Assembling the Puzzle of an Enigmatic Shark Tapeworm: a Comparative Genomic and Transcriptomic Approach

Kaitlin Gallagher

University of Connecticut - Storrs, kaitlin.gallagher@uconn.edu

Follow this and additional works at: https://opencommons.uconn.edu/dissertations

Recommended Citation
https://opencommons.uconn.edu/dissertations/2252
Assembling the Puzzle of an Enigmatic Shark Tapeworm: a Comparative Genomic and Transcriptomic Approach

Kaitlin Gallagher, Ph.D.

University of Connecticut, 2019

The evolution of novelty among closely related taxa is a phenomenon that has occurred repeatedly within parasitic clades, however, the mechanisms that underlie the evolution of such innovations is not fully understood. This dissertation explores one example of this phenomenon that occurs in a monogeneric order of tapeworms, the Litobothriidea, that infect lamniform sharks. The most recently described member of this group, *Litobothrium aenigmaticum*, lacks all the characters typically demonstrated by members of this group, yet it robustly nests within the genus *Litobothrium* with the barcoding gene 28S rDNA (D1-D3). The aim of this dissertation was to investigate which mechanisms may have allowed for the bizarre morphology and anatomy of *L. aenigmaticum*. To do this, I first needed to better understand the differences between *L. aenigmaticum* and the typical litobothrindaeans. Therefore, in chapter 1 I used transmission electron microscopy to characterize the internal anatomy of *L. aenigmaticum*. This work revealed that there are 11 novel cell types within the anterior region of the scolex of *L. aenigmaticum*, all of which contain secretory vesicles; nothing similar to this structure has been observed in any other cestode. With a more complete picture of the novelty of *L. aenigmaticum*, I next needed to generate the resources necessary to examine the mechanisms that led to the evolution of this species. In chapter 2 I assembled and annotated genomes for *L. aenigmaticum, L. daileyi,* and *L. amplifica* (338–406 Mb in size) and assembled transcriptomes for *L. aenigmaticum, L. daileyi,* and *L. nickoli*. Finally, in chapter 3 I used the genomic resources to perform a syntenic and gene family evolution analysis and the transcriptomic resources to perform a differential expression analysis. These analyses revealed that gene family expansions and contractions, differential regulation of translation, up-
and downregulation of specific transcripts, and coding region changes may underlie the novelty of *L. aenigmaticum*. In the future, I hope to broaden this study by also examining whether co-option and/or developmental changes may also be involved in this system.
Assembling the Puzzle of an Enigmatic Shark Tapeworm: a Comparative Genomic and Transcriptomic Approach

Kaitlin Gallagher

B.A., Florida Atlantic University, 2013

A Dissertation
Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy
at the
University of Connecticut
2019
APPROVAL PAGE

Doctor of Philosophy Dissertation

Assembling the Puzzle of an Enigmatic Shark Tapeworm: a Comparative Genomic and Transcriptomic Approach

Present by:

Kaitlin A. Gallagher, B.A.

Major Advisor:

__________________________________________
Janine N. Caira

Associate advisor:

__________________________________________
Marie Cantino

Associate advisor:

__________________________________________
Elizabeth L. Jockusch

Associate advisor:

__________________________________________
Jill Wegrzyn

University of Connecticut
2019
Acknowledgements

This dissertation was only possible due to the help and support of many wonderful individuals. I first need to recognize my advisor, Janine N. Caira, who has been incredibly supportive as I continually explored new methods. She has encouraged me every step of the way and she has helped me grow as a scientist, writer, teacher, and leader. I was very fortunate to have the chance to work with and learn from her and I am incredibly grateful for all the opportunities she has provided me. I was also fortunate to have a remarkable committee, all of whom were always incredibly enthusiastic about my project. Marie Cantino taught me everything I know about transmission electron microscopy; the first chapter of my dissertation would not be possible without her. Jill Wegrzyn taught me the theory and methods of genomics and spent many hours with me talking through study designs and troubleshooting coding problems. Her guidance made the third chapter of my dissertation possible. Elizabeth Jockusch taught me about molecular biology, graciously allowed me to use lab space, and spent hours talking to me about molecular and genomic methods and data visualization. She also was always willing to assist me with my writing and she challenged me to look at things differently, making me a better writer and scientist. I also want to thank David Wagner. He may not have been a formal member of my committee but he was always happy to help me with my dissertation. I greatly appreciate the time and thought that he put into my project over the years.

Thank you to the Bioscience Electron Microscopy Laboratory for the access to equipment and support. In particular, I need to thank Stephen Daniels and Xuanho Sun, who both assisted me with the first chapter of my dissertation. I also want to thank the Computational Biology Core, in particular Stephen King, Michael Wilson, Vijender Singh, and Neranjan Perera, for the computational support they provided. Thank you to Kirsten Jensen, Kaylee Herzog, and Hsaun-
Chieng Ho for assisting with the collection of specimens. I also owe many thanks to Cera Fisher, Hannah Ralicki, Sumaira Zaman, Alex Trouern-Trend, Madison Caballero, Nasim Rahmatpour, and Katie Taylor for their willingness to help me with all my coding woes. I also am extremely grateful to Elizabeth Barbeau for all her assistance with imaging and data basing and for always being willing to give me feedback on my presentations.

Many thanks to the wonderful Caira lab, past, present, and extended! I have had many adventures with the Caira lab that I will never forget! Veronica Bueno, Maria Pickering, Jimmy Bernot, Pancho Concha, and Doug Stephan were the best labmates I could have asked for. My discussions with them allowed me to become not just a better a scientist but also a better teacher. I also would like to thank Florian Reyda and Anna Phillips for their support as I figured out the next step in my career.

Many thanks to my fellow EEB graduate students and to the EEB department. I was very fortunate to be able to complete my dissertation in such an encouraging and collegial environment! In particular, I would like to thank Lauren Stanley, Tanisha Williams, Kristen Nolting, and Katie Taylor for their enthusiastic support. My discussions and interactions with them have made me a better scientist and I look forward to collaborating with them in the future.

I need to thank my family, particularly my mother, Katherine Gallagher. There are many times when they believed in me more than I believed in myself. I cannot express how much I appreciate their love and encouragement. Finally, I need to thank my partner, James Aguilar. He has been my rock these past two years. Thank you for always listening, for being incredibly supportive of my career, and for always believing in me. Most of all, thank you for all the wonderful adventures.
Dedication

This dissertation is dedicated to my father, John Gallagher (1961–2011), who inspired my love of the ocean and set me on the path that led me to where I am today. He is greatly missed.
# TABLE OF CONTENTS

Approval Page.................................................................................................................. iii
Acknowledgements........................................................................................................ iv
Dedication......................................................................................................................... vi
List of Tables.................................................................................................................. ix
List of Figures................................................................................................................ xi
Overall Introduction...................................................................................................... 1

Chapter 1: The internal anatomy of *L. aenigmaticum*......................................................... 8
  Introduction...................................................................................................................... 8
  Materials and Methods................................................................................................. 10
  Results............................................................................................................................ 12
  Discussion..................................................................................................................... 17

Chapter 2: *Litobothrium* genomic evolution................................................................. 28
  Introduction...................................................................................................................... 28
  Materials and Methods................................................................................................. 32
  Results............................................................................................................................ 46
  Discussion..................................................................................................................... 70

Chapter 3: Determining the mechanisms behind the novelty of *L. aenigmaticum*........... 85
  Introduction...................................................................................................................... 85
  Materials and Methods................................................................................................. 87
  Results............................................................................................................................ 91
  Discussion..................................................................................................................... 105
LIST OF TABLES

Table 1. Characteristics of the 11 new cell types in the cephalic peduncle of *Litobothrium aenigmaticum*.................................140

Table 2. Statistics on transcriptomic raw reads.................................................................141

Table 3. Statistics on the Trinity *de novo* assemblies for 14 individual litobothriidean transcriptomes........................................142

Table 4. Statistics on the clustered Trinity *de novo* assemblies for 14 individual litobothriidean transcriptomes........................................143

Table 5. BUSCO Completeness scores for the Trinity *de novo* assemblies....................144

Table 6. The statistics on the clustered, master transcriptomes........................................145

Table 7. The BUSCO completeness scores for the clustered, master transcriptomes........145

Table 8. The statistics on the raw sequencing reads for each of the genomic libraries.......146

Table 9. Statistics on the genomic reads after they have undergone trimming..............146

Table 10. Genome size estimation with the program Jellyfish........................................147

Table 11. Genome size estimation with BBMap...............................................................147

Table 12. SOAPdenovo assemblies without mate pair libraries.......................................148

Table 13. SOAPdenovo assemblies with mate-pair libraries............................................149

Table 14. ABySS2 assemblies without mate pair libraries..............................................150

Table 15. ABySS2 assemblies with mate-pair libraries..................................................151

Table 16. SPAdes assembly statistics...............................................................................152

Table 17. SPAdes assemblies scaffolded with master transcriptomes using tranScaff.......152

Table 18. MaSURCA assembly statistics.........................................................................153

Table 19. MaSURCA assemblies scaffolded with master transcriptomes using tranScaff...153
Table 20. The statistics for genome assemblies selected as final assemblies.....................154
Table 21. MAKER Annotation statistics for L. aenigmaticum.......................................154
Table 22. MAKER Annotation statistics for L. daileyi..................................................155
Table 23. MAKER Annotation statistics for SPAdes assembly of L. amplifica..............155
Table 24. MAKER Annotation statistics for MaSURCA assembly of L. amplifica..........156
Table 25. Braker annotation statistics for L. aenigmaticum.......................................156
Table 26. Braker annotation statistics for L. daileyi..................................................157
Table 27. GMAP annotation statistics........................................................................157
Table 28. Final annotation gene model statistics.........................................................158
Table 29. Genome annotation gene model statistics for multiexonic genes...............158
Table 30. Genome annotation gene model statistics for monoexonic genes...............159
Table 31. Genome annotation gene model statistics....................................................159
Table 32. Number of genes in gene families associated with hypotheses of Tsai et al. (2013)..160
LIST OF FIGURES

Figure 1. Basic body of a typical tapeworm illustrated on *Echinococcus granulosus*.........161

Figure 2. Basic body plan of a typical litobothriidean, *L. nickoli* (A), and the unusual *L. aenigmaticum* (B).................................................................162

Figure 3. Phylogenetic tree of litobothriidean relationships from Caira et al. (2014a)........163

Figure 4. Comparison of larval and adult forms of *Litobothrium daileyi* (A) and *Litobothrium aenigmaticum* (B).........................................................164

Figure 5. Phylogenetic tree of litobothriidean relationships from Caira et al. (2017)........165

Figure 6. Schematic of the four “tissue” types described by Caira et al. (2014a)..............166

Figure 7. Schematic of the location and associations of each of the 11 cell types in a lateral section of the scolex of *L. aenigmaticum*........................................167

Figure 8. Transmission electron micrographs of cross sections of cell type A, B, and C of *Litobothrium aenigmaticum*. .....................................................168

Figure 9. Transmission electron micrographs of cross sections of cell types D, E, and F of *Litobothrium aenigmaticum*.............................................................169

Figure 10. Transmission electron micrographs of cross sections of cell types G, H, and I of *Litobothrium aenigmaticum*.........................................................170

Figure 11. Transmission electron micrographs of cross sections of cell types J and K and the excretory system of *Litobothrium aenigmaticum*.................................171

Figure 12. Transmission electron micrographs of the tegument of *L. aenigmaticum*..........172

Figure 13. Periodic acid-Schiff (PAS) stained sections of different levels (A–C) of the cephalic peduncle of *L. aenigmaticum*.......................................................173

Figure 14. Species tree that resulted from clustering the 14 individual litobothriidean
transcriptomes with Orthofinder…………………………………………………………174

Figure 15. Species tree from clustering the litobothriidean and *E. multilocularis* genome annotation gene models in Orthofinder…………………………………………………………175

Figure 16. Comparison of number of genes in gene families that were hypothesized to have flatworm specific expansions by Tsai et al. (2013)……………………………………176

Figure 17. Comparison of number of genes in gene families that were hypothesized to have expansions in trematodes and cestodes by Tsai et al. (2013)…………………………177

Figure 18. Comparison of number of genes in gene families hypothesized to be specific to tapeworms by Tsai et al. (2013). …………………………………………………178

Figure 19. Comparison of number of genes in gene families hypothesized to have tapeworm-specific expansions by Tsai et al. (2013)……………………………………179

Figure 20. Comparison of number of genes in gene families hypothesized to be expanded in the *Taenia* genomes by Tsai et al. (2013)……………………………………180

Figure 21. Comparison of number of genes in gene families hypothesized to have species-specific expansions by Tsai et al. (2013)……………………………………181

Figure 22. Revigo biological process treemap for litobothriidean-specific orthogroups with three different species backgrounds…………………………………………………………182

Figure 23. Revigo cellular component treemap for litobothriidean-specific orthogroups with three different species background………………………………………………183

Figure 24. Revigo molecular function treemap for litobothriidean-specific orthogroups with three different species backgrounds………………………………………………184

Figure 25. Revigo treemaps for *Litobothrium daileyi*-specific orthogroups with three different species background…………………………………………………………185
Figure 26. Hierarchical correlation matrix of 14 transcriptomes from DESeq2

Figure 27. Plot of first and second principal components from principal components analysis of 14 litobothriidean transcriptomes.

Figure 28. Plot of second and third principal components from principal components analysis of 14 litobothriidean transcriptomes.

Figure 29. Clustering of 1,000 most expressed genes from DESeq2 analysis of 14 litobothriidean transcriptomes.

Figure 30. Biological process treemap for transcripts down regulated in L. aenigmaticum with three different species backgrounds.

Figure 31. Cellular component treemap for transcripts down regulated in L. aenigmaticum with three different species backgrounds.

Figure 32. Molecular function treemap for transcripts down regulated in L. aenigmaticum with three different species backgrounds.

Figure 33. Biological process treemap for transcripts downregulated in L. daileyi with three different species backgrounds.

Figure 34. Cellular component treemap for transcripts downregulated in L. daileyi with three different species backgrounds.

Figure 35. Molecular function treemap for transcripts downregulated in L. daileyi with three different species backgrounds.

Figure 36. Biological process treemap for transcripts downregulated in L. nickoli with three different species backgrounds.

