can be used for mitochondrial protein synthesis[40], or be catabolized into precursors or intermediates for processes like the tricarboxylic acid cycle[41].

Conclusion

Untargeted comparative metabolomics based on high resolution mass spectrometry has emerged as a powerful tool in the annotation of the metabolomes of various organisms. In this chapter, the author has shown that investigating the mitochondrion-mediated lifespan extension and reduction in *C. elegans* using this technique can reveal several differentially regulated metabolites, which could potentially be biomarkers of aging in nematodes. These metabolites could be synthesized for use in biological assays to study mechanisms of healthy aging, resistance to ROS etc. Mitochondrial dysfunction has been implicated in several age-related disorders in humans such as Alzheimer's disease[42-43], diabetes[42, 44], Parkinson's disease[42-43], as well as in rare juvenile diseases such as Leigh syndrome[45]. The metabolites discovered in this study appear to be derived from primary metabolic pathways that are highly conserved across all kingdoms of life. Such compounds, and their mammalian counterparts, could be used to study the role of transcription factors such as the oncogene p53 in aging and mitochondrial diseases in humans.

REFERENCES

1. Nussbaum, R. L., Mining yeast in silico unearths a golden nugget for mitochondrial biology. *J Clin Invest* **2005,** *115* (10), 2689-91.

2. Lapointe, J.; Hekimi, S., Early mitochondrial dysfunction in long-lived *Mclk*1+/- mice. *The Journal of biological chemistry* **2008,** *283* (38), 26217-27.

3. Lee, S. S.; Lee, R. Y.; Fraser, A. G.; Kamath, R. S.; Ahringer, J.; Ruvkun, G., A systematic RNAi screen identifies a critical role for mitochondria in *C. elegans* longevity. *Nature genetics* **2003,** *33* (1), 40-8.

4. Liu, X.; Jiang, N.; Hughes, B.; Bigras, E.; Shoubridge, E.; Hekimi, S., Evolutionary conservation of the *clk-1*-dependent mechanism of longevity: loss of mclk1 increases cellular fitness and lifespan in mice. *Genes & development* **2005,** *19* (20), 2424-34.

5. Dillin, A.; Hsu, A. L.; Arantes-Oliveira, N.; Lehrer-Graiwer, J.; Hsin, H.; Fraser, A. G.; Kamath, R. S.; Ahringer, J.; Kenyon, C., Rates of behavior and aging specified by mitochondrial function during development. *Science* **2002,** *298* (5602), 2398-401.

6. Copeland, J. M.; Cho, J.; Lo, T., Jr.; Hur, J. H.; Bahadorani, S.; Arabyan, T.; Rabie, J.; Soh, J.; Walker, D. W., Extension of *Drosophila* life span by RNAi of the mitochondrial respiratory chain. *Current biology : CB* **2009,** *19* (19), 1591-8.

7. Kirchman, P. A.; Kim, S.; Lai, C. Y.; Jazwinski, S. M., Interorganelle signaling is a determinant of longevity in *Saccharomyces cerevisiae*. *Genetics* **1999,** *152* (1), 179-90.

8. Ishii, N.; Fujii, M.; Hartman, P. S.; Tsuda, M.; Yasuda, K.; Senoo-Matsuda, N.; Yanase, S.; Ayusawa, D.; Suzuki, K., A mutation in succinate dehydrogenase cytochrome b causes oxidative stress and ageing in nematodes. *Nature* **1998,** *394* (6694), 694-7.

9. Walker, D. W.; Hajek, P.; Muffat, J.; Knoepfle, D.; Cornelison, S.; Attardi, G.; Benzer, S., Hypersensitivity to oxygen and shortened lifespan in a *Drosophila* mitochondrial complex II mutant. *Proc Natl Acad Sci U S A* **2006,** *103* (44), 16382-7.

10. Feng, J.; Bussiere, F.; Hekimi, S., Mitochondrial electron transport is a key determinant of life span in *Caenorhabditis elegans*. *Dev Cell* **2001,** *1* (5), 633-44.

11. Yang, W.; Hekimi, S., Two modes of mitochondrial dysfunction lead independently to lifespan extension in *Caenorhabditis elegans*. *Aging Cell* **2010,** *9* (3), 433-47.

12. Kondo, M.; Senoo-Matsuda, N.; Yanase, S.; Ishii, T.; Hartman, P. S.; Ishii, N., Effect of oxidative stress on translocation of DAF-16 in oxygen-sensitive mutants, *mev-1* and *gas-1* of *Caenorhabditis elegans*. *Mech Ageing Dev* **2005,** *126* (6-7), 637-41.

13. Baruah, A.; Chang, H.; Hall, M.; Yuan, J.; Gordon, S.; Johnson, E.; Shtessel, L. L.; Yee, C.; Hekimi, S.; Derry, W. B.; Lee, S. S., CEP-1, the *Caenorhabditis elegans* p53 homolog, mediates opposing longevity outcomes in mitochondrial electron transport chain mutants. *PLoS genetics* **2014,** *10* (2), e1004097.

14. Chang, H. W.; Pisano, S.; Chaturbedi, A.; Chen, J.; Gordon, S.; Baruah, A.; Lee, S. S., Transcription factors CEP-1/p53 and CEH-23 collaborate with AAK-2/AMPK to modulate longevity in *Caenorhabditis elegans*. *Aging Cell* **2017,** *16* (4), 814-824.

15. Walter, L.; Baruah, A.; Chang, H. W.; Pace, H. M.; Lee, S. S., The homeobox protein CEH-23 mediates prolonged longevity in response to impaired mitochondrial electron transport chain in *C. elegans*. *PLoS Biol* **2011,** *9* (6), e1001084.

16. Derry, W. B.; Putzke, A. P.; Rothman, J. H., *Caenorhabditis elegans* p53: role in apoptosis, meiosis, and stress resistance. *Science* **2001,** *294* (5542), 591-5.

17. Greiss, S.; Schumacher, B.; Grandien, K.; Rothblatt, J.; Gartner, A., Transcriptional profiling in *C. elegans* suggests DNA damage dependent apoptosis as an ancient function of the p53 family. *BMC Genomics* **2008,** *9*, 334.

18. Yanase, S.; Suda, H.; Yasuda, K.; Ishii, N., Impaired p53/CEP-1 is associated with lifespan extension through an age-related imbalance in the energy metabolism of *C. elegans*. *Genes Cells* **2017**.

19. Shen, L.; Hu, Y.; Cai, T.; Lin, X.; Wang, D., Regulation of longevity by genes required for the functions of AIY interneuron in nematode *Caenorhabditis elegans*. *Mech Ageing Dev* **2010,** *131* (11-12), 732-8.

20. Butler, J. A.; Mishur, R. J.; Bhaskaran, S.; Rea, S. L., A metabolic signature for long life in the *Caenorhabditis elegans* Mit mutants. *Aging Cell* **2013,** *12* (1), 130-8.

21. Butler, J. A.; Ventura, N.; Johnson, T. E.; Rea, S. L., Long-lived mitochondrial (Mit) mutants of *Caenorhabditis elegans* utilize a novel metabolism. *FASEB J* **2010,** *24* (12), 4977-88.

22. Durieux, J.; Wolff, S.; Dillin, A., The cell-non-autonomous nature of electron transport chain-mediated longevity. *Cell* **2011,** *144* (1), 79-91.

23. Smith, C. A.; Want, E. J.; O'Maille, G.; Abagyan, R.; Siuzdak, G., XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem* **2006,** *78* (3), 779-87.

24. Coburn, C.; Allman, E.; Mahanti, P.; Benedetto, A.; Cabreiro, F.; Pincus, Z.; Matthijssens, F.; Araiz, C.; Mandel, A.; Vlachos, M.; Edwards, S. A.; Fischer, G.; Davidson, A.; Pryor, R. E.; Stevens, A.; Slack, F. J.; Tavernarakis, N.; Braeckman, B. P.; Schroeder, F. C.; Nehrke, K.; Gems, D., Anthranilate fluorescence marks a calcium-propagated necrotic wave that promotes organismal death in *C. elegans*. *PLoS Biol* **2013,** *11* (7), e1001613.

25. Stupp, G. S.; von Reuss, S. H.; Izrayelit, Y.; Ajredini, R.; Schroeder, F. C.; Edison, A. S., Chemical detoxification of small molecules by *Caenorhabditis elegans*. *ACS chemical biology* **2013,** *8* (2), 309-13.

26. Thiaville, P. C.; Legendre, R.; Rojas-Benitez, D.; Baudin-Baillieu, A.; Hatin, I.; Chalancon, G.; Glavic, A.; Namy, O.; de Crecy-Lagard, V., Global translational impacts of the loss of the tRNA modification t6A in yeast. *Microb Cell* **2016,** *3* (1), 29-45.

27. Bose, N.; Ogawa, A.; von Reuss, S. H.; Yim, J. J.; Ragsdale, E. J.; Sommer, R. J.; Schroeder, F. C., Complex small-molecule architectures regulate phenotypic plasticity in a nematode. *Angew Chem Int Ed Engl* **2012,** *51* (50), 12438-43.

28. Miyauchi, K.; Kimura, S.; Suzuki, T., A cyclic form of N6-threonylcarbamoyladenosine as a widely distributed tRNA hypermodification. *Nat Chem Biol* **2013,** *9* (2), 105-11.

29. Wang, M.; Carver, J. J.; Phelan, V. V.; Sanchez, L. M.; Garg, N.; Peng, Y.; Nguyen, D. D.; Watrous, J.; Kapono, C. A.; Luzzatto-Knaan, T.; Porto, C.; Bouslimani, A.; Melnik, A. V.; Meehan, M. J.; Liu, W. T.; Crusemann, M.; Boudreau, P. D.; Esquenazi, E.; Sandoval-Calderon, M.; Kersten, R. D.; Pace, L. A.; Quinn, R. A.; Duncan, K. R.; Hsu, C. C.; Floros, D. J.; Gavilan, R. G.; Kleigrewe, K.; Northen, T.; Dutton, R. J.; Parrot, D.; Carlson, E. E.; Aigle, B.; Michelsen, C. F.; Jelsbak, L.; Sohlenkamp, C.; Pevzner, P.; Edlund, A.; McLean, J.; Piel, J.; Murphy, B. T.; Gerwick, L.; Liaw, C. C.; Yang, Y. L.; Humpf, H. U.; Maansson, M.; Keyzers, R. A.; Sims, A. C.; Johnson, A. R.; Sidebottom, A. M.; Sedio, B. E.; Klitgaard, A.; Larson, C. B.; P, C. A. B.; Torres-Mendoza, D.; Gonzalez, D. J.; Silva, D. B.; Marques, L. M.; Demarque, D. P.; Pociute, E.; O'Neill, E. C.; Briand, E.; Helfrich, E. J. N.; Granatosky, E. A.; Glukhov, E.; Ryffel, F.; Houson, H.;

Mohimani, H.; Kharbush, J. J.; Zeng, Y.; Vorholt, J. A.; Kurita, K. L.; Charusanti, P.; McPhail, K. L.; Nielsen, K. F.; Vuong, L.; Elfeki, M.; Traxler, M. F.; Engene, N.; Koyama, N.; Vining, O. B.; Baric, R.; Silva, R. R.; Mascuch, S. J.; Tomasi, S.; Jenkins, S.; Macherla, V.; Hoffman, T.; Agarwal, V.; Williams, P. G.; Dai, J.; Neupane, R.; Gurr, J.; Rodriguez, A. M. C.; Lamsa, A.; Zhang, C.; Dorrestein, K.; Duggan, B. M.; Almaliti, J.; Allard, P. M.; Phapale, P.; Nothias, L. F.; Alexandrov, T.; Litaudon, M.; Wolfender, J. L.; Kyle, J. E.; Metz, T. O.; Peryea, T.; Nguyen, D. T.; VanLeer, D.; Shinn, P.; Jadhav, A.; Muller, R.; Waters, K. M.; Shi, W.; Liu, X.; Zhang, L.; Knight, R.; Jensen, P. R.; Palsson, B. O.; Pogliano, K.; Linington, R. G.; Gutierrez, M.; Lopes, N. P.; Gerwick, W. H.; Moore, B. S.; Dorrestein, P. C.; Bandeira, N., Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat Biotechnol* **2016,** *34* (8), 828-837.

30. Harrison, A. G., Characterization of alpha- and gamma-glutamyl dipeptides by negative ion collision-induced dissociation. *J Mass Spectrom* **2004,** *39* (2), 136-44.

31. Kim, S.; Zhang, X., Comparative analysis of mass spectral similarity measures on peak alignment for comprehensive two-dimensional gas chromatography mass spectrometry. *Comput Math Methods Med* **2013,** *2013*, 509761.

32. Stein, S. E.; Scott, D. R., Optimization and testing of mass spectral library search algorithms for compound identification. *J Am Soc Mass Spectrom* **1994,** *5* (9), 859-66.

33. Senoo-Matsuda, N.; Yasuda, K.; Tsuda, M.; Ohkubo, T.; Yoshimura, S.; Nakazawa, H.; Hartman, P. S.; Ishii, N., A defect in the cytochrome b large subunit in complex II causes both superoxide anion overproduction and abnormal energy metabolism in *Caenorhabditis elegans*. *The Journal of biological chemistry* **2001,** *276* (45), 41553-8.

34. Harman, D., Aging: a theory based on free radical and radiation chemistry. *J Gerontol* **1956,** *11* (3), 298-300.

35. Schiavi, A.; Maglioni, S.; Palikaras, K.; Shaik, A.; Strappazzon, F.; Brinkmann, V.; Torgovnick, A.; Castelein, N.; De Henau, S.; Braeckman, B. P.; Cecconi, F.; Tavernarakis, N.; Ventura, N., Iron-Starvation-Induced Mitophagy Mediates Lifespan Extension upon Mitochondrial Stress in *C. elegans*. *Current biology : CB* **2015,** *25* (14), 1810-22.

36. Ryu, D.; Mouchiroud, L.; Andreux, P. A.; Katsyuba, E.; Moullan, N.; Nicolet-Dit-Felix, A. A.; Williams, E. G.; Jha, P.; Lo Sasso, G.; Huzard, D.; Aebischer, P.; Sandi, C.; Rinsch, C.; Auwerx, J., Urolithin A induces mitophagy and prolongs lifespan in *C. elegans* and increases muscle function in rodents. *Nat Med* **2016,** *22* (8), 879-88.

37. Palikaras, K.; Lionaki, E.; Tavernarakis, N., Coordination of mitophagy and mitochondrial biogenesis during ageing in *C. elegans*. *Nature* **2015,** *521* (7553), 525-8.

38. Thompson, D. M.; Lu, C.; Green, P. J.; Parker, R., tRNA cleavage is a conserved response to oxidative stress in eukaryotes. *RNA* **2008,** *14* (10), 2095-2103.

39. Orlowski, M.; Meister, A., The gamma-glutamyl cycle: a possible transport system for amino acids. *Proc Natl Acad Sci U S A* **1970,** *67* (3), 1248-55.

40. Fox, T. D., Mitochondrial protein synthesis, import, and assembly. *Genetics* **2012,** *192* (4), 1203-34.

41. Hildebrandt, T. M.; Nunes Nesi, A.; Araujo, W. L.; Braun, H. P., Amino Acid Catabolism in Plants. *Mol Plant* **2015,** *8* (11), 1563-79.

42. Pieczenik, S. R.; Neustadt, J., Mitochondrial dysfunction and molecular pathways of disease. *Exp Mol Pathol* **2007,** *83* (1), 84-92.

43. Stavrovskaya, I. G.; Kristal, B. S., The powerhouse takes control of the cell: is the mitochondrial permeability transition a viable therapeutic target against neuronal dysfunction and death? *Free Radic Biol Med* **2005,** *38* (6), 687-97.

44. Wallace, D. C., A mitochondrial paradigm of metabolic and degenerative diseases, aging, and cancer: a dawn for evolutionary medicine. *Annu Rev Genet* **2005,** *39*, 359-407.

45. Lake, N. J.; Compton, A. G.; Rahman, S.; Thorburn, D. R., Leigh syndrome: One disorder, more than 75 monogenic causes. *Ann Neurol* **2016,** *79* (2), 190-203.

CHAPTER 5: COMBINING UNTARGETED METABOLOMICS WITH GENOME-WIDE ASSOCIATION STUDIES REVEALS CANDIDATE GENES FOR METABOLITE BIOSYNTHESES

Introduction

The size of metabolome, the collection of small molecules produced by an organism, rivals that of the genome and proteome[1-3]. Even for well-studied model organisms such as *Caenorhabditis elegans*, *Drosophila*, *Arabidopsis*, or mouse, the structures and biosynthetic pathways of most endogenously produced small molecules remain to be elucidated. As the functions of roughly half of the genes in model organisms remain unknown[4], it seems reasonable to assume that many of these are involved in the biosynthesis of known as well hitherto unidentified small molecules. However, effective tools are lacking that could facilitate associating the vast numbers of new metabolites currently being discovered with their respective biosynthetic pathways. Natural genetic variation provides an important resource that can be leveraged to identify diverse genetic traits using genome-wide association (GWA) mappings[5]. The utility of GWA mappings for identifying genetic loci involved in a variety of phenotypes in *C. elegans* and other model systems is well established[6-9], and genome-wide association studies (GWAS) have contributed to the characterization of biosynthetic pathways in plants[10-12]. However, the potential of GWA to uncover the biosyntheses of the vast numbers of new metabolites currently being identified in animals has not been extensively explored.

Recent efforts by the Schroeder Research Group (Cornell University) combined intra-specific (natural) variation in ascaroside production in 264 wild isolates of *Pristionchus pacificus* with genome-wide association studies (GWAS), CRISPR-Cas9-induced genetic engineering, and

detailed metabolomic analysis of mutant lines (Falcke, Bose *et al.*, *manuscript submitted*). Targeted metabolomics-based screening for known ascarosides in the *exo*-metabolome of these *P. pacificus* natural isolates identified several strains that do not synthesize the dauer-inducing ascarosides ubas#1 and ubas#2 (Figure 5.1a). All these natural isolates were genotyped, and the variation in the abundances of the ubas ascarosides were linked to genotypic variance. An approximately 30 kb genomic locus was identified on Chromosome I that significantly correlated with variation in ubas production. Careful resequencing of this region resulted in the detection of an approximately 2.6 kb-deletion that is specific to all "no ubas" strains and contains a putative carboxylesterase-coding gene. A deletion mutant for this gene was created using CRISPR/Cas9 in a "high ubas" reference strain RS2333. The author compared the *endo*- and *exo*-metabolomes of this mutant with RS2333 (*P. pacificus* wildtype) and observed over 100-fold reduction in the production of ubas#1 and ubas#2 (Figure 5.1b). This gene was named *Ppa-uar-1* (**u**bas-**a**scaroside-**r**equired).

To further characterize the effect of *Ppa-uar-1* on metabolism, Dr. Alexander B. Artyukhin (Schroeder research group, Cornell University) compared the *exo*-metabolomes of the mutant and RS2333 using the untargeted comparative metabolomics technique described in Chapter 4. This comparative metabolomic analysis revealed seven compounds, in addition to ubas#1 and ubas#2, whose production was abolished in *Ppa-uar-1* mutants (Figure 5.1c). Based on their MS/MS fragmentation spectra, structures were proposed for these newly detected metabolites, named ubas#3 through ubas#9, all of which represent ascaroside derivatives that feature an ureidoisobutyryl moiety attached to ascarylose. This strongly suggests that *Ppa-uar-1* is indeed involved in attaching a ureidoisobutyryl moiety to ascarosides of diverse chemical structures. Thus, natural variation in *P. pacificus* revealed the putative esterase-encoding gene

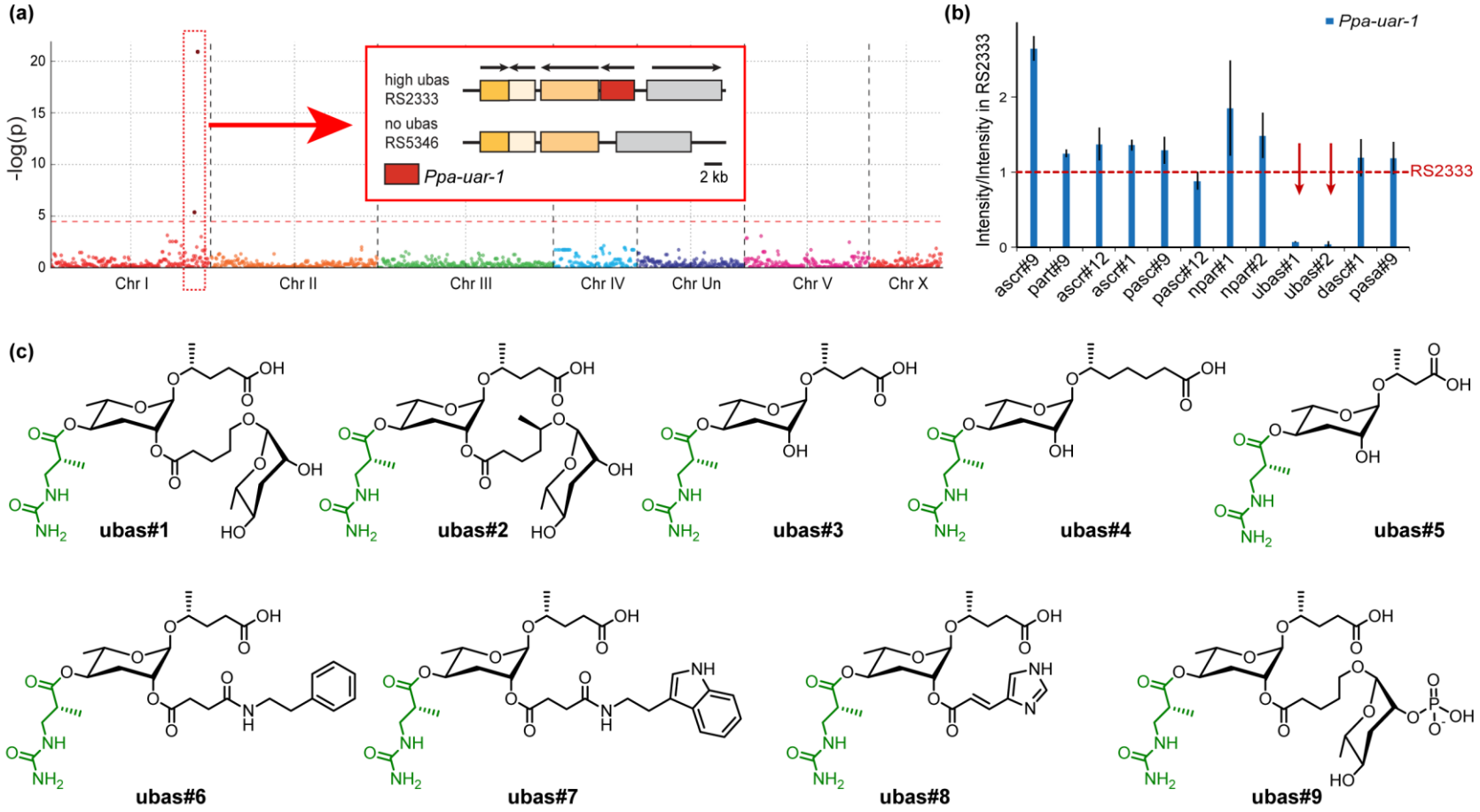

Figure 5.1. Identification of *Ppa-uar-1*. **(a)** Combining targeted metabolomics with GWAS reveals a genomic locus in *P. pacificus* that correlates significantly with variation in ubas#1 production across 264 natural isolates. **(b)** Deletion of this region by CRISPR/Cas9 results in almost complete abolishment of hitherto known ascarosides ubas#1 and ubas#2. Production of all other known ascarosides is similar to the *P. pacificus* wildtype strain RS2333. **(c)** Untargeted comparative metabolomics revealed seven additional ascarosides that are abolished in *Ppa-uar-1*, all of which contain a ureidoisobutyryl moiety highlighted in green.

Ppa-uar-1 to participate in the biosynthesis of a family of multi-modular ascarosides by the highly specific attachment of an ureidoisobutyryl side chain.

Genome-metabolome correlation in C. elegans

The aforementioned study in *Pristionchus pacificus* focused on natural variation in the abundances of a dozen known ascarosides only. Encouraged by the results from the targeted metabolomics approach, the author sought to extend this use of comparative metabolomics analysis to define functions of hitherto unannotated enzymes in *C. elegans*. Two aspects of this study in *C. elegans* are described in this chapter: (a) using variation in the abundances of known ascarosides in several wild isolates of *C. elegans* to arrive at possible candidate genes for their biosyntheses, *i.e.*, a targeted approach, and (b) coupling untargeted comparative metabolomics with GWA techniques to see which features map with high significance to different genomic loci, then using MS/MS data for those features to arrive at tentative structures, as well as looking for candidate genes in those loci that may be involved in their metabolism.

Dr. Mostafa Zamanian (Andersen research group, Northwestern University) provided 114 natural isolate strains of *C. elegans* (Appendix Table E1). In *C. elegans*, over 500 wild isolates that represent diverse genetic individuals are available and the 114 most genomically diverse selected for this study project represent over 85% of the total genetic diversity of the entire species[13]. In fact, these strains capture as much genetic diversity as some human sub-populations (*e.g.* Europeans), and therefore genome-wide association (GWA) mappings are readily performed to identify the genes that underlie quantitative trait differences[8, 13-14]. The author grew mixed stage liquid cultures for all these strains and obtained high resolution UHPLC-MS data for their *exo*-metabolome extracts. The data presented in this chapter come from a single replicate of all these natural isolate strains. Nonetheless, they lay the foundation for potential elucidation of several

biochemical pathways that will contribute to the structural and functional annotation of the *C. elegans* metabolome.

Natural variation in C. elegans ascaroside production

Abundances of known ascarosides in all the natural isolates were manually quantified using Thermo XCalibur QualBrowser software. Dr. Daehan Lee (Andersen research group, Northwestern University) then carried out a GWA mapping on the peak integration areas for these metabolites. It was found that ratios of oscr#9, an ascaroside with a 5-carbon side chain ω-linked to the ascarylose, to the ω-1-linked ascr#1 and ascr#2, mapped with high significance to a genomic locus on Chromosome I (Figure 5.2a, b). Fine mapping of this locus revealed variants in ACOX-1.1, ACOX-1.2 and ACOX-1.4 (Appendix Figure E1).

The ACOX family of enzymes catalyze the first step of the peroxisomal β-oxidation pathway[15], whereby longer-chained ascarosides are shortened iteratively, two carbons at a time (see Chapters 2 and 3). Different ACOX enzymes have different substrate specificities, and some of these enzymes form homo- and heterodimers to introduce a trans C2-C3 double bond in the acyl-CoA side chain of ascarosides[16-17]. The involvement of various acyl-CoA oxidases with different substrate specificities in the peroxisomal β-oxidation of fatty acids is conserved in plants and mammals as well[18]. In *C. elegans,* an ACOX-1.1/ACOX-1.3 heterodimer specifically oxidizes the ω-1-linked 7-carbon side chained ascr#1, which gets further modified to form the methylketone derivative ascr#2. The ACOX-1.2 homodimer specifically processes the ω-linked oscr#9, although the ACOX-1.1/ACOX-1.2 heterodimer may also play a role. Based on similarity of active site residues, ACOX-1.4 is predicted to prefer substrates similar in size to those of ACOX-1.1, such as the ω-1-linked 9-carbon side chained ascr#10. All the ACOX enzymes act on the Coenzyme A thioesters of these ascarosides[17]. Taken together, the GWAS results establish the involvement of

C. elegans ACOX-1.2 in oscr#9 biogenesis as a strong proof of concept.

In *P. pacificus*, the 'low ubas' natural isolates bear a deletion in the genomic region containing *uar-1*, whereas the 'high ubas' strains do not. A deletion allele was created in a 'high ubas' genetic background, and this *allele swap* resulted in the complete reversal of the metabotype, *i.e.*, the 'high ubas' strain RS2333 was converted to a 'low ubas' one. In the present study using *C. elegans*, a similar allele swap between a 'high' and 'low oscr#9: ascr#1 (or ascr#2)' strain will be the next step to unambiguously establish that the observed metabotype is due to the involvement of the ACOX enzymes.

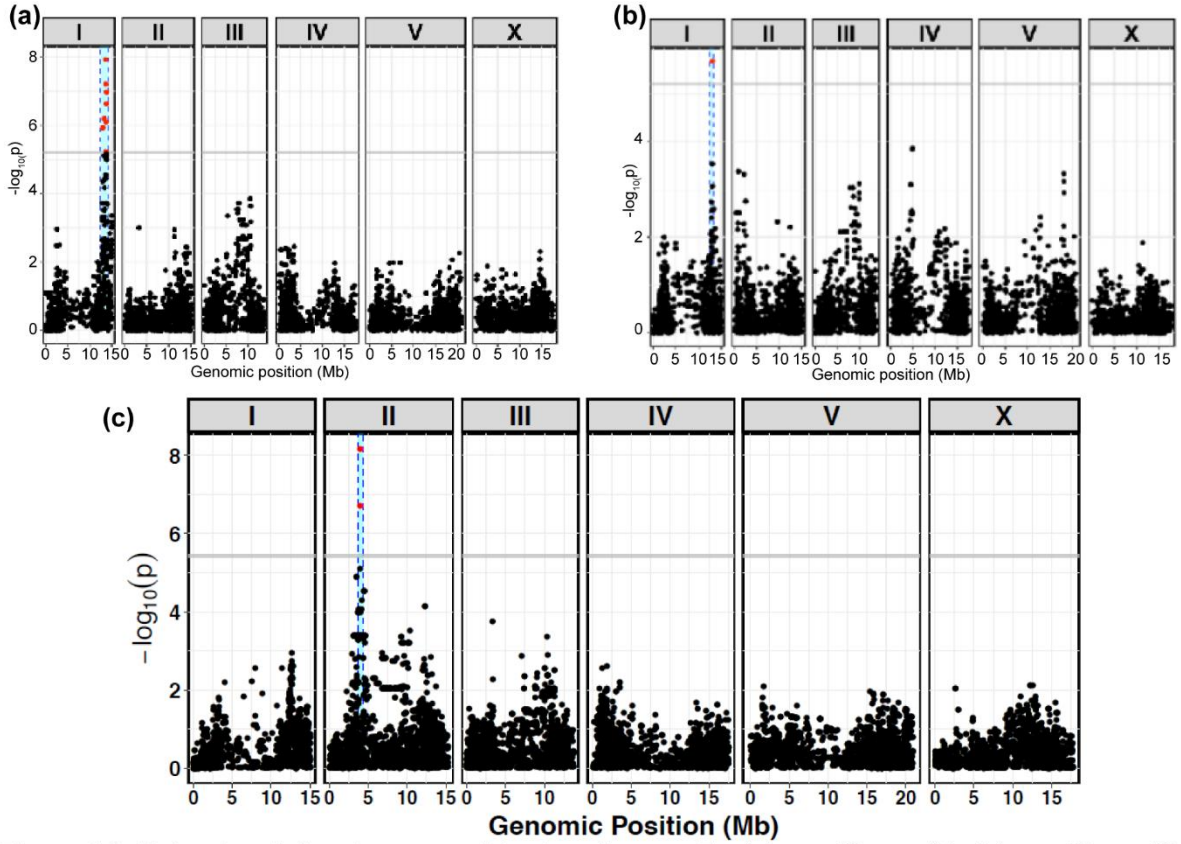

Figure 5.2. Natural variation in ascaroside abundances. For **(a)** oscr#9: ascr#1, **(b)** oscr#9:ascr#2 and **(c)** ascr#10:ascr#3. Blue boxes indicate mQTLs, and horizontal grey lines show the Bonferroni correction threshold.

In collaboration with Dr. Lee, the author also found that the ratio of ascr#10 to ascr#3 maps with high significance to a QTL on Chromosome II (Figure 5.2c), and fine mapping revealed

several variants in a locus containing the gene *sea-2* (Appendix Figure E2). In addition to being important components of the dauer pheromone, ascr#3 has been shown to extend the lifespan of *C. elegans*[19], whereas ascr#10 reduces nematode lifespan and acts as a sex pheromone (Ludewig *et al*., *manuscript in preparation*). The ratio of ascr#10 to ascr#3 has gained a lot of attention in the recent years because of the sex-specificity of this trait[20-23]. While both ascarosides are quite abundant in males and hermaphrodites, males produce over three times more ascr#10 than ascr#3, and hermaphrodites produce over three times more ascr#3 than ascr#10. Recent studies have shown that feminization of the *C. elegans* male intestine decreases ascr#10 production (Izrayelit *et al.*, *unpublished results*). Ascr#3 and ascr#10 structurally differ from each other by only one double bond (Chapter 3, Figure 3.1).

The RNA-binding zinc-finger protein SEA-2 controls the X:A chromosomal ratio, and acts directly upstream of several masculinizing or feminizing enzymes (TRA-2, FEM-2 etc.)[24]. Therefore, *sea-2* plays a major role in sex determination, which is reflected in the ascr#10: ascr#3 ratio. Loss of function of *sea-2* has also been shown to extend adult lifespan in a *daf-16*-dependent manner[25]. It is possible that ascr#10 shortens lifespan via SEA-2 signaling. However, Dr. Lee found that while there are several single nucleotide variants in the *sea-2* coding region, none of these appears to correlate significantly with variation in the ascr#10: ascr#3 ratio. It is difficult to precisely nail down candidate variants because of highly divergent *sea-2* genomic region. The Andersen research group (Northwestern University) is using CRISPR/Cas9 technology to edit several target sites in this locus, in natural isolate backgrounds with high and low ascr#10: ascr#3 ratios. This experiment will be used to determine whether in fact natural variation at the *sea-2* locus influences the ascr#10:ascr#3 ratio, which may affect behavioral phenotypes and lifespan in *C. elegans*.

Candidates for 4′-modified ascaroside biosynthesis

Two wild isolate strains of *C. elegans* were found in which icas#3, icas#9 and icas#10 production was abolished. ECA36 and QX1794 were somewhat deficient in other 4′-modified ascarosides, while the abundances of simple ascarosides were similar to those of the average of all other strains (Figure 5.3). The icas ascarosides are among the most abundantly produced ascarosides, so the deficiency in icas production in these ECA36 and QX1794 is particularly conspicuous. These two strains were not drastically deficient in the unmodified (parent) ascaroside abundances for the icas ascarosides (ascr#3, ascr#9 and ascr#10, see Chapter 3, Figures 3.1-3.2 for structures), suggesting that these two strains possibly lack a 4′-modifying enzyme. Moreover, the difference in the abundances of other families of 4′-modified ascarosides such as osas, hbas and mbas was not as severe as the icas family (Figure 5.3), which suggests that an enzyme responsible for specifically attaching an indole-3-carboxyl moiety to the parent ascarosides is defective in these strains. Interestingly, upon careful examination of the XCMS-generated lists as well as manually searching for other indole derivatives, the author found no other features that were severely up- or downregulated only in these two strains, which further supports the hypothesis that a putative enzyme mutated in these two strains is highly specific in its biosynthetic function.

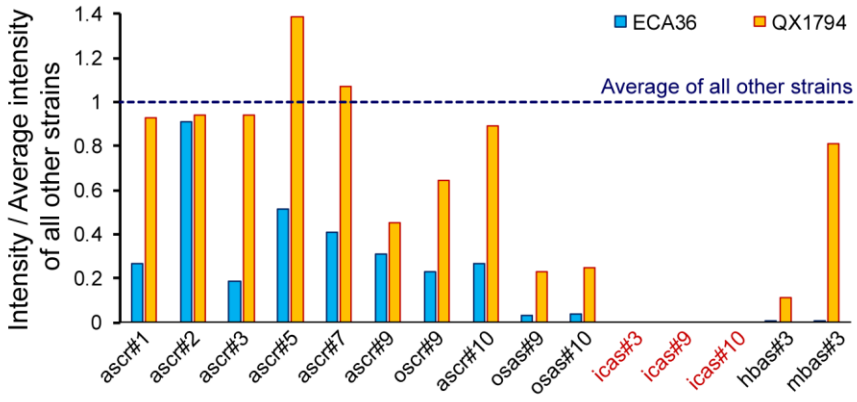

Figure 5.3. Ascaroside abundances in icas deficient strains. Peak intensities in ESI negative (ESI positive mode for ascr#2) mode for all ascarosides in ECA36 and QX1794 have been normalized to the average peak intensities of the other 112 wild isolate strains. icas#3, icas#9 and icas#10 are abolished in these two strains.

In *P. pacificus*, the only metabolomic differences between RS2333 (‘high ubas’) and the *Ppa-uar-1* deletion mutant were the presence or absence of nine ascarosides, all containing a 4′-ureidoisobutyryl moiety, the amounts of all other detected metabolites were unchanged. In *C. elegans*, the author discovered a similar situation where the only metabolomic differences between ECA36 and QX1794 on the one hand, and the remaining 112 natural isolates on the other, is the absence or presence of all ascarosides bearing a 4′-indole-3-carboxyl moiety, respectively. This highly specific metabotype is also reminiscent of the abolishment of only icas#9 and osas#9 ascarosides in the *acs-7* mutant described in Chapter 3. This evidence strongly suggests that ECA36 and QX1794 contain a non-functional variant of one very specific enzyme involved in attaching the indole-3-carboxyl moiety to the 4′-hydroxyl group of different ascarosides. *acs-7* is likely not a candidate for such a gene, since osas#9 is produced by these two strains.

A nonsense mutation in any gene resulting in a premature stop codon, that causes a large portion of the protein (specifically including its active site) to not be made, could be a likely cause behind the observed phenotype. Initial bioinformatic analysis by the Andersen research group revealed seven genes which have variants unique to ECA36 and QX1794 (Appendix Table E2). However, none of these variants point to a nonsense mutation these genes. Of these seven genes, only four have known or predicted functions, and none of them appear to have expected biosynthetic function of attaching the indole-3-carboxyl moiety to different ascarosides. Stefan Zdraljevic (Andersen research group, Northwestern University) also looked for genes that are affected by small indels, large deletions, large insertions and translocations in both ECA36 and QX1794, and not in any of the other 112 strains (Appendix Table E3). As the immediate next step, the laboratory wildtype strain N2, which is abundant in the production of icas ascarosides, will serve as an excellent background for editing out suitable candidate genes from this list using

CRISPR/Cas9 technology. This will be followed by a comparative metabolomic analysis to study the mutant's icas profile relative to N2. If such a mutant is indeed found to be deficient in the icas family of ascarosides, it would then be established as lacking the enzyme responsible for attaching an indole-3-carboxyl moiety to the 4′-hydroxyl position of different ascarosides. This would also suggest that this enzyme is an interacting partner of ACS-7 (see Chapter 3) in the biosynthesis of 4′-modified ascarosides[26]. To test this hypothesis, an assay with both these enzymes (recombinantly expressed and purified), indole-3-carboxylic acid and ascr#9 (or ascr#9-SCoA) should then be done to see whether icas#9 (or icas#9-SCoA) is formed *in vitro*.

Untargeted comparative metabolomics and GWAS

The UHPLC-HRMS data for all the *exo*-metabolome samples was subjected to the 'profile' mode MS peak peaking algorithm ('matched filter', see Appendix for R script) for the Bioconductor R package XCMS[27], to obtain a list of detected features and their peak areas (which is indicative of their abundances). This untargeted aspect of this study focused on ~7,800 features form the ESI negative ionization mode data whose XCMS-detected peak areas were more than 10,000,000 in at least one sample. From each of these features in worm samples, the average values of the same features in all the methanol blank runs were subtracted. To eliminate possible errors arising from differences in sample preparation or biomass, the data was then normalized by dividing the XCMS-detected intensity for each feature by the total intensity of all features in that sample (proportional to the total ion chromatogram area), and that value was multiplied by the average of total intensities across all samples. Features where more than 80 strains had intensity less than 100,000 were filtered out. These filtering criteria reduced chances of inadvertently mapping features that eventually turn out to be false positives from the XCMS output.

For the remaining features (~7,400), the normalized and rescaled intensity values were log

transformed and mapped to the *C. elegans* genome using the cegwas package in R[28]. All features that mapped with high significance i.e., exceeded the Bonferroni threshold to correct for multiple hypotheses, as well as had over 20% variance explained, were used for further analysis. Manhattan plots and phenotypic splits were generated for each of these features, and were manually checked to make sure the splits were not outlier driven. Of the total features in the ESI negative ionization mode data, about 150 mapped with high significance to various loci on the genome (Appendix Table E4). The author focused on features which mapped with high significance to common loci, a few such examples are discussed in the next few sections. The tentative structures for these features were assigned based on MS/MS fragmentation (see Appendix Figures).

Indole derivatives

Several features mapped to a common locus on Chromosome II between 10 and 15 Mb. MS/MS fragmentation in ESI negative ionization mode revealed that all these features have a common fragment at m/z 116.05057 ($C_8H_6N^-$), which suggests that the compounds contain an indole moiety. Thus, these metabolites were tentatively identified as indole derivatives, possibly resulting from the oxidative breakdown of the side chain in tryptophan (Figure 5.4).

Indole and some related compounds are known products of tryptophan metabolism in several strains of *E. coli*, including the laboratory *C. elegans* diet OP50. Several indole derivatives can regulate egg-laying behavior and survival in *C. elegans*[29]. A combination of *E. coli*-produced indole, indole-3-carboxaldehyde and indole-3-acetic acid is toxic to the nematodes[30]. On the other hand, glycosylation of the *E. coli* produced indole has been shown to be a chemical detoxification mechanism in *C. elegans* for microbial toxins[31]. Additionally, *C. elegans* itself can metabolize tryptophan to indole-3-carboxylic acid and, eventually, the icas family of ascarosides[32], even in the absence of bacteria.

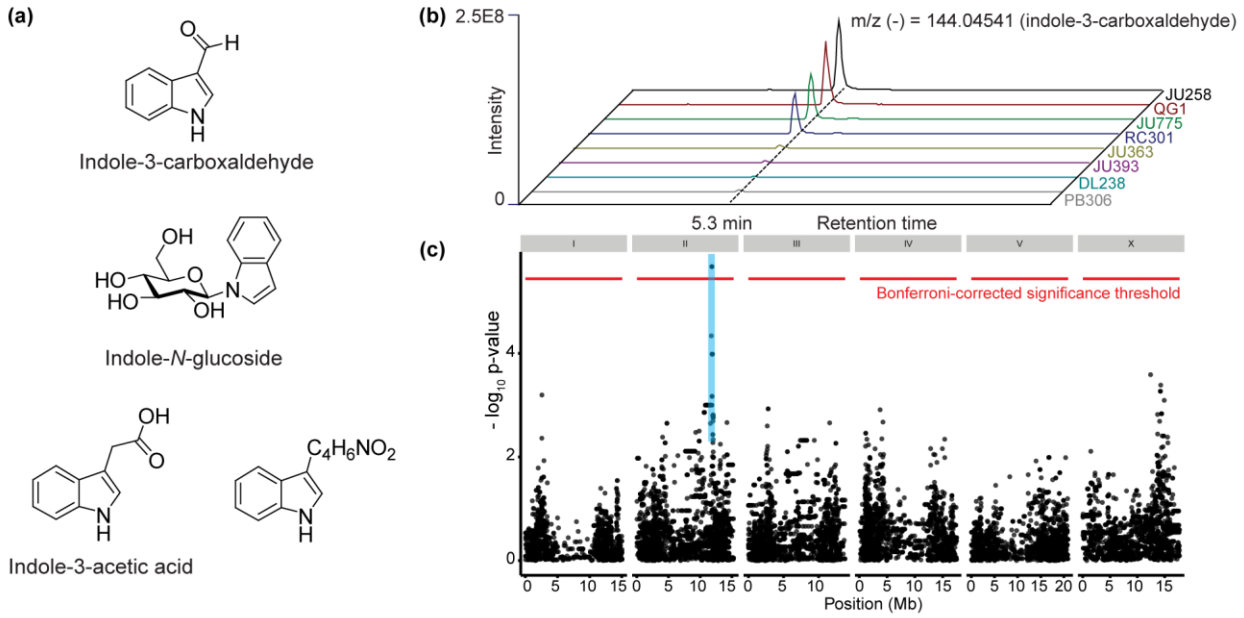

Figure 5.4. mQTL associated with indole metabolism. (a) Indole metabolites that map to a common locus. **(b)** Extracted ion chromatograms showing variation in abundance of an indole metabolite in natural isolates. **(c)** Manhattan plot showing the significance of association on the y-axis as the -$\log_{10}$(p-value) and the position of each variant tested on the x-axis in Mb. The blue box denotes the genomic region where a causal variant for altered tryptophan metabolism likely resides.

Dr. Mostafa Zamanian (Andersen research group, Northwestern University) initially found that the mQTL (**m**etabolite **Q**uantitative **T**rait **L**ocus) for the indole derivatives contains 19 *C. elegans* genes, only a few of which have been functionally annotated (Appendix Table E5). Of the genes that have predicted functions based on homology, none appear to be obviously implicated in chemical detoxification, tryptophan catabolism or icas biosynthetic pathways.

It should be pointed out that these mappings were done with only replicate of each natural isolate strain. Assaying more wild isolates and analyzing more replicates of every strain will yield more features that map with higher significance, as well as improve the signal-to-noise ratio for mappings. Such efforts are currently underway in the Schroeder research group (Cornell University). From these studies, if several other indole derivatives map to this particular mQTL, one will gain a better understanding of exactly which biochemical pathways are implicated. The next step would be to carry out fine mappings for the indole metabolites to come up with candidate genes for those pathways. In addition to single nucleotide variants, one can also look for common

indels and large deletions in genes in the 'low indole derivatives' strains compared to the 'high indole derivatives' strains (or vice versa) to arrive at candidate genes, following the approach adopted for the icas-deficient strains. Once a promising list of candidate genes is obtained, allele swaps or knockout mutants will be made to validate the functions of the genes.

Amino acid derivatives

A group of features corresponding to phenylalanine derivatives mapped to a common locus on Chromosome IV between 0 and 5 Mb. Based on their MS/MS fragmentation, these features were found to be *N*-acetylphenylalanine, *N*-succinylphenylethylamine and *N*-succinyltyramine (Figure 5.3). These compounds are structurally very similar to *N*-succinyloctopamine, which is the biosynthetic precursor to the osas family of ascarosides (Figure 5.5, also see Chapter 3). It would be interesting to see if *N*-succinyloctopamine, as well as the osas family of ascarosides, which are abundant in L1 worm pellets and media, map to a similar position on the genome, as it could indicate the presence of candidate biosynthetic genes for osas ascarosides. These derivatives of aromatic amino acids are known biosynthetic precursors of neurotransmitters or excretory metabolites even in higher organisms such as mammals[33-34]. This mQTL contains 11 *C. elegans* genes, only a handful of which have been functionally annotated (Appendix Table E6), and do not point to an obvious biosynthetic pathway.

Current efforts in the Schroeder research group are focused on growing all 114 wild isolate strains once again, this time by synchronizing the worms so that they are all at the same life stage (gravid adults) at the time of harvest. Once harvested, the worm media will be processed as a second biological replicate for the GWA mappings. The worm pellet will be bleached to dissolve the adult bodies, and the resulting eggs allowed to hatch in M9 buffer for 24 hours. The arrested (starved) L1s and their media will then be processed and used in GWAS. Such an experimental set

up will yield data for variation in osas ascaroside abundance, among others, across the wild isolate panel. As described in the preceding sections, lists of candidate genes will then be generated from significant mQTLs, and genome editing techniques used to validate their potential biosynthetic functions.

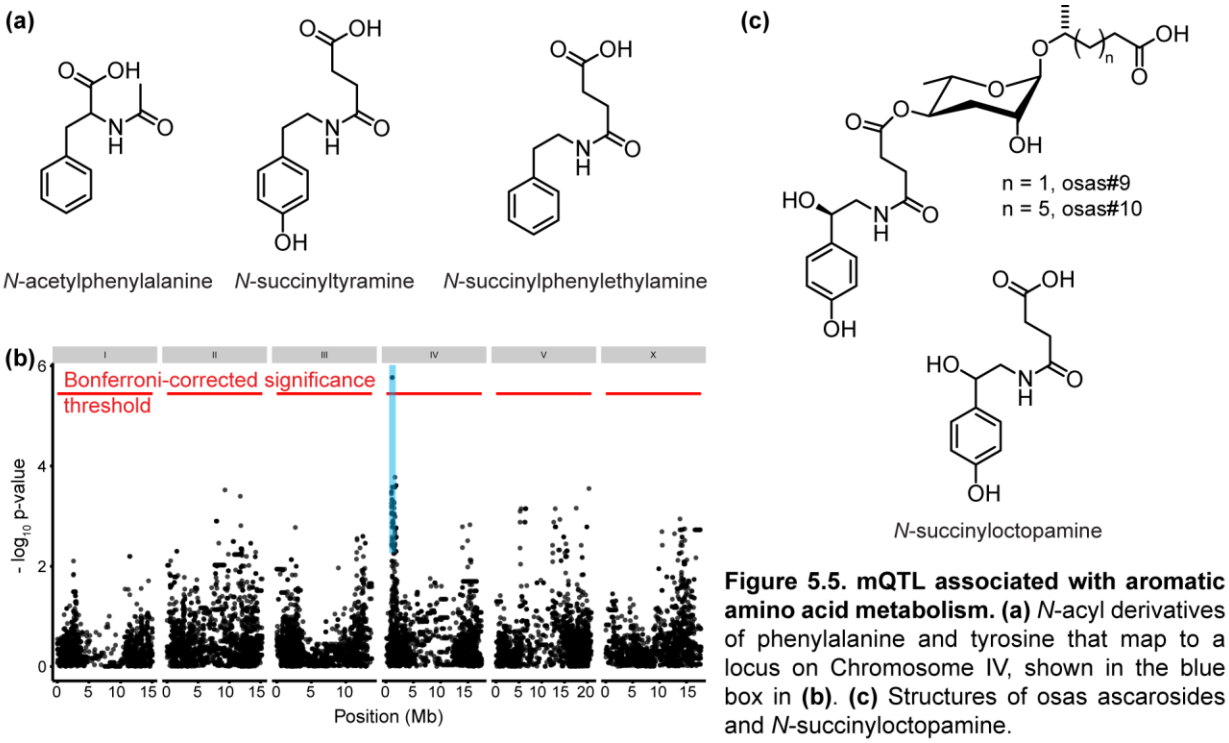

Figure 5.5. mQTL associated with aromatic amino acid metabolism. (a) *N*-acyl derivatives of phenylalanine and tyrosine that map to a locus on Chromosome IV, shown in the blue box in **(b)**. **(c)** Structures of osas ascarosides and *N*-succinyloctopamine.

Novel metabolites in C. elegans

For most features in the metabolome that are detected by the UHPLC-HRMS, there is no prior knowledge available regarding (a) what metabolite structures they correspond to, or (b) how they are synthesized by the worm, or (c) what biological functions they perform. From the high resolution m/z values for features obtained from any untargeted metabolomics study, one can derive molecular formulae. MS/MS fragmentation patterns can point to plausible building blocks for these compounds, and help develop hypotheses regarding the implicated biosynthetic pathways. Using the untargeted comparative metabolomics combined with GWAS approach

described in this chapter, one can map such features (about which very little is known), onto a genome (which is much better characterized than the metabolome). Thus, this approach offers a potential shortcut to the structural, biosynthetic and functional characterization of a part of the (largely uncharacterized) nematode metabolome. A few examples of metabolites whose structures and (eventually) candidate gene lists point to some known pathways have been described in the preceding sections. However, most features that map with high significance to different mQTLs represent entirely unknown compounds. For example, the author found a feature of m/z (-) 466.16891 that maps well to an mQTL to the left of Chromosome V. This feature corresponds to a chemical formula of $C_{18}H_{32}O_7N_3S_2^-$.

MS/MS fragmentation analysis (Appendix Figure E5) revealed a compound in which an *N*-propionylhomocysteine is conjugated to pantetheine via a disulfide bond (Figure 5.6). To the best of the author's knowledge, no such natural product has been previously described in the literature. The non-propionylated (free amine on homocysteine) counterpart of this compound was not detected, neither were Coenzyme A or 4′-phophospantetheinyl (P-pant) derivatives. A possible biosynthetic route can be proposed for this metabolite as outlined in Figure 5.4c. A SAM-dependent methyltransferase could convert S-adenosylmethionine (SAM) to S-adenosylhomocysteine (SAH). The adenosine moiety could be hydrolyzed and the resulting homocysteine propionylated at the amino group. This could then be conjugated to a pantetheine molecule. Genes containing single nucleotide variants in this mQTL are listed in Appendix Table E7. These could be candidates for any of the biosynthetic genes in the proposed pathway, or could point to the biological function of this metabolite. Extensive feature lists from more wild isolates, more biological replicates, life stage-specific worm media etc. could show other compounds structurally similar to pantetheinyl-*N*-propionylhomocysteine, and their common mQTLs (if any)

will be useful in the biosynthetic and functional characterization of this new class of metabolites.

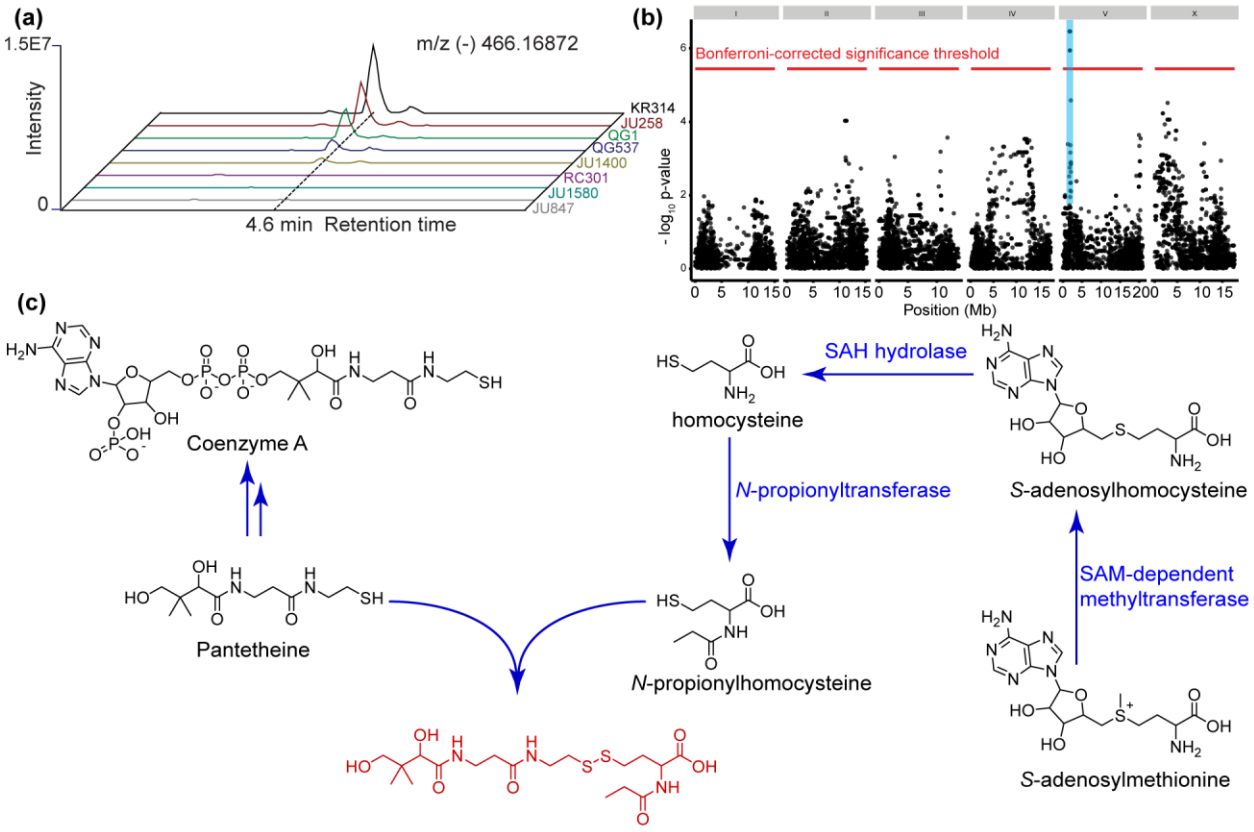

Figure 5.6. Novel compounds in the metabolome. (a) Variation in abundances of m/z (-) 466.16872, blue box in **(b)** shows the corresponding mQTL. **(c)** Proposed biosynthetic pathway for this compound, pantetheinyl-*N*-propionylhomocysteine (highlighted in red). Possible enzymes are shown in blue.

Validating genes in mQTLs

The mQTLs discovered for various metabolite families described above possibly control differences in biosynthesis, signal transduction or other aspects of metabolism of these compounds across the *C. elegans* natural isolate panel. These genomic intervals of interest may span a few genes to several hundred genes contained in over 1 Mb. As such, these intervals need to be 'fine-mapped' (*i.e.*, mapped with high resolution) to define the exact gene and variant that correlates with each specific metabolite family. Examples of fine mapping are shown for the QTLs for oscr#9:ascr#1, oscr#9:ascr#2 (Appendix Figure E1), and ascr#10:ascr#3 ratios (Appendix Figure E2). For unannotated candidate genes, homology searches will help in determining putative

functions and evaluating the likelihood of that gene being the cause behind the observed phenotype. Such analyses are important for prioritizing candidates for the gene validation step.

QTLs are typically narrowed down by generating recombinant strains targeting the interval[7-8, 35-39]. Independent near-isogenic lines (NILs), which contain the QTL genomic interval from one strain in the background of another strain, and tile across the QTL interval, will be generated. For example, the QTL intervals of a strain with abundant indole derivative production such as JU258 can be crossed into the background of a strain like PB306, which produces significantly reduced amounts of these compounds (Figure 5.4b), and vice versa. Then, a change in the indole derivative production phenotype would validate the mQTL and narrow the genomic interval harboring the relevant variant. This method is particularly useful for genomic intervals for which good sequencing data are not available, as it helps narrow down the intervals that require resequencing. The discovery of *Ppa-uar-1* highlights the success of such an approach.

Once mQTLs have been narrowed to ten or fewer candidate genes, genome-editing techniques can be used to either perform allele replacements or create putative loss-of-function alleles. For example, CRISPR/Cas9 technology was used to identify and validate the causal gene in ubas ascaroside production in *P. pacificus*. In some cases, viable deletion mutants for some of the candidate genes mentioned in the Appendix E Tables are readily available from the Caenorhabditis Genetics Center or the National BioResource Project, *e.g.* RB865 *(ok703)* for *basl-1*. These deletion mutants can be readily tested to see if the low abundance metabotype can be reproduced, which would quickly validate the gene for a metabolite production trait.

It should be noted, however, that testing deletion alleles of these genes in an N2 background (as are available from CGC and NBRP) assumes that N2 normally has the functioning allele. It is quite possible to not see a difference in metabolite abundance in the N2 deletion strain, even though

the candidate in question is indeed the gene naturally causing the metabotype difference. The initial natural isolate strain lacking the metabolite could, for example, have a gain-of-function allele in an inhibitory pathway, and loss of function of the N2 version of this enzyme (in the N2 deletion strain) might not change the metabolite level, depending on the dose effect. Dominance relationships and wildtype activity can, in principle, be completely switched in different natural isolates compared to genetics in the lab domesticated N2 strain. Thus, CRISPR/Cas9 based allele swaps and deletions in both high and low metabolite backgrounds, although time-consuming, are the best way to establish causality.

Conclusion

C. elegans has been shown to produce diverse multi-modular metabolites, derived from highly specific assembly of different building blocks. For the identification of genes involved in modular metabolite assembly, traditional mutant screens of candidate genes, selected based on functional predictions, have been of limited utility. The results discussed in this chapter highlight the potential of a natural variation-based unbiased approach for the study of novel biosynthetic pathways without any direct precedent. The *C. elegans* Million Mutation Project (MMP)[40], which has a collection of over 2,000 mutagenized, sequenced strains bearing indels and single nucleotide variants (several of which give rise to nonsense mutations) in over 20,000 genes, can be an excellent additional resource for such a study, to improve significance of mappings for various metabolites. This approach, however, would require greater genomic sequencing depth for the MMP strains than is currently available, to reduce the number of false positives in determining candidates for causal genes. One would also need to ascertain whether certain metabotypes in the MMP strains are actually due to non-functional biosynthetic genes, or overall poor health of the animal bearing too many mutations. This can be done by creating deletion mutants for candidate

genes from an MMP strain with a specific metabotype in an N2 background. Considering the wealth of natural variants isolated for *P. pacificus*, *C. elegans*[6, 14], *Drosophila*[41], and other animal models, this high resolution UHPLC-MS/MS, GWAS, and allele replacement using a CRISPR-Cas9-based workflow could serve as a blueprint for elucidating the biosynthetic pathways of the vast numbers of yet uncharacterized metabolites in these species.

REFERENCES

1. Viant, M. R.; Kurland, I. J.; Jones, M. R.; Dunn, W. B., How close are we to complete annotation of metabolomes? *Curr Opin Chem Biol* **2017,** *36*, 64-69.

2. Zampieri, M.; Sekar, K.; Zamboni, N.; Sauer, U., Frontiers of high-throughput metabolomics. *Curr Opin Chem Biol* **2017,** *36*, 15-23.

3. Schrimpe-Rutledge, A. C.; Codreanu, S. G.; Sherrod, S. D.; McLean, J. A., Untargeted Metabolomics Strategies-Challenges and Emerging Directions. *J Am Soc Mass Spectrom* **2016,** *27* (12), 1897-1905.

4. Niehaus, T. D.; Thamm, A. M.; de Crecy-Lagard, V.; Hanson, A. D., Proteins of Unknown Biochemical Function: A Persistent Problem and a Roadmap to Help Overcome It. *Plant Physiol* **2015,** *169* (3), 1436-42.

5. Bush, W. S.; Moore, J. H., Chapter 11: Genome-wide association studies. *PLoS computational biology* **2012,** *8* (12), e1002822.

6. Andersen, E. C.; Gerke, J. P.; Shapiro, J. A.; Crissman, J. R.; Ghosh, R.; Bloom, J. S.; Felix, M.-A.; Kruglyak, L., Chromosome-scale selective sweeps shape *Caenorhabditis elegans* genomic diversity. *Nat Genet* **2012,** *44* (3), 285-290.

7. McGrath, P. T.; Rockman, M. V.; Zimmer, M.; Jang, H.; Macosko, E. Z.; Kruglyak, L.; Bargmann, C. I., Quantitative mapping of a digenic behavioral trait implicates globin variation in *C. elegans* sensory behaviors. *Neuron* **2009,** *61* (5), 692-9.

8. Ghosh, R.; Andersen, E. C.; Shapiro, J. A.; Gerke, J. P.; Kruglyak, L., Natural variation in a chloride channel subunit confers avermectin resistance in *C. elegans*. *Science* **2012,** *335* (6068), 574-8.

9. Rhee, E. P.; Ho, J. E.; Chen, M. H.; Shen, D.; Cheng, S.; Larson, M. G.; Ghorbani, A.; Shi, X.; Helenius, I. T.; O'Donnell, C. J.; Souza, A. L.; Deik, A.; Pierce, K. A.; Bullock, K.; Walford, G. A.; Vasan, R. S.; Florez, J. C.; Clish, C.; Yeh, J. R.; Wang, T. J.; Gerszten, R. E., A genome-wide association study of the human metabolome in a community-based cohort. *Cell Metab* **2013,** *18* (1), 130-43.

10. Wu, S.; Alseekh, S.; Cuadros-Inostroza, A.; Fusari, C. M.; Mutwil, M.; Kooke, R.; Keurentjes, J. B.; Fernie, A. R.; Willmitzer, L.; Brotman, Y., Combined Use of Genome-Wide Association

Data and Correlation Networks Unravels Key Regulators of Primary Metabolism in *Arabidopsis thaliana*. *PLoS Genet* **2016,** *12* (10), e1006363.

11. Richter, A.; Schaff, C.; Zhang, Z.; Lipka, A. E.; Tian, F.; Kollner, T. G.; Schnee, C.; Preiss, S.; Irmisch, S.; Jander, G.; Boland, W.; Gershenzon, J.; Buckler, E. S.; Degenhardt, J., Characterization of Biosynthetic Pathways for the Production of the Volatile Homoterpenes DMNT and TMTT in *Zea mays*. *Plant Cell* **2016,** *28* (10), 2651-2665.

12. Wen, W.; Li, D.; Li, X.; Gao, Y.; Li, W.; Li, H.; Liu, J.; Liu, H.; Chen, W.; Luo, J.; Yan, J., Metabolome-based genome-wide association study of maize kernel leads to novel biochemical insights. *Nat Commun* **2014,** *5*, 3438.

13. Cook, D. E.; Zdraljevic, S.; Tanny, R. E.; Seo, B.; Riccardi, D. D.; Noble, L. M.; Rockman, M. V.; Alkema, M. J.; Braendle, C.; Kammenga, J. E.; Wang, J.; Kruglyak, L.; Felix, M. A.; Lee, J.; Andersen, E. C., The genetic basis of natural variation in *Caenorhabditis elegans* telomere length. *Genetics* **2016,** *204* (1), 371-83.

14. Cook, D. E.; Zdraljevic, S.; Roberts, J. P.; Andersen, E. C., CeNDR, the *Caenorhabditis elegans* natural diversity resource. *Nucleic Acids Res* **2017,** *45* (D1), D650-D657.

15. von Reuss, S. H.; Bose, N.; Srinivasan, J.; Yim, J. J.; Judkins, J. C.; Sternberg, P. W.; Schroeder, F. C., Comparative metabolomics reveals biogenesis of ascarosides, a modular library of small molecule signals in *C. elegans*. *J Am Chem Soc* **2012,** *134* (3), 1817–1824.

16. Zhang, X.; Feng, L.; Chinta, S.; Singh, P.; Wang, Y.; Nunnery, J. K.; Butcher, R. A., Acyl-CoA oxidase complexes control the chemical message produced by *Caenorhabditis elegans*. *Proceedings of the National Academy of Sciences* **2015,** *112* (13), 3955-3960.

17. Zhang, X.; Li, K.; Jones, R. A.; Bruner, S. D.; Butcher, R. A., Structural characterization of acyl-CoA oxidases reveals a direct link between pheromone biosynthesis and metabolic state in *Caenorhabditis elegans*. *Proc Natl Acad Sci U S A* **2016,** *113* (36), 10055-60.

18. Poirier, Y.; Antonenkov, V. D.; Glumoff, T.; Hiltunen, J. K., Peroxisomal β-oxidation—A metabolic pathway with multiple functions. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **2006,** *1763* (12), 1413-1426.

19. Ludewig, A. H.; Izrayelit, Y.; Park, D.; Malik, R. U.; Zimmermann, A.; Mahanti, P.; Fox, B. W.; Bethke, A.; Doering, F.; Riddle, D. L.; Schroeder, F. C., Pheromone sensing regulates *Caenorhabditis elegans* lifespan and stress resistance via the deacetylase SIR-2.1. *Proc Natl Acad Sci U S A* **2013,** *110* (14), 5522-7.

20. Izrayelit, Y.; Srinivasan, J.; Campbell, S. L.; Jo, Y.; von Reuss, S. H.; Genoff, M. C.; Sternberg, P. W.; Schroeder, F. C., Targeted metabolomics reveals a male pheromone and sex-specific ascaroside biosynthesis in *Caenorhabditis elegans*. *ACS chemical biology* **2012,** *7* (8), 1321-5.

21. Maures, T. J.; Booth, L. N.; Benayoun, B. A.; Izrayelit, Y.; Schroeder, F. C.; Brunet, A., Males shorten the life span of *C. elegans* hermaphrodites via secreted compounds. *Science* **2014,** *343* (6170), 541-4.

22. Aprison, E. Z.; Ruvinsky, I., Sexually Antagonistic Male Signals Manipulate Germline and Soma of *C. elegans* Hermaphrodites. *Current biology : CB* **2016,** *26* (20), 2827-2833.

23. Aprison, E. Z.; Ruvinsky, I., Counteracting Ascarosides Act through Distinct Neurons to Determine the Sexual Identity of *C. elegans* Pheromones. *Current biology : CB* **2017,** *27* (17), 2589-2599 e3.

24. Zarkower, D., Somatic sex determination. *WormBook* **2006**, 1-12.

25. Huang, X.; Zhang, H.; Zhang, H., The zinc-finger protein SEA-2 regulates larval developmental timing and adult lifespan in *C. elegans*. *Development (Cambridge, England)* **2011,** *138* (10), 2059-68.

26. Panda, O.; Akagi, A. E.; Artyukhin, A. B.; Judkins, J. C.; Le, H. H.; Mahanti, P.; Cohen, S. M.; Sternberg, P. W.; Schroeder, F. C., Biosynthesis of Modular Ascarosides in *C. elegans*. *Angew Chem Int Ed Engl* **2017,** *56* (17), 4729-4733.

27. Smith, C. A.; Want, E. J.; O'Maille, G.; Abagyan, R.; Siuzdak, G., XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem* **2006,** *78* (3), 779-87.

28. Evans, K. S.; Zhao, Y.; Brady, S. C.; Long, L.; McGrath, P. T.; Andersen, E. C., Correlations of Genotype with Climate Parameters Suggest *Caenorhabditis elegans* Niche Adaptations. *G3 (Bethesda)* **2017,** *7* (1), 289-298.

29. Lee, J. H.; Kim, Y. G.; Kim, M.; Kim, E.; Choi, H.; Kim, Y.; Lee, J., Indole-associated predator-prey interactions between the nematode *Caenorhabditis elegans* and bacteria. *Environ Microbiol* **2017,** *19* (5), 1776-1790.

30. Bommarius, B.; Anyanful, A.; Izrayelit, Y.; Bhatt, S.; Cartwright, E.; Wang, W.; Swimm, A. I.; Benian, G. M.; Schroeder, F. C.; Kalman, D., A family of indoles regulate virulence and Shiga toxin production in pathogenic *E. coli*. *PLoS One* **2013,** *8* (1), e54456.

31. Stupp, G. S.; von Reuss, S. H.; Izrayelit, Y.; Ajredini, R.; Schroeder, F. C.; Edison, A. S., Chemical detoxification of small molecules by *Caenorhabditis elegans*. *ACS chemical biology* **2013,** *8* (2), 309-13.

32. Srinivasan, J.; von Reuss, S. H.; Bose, N.; Zaslaver, A.; Mahanti, P.; Ho, M. C.; O'Doherty, O. G.; Edison, A. S.; Sternberg, P. W.; Schroeder, F. C., A Modular Library of Small Molecule Signals Regulates Social Behaviors in *Caenorhabditis elegans*. *PLoS Biol* **2012,** *10* (1), e1001237.

33. Jansen, R. S.; Addie, R.; Merkx, R.; Fish, A.; Mahakena, S.; Bleijerveld, O. B.; Altelaar, M.; L, I. J.; Wanders, R. J.; Borst, P.; van de Wetering, K., N-lactoyl-amino acids are ubiquitous metabolites that originate from CNDP2-mediated reverse proteolysis of lactate and amino acids. *Proc Natl Acad Sci U S A* **2015,** *112* (21), 6601-6.

34. Okajima, K.; Inoue, M.; Morino, Y., Studies on the mechanism for renal elimination of N-acetylphenylalanine: its pathophysiologic significance in phenylketonuria. *J Lab Clin Med* **1985,** *105* (1), 132-8.

35. Large, E. E.; Xu, W.; Zhao, Y.; Brady, S. C.; Long, L.; Butcher, R. A.; Andersen, E. C.; McGrath, P. T., Selection on a Subunit of the NURF Chromatin Remodeler Modifies Life History Traits in a Domesticated Strain of *Caenorhabditis elegans*. *PLoS Genet* **2016,** *12* (7), e1006219.

36. Glater, E. E.; Rockman, M. V.; Bargmann, C. I., Multigenic natural variation underlies *Caenorhabditis elegans* olfactory preference for the bacterial pathogen *Serratia marcescens*. *G3 (Bethesda)* **2014,** *4* (2), 265-76.

37. Bendesky, A.; Tsunozaki, M.; Rockman, M. V.; Kruglyak, L.; Bargmann, C. I., Catecholamine receptor polymorphisms affect decision-making in *C. elegans*. *Nature* **2011,** *472* (7343), 313-8.

38. Bendesky, A.; Pitts, J.; Rockman, M. V.; Chen, W. C.; Tan, M. W.; Kruglyak, L.; Bargmann, C. I., Long-range regulatory polymorphisms affecting a GABA receptor constitute a quantitative trait locus (QTL) for social behavior in *Caenorhabditis elegans*. *PLoS Genet* **2012,** *8* (12), e1003157.

39. Greene, J. S.; Brown, M.; Dobosiewicz, M.; Ishida, I. G.; Macosko, E. Z.; Zhang, X.; Butcher, R. A.; Cline, D. J.; McGrath, P. T.; Bargmann, C. I., Balancing selection shapes density-dependent foraging behaviour. *Nature* **2016,** *539* (7628), 254-258.

40. Thompson, O.; Edgley, M.; Strasbourger, P.; Flibotte, S.; Ewing, B.; Adair, R.; Au, V.; Chaudhry, I.; Fernando, L.; Hutter, H.; Kieffer, A.; Lau, J.; Lee, N.; Miller, A.; Raymant, G.; Shen, B.; Shendure, J.; Taylor, J.; Turner, E. H.; Hillier, L. W.; Moerman, D. G.; Waterston, R. H., The million mutation project: a new approach to genetics in *Caenorhabditis elegans*. *Genome Res* **2013,** *23* (10), 1749-62.

41. Lack, J. B.; Lange, J. D.; Tang, A. D.; Corbett-Detig, R. B.; Pool, J. E., A Thousand Fly Genomes: An Expanded *Drosophila* Genome Nexus. *Mol Biol Evol* **2016,** *33* (12), 3308-3313.

CONCLUSIONS AND OUTLOOK

In this dissertation, the author has demonstrated the use of traditional activity-guided fractionation as well as modern comparative metabolomics techniques to elucidate the structures and biosynthetic pathways of various signaling molecules in the pathogenic bacterium *P. aeruginosa*, the free-living model nematode *C. elegans* and the necromenic satellite model organism *P. pacificu*s. In Chapter 1, the author has discussed the pros and cons of using iterative chromatographic separation of complex metabolome mixtures to elucidate pathogen-produced chemical cues that are perceived by host nematodes, which causes them to avoid the pathogen and thereby avert infection[1]. In Chapter 2, the author demonstrates the evolutionary divergence of biosynthetic pathways in *P. pacificus*, for modular metabolites similar in structure and function to those produced by *C. elegans*[2]. Chapter 3 demonstrates the utility of candidate-based approaches in the identification of biosynthetic enzymes for nematode-derived modular metabolites, their modes of action as well as sub-cellular localization[3]. The author expands the comparative metabolomics technique to study the metabolome of nematodes in an unbiased manner in Chapter 4, using an interesting lifespan phenotype to guide the quest for novel metabolites that may be involved in aging and age-related diseases. In Chapter 5, the author combines untargeted comparative metabolomics with natural genetic variation in *C. elegans* via a genome-wide association study, to arrive at candidate biosynthetic genes for nematode-derived modular metabolites.