Figure 37. Chemical component treemap for transcripts downregulated in L. nickoli with three different species backgrounds.
Figure 38. Molecular function treemap for transcripts downregulated in *L. nickoli* with three
different species backgrounds.................................................................198

Figure 39. Biological process treemap for transcripts upregulated in *L. nickoli* with three
different species backgrounds.................................................................199

Figure 40. Cellular component treemap for transcripts upregulated in *L. nickoli* with *L. aenigmaticum* and *L. daileyi* backgrounds.................................................................200

Figure 41. Molecular component treemap for transcripts upregulated in *L. nickoli* with three
different species backgrounds.................................................................201

Figure 42. Dot plot resulting from SynMap comparing *L. aenigmaticum* and *L. daileyi*........202

Figure 43. Dot plot resulting from SynMap comparing *L. aenigmaticum* and *L. amplifica*........203

Figure 44. Dot plot resulting from SynMap analysis comparing *L. amplifica* and *L. daileyi*.....204

Figure 45. Histograms of synonymous substitution rates, nonsynonymous substitution rates, and
nonsynonymous to synonymous substitution rates ratio................................205

Figure 46. Transformed species tree from ETE3 Toolkit used for CAFE analysis..................206

Figure 47. Phylogeny resulting from CAFE analysis........................................207

Figure 48. Biological process, cellular component, and molecular function GO terms associated
with gene families CAFE found to be expanded in *Litobothrium aenigmaticum*........208

Figure 49. Biological process, cellular component, and molecular function GO terms associate
with gene families CAFE found to be contracted in *Litobothrium aenigmaticum*........209

Figure 50. Biological process, cellular component, and molecular function GO terms associated
with gene families CAFE found to be expanded in *Litobothrium daileyi*.................210

Figure 51. Biological process, cellular component, and molecular function treemaps for gene
families CAFE analysis found to be rapidly expanding in *L. amplifica*...............211
Figure 52. Biological process, cellular component, and molecular function treemaps for gene families CAFE analysis found to be rapidly contracted in *L. amplifica*...

Figure 53. Biological process tree map for gene families CAFE analysis identified as significantly rapidly expanding in both *L. aenigmaticum* and *L. daileyi*...

Figure 54. Cellular component treemap for gene families CAFE analysis identified as significantly rapidly expanding in both *L. aenigmaticum* and *L. daileyi*...

Figure 55. Molecular function treemap for gene families CAFE analysis identified as significantly rapidly expanding in both *L. aenigmaticum* and *L. daileyi*...

Figure 56. Biological process, cellular component, and molecular function treemaps for gene families CAFE analysis found to be contracted in all three litobothriideans...

Figure 57. Biological process and cellular component treemaps for gene families CAFE analysis found to be expanded in *E. multilocularis*...

Figure 58. Biological process treemaps for gene families CAFE analysis found to be contracted in *E. multilocularis*...

Figure 59. Cellular component treemaps for gene families CAFE analysis found to be contracted in *E. multilocularis*...

Figure 60. Molecular function treemaps for gene families CAFE analysis found to be contracted in *E. multilocularis*...

Figure 61. Significantly enriched GO terms from DESeq2 analysis...

Figure 62. Significantly enriched GO terms from CAFE analysis...
Overall Introduction

Parasitism is one of the most successful life style strategies in the animal kingdom, evidenced by the fact that approximately 40% of currently known species are parasitic (Dobson et al., 2008). The evolution from a free-living to a parasitic life style within metazoan animals alone has independently occurred at least 60 times; if the protozoans were included this number would be substantially higher (Poulin and Morand, 2000). Along with this switch to parasitism have come suites of novel structures and adaptations that allow these species to attach to a host and, for endoparasitic species, to survive the harsh environment within a host’s body. Lineages that evolve parasitism tend to undergo remarkable radiations so often parasitism occurs in clades (Poulin and Morand, 2000). The members of such radiations tend to be morphologically very similar to one another since they all typically display the same types of adaptations. However, these adaptations often result in the morphology of the parasitic species diverging substantially from that of their nonparasitic relatives. Examples of this are seen throughout the Platyhelminthes, a phylum in which three of the four classes are entirely parasitic.

It is, however, rare that substantially different body plans are seen among taxa within closely related groups of parasites. Examples of this phenomenon are key to understanding the evolution of novelty in shorter time frames. This study focuses on one such example and what it can tell us about some of the factors that might account for major changes in body form among closely related species.

The focal taxa of this study are the litobothriidean tapeworms that parasitize the pelagic thresher shark, Alopias pelagicus Nakamura, 1935. The basic body of a typical tapeworm consists of two major regions: the scolex and the strobila (Fig. 1). The anterior region of the tapeworm body is referred to as the scolex; it typically bears structures that allow the worm to attach to the
A great diversity of types and forms of attachment structures are found on the scoleces across the various groups of tapeworms. These include bothria, acetabula of a variety of forms (e.g., suckers, bothridia), hooks, and tentacles that bear hooks. The posterior region of the tapeworm body is referred to as the strobila. This structure consists of a chain of proglottids, each of which is essentially a compartment filled with reproductive organs. Since tapeworms are hermaphroditic, each proglottid contains both male and female reproductive organs. As of 2014, all eight described species of the Litobothriidea had been reported to exhibit the basic tapeworm body plan (Fig. 2A). The scolex of these eight species consists of an apical sucker and three to six cruciform pseudosegments; all lack the neck region of the scolex seen in some cestodes referred to as a cephalic peduncle (Fig. 2A) (Dailey, 1969; Dailey, 1971; Kurochkin and Slankis, 1973; Caira and Runkle, 1993; Olson and Caira, 2001). The litobothriidean strobila consists of 50 to 70 hermaphroditic proglottids that range in length from 0.5 to 1.5 mm. All eight species are euapolytic, in that their proglottids detach when mature (Dailey, 1969; Dailey, 1971; Kurochkin and Slankis, 1973; Caira and Runkle, 1993; Olson and Caira, 2001).

However, in 1996 some intrigue developed that would be found to be relevant to this genus. As part of a survey of the tapeworms of sharks and rays in the Gulf of California, Caira and collaborators discovered multiple specimens of tapeworms in the spiral intestines of pelagic thresher sharks that were unlike any known members of the nine orders of tapeworms known to parasitize elasmobranchs. These unusual cestodes bore a dome-like anterior region of the body (Fig. 2B) that was followed by an extremely long cephalic peduncle and then an undivided (i.e., unproglottized) region of the body (Fig. 2B). A small subset of the specimens was found to bear a terminal chain of up to 20 tiny (i.e., 50–60 µm in length), immature proglottids (Fig. 2B).
Furthermore, the cephalic peduncle was found to contain an odd collection of “tissues” of unknown origin (Caira et al., 2014a).

For almost two decades, these enigmatic specimens were circulated among tapeworm taxonomic experts around the globe for opinions on their identity. Ultimately, community consensus was that they resembled no known order of tapeworms and thus, given their host, must belong to one of the orders that parasitizes elasmobranchs. As a result, in 2013 Caira and collaborators formally took up the challenge of identifying the phylogenetic affinities of these specimens.

Fortunately, by 2013, molecular phylogenetics had advanced the point at which molecular data could be employed to help inform those efforts. In 2014(a), Caira et al. reported on their results of phylogenetic analyses of partial (D1–D3) 28S rDNA sequence data, which revealed that these bizarre cestodes nested deeply among litobothriidean species, as the sister taxon to Litobothrium nickoli Olson and Caira, 2001 (Fig. 3), which also parasitizes the pelagic thresher shark. Concerned that these unusual worms might merely represent the juvenile stage of one of the three previously described species of litobothriideans from the pelagic thresher shark previously (Kurochkin and Slankis, 1973; Olson and Caira, 2001), Caira et al. (2014a) also characterized the early juvenile stages of typical litobothriideans for the first time. However, Caira et al. (2014a) found that like juveniles of other elasmobranch-hosted cestodes groups (Jensen and Bullard, 2010), the juveniles of Litobothrium daileyi Kurochkin and Slankis, 1973 and Litobothrium janovyi Olson and Caira, 2001 exhibited features (i.e., the cruciform pseudosegments) typical of the scoleces of the adult forms (Figs. 4A, B). No evidence of the unusual morphology of the bizarre worm was seen. Although juveniles that might be attributed to this unusual form were not yet known, this evidence seemed sufficiently convincing to eliminate the idea that this form was a developmental
stage of one of the other species of *Litobothrium*. In combination with evidence from scanning electron microscopy, Caira et al. (2014a) used these molecular and developmental data to justify description of these worms as the new species *Litobothrium aenigmaticum*—so named in reference to its enigmatic form. In that same paper they characterized what they referred to as four distinct “tissue” types within the cephalic peduncle of these worms.

The addition of *L. aenigmaticum* to the order Litobothriidea was unsettling for a number of reasons, none the least of which was the fact that no satisfactory explanation for what might have led to the unconventional morphology of this species related to its congeners has been identified. The primary goal of this dissertation was to explore some of the mechanisms that might account for the evolution of the unusual body plan of *L. aenigmaticum*. The approach taken was to expand the understanding of the morphology, development, and genetics of this species in a comparative framework, that ideally included the sister taxon of *L. aenigmaticum*. But, the phylogenetic analyses of Caira et al. (2014a) included only four of the nine species of *Litobothrium*, and only three of the four species that parasitize the pelagic thresher shark. As a consequence, the sister taxon relationship they found between *L. aenigmaticum* and *L. nickoli* required confirmation. In 2017, Caira et al. expanded their analyses to include 28S rDNA (D1-D3 region) data for *Litobothrium daileyi* Kurochkin and Slankis, 1973, the fourth species that parasitizes the pelagic thresher shark. The resulting tree suggests that *L. daileyi*, rather than *L. nickoli*, is the sister taxon to *L. aenigmaticum* (Fig. 5); in fact, they found *L. aenigmaticum* and *L. daileyi* to be identical in sequence for this particular gene region. Given that this gene region is generally considered to differ at least somewhat between even closely related species, this result was puzzling. Although, it is possible that if the divergence between *L. aenigmaticum* and *L.
daileyi occurred so recently that this region does not reflect this divergence, the use of additional, more rapidly evolving gene markers is required to evaluate the relationship more thoroughly.

Any lingering concern that L. aenigmaticum may represent a juvenile form of one of the other members of the genus that parasitizes the pelagic thresher shark was eliminated in 2017, when Caira et al. described the juvenile form of L. aenigmaticum. Similar to the juvenile forms of typical litobothriideans such as L. daileyi (Fig. 4A), the scolex of the juvenile of L. aenigmaticum was found to exhibit characteristics of the adult scolex (i.e., it bore a dome-shaped scolex) (Fig. 4B).

It should be noted that some have suggested that L. aenigmaticum is not a separate species from L. daileyi and that instead the unusual morphology of these specimens is due to phenotypic plasticity. However, this explanation seems unlikely since these two species are found not only in the same host species, A. pelagicus, but have been repeatedly found side by side within the spiral valve of the same host individual. Therefore, they are not experiencing different environmental pressures. Furthermore, the discovery of juvenile forms of L. aenigmaticum indicate that this species already bears aspects of the bizarre scolex morphology when it enters its final host (Caira et al., 2017).

The morphological and molecular incongruencies in this system offer a unique opportunity to explore the factors that underlie dramatic changes in morphology within a group of closely related taxa. To develop a better understanding of morphological novelty in this system, a more thorough understanding of the anatomy of L. aenigmaticum was required. The anatomy of L. aenigmaticum was characterized using transmission electron microscopy in chapter one. Also necessary was a more detailed knowledge of the genetic differences among L. aenigmaticum and its close relatives. The processes of assembling, annotating, and comparing the litbothriidean
genomes are described in chapter two. Using the results of these two chapters, chapter three examines the mechanisms that may have led to the evolution of the novel body form of *L. aenigmaticum*.

Understanding the origin and evolution of novelty is one of the fundamental questions in evolutionary biology. However, this field is rife with disagreement. There is disagreement about everything from the definition of novelty to the various kinds of mechanisms that allow for the evolution of such innovations. Definitions of novelty generally fall into two main categories. The first relates to function, while the second pertains to structure. Definitions relating to function consider a novelty to be a structure that allows an organism to gain a new function (Mayr, 1960). In contrast, definitions that relate to structure state that a novelty must either be a structure that differs conspicuously from that of the ancestral state (Müller, 1990), or that has no homolog in the ancestral species (Müller and Wagner, 1991). The two general types of changes that have been theorized to lead to the evolution of morphological novelty are genetic and developmental (Müller and Wagner, 1991; Carroll et al., 2005; Wagner and Lynch, 2010). Examples of candidate genetic changes are structural gene mutations, chromosomal rearrangements, genome size change, regulatory mutations, gene duplications, transposable element insertions, and the evolution of novel genes (Müller and Wagner, 1991; Lynch and Conery, 2000; Carroll et al., 2005; Wagner and Lynch, 2010). Examples of candidate developmental changes are heterochrony (changes in the timing and rate of ontogenetic processes), ontogenetic re-patterning, changes in epigenetic cascades, changes in reaction norms, and co-option (Müller and Wagner, 1991; Carroll et al., 2005; Glassford et al., 2015). Although it is agreed that all the above changes could theoretically lead to the evolution of novelty, there is much debate regarding which of these often actually lead to the origination of novelty. Some argue that novelty is more likely to arise through changes in coding
regions (Hoekstra, 2006; Hoekstra and Coyne, 2007), while others argue it is often the result of gene duplications (Lynch and Conery, 2000; 2003), and still others believe that it is most likely changes in gene regulation that underlie this phenomenon (Wagner, 2007; Carroll, 2008). While it is beyond the scope of this dissertation to examine all possible mechanisms that might account for novelty in *L. aenigmaticum*, in chapter three I explore the genetic changes that may have contributed to the novel morphology of this exceptional litobothriidean.

This dissertation has three primary goals: (1) to characterize the ultrastructure of the cells that make up the novel anatomy of *L. aenigmaticum*, (2) to examine the genomic evolution of the litobothriidean tapeworms, and (3) to explore the mechanisms that may have led to the evolution of the unique morphology and anatomy of *L. aenigmaticum*. In addition to these, my research yielded new genomic and transcriptomic resources for studying the evolution and biosystematics of cestodes, and enhanced our understanding of the genetic mechanisms underlying novelty among closely related species.
**Chapter 1: The internal anatomy of *L. aenigmaticum***

Published in the journal of Invertebrate Biology by Gallagher et al. (2017)

**Introduction**

The litobothriidean tapeworm *Litobothrium aenigmaticum* exhibits a body plan unlike any other litobothriidean tapeworm. Not only does it lack all of the morphological characteristics that are typical of this order, it also seems to diverge from its congeners in regards to its anatomy. In an attempt to understand the morphology of *L. aenigmaticum*, Caira et al. (2014a) examined whole mounts and histological sections using light microscopy. Based on that work they reported a series of four unusual types of “tissues” within the scolex (Fig. 6). “Tissue” type one, located in the scolex proper and anterior region of the cephalic peduncle, consisted of large cells with small nuclei, a large cytoplasm to nucleus ratio, and numerous inclusions. They found this “tissue” to stain positive with Periodic acid-Schiff (PAS), and thus suggested that it contains mucopolysaccharides and may be responsible for producing material that assists with adhesion to the host mucosal surface. They reported “tissue” type two in the middle of the cephalic peduncle, partially surrounded by “tissue” type one. This “tissue” consisted of moderately sized cells with large nuclei, and a cytoplasm that is possibly vacuolated. “Tissue” type three was found posterior to, and partially surrounding, “tissue” type two and consisted of small, densely packed cells with small nuclei occupying half the area of the cytoplasm. The fourth “tissue” type was located in the posterior region of the cephalic peduncle partially surrounding “tissue” type three. This final “tissue” consisted of moderately sized cells with large nuclei that occupy a fourth of the area of the cytoplasm. Caira et al. (2014a) refrained from attributing functions to the last three “tissue” types. Caira et al. (2014a) also reported two sublateral pairs of regularly perforated sinuous ducts, surrounded by dense cells extending throughout the entire length of the worm and were similarly
unable to assign a function to these ducts. None of these features have been observed in the typical litobothriidean species, in fact such features do not appear to have been reported in any other tapeworms!

The primary aim of this chapter was to examine the ultrastructure of the anatomy of the scolex of *L. aenigmaticum* in more detail in the hope of shedding some additional light on the structure and function(s) of these cell types and the sublateral pairs of ducts. In order to achieve this, three specimens were examined with transmission electron microscopy. This work allowed for the identification and characterization of cell types. Light microscopy was then used to determine whether each of the identified cells types was PAS positive and to gain a better understanding of the location of each cell type.

This work revealed that the internal anatomy of *L. aenigmaticum* is even more complex than originally thought based on light microscopy alone and that there may be up to 11 distinct cell types in the scolex proper and cephalic peduncle of this species. These cell types and ducts and their potential function(s) are described in detail below. It should be noted that the terminology used here deviates from that used by Caira et al. (2014a) in two major respects. First, whereas they referred to the four aggregates of cells as “tissues,” here they are referred to as cell types. This is because a tissue is typically made up of several different cell types that work together to perform a specific function. The present work revealed aggregates of cells that share the same ultrastructural characteristics, rather than collections of different cell types and thus use of cell types is more appropriate. Second, since many more than four cell types were found in the course of this study, the exact correspondence between these cell types and the four “tissues” reported by Caira et al. (2014a) is difficult to assess. Therefore the 11 cell types described here are referred to with the letter designations A–K instead of the number system used by Caira et al. (2014a).
Materials and Methods

Cestode specimens were collected in November of 2013 from a female pelagic thresher shark, *Alopias pelagicus*, in Chenggong, Taiwan. Additional information on the host (No. TW-102) is available in the Global Cestode Database (www.tapewormsdb.uconn.edu) (Caira et al., 2018). The spiral intestine was removed, opened with a medial longitudinal incision and examined under a dissecting microscope. Of the specimens of *L. aenigmaticum* collected, those for transmission electron microscopy (TEM) were fixed in a solution of 1.5% paraformaldehyde, 1.5% glutaraldehyde, 0.1M HEPES, 0.08M sodium chloride, and 3mM magnesium chloride at pH 7.3; others were fixed in 10% formalin buffered in seawater for light microscopy.

Transmission Electron Microscopy

Specimens for TEM were stored in fixative for approximately two weeks. They were then washed in 0.1M HEPES with 0.08M sodium chloride and 3mM MgCl three times and post-fixed in 1% OsO₄ and 0.8% K₃Fe(CN)₆ in the same buffer. They were dehydrated in a graded ethanol series and placed in epoxy resin overnight (Araldite 506 epoxy resin, Embed 812 epoxy resin, dodecyl succinic anhydride, and 1.5% DMP30). Each specimen was then cut into three pieces (scolex proper, anterior cephalic peduncle, and posterior cephalic peduncle) and photographed using a Zeiss 47 50 52 – 9901 Stereo Zoom microscope and Canon EOS 700D camera. Each piece was then placed in a flat mold and oriented for either cross (two specimens) or longitudinal (one specimen) sectioning. Blocks were subsequently polymerized for 36 hours at 60ºC, and sectioned using a Leica UTC ultramicrotome. A diamond knife was used to cut thick (1 μm) and thin (90 nm) sections from each block. Two to four thick sections were taken before every thin section to serve as a macroscopic reference point. Thick sections were placed on Fisherbrand SuperFrost plus slides and stained with Methylene Blue Azure II. The first two specimens were sectioned at
intervals; in each case the number of thin and thick sections taken was recorded so that the position of the section along the length of the scolex could be determined. The third specimen was sectioned serially; a thick section was taken every 15 µm to gauge which type of cells were present. A thin section was taken at least every 30 µm along the length of the specimen. Thick sections were mounted on glass slides and covered with Fisherfinest Premium coverslips. Thin sections were placed on 200/300 mm mesh grids and stained with 2% uranyl acetate and Sato’s lead citrate. They were examined using a Tecnai 12 Biotwin TEM operated at 80 kV and equipped with an AMT 2k XR40 CCD camera.

**Light Microscopy**

After approximately two weeks, specimens for light microscopy were transferred to 70% ethanol for storage. Two specimens were prepared for histological sectioning as follows. They were dehydrated in a graded ethanol series, cleaned in xylene, transferred to a 1:1 mixture of xylene and paraffin in an oven for an hour, and then transferred to paraffin overnight. Specimens were embedded in paraffin in blocks and allowed to cool overnight. Blocks were trimmed using a razor blade and serially sectioned at ~6 µm intervals using an Olympus CUT 4060 microtome. Approximately 30 sections were mounted per glass slide using sodium silicate. Every other slide in the series was stained with Delafield’s hematoxylin and eosin; the remaining slides were stained with PAS. This made it possible for sections of each of the 11 cell types to be stained with both staining protocols. Longitudinal TEM sections were used to confirm the position and association of the cell types that were inferred from light microscopy cross sections.

Slides of histological sections have been deposited in the Lawrence R. Penner (LRP) parasite collection at the University of Connecticut (Nos. 8925–8985). The terminology of Caira et al. (2014a) has been employed for regions of the body of *L. aenigmaticum*. Following Khalil et
al. (1994), the outer region of the body is referred to as the cortex and the region inside the longitudinal muscles as the medulla (Khalil, 1994).

Results

Cell Types

The scolex proper and cephalic peduncle of *L. aenigmaticum* were found to house a total of at least 11 distinct types of cells, here given the letter designations A–K. The arrangement of these cell types relative to one another throughout the length of the scolex and cephalic peduncle of *L. aenigmaticum* are illustrated in Figure 7. This figure was constructed using TEM cross and longitudinal sections as well as serial sections examine by light microscopy. The serial sections examined by light microscopy allowed us to establish the range and associations of each cell type distinguished in ultrathin sections examined by TEM. The distinguishing features of the 11 cell types are summarized in Table 1. The ultrastructure of each cell type is described separately below.

Cell Type A (Figs. 8A–C): This cell type consists of large (23.54 ± 4.6 μm, n=8), irregular cells with nuclei that are moderate in size (3.63 ± 1.06 μm, n=5) and often contain a circular, eccentric nucleolus (Fig. 8A). These cells have a high cytoplasm to nucleus ratio (1:7.62 ± 2.26, n=5). The cytoplasm is electron lucent and is densely packed with small vesicles of varying electron densities (Fig. 8B) as well as small circular, electron dense structures that resemble microvesicles (Fig. 8C). No mitochondria, rough endoplasmic reticulum, or Golgi apparatus were observed. The cell type A was found to be PAS positive. This cell type extends from near the anterior-most region of the scolex proper to approximately the mid-level of the cephalic peduncle. It occupies the medulla in the scolex proper and anterior regions of the cephalic peduncle, where it is the only cell type present. It extends in dorsal and ventral cortical sheets in the region of cell type B (Fig. 7).
Cell Type B (Figs. 8D–G): This cell type consists of moderately sized (11.78 ± 2.83 µm, n=10), irregular cells that are densely packed and contain moderately sized nuclei (3.2 ± 0.38 µm, n=10) with conspicuous marginal chromatin (Fig. 8D). These cells have a moderate cytoplasm to nucleus ratio (1:3.65 ± 0.72, n=10). The cytoplasm is electron lucent and contains numerous, small, mostly electron dense vesicles (Fig. 8F). Mitochondria, free ribosomes, and Golgi apparatus were seen in these cells (Fig. 8E), however no rough endoplasmic reticulum was observed. The majority of cell type B was found to be PAS positive, except that the cells occupying the central medullary region of these cells were found to be PAS negative; this region was also found to contain electron lucent inclusions (Fig. 8G). Cell type B begins approximately in the anterior fifth of the cephalic peduncle, and surrounds cell type C, D, and E. It extends approximately to the mid-level of the cephalic peduncle (Fig. 7).

Cell Type C (Figs. 8H–J): This cell type consists of small (10.05 ± 2.11 µm, n=6), densely packed, irregularly oval cells (Fig. 8H). The nuclei are moderately sized (3.24 ± 0.51 µm, n=6) and contain irregular patches of marginal chromatin (Fig. 8H). These cells have a small cytoplasm to nucleus ratio (1:3.1 ± 0.6, n=6). The cytoplasm is electron dense and contains some electron dense vesicles (Fig. 8H,I) and numerous electron lucent vesicles (Fig. 8H,J). Mitochondria, Golgi apparatus, and free ribosomes were visible in these cells (Fig. 8I,J). No endoplasmic reticulum was observed. Cell type C was found to be PAS positive. This cell type is restricted to a short region of the cephalic peduncle, essentially in the middle third of the length of cell type B. This cell type is medullary and surrounds cell types D and E (Fig. 7).

Cell Type D (Figs. 9A–C): This cell type consists of small (9.32 ± 1.74 µm, n=11), densely arranged cells with moderately sized nuclei (3.22 ± 0.25 µm, n=11) and a small cytoplasm to nucleus ratio (1:2.89 ± 0.64, n=11) (Fig. 9A). The cytoplasm is electron lucent and generally
contains small vesicles most of which are electron dense (Fig. 9A). Mitochondria, Golgi apparatus, and free ribosomes were seen in this cell type (Figs. 9B,C). No endoplasmic reticulum was observed. Cell type D was found to be PAS positive. This cell type is coincident in length with cell type B and occupies the medulla immediately surrounding cell type E. This cell type is surrounded by cell type B for its entire length and also by cell type C for the middle third of its length (Fig. 7).

**Cell Type E (Figs. 9D–F):** This cell type consists of densely arranged interdigitating cells with very small nuclei (2.14 ± 0.68 µm, n=4) (Fig. 9D). The full size of the cells was difficult to assess given their interdigitated configuration. The cytoplasm is packed with large electron dense vesicles interspersed with some electron lucent vesicles (Fig. 9D). Endoplasmic reticulum and microtubules that line the cell membranes were visible in the cytoplasm (Figs. 9E,F). No mitochondria or Golgi apparatus were observed. Cell type E was found to be PAS positive. This cell type occupies the central core of the medulla and extends slightly more anterior and slightly more posterior than cell types B and D (Fig. 7).

**Cell Type F (Figs. 9G–J):** This cell type consists of loosely arranged, moderately sized cells (13.24 ± 2.20 µm, n=15) with large nuclei (3.81 ± 0.78 µm, n=15) and a moderate cytoplasm to nucleus ratio (1:3.66 ± 1.04, n=15) (Fig. 9G). The cytoplasm is electron lucent and is loosely packed with electron dense vesicles of varying sizes and shapes (Fig. 9G). Mitochondria, Golgi apparatus, and endoplasmic reticulum were observed in these cells (Figs. 9H–J). Cell type F was found to be PAS positive. This cell type begins immediately posterior to cell type A, and extends in dorsal and ventral cortical sheets throughout much of the posterior half of the cephalic peduncle. In the posterior one-sixth of the cephalic peduncle it extends into the medulla to surround cell type I and then more posteriorly, cell type K (Fig. 7).
Cell Type G (Figs. 10A–D): This cell type consists of densely arranged, moderately sized cells (11.85 ± 2.35 μm, n=9) with large nuclei (3.89 ± 0.53 μm, n=9) and a small cytoplasm to nucleus ratio (1:3.09 ± 0.74, n=9) (Fig. 10A). The cytoplasm contains some electron dense vesicles and numerous, large electron lucent vesicles (Figs. 10A,C). Mitochondria and free ribosomes were seen in these cells (Fig. 10B,D). No rough endoplasmic reticulum or Golgi apparatus were observed. Cell type G was found to be PAS positive. This cell type begins at the posterior most extensions of cell type A. It occupies much of the medulla in the mid region of the cephalic peduncle and is reduced to occupy only the central core of the medulla in the posterior four-fifths of the cephalic peduncle (Fig. 7).

Cell Type H (Figs. 10E–G): This cell type consists of densely arranged, moderately sized cells (12.76 ± 1.87 μm, n=8) with moderately sized nuclei (3.22 ± 0.32 μm, n=8) and a small cytoplasm to nucleus ratio (1:3.44 ± 0.43, n=8) (Fig. 10E). The cytoplasm contains some electron dense vesicles (Fig. 10E). Mitochondria, Golgi apparatus, and a large amount of endoplasmic reticulum were seen in these cells (Figs. 10F,G). Cell type H was found to be PAS positive. This cell type extends throughout the length of the reduced region of cell type G and is medullary to cell type F (Fig. 7).

Cell Type I (Figs. 10H–J): This cell type consists of densely arranged, spindle-shaped cells (Fig. 10H) that are moderately long (12.28 ± 1.47 μm, n=10); the nuclei are small (2.94 ± 0.36 μm, n=10) and have a moderate cytoplasm to nucleus ratio (1:4.23 ± 0.73, n=10) (Fig. 10H). The cytoplasm is electron lucent and contains electron dense vesicles (Fig. 10H–J). Mitochondria, but no rough endoplasmic reticulum or Golgi apparatus were observed in these cells (Fig. 10I). Cell type I was found to be PAS positive. This cell type is medullary to cell types F and H. It
extends throughout much of the length of cell type H, and occupies the central medulla of the cephalic peduncle for a short region near its posterior end (Fig. 7).