In Chapter 3, the author suggested the use of this untargeted metabolomics technique at the sub-cellular level, to study the metabolomes of various organelles. This is particularly important, given the origin of several signaling molecules and the localization of at least one biosynthetic enzyme in the acidic lysosome-related organelles in the *C. elegans* intestine. The author assessed

the utility of this technique to study the mitochondrial metabolome in Chapter 4. Extension of this method to the metabolomics of acidic gut granules could reveal new signaling molecules, intermediate molecules in their biosynthetic pathways as well as their biosynthetic genes. A thorough understanding of mechanisms for the biosynthesis of signaling molecules in nematodes, combined with mechanistic knowledge of perception of pathogenic cues by host organisms, could enable efficient anthelmintic drug design to mitigate various diseases affecting both crop plants and humans worldwide.

The newly discovered metabolites described in Chapter 4, that correlate with aging in nematodes, could eventually lead to the identification of biomarkers for aging and age-related diseases in higher organisms. Given their structural similarity to highly conserved molecules in mammals, they may be used to understand regulatory mechanisms of ageing and disease pathways in humans. As highlighted in Chapter 5, correlating the variation in the production of nematode-derived metabolites to different genomic loci can be used to characterize pathways for their biosynthesis.

For any perturbation introduced to a given biological system, there is a resulting molecular signature. High-resolution UHLPC-MS/MS combined with bioinformatics approaches in the untargeted comparative metabolomics pipeline described in this dissertation can be very useful in studying such perturbations. As demonstrated by the author in this dissertation, this approach can be used to study various organisms, from bacteria to humans, to understand the roles of biogenic signaling molecules at the sub-cellular, cellular, tissue-specific and organismal levels.

REFERENCES

1. Meisel, J. D.; Panda, O.; Mahanti, P.; Schroeder, F. C.; Kim, D. H., Chemosensation of bacterial secondary metabolites modulates neuroendocrine signaling and behavior of *C. elegans*. *Cell* **2014,** *159* (2), 267-80.

2. Markov, G. V.; Meyer, J. M.; Panda, O.; Artyukhin, A. B.; Claassen, M.; Witte, H.; Schroeder, F. C.; Sommer, R. J., Functional Conservation and Divergence of daf-22 Paralogs in *Pristionchus pacificus* Dauer Development. *Mol Biol Evol* **2016,** *33* (10), 2506-14.

3. Panda, O.; Akagi, A. E.; Artyukhin, A. B.; Judkins, J. C.; Le, H. H.; Mahanti, P.; Cohen, S. M.; Sternberg, P. W.; Schroeder, F. C., Biosynthesis of Modular Ascarosides in *C. elegans*. *Angew Chem Int Ed Engl* **2017,** *56* (17), 4729-4733.

APPENDIX A: SUPPLEMENTAL INFORMATION FOR CHAPTER 1

Materials and methods

Genetic Screening and Mapping. *daf-7p::gfp* animals were EMS mutagenized and 500 P0s were egg-laid to produce 32 independent pools each containing 400 F1 animals (approximately 25,000 haploid genomes). The F1 pools were egg-prepped to create synchronized populations of F2s, which were grown on *E. coli* OP50 until the L4 larval stage. Mutagenized F2s were then transferred onto 10 cm NGM plates seeded with *P. aeruginosa* PA14, incubated overnight at 25˚C, and screened the following day for a lack of GFP fluorescence in the ASJ neurons. After retesting and backcrossing, mutants were SNP mapped using a Hawaiian strain carrying the *daf-7p::gfp* transgene and sequenced using Illumina technology.

Protocol used for Fractionation of *P. aeruginosa* supernatant extract. *P. aeruginosa* supernatant extract (1.8 g) in dichloromethane:methanol (4:1, 100 mL) was added to Celite powder (10 g), prewashed with ethyl acetate. After rotary evaporation the Celite was dry-loaded into an empty 25 g RediSepRf loading cartridge. Fractionation was performed using a Teledyne ISCO CombiFlash system over a RediSep*Rf* GOLD 40 g HP Silica Column using a normal phase dichloromethane-methanol solvent system, starting with 100% dichloromethane and a linear increase in methanol content up to 10% at 7 column volumes, followed by another linear increase up to 30% at 11 column volumes, followed by a third linear increase in methanol content up to 100% at 12 column volumes, which was then continued up to 18.2 column volumes. 42 fractions (25 mL each) generated from the CombiFlash run were individually evaporated *in vacuo* and prepared for bio-assays and analyses by NMR spectroscopy (^{1}H NMR, DQF-COSY, HSQC and HMBC) and HPLC-MS.

Protocol for further fractionation via HPLC. Extract fractions of interest were evaporated *in*

vacuo, resuspended in 450 µL of methanol and submitted to HPLC analysis, using an Agilent 1100 Series HPLC system equipped with an Agilent Eclipse XDB C-18 column (25 cm x 9.4 cm, 5 µm particle diameter. A 0.1% acetic acid in water ("aqueous") – acetonitrile ("organic") solvent system was used, starting with 5% organic solvent for 2 min, which was increased linearly to 100% over a period of 28 min and continued at 100% organic solvent for 10 min. Fraction collection times were decided based on the UV trace of the chromatogram. Fractions were collected using a Teledyne ISCO Foxy 200 X-Y fraction collector from 2 to 29 min. Collected fractions were individually evaporated *in vacuo* for further analysis by NMR spectroscopy.

Supplemental figures

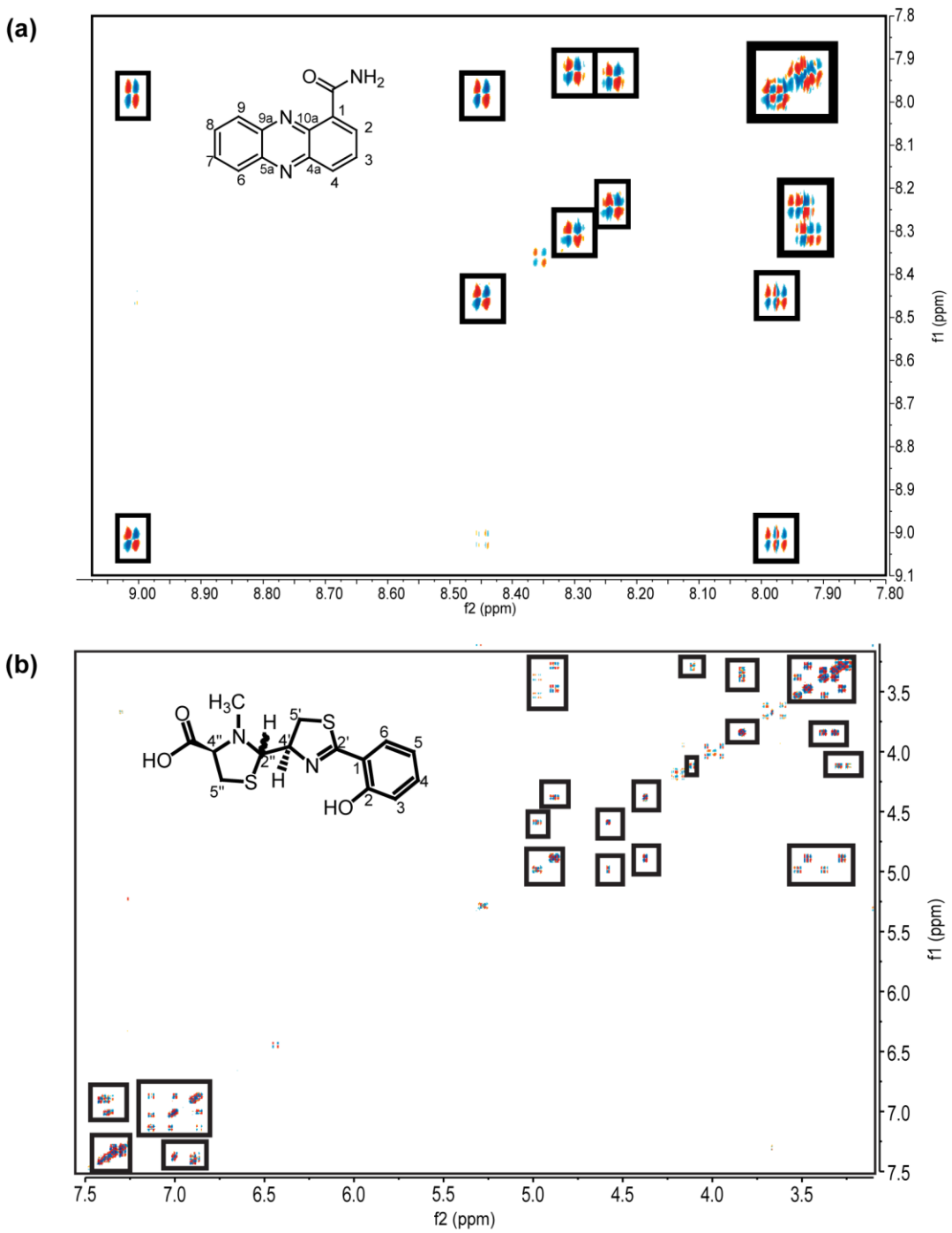

Figure A1. (a) Section of DQF-COSY spectrum (600 MHz, CDCl3) with boxed cross-peaks indicative of major component phenazine-1-carboxamide. **(b)** Section of DQF-COSY spectrum (600 MHz, CDCl3) with boxed crosspeaks indicative of major component pyochelin.

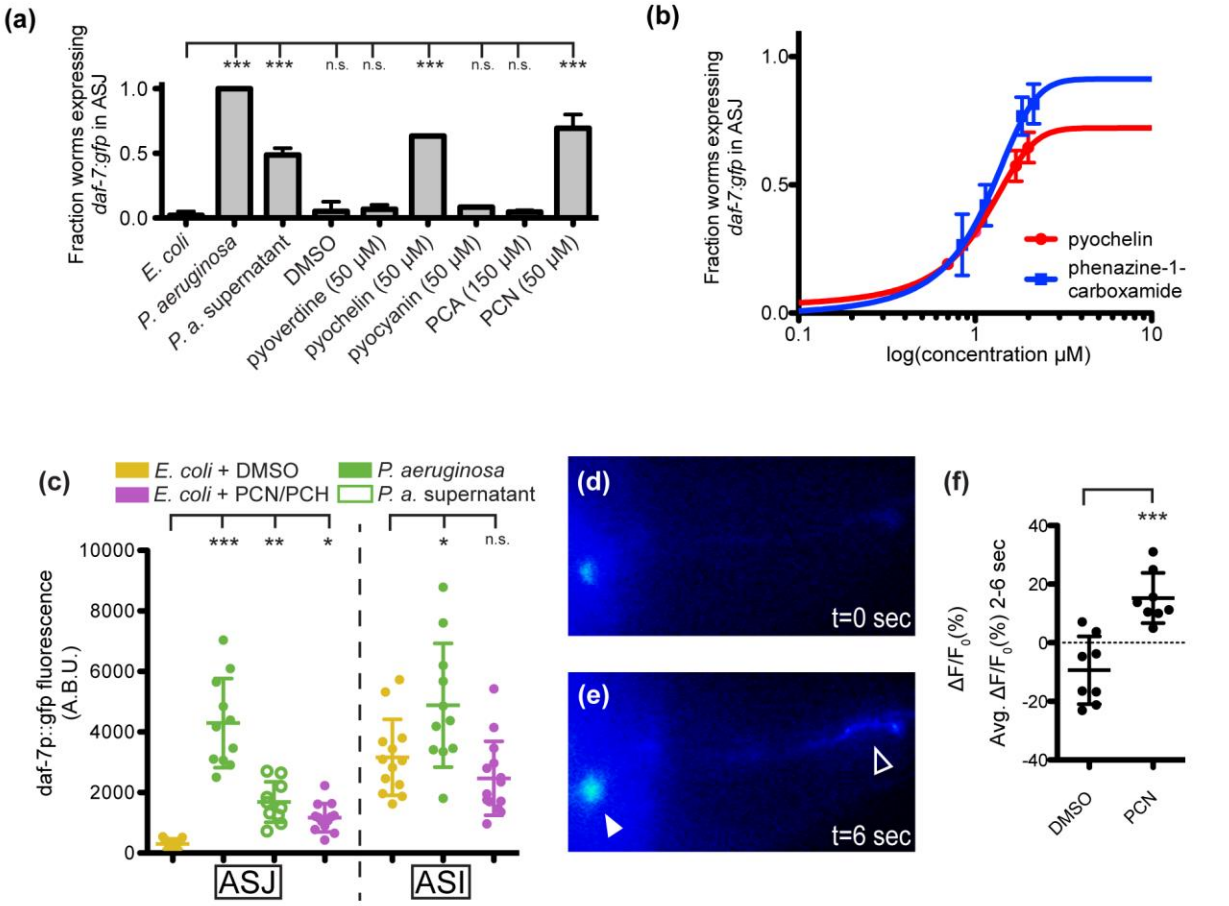

Figure A2. *daf-7p::gfp* expression in the ASJ or ASI neurons after 16 hr exposure to *E. coli*, *P. aeruginosa*, or *E. coli* supplemented with indicated material. Data represent **(a)** the fraction of animals expressing *daf-7p::gfp* in ASJ above background or **(b)** the maximum fluorescence values of *daf-7p::gfp* in ASJ and ASI. All compounds were dissolved in DMSO. ***p < 0.001, **p < 0.01, and *p < 0.05 as determined by one-way ANOVA, followed by Dunnett's multiple comparison test. ns, not significant. Error bars indicate standard deviation. **(c)** Fraction of animals expressing *daf-7::gfp* in the ASJ neurons after exposure to *E. coli* supplemented with either PCN or pyochelin. Data represent the average of three independent experiments and error bars indicate standard error. The data have been fit using a three-parameter dose-response curve and plotted on a semi-log scale. EC_{50} values for pyochelin and PCN are 11.7 mM and 13.4 mM, respectively, and maximum response values for pyochelin and PCN are 0.7 and 0.9, respectively. GCaMP5 expression in the ASJ neurons immediately prior to **(d)** or following **(e)** addition of PCN. Filled triangle indicates the ASJ cell body and open triangle indicates the ASJ sensory projection. **(f)** Average changes in GCaMP fluorescence in the 2–6 s following addition of either PCN or DMSO. ***p < 0.001 as determined by unpaired t-test. Error bars indicate standard deviation.

Supplemental tables

Table A1. NMR spectroscopic data of Phenazine-1-Carboxamide, related to Figure A1. ^{1}H (600 MHz), ^{13}C (151 MHz), and important HMBC NMR spectroscopic data for PCN in $CDCl_3$. Chemical shifts were referenced to ($C\underline{H}Cl_3$) = 7.26 ppm and ($\underline{C}DCl_3$) = 77.16 ppm. Multiplicities are indicated in parentheses.

Carbon No.	δ (ppm)	Proton No.	δ (ppm)	J (Hz)	HMBC correlations
-		NH_2	10.80 (bs)		-
Carbonyl	170.01	-	-		-
1	128.41	-	-		-
2	134.38	2	8.45 (d)	$J_{2,3}$ = 8.8	C-4, C-10a
3	129.88	3	7.98 (dd)		C-4, C-1
4	135.99	4	9.01 (d)	$J_{3,4}$ = 8.8	C-3, C-2, C-4a
4a	140.67	-	-		-
5a	141.43	-	-		-
6	129.60	6	8.30 (m)		C-5a, C-7
7	131.05	7	7.93 (m)		C-6, C-5a
8	129.02	8	7.94 (m)		C-9, C-9a
9	128.02	9	8.24 (d)	$J_{8,9}$ = 8.8	C-7, C-9a
9a	143.09	-	-		
10a	140.80	-	-		

Table A2. NMR spectroscopic data of Pyochelin, related to Figure A1. ^{1}H (600 MHz), ^{13}C (151 MHz), and important HMBC NMR spectroscopic data for pyochelin in $CDCl_3$. Chemical shifts were referenced to ($C\underline{H}Cl_3$) = 7.26 ppm and ($\underline{C}DCl_3$) = 77.16 ppm. Multiplicities are indicated in parentheses.

Carbon No.	δ (ppm)	Proton No.	δ (ppm)	J (Hz)	HMBC correlations
1	115.97	-	-		-
2	159.14	-	-		-
3	117.17	3	7.01 (m)		C-1, C-5, C-2
4	130.65	4	7.38 (m)		C-6
5	120.44	5	6.90 (m)		C-1, C-3, C-4, C-6
6	133.50	6	7.42 (d)	$J_{5,6}$ = 8.5	C-2
2’	173.75	-	-		-
4’	80.02	4’	4.89 (m)		C-5’, C-2’, C-2”
5’	33.19	5a’, 5b’	3.28, 3.47 (dd)		C-2’, C-4’, C-2”
2”	77.74	2”	4.38 (d)	$J_{2",4'}$ = 8.7	C-4’, C-5’
4”	73.51	4”	3.84 (t)		C-5”, C-2”, COOH
5”	33.24	5a”, 5b”	3.39, 3.32 (dd)		C-4”, C-2”, COOH
Methyl	43.62	Methyl	2.71 (s)		C-2”, C-4”
Carboxylic acid	172.53	-	-		-

Table A3. Complete list of *C. elegans* strains used in this study

STRAIN	GENOTYPE
N2	wild type
ZD715	*daf-7(ok3125)* 5x Backcrossed
CB1372	*daf-7(e1372)*
DR40	*daf-1(m40)*
CB1393	*daf-8(e1393)*
DR1572	*daf-2(e1368)*
DA650	*npr-1(215F)*
NL348	*gpa-2(pk16) gpa-3(pk35)*
ZD907	*daf-7(ok3125); daf-3(e1376)*
ZD889	*daf-7(ok3125); dgk-1(nu62)*
ZD890	*goa-1(n1134); daf-7(ok3125)*
ZD910	*tdc-1(n3420); daf-7(ok3125)*
ZD911	*daf-7(ok3125); tbh-1(n3722)*
ZD634	*gcy-35(ok769); daf-7(ok3125)*
KQ143	*daf-7(e1372); daf-12 (m20)*
ZD919	*daf-7(ok3125); npr-1(215F)*
ZD695	*daf-7(ok3125); qdEx34[trx-1p::daf-7 + ges-1p::gfp]*
ZD696	*daf-7(ok3125); qdEx35[trx-1p::daf-7 + ges-1p::gfp]*
ZD729	*daf-7(ok3125); qdEx37[daf-7p::daf-7 + ges-1p::gfp]*
ZD730	*daf-7(ok3125); qdEx38[daf-7p::daf-7 + ges-1p::gfp]*
ZD735	*daf-7(ok3125); qdEx43[str-3p::daf-7 + ges-1p:gfp]*
ZD736	*daf-7(ok3125); qdEx44[str-3p::daf-7 + ges-1p:gfp]*
ZD732	*daf-7(ok3125); qdEx40[trx-1p::daf-7 + ges-1p::gfp]*
ZD733	*daf-7(ok3125); qdEx41[trx-1p::daf-7 + ges-1p::gfp]*
KQ251	*daf-1(m40); ftEX69[pegl-3::daf-1::GFP; odr-1::dsRED]*
KQ252	*daf-1(m40); ftEX70[pbbs-1::daf-1::GFP; odr-1::dsRED]*
KQ265	*daf-1(m40); ftEX83[posm-6::daf-1::GFP; odr-1::dsRED]*
KQ280	*daf-1(m40); ftEX98[pdaf-1::daf-1::GFP; odr-1::dsRED]*
KQ380	*daf-1(m40); ftEX205[ptdc-1::daf-1-gfp; odr-1::dsRED]*
ZD762	*mgIs40[daf-28p::nls-GFP]; jxEx100[trx-1::ICE + ofm-1::gfp]*
ZD763	*mgIs40[daf-28p::nls-GFP]; jxEx102[trx-1::ICE + ofm-1::gfp]*
ZD818	*mgIs40; daf-3(e1376); jxEx100*
ZD819	*mgIs40; daf-3(e1376); jxEx102*
ZD820	*mgIs40; daf-3(ok3610); jxEx100*
ZD821	*mgIs40; daf-3(ok3610); jxEx102*
FK181	*ksIs2[daf-7p::GFP + rol6(su1006)]*
ZD682	*ksIs2; daf-11(m47)*
ZD714	*ksIs2; tax-4(p678)*
ZD726	*tax-2(p671) ksIs2*
ZD884	*ksIs2; gpa-3(qd262)*

ZD887	*ksIs2; gpa-2(pk16) gpa-3(pk35)*
ZD1083	*ksIs2; gpa-2(pk16)*
ZD1047	*ksIs2; gpa-3(pk35)*
ZD971	*ksIs2; gpa-2(pk16) gpa-3(pk35); qdEx64[bbs-1p::gpa-3 + ofm-1::gfp]*
ZD974	*ksIs2; gpa-2(pk16) gpa-3(pk35); qdEx67[trx-1p::gpa-3 + ofm-1::gfp]*
ZD1059	*ksIs2; gpa-3(qd262); qdEx88[bbs-1p::gpa-3 + ofm-1::gfp]*
ZD1061	*ksIs2; gpa-3(qd262); qdEx90[trx-1p::gpa-3 + ofm-1::gfp]*
ZD1111	*ksIs2; gpa-2(pk16) gpa-3(pk35); qdEx92[bbs-1p::gpa-2 + ofm-1::gfp]*
ZD1114	*ksIs2; gpa-2(pk16) gpa-3(pk35); qdEx95[trx-1p::gpa-2 + ofm-1::gfp]*
ZD1115	*ksIs2; gpa-2(pk16) gpa-3(pk35); qdEx96[ceh-36p::gpa-2 + ofm-1::gfp]*
OE3010	*lin-15B(n765); ofEx4[pBLH98: lin-15(+) + trx-1::GFP]*
CX5478	*lin-15B(n765); kyEx581[ocr-4::GFP + lin-15(+)]*
OH7547	*otIs199[cat-2::GFP + rgef-1::dsRed + rol-6(su1006)]*
ZD1184	*qdEx103[trx-1p::GCaMP5 + ofm-1::gfp]*

Table A4. Complete list of oligos used in this study

Oligos used for *daf-7* smFISH Probe	**Oligos used for *gpa-2* smFISH Probe**
5'-ttcatttctgggtccattgg-3'	5'-cgttgactttccacattctc-3'
5'-atacatctccaggtagactg-3'	5'-agcaatctcatctgcttgag-3'
5'-tcgtccttctccagtaagtc-3'	5'-atctgtgtactgttttgacg-3'
5'-catttcaacgcccatatcct-3'	5'-ttggcctgcgtcaaaagttc-3'
5'-gatctttggcggtgtagaat-3'	5'-aactatgtttgtgtatacca-3'
5'-gacgggttctctccataact-3'	5'-tgctttgactaaatgatcca-3'
5'-gtcaaacttggcaacaagct-3'	5'-aaattcaatccggcagctgg-3'
5'-gcttctttccaaatcatttg-3'	5'-atcatgttctctcatcggat-3'
5'-tgagtgtggcctgaagaata-3'	5'-aatgtagagggtcagcatgt-3'
5'-gcaggaatctcaattgatac-3'	5'-gttggaagtttttatgttgc-3'
5'-ttgaagcatccctgaatcct-3'	5'-tttctccacatgatctgcag-3'
5'-cgtaaacttgaacttgaaca-3'	5'-ttcaattctttttctctcggc-3'
5'-attgatccatcctcgttctt-3'	5'-aagtattccgtgttatcacc-3'
5'-tccagaagtgaccatctctc-3'	5'-atattcttggcaggttttca-3'
5'-cggatccctttgtagcgaaa-3'	5'-tcgcatttgggtgataatct-3'
5'-ggaagttgaatgctgatacg-3'	5'-gtccgcaaaagaagggtgtc-3'
5'-ccaacttttgacagtatcaa-3'	5'-tgaatattctccttcttgat-3'
5'-cttggatcggagaaattgtg-3'	5'-tcgatattccggaaacgcct-3'
5'-agcattgccttgacgaagat-3'	5'-tttcagcgtaattttgctct-3'
5'-atgaagggcaacgttgcgcc-3'	5'-ttcgtcttgataaatgcgac-3'
5'-aacgtcagcagtggtctgtt-3'	5'-gttattggaaagcgcctcga-3'
5'-agagctgaagacgcatgttg-3'	5'-gcacgtagaatgtcttcttc-3'
5'-cgggacccctttggacgagt-3'	5'-tgtgtctgttgcgcatgttt-3'
5'-tttggcatgagaacggcgtt-3'	5'-ccaaaatcttttgcacttga-3'
5'-tgtgcttcggcattgcaaac-3'	5'-agagtccagatttgtgtagg-3'

Oligos used for *daf-7* rescue constructs	
5'-ATGGAAGCTTCGGCAACTAA-3'	*daf-7p* forward
5'-GAGTGAAGATGCCATGAACATGGCTGAACTTCAAGCGGGCTGAACC-3'	*daf-7p* reverse + *daf-7* homology
5'-ATATTTTGTTGGCCCATCGT-3'	*str-3p* forward
5'-GAGTGAAGATGCCATGAACATGTTCCTTTTGAAATTGAGGC-3'	*str-3p* reverse + *daf-7* homology
5'-ATGGATACCTGATCATTC-3'	*trx-1p* forward
5'-GAGTGAAGATGCCATGAACATAAGTGAGAAGATGAAGAG-3'	*trx-1p* reverse + *daf-7* homology
5'-ATGTTCATGGCATCTTCACTC-3'	*daf-7* cDNA forward
5'-TAGGGATGTTGAAGAGTAATTGGACTTATGAGCAACCGCATTTCTT-3'	*daf-7* cDNA reverse + unc-54 homology
5'-GTCCAATTACTCTTCAACATCCCTA-3'	*unc-54* 3'UTR forward
5'-CAGTTATGTTTGGTATATTGGGAATG-3'	*unc-54* 3'UTR reverse
Oligos used for *gpa-3* and *gpa-2* rescue constructs	
5'-TTGCAAAGAAGCACATCAAAA-3'	*bbs-1p* forward
5'-GCAGATTGGCATAATCCCATTTTTTGTTAATTTTGGAGCAC-3'	*bbs-1p* reverse + *gpa-3* homology
5'-TCACTTTGGCACAGACCCATTTTTTGTTAATTTTGGAGCAC-3'	*bbs-1p* reverse + *gpa-2* homology
5'-TGAGTTGGGCACTTCGTAGA-3'	*trx-1p* forward
5'-GCAGATTGGCATAATCCCATGATCAATTGCTCAAAGTCAC-3'	*trx-1p* reverse + *gpa-3* homology
5'-TCACTTTGGCACAGACCCATGATCAATTGCTCAAAGTCAC-3'	*trx-1p* reverse + *gpa-2* homology
5'-acgtggttttcacatttgtgca-3'	*ceh-36p* forward
5'-TCACTTTGGCACAGACCCATtgtgcatgcgggggcagg-3'	*ceh-36p* reverse + *gpa-2* homology
5'-ATGGGATTATGCCAATCTGC-3'	*gpa-3* cDNA forward
5'-TAGGGATGTTGAAGAGTAATTGGACTCAGTACAAACCGCATCCTT-3'	*gpa-3* cDNA reverse + *unc-54* homology
5'-ATGGGTCTGTGCCAAAGTGA-3'	*gpa-2* cDNA forward
5'-	*gpa-2* cDNA reverse + *unc-54* homology

TAGGGATGTTGAAGAGTAATTGGACTTA ATAGAGTCCAGATTTGTGTAGGT-3'	
5'-GTCCAATTACTCTTCAACATCCCTA-3'	*unc-54* 3'UTR forward
5'-ATTTGCGCGGGAATTCAA-3'	*unc-54* 3'UTR reverse
Oligos used for GCaMP5 construct	
5'-TGAGTTGGGCACTTCGTAGA-3'	*trx-1p* forward
5'-GATGATGATGATGATGAGAACCCATGAT CAATTGCTCAAAGTCAC-3'	*trx-1p* reverse + GCaMP homology
5'-ATGGGTTCTCATCATCATCATCATC-3'	GCaMP5 Forward
5'-TAGGGATGTTGAAGAGTAATTGGACTCA CTTCGCTGTCATCATTTGTAC-3'	GCaMP5 Reverse + *unc-54* homology
5'-GTCCAATTACTCTTCAACATCCCTA-3'	*unc-54* 3'UTR forward
5'-CAGTTATGTTTGGTATATTGGGAATG-3'	*unc-54* 3'UTR reverse

APPENDIX B: SUPPLEMENTAL INFORMATION FOR CHAPTER 2

Materials and methods

Nematode Strains. All experimental analyses were done with the *P. pacificus* strain RS2333, a laboratory derivative of the original strain PS312 isolated in Pasadena, CA (USA) in 1988.

Sequence Data Sampling and Phylogenetic Analyses. For phylogenetic analyses, protein sequences for *Caenorhabditis elegans*, *Caenorhabditis briggsae*, *Caenorhabditis remanei*, *Haemonchus contortus*, *Brugia malayi*, *Loa loa*, *Ascaris suum*, and *Trichinella spiralis* were collected from BLAST searches in GenBank. Sequences for *Bursaphelenchus xylophilus*, *Meloidogyne hapla*, *Panagrellus redivivus* and *Strongyloides ratti* were collected from BLAST searches in Wormbase version WS240. Sequences of *P. pacificus* are from the HYBRID1 proteomics gene model dataset that is available on the website www.pristionchus.org (last accessed May 10, 2016) and were refined when necessarily by the new assembly dataset (SNAP2012). Sequences from the sister species *Pristionchus exspectatus* also come from the published genome assembly. Augustus predictions are available on www.pristionchus.org (last accessed May 10, 2016), whereas the SNAP predictions are available on http://parasite.wormbase.org/ (last accessed May 10, 2016). One additional experimental sequence from the Clade IV nematode *Heterodera glycines* was also incorporated. Some corrected cDNA predictions for *P. pacificus* were added in the Hybrid1 community track database that is displayed in the genome browser from www.pristionchus.org. Collected sequences were aligned with Muscle and alignments were checked by eye and edited with Seaview. Phylogenetic trees were made using PHYML, a fast and accurate maximum likelihood heuristic method, using the best estimated substitution models. Reliability of nodes was assessed by the likelihood-ratio test.

CRISPR/Cas9 Deletion Mutants. CRISPR/Cas9-induced gene inactivation was performed as

described in Witte *et al*., 2015. For *daf-22.1*, 288 F1 animals were screened for deletions by Sanger sequencing after injection. For *daf-22.2*, a coinjection procedure with the *dpy-1* marker gene enabled a reduction of screened animals to the progeny of 25 F1 animals. Two mutant lines *Ppa-daf-22.1(tu489)* and *Ppa-daf-22.2(tu504)* were obtained, with a 7-bp deletion in *tu489* and a 7-bp insertion in *tu504*. All mutant lines including the double mutant were backcrossed twice.

RNA Extraction. Nematodes were separated by centrifugation from the culture supernatant, which was extracted with methanol and analyzed for ascarosides using HPLC-MS. The worm pellet was frozen at −80 °C and then immediately transferred into Tri Reagent® (Sigma Aldrich, Munich, Germany). After three rounds of freezing in liquid nitrogen and thawing at 37 °C to break open cells, RNA was extracted with the Direct-zolM RNA MiniPrep (Zymo, Freiburg, Germany) following the manufacturer's protocol. RNA quality was determined with a NanoDrop ND 1000 spectrometer (PeqLab, Erlangen, Germany) and RNA integrity was checked with the Agilent RNA Nano Chip Assay (Agilent, Santa Clara, USA). Only RNA samples of high quality (OD 260/280 >2 and OD 260/230 > 1.8) and high integrity were used for quantitative RT-PCR.

Quantitative RT-PCR. In brief, cDNA was synthesized from complex RNA with oligoDT primers to enrich for mRNA with the Superscript II kit (Invitrogen, Carlsbad, CA, USA) following the manufacturer's protocol. RT-PCR was performed as described in Schuster *et al.*, 2012, using the LC-480 SybrGreen Mix and a LC480 light cycler (Roche, Mannheim, Germany). As reference genes *Ppa-β-tubulin* and *Ppa-actin* were used.

Dauer Pheromone Extraction and Dauer Assays. Pure cultures of all wild-type and mutant worms were grown in 500 ml liquid cultures. The supernatant was sterile-filtered and incubated with 30 g of activated charcoal under stirring. The charcoal was washed five times with water and the dauer pheromone was eluted from the charcoal with ethanol Ph. Eur. (Sigma Aldrich). Ethanol

was evaporated using a Rotavapor® R210 with cold trap (Buchi, Essen, Germany) and the remaining yellow pellet was re-suspended in 3 ml of water. For dauer formation assays, one hundred microliters of the pheromone extract were applied to the 6-cm NGM agar plates without peptone or cholesterol but with 50 μg/ml kanamycin (final concentration). Afterwards, kanamycin-inactivated OP50 and either *P. pacificus* freshly hatched J2 larvae or *C. elegans* L1 larvae were transferred to the assay plates. Note that in *P. pacificus*, the J1 stage hatches inside the eggshell. Therefore, the J2 is the first larval stage that can be collected and manipulated. Plates were incubated at 25 °C, and after 2 days dauer versus non-dauer larvae were counted. For all experiments, three independent replicates (three assay plates) for each treatment were performed.

Metabolomic Profiling. To grow mixed-stage cultures, worms from a populated 10 cm NGM agar plate seeded with *Escherichia coli* OP50 were washed into 25 ml of S-complete medium and fed OP50 on Days 1 and 3 (starved) or on Days 1, 3, 5, and 8 (ad lib) for an 11-day culture period, whereas shaking at 22°C, 220 rpm. The cultures were then centrifuged and worm pellets and supernatant frozen separately, lyophilized and extracted with 95:5 ethanol:water for 12 h. The extract was dried *in vacuo*, resuspended in methanol and analyzed by LC-MS as described below. For synchronized cultures, gravid adults and eggs were collected from six 10 cm NGM agar plates fed with *E. coli* HB101, isolated eggs by alkali hypochlorite treatment, and let hatch in M9 buffer overnight at room temperature. Hatched J2 larvae were transferred to S-complete medium with HB101 (70,000 worms per 25 ml) and grown at 22°C. After 3 days, the medium was collected, frozen and lyophilized. After overnight methanol extraction, the extract was dried *in vacuo*, resuspended in methanol and analyzed by LC-MS. LC-MS analysis was performed on a Dionex 3000 UPLC coupled with a Thermo Q Exactive high resolution mass spectrometer. In the case of analyses targeted at short-chain ascarosides, metabolites were separated using water-acetonitrile

gradient on Agilent Zorbax Eclipse XDB-C18 column, 150 mm×2.1 mm, particle size 1.8 μm, and maintained at 40°C. Solvent A: 0.1% formic acid in water; Solvent B: 0.1% formic acid in acetonitrile. A/B gradient started at 5% B at 1.5 min after injection and increased linearly to 100% B at 12.5 min. Long-chain ascarosides were analyzed on the same column using an acetonitrile-propanol gradient. Solvent B: As above. Solvent C: 100% propanol. B/C gradient started at 5% C at 1.5 min after injection and increased linearly to 75% C at 12.5 min. Most ascarosides were detected as $[M\text{-}H]^-$ ions in the negative ionization mode. Methylketones were observed as $[M+Na]^+$ adducts in the positive ionization mode. Metabolites were identified based on their high resolution masses (< 1 ppm), fragmentation spectra, and comparison of retention times with those of synthetic standards.

Supplemental figures

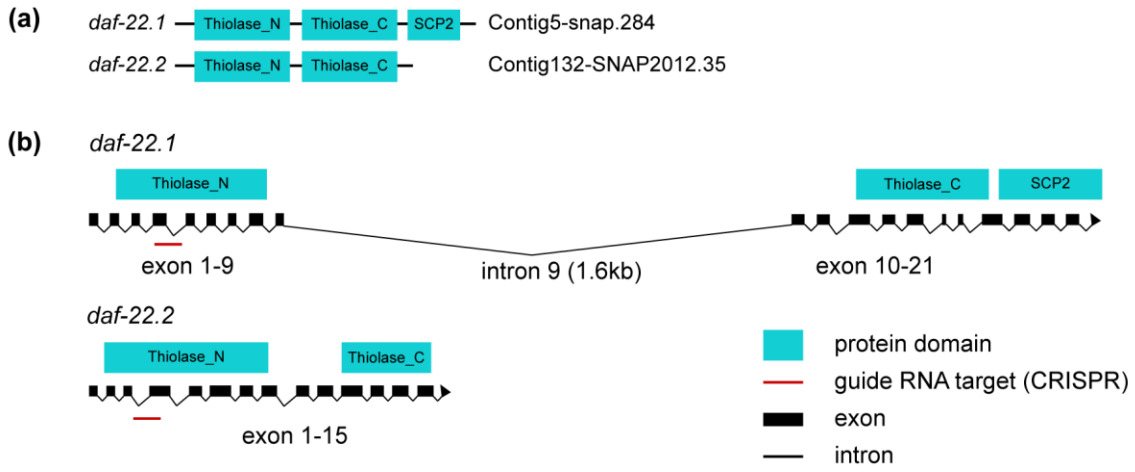

Figure B1. (a) Closest paralogs of *C. elegans daf-22* gene in *P. pacificus* encoding the thiolase of the peroxisomal β-oxidation pathway. **(b)** *T*he gene structure and protein domains of the *daf-22.1* and *daf-22.2* genes respectively.

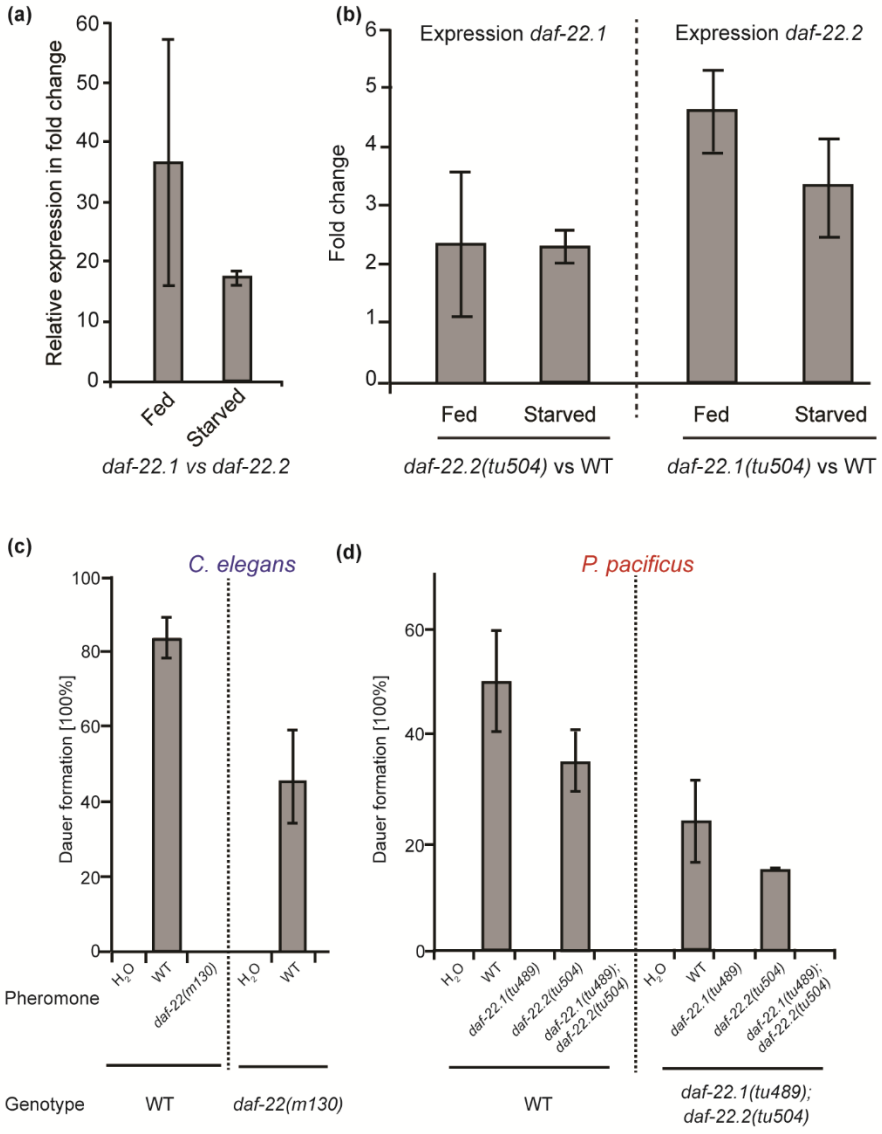

Figure B2. Relative expression of *daf-22 genes* in *P. pacificus*. **(a)** Quantitative RT-PCR experiment of *daf-22* genes reveals that *daf-22.1* is expressed 35-fold higher than *daf-22.2* under fed (*ad lib*) and approximately 18-fold higher under starved conditions. **(b)** Fold change relative to wildtype worms of *daf-22.1* and *daf-22.2* mutant worms under fed (*ad lib*) and starved conditions (mean and standard error for three biological replicates). **(c)** *C. elegans* dauer induction assay using pheromone extracts from wildtype and *daf-22(m130)* on wildtype worms. Bars represent the mean with standard error of three independent biological replicates. **(d)** *P. pacificus* dauer induction assay using pheromone extracts from wildtype, *daf-22.1*, *daf-22.2*, and the *daf-22.1;daf-22.2* double mutant on wildtype and double mutant worms. Bars represent the mean with standard error of three independent biological replicates.

APPENDIX C: SUPPLEMENTAL INFORMATION FOR CHAPTER 3

Materials and methods

***C. elegans* Strains.** Wild type (N2, Bristol), FX04381 *dhs-13*(*tm4381*), FX06263 *C24A3.4*(*tm6263*), FX03584 *ndx-9*(*tm3584*), VC754 *ctl-2*(*ok1137*), RB2147 *acs-13*(*ok2861*), RB2452 *acs-14*(*ok3391*), GS2477 arIs37; *cup-5*(*ar465*); *dpy-20*(*e1282*), RB1080 *haf-4(ok1042)*, VC32 *haf-9*(*gk23*), GH10 *glo-1* (*zu437*), RB811 *glo-4*(*ok623*), RB662 *apb-3*(*ok429*), FX06781 *acs-7*(*tm6781*), FCS1 *daf-22(ok693)*. Some strains were obtained from the Caenorhabditis Genetics Center (CGC), USA, and the National BioResource Project (NBRP), Japan. GH10 was kindly provided by D. Gems and *daf-22(ok693)* was a gift from H. Y. Mak. See Table C1 for a list of *O*-acyltransferase mutants. FCS10 *acs-7*(*tm6781*) was obtained by outcrossing FX06781 10x against GE1710 *rol-6*(*e187*);*unc-4*(*e120*). FCS10 was used for all experiments reported for *acs-7*(*tm6781*). Worms were maintained on Nematode Growth Medium (NGM) plates seeded with *E. coli* OP50 or HB101.

Nematode Culture and Extraction. Mixed stage worms from a populated 10 cm NGM agar plate seeded with *E. coli* OP50 were washed into 25 ml of S-complete medium and fed OP50 on days 1, 3 and 5 for a 7-day culture period, while shaking at 22 °C, 220 rpm. The cultures were then centrifuged and worm pellets and supernatant frozen separately, lyophilized and extracted with 35 mL of 95% ethanol at room temperature for 12 h. The extracts were dried *in vacuo*, resuspended in 200 µL methanol and analyzed by LC/MS. All cultures were grown in at least two biological replicates.

Mass Spectrometric Analysis. High resolution LC-MS analysis was performed on a Dionex 3000 UPLC coupled with a Thermo Q Exactive high-resolution mass spectrometer. Metabolites were separated using water–acetonitrile gradient on Agilent Zorbax Eclipse XDB-C18 column (150 mm

x 2.1 mm, particle size 1.8 µm) maintained at 40 °C. Solvent A: 0.1% formic acid in water; Solvent B: 0.1% formic acid in acetonitrile. A/B gradient started at 5% B for 5 min after injection and increased linearly to 100% B at 12.5 min. Most ascarosides were detected as $[M-H]^-$ ions or $[M+Cl]^-$ adducts in the negative ionization mode (spray voltage 3 kV) and confirmed based on their high-resolution masses (< 1 ppm), fragmentation spectra, and comparison of retention times with those of synthetic standards.

Low resolution LC-MS was performed using the Agilent 1100 Series HPLC system equipped with an Agilent Eclipse XDB-C18 column (250 mm x 9.4 mm, particle size 5 µm), connected to a Quattro II or Quattro Ultima mass spectrometer. Solvent A: 0.1% acetic acid in water; Solvent B: 0.1% acetic acid in acetonitrile. A/B gradient started at 5% B for 5 min after injection and increased linearly to 100% B over a period of 40 min. Ascarosides were detected as $[M-H]^-$ ions in the negative ionization mode (spray voltage 3.5 kV, cone voltage -40 V) and confirmed based on comparison of retention times with those of synthetic standards.

Ascr#1 Feeding Experiment. Mixed stage *daf-22(ok693)* worms from a populated 10 cm NGM agar plate seeded with *E. coli* OP50 were washed into two flasks containing 10 ml of S-complete medium and 2% HB101. One flask additionally contained 10 µM of synthetic ascr#1. HB101 was added on days 1, 3 and 5 for a 7-day culture period, while shaking at 22 °C, 220 rpm. The medium was then collected, processed, and analyzed by high-resolution HPLC-MS as described above, revealing production of icas#1. None of the other known ascarosides were observed. These results are consistent with the results reported in von Reuss *et al.*, 2012, in which ascr#3 was fed to *daf-22* worms and production of icas#3 was observed.

Heterologous Protein Expression and Purification. The protein coding genetic sequence of *acs-7* was synthesized (Biomatik) and cloned into pET-21a(+) with primers 3′-

ATAATTTTGTTTAACTTTAAGAAGGAGATATACATATGATATTTCACGGTGAACAAC TTG-5′ and 5′-GTTAGCAGCCGGATCTCAGTGGTGGTGGTGGTGGTGCAATTTAGCC TTTTTTGCATCCA-3 ′ using ligase-free PCR cloning to construct C-terminally tagged hexa-histidine ACS-7. The expression vector was then transformed into BL21(DE3) (New England Biolabs) *E. coli* and grown in Terrific Broth (TB) supplemented with 10 mM $MgCl_2$ and selected with 100 µg/mL ampicillin. 10 mL overnight cultures were diluted into 1 L of TB in a 4 L Erlenmeyer flask and shaken at 200 RPM at 37 °C to an OD of approximately 0.75, cooled to 16 °C and further grown to an OD of roughly 1.0-1.2 and induced with 100 µM IPTG. Cultures were maintained at 16 °C at 200 RPM for an additional 18 hours before harvesting at 5000xg, 4 °C for 10 min and stored at -80 °C until purification. All further steps occur at 4 °C unless otherwise noted.

10 g of frozen pellets were resuspended in 100 mL of 50 mM sodium phosphate pH 7.6, 300 mM NaCl, 0.3 mM PMSF and sonicated. Lysed cells were spun at 20,000x*g* for 20 min and the supernatant was collected and gently stirred with 1 mL HisPur Ni-NTA Resin (Thermo Fisher Scientific) for 30 min. During incubation, 5 uL of Benzonase (EMD Millipore) was added along with 1 mM $MgCl_2$. The slurry was loaded and passed through a column and the resin was washed with 20 column volumes of fresh lysis buffer without PMSF. The protein was then eluted with 20 mL lysis buffer containing 200 mM imidazole and 10% glycerol. The elution was concentrated with an Amicon Ultra-15 30 kDa spin filter (EMD Millipore), flash frozen in liquid nitrogen and stored at -80 °C until further purification. FPLC purification was performed using a HiLoad 16/600 Superdex 200 prep grade column run on a Amersham Biosciences P-920 pump equipped with a UPC-900 detector and a Frac-950 fraction collector (GE Healthcare) with a running buffer of 50 mM sodium phosphate pH 7.6, 300 mM NaCl, and 10% glycerol. Fractions containing ACS-7

were combined and concentrated with an Amicon Ultra-15 30 kDa spin filter and flash frozen in liquid nitrogen and stored at -80 °C until further analysis.

***In vitro* assays.** All assays were performed with ACS-7 assay buffer containing: 100 mM potassium phosphate pH 7.0, 5 mM ATP, 5 mM $MgCl_2$. For representative HPLC chromatograms, assay buffer was incubated with 1.7 μM ACS-7 and 100 μM putative substrate(s) (pentanoic acid, ascr#9, ascr#9-SCoA, indole-3-carboxylic acid, *N*-succinyl octopamine, as well as combinations of indole-3-carboxylic acid or *N*-succinyl octopamine with ascr#9 or ascr#9-SCoA), which were added from concentrated ethanolic stock solutions so that the final ethanol concentration remained below 1%. The reaction was allowed to incubate for 2 hours before analysis by HPLC/MS. For kinetic analysis of ACS-7, various concentration of indole-3-carboxylic acid were added to 170 nM ACS-7 in ACS-7 assay buffer in 1 mL reaction volumes and timed injections were collected and analyzed by HPLC/MS. Kinetic data analysis was performed using GraphPad Prism version 6.0 for Windows.

Synthesis of the CoA-thioester of ascr#9 (ascr#9-SCoA). Synthetic ascr#9 (2 μmol), tetrafluorophenol (6 μmol), DIPEA (12 μmol), EDC·HCl (4 μmol), and DMAP (0.2 μmol) were dissolved in 0.5 mL of DMF and stirred for 36 hours at room temperature under argon. The reaction mixture was diluted with a 1:1 mixture of ethyl acetate and 0.1 M aqueous HCl (2 mL), and the organic layer was collected and then washed two additional times with 1 mL of 0.1 M HCl. The organic layer was dried and dissolved in 625 μL of a mixture of DMF and aqueous 10 mM potassium phosphate (4:1), containing the sodium salt of coenzyme A (2 μmol), and stirred for 16 hours at room temperature under argon. Subsequently, ascr#9-SCoA was isolated by reverse phase HPLC, using acetonitrile and water, both containing 0.1% acetic acid. A gradient was used starting with 1% acetonitrile for 5 minutes, followed by a linear gradient to 100% acetonitrile over 27.5

minutes. After 2.5 minutes at 100% acetonitrile, the column was re-equilibrated at 1% acetonitrile for 5 minutes. NMR spectra of the purified compound were identical to those reported by Zhang *et al.*, 2015.

Microscopy. For LysoTracker staining, 0.5 ml of 2 μM Lysotracker Deep Red (obtained from Thermo Fisher as 1 mM stock solution in DMSO) diluted in M9 buffer was added to an NGM plate seeded with *E. coli* OP50 and incubated in the dark at 20 ºC for 24 h. Worms were then added to the plate and allowed to grow overnight in the dark. For imaging, worms were removed from the plate and transferred onto a glass slide with a thin agarose pad containing sodium azide or levamisole to immobilize worms during imaging. Microscopic analysis was performed with Leica TCS SP5 laser scanning confocal microscope. GFP was excited with 488 nm Argon laser line and the emission detector was set at 500-550 nm. LysoTracker Red stain was excited with 561 nm diode-pumped solid-state laser, while the detector was set at 570-650 nm. Images were taken at 1024x1024 pixel resolution and 100-400 Hz scanning rate. Dry 40x/0.85 and oil 63x/1.4 objectives were used.

Supplemental figures

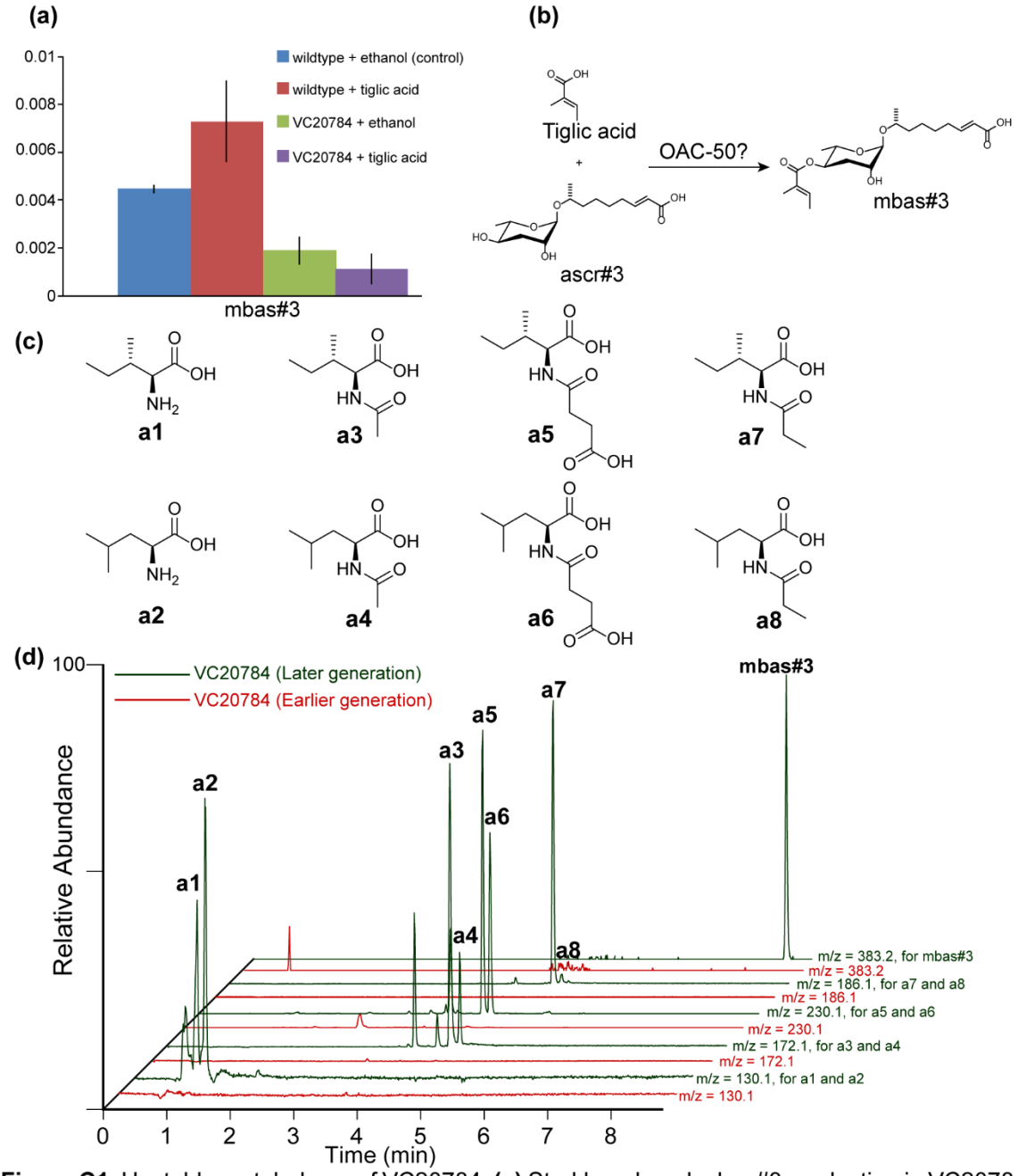

Figure C1. Unstable metabolome of VC20784. **(a)** Starkly reduced mbas#3 production in VC20784 worms, which bear *oac-50(gk402144)*. Production of mbas#3 could not be recovered upon exogenous addition of 2μM tiglic acid. Error bars represent standard deviation of 2 biological replicates. **(b)** Proposed biosynthesis of mbas#3, where tiglic acid and ascr#3 are combined, possibly by OAC-50. **(c)** Structures of isoleucine **(a1)**, leucine **(a2)**, *N*-acetylisoleucine **(a3)**, *N*-acetylleucine **(a4)**, *N*-succinylisoleucine **(a5),** *N*-succinylleucine **(a6),** *N*-propionylisoleucine **(a7)**, *N*-propionylleucine **(a8)**. **(d)** HPLC-MS (ESI-) ion chromatograms showing **a1-a8** and mbas#3 in VC20784 (earlier generation) and VC20784 (later generation).

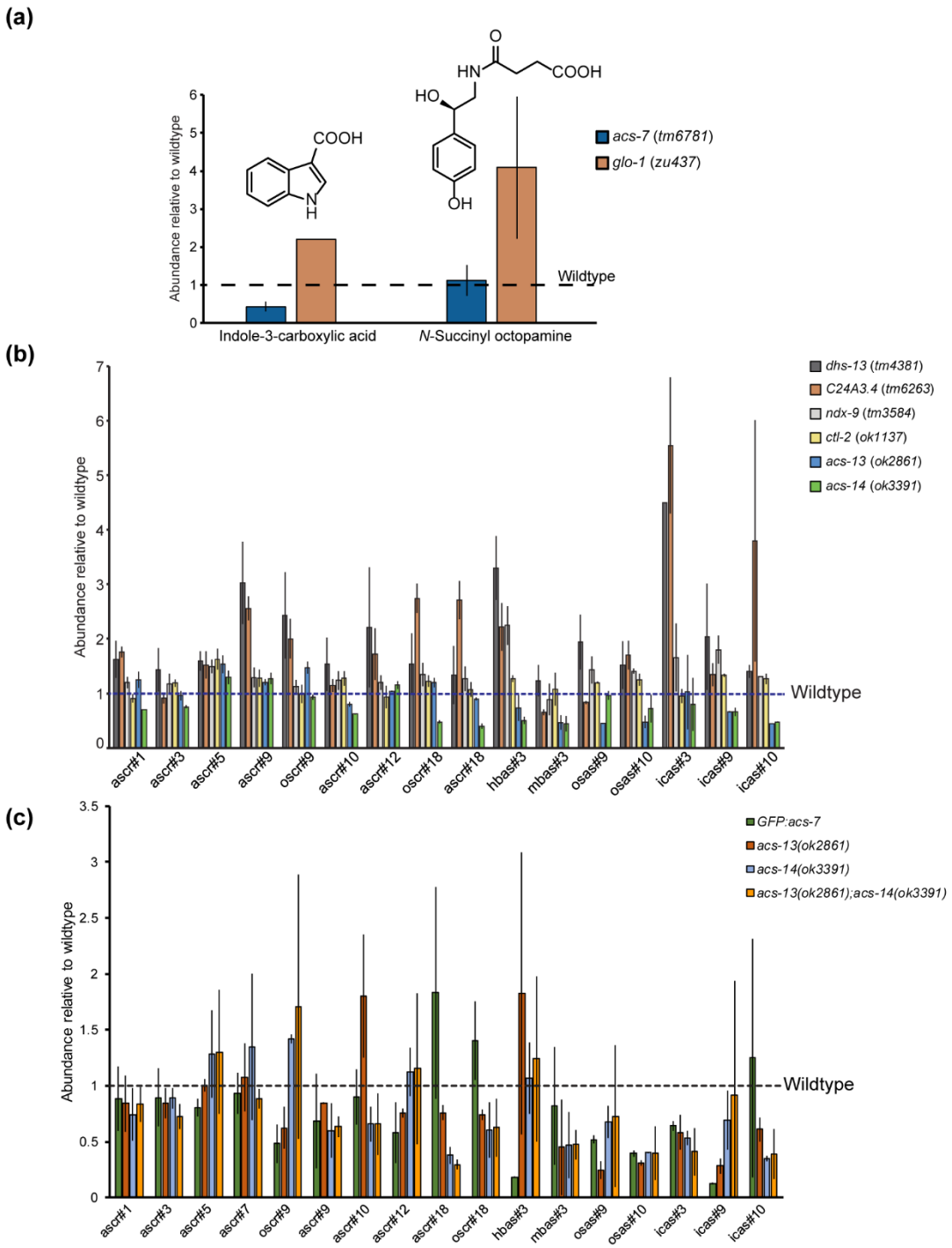

Figure C2. Quantification of metabolites by HPLC/MS. **(a)** Relative abundances of indole-3-carboxylic acid and *N*-succinyl octopamine in *acs-7* and *glo-1* mutants, as determined by negative-ion ESI HPLC-MS. Samples were prepared from synchronized mixed stage cultures as described above (for measurement of indole-3-carboxylic acid) or L1-stage larvae that were incubated for 2.5 days (for measurement of *N*-succinyl octopamine). Error bars represent standard deviations of two biological replicates. Relative abundances of ascarosides in knock-out mutants of putative peroxisome-targeted genes **(b,c)** and transgenic worms carrying *acs-7p::gfp::acs-7* in *acs-7* mutant background **(c)**, as determined by negative-ion ESI HPLC-MS. Error bars represent standard deviations of two technical replicates.

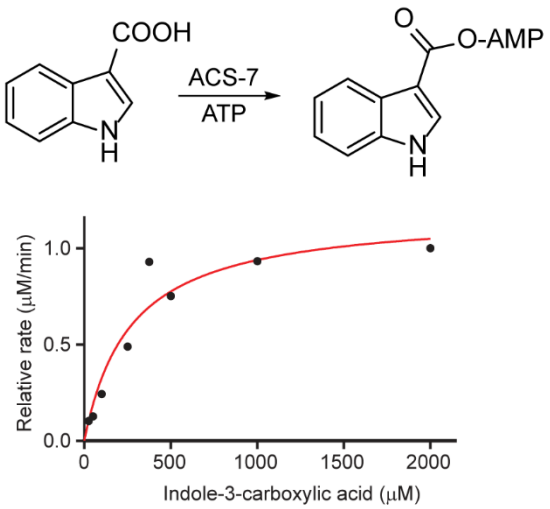

Figure C3. Steady-state kinetics evaluated for ACS-7 operating on indole-3-carboxylic acid; Km = 270±90 μM at 25 °C.

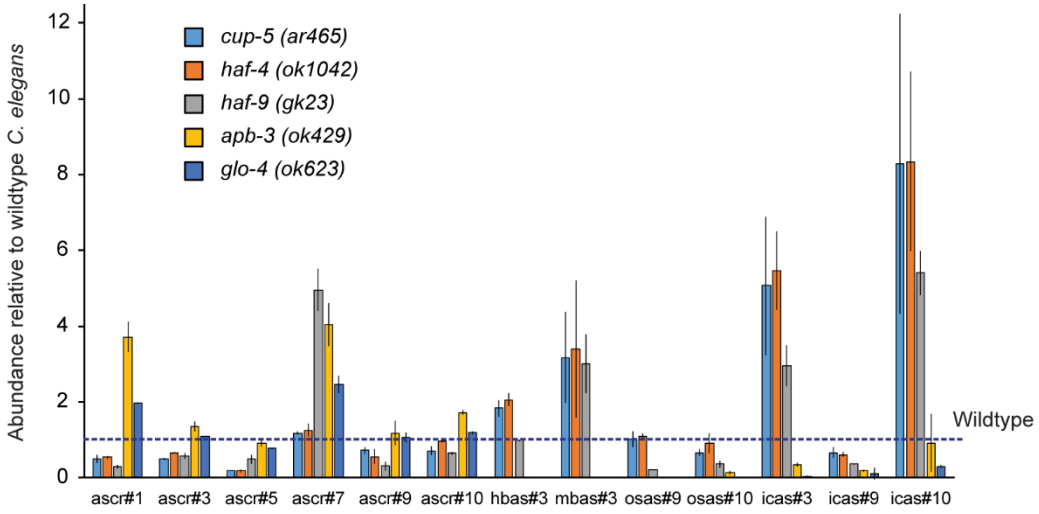

Figure C4. Relative abundances of ascarosides, as determined by negative-ion ESI HPLC-MS, in *glo-4(ok623)* and *apb-3(ok429)* mutants, in which acidic LRO formation is reduced, but not abolished, as well as *cup-5(ar465)*, *haf-4(ok1042)* and *haf-9(gk23)* mutants, which are defective in the formation of non-acidic gut granules, but have normal acidic LROs. Error bars represent standard deviations of at least two biological replicates.

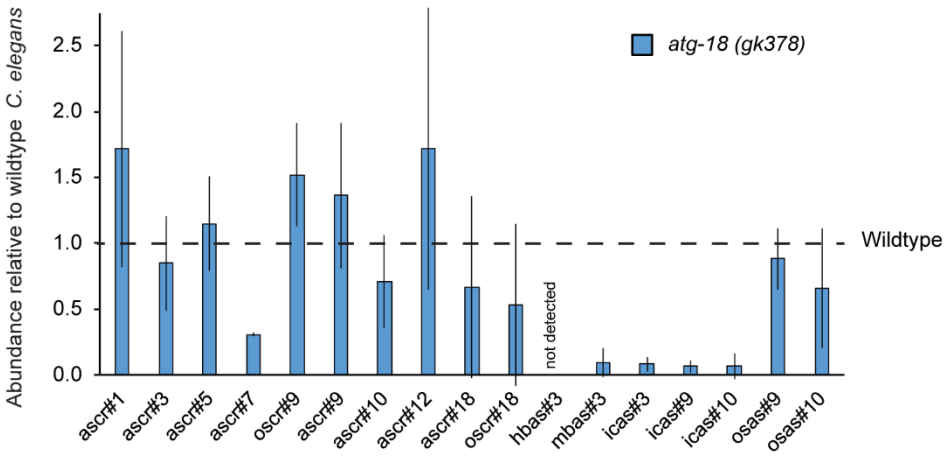

Figure C5. Relative abundances of ascarosides in autophagy-deficient *atg-18* mutants, as determined by negative-ion ESI HPLC-MS. Error bars represent standard deviations of two biological replicates.

Supplemental table

Table C1. List of O-acyltransferase mutants screened.

Gene	**Strain**	**Source**
ndg-4 (*sa529*)	JT529	Deletion mutant from CGC
nrf-6 (*sa525*)	JT525	Deletion mutant from CGC
oac-11 (*gk531381*)	VC40243	Million Mutant Project
oac-14 (*gk519224*)	VC40217	Million Mutant Project
oac-14 (*gk786954*)	VC40738	Million Mutant Project
oac-16 (*gk914989*)	VC40988	Million Mutant Project
oac-20 (*gk256989*)	VC10128	Million Mutant Project
oac-23 (*gk445127*)	VC30240	Million Mutant Project
oac-27 (*gk694121*)	VC40561	Million Mutant Project
oac-29 (*gk646323*)	VC40455	Million Mutant Project
oac-3 (*gk252641*)	VC20209	Million Mutant Project
oac-34 (*gk652397*)	VC40469	Million Mutant Project
oac-35 (*gk883174*)	VC40922	Million Mutant Project
oac-36 (*gk124636*)	VC20551	Million Mutant Project
oac-38 (*gk648702*)	VC40461	Million Mutant Project
oac-39 (*gk145*)	VC247	Deletion mutant from CGC
oac-4 (*gk363869*)	VC20633	Million Mutant Project
oac-40 (*gk242459*)	VC20235	Million Mutant Project
oac-41 (*gk242464*)	VC20211	Million Mutant Project
oac-41 (*gk766757*)	VC40696	Million Mutant Project

Gene	**Strain**	**Source**
oac-42 (*WBVar00026015*)	CB4856	Wild isolate
oac-43 (*gk737013*)	VC40638	Million Mutant Project
oac-49 (*gk264099*)	VC20294	Million Mutant Project
oac-5 (*gk398429*)	VC30020	Million Mutant Project
oac-50 (*gk402144*)	VC20784	Million Mutant Project
oac-51 (*gk533438*)	VC40246	Million Mutant Project
oac-54 (*gk684785*)	VC40540	Million Mutant Project
oac-6 (*gk735518*)	VC40635	Million Mutant Project
oac-7 (*gk586689*)	VC40345	Million Mutant Project
oac-8 (*gk211086*)	VC20046	Million Mutant Project
oac-9 (*gk662463*)	VC40490	Million Mutant Project
oac-26 (*WBVar00158777*)	CB4856	Wild isolate

APPENDIX D: SUPPLEMENTAL INFORMATION FOR CHAPTER 4

Materials and methods

***C. elegans* Strains.** N2 (Bristol), IU279.1 *ceh-23(ms23)*, IU445 *cep-1(gk138)*, IU291.4 *isp-1(qm150)*, IU398.1 *nuo-6(qm200)*, IU231.2 *mev-1(kn1)*, IU293.2 *ceh-23(ms23); isp-1(qm150)*, IU448 *cep-1(gk138); isp-1(qm150)*, IU450.1 *mev-1(kn1); cep-1(gk138)*, IU399.2 *nuo-6(qm200);ceh-23(ms23)*, IU481.3 *ceh-23(ms23); mev-1(kn1)* and *nuo-6(qm200);cep-1(gk138)* were obtained from the Lee research group at Cornell University. Worms were maintained on Nematode Growth Medium (NGM) plates seeded with *E. coli* OP50.

Nematode Culture and Extraction. Mixed stage worms from a populated 10 cm NGM agar plate seeded with *E. coli* OP50 were washed into 25 ml of S-complete medium and fed OP50 twice during the total culture period, whenever the nematodes appeared to have consumed remaining food in the culture media, while shaking at 22 °C, 220 rpm. A quarter of this culture volume was then diluted to 25 mL in S-complete in a fresh Erlenmeyer flask, and subjected to a similar feeding and shaking regimen for the total culture period. When the OP50 had been completely consumed, the cultures were then centrifuged and worm pellets and supernatant frozen separately and lyophilized. The lyophilized supernatant was extracted with 35 mL of 95% ethanol at room temperature for 12 h. The lyophilized worm pellet was crushed over dry ice and extracted with 10 mL of methanol at room temperature for 12 h. The extracts were dried *in vacuo*, resuspended in 200 µL methanol and analyzed by LC-MS. All cultures were grown in at least three independent biological replicates.

Mass Spectrometric Analysis. High resolution LC-MS analysis was performed on a Dionex 3000 UPLC coupled with a Thermo Q Exactive high-resolution mass spectrometer as described in Appendix C. Metabolites from worm media samples were separated using water-acetonitrile

gradient on Agilent Zorbax Eclipse XDB-C18 column (150 mm x 2.1 mm, particle size 1.8 µm) maintained at 40 °C. Solvent A: 0.1% formic acid in water; Solvent B: 0.1% formic acid in acetonitrile. A/B gradient started at 5% B for 5 min after injection and increased linearly to 100% B at 12.5 min. For worm pellet and mitochondrial samples, A/B gradient started at 1% B for 5 min after injection and increased linearly to 100% B at 13.5 min.

Supplemental figures

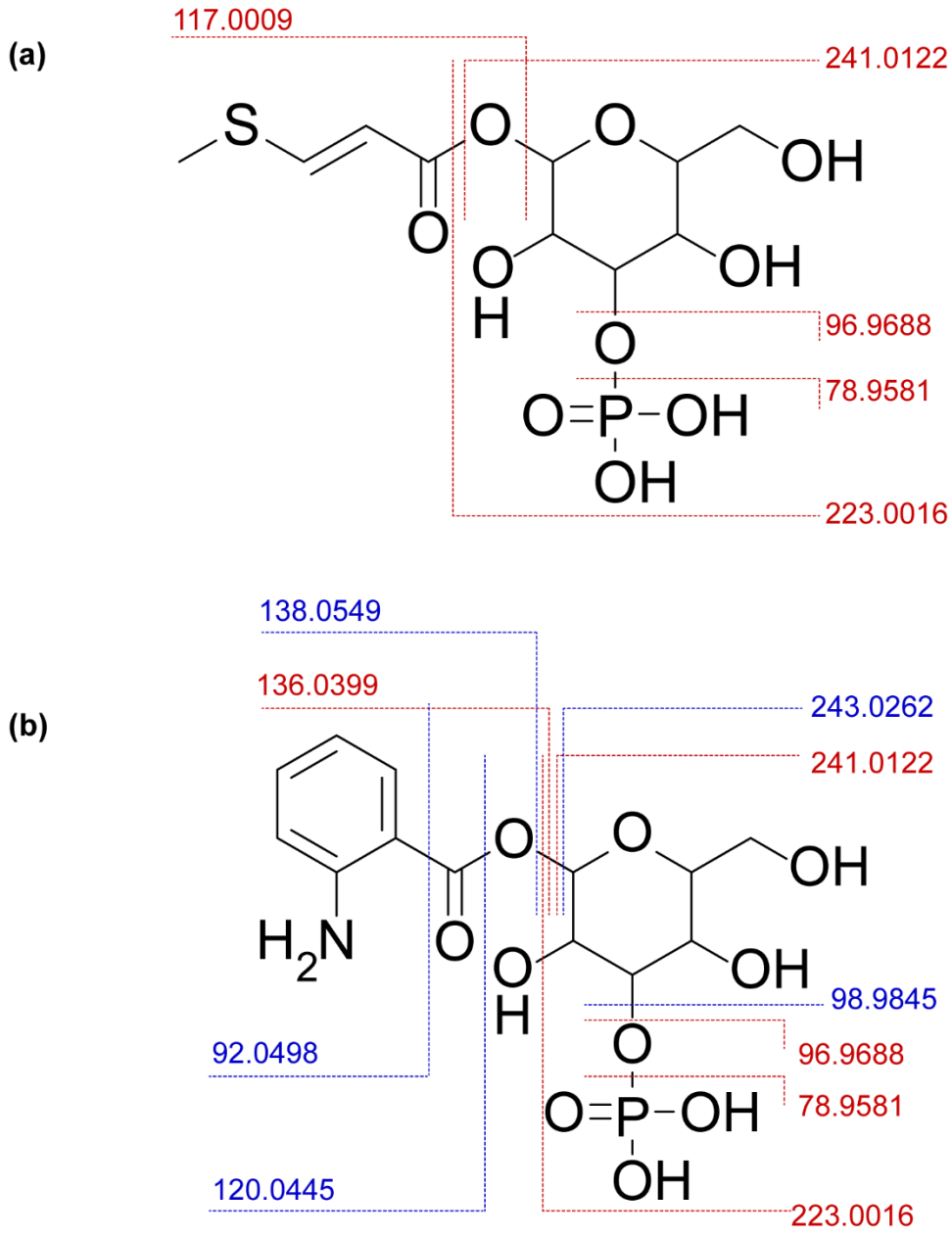

Figure D1. MS/MS fragmentation patterns of differential compounds. (a) Fragments for m/z (-) 359.02073, $C_{10}H_{16}O_{10}PS^-$, which ionizes poorly in ESI positive mode. **(b)** Fragments for m/z (-) 378.05956, $C_{13}H_{17}O_{10}NP^-$, a metabolite known as angl#2. ESI negative fragments are shown in red, ESI positive fragments are shown in blue.