**Cell Type J (Figs. 11A–B):** This cell type consists of small, spindle-shaped, densely arranged cells (9.35 ± 2.12 µm, n=13) with small nuclei (2.82 ± 0.41 µm, n=13) and a small cytoplasm to nucleus ratio (1:3.38 ± 0.91, n=13) (Fig. 11A). The cytoplasm contains numerous electron dense vesicles of various shapes and sizes (Fig. 11A). Mitochondria and numerous free ribosomes, but no endoplasmic reticulum or Golgi apparatus, were observed (Fig. 11B). Cell type J was found to be PAS positive. This cell type extends for much of the length of cell type H. It is medullary to cell types F, H, and I and surrounds cell type G in the latter’s posterior-most region (Fig. 7).

**Cell Type K (Figs. 11C–E):** This cell type consists of loosely arranged, moderately sized cells (14.51 ± 3.79 µm, n=9) with large nuclei (3.46 ± 0.59 µm, n=9) and a moderate cytoplasm to nucleus ratio (1:4.23 ± 0.95, n=9) (Fig. 11C). The cytoplasm contains both electron dense vesicles and some electron lucent vesicles (Figs. 11D,E). Mitochondria, Golgi apparatus, and free ribosomes, but no rough endoplasmic reticulum were observed in these cells (Fig. 11D). Cell type K was found to be PAS positive. This cell type replaces cell type I in the central medulla of the posterior most region of the cephalic peduncle and is surrounded by cell type F (Fig. 7).

*Paired Sublateral Ducts*

TEM revealed that the two sublateral pairs of sinuous ducts that extend through the entire length of the worm are elements of the excretory system. The position of the ducts in relation to the 11 cell types is illustrated in the cross sections presented in Figure 2. Details of the ducts and surrounding tissues are provided in Figures 11F–I. Ducts in a pair are unequal in diameter and are closely associated with one another. The larger of the ducts in each pair is approximately 7–12 µm
in diameter and is located more medullary while the smaller ducts in each pair, which are approximately 4–5 µm in diameter, are located more cortically. Ducts in each pair were found to be surrounded by dense layers of cells. The larger ducts were surrounded by a layer of cells approximately 3–5 µm thick; the smaller ducts were surrounded by a layer of cells approximately 2 µm thick. Each pair of ducts was found to be closely associated with numerous protonephridia (i.e., flame cells) and assorted collecting ducts. In several instances protonephridia were seen to connect to these collecting ducts (Fig. 11H). The layers of cells surrounding the excretory ducts stained negatively with PAS.

**Tegument**

The tegument of *L. aenigmaticum* is approximately 5–7 µm thick. The distal cytoplasm was found to be extremely densely packed with numerous small, electron dense inclusions that although rod-shaped in section, were revealed to be discs by tilting the TEM stage (Fig. 12A–C). Both electron dense and electron lucent vesicles were observed within the distal and perinuclear cytoplasm of the tegument (Fig. 12C–E). Some evidence of both types of vesicles, presumed to be moving towards the outer membrane of the tegument, was observed. The apex of the scolex bears an aperture that is surrounded by numerous pores. TEM revealed electron dense products in both the aperture and pores (Fig. 12J–M). It also appears that electron lucent products are being released from the aperture via apocrine release (Fig.12J).

**Discussion**

This TEM work has led to a much more thorough understanding of the cell type architecture and ultrastructure of *L. aenigmaticum* than first described by Caira et al. (2014a) using light microscopy alone. In some cases these structural details yielded clues to the function of the cell types. First, the two sublateral pairs of sinuous ducts are excretory in function. These ducts
were found to be surrounded by numerous protonephridia and their associated collecting ducts. This suggests that rather than separate dorsal and ventral excretory ducts on each side of the body, as is typical of other litobothriideans (Dailey, 1969; Dailey, 1971; Kurochkin and Slankis, 1973; Caira and Runkle, 1993; Olson and Caira, 2001), the two excretory ducts on each side of the body of *L. aenigmaticum* are intertwined with one another. What remains to be determined is the relationship between these sublateral excretory ducts and the subtegumental anastomosing network of canals postulated by Caira et al. (2014a) to be excretory in function. Although subtegumental anastomosing canals have not been reported in any of the other eight species of *Litobothrium*, an excretory system consisting of an anastomosing network of canals has been reported, for example in cestodes of the orders Amphilinidea (Joyeux and Baer, 1961) and Caryophyllidea (Mackiewicz, 1972). However, in these taxa sublateral pairs of excretory ducts were not also observed. Examination of the anastomosing canals of *L. aenigmaticum* using TEM revealed that some protonephridia occur in the vicinity of these subtegumental canals but they were sparse and no connection between these protonephridia and the canals was observed. The protonephridia were also found to be located much farther away from the anastomosing canals than they were from the sublateral paired ducts.

TEM also revealed that the syncytial cytoplasm of the neodermis in *L. aenigmaticum* is not only much thicker (5–7 µm) than seen in other cestodes, but is also extremely densely packed with electron dense discoidal inclusions (Figs. 12A,B). Similar inclusions, referred to differentially as rhabdiform organelles (Smyth, 1969), discoidal granules (Threadgold, 1984), and discoidal bodies (Osaki, 1990), have been reported in other cestode taxa. The latter term seems most appropriate as these bodies do not appear to be organelles, nor are they consistent in form with the electron dense vesicles seen throughout the various cell types of the scolex of *L.*
*aenigmaticum*. Although by far the majority of these bodies were located in the distal cytoplasm of the neodermis, some were observed within the submuscular cytons and their associated cytoplasmic extensions. Several functions have been proposed for these bodies. They have been hypothesized: (1) to contribute to the turnover of the plasma membrane of the tegument (Lumsden et al., 1974; Wilson and Barnes, 1974; Osaki, 1990), (2) to assist with the synthesis and secretion of proteins (Smyth, 1969), or (3) to contribute to the formation of the electron dense cap of microtriches (Lumsden et al., 1974; Shivers et al., 1986). The results provide some evidence that call the latter function into question because discoidal bodies were not seen either aligning with the base of, or passing into, microtriches as described by Shivers et al. (1986). Furthermore, the microtriches of the cephalic peduncle of *L. aenigmaticum* are filitriches (Caira et al., 2014a) and thus have extremely small electron dense caps (Fig. 12A insert). This is in contrast to the spinitriches that bear large, electron dense caps that were observed by Lumsden et al. (1974) in *Spirometra mansonoides* (Mueller, 1935) Wardle, McLeod, & Stewart, 1947 and by Shivers et al. (1986) in *Hymenolepis diminuta* (Rudolphi, 1819) Weinland, 1858. If the discoidal bodies are responsible for the formation of the microthrix cap we would expect a much greater number of these structures to occur in cestodes with spinitriches than in cestodes with only filitriches, but this was not found to be the case. *Litobothrium aenigmaticum* appears to have a much greater density of discoidal bodies in its distal cytoplasm than does any other cestode reported to date.

The most striking result, however, was the discovery that the scolex of *L. aenigmaticum* houses many more than the four “tissues,” in a much more complex arrangement, than described by Caira et al. (2014a) with light microscopy alone. Although recognition of the 11 distinct cell types was based on some criteria that may be transitory in nature (e.g., density of vesicle content), the majority of the criteria used (e.g., cell size and shape, nucleus size, cytoplasm to nucleus ratio,
and whether Golgi apparatus, mitochondria, endoplasmic reticulum, and free ribosomes were readily observed) are more stable. Therefore, it is believed that the combination of these criteria justifies the recognition of 11 distinct cell types. Furthermore, the aggregations of each cell type were found to be concentrated in specific regions of the scolex and the position and association of these cell types to one another was consistent across all five specimens examined. This leads us to believe that these are distinct cell types rather than developmental stages of a smaller number of cell types.

Unfortunately, it is difficult to reconcile the 11 cell types described here with the four “tissues” recognized by Caira et al. (2014a). The exception is that of “tissue one.” Given only a single “tissue” or cell type is located in the scolex proper and anterior-most regions of the cephalic peduncle it is thought that cell type A is the equivalent of “tissue” one. This assertion is further supported by the fact that both cell type A and “tissue” one were described as consisting of large cells bearing small nuclei. All equivalencies beyond that are purely speculative because most, or possibly all, of the remaining three “tissues” recognized by Caira et al. (2014a) appear to be composed of aggregations of more than one of the cell types, the majority of which are not identifiable with light microscopy alone. Nonetheless, based on position, we believe that “tissue” two may include some combination of cell types A, B, C, D, and E; “tissue” three may include some combination of cell types F, G, and possibly also H, I, and J; “tissue” four is either the equivalent of just cell type F, or a combination of F and cells types I and K. We would, however, caution against considering these combinations of cells as true tissues until the functions and interdependencies of the various cell types is more fully understood.

Both electron dense and electron lucent inclusions have been referred to as vesicles because the large majority of them appear to be membrane-bound. Although it is possible that the inclusions
that show less evidence of membranes may be lipid droplets, we are hesitant to refer to them as such for two reasons. First, there is evidence (e.g., see the Golgi apparati in Fig. 10F) that not all membranes in our specimens were equally well preserved. Second, even if the membranes were properly preserved they can be difficult to distinguish from the contents of electron dense vesicles whose density often matches that of a limiting membrane (Fawcett, 1981).

It is believed that the electron dense and electron lucent vesicles are different types of vesicles that likely contain different products, rather than being developmental stages of the same type of vesicle. This is because, beyond the differences in the density of their contents, they also vary in size and shape, with the electron lucent vesicles typically being much larger and more irregular in shape than the electron dense vesicles. As a consequence, the presence of both electron dense and electron lucent vesicles within the cells of cell types C, G, and K suggests these cells are capable of producing more than one product. However, exactly what those products might be and the functions those products might serve are currently unknown.

The presence of vesicles in all 11 cell types suggests that either each cell type is capable of producing its own products, or that the products produced by a subset of cell types are being transported between cell types. Our observations lead us to believe the former is the case. Evidence supporting this conclusion is as follows. Six of the 11 cell types (i.e., B, C, D, F, H, and K) contain some combination of endoplasmic reticulum, free ribosomes, and Golgi apparati — organelles known to be required to sort, modify, and appropriately package proteins (Urbé et al., 1997; Alberts et al., 2014). Although the remaining five cell types were not observed to contain all these organelles, they may contain them, but in such small numbers as to be difficult to observe in TEM sections. This explanation seems plausible given that essentially all eukaryotic cells are known to contain endoplasmic reticulum, Golgi apparati, and free ribosomes (Alberts et al. 2014).
Although in most cases we found no clues indicating how either type of vesicle is formed, the exception was cell type A in which tiny, electron dense vesicles were observed. These vesicles closely resemble the microvesicles reported in the secretory glands of the mollusk *Wirenia argentea* Odhner, 1920 (Todt and Salvini-Plawen, 2004) and it was suggested that microvesicles may fuse to form larger vesicles. In fact, vesicle growth through the condensation of smaller vesicles has been reported in many other organisms (Alberts et al. 2014). However, no evidence of the actual fusion of microvesicles was observed in *L. aenigmaticum*. Again, it is possible that this phenomenon occurs in this species but was not observed in the sections prepared.

Another somewhat unexpected finding was that many more regions of the scolex than originally thought were found to be PAS positive. When *L. aenigmaticum* was originally described by Caira et al. (2014a) only sections of the anterior-most region of the scolex were subjected to PAS staining because it was the only region in which vesicles were clearly visible with light microscopy. The discovery that all 11 cell types contain vesicles led us to stain light microscopical sections of all 11 cell types with PAS, and indeed all but the cells in the central region of the aggregation of cell type B were found to be PAS positive (Fig. 13). As all 11 cell types contain electron dense vesicles we believe the products in the electron dense vesicles are staining with PAS, however we cannot be certain since we did not use a PAS staining protocol at the TEM level. Currently, it is unclear as to why certain regions of cell type B are negative for PAS. TEM revealed that some cells of cell type B contain electron lucent vesicles. However, since it is not possible to tell what kind of vesicles are present in the PAS negative region with light microscopy the cause of this result cannot be determined.

Products that have been suggested to have an affinity for PAS stain include those that contain glycogen and/or mucin (Bogitsh, 1962; 1963). In cestodes a positive reaction to PAS has
been associated with the presence of mucoproteins (Bogitsh, 1962; Whittington and Cribb, 2001), glycoproteins (Brockerhoff and Jones, 1995), glycoproteins and mucopolysaccharides (Whittington and Cribb, 2001; Caira et al., 2005; Koch et al., 2012), and glycogen or neutral mucosubstances (Jensen and Russell, 2014). However, with PAS alone it is not possible to distinguish among these substances. Since we did not further characterize the products within either type of vesicle we cannot be certain of the nature of the products being produced by any of the 11 cell types. However, previous work on cestodes involving use of PAS leads us to hypothesize that one or possibly both types of vesicles contain some form of mucin and/or glycoprotein. What is puzzling, however, is that mucous and its derivatives are typically electron lucent when examined with TEM. If the products within vesicles in the scolex of cestodes are mucopolysaccharides or mucoproteins, we would expect them to be electron lucent.

Previous work includes examples of both electron dense and electron lucent products occurring within the scolex glands of cestodes. Hayunga (1979b) found that the frontal glands of *Hunterella nodulosa* Mackiewicz and McCrae, 1962 and the Faserzellen of *Gladridacris laruei* (Lamont, 1921) Hunter, 1927 were PAS positive and contained only electron lucent vesicles. McCullough and Fairweather (1989) found that the scolex glands of *Trilocularia gracilis* Olson 1867 (as *Trilocularia acanthiaevulgaris* Olson 1866) contained electron dense vesicles; a strong positive reaction to silver methenamine suggests that the contents of these vesicles were glycoproteins. Other studies (Žďarška and Nebesářová, 1997; Žďáršká et al., 2004) have reported the presence of electron dense vesicles in cestode scolex glands but they did not use stains that allowed them to characterize the products in these vesicles. Overall it has been suggested that electron lucent vesicles may produce a PAS positive reaction but the same has not been shown for electron dense vesicles. This suggests that PAS positive reaction exhibited by cell types C, G, and
K may be caused by the electron lucent vesicles in the cells. However, the rest of the cell types do not contain electron lucent vesicles yet still exhibit a PAS positive result, leading us to believe that the electron dense vesicles would also be PAS positive.