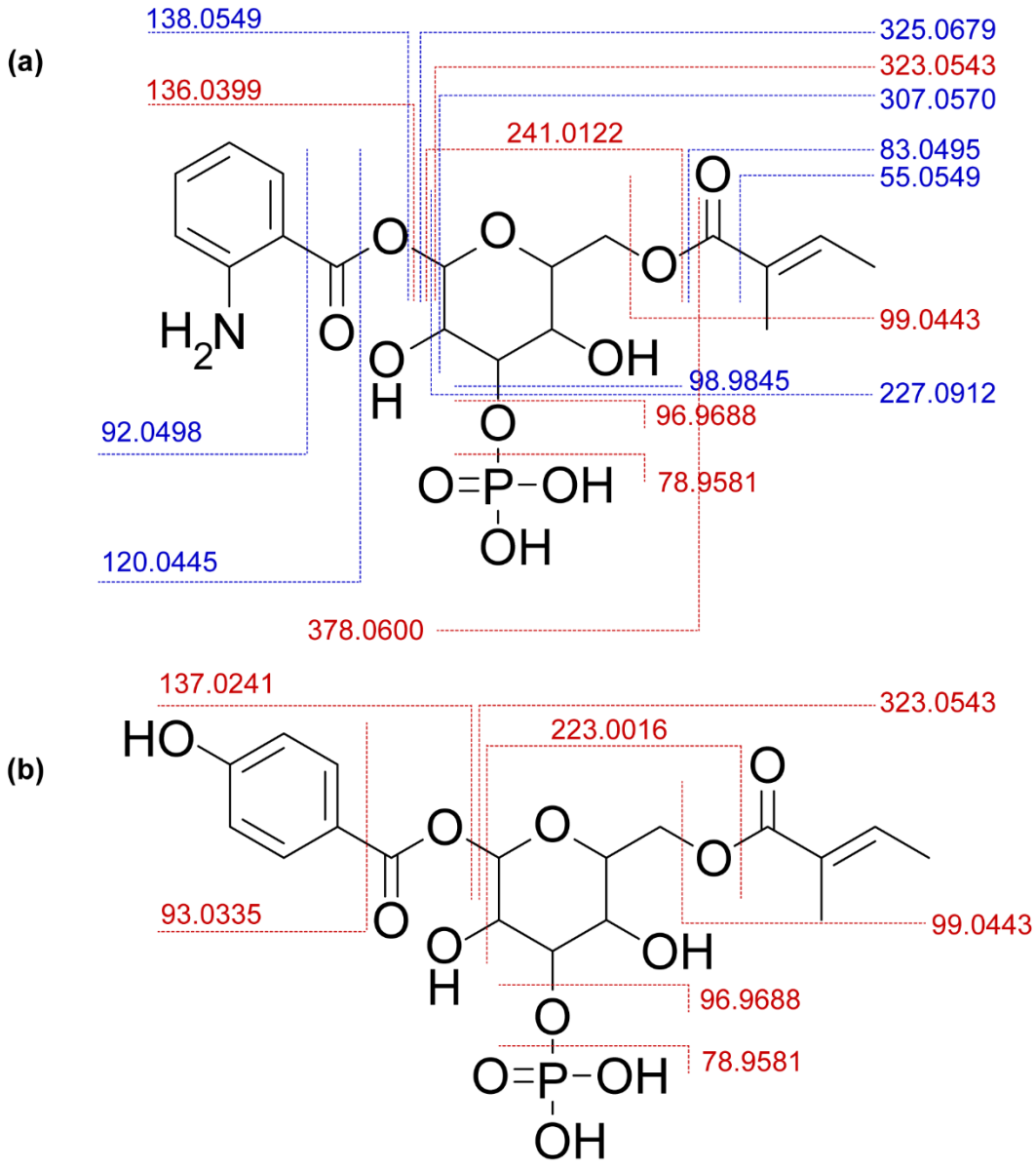

Figure D2. MS/MS fragmentation patterns of differential compounds. (a) Fragments for m/z (-) 460.10142, $C_{18}H_{23}O_{11}NP^-$, ESI negative fragments are shown in red, ESI positive fragments are shown in blue. **(b)** Fragments for m/z (-) 461.08544, $C_{18}H_{22}O_{12}P^-$, which ionizes poorly in ESI positive mode. MS/MS fragments alone are insufficient to distinguish between isomers.

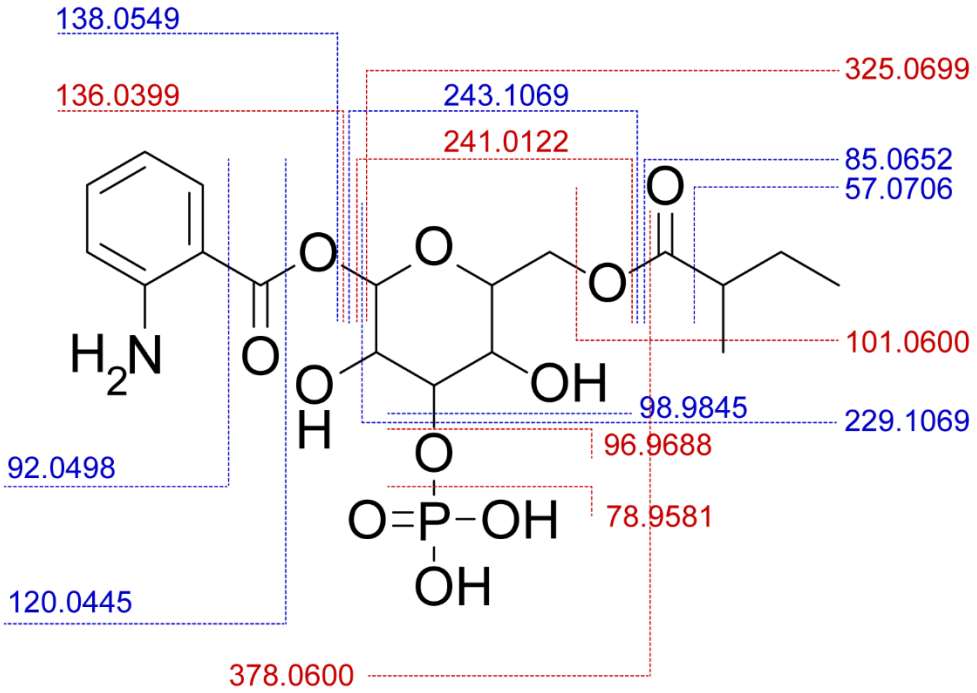

Figure D3. MS/MS fragmentation patterns of differential compounds. Fragments for m/z (-) 462.11707, $C_{18}H_{25}O_{11}NP^-$, ESI negative fragments are shown in red, ESI positive fragments are shown in blue. MS/MS fragments alone are insufficient to distinguish between isomers.

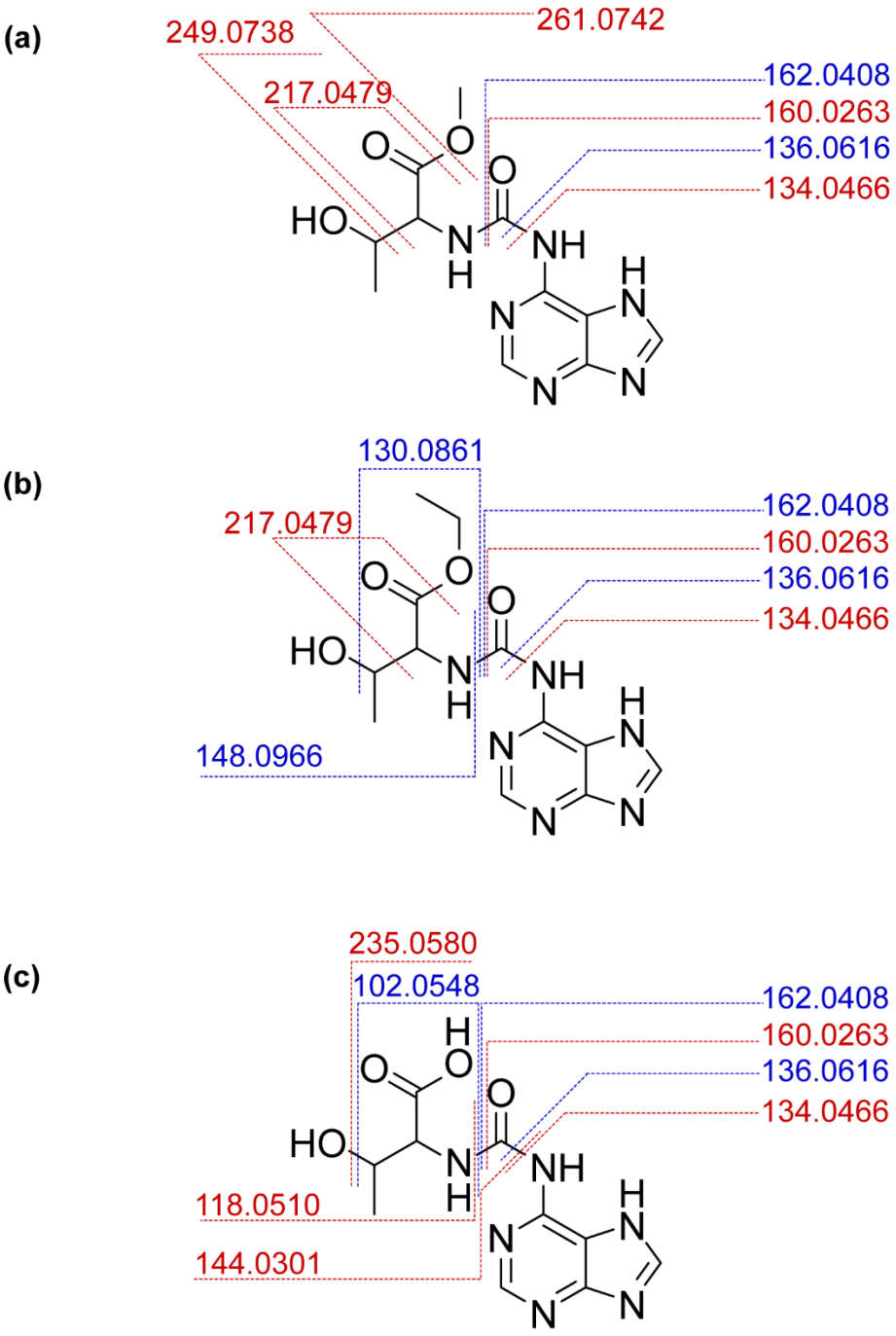

Figure D4. MS/MS fragmentation patterns of differential compounds. Fragments for **(a)** tade#1, **(b)** tade#2 and **(c)** free acid of tade, ESI negative fragments are shown in red, ESI positive fragments are shown in blue.

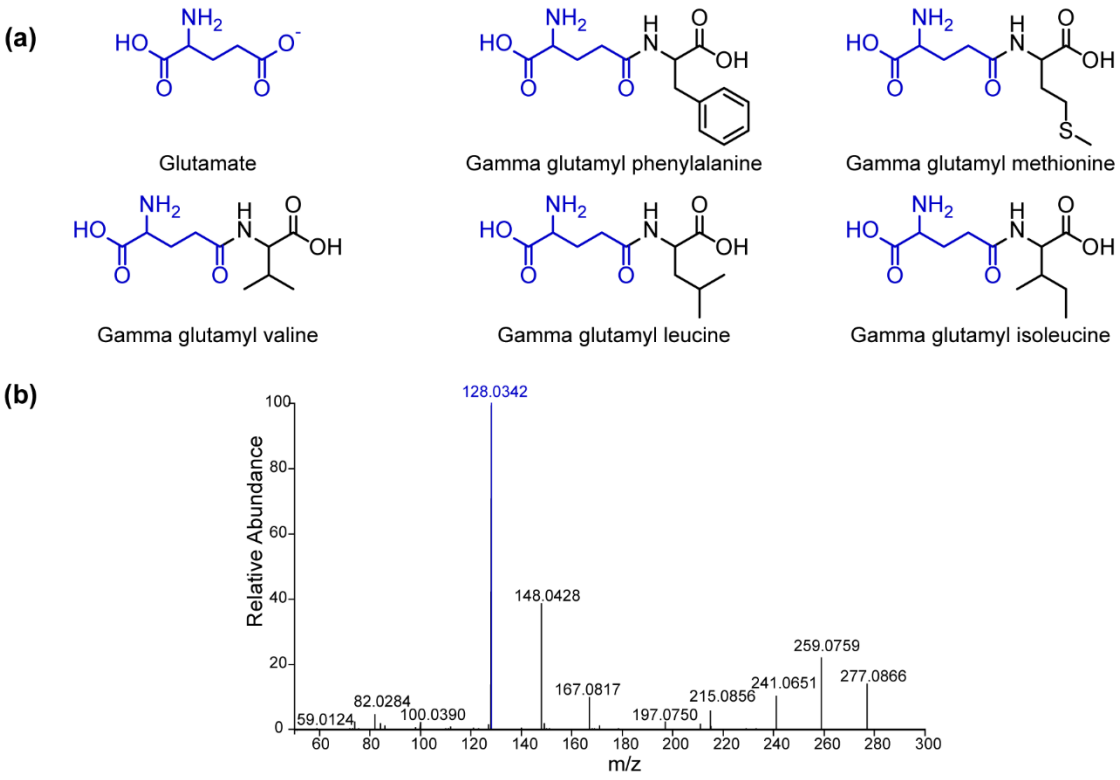

Figure D5. Gamma glutamyl dipeptides enriched in the mitochondria. (a) Tentative structures based on MS/MS fragmentation. **(b)** ESI negative spectrum for putative gamma glutamyl methionine showing prominent peak at m/z 128.0342, which is characteristic of gamma glutamyl dipeptides.

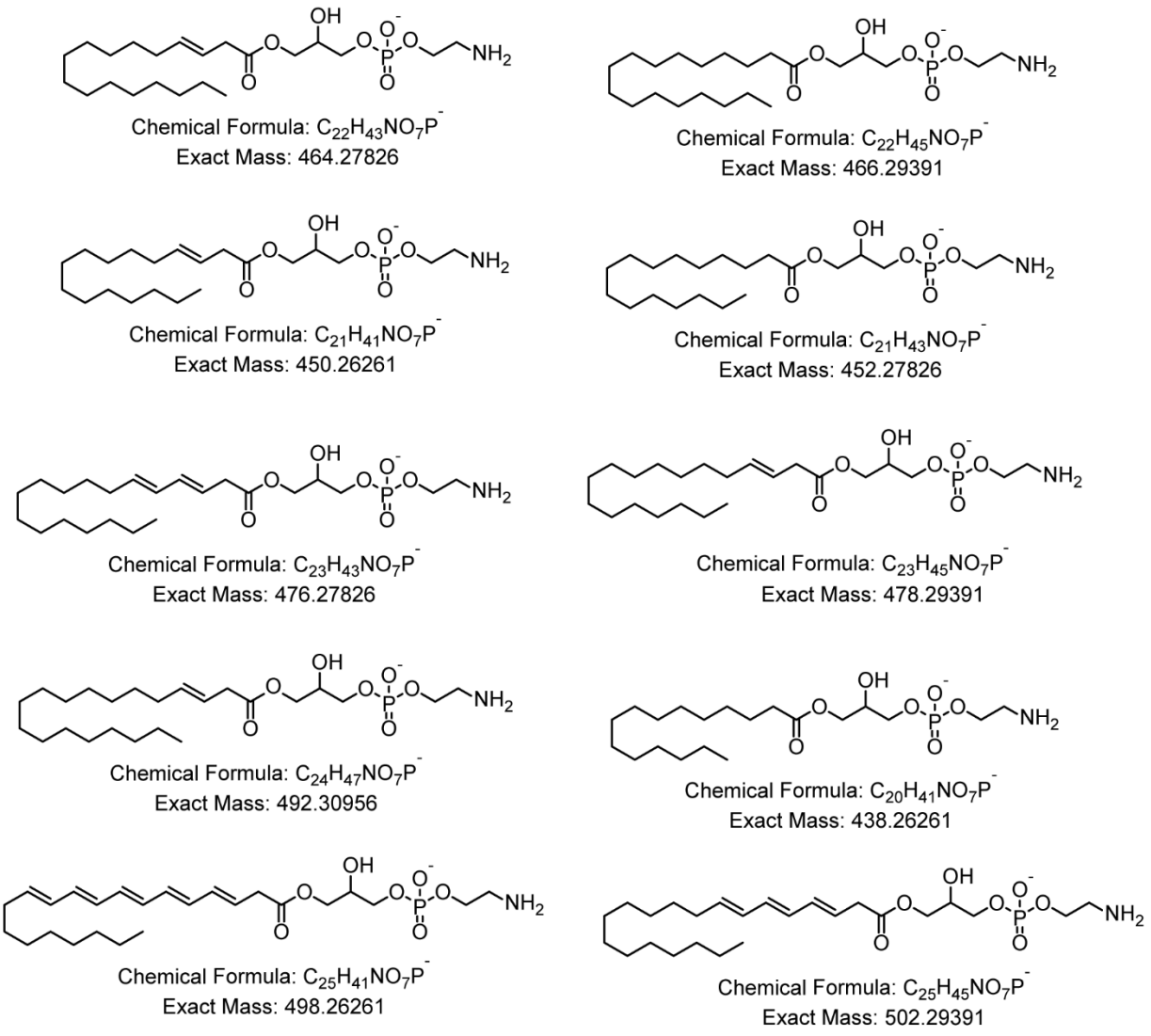

Figure D6. Tentative structures of phospholipids enriched in the mitochondria (clustered in the blue box of Figure 4.5). Structures are assigned based on fragmentation patterns that match with www.metlin.scripps.edu. Positions and stereochemistry of double bonds remain to be determined.

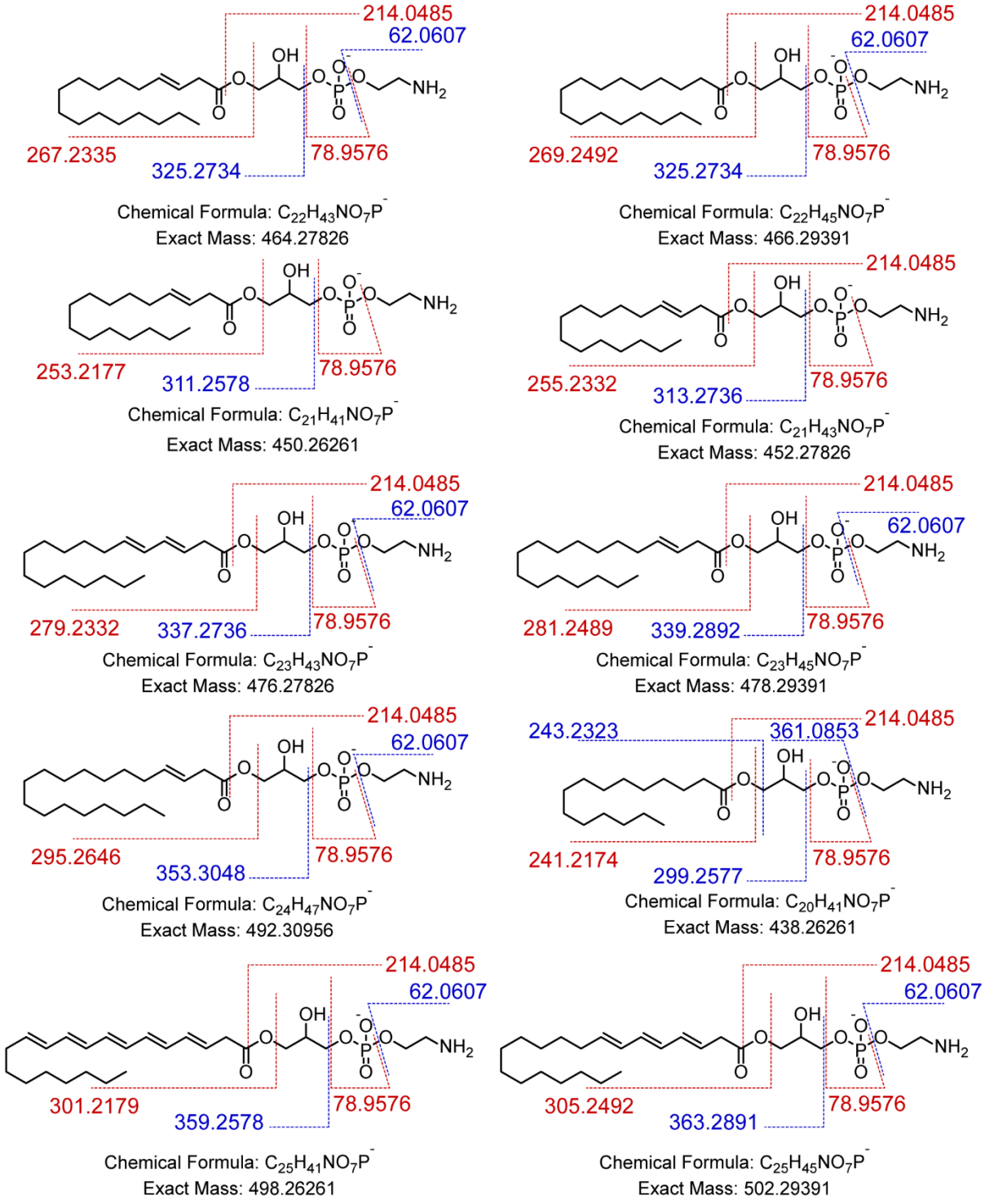

Figure D7. MS/MS fragmentation patterns of phospatidylethanolamines enriched in the mitochondria. ESI negative fragments are shown in red, ESI positive fragments are shown in blue.

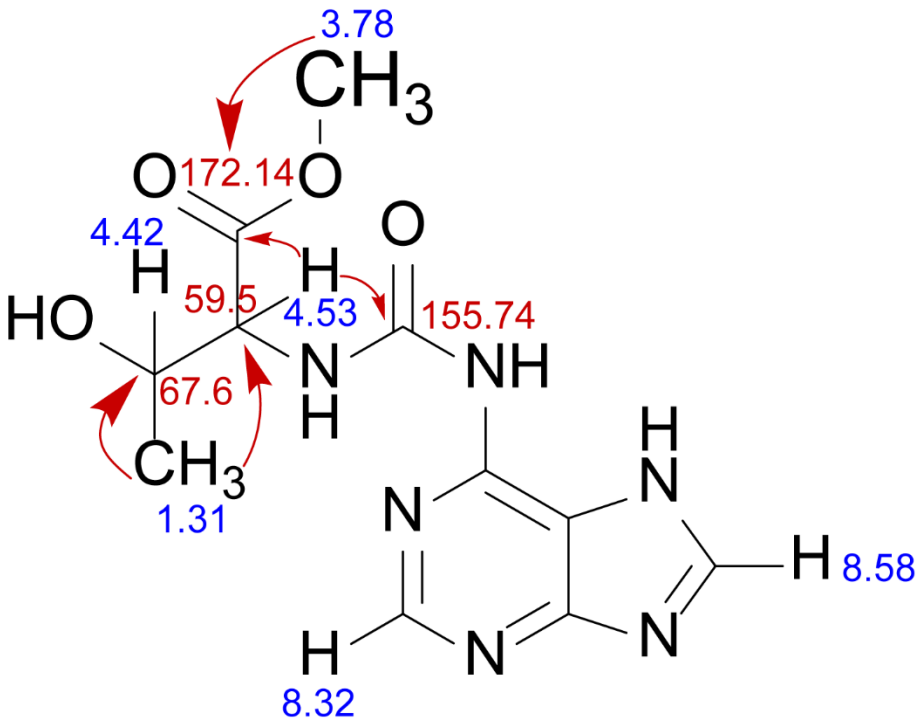

Figure D8. NMR chemical shifts for tade#1. ^{1}H chemical shifts are shown in blue, ^{13}C chemical shifts are shown in red. Red arrows indicate relevant HMBC correlations.

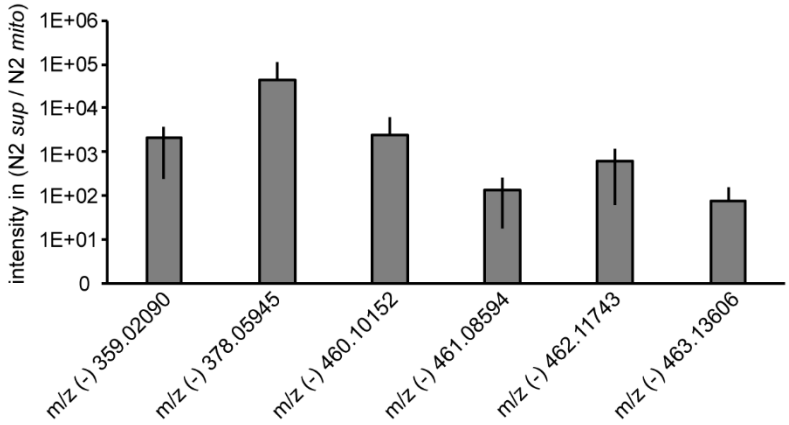

Figure D9. Phosphohexose derivatives in N2 *sup* relative to N2 *mito*. The phosphohexose derivatives are enriched in the N2 sup. Error bars = s.d. of three biological replicates. The y-axis was plotted in the logarithmic scale (base 10).

Supplemental tables

Table D1. MS/MS fragments for putative gamma glutamyl dipeptides

Observed fragments for putative gamma glutamyl isoleucine

259.13004	$C_{11}H_{19}O_5N_2^-$	261.14404	$C_{11}H_{21}O_5N_2^+$
241.11928	$C_{11}H_{17}O_4N_2^-$	244.11761	$C_{11}H_{18}O_5N^+$
223.10863	$C_{11}H_{15}O_3N_2^-$	198.11230	$C_{10}H_{16}O_3N^+$
197.12898	$C_{10}H_{17}O_2N_2^-$	170.11734	$C_9H_{16}O_2N^+$
179.11813	$C_{10}H_{15}ON_2^-$	132.10181	$C_6H_{14}O_2N^+$
130.08614	$C_6H_{12}O_2N^-$	86.09679	$C_5H_{12}N^+$
128.03410	$C_5H_6O_3N^-$	84.04481	$C_4H_6ON^+$
84.04411	$C_4H_6ON^-$		
82.02846	$C_4H_4ON^-$		
74.02334	$C_2H_4O_2N^-$		

Observed fragments for putative gamma glutamyl leucine

259.13028	$C_{11}H_{19}O_5N_2^-$	261.14441	$C_{11}H_{21}O_5N_2^+$
241.11955	$C_{11}H_{17}O_4N_2^-$	244.11783	$C_{11}H_{18}O_5N^+$
223.10881	$C_{11}H_{15}O_3N_2^-$	198.11229	$C_{10}H_{16}O_3N^+$
197.12923	$C_{10}H_{17}O_2N_2^-$	142.05009	$C_6H_8O_3N^+$
179.11832	$C_{10}H_{15}ON_2^-$	132.10191	$C_6H_{14}O_2N^+$
130.08623	$C_6H_{12}O_2N^-$	114.05508	$C_5H_8O_2N^+$
128.03418	$C_5H_6O_3N^-$	86.09687	$C_5H_{12}N^+$
112.03924	$C_5H_6O_2N^-$	84.04486	$C_4H_6ON^+$
84.04409	$C_4H_6ON^-$		
82.02840	$C_4H_4ON^-$		
74.02329	$C_2H_4O_2N^-$		

Observed fragments for putative gamma glutamyl methionine

277.08658	$C_{10}H_{17}O_5N_2S^-$	279.10065	$C_{10}H_{19}O_5N_2S^+$
259.07587	$C_{10}H_{15}O_4N_2S^-$	262.07404	$C_{10}H_{16}O_5NS^+$
241.06512	$C_{10}H_{13}O_3N_2S^-$	231.09740	$C_9H_{15}O_5N_2^+$
215.08569	$C_9H_{15}O_2N_2S^-$	216.06885	$C_9H_{14}O_3NS^+$
197.07494	$C_9H_{13}O_2NS^-$	168.06535	$C_8H_{10}O_3N^+$
167.08162	$C_8H_{11}O_2N_2^-$	150.05826	$C_5H_{12}O_2NS^+$
148.04277	$C_5H_{10}O_2NS^-$	133.03171	$C_5H_9O_2S^+$
128.03409	$C_5H_6O_3N^-$	130.04984	$C_5H_8O_3N^+$
112.03898	$C_5H_6O_2N^-$	104.05309	$C_4H_{10}NS^+$
100.03895	$C_4H_6O_2N^-$	87.02670	$C_4H_7S^+$
86.02335	$C_3H_4O_2N^-$	84.04482	$C_4H_6NO^+$
84.04417	$C_4H_6ON^-$	74.06055	$C_3H_8NO^+$
82.02844	$C_4H_4ON^-$	61.01129	$C_2H_5S^+$
74.02335	$C_2H_4O_2N^-$	56.05019	$C_3H_6N^+$

Observed fragments for putative gamma glutamyl phenylalanine

293.11481	$C_{14}H_{17}O_5N_2^-$	295.12860	$C_{14}H_{19}O_5N_2^+$
275.10422	$C_{14}H_{15}O_4N_2^-$	278.10199	$C_{14}H_{16}O_5N^+$
257.09332	$C_{14}H_{13}O_3N_2^-$	232.09660	$C_{13}H_{14}O_3N^+$
231.11366	$C_{13}H_{15}O_2N_2^-$	186.09116	$C_{12}H_{12}ON^+$
213.10303	$C_{13}H_{13}ON_2^-$	166.08607	$C_9H_{12}O_2N^+$
185.07115	$C_{11}H_9ON_2^-$	149.05959	$C_9H_9O_2^+$
164.07083	$C_9H_{10}O_2N^-$	131.04907	$C_9H_7O^+$
147.04419	$C_9H_7O_2^-$	130.04979	$C_5H_8O_3N^+$
128.03415	$C_5H_6O_3N^-$	120.08081	$C_8H_{10}N^+$
127.05016	$C_5H_7O_2N_2^-$	103.05444	$C_8H_7^+$
109.03947	$C_5H_5ON_2^-$	84.04481	$C_4H_6ON^+$
91.05392	$C_7H_7^-$	79.05463	$C_6H_7^+$
72.00765	$C_2H_2O_2N^-$	56.05019	$C_3H_6N^+$
58.02844	$C_2H_4ON^-$		

Observed fragments for putative gamma glutamyl valine

245.11452	$C_{10}H_{17}O_5N_2^-$	247.12871	$C_{10}H_{19}O_5N_2^+$
227.10384	$C_{10}H_{15}O_4N_2^-$	230.10223	$C_{10}H_{16}O_5N^+$
209.09294	$C_{10}H_{13}O_3N_2^-$	184.09676	$C_9H_{14}O_3N^+$
183.11319	$C_9H_{15}O_2N_2^-$	156.10188	$C_8H_{14}O_2N^+$
165.10246	$C_9H_{13}ON_2^-$	138.09128	$C_8H_{12}ON^+$
128.03421	$C_5H_6O_3N^-$	130.04987	$C_5H_8O_3N^+$
116.07057	$C_5H_{12}O_2N^-$	118.08640	$C_5H_{12}O_2N^+$
112.03935	$C_5H_6O_2N^-$	102.05520	$C_4H_8O_2N^+$
84.04417	$C_4H_6ON^-$	84.04485	$C_4H_6ON^+$
82.02853	$C_4H_4ON^-$	72.08134	$C_4H_{10}N^+$
74.02339	$C_2H_4O_2N^-$	56.05020	$C_3H_6N^+$
		55.05496	$C_4H_7^+$

Observed fragments for glutamate

146.04500	$C_5H_8O_4N^-$	148.06035	$C_5H_{10}O_4N^+$
128.03426	$C_5H_6O_3N^-$	130.04985	$C_5H_8O_3N^+$
102.05477	$C_4H_8O_2N^-$	102.05522	$C_4H_8O_2N^+$
100.03908	$C_4H_6O_2N^-$	85.02882	$C_4H_5O_2^+$
85.02807	$C_4H_5O_2^-$	84.04482	$C_4H_6ON^+$
74.02336	$C_2H_4O_2N^-$	56.05020	$C_3H_6N^+$

Table D2. Numbers of features for all comparisons in this study

In Mit mutants worm pellet samples

Number of features	ESI negative	ESI positive

After high mass defect filter	90,465	95,154
Where less than half of total samples have intensity = 0	62,112	57,627
2 x UP in long-lived, DOWN in short-lived Mit mutants	134 (NONE after manual verification of raw files)	111 (NONE after manual verification of raw files)
2 x DOWN in long-lived, UP in short-lived Mit mutants	272 (11 after manual verification of raw files)	89 (NONE after manual verification of raw files)

In Mit mutants worm media samples

Number of features	**ESI negative**	**ESI positive**
After high mass defect filter	97,992	186,525
Where less than half of total samples have intensity = 0	85,406	175,357
2 x UP in long-lived, DOWN in short-lived Mit mutants	4 (NONE after manual verification of raw files)	1 (NONE after manual verification of raw files)
2 x DOWN in long-lived, UP in short-lived Mit mutants	24 (10 after manual verification of raw files)	42 (TBD after manual verification of raw files)

In *mito* and *sup* samples

Number of features	**ESI negative**	**ESI positive**
After high mass defect filter	55,599	74,267
Present in all *mito* replicates	16,778	22,782
2 x UP in *mito* over *sup*	75	179 (to be verified in Cytoscape)
2 x DOWN in *mito* over *sup*	187	81 (to be verified in Cytoscape)

Table D3. Lists of Differential Features

DOWN in long-lived, UP in short-lived Mit mutant worm pellets (ESI-)

m/z	rt (min)	Description
359.0209	3.64	phosphohexose derivative
378.05945	3.94	angl#2
430.15762	0.74	very small peak
457.05031	0.78	UMP-derivative
459.10606	6.79	very small peak
460.10152	5.94	phosphohexose derivative
460.10152	6.61	phosphohexose derivative
461.08594	5.22	phosphohexose derivative
462.11743	6.16	phosphohexose derivative
463.13606	6.94	phosphohexose derivative
502.05927	0.79	very small peak

DOWN in long-lived, UP in short-lived Mit mutant worm media (ESI-)

m/z	rt (min)	Description
134.0467	4.31	Fragment of tade#2
376.2092	4.1	$C_{16}H_{30}O_7N_3^-$
356.1289	4.28	$C_{15}H_{22}O_5N_3S^-$
551.1856	4.68	$C_{21}H_{35}O_9N_4S_2^-$, 2 isomers
631.2156	4.95	formate adduct of 585.20996, $C_{19}H_{42}O_{14}N_2PS^-$
244.0651	4.98	$C_{10}H_{14}O_4NS^-$, 3 isomers
516.0974	5.57	very small peak
401.0884	5.85	formate adduct of 355.08270 which ionizes poorly
355.0465	3.96	very small peak
370.1621	3.96	very small peak

Networking parameters and R codes

MS/MS Networking workflow

A molecular network was created using the online workflow at GNPS. The data was then clustered with MS-Cluster with a parent mass tolerance of 0.0075 Da and a MS/MS fragment ion tolerance of 0.0075 Da to create consensus spectra. Further, concensus spectra that contained less than 2 spectra were discarded. A network was then created where edges were filtered to have a cosine score above 0.4 and more than 2 matched peaks. Further edges between two nodes were kept in the network if and only if each of the nodes appeared in each other's respective top 20 most similar nodes. The spectra in the network were then searched against GNPS' spectral libraries. All matches kept between network spectra and library spectra were required to have a score above 0.5 and at least 2 matched peaks.

Parameter	Value	Parameter	Value
PAIRS_MIN_COSINE	0.4	FILTER_STDDEV_PEAK_INT	0
ANALOG_SEARCH	0	RUN_MSCLUSTER	on
tolerance.PM_tolerance	0.0075	FILTER_PRECURSOR_WINDOW	0
tolerance.Ion_tolerance	0.0075	FILTER_LIBRARY	0
MIN_MATCHED_PEAKS	2	WINDOW_FILTER	0
TOPK	20	SCORE_THRESHOLD	0.5
CLUSTER_MIN_SIZE	2	MIN_MATCHED_PEAKS_SEARCH	2
MAXIMUM_COMPONENT_SIZE	100	MAX_SHIFT_MASS	100
MIN_PEAK_INT	0		

List of features from MS/MS network (Figure 4.7) obtained from Cytoscape 3.5.1

m/z	rt (min)	Enriched in	m/z	rt (min)	Enriched in
131.043	0.783704	*mito*	288.12	1.30193	*mito*
134.046	4.629285	*mito*	293.113	3.921074	*mito*
137.02	1.244651	*mito*	297.067	4.030192	*mito*
146.044	0.758701	*mito*	300.12	2.536802	*mito*
150.04	1.284273	*mito*	302.136	3.628083	*mito*
165.05	7.000766	*mito*	350.07	2.943417	*mito*
166.039	1.333623	*mito*	397.19	6.715742	*mito*
171.1	7.234689	*mito*	431.184	9.381025	*mito*
174.04	1.178044	*mito*	438.262	9.602024	*mito*
174.08	1.937778	*mito*	441.126	3.093317	*mito*
175.06	4.288654	*mito*	450.262	9.637756	*mito*
187.109	2.886546	*mito*	452.28	10.33251	*mito*
188.092	3.81023	*mito*	455.098	4.2049	*mito*
190.053	4.142957	*mito*	464.279	9.676986	*mito*
201.124	3.188837	*mito*	466.29	10.08392	*mito*
206.08	5.346167	*mito*	466.294	10.86463	*mito*
208.07	4.265883	*mito*	469.26	9.745386	*mito*
211.072	3.268017	*mito*	476.279	9.218324	*mito*
221.093	3.79345	*mito*	477.107	6.552425	*mito*
235.048	1.643072	*mito*	478.294	9.848335	*mito*
241.08	5.024058	*mito*	478.295	10.543	*mito*
241.082	2.531733	*mito*	480.31	11.63501	*mito*
243.17	4.347214	*mito*	488.162	0.759453	*mito*
245.04	0.645457	*mito*	488.28	9.938429	*mito*
245.09	5.550017	*mito*	490.29	10.47274	*mito*
245.113	1.849744	*mito*	492.31	10.98777	*mito*
251.104	3.717717	*mito*	498.262	9.152276	*mito*
257.029	1.654947	*mito*	502.293	10.24537	*mito*
259.13	3.135492	*mito*	526.315	9.790274	*mito*
261.12	4.448539	*mito*	532.267	10.13999	*mito*
263.14	4.221558	*mito*	536.25	10.64206	*mito*
266.09	3.012917	*mito*	562.316	9.578635	*mito*
274.094	3.2976	*mito*	564.331	10.17511	*mito*
277.087	3.178163	*mito*	586.316	9.602993	*mito*
277.158	4.853706	*mito*	588.331	10.05334	*mito*
279.01	1.63121	*mito*	590.347	10.57314	*mito*
287.089	3.302458	*mito*	784.151	3.926708	*mito*
288.12	2.704467	*mito*			

m/z	rt (min)	Enriched in	m/z	rt (min)	Enriched in
103.04	1.706775	*sup*	145.086	6.03205	*sup*
129.02	19.73108	*sup*	147.029	1.370986	*sup*
129.02	19.83767	*sup*	147.03	19.68607	*sup*
129.02	19.73467	*sup*	147.03	4.807012	*sup*
129.02	19.39646	*sup*	147.07	4.895111	*sup*
129.02	15.94229	*sup*	158.081	1.362362	*sup*
129.02	15.48403	*sup*	158.081	1.876335	*sup*
129.02	12.42074	*sup*	158.081	0.694185	*sup*
129.02	19.47917	*sup*	160.06	1.34529	*sup*
129.02	19.92367	*sup*	160.06	0.706973	*sup*
129.02	0.084605	*sup*	161.04	10.11387	*sup*
129.02	5.127789	*sup*	161.04	16.60939	*sup*
129.02	6.69257	*sup*	165.05	12.89811	*sup*
129.02	4.440682	*sup*	165.06	12.85602	*sup*
129.02	2.911897	*sup*	167.02	1.047048	*sup*
129.02	7.538944	*sup*	171.028	2.069026	*sup*
135.03	1.215197	*sup*	181.07	0.572514	*sup*
136.04	5.073421	*sup*	181.071	11.13603	*sup*
137.02	0.452164	*sup*	188.06	1.833466	*sup*
137.02	19.922	*sup*	189.04	19.56131	*sup*
137.02	19.8465	*sup*	191.02	1.190284	*sup*
137.02	19.75819	*sup*	192.068	5.818917	*sup*
137.02	19.57372	*sup*	194.081	7.343007	*sup*
137.02	7.769817	*sup*	199.13	8.864642	*sup*
137.02	7.872542	*sup*	199.13	5.363894	*sup*
137.02	7.362733	*sup*	199.13	6.874188	*sup*
137.02	7.323096	*sup*	199.13	6.761842	*sup*
137.02	6.8586	*sup*	202.109	4.475342	*sup*
137.02	5.681456	*sup*	203.082	3.853242	*sup*
137.02	6.705024	*sup*	215.13	5.764242	*sup*
137.02	2.477928	*sup*	215.13	19.71027	*sup*
137.02	3.818879	*sup*	215.13	7.744713	*sup*
137.02	5.70405	*sup*	215.13	6.280675	*sup*
137.02	3.420077	*sup*	215.13	5.571275	*sup*
137.021	6.336455	*sup*	215.13	5.056178	*sup*
137.021	6.569744	*sup*	215.13	1.713232	*sup*
137.021	6.41484	*sup*	215.13	1.897269	*sup*
137.021	6.085739	*sup*	215.13	6.763952	*sup*
137.021	5.854242	*sup*	215.13	7.125267	*sup*
137.022	6.261692	*sup*	217.12	2.60831	*sup*
144.066	0.840336	*sup*	217.12	1.773303	*sup*

m/z	rt (min)	Enriched in	m/z	rt (min)	Enriched in
218.103	3.547167	*sup*	282.085	2.929187	*sup*
220.099	5.88515	*sup*	291.15	6.13912	*sup*
221.06	0.594645	*sup*	293.151	4.34975	*sup*
221.09	6.318978	*sup*	300.13	0.634233	*sup*
223.1	7.429453	*sup*	301.168	6.005233	*sup*
225.11	6.997513	*sup*	301.168	5.931917	*sup*
227.14	3.768617	*sup*	302.101	0.641814	*sup*
233.103	3.653492	*sup*	303.182	6.207889	*sup*
236.09	4.843853	*sup*	309.11	4.881642	*sup*
237.09	13.17417	*sup*	309.11	2.56243	*sup*
237.091	1.906023	*sup*	309.113	3.147044	*sup*
237.11	7.534163	*sup*	313.1	4.228108	*sup*
241.01	4.659158	*sup*	315.25	10.12735	*sup*
241.01	13.26374	*sup*	321.05	1.876382	*sup*
241.01	4.509562	*sup*	323.029	0.94649	*sup*
241.01	19.53011	*sup*	331.211	7.291528	*sup*
241.01	19.32442	*sup*	341.064	3.925489	*sup*
241.01	0.563816	*sup*	342.131	3.569022	*sup*
241.011	0.591508	*sup*	352.064	2.906902	*sup*
243.16	7.351971	*sup*	358.07	4.803085	*sup*
245.04	19.77717	*sup*	361.089	4.172625	*sup*
245.04	16.01959	*sup*	361.09	1.844039	*sup*
245.11	3.666675	*sup*	378.06	3.944008	*sup*
247.119	4.087983	*sup*	379.18	14.12238	*sup*
247.12	4.919717	*sup*	379.18	6.878417	*sup*
249.06	3.430067	*sup*	387.11	12.80428	*sup*
253.08	1.411702	*sup*	387.115	1.979141	*sup*
259.089	7.737308	*sup*	387.12	13.18515	*sup*
261.13	5.566967	*sup*	395.075	4.306825	*sup*
261.134	4.562533	*sup*	395.19	14.65424	*sup*
266.1	6.410642	*sup*	426.11	3.694725	*sup*
269.211	11.53366	*sup*	453.198	4.326492	*sup*
273.135	4.897322	*sup*	455.102	1.818582	*sup*
275.15	5.48385	*sup*	463.092	5.350325	*sup*
276.992	15.30276	*sup*	478.29	12.38351	*sup*
277.07	0.643424	*sup*	478.331	11.93668	*sup*
277.14	10.55347	*sup*	495.184	2.806489	*sup*
279.009	7.432503	*sup*	497.134	4.051575	*sup*
279.01	13.8506	*sup*	498.262	19.38333	*sup*
279.06	5.719717	*sup*	500.28	9.3357	*sup*
279.06	1.283275	*sup*	502.3	9.695925	*sup*

m/z	rt (min)	Enriched in	m/z	rt (min)	Enriched in
513.128	3.975783	*sup*	610.301	8.492857	*sup*
525.128	3.9851	*sup*	617.274	9.601374	*sup*
551.236	5.620617	*sup*	628.311	8.329483	*sup*
552.331	10.27476	*sup*	633.306	7.769379	*sup*
564.33	11.68212	*sup*	634.301	8.256764	*sup*
564.33	14.77455	*sup*	660.316	9.099329	*sup*
566.347	10.96183	*sup*	667.188	4.174672	*sup*
580.362	11.52325	*sup*	693.204	5.173256	*sup*
584.284	8.01957	*sup*	719.202	0.786149	*sup*
597.305	11.40258	*sup*	743.231	0.786202	*sup*
598.301	8.471847	*sup*	792.272	5.262692	*sup*
598.31	11.49167	*sup*			

R scripts used for untargeted metabolomics analysis.