In general, four functions have been proposed for the products found in secretory vesicles of cestode scoleces. A number of authors have suggested that some products are proteolytic and thus are involved in nutrition and penetration of host mucosa by enzymatic activity (Farooqi, 1958; Slais, 1961; Thompson et al., 1979; Kuperman and Davydov, 1982). Davey and Breckenridge (1967) and Smyth (1971) suggested that some products may assist in strobilization (i.e., the process of production of cestode proglottids) (Davey and Breckenridge, 1967; Smyth, 1971). It also has been proposed that some products may protect the cestode from the immune system of its host (Farooqi, 1958; Thompson et al., 1979; Kuperman and Davydov, 1982). However, the most common function attributed to the substances associated with cestode scolex glands is that they assist with adhesion to host mucosa (Mackiewicz, 1972; Hayunga, 1979b; Kuperman and Davydov, 1982; McCullough and Fairweather, 1989). Evidence supporting these different hypotheses varies. Since enzymatic activity has yet to be demonstrated in the scolex glands of any tapeworm species a proteolytic function for these products is questionable (Hayguna 1979b; McCollough and Fairweather 1989). Similarly, the suggestion that these products assist with strobilization is unlikely since the products would have to be neurosecretory in nature and no evidence for this has been observed (McCollough and Fairweather, 1989). Although an intriguing idea, little evidence has been presented to support the suggestion that the secretory products could be used for protection (McCollough and Fairweather, 1989). Furthermore, _L. aenigmaticum_ elicits an extreme inflammatory host reaction (see Caira et al., 2014a) so we doubt that the products of this species serve to protect the worm from an immune reaction. In contrast, evidence for an
adhesive function for scolex secretory products continues to mount. McCollough and Fairweather (1989) found that the scolex glands of plerocercoids of *T. acanthiaevulgaris* contain much more product than seen in the adults and believed this was because plerocercoids would require more products for the initial stages of attachment than would adults. Furthermore, cestode taxa that lack, or possess weak attachment organs, such as a number of the Caryophyllidea, tend to have more prominent scolex glands (Mackiewicz, 1972). Typically these cestodes have frontal glands located near the apex of the scolex, but some species in this order also bear Faserzellen, which occur as columns of cells in the medullary parenchyma of the anterior regions of the scolex (Hayunga, 1979a; Hayunga, 1979b; Dezfuli et al., 2011). These glands are believed to assist with attachment (Hayunga, 1979b; McCollough and Fairweather, 1989). Therefore, this final hypothesis seems the most likely, but it should be noted that we can only speculate about the function of the vesicle products since our data do not provide conclusive indication as to what they may be used for.

We remain uncertain about how the products in vesicles of most of the cell types are being released from the body of *L. aenigmaticum*. The exception is cell type A. The apex of the scolex proper bears an aperture and apical pores that Caira et al. (2014a) hypothesized are being used to release secretory products. The TEM conducted here suggests that products are being released from the aperture via apocrine release (Figs. 12J–K). Our work also revealed electron dense products in the pores surrounding this aperture (Figs. 12L–M), and suggest that these products are produced by cell type A since this is the only cell type that occurs in that region of the scolex. What remains unclear is how the products from the other ten cell types are being released from the scolex. Some may release their products directly through the adjacent tegument since electron dense vesicles and electron lucent vesicles were observed within the distal cytoplasm. However, we found no evidence of vesicles actually being transported to the tegument. No vesicles were
observed moving through the muscle layers and none of the cell types are located immediately adjacent to the cytons or the syncytial layer of the tegument. In several sections vacuole-like structures were observed between two cell types that may be collecting and moving products from the secretory vesicles but they were only observed between cell types G and F, H and F, and I and F (Fig. 12F–I). Thus, even if these structures were responsible for moving products between these cell types it does not explain how the products of the remaining cell types are dispersed, nor how the products are ultimately released.

We can speculate as to why *L. aenigmaticum* requires such an extensive and complex set of associated cell types. In their description of *L. aenigmaticum*, Caira et al. (2014a) reported that this cestode appears to induce the mucosa of its host to form a papilla around the entire scolex (i.e., scolex proper and cephalic peduncle) of this relatively large worm. The 11 cell types described here occupy essentially the portion of the body of the worm that is surrounded by this host mucosal papilla. Although only conjecture at this point, it thus seems possible that at least some of the 11 cell types play a role in the production of substances that elicit this host response. If the products of some or all 11 cell types are interacting to produce this reaction then they are cooperating to perform a single function or a set of related functions and may be correctly categorized as a tissue.

The strategy of eliciting an inflammatory response in the host to aid in attachment has been reported in other cestodes. The caryophyllideans *Hunterella nodulosa* and *Monobothrium wageneri* Nybelin, 1922 are similar to *L. aenigmaticum* in that they lack holdfast organs, such as hooks or bothridia, and induce an inflammatory reaction of the host mucosa that is considered to aid in attachment (Hayunga 1979a; Dezful et al. 2011). However, in the case of both species the inflammatory reaction differs from that seen in *L. aenigmaticum* in that the inflammation of host
tissue occurs around more than one individual—typically two or more individuals of *H. nodulosa* (see Hayunga 1979a) and up to 100 individuals of *M. wagneri* (see Dezfuli et al. 2011). Furthermore, neither *H. nodulosa* nor *M. wagneri* have the complexity of cell types seen in *L. aenigmaticum*. Instead, as discussed earlier, these caryophyllideans have only frontal glands and Faserzellen (Hayunga, 1979a; Hayunga, 1979b; Dezfuli et al., 2011).

Previous light microscopy and molecular work brought to light the unusual morphology and mode of reproduction of *L. aenigmaticum* not only relative to its congeners, but also relative to cestodes overall. Although the TEM conducted here has resolved the issues of the configuration of the excretory system in this unusual cestode, it has raised questions about its novelty in other respects. The cell types seen in what is interpreted as the scolex of *L. aenigmaticum* are unlike those described in any other cestode. Although we have speculated as to the role of these cell types, further studies will be needed to determine their roles in this extremely enigmatic animal.
Chapter 2: Litobothrium genomic evolution

Introduction

The primary aim of this chapter is to compare the genome of *L. aenigmaticum* with those of some of its congeneres. However, this task is complicated by the fact that genomic data were unavailable for any member of this genus and thus needed to be generate *de novo*. Although the field of genomics has been around since the early 1980s (for microbes), these methodologies were not applied to the Platyhelminthes until the early 2000s (Olson et al., 2012). The first flatworm genomes generated were of the blood flukes *Schistosoma mansoni* and *Schistosoma japonicum* (Berriman et al., 2009; Zhou et al., 2009). The first tapeworm genomes were not available until 2013 when Tsai et al. (2013) published genomes for the following cyclophyllidean species: *Echinococcus granulosus* (Batsch, 1786) Rudoplhi 1801, *Echinococcus multilocularis* Leuckart, 1863, *Hymenolepis microstoma* (Dujardin, 1845), and *Taenia solium* Linnaeus, 1758. Since 2013 the number of published tapeworm genomes has quadrupled. Genomes are currently available for 12 addition species: *Dibothriocephalus latus* Linnaeus, *Schistocephalus solidus* (Mueller, 1776) Steenstrup, 1857, *Spirometra erinaceieuropaei*, *Echinococcus canadensis* (Cameron, 1960), *Hydatigera taeniaeformis* (Batsch, 1786) Lamarck, 1816, *Hymenolepis diminuta* (Rudolphi, 1819) Weinland, 1858, *Hymenolepis nana* Stiles, 1906, *Mesocestoides corti* Hoeppli, 1925, *Taenia asiatica* Eom and Rim,1993, *Taenia multiceps* Leske, 1780, and *Taenia saginata* Goeze, 1782 (Tsai et al., 2013; Zheng et al., 2013; Bennett et al., 2014; Maldonado et al., 2017; International Helminth Genomes Consortium, 2019). However, all these genomic sequencing efforts have been focused on the members of medically important tapeworm orders, largely the Cyclophyllidea and Diphyllobothriidea. Furthermore, the Diphyllobothriidea represents an early diverging group while the Cyclophyllidea represents a crown group, meaning that both are distantly related to the
Litobothriidea. As a consequence, the published genomic resources available for tapeworms cannot be used to assist with assembly of the litobothriidean genomes; instead the litobothriidean genomes needed to be assembled de novo. These resources were, however, used to assist with the annotation of the litobothriidean genomes. They were also used to inform hypotheses regarding the potential size, structure, and content of the litobothriidean genomes. The diphyllobothriidean genomes were reported to range in size from 531–1,259 Mb and to contain 19,966–39,557 genes (Bennett et al., 2014; International Helminthes Consortium, 2019). The cyclophyllidean genomes were reported to range in size from 104–166 Mb and to contain 10,614–13,777 genes (Tsai et al., 2013; Maldonando et al., 2017; International Helminthes Consortium, 2019). These data suggest that earlier diverging orders (e.g., the Diphyllobothriidea) have larger genome sizes than later diverging groups (e.g., the Cyclophyllidea). Based on this assumption and the phylogenetic position of Litobothriidea (Caira et al., 2014b), it was hypothesized that the litobothriideans would have genome sizes that were smaller than those of the diphyllobothriideans but larger than those of the cyclophyllideans. It was also hypothesized that the gene numbers in the litobothriideans would range somewhere between those of the diphyllobothriideans and those of the cyclophyllideans, thus approximately 10,000–20,000 genes. Given that both the diphyllobothriideans and cyclophyllideans have reduced repeat content relative to the other platyhelminthes (Tsai et al., 2013; Zheng et al., 2013; Bennett et al., 2014; Maldonando et al., 2017; International Helminthes Consortium, 2019), it was expected that this would also be seen to be the case in the litobothriideans.

As a result of their comparative genomic analyses, Tsai et al. (2013) also provided some intriguing hypotheses about gene family evolution in the tapeworms. The hypotheses are as follows: (1) The laminin family, thrombospondin containing family, and novel protocadherin
family were expanded in Platyhelminthes overall. (2) The LDL receptor family and galactosyl transferases family were expanded in only the digeneans and tapeworms. (3) The CD2 domain containing protein family and novel transmembrane family are tapeworm-specific. (4) The diagnostic antigen 50, tegumental dynein light I antigen, BTB, BACK, Kelch protein family, ortoperin like genes, novel repeat domain family, and ubiquitin conjugating enzyme families underwent tapeworm-specific expansions. (5) The novel flatworm gene family (that is superficially similar to the Zona pellucida-like domain) and the novel taeniid protein are present in all cyclophyllideans, but they are especially expanded in the *Taenia* genomes. (6) The novel *E. multilocularis* gene family (novel domains: novel_000011, novel_000049), novel *E. multilocularis* gene family (novel domains: novel_00051, novel_002642), novel *E. multilocularis* gene family (similar to Chromo-domain family), novel *H. microstoma* family (gag-pol transposable element), novel *E. multilocularis* gene family (similar to Chromo-domain family), novel *H. microstoma* family (transposable element), and protein kinase family have undergone species-specific expansions in each of the above taxa. However, the taxon limitation of the samples used in the study limits the population of inference of these hypotheses. The inclusion of other divergent taxa provides an opportunity to test the generality of their results. Therefore, a secondary aim of this study was to test the conclusions of Tsai et al. (2013), as well as to investigate the existence of litobothriidean-specific gene families.

In order to achieve these aims, genomes were generated for three litobothriidean species, *L. aenigmaticum, L. daileyi,* and *L. amplifica.* All three species infect the same species of definitive host, the pelagic thresher shark, *Alopias pelagicus.* As the hypothesized sister taxon of *L. aenigmaticum,* *L. daileyi* provides the opportunity for comparison to a close relative; *L. amplifica* allows for the comparison to a slightly more distant relative.
Transcriptomes were also generated not only for use in scaffolding genomic assemblies and assisting in gene space annotation, but also for identifying genes that are differentially expressed in *L. aenigmaticum* compared to its congeners (see Chapter 3). The ideal sampling design at the outset of the project was to generate transcriptomes for all four of the species (i.e., *L. aenigmaticum, L. daileyi, L. nickoli*, and *L. amplifica*) that infect *A. pelagicus*, therefore allowing for some control of environmental variation. To account for individual variation the original study design was to generate transcriptomes for three individuals for each species within a single host specimen. To further control for environmental variation, we originally planned to replicate this work on worms from three host individuals. That original, somewhat ambitious, study design required collection of three specimens of each of the four litobothriidean species in each of the three individual pelagic thresher sharks, allowing for the generation of a total of 36 transcriptomic libraries. A major limitation of achieving this sampling was that in order to generate the transcriptomes, the tapeworm specimens needed to be collected while they were still alive so as to ensure RNA was not degraded. In the case of the cestodes of the pelagic thresher shark this proved to be somewhat challenging for it required the necropsy of extremely fresh sharks. The pelagic thresher shark is a large, oceanic species that is targeted by fishermen in regions of the globe such as Taiwan. However, fishing often spreads over several days at sea before returning to market with their catch. As a consequence, by the time they return to market the sharks in this catch may have been dead for up to several days – greatly decreasing the chances that the tapeworms hosted by these sharks are still alive. Ultimately, live specimens of three of the four target species (i.e., *L. aenigmaticum, L. daileyi*, and *L. nickoli*) were collected from each of two host specimens. Therefore, the final study design for the transcriptomic work consisted of three replicates of each
of the three species in each of two host individuals, for a total of 9 transcriptomes per host and 18 transcriptomes overall.