Source code (Developed by Dr. Maximilian Helf)

```
install.packages(c("graphics", "Hmisc", "MASS", "plot3D", "heatmaply",
                   "pvclust","data.table", "DT", "shiny", "rhandsontable", "TeachingDemos",
"jsonlite","rcdk"))
source("https://bioconductor.org/biocLite.R")
biocLite()
biocLite("xcms")
biocLite(c("CAMERA","BiocParallel"))
library(xcms)
library(graphics)
library(Hmisc)
library(zoo)
library(peakPick)
library(tcltk)
library(MASS)
library(plot3D)
library(heatmaply)
library(pvclust)
library(data.table)
library(DT)
library(shiny)
library(rhandsontable)
library(plotly)
library(TeachingDemos) # for non-overlapping label placement
library(jsonlite)
library(BiocParallel)
library(rcdk) #for molecule drawing
EICfiles<-function(binvar1 = c("all"),
                   folders = "",
                   type="mzXML"){
  if(folders==""){
    fileList=list()
    for(i in c(1:length(binvar1))){
      fileList[[i]] <- choose.files(default= if(i>1){dirname(fileList[[i-1]][1])}else{""},
caption=paste("Add files to group", binvar1[i]), multi=T)
      repeat{
        if(winDialog(type="yesno", "Do you want to add more files to this group?")=="YES")
        {fileList[[i]] <- c(fileList[[i]], choose.files(default=
dirname(fileList[[i]][1]),caption=paste("Add more files to group", binvar1[i]),
multi=T))}else{break}}}
    bin1_p <- fileList
  }else{
      mzxml_p <- list.files(folders[1], pattern=type, recursive = TRUE, include.dirs=T,
full.names=T)
    if(length(folders)>1){for(i in c(2:length(folders))){
      mzxml_p <- c(mzxml_p,list.files(folders[i], pattern=type, recursive = TRUE,
include.dirs=T, full.names=T))
    }}
    bin1_p <- list()
    for(i in c(1:length(binvar1))){
      bin1_p[[i]] <- mzxml_p[grepl(binvar1[i], mzxml_p, ignore.case = T)]
    }
  }
  return(bin1_p)
}
EICraw <- function (bin1_p, MSn = T){
  rawlist=list()
```

```
    rawcoll=list()
  for(t in c(1:length(bin1_p))){
    for (numb in c(1:length(bin1_p[[t]]))){
      rawlist[[numb]] <- xcmsRaw(bin1_p[[t]][numb], profstep=0, includeMSn = MSn)
      cat(basename(bin1_p[[t]][numb]), " ")}
    rawcoll[[t]] <-rawlist
    rawlist=list()}
  names(rawcoll)<-names(bin1_p)
  return(rawcoll)}
EICrawP <- function (bin1_p, MSn = T, workers=1){
        rawcoll=list()
        if (length(unlist(bin1_p))<=10){workers<-1}
        param <- SnowParam(workers = workers)
        rawcoll <- bplapply(unlist(bin1_p),xcmsRaw,  profstep=0, includeMSn = MSn, BPPARAM=
param)
        #cat(names(bin1_p)[t], " ")
        rawlist1 <- list()
        rawlist2 <- list()
        cl <- vector()
        for (f in rawcoll){
            cl <- c(cl,f@filepath[1])
        }
        for (i in c(1:length(bin1_p))){
            rawlist1[[i]] <- rawcoll[cl %in% bin1_p[[i]]]}
        names(rawlist1)<-names(bin1_p)
        return(rawlist1)}
loadfeats <- function (seper ="\t"){
  fileNamex <<- choose.files(caption="Select table with m/z and rt values")
  #reading the table, making column names into callable variables
  if(identical(fileNamex,character(0))){alli <- NULL}else{
  alli <- read.table(fileNamex, header= T, sep= seper, dec=".", fill= T, skip=0)
  if(is.null(alli$rt)){alli$rt <- alli$rtmed}
  if(is.null(alli$mz)){alli$mz <- alli$mzmed}
  }
  return(alli)}
EICplot <- function(bin1_p, rawcoll, binvar1 = c("all"), alli = NULL, ppr = 10, rtw=90,
                    grn= 2,  swABS=F, hi=5, wi=3.5, labex = F, rtline=T, evalmode=F,
                    pdfout=T, fileNamex="./defaultoutput",
                    deffile = F, #use if the filename should be used exactly as put in in
fileNamex, will also suppress automatic pdf opening (for use in mosaic)
                    cx = 1,
                    rtmin=NULL,
                    rtmax=NULL){
  if(!is.null(alli)){
     mzr  <- data.frame(alli$mz-0.000001*ppr*alli$mz,alli$mz+0.000001*ppr*alli$mz)
    if(labex){mzriso <- data.frame(alli$mziso-
0.000001*ppr*alli$mziso,alli$mziso+0.000001*ppr*alli$mziso)}
    if(swABS){mzr  <- data.frame(alli$mzmin,alli$mzmax)}
    #set the RT window for EICs
    if (is.null(alli$rt) & (is.null(rtmin) | is.null(rtmax)) ){rtr <-
data.frame(0,max(sapply(sapply(unlist(bin1_p),slot,"scantime"),max)))}
    else
    {if (!is.null(rtmin) & !is.null(rtmax)){rtr <- data.frame(rtmin,rtmax)}
        else
        {rtr <- data.frame(alli$rt-rtw,alli$rt+rtw)}
            }
    }
    if(evalmode==T) {cat("Enter evaluation in console window. Presets: 1= Good, 2= OK, 3=Bad,
4= Unknown. Enter any value or string and press Enter. \n")}
   if(deffile){
    nakedFile <- sub("^([^.]*).*", "\\1",fileNamex)
```

```
    pdfFile <- fileNamex
  }else{
  if(is.null(alli)){
    nakedFile <- paste0(dirname(bin1_p[[1]][1]),"/TICs_only")
    pdfFile <- paste0(nakedFile,".pdf")
  }else{
    nakedFile <- sub("^([^.]*).*", "\\1",fileNamex)
    pdfFile <- paste0(nakedFile,"_EICv8_output.pdf")
  }
  if(file.exists(pdfFile)){pdfFile <-paste0(nakedFile,"_",round(runif(1,min=0.1,max=1)*10000,
digits=0),"_EICv8_output.pdf")}  #{file.remove(pdfFile)}
  }
    nr <- if(length(bin1_p)>=grn){ceiling((length(bin1_p))/grn)}else{1}
  Fighi <- hi*nr+2
  Figwi <- wi*grn
 if(evalmode==F & pdfout==T) {pdf(pdfFile,Figwi,Fighi)}
  par(mfrow=c(nr,grn), oma=c(0,0,10,0),xpd=NA, bg=NA)
    eiclist=list()
  maxlist=list()
  eiccoll=list()
  maxcoll=list()
  for(t in c(1:length(bin1_p))){
    for (numb in c(1:length(bin1_p[[t]]))){
      eiclist[[numb]] <- rawEIC(rawcoll[[t]][[numb]],mzrange =c(0,5000),rtrange =
c(0,max(rawcoll[[t]][[numb]]@scantime)))
      maxlist[[numb]] <- max(eiclist[[numb]]$intensity)
      maxlist[[numb]]$loop <- numb}
    eiccoll[[t]] <- eiclist
    eiclist=list()
    maxcoll[[t]] <- maxlist
    maxcoll[[t]]$cloop <- t
    maxlist=list()}
    maxsi=list()
  for(t in c(1:length(bin1_p))){
    for (numb in c(1:length(bin1_p[[t]]))){
      maxsi[[t]] <- max(unlist(maxcoll[[t]]))#,maxcoll[[t]][[numb]][[1]]))
    }}
    for(t in c(1:length(bin1_p))){
    RTwr <- c(min(rawcoll[[t]][[1]]@scantime)/60,max(rawcoll[[t]][[1]]@scantime)/60)
    RT1 <- (eiccoll[[t]][[1]]$scan-eiccoll[[t]][[1]]$scan+RTwr[[1]]+(eiccoll[[t]][[1]]$scan-
min(eiccoll[[t]][[1]]$scan))*((RTwr[[2]]-RTwr[[1]])/(max(eiccoll[[t]][[1]]$scan)-
min(eiccoll[[t]][[1]]$scan))))
    options(scipen=-20)
    plot(RT1,eiccoll[[t]][[1]]$intensity, type= "n", xlim=
c(0,max(rawcoll[[t]][[1]]@scantime)/60), ylim =c(0,maxsi[[t]]), axes=F,
ylab="Intensity",xlab="")
    axis(side=2, lwd=1, las=2, mgp=c(0.7,0.6,0))
    options(scipen=20)
    axis(side=1, lwd=1, minor.tick(nx=10,ny=5, tick.ratio=0.5), mgp=c(1,0.4,0))#x-axis mgp[2]
controls distance of tick labels to axis
    mtext(side=1, text= "RT (min)", line=1.2, cex=cx*0.7)
    title(main=binvar1[t], line=4+0.1*length(eiccoll[[t]]))
    colr <- rainbow(length(eiccoll[[t]]), s = 1, v = 1, start = 0, end = max(1,
length(eiccoll[[t]]) - 1)/length(eiccoll[[t]]), alpha = 0.7)#topo.colors(length(eiccoll[[t]]),
alpha=1)
    for (numb in c(1:length(eiccoll[[t]]))){
      RTa <- (eiccoll[[t]][[numb]]$scan-
eiccoll[[t]][[numb]]$scan+RTwr[[1]]+(eiccoll[[t]][[numb]]$scan-
min(eiccoll[[t]][[numb]]$scan))*((RTwr[[2]]-RTwr[[1]])/(max(eiccoll[[t]][[numb]]$scan)-
min(eiccoll[[t]][[numb]]$scan))))
      lines(RTa,eiccoll[[t]][[numb]]$intensity, col=colr[numb], lwd=1)}
```

```
    legend("topright", inset=c(0,-0.025*length(eiccoll[[t]])),basename(bin1_p[[t]]),
lty=1,lwd=2.5, col=colr, bty="n",  cex=cx*0.5)}
  mtext("TICs", side=3, outer=T,line=7.2, cex=cx*1)
  if(is.null(alli)){
    dev.off()
    if (pdfout==T){shell.exec(pdfFile)}}
  eiclist=list()
  maxlist=list()
  eiccoll=list()
  maxcoll=list()
  if(labex){
  eiclistiso=list()
  maxlistiso=list()
  eiccolliso=list()
  maxcolliso=list()}
  for(m in c(1:NROW(mzr))){
  RTwr <- if(nrow(rtr)==nrow(mzr)){rtr[m,]}else{rtr[1,]}
    for(t in c(1:length(bin1_p))){
      for (numb in c(1:length(bin1_p[[t]]))){
        eiclist[[numb]] <- rawEIC(rawcoll[[t]][[numb]],mzrange = mzr[m,],rtrange = RTwr)
        maxlist[[numb]] <- max(eiclist[[numb]]$intensity)
        maxlist[[numb]]$loop <- numb}
      eiccoll[[t]] <- eiclist
      eiclist=list()
      maxcoll[[t]] <- maxlist
      maxcoll[[t]]$cloop <- t
      maxlist=list()}
        if(labex){
    for(t in c(1:length(bin1_p))){
      for (numb in c(1:length(bin1_p[[t]]))){
        eiclistiso[[numb]] <- rawEIC(rawcoll[[t]][[numb]],mzrange = mzriso[m,],rtrange = RTwr)
        maxlistiso[[numb]] <- max(eiclistiso[[numb]]$intensity)
        maxlistiso[[numb]]$loop <- numb}
      eiccolliso[[t]] <- eiclistiso
      eiclistiso=list()
      maxcolliso[[t]] <- maxlistiso
      maxcolliso[[t]]$cloop <- t
      maxlistiso=list()}  }
    maxsi=list()
    for(t in c(1:length(bin1_p))){
      for (numb in c(1:length(bin1_p[[t]]))){
        maxsi[[t]] <- max(unlist(maxcoll[[t]]))#,maxcoll[[t]][[numb]][[1]]))
      }}
    if(labex){
    maxsiiso=list()
    for(t in c(1:length(bin1_p))){
      for (numb in c(1:length(bin1_p[[t]]))){
        maxsiiso[[t]] <- max(unlist(maxcolliso[[t]]))#,maxcoll[[t]][[numb]][[1]]))
      }}}
    par(mfrow=c(nr,grn), oma=c(0,0,10,0),xpd=NA, bg=NA)
    for(t in c(1:length(bin1_p))){
     RT1 <- (eiccoll[[t]][[1]]$scan-eiccoll[[t]][[1]]$scan+RTwr[[1]]+(eiccoll[[t]][[1]]$scan-
min(eiccoll[[t]][[1]]$scan))*((RTwr[[2]]-RTwr[[1]])/(max(eiccoll[[t]][[1]]$scan)-
min(eiccoll[[t]][[1]]$scan))))/60
      options(scipen=-20)
      plot(RT1,eiccoll[[t]][[1]]$intensity, type= "n",
           ylim =if (labex){c(0,max(c(maxsi[[t]],maxsiiso[[t]])))}else{c(0,maxsi[[t]])},
axes=F, ylab="Intensity",xlab="")
      axis(side=2, lwd=1, las=2, mgp=c(0.7,0.6,0))
      options(scipen=20)
      axis(side=1, lwd=1, minor.tick(nx=10,ny=5, tick.ratio=0.5), mgp=c(1,0.4,0))#x-axis
```

```
mgp[2] controls distance of tick labels to axis
      mtext(side=1, text= "RT (min)", line=1.2, cex=cx*0.7)
      title(main=binvar1[t], line=4+0.1*length(eiccoll[[t]]))
      colr <- topo.colors(length(eiccoll[[t]])+if(labex){length(eiccolliso[[t]])}else{0},
alpha=1)#rainbow(length(eiccoll[[t]]), s = 1, v = 1, start = 0, end = max(1,
length(eiccoll[[t]]) - 1)/length(eiccoll[[t]]), alpha = 0.7)#topo.colors(length(eiccoll[[t]]),
alpha=1)
      for (numb in c(1:length(eiccoll[[t]]))){
        RTa <- (eiccoll[[t]][[numb]]$scan-
eiccoll[[t]][[numb]]$scan+RTwr[[1]]+(eiccoll[[t]][[numb]]$scan-
min(eiccoll[[t]][[numb]]$scan))*((RTwr[[2]]-RTwr[[1]])/(max(eiccoll[[t]][[numb]]$scan)-
min(eiccoll[[t]][[numb]]$scan))))/60

        lines(RTa,eiccoll[[t]][[numb]]$intensity, col=colr[numb], lwd=1)
        }
      if(labex){
       for (numb in c(1:length(eiccolliso[[t]]))){
        RTa <- (eiccolliso[[t]][[numb]]$scan-
eiccolliso[[t]][[numb]]$scan+RTwr[[1]]+(eiccolliso[[t]][[numb]]$scan-
min(eiccolliso[[t]][[numb]]$scan))*((RTwr[[2]]-RTwr[[1]])/(max(eiccolliso[[t]][[numb]]$scan)-
min(eiccolliso[[t]][[numb]]$scan))))/60
        lines(RTa,eiccolliso[[t]][[numb]]$intensity, col=colr[length(eiccoll[[t]])+numb],
lwd=1)}}
       legendtext <- if(labex){
        c(paste(sub("^([^.]*).*", "\\1",basename(bin1_p[[t]]))),
          paste0(sub("^([^.]*).*", "\\1",basename(bin1_p[[t]])),"+LABEL"))
        }else{paste(sub("^([^.]*).*", "\\1",basename(bin1_p[[t]])))}
      legend("topright", inset=c(0,-0.025*length(eiccoll[[t]])),legendtext, lty=1,lwd=2.5,
col=colr, bty="n",  cex=cx*0.5)
      if(rtline){segments(alli$rt[m]/60,0,alli$rt[m]/60,if
(labex){max(c(maxsi[[t]],maxsiiso[[t]]))}else{maxsi[[t]]},lty=2,lwd=1, col="black")}
      }
    if(labex){
      mtext(paste("m/z range:", round(alli$mz[m],5), "/", round(alli$mziso[m],5), "(labeled)
+/-", ppr," ppm @ RT",round(alli$rt[m],1),"sec /",round(alli$rt[m]/60,2), "min"), side=3,
outer=T,line=7.2, cex=cx*1)
    }else{
      mtext(paste("m/z range:", round(mzr[m,1],5),"-
",round(mzr[m,2],5),"(",round(alli$mz[m],5),"+/-", ppr, "ppm @ RT",round(alli$rt[m],1),"sec
/",round(alli$rt[m]/60,2), "min"), side=3, outer=T,line=7.2, cex=cx*1)
    }
      mtext(paste("Category:", alli$Sample[m], alli$Medium[m], alli$Extraction[m],
alli$Tool[m]), side=3, outer=T,line=6, cex=cx*1)
    mtext(paste("Comment:",alli$Comments[m]), side=3, outer=T,line=4.8, cex=cx*0.5)
    if(evalmode==T) {
    cat("Manual evaluation of ", paste0("m/z range:", round(alli$mz[m],5),
if(labex){paste0("/", round(alli$mziso[m],5), "(labeled)")}," +/-", ppr," ppm @
RT",round(alli$rt[m],1),"sec /",round(alli$rt[m]/60,2), "min"))
      doneVal <- readline(prompt = "Enter here > ")
      alli$Comments[m] <- doneVal
   #dev.off()
    }
    }
   if(evalmode==T) {return(alli)}
   if(pdfout){ dev.off()
  #open your output pdf file
     if(!deffile){
  shell.exec(pdfFile)}}
}
multiEIC <- function (rawfile= rawdata[[1]][[39]] ,
                          featuretable=alli,
```

```
                        mz = alli$mz,
                        ppm=5,
                        rtw= data.frame(alli$rtmin-5,alli$rtmax+5),
                        mini=5000,
                        coltag="test",
                        pval=0.05,
                        gauss=T,
                        XIC=T,
                        abovex=T){
    mx <- matrix(data= c(mz-ppm*(mz/1000000),
                         mz+ppm*(mz/1000000),
                         rowMin(as.matrix(rtw)),
                         rowMax(as.matrix(rtw))), nrow= length(mz), ncol=4)
    mxl <-unname(as.list(data.frame(t(mx[,1:2]))))
  rxl <-unname(as.list(data.frame(t(mx[,3:4]))))
 summe <- mapply(rawEIC, mzrange = mxl,
                  rtrange = rxl, MoreArgs=list(object=rawfile), SIMPLIFY = F)
  fx <- function(x) x$intensity-min(x$intensity)
  summe <- lapply(summe, fx )
  if(XIC){
    import <- sapply(summe, mean)
    featuretable$importxx<-import
    colnames(featuretable)[which(names(featuretable) == "importxx")] <-
paste0(sub("^([^.]*).*", "\\1",basename(coltag)),"_XIC")
  }
  if(abovex){
    #how many consecutive intensity values are above mini (note this is after baseline
correction!)
    fx <- function(x) max(rle(x>mini)$lengths[which(rle(x>mini)$values ==T)])
    abovemini <- sapply(summe, fx)
    abovemini[abovemini==-Inf] <-0
    #how many consecutive intensity values are above average of values in rt window (note this
is after baseline correction!)
    fy <- function(x) max(rle(x>mean(x))$lengths[which(rle(x>mean(x))$values ==T)])
    abovemean <- sapply(summe, fy)
    abovemean[abovemean==-Inf] <-0
     featuretable$aboveminixx <- abovemini
    featuretable$abovemeanxx <- abovemean
    colnames(featuretable)[which(names(featuretable) == "aboveminixx")] <-
paste0(sub("^([^.]*).*", "\\1",basename(coltag)),"_Abovemini")
    colnames(featuretable)[which(names(featuretable) == "abovemeanxx")] <-
paste0(sub("^([^.]*).*", "\\1",basename(coltag)),"_Abovemean")
  }
  if(gauss){
    fx <- function(x) if(max(x)>0){x/max(x)}else{x}
    summe <- sapply(summe, fx)
    fy <- function(sm){
      fit <- try(nls(y ~ SSgauss(x, mu, sigma, h), data.frame(x =
                                                              1:length(sm), y =
sm)),silent=T)
      if(class(fit) == "try-error")
      {
        0
      } else
      {
      if(length(which(!is.na(sm-fitted(fit)))) > 4 &&
           length(!is.na(unique(sm)))>4 && length(!is.na(unique(fitted(fit))))>4)
        {
          cor <- NULL
          options(show.error.messages = FALSE)
          cor <- try(cor.test(sm,fitted(fit),method="pearson",use="complete"))
```

```
          options(show.error.messages = TRUE)
          if (!is.null(cor))
          {
            if(cor$p.value <= pval) cor$estimate else 0
          } else 0
        } else 0
      }}
    gauss <- sapply(summe, fy)
     featuretable$gaussxx <-gauss
    colnames(featuretable)[which(names(featuretable) == "gaussxx")] <-
paste0(sub("^([^.]*).*", "\\1",basename(coltag)),"_Gauss")
    }
  return(featuretable)}
featlistCompare <- function(reflist=unlabpeaks,
                            complist=labpeaks,
                            mzdiff=2*1.00335,
                            pwi= reflist$rtmax-reflist$rtmin,
                            pktol = 2,#fold difference in peak width allowed
                            rtd=10,
                            ppm=5
){
   cat(paste0("Reference list with ",length(reflist[,1])," features, iterating through list,
feature #" ))
   collector <- vector()
  collectlist <- list()
  for (i in c(1:length(reflist[,1]))){
    peakwi <- pwi[i]
    selection1 <- complist[which(abs(reflist$mz[i]+mzdiff-
complist$mz)<reflist$mz[i]*ppm*0.000001
                                 & abs(reflist$rt[i]-complist$rt)<rtd
                                 & complist$rtmax-complist$rtmin < peakwi*pktol
                                 & complist$rtmax-complist$rtmin > peakwi/pktol
    ),]
    selection2 <- reflist[which(abs(reflist$mz[i]+mzdiff-
reflist$mz)<reflist$mz[i]*ppm*0.000001
                                & abs(reflist$rt[i]-reflist$rt)<rtd
                                # & reflist$rtmax-reflist$rtmin < peakwi*pktol
                                #  & reflist$rtmax-reflist$rtmin > peakwi/pktol
    ),]
  selection3 <- complist[which(abs(reflist$mz[i]-complist$mz)<reflist$mz[i]*ppm*0.000001
                                 & abs(reflist$rt[i]-complist$rt)<rtd
                                 #  & complist$rtmax-complist$rtmin < peakwi*pktol
                                 #  & complist$rtmax-complist$rtmin > peakwi/pktol
    ),]
       if(length(selection1[,1])>0){
      collectlist$iso1sam1[[length(collectlist$iso1sam1)+1]] <- reflist[i,]
      collectlist$iso2sam2[[length(collectlist$iso1sam1)]] <-
selection1[order(abs(selection1$rt-reflist$rt[i])),]
      collectlist$iso2sam1[[length(collectlist$iso1sam1)]] <-
selection2[order(abs(selection1$rt-reflist$rt[i])),]
      collectlist$iso1sam2[[length(collectlist$iso1sam1)]] <-
selection3[order(abs(selection1$rt-reflist$rt[i])),]
      }
    if(i %% 200==0){cat(i, " ")}}
  return(collectlist)}
mergecollectlist <- function(collectlist){
  #combining the lists into one file, using only the features as filtered with the smallest RT
difference (1 to 1 to 1 to 1 feature)
  combinat2 =vector()
  for (n in c(1:length(collectlist$iso1sam1))){
    collectlist$iso1sam1[[n]]$feat <- n
```

```
    collectlist$iso2sam2[[n]]$feat <- n
    collectlist$iso2sam1[[n]]$feat <- n
    collectlist$iso1sam2[[n]]$feat <- n
    colnames(collectlist$iso1sam1[[n]]) <- paste0("I1S1.",colnames(collectlist$iso1sam1[[n]]))
    colnames(collectlist$iso2sam2[[n]]) <- paste0("I2S2.",colnames(collectlist$iso2sam2[[n]]))
    colnames(collectlist$iso2sam1[[n]]) <- paste0("I2S1.",colnames(collectlist$iso2sam1[[n]]))
    colnames(collectlist$iso1sam2[[n]]) <- paste0("I1S2.",colnames(collectlist$iso1sam2[[n]]))
    combinat <-cbind(collectlist$iso1sam1[[n]][1,],collectlist$iso2sam2[[n]][1,],
                     collectlist$iso2sam1[[n]][1,],collectlist$iso1sam2[[n]][1,])
    if( length(combinat2)>0){
      combinat2 <-rbind(combinat2,combinat)}
    else{
      combinat2 <-combinat
    }
  }
  return(combinat2)}
redundfind <- function(pl = xset3.pos.pl,
                       mzppm = 3,
                       rtsec = 3){
   pl <- pl[order(pl$mz),]
  pl$diff <-c(diff(pl$mz),1000)
  plnc <- vector()
  plvc <- vector()
  counter<-1
  selector <- c(0,which(pl$diff > pl$mz*mzppm*1e-6))
  for (n in c(1:(length(selector)-1))){
    pls <-
data.frame(mz=pl$mz[(selector[n]+1):(selector[n+1])],rt=pl$rt[(selector[n]+1):(selector[n+1])]
) #shrinking version of pl
    plv <- vector()
    plv[1:nrow(pls)] <- 0
    pln <- vector()
    pln[1:nrow(pls)] <- 1
    for (i in c(1:nrow(pls))){#
    plv[which(abs(pls$mz-pls$mz[i])< pls$mz[i]*1e-6*mzppm
                & abs(pls$rt-pls$rt[i])< rtsec
        )]  <- counter
        pln[which(plv==counter)]  <- length(pln[which(plv==counter)])
        counter <- counter+1
      }
    }
    plvc<-c(plvc,plv)
    plnc<-c(plnc,pln)
    if (n%%1000==0){cat(selector[n], " ")}}
  pl$redundancy<-plnc
  pl$redgroup<-plvc
  return(pl)
}
redundfind2 <- function(pl,
                        coln = "redgroup",
                        ...
                       #mzppm = 3,
                       #rtsec = 3,
                       #shifts = c(Na= 21.98249281,
                        #          NH3= 17.0265491,
                         #         C13=1.003354838),
                       #charges = c(1,2),
                       #polarity = "positive"
    ){
    constants <- c(e=0.0005485799, proton=1.00727646681)
        pt2<-proc.time()
```

```
        groups <- integer(nrow(pl))
        counter <- 1
            selection <- apply(pl,1,
                               redundhelp,
                               pls = pl,
                               ...)
            for(n in selection){
            if (max(groups[n])>0){
                groups[n] <- min(groups[n][groups[n]!=0])
            }else{
                groups[n] <- counter
                counter <- counter + 1
            }
            }
    pt2<-proc.time()-pt2
    pl[[coln]] <- groups
    return(pl)}
  redundhelp <- function(     item = pl[1,],
                              pls = pl,
                              mzppm = 3,
                              rtsec = 3,
                              shifts = c(Na= 21.98249281,
                                         NH3= 17.0265491,
                                         C13=1.003354838,
                                         Noshift = 0 ),
                              charges = c(1,2),
                              polarity = "positive",
                              colnums = list(mz= which(colnames(pls)=="mz"),
                                             rt= which(colnames(pls)=="rt"))){
    constants <- c(e=0.0005485799, proton=1.00727646681, H=1.007825032)
    if(polarity=="positive"){
        uncharged = (as.numeric(item[colnums$mz])*charges)-charges*constants['proton']
        charged =
unique(as.vector(t(outer(uncharged,charges*constants['proton'],FUN="+")))/charges)
        shiftvars = unique(as.vector(outer(shifts,charges,"/")))
        combinat = unique(as.vector(outer(charged,shiftvars,"+")))
        }
    if(polarity=="negative"){
        uncharged = (as.numeric(item[colnums$mz])*charges)+charges*(constants['H']-
constants['e'])
        charged = unique(as.vector(t(outer(uncharged,charges*(constants['H']-
constants['e']),FUN="-")))/charges)
        shiftvars = unique(as.vector(outer(shifts,charges,"/")))
        combinat = unique(as.vector(outer(charged,shiftvars,"+")))
            }
    tb <- which(abs(pls$rt - as.numeric(item[colnums$rt])) < rtsec)
    if(length(tb)>0){
    return(tb[which(rowMin(t(abs(outer(combinat,pls$mz[tb],"-"))/combinat)) < mzppm*1e-6)])}
    else
    {return(integer(0))}}
redundrank <- function(pl,
                       groupingcolumn=pl$pcgroupown,
                        intcols = c(12:16),
                       colnls = list(rank="pcgrank",membernum = "nummembers")){
    pcgmembers <- integer(nrow(pl))
    pcgrank <- integer(nrow(pl))
    intens <- rowMeans(as.matrix(pl[,intcols]))
    for(i in c((min(groupingcolumn)):(max(groupingcolumn)))){
    selection <- which(groupingcolumn==i)
        pcgrank[selection] <- rank(intens[selection])
        pcgmembers[selection] <- length(selection)
```

```
    }
    pl[[colnls$rank]] <- pcgrank
    pl[[colnls$membernum]] <- pcgmembers
    return(pl)}
isofind <- function(pl = xset3.pos.pl,
                    mztol = 0.001,
                    rtsec = 3){
  pt1<-proc.time()
  #sort dataset by mz and see how large the mz gaps are between closest mz values
  pl <- pl[order(pl$mz),]
  pl$diff <-c(diff(pl$mz),0.1)
  pl$preselect1 <- cumsum(pl$diff)
  pl$preselect2 <-0
  for(z in c(1:floor(max(pl$preselect1)))){
    pl$preselect2[which.min(abs(z*1.00335-pl$preselect1))]<-1
  }
  plnc <- vector()
  plvc <- vector()
  counter<-1
  selector <- c(0,which(pl$preselect2==1,nrow(pl)))
  for (n in c(1:(length(selector)-2))){
   pls <-
data.frame(mz=pl$mz[(selector[n+1]+1):(selector[n+2])],rt=pl$rt[(selector[n+1]+1):(selector[n+
2])])
    srch <-
data.frame(mz=pl$mz[(selector[n]+1):(selector[n+2])],rt=pl$rt[(selector[n]+1):(selector[n+2])]
)
    plv <- vector()
    plv[1:nrow(pls)] <- 0
    pln <- vector()
    pln[1:nrow(pls)] <- 1
    for (i in c(1:nrow(pls))){
   temp    <- srch[which(srch$mz-pls$mz[i]>1.00335-mztol & srch$mz-pls$mz[i]<1.00335+mztol),]
      plv[i]  <-  if(sum(abs(temp$rt-pls$rt[i])<rtsec)>0){temp$mz[which.min(abs(temp$rt-
pls$rt[i])<rtsec)]}else{0}#sum(abs(temp$rt-pls$rt[i])<rtsec)
    }
    plvc<-c(plvc,plv)
    if (n%%10==0){cat(selector[n], " ")}}
    filler <-vector()
  filler[1:selector[2]] <-0
  pl$isocheck<-c(filler,plvc)
  pt1 <- proc.time()-pt1
  cat(pt1)
  return(pl)
}
sampid <- function(pl=filtrate1,
                   sfactor = 5,#fold change over max wiltype intensity value
                   wtc= c(10,11),# columns in pl which contain control sample intensities
                   scol= c(12:50), # colums in pl which contain test sample intensities
                   rem = "_XIC",
                   snames=NULL) # part of columnnames to be removed to get sample name
{
  samples <- vector()
  samples[1:nrow(pl)]<- ""
  nsamples <- vector()
  nsamples[1:nrow(pl)]<- 0
  cnames <- if (is.null(snames)) {colnames(pl)[scol]}else{snames}
  for (n in c(1:length(scol))){
    sc <- scol[n]
    rat <- pl[,sc]/rowMax(as.matrix(pl[,wtc]))
    sname <- gsub(rem,"",cnames[n])
```

```
    samples[which(rat>=sfactor)] <- paste(samples[which(rat>=sfactor)],sname)
    nsamples[which(rat>=sfactor)] <- nsamples[which(rat>=sfactor)]+1
  }
   samples <- gsub("[[:space:]]*$","", samples)
  samples <- gsub("^*[[:space:]]","", samples)
  #pl$maxfold <- rowMax(as.matrix(pl[,scol]))/rowMax(as.matrix(pl[,wtc]))
  pl$numsamples <- nsamples
  pl$upinsamples <- samples
  return(pl)}
densplot <-function(densin = log10(as.numeric(unlist(filtrate3[,scol]))),#filtrate3$rt,#input
values
                    perc = c(0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9, 0.95, 0.99),
                    ...
                    ){
   densin[densin==Inf] <- 1.1*max(densin[densin!=Inf])
  densin[densin==-Inf] <- 0.9*min(densin[densin!=-Inf])
  densin <- na.omit(densin)
  dens <- density(densin,from=min(densin),to=max(densin), cut=0, n=4096, na.rm = T)
  plot(dens, type= "l", ...)
  quan <- quantile(densin,perc)
  colr <- rainbow(length(perc), s = 1, v = 1, start = 0, end = max(1, length(perc) -
1)/length(perc), alpha = 0.5)
  segments(quan,min(dens$y),quan,max(dens$y), col=colr, lwd=0.8)
  legendtext <- paste0(perc,": ",round(quan,4) )
  legend("topright", inset=c(0,0.03*max(dens$y)),legendtext, lty=1,lwd=2.5, col=colr, bty="n",
cex=.5)
}
densplot2D <-function(densin = log10(rowvars),
                      densin2 = filtrate1$numsamples,#filtrate3$rt,#input values
                      quants = T, #print quantile lines
                      points = T, #print data points
                      perc = c(0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9, 0.95, 0.99),
                      ...){#arguments to be passed to image2D
    nas <- unique(c(which(is.na(densin)),which(is.na(densin2))))
  #densin <- log10(filtrate3$maxfold)
  if (length(nas)>0) {densin <- densin[-nas]}
  densin[densin==Inf] <- 1.1*max(densin[densin!=Inf])
  densin[densin==-Inf] <- 0.9*min(densin[densin!=-Inf])
  if (length(nas)>0) {densin2 <- densin2[-nas]}
  densin2[densin2==Inf] <- 1.1*max(densin2[densin2!=Inf])
  densin2[densin2==-Inf] <- 0.9*min(densin2[densin2!=-Inf])
  dens2d <- kde2d(densin,densin2, n = 512)
  image2D(dens2d, ...)
  if(points){points(densin,densin2, pch=3, cex = 0.5)}
  if(quants){
  dens <- density(densin,from=min(densin),to=max(densin), cut=0, n=4096, na.rm = T)
  dens2 <- density(densin2,from=min(densin2),to=max(densin2), cut=0, n=4096, na.rm = T)
  quan <- quantile(densin,perc)
  quan2 <- quantile(densin,perc)
 colr <- rainbow(length(perc), s = 1, v = 1, start = 0, end = max(1, length(perc) -
1)/length(perc), alpha = 0.5)
  par(xpd=F)
  segments(quan,min(dens2$x),quan,max(dens2$x), col=colr, lwd=0.8)
  segments(min(dens$x),quan2,max(dens$x),quan2, col=colr, lwd=0.8)
  legendtext <- paste0(perc,": ",round(quan,4),"/",round(quan2,4) )
  legend("topright", inset=c(0,0),legendtext, lty=1,lwd=2.5, col=colr, bty="n",  cex=.5)
  }
  }
gausscheck <- function(inp=allfilt,#an input dataframe with a column $upinsamples from the
sampid function
                       rawdata,
```

```
                      gauss=T,
                      abovex=F){ #a list object with xcmsraw objects, no list substructures
allowed
  filesel <- vector()
  rawdata <- unlist(rawdata, recursive=T)
  for (n in c(1:length(rawdata))){
    filesel[n] <-  sub("^([^.]*).*", "\\1",basename(rawdata[[n]]@filepath@.Data))}
  filesel <- gsub("-",".",filesel)
  if(gauss){inp$gauss <- " "}
  if(abovex){inp$abovemean <- " "}
  for(n in c(1:length(filesel))){
  subsel <- inp[which(inp$upinsamples %like% filesel[n]),]
  if(nrow(subsel)>0){
  microl <-    multiEIC(rawfile= rawdata[[n]] ,
                            featuretable=data.frame(numeric(nrow(subsel))),
                            mz = subsel$mz,
                            ppm=5,
                            rtw= data.frame(subsel$rtmin-5,subsel$rtmax+5),
                            mini=5000,
                            coltag="exp",
                            pval=0.05,
                            XIC=F,
                            abovex = abovex,
                            gauss= gauss)
    if(gauss){inp$gauss[which(inp$upinsamples %like% filesel[n])] <-
paste(inp$gauss[which(inp$upinsamples %like% filesel[n])],
round(microl$exp_Gauss,4),collapse=NULL)}
    if(abovex){inp$abovemean[which(inp$upinsamples %like% filesel[n])] <-
paste(inp$abovemean[which(inp$upinsamples %like% filesel[n])],
round(microl$exp_Abovemean,4),collapse=NULL)}
   if (n%%500==0){cat(n, " ")}}
    }
  inp$gauss <- trimws(inp$gauss)
  inp$abovemean <- trimws(inp$abovemean)
  if(gauss){
  inp$gauss <- trimws(inp$gauss)
  inp$gaussmax <- strmax(inp$gauss)}
  if(abovex){
  inp$abovemean <- trimws(inp$abovemean)
  inp$abovemeanmax <- strmax(inp$abovemean)}
  return(inp)
}
quanquot <- function(inp, #input dataframe
                     cols = c(1,2), #columns of dataframe to use for calculation
                     quanti = 0.75){ #define the quantile over which to calculate the fold
change
  quanx <-vector()
  for (n in c(1:nrow(inp))){
    quanx[n] <- (rowMax(as.matrix(inp[n,cols]))/(quantile(inp[n,cols],quanti)))
  }
  return(quanx)
}
rowVar <- function(inp, #input dataframe
                     cols = c(1,2) #columns of dataframe to use for calculation
                     ){
  quanx <-vector()
  for (n in c(1:nrow(inp))){
    quanx[n] <- var(as.numeric(inp[n,cols]))
  }
  return(quanx)
}
```

```
NormDataCols <- function(inp,
                     cols = c(1),
                     replace=T,
                     suffix="_norm"){
  if (replace){
  for (n in c(1:length(cols))){
    cl <-cols[n]
    inp[,cl] <- inp[,cl]/(mean(inp[,cl]))
    colnames(inp)[cl]<- paste0(colnames(inp)[cl],suffix)
  } }else{
    for (n in c(1:length(cols))){
      cl <-cols[n]
      inp$normxx <- inp[,cl]/(mean(inp[,cl]))
      colnames(inp)[which(names(inp)=="normxx")]<- paste0(colnames(inp)[cl],suffix)
    }}
  return(inp)
}
NormDataCols <- function(inp,
                         cols = c(1),
                         replace=T,
                         suffix="_norm"){
  if (replace){
    for (n in c(1:length(cols))){
      cl <-cols[n]
      inp[,cl] <- inp[,cl]/(mean(inp[,cl]))
      colnames(inp)[cl]<- paste0(colnames(inp)[cl],suffix)
    } }else{
      for (n in c(1:length(cols))){
        cl <-cols[n]
        inp$normxx <- inp[,cl]/(mean(inp[,cl]))
        colnames(inp)[which(names(inp)=="normxx")]<- paste0(colnames(inp)[cl],suffix)
      }}
  return(inp)
}
ttestx <- function(x=c(1,2,2,3,3,3,4,4,4,4,4,5,5,5,6,6,7) #input vector
                       , ltail=T,
                     over= NULL,
                     calc = NULL){
  if (is.null(over)){over <- c(1:length(x))}
  if (is.null(calc)){calc <- c(1:length(x))}
  m <- mean(x[over])
  s <-sd(x[over])
  n <- length(x[over])
  fx <- function(x) (m-x)/(s/sqrt(n))
  t <- sapply(x[calc], fx )

  fy <- function(x) pt(x, df=n-1, lower.tail = ltail)
  pvals <- sapply(t, fy)
  return(pvals)
}
heatmapx <- function(inp = filtrate2,
                     cols = c(10:50),
                     rel = T,
                     log = T,
                     ...){
  heatq <- as.matrix(inp[,cols])
  rownames(heatq) <- paste0(round(inp$mz,5),"_",(round(inp$rt,1)))
  if(rel){heatq <- heatq/(rowMeans(heatq))}
  if (log){heatq <- log10(heatq)}
  heatq[heatq==Inf] <- 1.1*max(heatq[heatq!=Inf])
  heatq[heatq==-Inf] <- 0.9*min(heatq[heatq!=-Inf])
```

```
  heatmaply(heatq,
            scale_fill_gradient_fun = ggplot2::scale_fill_gradient2(
              low = "blue", high = "red", midpoint = mean(heatq), limits = c(min(heatq),
max(heatq))), ...)
}
strmax <- function(x){
  fil<-  strsplit(x," ")
  tt<- as.numeric(sapply(fil, max))
return(tt)
}
kth <- function(x, k){
  k <- ncol(x)+1-k
  return(as.vector(unlist(sapply(split(unname(x), seq(nrow(x))),sort)[k,])))
}
mttest <- function (x,y){
  xl <- as.list(data.frame(t((x))))
  yl <- as.list(data.frame(t((y))))
  listo <- mapply(sttest,xl,
                  yl, SIMPLIFY = F)
  return(sapply(listo,"[[",3))}
msana1 <- function (pl,
                    groups = list(wtc,scol),
                    groupnames = c("N2", "Mut"),
                    norm=T,
                    ttest=T){
  if(norm){pl <- NormDataCols(pl,cols=unlist(groups),replace=T,suffix = "")}
  pl$maxfold <- rowMax(as.matrix(pl[,groups[[2]]]))/rowMax(as.matrix(pl[,groups[[1]]]))
  pl$group_up <- ""
  pl$group_up[which(pl$maxfold>1)] <- groupnames[2]
  pl$group_up[which(pl$maxfold<1)] <- groupnames[1]
  pl$maxfold[which(pl$maxfold<1)] <- 1/(pl$maxfold)[which(pl$maxfold<1)]
  pl$massdefppm <- ((pl$mz-floor(pl$mz))/pl$mz)*1e6
  if(ttest){pl$pval <- mttest(pl[,wtc],pl[,scol])}
  return(pl)
}
msampid <- function(pl,
                     groups = list(wtc,scol),
                     groupnames = c("N2", "Mut"),
                     ...){
pl2 <- sampid(pl=pl[which(pl$group_up==groupnames[1]),], #sampid checks the values of which
samples are above the control*sfactor
                       wtc= unlist(groups)[which(!unlist(groups) %in% groups[[1]])],# columns
in pl which contain control sample intensities
                       scol= groups[[1]], # colums in pl which contain test sample
intensities (upregulated features)
                       snames = gsub("_XIC","",colnames(pl)[groups[[1]]]),
                       ...)
for (i in c(2:length(groups))){
  if(sum(pl$group_up==groupnames[i])>0){
  pl2 <- rbind(pl2, sampid(pl=pl[which(pl$group_up==groupnames[i]),], #sampid checks the
values of which samples are above the control*sfactor
                  wtc= unlist(groups)[which(!unlist(groups) %in% groups[[i]])],# columns in pl
which contain control sample intensities
                  scol= groups[[i]], # colums in pl which contain test sample intensities
(upregulated features)
                  snames = gsub("_XIC","",colnames(pl)[groups[[i]]]),
                  ...))}
return (pl2)
}
}
msana2 <- function (pl = f2gf[1:20,],
```

```
                    groups = glist,
                    norm=T,
                    ttest=T,
                    adjmethod='holm',
                    fold =10){
  if(norm){pl <- NormDataCols(pl,cols=unlist(groups),replace=T,suffix = "")}
 count <-1
  for (i in groups){
    noni <- unlist(groups)[which(!unlist(groups) %in% i)]
    meanint <- rowMeans(as.matrix(pl[,i]))
    maxint <- rowMax(as.matrix(pl[,i]))
    fold2 <- rowMeans(as.matrix(pl[,i]))/rowMeans(as.matrix(pl[,noni]))
    sdev <- sapply(as.list(data.frame(t((pl[,i])))),sd)/rowMeans(as.matrix(pl[,i]))
    sdev2 <- sapply(split(pl[,i],seq(nrow(pl))),sd)/rowMeans(as.matrix(pl[,i]))
    sdev[which(is.na(sdev))] <-0
    if(ttest){if (min(length(i),length(noni)) > 1){
                        pval <- mttest(pl[,i],pl[,noni])}
                 else{ pval <- sapply(as.list(data.frame(t((pl[,c(i,noni)])))),
                                      ttestx,calc=1,over=c(2:(length(i)+length(noni))))
                        pval[which(pval>0.5)] <- 1-pval[which(pval>0.5)]}
             padj <- p.adjust(pval, method = adjmethod)}
        overg1 <-
mapply(minfolds,as.list(data.frame(t((pl[,groups[[1]]])))),as.list(data.frame(t((pl[,i])))),Mo
reArgs = list(fold=fold, mode="Max"))
    overall <-
mapply(minfolds,as.list(data.frame(t((pl[,noni])))),as.list(data.frame(t((pl[,i])))),MoreArgs
= list(fold=fold, mode="Average"))
   options(scipen = -100, digits = 4)
    pl[[paste0(names(groups)[count],"_stats")]] <- gsub("\\s+", " ",
trimws(paste(format(fold2,digits=5),format(sdev,digits=5),

if(ttest){format(pval,digits=5)},if(ttest){format(padj,digits=5)},

format(overg1,digits=3),format(overall,digits=3),

format(meanint,digits=3),format(maxint,digits=3))))
    options(scipen = 1, digits = 10)
    count=count+1
  }
  pl$massdefppm <- ((pl$mz-floor(pl$mz))/pl$mz)*1e6
  return(pl)
  }
groupfilter <- function (pl,
                         groups = glist,
                         minfold=0,
                         maxfold=Inf,
                         maxsd=1,
                         maxpval=1,
                         maxpadj=1,
                         overg1=0, #how many samples (fraction of total in group) have to be
above control*fold
                         overall=0){ #how many samples (fraction of total in group) have to be
above allgroups*fold
  pl$groups <- ""
  for (i in names(groups)){
    pl[[paste0(i,"_stats")]][1]
    tb <- do.call(rbind,strsplit(pl[[paste0(i,"_stats")]], split=" "))
    sel <- which(as.numeric(tb[,1])>=minfold
                 & as.numeric(tb[,1])<=maxfold
                 & as.numeric(tb[,2])<=maxsd
                 & as.numeric(tb[,3])<=maxpval
```

```
                & as.numeric(tb[,4])<=maxpadj
                & as.numeric(tb[,5])>=overg1*length(groups[[i]])
                & as.numeric(tb[,6])>=overall*length(groups[[i]])
                )
    pl$groups[sel] <- paste(pl$groups[sel],i)
  }
  return(pl)
}
minfolds <- function (x, # vector with values to compare y values to
                      y, # values
                      fold=10, #min fold change
                      mode='Max'){
  max(0,0)
    if (mode=="Max"){
    vec <- y/max(x)
    vec[which(is.na(vec))]<-0
    return(sum(vec>=fold))
  }
  if (mode=="Average"){
  vec <- y/mean(x)
  vec[which(is.na(vec))]<-0
  return(sum(vec>=fold))
}
}
splitter <- function (x,
                      sep=" ",
                      coln = NULL){
  x<- as.character(x)
  x <- do.call(rbind,strsplit(x, split=sep))
  colnames(x) <- coln
  return(x)
}
foldgr <- function (pl = f1[1:2,],
                    groups = glist,
                    mode ="Max"){ #or "Average"
if (mode == "Max"){f <-rowMax}else{f<-rowMeans}
  count <-1
  for (i in groups){
    fold <- vector(mode='character', length=nrow(pl))
    noni <- unlist(groups)[which(!unlist(groups) %in% i)]
    for (n in i){
      options(scipen = -100, digits = 4)
      fold <- paste(fold, trimws(format(pl[,n]/f(as.matrix(pl[,noni])),digits=3)))}
    pl[[paste0(names(groups)[count],"_fold_",mode)]] <- trimws(fold)
    count=count+1
  }
  return(pl)}
whichvec <- function(obj=d,
                     fil=t[1,which(t[1,]!="")]){
  cn <- colnames(fil)[which(fil[1,]!="")]
  whi=1:nrow(obj)
  whit <- whi
  for (i in cn[which(cn %in% colnames(obj))]){
    mm <- splitter(as.character(fil[,i]))
    if(ncol(mm)==1){
    whi <- which(as.numeric(obj[,i]) >= as.numeric(mm[1]))
    }else{
      if(mm[1]==''){
      whi <- which(as.numeric(obj[,i]) <= as.numeric(mm[2]))}
    else{
    whi <- which( as.numeric(obj[,i]) >= as.numeric(mm[1])
```

```
                & as.numeric(obj[,i]) <= as.numeric(mm[2]))
    }
    }
    wx <- c(whit,whi)
    whit <- as.numeric(names(table(wx)[which(table(wx)==2)]))
  }
  return(unique(whit))
}
dftrans <- function(a,b){
    for (i in colnames(a)[which(colnames(a) %in% colnames(b))]){
      b[c(1:(length(a[,i]))),i] <- as.character(a[,i])}
  print(paste('DFtrans from',deparse(substitute(a)),'to',deparse(substitute(b))))
  return(b)
}
rawread <- function(x, ...){
sc <- rawEIC(x, ...)
sc$rt <- x@scantime[sc$scan]
sc$tic <- x@tic[sc$scan]
sc$fname <- x@filepath[[1]]
sc$maxint <- max(sc$intensity)
sc$maxtic <- max(sc$tic)
return(sc)}
specplot <- function (x=tt2[,1],
                      y=tt2[,2],
                      norm=max(y)/100,
                      cx=1.5,
                      k = 10,
                      fileName = basename(rawdata[[1]][[1]]@filepath),
                      yrange = c(0,100),
                      xrange = range(x),
                      maxi = max(y),
                      ...
                      ){
 pd <- data.frame(x=x,y=y/norm)
  par(mar=c(5,4,10,2))
plot(pd$x,pd$y,type="h", bty="n", axes=F,lwd=0.8,
     main=fileName, cex.main=0.5*cx, ann=FALSE, ylab="Relative intensity",
     xlab= expression(italic(m/z)),
     xaxs="i",yaxs="i",
     xlim=xrange,
     ylim=yrange,
     ...)
  currview <- pd[which(pd$y <= max(yrange)
                       & pd$y >= min(yrange)
                       & pd$x <= max(xrange)
                       & pd$x >= min(xrange)),]
  if (length(currview$y) >= k){
  kn <-  sort(currview$y, decreasing = T)[k]
  labs <- currview[which(currview$y>=kn),]
  }else{
  labs <- currview}
  if(nrow(labs) > 0 ){
  labs$xcorr <- spread.labs(labs[,1],1.05*strwidth("A"), maxiter=1000, min=min(labs[,1]),
max=max(labs[,1]))
  par(xpd=NA)
  segments(labs[,1],labs[,2]+0.01*max(yrange),labs$xcorr,labs[,2]+0.05*max(yrange), col="red",
lwd=0.8)
  text(labs$xcorr,labs[,2]+0.055*max(yrange),labels=labs[,1], col="blue3",
srt=90,adj=c(0,0.3), cex=0.5*cx)
  mtext(side=3, text=format(maxi*(max(labs$y)/100), scientific = T, digits =4), line=0,
cex=0.5*cx, adj=1)
```

```
  }
mtext(side=1, text= expression(italic(m/z)), line=0.7, cex=0.5*cx)
mtext(side=2, text="Relative intensity (%)", line=1.1, cex=0.5*cx)
mtext(side=1, text=fileName, line=1.2, cex=0.5*cx)
#mtext(side=3, text=Ptext, line=0.6, cex=0.5, adj=1)
par(cex.axis=0.5*cx, tcl=-0.3)
axis(side=1, lwd=1, minor.tick(nx=10,ny=5, tick.ratio=0.5), mgp=c(0.5,0,0)) #x-axis mgp[2]
controls distance of tick labels to axis
axis(side=2, lwd=1, las=2, mgp=c(0.5,0.4,0)) #y-axis
}
rawselect <- function(namelist,rawlist){
  cl <- vector()
  for (i in rawlist)
    cl<- c(cl, basename(i@filepath[1]) %in% namelist)
  return(rawlist[which(cl)])
}
rawselect2 <- function(namex,rawlist){
     for (i in unlist(rawlist)){
        if(basename(i@filepath[[1]]) == namex){return (i)}}
    print("rawselect2: file not found!")
}
massquery <- function(mz, range=0.01, ppm=5,
                      elem= "C0-100H0-202N0-10O0-10F0-3Cl0-3Br0-1",
                      charge = 1){
    charge <- as.numeric(charge)
    if (charge > 0 ){charge2 <- paste0("%2B",abs(charge))}
    if (charge < 0 ){charge2 <- paste0("-",abs(charge))}
    if (charge == 0 ){charge2 <- paste0("")}
    if (!is.null(ppm)){range <- as.numeric(mz)*ppm*1e-6}
    mzq <-  paste0("http://www.chemcalc.org/service?action=em2mf&monoisotopicMass=",
                   mz,"&massRange=",range,"&mfRange=",
                   elem,
                   "(",charge2,")")
    res <- fromJSON(mzq)
    return(res$results)}
MSTana <- function(pl=plist_own,
                    glist = glist,
                    ctrl = glist[1],
                    normalize = T,
                    pval = T){
    wtc <- unlist(ctrl)
    scol <- unlist(glist)[which(!unlist(glist) %in% wtc)]
    pl <- msana1(pl,
                 groups = glist,
                 groupnames = names(glist),
                 norm=normalize,
                 ttest=F)
    pl[which(pl==0, arr.ind=T)[which(which(pl==0, arr.ind=T)[,2] %in% c(wtc,scol)),]]<-
min(as.matrix(pl[,c(wtc,scol)])[which(as.matrix(pl[,c(wtc,scol)])>0)])
    pl <- msana1(pl,
                 groups = glist,
                 groupnames = names(glist),
                 norm=F,
                 ttest=F)
    pl <- foldgr(pl, groups = glist, mode="Max")
    pl <- foldgr(pl, groups = glist, mode="Average")
    pl$maxint <- rowMax(as.matrix(pl[,unlist(glist)]))
    pl$maxfoldgr_max <-
rowMax(apply(do.call(cbind,lapply(pl[,grep("fold_Max",colnames(pl))],splitter)),2,as.numeric))
    pl$maxfoldgr_ave <-
rowMax(apply(do.call(cbind,lapply(pl[,grep("fold_Average",colnames(pl))],splitter)),2,as.numer
```

```
ic))
    pl <- msana2(pl,
                 groups = glist,
                 norm=F,
                 ttest=pval,
                 adjmethod='holm',
                 fold = 10)
    if(pval){
        groupsplitnames <- c("mean_fold_over_other_groups", "sdev", "pval", "pval_adjusted",
"samples_10fold_over_ctrl", "samples_10fold_over_other_groups", "mean_intensity",
"max_intensity")
    }else{
        groupsplitnames <- c("mean_fold_over_other_groups", "sdev",
"samples_10fold_over_ctrl", "samples_10fold_over_other_groups", "mean_intensity",
"max_intensity")
    }
    return(list(data=pl,groupsplitnames=groupsplitnames))
}
sttest <- function(out=NA,...){
    res <-try(t.test(...), silent = T)
    if(is(res,"try-error")){return(out)}else{return(res)}
}
setClass("MSps", slots= c( featuremz="vector",featurert="vector",
                           MS1meta = "data.frame",
                           MS2meta="data.frame"))
Parentsearch <- function (bin1_p, mzs = c(702.3249,451.26903),rts = c(400,300), partol= 0.02,
rttol=200, MSn = T, MS1 = T){
    if(MS1==T){
        mzlist=list()
        rawlist=list()
        rawcoll=list()
        for (n in c(1:length(mzs))){
            for(t in c(1:length(bin1_p))){
                for (numb in c(1:length(bin1_p[[t]]))){
                    rawlist[[numb]] <-
data.frame(getScan(bin1_p[[t]][[numb]],which.min(abs(bin1_p[[t]][[numb]]@scantime-
rts[n]))))}#, mzrange= c(mzs[n]-2,mzs[n]+5) )}
                rawcoll[[t]] <-rawlist
                rawlist=list()}
            mzlist[[n]] <-rawcoll
            rawcoll=list()}
        MS1speclist <- mzlist
        mzlist=list()
        rawlist=list()
rawcoll=data.frame(scan=integer(),rt=numeric(),scannum=integer(),MS1target=numeric(),MS1hit=nu
meric(),MS1hitI=numeric(),MS1ppm=numeric(),file=character(), stringsAsFactors = F)
        for (n in c(1:length(mzs))){
            for(t in c(1:length(bin1_p))){
                for (numb in c(1:length(bin1_p[[t]]))){
                    scan <- which.min(abs(bin1_p[[t]][[numb]]@scantime-rts[n]))
                    rawlist[[numb]] <- data.frame(scan)
                    rawlist[[numb]]$rt <- bin1_p[[t]][[numb]]@scantime[scan]
                    rawlist[[numb]]$scannum <- bin1_p[[t]][[numb]]@acquisitionNum[scan]
                    rawlist[[numb]]$MS1target <- mzs[n]
                    rawlist[[numb]]$MS1hit <-
MS1speclist[[n]][[t]][[numb]]$mz[which.min(abs(mzs[n]-MS1speclist[[n]][[t]][[numb]]$mz))]
                    rawlist[[numb]]$MS1hitI <-
MS1speclist[[n]][[t]][[numb]]$intensity[which.min(abs(mzs[n]-
MS1speclist[[n]][[t]][[numb]]$mz))]
                    rawlist[[numb]]$MS1ppm <-
(abs((MS1speclist[[n]][[t]][[numb]]$mz[which.min(abs(mzs[n]-
```

```
MS1speclist[[n]][[t]][[numb]]$mz))]-mzs[n]))/mzs[n])*1000000
                    rawlist[[numb]]$file <- basename(bin1_p[[t]][[numb]]@filepath[[1]])
                }
                rawcoll <-rbind(rawcoll,rbindlist(rawlist))
                rawlist=list()}
            mzlist[[n]] <-rawcoll
            rawcoll=list()}
        MS1metalist <- rbindlist(mzlist)
    }else{
        MS1metalist =data.frame()}
    if(MSn==T){
        rawlist=list()
        msnlist=list()
        metalist=list()
        rawcoll=list()
        for (n in c(1:length(mzs))){
            for(t in c(1:length(bin1_p))){
                for (numb in c(1:length(bin1_p[[t]]))){
                    scans <- which(abs(bin1_p[[t]][[numb]]@msnPrecursorMz-mzs[n])<partol
                                   & abs(bin1_p[[t]][[numb]]@msnRt-rts[n])<rttol)
                    if(length(scans)>0){
                        for(s in c(1:length(scans))){
                            msnlist[[s]]<-data.frame(scans[s], stringsAsFactors = F)
                            msnlist[[s]]$msnPrecursorMz <-
bin1_p[[t]][[numb]]@msnPrecursorMz[scans[s]]
                            msnlist[[s]]$rt <- bin1_p[[t]][[numb]]@msnRt[scans[s]]
                            msnlist[[s]]$scannum <-
bin1_p[[t]][[numb]]@msnAcquisitionNum[scans[s]]
                            msnlist[[s]]$msnPrecursorIntensity <-
bin1_p[[t]][[numb]]@msnPrecursorIntensity[scans[s]]
                            msnlist[[s]]$msnCollisionEnergy <-
bin1_p[[t]][[numb]]@msnCollisionEnergy[scans[s]]
                            msnlist[[s]]$file <- basename(bin1_p[[t]][[numb]]@filepath[[1]])
                        }}else{msnlist[[1]] <-
data.frame(scans=integer(),msnPrecursorMz=numeric(),rt=numeric(),scannum=numeric(),msnPrecurso
rIntensity=numeric(),msnCollisionEnergy=numeric(),file=character(), stringsAsFactors = F)}
                    rawlist[[numb]] <- rbindlist(msnlist)
                    msnlist=list()
                }
                rawcoll[[t]] <-rbindlist(rawlist)
                rawlist=list()}
            metalist[[n]] <-rbindlist(rawcoll)
            rawcoll=list()}
    }else{scanlist=list()
    metalist=data.frame()}
    tata <- new("MSps",  featuremz = mzs, featurert = rts,
                MS1meta=MS1metalist,
                MS2meta=rbindlist(metalist)
    )
  return(tata)}
psmile <- function(SMILE, width=500, height=500){
    mols <- parse.smiles(SMILE)
    par(mar=c(0,0,0,0)) # set margins to zero since this isn't a real plot
    plot(NA,NA,xlim=c(1,10),ylim=c(1,10),xaxt='n',yaxt='n',xlab='',ylab='')
    temp1 = view.image.2d(mols[[1]],width,height)
    rasterImage(temp1,1,1,10,10, col = "black")
}
decodeanno <- function(x, sep1="*", sep2 = " "){
    tinp2 <- strsplit(x, sep1, fixed = T)[[1]]
    mz <- numeric(0)
    sum_formula <- character(0)
```

```
    SMILE <- character(0)
    for (i in tinp2){
        tinp3 <- strsplit(i, sep2)[[1]]
        mz <- c(mz,as.numeric(tinp3[1]))
        sum_formula <- c(sum_formula,as.character(tinp3[2]))
        SMILE <- c(SMILE,as.character(tinp3[3]))
    }
    return(data.frame(mz,sum_formula,SMILE, stringsAsFactors = F))}
encodeanno <- function(x, sep1="*", sep2 = " "){
    coll <-""
    if (nrow(x)==0){return("")}
    item <- paste(x[1,1],x[1,2],x[1,3], sep = sep2)
    if (nrow(x)>1){
    for ( n in 2:nrow(x)){
       itemt <- paste(x[n,1],x[n,2],x[n,3], sep = sep2)
       item <- paste(item,itemt, sep=sep1)
    }}
    return(item)}
getMSnOwn <-  function(object,
                       scannum,
                       mzrange = numeric()) {
    scan <- which(object@msnAcquisitionNum==scannum)
    if (scan < 0)
        scan <- length(object@msnRt) + 1 + scan
    idx <- seq(object@msnScanindex[scan]+1, min(object@msnScanindex[scan+1],
                                                length(object@env$msnMz), na.rm=TRUE))
    if (length(mzrange) >= 2) {
        mzrange <- range(mzrange)
        idx <- idx[object@env$msnMz[idx] >= mzrange[1] & object@env$msnMz[idx] <= mzrange[2]]
    }
    points <- cbind(mz = object@env$msnMz[idx], intensity = object@env$msnIntensity[idx])
    invisible(points)
}
```

R code for XCMS (with inputs from Dr. Susan Strickler and Dr. Maximilian Helf)

```
INPUTFOLDER <- "F:/Users$/op56/Desktop/Mit Mutants WM centroid/WM pos/"#to be set by user
OUTPUTFOLDER <- "F:/Users$/op56/Desktop/Mit Mutants WM centroid/WM pos/"#to be set by user
camera <-FALSE
setwd(OUTPUTFOLDER)
f1 <- list.dirs(INPUTFOLDER,full.names = T,recursive = F)
for (fo in f1){
if(length(grep("pos",fo))>0){
#set polarity
pol <- "positive"
}else{
pol <- "negative"
}
tag1 <- basename(fo)
f2 <- list.dirs(fo,full.names = T,recursive = F)
for (fo2 in f2){
    tag2 <- paste0(tag1,"_",basename(fo2),"_")
mzxml_pos <- list.files(fo2, pattern="mzXML", recursive = TRUE, full.names = T)
print(tag2)
print(mzxml_pos)
bothl <- list()
bothl[[1]] <- mzxml_pos
rawdata <- EICrawP(bothl)
param <- SnowParam(workers = 4)
  xset.raw.pos <- xcmsSet(mzxml_pos, method="centWave",
                          ppm=4, peakwidth=c(3,20),
                          mzdiff=0.01, snthresh=10,prefilter=c(3,100),
```

```
                          BPPARAM=param)
  save(xset.raw.pos, file=paste0(tag2,"xsetraw"))
  xset.pos <- group(xset.raw.pos, minfrac=0.2, bw=5, mzwid=0.005, max=500, minsamp=1)
  save(xset.pos, file=paste0(tag2,"xsetgrouped"))
  xset.pos.pl <- peakTable(xset.pos)
  xset.pos.pl[is.na(xset.pos.pl)]<-0
  write.csv(xset.pos.pl,file = paste0(tag2,"peaklist_nonRTcorr.csv"))
  xset2.pos <- retcor(xset.pos, method="obiwarp", profStep=0.1, plottype = "deviation")
  save(xset2.pos, file=paste0(tag2,"xset2_pos"))
  xset3.pos <- group(xset2.pos, minfrac=0, bw=5, mzwid=0.005, max=500, minsamp=1)
  save(xset3.pos, file=paste0(tag2,"xset3_pos"))
  xset3.pos.pl <- peakTable(xset3.pos)
  xset3.pos.pl[is.na(xset3.pos.pl)]<-0
  write.csv(xset3.pos.pl,file = paste0(tag2,"peaklist_RTcorr.csv"))
  xset3.pos.filled <- fillPeaks.chrom(xset3.pos, BPPARAM=param)
  xset3.pos.filled.pl <- peakTable(xset3.pos.filled)
  write.csv(xset3.pos.filled.pl,file = paste0(tag2,"peaklist_filled_RTcorr.csv"))
  plist_own <- xset3.pos.filled.pl
  pt2<-proc.time()
  for (n in c(1:length(rawdata[[1]]))){
    plist_own <-    multiEIC(rawfile= rawdata[[1]][[n]] ,
                             featuretable=plist_own,
                             mz = plist_own$mz,
                             ppm=5,
                             rtw= data.frame(plist_own$rtmin-5,plist_own$rtmax+5),
                             mini=5000,
                             coltag=paste0(rawdata[[1]][[n]]@filepath@.Data),
                             pval=0.05,
                             XIC=T,
                             abovex = F,
                             gauss= F)
    print(n)}
  pt2<-proc.time()-pt2
  print(pt2)
  write.csv(plist_own,file = paste0(tag2,"_plist_own_RTcorr.csv"))
}
}
```

R code for EIC Reader (developed by Dr. Maximilian Helf)

```
library(xcms)
library(graphics)
library(Hmisc)
binvar1 <- c("Everything")
grn <- 3
ppr <- 5 #ppm
rtw <- 20 #sec
#input a tsv file with AT LEAST two columns, "mz" and "rt". (separator currently set to tabs;
fileName <- choose.files()
alli <- read.table(fileName, header= T, sep= "\t", dec=".", fill= T, skip=0)
mzr  <- data.frame(alli$mz-0.000005*ppr*alli$mz,alli$mz+0.000005*ppr*alli$mz)
rtr <- data.frame(alli$rt-rtw,alli$rt+rtw)
#select mzXML files
fileList=list()
for(i in c(1:length(binvar1))){
fileList[[i]] <- choose.files(default= if(i>1){dirname(fileList[[i-1]][1])}else{""},
caption=paste("Add files to group", binvar1[i]), multi=T)
repeat{
if(winDialog(type="yesno", "Do you want to add more files to this group?")=="YES")
{fileList[[i]] <- c(fileList[[i]], choose.files(default=
dirname(fileList[[i]][1]),caption=paste("Add more files to group", binvar1[i]),
multi=T))}else{break}}}
bin1_p <- fileList
```

```
rawlist=list()
rawcoll=list()
for(t in c(1:length(bin1_p))){
  for (numb in c(1:length(bin1_p[[t]]))){
    rawlist[[numb]] <- xcmsRaw(bin1_p[[t]][numb], profstep=0)}
  rawcoll[[t]] <-rawlist
  rawlist=list()}
nakedFile <- sub("^([^.]*).*", "\\1",fileName)
pdfFile <- paste0(nakedFile,"_EICv6_output.pdf")
if(file.exists(pdfFile)){pdfFile <-paste0(nakedFile,"_",round(runif(1,min=0.1,max=1)*10000,
digits=0),"_EICv6_output.pdf")}  #{file.remove(pdfFile)}
nr <- if(length(bin1_p)>=grn){ceiling((length(bin1_p))/grn)}else{1}
Fighi <- 5*nr+2
Figwi <- 3.5*grn
pdf(pdfFile,Figwi,Fighi)
par(mfrow=c(nr,grn), oma=c(0,0,10,0),xpd=NA, bg=NA)
eiclist=list()
maxlist=list()
eiccoll=list()
maxcoll=list()
for(t in c(1:length(bin1_p))){
  for (numb in c(1:length(bin1_p[[t]]))){
    eiclist[[numb]] <- rawEIC(rawcoll[[t]][[numb]],mzrange =c(0,5000),rtrange =
c(0,max(rawcoll[[t]][[numb]]@scantime)))
    maxlist[[numb]] <- max(eiclist[[numb]]$intensity)
    maxlist[[numb]]$loop <- numb}
  eiccoll[[t]] <- eiclist
  eiclist=list()
  maxcoll[[t]] <- maxlist
  maxcoll[[t]]$cloop <- t
  maxlist=list()}


maxsi=list()
for(t in c(1:length(bin1_p))){
  for (numb in c(1:length(bin1_p[[t]]))){
    maxsi[[t]] <- max(unlist(maxcoll[[t]]))#,maxcoll[[t]][[numb]][[1]]))
  }}
for(t in c(1:length(bin1_p))){
  #making RT out of scan#
  RTwr <- c(min(rawcoll[[t]][[1]]@scantime)/60,max(rawcoll[[t]][[1]]@scantime)/60)
  RT1 <- (eiccoll[[t]][[1]]$scan-eiccoll[[t]][[1]]$scan+RTwr[[1]]+(eiccoll[[t]][[1]]$scan-
min(eiccoll[[t]][[1]]$scan))*((RTwr[[2]]-RTwr[[1]])/(max(eiccoll[[t]][[1]]$scan)-
min(eiccoll[[t]][[1]]$scan))))
  plot(RT1,eiccoll[[t]][[1]]$intensity, type= "n", xlim=
c(0,max(rawcoll[[t]][[1]]@scantime)/60), ylim =c(0,maxsi[[t]]), axes=F,
ylab="Intensity",xlab="")
  axis(side=1, lwd=1, minor.tick(nx=10,ny=5, tick.ratio=0.5), mgp=c(1,0.4,0))#x-axis mgp[2]
controls distance of tick labels to axis
  axis(side=2, lwd=1, las=2, mgp=c(0.7,0.6,0))
  mtext(side=1, text= "RT (min)", line=1.2, cex=0.7)
  title(main=binvar1[t], line=4+0.1*length(eiccoll[[t]]))
  colr <- rainbow(length(eiccoll[[t]]), s = 1, v = 1, start = 0, end = max(1,
length(eiccoll[[t]]) - 1)/length(eiccoll[[t]]), alpha = 0.7)#topo.colors(length(eiccoll[[t]]),
alpha=1)
  for (numb in c(1:length(eiccoll[[t]]))){
    RTa <- (eiccoll[[t]][[numb]]$scan-
eiccoll[[t]][[numb]]$scan+RTwr[[1]]+(eiccoll[[t]][[numb]]$scan-
min(eiccoll[[t]][[numb]]$scan))*((RTwr[[2]]-RTwr[[1]])/(max(eiccoll[[t]][[numb]]$scan)-
min(eiccoll[[t]][[numb]]$scan))))
    lines(RTa,eiccoll[[t]][[numb]]$intensity, col=colr[numb], lwd=1)}
```

```
  legend("topright", inset=c(0,-0.025*length(eiccoll[[t]])),basename(bin1_p[[t]]),
lty=1,lwd=2.5, col=colr, bty="n",  cex=.5)}
mtext("TICs", side=3, outer=T,line=7.2, cex=1)
eiclist=list()
maxlist=list()
eiccoll=list()
maxcoll=list()
for(m in c(1:NROW(mzr))){
  RTwr <- if(nrow(rtr)==nrow(mzr)){rtr[m,]}else{rtr[1,]}
  for(t in c(1:length(bin1_p))){
    for (numb in c(1:length(bin1_p[[t]]))){
      eiclist[[numb]] <- rawEIC(rawcoll[[t]][[numb]],mzrange = mzr[m,],rtrange = RTwr)
      maxlist[[numb]] <- max(eiclist[[numb]]$intensity)
      maxlist[[numb]]$loop <- numb}
    eiccoll[[t]] <- eiclist
    eiclist=list()
    maxcoll[[t]] <- maxlist
    maxcoll[[t]]$cloop <- t
    maxlist=list()}
  maxsi=list()
  for(t in c(1:length(bin1_p))){
    for (numb in c(1:length(bin1_p[[t]]))){
      maxsi[[t]] <- max(unlist(maxcoll[[t]]))#,maxcoll[[t]][[numb]][[1]]))
    }}
 par(mfrow=c(nr,grn), oma=c(0,0,10,0),xpd=NA, bg=NA)
  for(t in c(1:length(bin1_p))){
    RT1 <- (eiccoll[[t]][[1]]$scan-eiccoll[[t]][[1]]$scan+RTwr[[1]]+(eiccoll[[t]][[1]]$scan-
min(eiccoll[[t]][[1]]$scan))*((RTwr[[2]]-RTwr[[1]])/(max(eiccoll[[t]][[1]]$scan)-
min(eiccoll[[t]][[1]]$scan))))/60
    plot(RT1,eiccoll[[t]][[1]]$intensity, type= "n", ylim =c(0,maxsi[[t]]), axes=F,
ylab="Intensity",xlab="")
    axis(side=1, lwd=1, minor.tick(nx=10,ny=5, tick.ratio=0.5), mgp=c(1,0.4,0))#x-axis mgp[2]
controls distance of tick labels to axis
    axis(side=2, lwd=1, las=2, mgp=c(0.7,0.6,0))
    mtext(side=1, text= "RT (min)", line=1.2, cex=0.7)
    title(main=binvar1[t], line=4+0.1*length(eiccoll[[t]]))
    colr <- rainbow(length(eiccoll[[t]]), s = 1, v = 1, start = 0, end = max(1,
length(eiccoll[[t]]) - 1)/length(eiccoll[[t]]), alpha = 0.7)#topo.colors(length(eiccoll[[t]]),
alpha=1)
    for (numb in c(1:length(eiccoll[[t]]))){
      RTa <- (eiccoll[[t]][[numb]]$scan-
eiccoll[[t]][[numb]]$scan+RTwr[[1]]+(eiccoll[[t]][[numb]]$scan-
min(eiccoll[[t]][[numb]]$scan))*((RTwr[[2]]-RTwr[[1]])/(max(eiccoll[[t]][[numb]]$scan)-
min(eiccoll[[t]][[numb]]$scan))))/60
    lines(RTa,eiccoll[[t]][[numb]]$intensity, col=colr[numb], lwd=1)}
    legendtext <- paste(sub("^([^.]*).*", "\\1",basename(bin1_p[[t]])))
    legend("topright", inset=c(0,-0.025*length(eiccoll[[t]])),legendtext, lty=1,lwd=2.5,
col=colr, bty="n",  cex=.5)}
  mtext(paste("m/z range:", round(mzr[m,1],5),"-
",round(mzr[m,2],5),"(",round(alli$mz[m],5),"+/- 10 ppm @ RT",round(alli$rt[m],1),"sec
/",round(alli$rt[m]/60,2), "min"), side=3, outer=T,line=7.2, cex=1)
  mtext(paste("Category:", alli$Sample[m], alli$Medium[m], alli$Extraction[m], alli$Tool[m]),
side=3, outer=T,line=6, cex=1)
  mtext(paste("Comment:",alli$Comments[m]), side=3, outer=T,line=4.8, cex=1)}
dev.off()
shell.exec(pdfFile)
```

APPENDIX E: SUPPLEMENTAL INFORMATION FOR CHAPTER 5

Materials and methods

Nematode Culture and Extraction. Mixed stage worms from a populated 10 cm NGM agar plate seeded with *E. coli* OP50 were washed into 25 ml of S-complete medium and fed OP50 on days 1, 3 and 5 of a seven-day total culture period, while shaking at 22 °C, 220 rpm. The cultures were then centrifuged and worm pellets and supernatant frozen separately and lyophilized. The lyophilized supernatant was extracted with 35 mL of 95% ethanol at room temperature for 12 h. The extracts were dried *in vacuo*, resuspended in 250 µL methanol and analyzed by LC-MS.

Mass Spectrometric Analysis. High resolution LC-MS analysis was performed on a Dionex 3000 UPLC coupled with a Thermo Q Exactive high-resolution mass spectrometer as described in Appendix C. Metabolites from worm media samples were separated using water-acetonitrile gradient on Agilent Zorbax Eclipse XDB-C18 column (150 mm x 2.1 mm, particle size 1.8 µm) maintained at 40 °C. Solvent A: 0.1% formic acid in water; Solvent B: 0.1% formic acid in acetonitrile. A/B gradient started at 5% B for 1.5 min after injection and increased linearly to 100% B at 12.5 min. The matched filter algorithm in XCMS for peak picking in the profile data was used for results in this chapter.

Supplemental figures

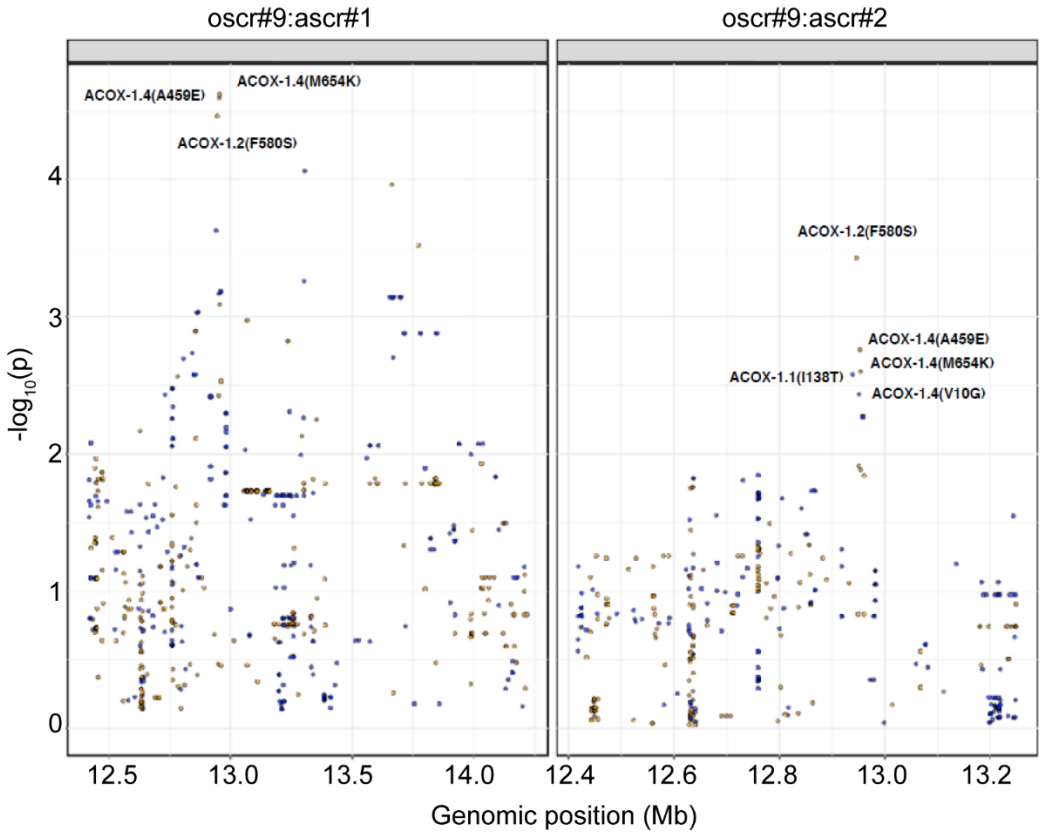

Figure E1. Fine mappings for oscr#9. oscr#9:ascr#1 ratio fine-maps to Met654Lys, Ala459Glu at *acox-1.4*, and Phe580Ser at *acox-1.2*. oscr#9:ascr#2 ratio fine-maps to Phe580Ser at *acox-1.2*, Met654Lys, Ala459Glu and Val10Gly at *acox-1.4*, and Ile138Thr at *acox-1.1*.

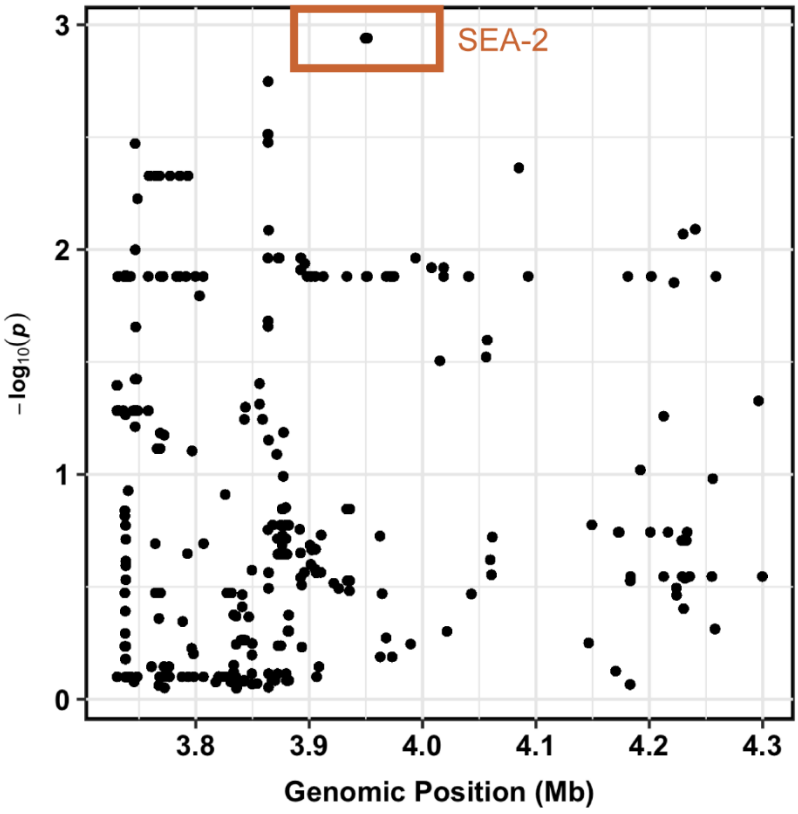

Figure E2. Fine mapping for ascr#10:ascr#3 ratio. The gene *sea-2* will be tested for causality.

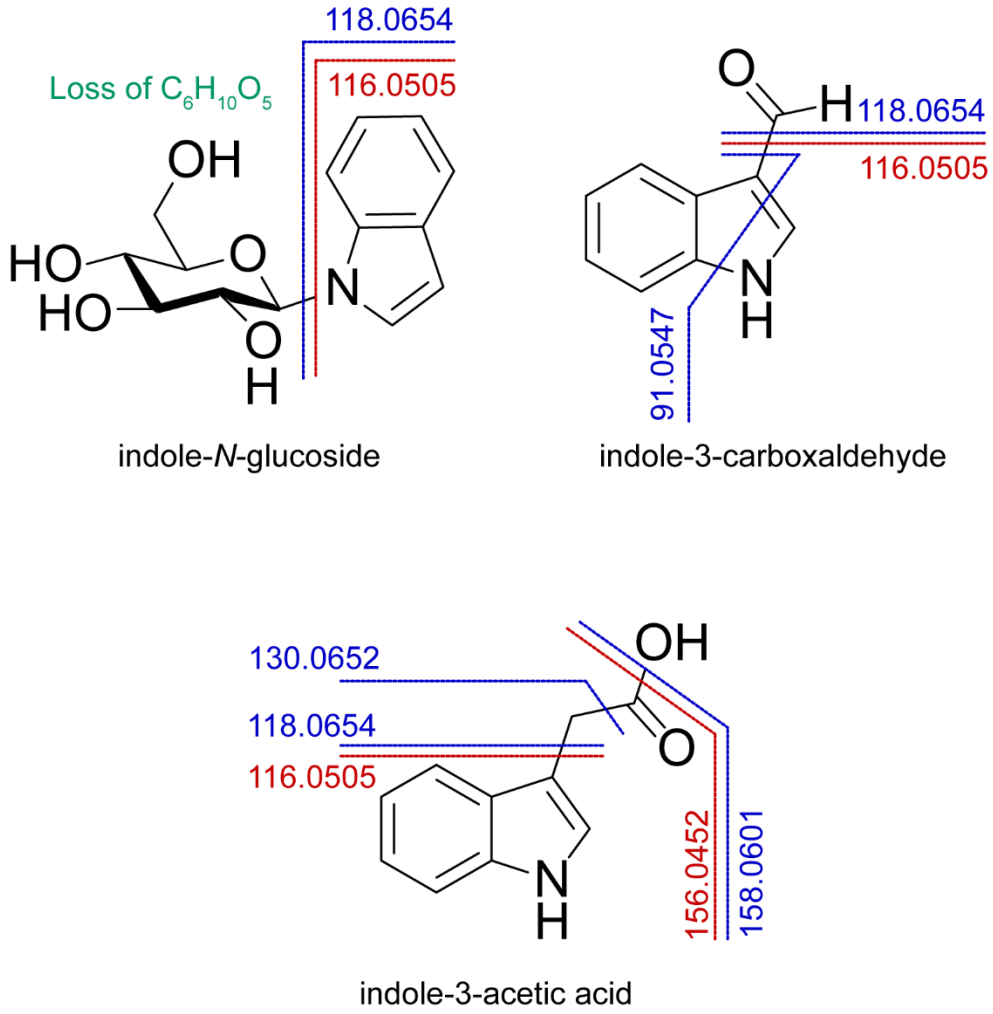

Figure E3. MS/MS fragmentation pattern for indole derivatives. ESI negative fragments are shown in red, ESI positive fragments are shown in blue, neutral losses are shown in green.

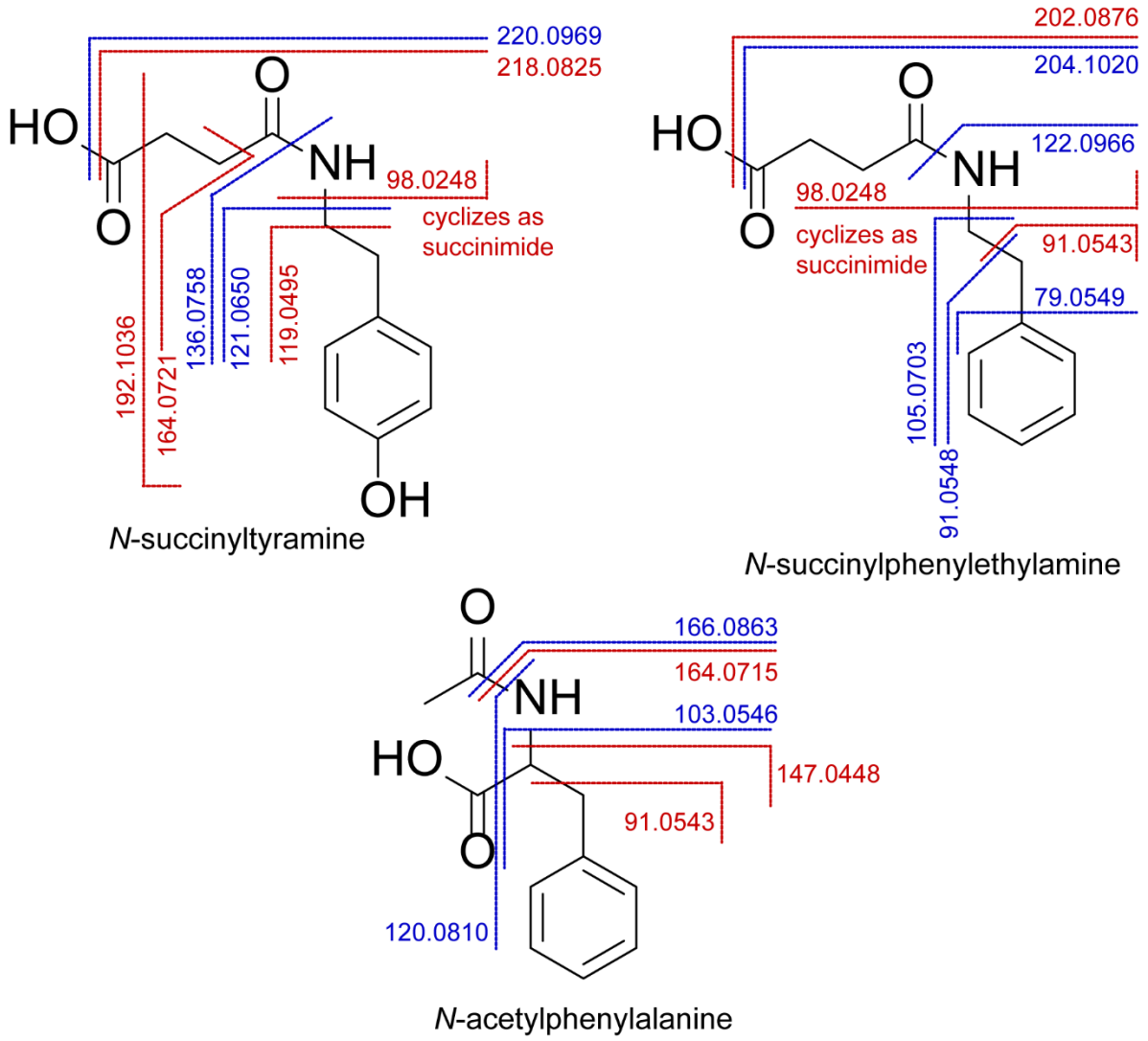

Figure E4. MS/MS fragmentation pattern for amino acid derivatives. ESI negative fragments are shown in red, ESI positive fragments are shown in blue.

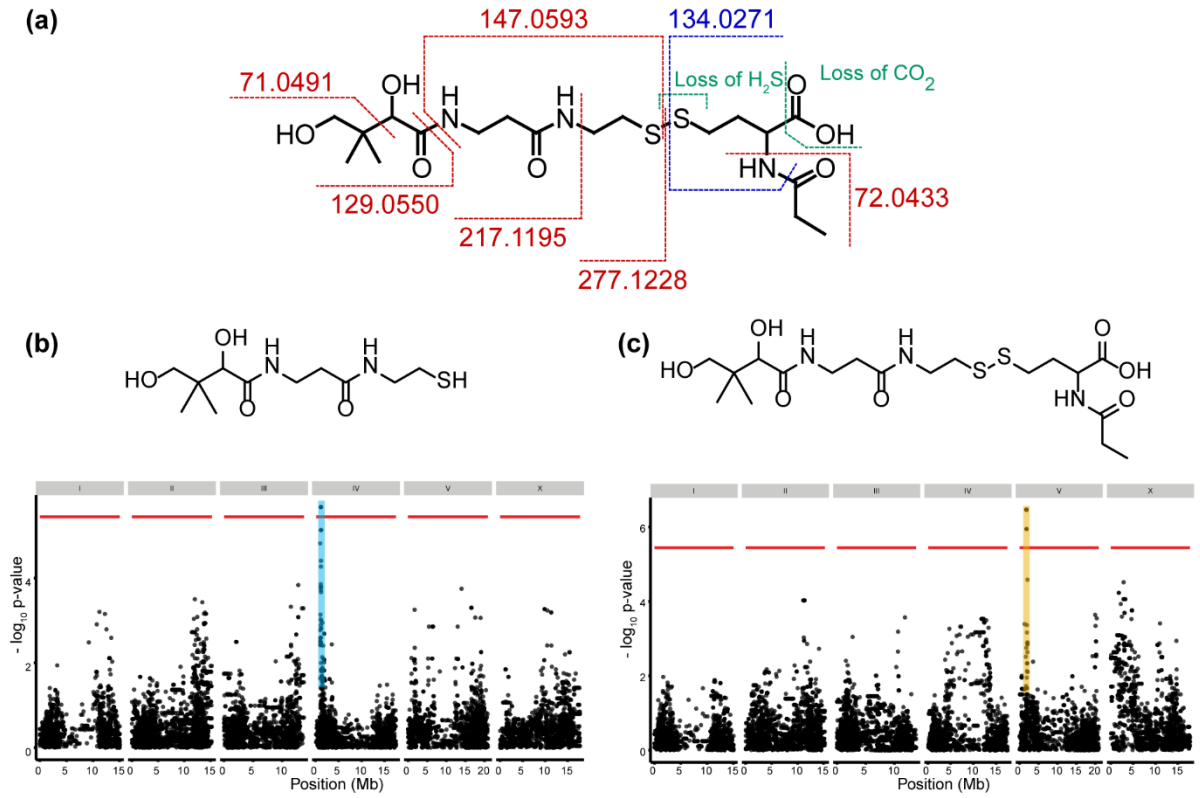

Figure E5. MS/MS fragmentation pattern for putative pantetheinyl-*N*-propionylhomocysteine. (a) ESI negative fragments are shown in red, ESI positive fragments are shown in blue, neutral losses are shown in green. **(b)** Pantetheine maps to a different QTL (blue box), suggesting the possibility of a disulfide-forming biosynthetic gene in the QTL (yellow box) for this novel compound **(c)**.

Supplemental tables

Table E1. Strains used in Chapter 5

AB1	CX11285	ED3012	EG4725	JU1395	JU1896
AB4	CX11292	ED3017	EG4946	JU1400	JU2001
BRC20067	CX11314	ED3046	GXW1	JU1409	JU2007
CB4855	CX11400	ED3048	JT11398	JU1516	JU2316
CB4856	DL200	ED3049	JU1172	JU1530	JU2464
CB4932	DL226	ED3052	JU1200	JU1568	JU2466
CX11254	DL238	ED3073	JU1212	JU1580	JU258
CX11264	ECA36	ED3077	JU1213	JU1581	JU310
CX11271	ED3005	EG4347	JU1242	JU1586	JU311
CX11276	ED3011	EG4349	JU1246	JU1652	JU323
JU346	JU642	LSJ1	NIC2	QG1	QX1793
JU360	JU751	MY1	NIC207	QG536	QX1794
JU363	JU774	MY10	NIC231	QG537	RC301
JU367	JU775	MY16	NIC236	QG538	WN2001
JU393	JU778	MY23	NIC256	QG556	
JU394	JU792	N2	NIC267	QG557	
JU397	JU830	NIC1	NIC3	QX1212	
JU406	JU847	NIC166	PB303	QX1233	
JU440	KR314	NIC195	PB306	QX1791	
JU491	LKC34	NIC199	PS2025	QX1792	

Table E2. Genes with single nucleotide variants unique to ECA36 and QX1794

Amino acid change	Gene ID	Gene name	Effect	Description
p.Gln76His	WBGene00021298	Y27F2A.6	missense variant	involved in lipid storage
p.Pro86Leu	WBGene00001571	gei-14	missense variant	GEX interacting protein
p.Ala1100Thr	WBGene00007789	C28A5.2	missense variant	not available
p.Lys176Gln	WBGene00018340	anmt-2	missense variant	amine *N*-methyltransferase
p.Val113Ala	WBGene00000665	col-90	missense variant	collagen
p.Ile77Val	WBGene00022583	mrps-18A	missense variant	mitochondrial ribosomal protein, small
p.Asn452Ile	WBGene00013540	Y75B8A.3	missense variant	ortholog of human neuroligin

Table E3. Genes with small indels and large deletions unique to ECA36 and QX1794

Gene ID	Gene name	Effect	Description
WBGene00013138	Y53C10A.10	small indel	not available
WBGene00018223	bath-13	small indel	BTB and MATH domain containing
WBGene00019150	H04J21.1	small indel	not available
WBGene00017099	E02H9.1	small indel	not available
WBGene00020598	T20B6.2	small indel	not available
WBGene00015467	basl-1	small indel	BAS-like
WBGene00015988	C18H2.1	small indel	involved in lipid storage and oviposition
WBGene00013318	Y57G11C.23	small indel	predicted transmembrane transporter activity
WBGene00017722	F22F7.4	small indel	not available
WBGene00022016	Y61A9LA.4	small indel	not available
WBGene00045493	F55G7.4	small indel	not available
WBGene00008902	phf-32	small indel	PHd Finger family
WBGene00015198	clec-2	large deletion	C-type lectin
WBGene00016577	clec-3	large deletion	C-type lectin
WBGene00022327	fbxa-15	large deletion	F-box A protein
WBGene00021152	mgl-3	large deletion	G-protein coupled metabotropic glutamate receptor family
WBGene00047634	21ur-4408	large deletion	21U-RNA
WBGene00013066	Y51A2B.2	large deletion	not available

Table E4. Features with significant mappings (features with same letter superscripts map to common loci)

m/z	rt (min)	m/z	rt (min)	m/z	rt (min)	m/z	rt (min)
154.03707	0.68	466.24830	3.58 [d]	206.08217	5.26 [b]	1165.5776	7.56
135.05062	0.70	290.07150	3.61 [c]	521.04964	5.32	598.29897	7.63
184.02105	0.72	139.04003	3.67	144.04541	5.34 [a]	490.21288	7.75
247.09311	0.74	349.14334	3.89	180.02211	5.34 [a]	1050.5346	7.77
274.10422	0.75	644.36262	3.96	220.09784	5.44 [b]	402.28549	7.84
182.02065	0.75	236.09263	4.00 [b]	288.08517	5.45	452.27317	7.88
254.01767	0.77	224.11518	4.02	324.19452	5.48	435.13847	8.00
166.01608	0.79	163.00363	4.04	1010.3963	5.50	404.17597	8.01
169.01976	0.87	717.38741	4.08	482.08504	5.52	873.50011	8.02
392.18138	0.89	416.02830	4.10	520.29049	5.58	917.42644	8.04
209.09786	0.89	222.07708	4.14 [b]	685.37805	5.58	916.42390	8.07
231.13462	0.92	345.05824	4.31	501.21070	5.67	534.27881	8.17
201.08787	1.03	432.12292	4.34	1051.4184	5.82	198.15790	8.72
317.09862	1.09 [f]	274.13758	4.42	363.05499	5.85	185.15457	8.77 [e]
339.08043	1.09 [f]	318.11749	4.47	756.24745	6.20	211.17020	9.28
812.31364	1.11	653.36266	4.48	293.15077	6.31 [g]	199.17024	9.34 [e]
365.09834	1.17	116.05049	4.49 [a]	498.04697	6.51	571.28168	9.57
160.05755	1.33	318.08295	4.49	627.43381	6.71	398.22793	9.71
542.33071	1.43	470.18517	4.56	672.33564	6.79	229.18066	9.76
537.19318	1.52	466.16827	4.63	295.16606	6.81 [g]	564.36621	10.1
349.04497	1.61	259.09079	4.67	918.37334	6.85	374.32733	10.1
205.07161	1.96	499.19530	4.74	442.28013	6.88	1039.6284	10.3
263.07042	2.08	452.13170	4.80	444.14909	6.94	427.23633	10.4 [c]
437.21350	2.40	452.11053	4.80	578.07720	6.95 [d]	578.38200	10.6
723.07279	2.61	951.48401	4.81	332.24387	6.96	534.32887	10.8
237.10109	2.74	246.04045	4.85	685.19976	7.02	594.39434	10.9
271.12398	3.06	174.05578	4.88 [a]	183.10052	7.08	607.41707	11.2
224.11514	3.07	333.14470	4.93	628.43660	7.09	711.42655	11.6
722.24433	3.09	251.05920	4.93	336.15587	7.11	347.25490	11.7
648.22604	3.15	376.14295	4.93	345.24699	7.12	725.10448	11.9
244.13159	3.17	215.08249	4.94 [a]	398.21285	7.15	625.51617	12.4
167.03245	3.20	251.07820	4.94	1095.5354	7.26	765.67197	13.7
201.07695	3.28	378.14034	4.95	333.09445	7.26	352.22387	13.8
135.04502	3.31	348.13845	5.00	315.17104	7.32	741.53005	13.9
610.12409	3.48	445.25510	5.02	415.24454	7.33	690.47908	14.0
610.08327	3.48	255.13279	5.02	359.21792	7.41	690.51727	14.0
258.08405	3.49	383.12447	5.11	640.30878	7.41	440.27657	14.3
465.25381	3.49	488.16900	5.20	656.46887	7.44	959.61687	14.9
234.12289	3.57	162.09239	5.26	1109.5519	7.48		

Table E5. Candidate genes for indole derivatives (all missense variants)

Amino acid change	Gene ID	Gene name	Spearman correlation coefficient	Description
p.Thr258Ala	WBGene00008140	xpf-1	0.300804	DNA repair gene homolog
p.Lys116Gln	WBGene00013923	ghi-1	0.287163	glycoprotein hormone interacting protein
p.Arg239Lys	WBGene00013925	ZK20.4	0.287163	ortholog of human cytochrome c oxidase assembly factor 7
p.Asp100Asn	WBGene00013926	nep-1	0.287163	neprilysin metallopeptidase family
p.Asp410Asn	WBGene00008743	F13D12.9	0.287163	involved in lipid storage
p.Val289Gly	WBGene00009690	F44E5.3	0.281804	not available
p.Gly126Ser	WBGene00007342	C05D12.5	0.27519	not available
p.Leu271Phe	WBGene00014666	C05D12.3	0.272422	Ortholog of human *N*-acetylglucosamine-1-phosphate transferase
p.Phe17Leu	WBGene00007340	C05D12.4	0.263979	not available
p.Thr8Met	WBGene00012157	adbp-1	0.263979	ADR-2 binding protein
p.Ser201Arg	WBGene00007143	B0334.3	0.263979	ortholog of human 2-hydroxyacyl-CoA lyase 1
p.Lys323Arg	WBGene00010055	F54D5.12	0.263979	ortholog of human D-2-hydroxyglutarate dehydrogenase
p.Asn285His	WBGene00008547	F07A11.4	0.263979	ortholog of human ubiquitin specific peptidase
p.Leu114Met	WBGene00008549	din-1	0.263979	DAF-12 corepressor
p.Asp3426Asn	WBGene00007028	trr-1	0.263979	transcription/transformation domain-associated protein
p.Glu293Lys	WBGene00008136	C47D12.2	0.263979	ortholog of human FLJ20071/FLJ90130
p.Ser47Gly	WBGene00012149	VF13D12L.3	0.263979	ortholog of human argininosuccinate synthetase
p.Ile245Met	WBGene00012395	Y6D1A.1	0.263979	not available
p.His211Tyr	WBGene00003979	pes-5	0.263979	patterned expression site

Table E6. Candidate genes for amino acid derivatives (all missense variants, Y55F3AM.13 also has a splice variant)

Amino acid change	Gene ID	Gene name	Spearman correlation coefficient	Description
p.Lys212Glu	WBGene00001103	dsl-1	0.321986	Delta/Serrate/Lag-2 domain
p.Gly126Ala	WBGene00002278	lep-2	0.295856	Leptoderan male tail
p.Arg236Gln	WBGene00021926	Y55F3AM.9	0.295856	predicted deubiquitinase activity
p.Ala45Val	WBGene00001104	dsl-2	0.28862	Delta/Serrate/Lag-2 domain
p.Lys309Asn	WBGene00005905	srx-14	0.279332	serpentine receptor, class X
p.Ala777Thr	WBGene00018776	F53H1.1	0.276589	ortholog of human DEAD-box helicase
p.Ser144Thr	WBGene00021121	W09G12.7	0.273648	not available
p.Asn83Ser	WBGene00009517	clec-167	0.269753	C-type lectin
p.Val447Ala	WBGene00009518	clec-166	0.265944	C-type lectin
p.Val44Leu	WBGene00021930	Y55F3AM.13	0.264118	not available
p.His52Gln	WBGene00022425	plst-1	0.254084	Plastin homolog

Table E7. Candidate genes for pantetheinyl-*N*-propionylhomocysteine (all missense variants)

Amino acid change	Gene ID	Gene name	Spearman correlation coefficient	Description
p.Ser136Ala	WBGene00022808	ZK697.8	0.693496	ortholog of human transthyretin
p.Asn177Lys	WBGene00017512	nhr-179	0.574537	nuclear hormone receptor
p.Val215Ile	WBGene00020340	T08B1.1	0.497930	predicted transmembrane transporter activity
p.Phe13Leu	WBGene00019388	K04F1.9	0.486187	not available
p.Thr37Pro	WBGene00021608	Y46H3D.1	0.336067	not available
p.Thr119Lys	WBGene00005604	srj-16	0.326107	serpentine receptor, class J
p.Pro96Ser	WBGene00044516	cc	0.321145	not available
p.Leu92Ser	WBGene00002018	hsp-16.41	0.288308	heat shock protein
p.Lys7Arg	WBGene00021486	lbp-9	0.278289	Lipid binding protein
p.Gly88Asp	WBGene00018098	F36F12.7	0.255235	not available
p.Lys326Arg	WBGene00006252	str-221	0.251544	seven TM receptor

R Code

XCMS for profile mode data

```
source("http://bioconductor.org/biocLite.R")
biocLite("mzR")
biocLite("xcms")
biocLite("MSnbase")
library(xcms)
library(MSnbase)
setwd("C:/Users/op56/Desktop/CNV profile neg/") #set working directory here
mzxml_pos <- list.files(".", pattern="mzXML", recursive = TRUE)
xr<-xcmsRaw(mzxml_pos[1])
xr
names(attributes(xr))
xset.raw.pos <- xcmsSet(mzxml_pos, step=0.1, fwhm=4, mzdiff=0.01, snthresh=10, max=100)
xset.pos <- group(xset.raw.pos, minfrac=0.5, bw=5, mzwid=0.025, max=100, minsamp=1)
xset2.pos <- retcor(xset.pos, method="obiwarp", profStep=1, plottype = "deviation")
xset.pos <- group(xset2.pos, minfrac=0.5, bw=5, mzwid=0.025, max=100, minsamp=1)
xset2.pos <- retcor(xset.pos, method="obiwarp", profStep=1, plottype = "deviation")
xsg.pos <- group(xset2.pos, minfrac=0.5, bw=5, mzwid=0.025, max=100, minsamp=1)
xsg.pos <- fillPeaks(xsg.pos)
xsg.pos
names(attributes(xsg.pos))
peaks(xsg.pos)[1:10, ]
table <- peakTable(xsg.pos)
write.csv(table, file="Cel_Nat_Var.csv")
```