**Materials and Methods**

**Specimen Collection**

Specimens of *L. aenigmaticum*, *L. daileyi*, and *L. amplifica* preserved in 95% ethanol resulting from a previous collecting trip to Taiwan in 2013 were available for sequencing. These specimens came from five different specimens of *A. pelagicus* that were landed by trawling vessels in Chenggong (2 specimens) and Nanfang-ao (3 specimens). Each animal was assigned a unique collection code (TW-55, TW-56, TW-57, TW-101, and TW-102, respectively) and photographed. Additional information can be accessed in the Global Cestode Database (www.tapewormsdb.uconn.edu; Caira et al., 2018) by searching for the Collection Code and Collection Number. Necropsies were performed on November 4 (TW-55, TW-56, TW-57) and November 9 (TW-101, TW-102) in 2013. Four specimens were female, 182–392 cm in total length (TL). One specimen was male, 301 cm in TL. The abdominal cavity of each animal was opened with a midventral incision and a small sample of liver was removed and preserved in 95% ethanol for molecular verification of host identity. The spiral intestine was then removed and opened with a longitudinal incision. Spiral intestines were then either examined for tapeworms in the field or preserved and examined for tapeworms upon return to the laboratory. Tapeworms examined in the field were sorted and one subset was preserved in 95% ethanol for molecular work and the other was preserved in 10% buffered formalin for examination with light and scanning electron microscopy (SEM). No litobothriidean specimens were preserved in RNAlater (Thermofisher Scientific) at that time.
In January of 2017, a second trip to Taiwan was made to collect and preserve specimens in RNA later for the generation of transcriptomes based on the study design outline above. During this collection trip four specimens of *A. pelagicus* were landed by trawling vessels in Donggang, Fugang, and Chenggong. Necropsies were performed on January 10 (TW-207) and January 12 (TW-211, TW-215, and TW-216) in 2017. Two specimens were male and two were female. These individuals were 170–228 cm in length. The tapeworms in two of these host specimens (TW-207 and TW-216) were still living at the time of necropsy. The spiral intestine of host individual TW-207 was brought back to the laboratory at the Institute of Marine Biology and Aquarium in Checheng, Taiwan where worms were sorted and preserved in RNA later. In total, 5 specimens of *L. aenigmaticum*, 15 specimens of *L. daileyi*, 13 specimens of *L. nickoli*, and 6 specimens of *L. amplifica* were collected from TW-207. The spiral intestine of host individual TW-216 was examined for tapeworms in the field. Worms were preliminarily sorted to species and were preserved in RNA later. In total, 4 specimens of *L. aenigmaticum*, 29 specimens of *L. daileyi*, 11 specimens of *L. nickoli*, and 6 specimens of *L. amplifica* were collected from TW-216.

**Transcriptomic Methods**

*RNA Extraction & Library Preparation*

RNA was extracted from whole specimens preserved in RNA later using Trizol following the manufacturer’s instructions. RNA quality and quantity was assessed first by fluorometry with a Qubit with RNA Broad Range Assay Reagents (Invitrogen) and then by fragment analysis with an Agilent Tape Station prior to library preparation. Transcriptomic libraries were prepared from extractions with adequate RNA at the University of Connecticut’s Center for Genomic Innovation using the Illumina TruSeq Stranded mRNA preparation kit. Dual indices were used in library preparation to address the known index hopping problem associated with the Illumina HiSeq4000
Six libraries were prepared for *L. aenigmaticum* (three from host individual TW-207 and three from host individual TW-216); six were prepared for *L. nickoli* (three from host individual TW-207 and three from host individual TW-216); five were prepared for *L. daileyi* (three from host individual TW-207 and two from host individual TW-216). No libraries were prepared for *L. amplifica* due to low RNA yield and lack of host species replication. All of the libraries for *L. aenigmaticum* and five of the libraries for *L. nickoli* were generated from RNA taken from single, individual worms. Because of low RNA yields, one library for *L. nickoli* and all five of the libraries for *L. daileyi* were generated from RNA pooled across individuals, ranging from two to three individuals per pool.

Libraries were shipped to the Vincent J. Coates Genomics Sequencing Laboratory at the University of California Berkeley (QB3 Berkeley) for sequencing on the HiSeq4000. Technicians at QB3 Berkeley checked the quality of the libraries using an Agilent Bioanalyzer and discovered primer contamination in all of the libraries. This was dealt with by pooling the libraries together and removing the contamination from the pool using size selection with Pippin Prep (Sage Science). They then performed another short sequencing run on the MiSeq Nano (Illumina) to ensure that the pools had equal representation of each library. This run revealed contamination of the libraries with material tagged with the TruSeq indices 10 and 12 which were not used in this study. Levels of contamination varied across samples; in all but two cases it was fairly low. However, in one library for *L. aenigmaticum* and one for *L. daileyi* it was extremely high. As a consequence, these two libraries were eliminated from further consideration. Without these libraries the contamination level was 25%. The remaining 14 libraries were sequenced across two lanes of the Illumina HiSeq4000 to ensure ample sequencing depth.

*Transcriptome Assembly*
FastQC (Andrews, 2010) was used to generate basic statistics, assess the quality of the raw reads, and identify levels of contamination in the raw reads in order to identify samples that needed trimming of poor quality reads. All samples were found to contain low levels of adapter contamination, which was removed with the program Trimmomatic (Bolger et al., 2014). FastQC was then run on the trimmed files to ensure the contamination had been removed.

Since no reference genomes are available for the litobothriideans, transcriptomes for each library were assembled *de novo* using the Trinity pipeline (Haas et al., 2013). Assembly quality was assessed using QUAST (Gurevich et al., 2013), which yielded basic summary statistics including assembly size, number of contigs, largest contig, and N50. Overall, 5 individual transcriptomes were generated for *L. aenigmaticum*, 3 for *L. daileyi*, and 6 for *L. nickoli*. Since none of these individual transcriptomes alone were likely to represent the entire transcriptome for any of the three species, a master transcriptome was generated from the individual transcriptomes for each species. These master transcriptomes were produced by clustering the individual transcriptomes for each species using the program CD-HIT (Li and Godzik, 2006) set at 90% identity. Thus, 5 assemblies were clustered to form the *L. aenigmaticum* master transcriptome, 3 assemblies were clustered to form the *L. daileyi* master transcriptome, and 6 assemblies were clustered to form the *L. nickoli* master transcriptome.

**Genomic Methods**

*DNA Extraction & Library Preparation*

Specimens of *L. aenigmaticum*, *L. amplifica*, and *L. daileyi* preserved in 95% ethanol were washed in lysis buffer to remove any traces of host tissue. Total genomic DNA was then extracted from these specimens using a MasterPure™ DNA Purification Kit (EpiCentre Technologies, Madison, Wisconsin) following the manufacturer’s instructions. Whole specimens were used for
these extractions. The final DNA extractions were left at 65°C with gentle shaking overnight to allow the DNA to go into solution. The quality of the extraction was assessed by fluorometry with a NanoDrop 2000 micro-volume spectrophotometer (Thermo Fisher Scientific, Waltham, Massachusetts) and Qubit with RNA Broad Range Assay Reagents (Invitrogen).

Overall, two short read libraries, with insert sizes averaging 350 and 550 bp, were generated for each of the three species. Two long read mate pair libraries (2.5 and 7.5 Kb) were generated for *L. aenigmaticum* to help with scaffolding the genomes during assembly. In order to prepare these libraries, the extractions of *L. daileyi* and *L. amplifica* were separated into two subsets, and the extractions of *L. aenigmaticum* were separated into three subsets. One subset of each species was sheared into 350 bp fragments using the Covaris sonication system. The second subset of each species was sheared into 550 bp fragments. The quality of the sheared samples was assessed by fluorometry with a Qubit with RNA Broad Range Assay Reagents (Invitrogen). Libraries were prepared from these sheared samples using the NEBNext Ultra II DNA Library Preparation kit for Illumina. The quality of the libraries was assessed with a fragment analysis using an Agilent Tape Station, which indicated that the sizes of the fragments in the libraries were too dispersed for effective sequencing. To deal with this, the libraries were further size selected using the PippinPrep Blue. These size-selected libraries were sent to QB3 Berkeley for multiplexing and sequencing on the HiSeq4000. The third subset of the extractions of *L. aenigmaticum* was also sent to QB3 Berkeley for the preparation of two mate pair libraries. Whole genome amplification was performed on this extraction subset to ensure there was enough DNA for the preparation of both the 2.5 and 7.5 kb libraries. Both mate pair libraries were also sequenced on the HiSeq4000.

*Genome Assembly*
Quality of the raw reads was assessed using the program FastQC (Andrews, 2010). Raw reads were then trimmed with the program Sickle (Joshi and Fass, 2011). Genome size estimation was performed with program Jellyfish (Marçais and Kingsford, 2011) and the kmercountexact.sh script from the program BBMap (Bushnell, 2014) using k-mer sizes 17, 19, 21, 25, and 27. Genome assembly was performed with four programs: ABySS 2 version 3.0 (Jackman et al., 2017), SOAPdenovo2 version 2.3 (Luo et al., 2012), SPAdes version 3.9.0 (Bankevich et al., 2012), and MaSURCA version 3.2.3 (Zimin et al., 2013).

For assembly with ABySS2, trimmed, paired and unpaired short read (350 and 550 bp) sequence data were used. The program was run with k-mer sizes 27, 31, and 35. The resulting assembly was assessed with the program QUAST (Gurevich et al., 2013). The assembly was then scaffolded with the short read libraries for all the species, and also with the two long read in the case of L. aenigmaticum, with the program SSPACE (Boetzer et al., 2010). The quality of the scaffolded assemblies was then assessed with QUAST.

For assembly with SOAPdenovo2, trimmed, paired short read (350 and 550 bp) sequence data and, for L. aenigmaticum, also long read mate pair (2.5 and 7.5 kb) sequence data, were used. The program was run with k-mer sizes 27, 31, and 35. The resulting assembly was assessed with QUAST and then scaffolded with the program SSPACE. The quality of the scaffolded assembly was then assessed with QUAST.

For assembly with SPAdes, trimmed, paired short read (350 and 550 bp) sequence data and, for L. aenigmaticum, also long read mate pair (2.5 and 7.5 kb) sequence data, were used. The resulting assembly was assessed with QUAST and then scaffolded with the program SSPACE. The quality of the scaffolded assembly was then assessed with QUAST.
For assembly with MaSURCA, untrimmed short read and, for *L. aenigmaticum*, long read mate pair (2.5 and 7.5 kb) sequence data were used. The resulting assembly was assessed with QUAST and then scaffolded with the program SSPACE. The quality of the scaffolded assembly was then assessed with QUAST.

The scaffolded assemblies across the four assembly programs were compared based on the following metrics: genome size, largest scaffold, number of scaffolds, and N50. These metrics were used to select the best two assemblies for each of the three species. The best assemblies for *L. aenigmaticum* and *L. daileyi* were then scaffolded with their respective species master transcriptomes using tranScaff scripts (Wegrzyn, unpublished) to generate more contiguous assemblies. This method could not be used with *L. amplifica* because transcriptome data were unavailable for this species. These transcriptome-scaffolded assemblies were then assessed using QUAST and the best of the two assemblies was chosen for *L. aenigmaticum* and *L. daileyi*.

**Genome Annotation**

In the cases of the best assembly for each of the three species, repeat regions were soft masked using the programs RepeatModeler and RepeatMasker (Tarailo-Graovac and Chen, 2009). RepeatModeler used RECON, RepeatScout, and Tandem Repeat Finder to identify repeat regions within each genome and to build a repeat library. RepeatMasker was run on the assemblies to “mask” regions matching sequences in the repeat library by replacing them with lower case letters (soft masking) (Tarailo-Graoac and Chen, 2009). This allowed downstream annotation pipelines to predict genes within what has been classified as a repeat region if there is evidence supporting the placement of a gene model at that location. The soft masked assemblies were then annotated using three different methods: using the program MAKER version 2.31.10 (Cantarel et al., 2008), using the program Braker2 version 2.0.5 (Hoff et al., 2015), and by mapping the master
transcriptomes to the genomes with the program GMAP version 2017-03-17 (Wu and Watanabe, 2005).

**MAKER pipeline**

The MAKER pipeline was run for each species in three iterative rounds in order to improve gene prediction with each round. For the first round, the pipeline was provided with alternative species expressed sequence tags (EST) evidence, species-specific EST evidence, and alternative species protein evidence. Alternative species EST evidence was obtained from GenBank using the following search terms: Cestoda, Eucestode, and Platyhelminthes. This evidence was then concatenated with the master transcriptomes for the other litobothriidean species; for *L. aenigmaticum* the GenBank EST data were combined with the *L. daileyi* and *L. nickoli* master transcriptomes; for *L. daileyi* the GenBank data were combined with the *L. aenigmaticum* and *L. nickoli* master transcriptomes; for *L. amplifica* the GenBank data were combined with the master transcriptomes of all three other *Litobothrium* species. All of these EST data were then mapped to each of the three species genomes with the program GMAP at 70% identity. For each species, the evidence that mapped to the genome was then clustered with the program USearch (Edgar, 2015) at 70% identity. The clustered sequences were then mapped to the genome again using GMAP. The resulting gff3 file was used as the alternative EST input file for running the first round of MAKER.

For the species-specific EST evidence, the transcriptomes of *L. aenigmaticum* and *L. daileyi* were mapped to their respective genomes using GMAP at 70% identity. For each species, the EST evidence that mapped to the genome was then clustered with the program USearch at 70% identity. The clustered sequences were then mapped to their respective genome again using GMAP. The resulting gff3 file for each species was used as the species-specific EST input file for
running the first round of MAKER. Since no transcriptome data were available for *L. amplifica*, this species did not have species-specific EST evidence.

Protein evidence was obtained from two main sources. Protein files were downloaded from WormBase ParaSite (https://parasite.wormbase.org/index.html) for the following platyhelminth species: the cestodes *Diphyllobothrium latum, Schistocephalus solidus, Spirometra erinaceieuropaei, Echinococcus canadensis, Echinococcus granulosus, Echinococcus multilocularis, Hydatigera taeniaeformis, Hymenolepis diminuta, Hymenolepis microstoma, Hymenolepis nana, Mesocestoides corti, Taenia asiatica, Taenia saginata*, and *Taenia solium*; and the digeneans *Clonorchis sinensis, Echinostoma caproni, Fasciola hepatica, Opisthorchis viverrini, Schistosoma curassoni, Schistosoma haematobium, Schistosoma japonicum, Schistosoma mansoni, Schistosoma margrebowiei, Schistosoma mattheei, Schistosoma rodhaini, and Trichobilharzia regent*. All curated platyhelminth proteins were also downloaded from Uniprot (https://www.uniprot.org/). These protein files were concatenated and then mapped to all three *Litobothrium* genomes using the program Blat at 50% identity. The mapped protein evidence was then clustered with USearch at 70% identity. The resulting centroids were used as the protein input file for running the first round of MAKER.

The first round of MAKER was run using the soft masked genomes, the alternate species EST evidence in a gff file, the species-specific EST evidence in a gff file, and the alternate protein evidence in a fasta file. The program was run with the est2genome, altest2genome, and protein2genome flags activated and with the following settings: min_contig=500, max_dna_length=100000, pred_flank=200, split_hit=25000, and single_length=200.

Once the first round of MAKER finished, the resulting gene model gff and FASTA files were merged using the gff3_merge and fasta_merge command. The resulting gff files were parsed
to create a cdna2genome gff file and a protein2genome gff file. The resulting FASTA files were parsed with Python scripts to generate the following transcript and protein FASTA files: all MAKER models with an annotation edit distance (AED) of 0.2, all MAKER models with AED of 0.1, all complete MAKER models, all complete and monoexonic MAKER models, all complete and multiexonic MAKER models, and all partial MAKER models. The program SNAP (version 2013-11-29) was then trained with the MAKER derived transcripts. The program Interproscan version 5.35–74.0 (Quevillon et al., 2005) was then run on the protein file containing all MAKER models with an AED limit of 0.2. The result of this process was then used to train the program Augustus version 3.2.3 (Stanke et al., 2006).

The second round of MAKER was run with the cdna2genome gff file, protein2genome gff file, SNAP zff.hmm file and Augustus species parameters generated as described above. The program was run with the following settings: min_contig=500, max_dna_length=100000, pred_flank=200, split_hit=25000, and single_length=200.

The resulting gff and FASTA files were merged using the gff3_merge and fasta_merge commands. The merged FASTA files were parsed with Python scripts to generate the following transcript and protein FASTA files: all MAKER models with AED of 0.2, all MAKER models with AED of 0.1, all complete MAKER models, all complete and monoexonic MAKER models, all complete and multiexonic MAKER models, and all partial MAKER models. SNAP was then trained with the MAKER-derived transcripts. Interproscan was then run on the protein file that contained all the MAKER models with an AED limit of 0.2. The result of this process was then used to train the program Augustus again.

The third round of MAKER was run with the cdna2genome gff file and protein2genome gff file from round one and the SNAP zff.hmm file and Augustus species parameters generated
from MAKER round two. The program was run with the following settings: min_contig=500, max_dna_length=100000, pred_flank=200, split_hit=25000, and single_length=200.

The resulting gff and FASTA files were merged using the gff3_merge and fasta_merge command. The resulting FASTA files were parsed with Python scripts to generate the following transcript and protein FASTA files: all MAKER models with AED of 0.2, all MAKER models with AED of 0.1, all complete MAKER models, all complete and mono-exonic MAKER models, all complete and multi-exonic MAKER models, and all partial MAKER models.

After each round of MAKER the program gFACs (Caballero and Wegrzyn, 2018) was run on the gff file that contained all the MAKER gene models to determine the number of canonical, complete genes identified by the program and to output a protein FASTA file containing those genes. The completeness of the resulting annotations was assessed using the program BUSCO (Simão et al., 2015) to examine the number of metazoan universal single-copy orthologs that are present, present and duplicated, present but fragmented, and missing from each the genome.

Braker Pipeline

The Braker2 pipeline was run on *L. aenigmaticum* and *L. daileyi*. It was not possible to run on *L. amplifica* because of the lack of species-specific EST evidence. This pipeline was run twice for each species, once with just EST evidence and once with both EST and protein evidence. For the EST only run, Braker2 was given the master transcriptome for each species as evidence. The resulting gene models were assessed with gFACs to determine the number of complete, canonical models identified by the pipeline. BUSCO was used to assess the number of metazoan universal single copy orthologs represented by the complete, canonical gene models. For the second run, Braker2 was given both EST and translated protein evidence from the relevant master transcriptome. The resulting gene models were assessed with gFACs to determine the number of
complete, canonical models identified by the pipeline. BUSCO was used to assess the number of metazoan universal single-copy orthologs represented by the complete, canonical gene models.

**Annotation with GMAP**

The transcriptomes of *L. aenigmaticum* and *L. daileyi* were mapped to their respective genomes to recover genes that may not have been predicted by either the MAKER or Braker annotation pipelines. Mapping was done with the program GMAP with canonical mode set to true, minimum identity set to 0.95, and minimum coverage set to 0.90. Both the *L. aenigmaticum* and *L. daileyi* transcriptomes were mapped to the *L. amplifica* genome with the same settings described above.

**Final Gene Models**

In order to obtain the final gene model annotation sets for the genomes of each of the three species, the gene models from MAKER, Braker, and GMAP were filtered for overlap. The output gff files from each program were used to create BED files which were fed into the program BEDTools version 2.27.1 (Quinlan and Hall, 2010) to look for gene models that overlapped either internally or partially. Gene models that overlapped were removed using custom Python scripts. Since the gff files from each of the annotation pipelines had a slightly different format, gFACs version 1.0.0 was used to create gtf files for the gene models from each program. This allowed the gtf files from the different programs to be merged into one gtf file that represented the final genome annotation set for each *Litobothrium* species. The program gFACs was also used to obtain basic statistics and distributions for each of these final gene sets. This final set of gene models was then functionally annotated with the program EnTAP in order to determine which of them had protein domains.

*Differences between L. daileyi and L. aenigmaticum*
The typical barcoding gene for cestodes, 28S rDNA (D1-D3 region), shows no phylogenetic distance between *L. daileyi* and *L. aenigmaticum*. However, all morphological data indicates they are separate species. To further investigate this issue, differences between the two species were examined using both the transcriptomic and genomic data. Two different methods were used. The first was to examine the gene trees produced by Orthofinder (Emms and Kelly, 2015) when the individual transcriptomes for the litobothriideans were clustered and when the genome annotation gene models for the litobothriideans and *E. multilocularis* were clustered. The second method was to directly compare the genomes of *L. aenigmaticum* and *L. daileyi* using the program QUAST LG (Mikheenko et al., 2018).

**Comparative Transcriptomics**

The individual transcriptomes and genomic gene models were clustered with Orthofinder to identify orthologous protein sequence families (referred to as orthogroups) to generate a gene tree for each of these families (using a maximum likelihood approach), and to then produce a species tree by generating a consensus tree from the gene trees that contain all of the species included in the analysis.

For the transcripts, the separate transcriptomes for each sampled individual for each species (5 individuals for *L. aenigmaticum*, 3 individuals for *L. daileyi*, and 6 individuals for *L. nickoli*) were clustered with Orthofinder. The single-copy orthogroup gene trees were then analyzed using the R package Phytools (Revell, 2012) and Phyloch (Heibl, 2013) to determine how many of the gene trees indicated that both *L. aenigmaticum* and *L. daileyi* were monophyletic. The orthogroups that contained all three species but were not single-copy were also analyzed to determine how often *L. aenigmaticum* and *L. daileyi* were each monophyletic.
The final annotation gene models for each of the three litobothriidean species and *E. multilocularis* were also clustered with Orthofinder. The gene trees for the single-copy orthogroups were then examined using the R package Phytools to determine in how many of them that *L. aenigmaticum* and *L. daileyi* were shown to be separately monphyletic.

**Comparative Genomics**

The genome assemblies of *L. aenigmaticum* and *L. daileyi* were also compared using QUAST LG to identify possible coding region differences. Specifically, this program was used to generate a report summarizing the translocations and relocations, inversions, possible transposable elements, mismatches, and indels present in one assembly but not the other.

**Mapping Cyclophyllidean genes**

In order to test the conclusions made by Tsai et al. (2013) regarding the tapeworm- and cyclophyllidean-specific or expanded gene families, the genes from those families were mapped to the litobothriidean genomes using the program Blat. Genes extracted from the genomes of *Echinococcus multilocularis, Echinococcus granulosus, Hymenoleis microstoma, Taenia solium,* and *Schistosoma mansoni* (Tsai et al., 2013) were mapped to the litobothriidean genomes. The protein sequences of these genes were mapped to the litobothriidean genomes with a minimum identity of 70%.

**Identifying litobothriidean specific gene families**

Following clustering of the genome annotation gene models for *L. aenigmaticum, L. daileyi, L. amplifica, E. multilocularis, H. microstoma, T. multiceps, S. solidus,* and *S. mansoni* using Orthofinder, a Python script was used to parse the Orthofinder results to identify the orthogroups that were specific to the litobothriideans and those that were specific to *L. aenigmaticum, L. daileyi,* and *L. nickoli.* The genes in these orthogroups were functionally
annotated with EnTAP and a functional enrichment analysis was performed using the R package GoSeq (version 1.34.1) with a p-value<0.05 indicating a significant result. Redundant GO terms were removed with the program REVIGO (Supek et al., 2011) and the results were visualized with the R package treemap version 2.4–2 (Tennekes and Ellis, 2017).

Results

Transcriptomics

RNA Extraction and Library Preparation

RNA extractions were performed on four specimens of *L. aenigmaticum* from host TW-216 and five specimens from host TW-207. Extractions yielded 7–407 ng of RNA per specimen. The three specimens with the highest RNA yield from each of the two host individuals were selected for library preparation. RNA yield for the specimens from TW-216 was 89–257 ng; RNA yield for the specimens from TW-216 was 93–407 ng. Contamination that occurred during library preparation caused one of the libraries from host TW-207 to be excluded from further consideration; the remaining 5 libraries were sequenced for *L. aenigmaticum*.

RNA extractions were performed on 16 specimens of *L. daileyi* from host TW-216 and 11 specimens from host TW-207. Extractions yielded 4–89 ng of RNA per specimen. Given the minimum input necessary for library generation was 100 ng with the Illumina TruSeq Stranded mRNA preparation kit, none of the specimens yielded RNA to allow a RNASeq library to be prepared from a single individual. Therefore, all the libraries for this species were generated from pools of individuals. Specimens of *L. daileyi* from host TW-216 had much lower RNA yield per specimen than those of TW-207 (only 2 samples yielded >20 ng from TW-216 while 6 specimens yielded >20 ng from TW-207). With such low yields, it would have been necessary to pool more than three individuals in order to get three libraries for this host. Since it is best to avoid this amount
of pooling, only two libraries were generated, each using a pool of three individuals from this host. With respect to the three libraries from host TW-207, two of these libraries were made from pools of two specimens and the third library was made from a pool of three specimens. Contamination that was present in the library preparation reagents caused one of the libraries from host TW-207 to be eliminated from further consideration.

RNA extractions were performed for eight specimens of *L. nickoli* from host TW-216 and 10 specimens from host TW-207. Extractions yielded 6–200 ng of RNA per specimen. The three specimens with the highest RNA yield, 138–231 ng, from host TW-216 were used to generate the RNASeq libraries. Two specimens from host TW-207 yielded enough RNA, 253–298 ng, to generate libraries from a single specimen. The third library was generated from a pool of two specimens (46 and 59 ng respectively).

RNA extractions were performed on six specimens of *L. amplifica* from host TW-216 and two specimens from host TW-207. Extractions yielded a range of 4–264 ng of RNA per specimen. Only 2 specimens, 1 from each host individual, yielded enough RNA for library preparation. This gave the necessary between host species replication but did not allow for within host species replication and thus it was not possible to control for individual variation. As a consequence, *L. amplifica* was excluded from the transcriptomic study.

Overall, 14 transcriptomic libraries were prepared, 5 for *L. aenigmaticum*, 3 for *L. daileyi*, and 6 for *L. nickoli*.

**Raw Read Quality**

The 5 libraries for *L. aenigmaticum* yielded a range of 33,717,790–41,318,436 raw, paired reads (Table 2). One of these libraries had adapter contamination below 1%, three had adapter contamination between 1–2%, and one had adapter contamination around 10%. Trimming reduced
all contamination to below 1% and this reduced the range of paired reads to 27,524,462–28,867,143 (Table 2).

The 3 libraries for *L. daileyi* yielded a range of 28,580,831–35,626,809 raw, paired reads (Table 2). All 3 libraries had adapter contamination between 1–3%. Trimming reduced all contamination to below 1% and this reduced the number of paired reads to 23,010,291–29,111,547 (Table 2).

The 6 libraries for *L. nickoli* yielded a range of 30,350,303–41,424,952 raw, paired reads (Table 2). Two of these libraries had adapter contamination below 1%, three had adapter contamination of 1–2%, and one had adapter contamination around 3%. Trimming reduced all contamination to below 1% and this reduced the number of paired reads to 24,662,602–33,973,280 (Table 2).

*Transcriptome Quality*

The Trinity *de novo* assemblies for *L. aenigmaticum* were 33,535,388–53,446,605 bp in size (Table 3). The N50 for these 5 transcriptomes was 1,476–2,022 (Table 3); the proportion of complete metazoan BUSCO genes represented by the transcriptomes was 60.3–74.9% (Tables 5). Clustering the transcriptomes with the program CD-HIT reduced the size of the transcriptomes to 24,400,348–37,320,552 bp (Table 4). The N50 for these clustered transcriptomes was 1,307–1,716 (Table 4) and the proportion of complete metazoan BUSCO gene represented by the transcriptomes was 60.3–74.9% (Tables 5). The clustered master species transcriptome for *L. aenigmaticum* was 20,992,134 bp in size, with an N50 of 1,308 (Table 6), and a complete BUSCO score of 66.3% (Tables 7).

The Trinity *de novo* assemblies for *L. daileyi* were 11,737,573–30,015,264 bp in size (Table 3). The N50 for these transcriptomes was 925–1,524 (Table 3); the proportion of complete
metazoan BUSCO genes represented by the transcriptomes was 42.4–73.6\% (Table 5). Clustering the transcriptomes with CD-HIT reduced the size of the transcriptomes to 9,345,770–21,784,688 bp (Table 4). The N50 for these clustered transcriptomes was 889–1,404 (Table 4); the proportion of complete metazoan BUSCO genes represented by the transcriptomes was 54.2–70.7\% (Table 5). The clustered master species transcriptome for *L. daileyi* was 13,950,882 bp in size, with an N50 of 1,122 (Table 6), and a complete BUSCO score of 65.1\% (Tables 7).

The Trinity *de novo* assemblies for *L. nickoli* were 17,888,152–33,377,702 bp in size (Table 3). The N50 for these transcriptomes was 1,230–1,666 (Table 3); the proportion of complete metazoan BUSCO genes represented by the transcriptomes was 62.1–78\% (Table 5). Clustering the transcriptomes with CD-HIT reduced the size of the transcriptomes to 14,155,628–23,651,108 bp (Table 4). The N50 for these clustered transcriptomes was 1,147–1,474; the proportion of complete metazoan BUSCO genes represented by the transcriptomes was 67.1–74.5\% (Table 4,5). The clustered master species transcriptome for *L. nickoli* was 15,887,406 bp in size, with an N50 of 1,116, and a complete BUSCO score of 67.1\% (Tables 6, 7).

**Genome Assembly**

*Raw Read and Genome Size Estimation*

Four libraries were sequenced for *L. aenigmaticum*: 350 bp, 550 bp, 2.5 kb, and 7.5 kb libraries. The 350 bp library yielded 81,426,186 paired reads (Table 8). The 550 bp library yielded 74,288,258 paired reads (Table 8). The 2.5 kb library yielded 21,227,661 paired reads (Table 8). The 7.5 kb library yielded 18,466,882 paired reads (Table 8). Trimming the 350 and 550 bp libraries with Sickle reduced the number of paired reads to approximately 58 million and 63 million respectively (Table 9). Trimming the 2.5 kb and 7.5 kb libraries with NxtTrim reduced the number of paired reads to approximately 1.4 million and 1.3 million respectively (Table 9).
Genome size estimation, using k-mer sizes 17–27, with the program Jellyfish yielded size estimates from approximately 319–350 Mb (Table 10). Genome size estimation with BBMap yielded size estimates from 334–355 Mb (Table 11).

Two libraries were sequenced for *L. daileyi*: 350 bp and 550 bp. The 350 bp library yielded 83,556,525 paired reads (Table 8). The 550 bp library yielded 48,534,695 paired reads (Table 8).Trimming the 350 and 550 bp libraries with Sickle reduced the number of paired reads to approximately 38 million and 64 million respectively (Table 9). Genome size estimation, using k-mer sizes 17–27, with the program Jellyfish yielded sizes estimates of approximately 330–399 Mb (Table 10). Genome size estimation with BBMap yielded size estimates of approximately 347–413 Mb (Table 11).

Two libraries were sequenced for *L. amplifica*: 350 bp and 550 bp. The 350 bp library yielded 76,212,112 paired reads (Table 8). The 550 bp library yielded 39,985,806 paired reads (Table 8). Trimming the 350 and 550 bp libraries with Sickle reduced the number of paired reads to approximately 58 million and 31 million respectively (Table 9). Genome size estimation using k-mer sizes 17–27, with Jellyfish yielded sizes estimates of approximately 358–440 Mb (Table 10). Genome size estimation with BBMap yielded size estimates of approximately 376–474 Mb (Table 11).

*Genome Assembly*

**SOAPdenovo2**

SOAPdenovo2 was run for all three species with k-mer sizes of 27, 31, and 35. In the case of *L. aenigmaticum* the program was run once with the mate pair libraries and once without them. When mate pair libraries were excluded, a k-mer size of 27 resulted in an assembled genome size of 300,629,526 bp and an N50 of 7,447 (Table 12). Scaffolding with SSPACE increased the
assembled genome size to 331,857,039 bp and the N50 to 11,965 (Table 12). Using a k-mer size of 31 resulted in a genome size of 308,445,946 bp and an N50 of 6,273 (Table 12). Scaffolding with SSPACE increased the assembled genome size to 342,101,097 bp and the N50 to 10,773 (Table 12). A k-mer size of 35 resulted in a genome size of 319,337,072 bp and an N50 of 5,347 (Table 12). Scaffolding with SSPACE increased the assembled genome size to 352,959,590 bp and the N50 to 9,900 (Table 12).

When the mate pair libraries were included for *L. aenigmaticum*, a k-mer size of 27 resulted in a genome size of 296,778,647 bp and an N50 of 7,511 (Table 13). Scaffolding with SSPACE increased the assembled genome size to 339,774,740 bp and the N50 to 12,520 (Table 13). Using a k-mer size of 31 resulted in a genome size of 305,528,263 bp and an N50 of 6,298 (Table 13). Scaffolding with SSPACE increased the assembled genome size to 352,544,282 bp and the N50 to 11,548 (Table 13). Using a k-mer size of 35 resulted in a genome size of 317,074,558 bp and an N50 of 5,367. Scaffolding with SSPACE increased the assembled genome size to 364,915,107 bp and the N50 to 10,854 (Table 13).

For *L. daileyi*, using a k-mer size of 27 resulted in a genome size of 325,050,051 bp and an N50 of 12,402 (Table 12). Scaffolding with SSPACE increased the assembled genome size to 345,424,829 bp and the N50 to 16,023 (Table 12). Using a k-mer size of 31 resulted in a genome size of 321,594,163 bp and an N50 of 10,366 (Table 12). Scaffolding with SSPACE increased the assembled genome size to 346,420,159 bp and the N50 to 14,333 (Table 12). Using a k-mer size of 35 resulted in a genome size of 321,008,597 bp and an N50 of 9,059 (Table 12). Scaffolding with SSPACE increased the assembled genome size to 349,220,283 bp and the N50 to 13,167 (Table 12).
For *L. amplifica*, using a k-mer size of 27 resulted in a genome size of 319,365,064 bp and an N50 of 15,803 (Table 12). Scaffolding with SSPACE increased the assembled genome size to 344,855,725 bp and the N50 to 20,948 (Table 12). Using a k-mer size of 31 resulted in a genome size of 332,680,477 bp and an N50 of 14,575 (Table 12). Scaffolding with SSPACE increased the assembled genome size to 347,301,910 bp and the N50 to 20,186 (Table 12). Using a k-mer size of 35 resulted in a genome size of 334,544,798 bp and an N50 of 14,178 (Table 12). Scaffolding with SSPACE increased the assembled genome size to 350,171,298 bp and the N50 to 20,118 (Table 12).

**ABySS2**

ABySS2 was run with the same three k-mer sizes: 27, 31, and 35 and, for *L. aenigmaticum*, both with and without the mate pair libraries. When the mate pair libraries were excluded, a k-mer size of 27 resulted in a genome size of 250,157,438 bp and an N50 of 9,612 (Table 14). Scaffolding with SSPACE increased the assembled genome size to 277,527,658 bp and the N50 to 14,639 (Table 14). Using a k-mer size of 31 resulted in a genome size of 262,405,385 bp and an N50 of 14,114 (Table 14). Scaffolding with SSPACE increased the assembled genome size to 280,432,751 bp and the N50 to 18,208 (Table 14). Using a k-mer size of 35 resulted in a genome size of 268,280,519 bp and an N50 of 18,050 (Table 14). Scaffolding with SSPACE increased the assembled genome size to 283,256,722 bp and the N50 to 21,870 (Table 14).

When the mate pair libraries were included for *L. aenigmaticum*, a k-mer size of 27 resulted in a genome size of 248,632,364 bp and an N50 of 5,273 (Table 15). Scaffolding with SSPACE increased the assembled genome size to 279,607,219 bp and the N50 to 13,211 (Table 15). Using a k-mer size of 31 resulted in a genome size of 261,230,091 bp and an N50 of 7,881 (Table 15). Scaffolding with SSPACE increased the assembled genome size to 282,177,161 bp and the N50
to 16,238 (Table 15). Using a k-mer size of 35 resulted in a genome size of 267,327,203 bp and an N50 of 10,306 (Table 15). Scaffolding with SSPACE increased the assembled genome size to 284,819,813 bp and the N50 to 19,319 (Table 15).

For *L. daileyi*, using a k-mer size of 27 resulted in a genome size of 224,128,984 bp and an N50 of 7,915 (Table 14). Scaffolding with SSPACE increased the assembled genome size to 274,533,532 bp and the N50 to 12,593 (Table 14). Using a k-mer size of 31 resulted in a genome size of 258,668,634 bp and an N50 of 12,135 (Table 14). Scaffolding with SSPACE increased the assembled genome size to 277,181,504 bp and the N50 to 15,899 (Table 14). Using a k-mer size of 35 resulted in a genome size of 265,010,818 bp and an N50 of 15,244 (Table 14). Scaffolding with SSPACE increased the assembled genome size to 279,899,663 bp and the N50 to 18,341 (Table 14).

For *L. amplifica*, using a k-mer size of 27 resulted in a genome size of 244,340,710 bp and an N50 of 6,649 (Table 14). Scaffolding with SSPACE increased the assembled genome size to 290,047,530 bp and the N50 to 9,320 (Table 14). Using a k-mer size of 31 resulted in a genome size of 268,993,684 bp and an N50 of 9,562 (Table 14). Scaffolding with SSPACE increased the assembled genome size to 298,829,428 bp and the N50 to 12,398 (Table 14). Using a k-mer size of 35 resulted in a genome size of 292,612,921 bp and an N50 of 12,441 (Table 14). Scaffolding with SSPACE increased the assembled genome size to 303,935,085 bp and the N50 to 15,389 (Table 14).

**SPAdes**

SPAdes was run only once for *L. aenigmaticum* with the mate pair libraries. The resulting assembly was 283,966,322 bp in size and had an N50 of 19,530 (Table 16). Scaffolding with SSPACE increased the assembled genome size to 302,424,677 bp and the N50 to 45,536 (Table 16).
Further scaffolding with the *L. aenigmaticum* master transcriptome resulted in a genome size of 302,510,833 bp and an N50 of 58,137 (Table 17).

For *L. daileyi*, the resulting assembly was 287,655,135 bp in size and had an N50 of 16,852 (Table 16). Scaffolding with SSPACE increased the assembled genome size to 296,102,134 bp and the N50 to 37,572 (Table 16). Further scaffolding with the *L. daileyi* master transcriptome resulted in a genome size of 296,177,473 bp and an N50 of 45,652 (Table 17).

For *L. amplifica*, the resulting assembly was 314,209,663 bp in size and had an N50 of 18,096 (Table 16). Scaffolding with SSPACE increased the assembled genome size to 320,039,359 bp and an N50 to 38,108 (Table 16).

**MaSURCA**

For *L. aenigmaticum*, the resulting assembly was 329,296,342 bp in total length and had an N50 of 4,533 (Table 18). Since this assembler performed so poorly compared to the other assemblers for this species, no further work was done with this genome assembly.

For *L. daileyi*, the resulting assembly was 332,106,365 bp in total length and had an N50 of 32,543 (Table 18). Scaffolding with SSPACE increased the assembled genome size to 332,304,163 bp and the N50 to 36,392 (Table 18). Further scaffolding with the *L. daileyi* master transcriptome resulted in a genome size of 325,242,964 bp and an N50 of 51,464 (Table 19).

For *L. amplifica*, the resulting assembly was 355,450,347 bp in total length and had an N50 of 37,643 (Table 18). Scaffolding with SSPACE increased the assembled genome size to 355,593,269 bp and the N50 to 42,370 (Table 18).

**Genome Annotation Quality**

*Repeat Content*
Genome annotation was completed using the SPAdes transcriptome scaffolded assembly for *L. aenigmaticum* and *L. daileyi*. Since the quality of the MaSURCA and SPAdes SSPACE scaffolded assemblies for *L. amplifica* were remarkably similar, both assemblies were annotated to see which assembly was best to use for the final product. Prior to annotation, repeats were soft-masked using RepeatMasker. RepeatModeler found that the genome of *L. aenigmaticum* had a GC content of 43.38% and a repeat content of 27.4% (Table 20). Approximately 27% of the repeats were unclassified while 1% were classified as simple repeats. RepeatModeler found that the genome of *L. daileyi* had a GC content of 43.5% and repeat content of 28.1% (Table 20). Approximately 28% of the repeats were unclassified while 1% were classified as simple repeats. RepeatModeler found that the genome of *L. amplifica* had a GC content of 42.9% and repeat content of 31.35% (Table 20). Approximately 31% of the repeats were unclassified while 1% were classified as simple repeats.

**MAKER Annotations**

Round one of the MAKER pipeline for *L. aenigmaticum* resulted in a total of 14,886 gene models, 11,096 of which were complete. The complete models had a BUSCO score of 49.8% (Table 21). Of those complete models, 7,618 were also canonical. The complete, canonical models had a BUSCO score of 35.3% (Table 21). The addition of non-canonical, complete gene models and partial gene models that had at least either a start or stop codon increased the BUSCO score to 72.9% (Table 21). Round two of MAKER resulted in 10,015 gene models, 4,221 of which were complete (Table 21). The complete models had a BUSCO score of 32.8%. Of these complete models, 4,151 were also canonical (Table 21). The canonical, complete models had a BUSCO score of 35.31% (Table 21). The addition of non-canonical, complete gene models and partial gene models that had at least either a start or stop codon increased the BUSCO score to 56.2% (Table
Round three of MAKER resulted in 10,196 gene models, 4,944 of which were complete. The complete models had a BUSCO score of 39.7% (Table 21). Of these complete models, 3,758 were also canonical. The canonical, complete models had a BUSCO score of 32.2% (Table 21). The additions of non-canonical, complete gene models and partial gene models that had at least either a start or stop codon increased the BUSCO score to 49.3% (Table 21).

For *L. daileyi*, round one of MAKER resulted in a total of 16,425 gene models, 12,234 of which were complete. The complete models had a BUSCO score of 50.6% (Table 22). Of those complete models, 8,743 were also canonical. The canonical, complete models had a BUSCO score of 35.6% (Table 22). The additions of non-canonical, complete gene models and partial gene models that have at least either a start or stop codon increased the BUSCO score to 72.1% (Table 22). Round two of MAKER resulted in 11,965 gene models, 7,400 of which were complete. The complete models had a BUSCO score of 49.9% (Table 22). Of these complete models, 6,878 were also canonical. These complete, canonical models had a BUSCO score of 53% (Table 22). The additions of non-canonical, complete gene models and partial gene models that had at least either a start or stop codon increased the BUSCO score to 69.3% (Table 22). Round three of MAKER resulted in 10,559 gene models, 4,357 of which were complete. These complete models had a BUSCO score of 35.7% (Table 22). Of these complete models, 4,161 were also canonical. These complete, canonical models had a BUSCO score of 33.9% (Table 22). The additions of non-canonical, complete gene models and partial gene models that had at least either a start or stop codon increased the BUSCO score to 53.8% (Table 22).

For the SPAdes assembly of *L. amplifica*, round one resulted in a total of 5,399 gene models, 4,183 of which were complete. The complete models had a BUSCO score of 10.9%. Of these complete models, 2,946 were also canonical. These canonical, complete models had a
BUSCO score of 7.3% (Table 23). The addition of non-canonical, complete gene models and partial gene models that had at least either a start or stop codon increased the BUSCO score to 18.7% (Table 23). Round two of MAKER resulted in 3,802 gene models, 1,846 of which were complete. The complete models had a BUSCO score of 7.5% (Table 23). Of these complete models, 1,329 were also canonical. These complete, canonical had a BUSCO score of 6.1%. The addition of non-canonical, complete gene models and partial gene models that had at least either a start or stop codon increased the BUSCO score to 10.2% (Table 23). Round three of MAKER resulted in 3,538 gene models, 1,401 of which were complete. The complete models had a BUSCO score of 5.9% (Table 23). Of those complete models, 1,386 were also canonical. These canonical, complete models have a BUSCO score of 6.2% (Table 23). The addition of non-canonical, complete gene models and partial gene models that had at least either a start or stop codon increased the BUSCO score to 11.1% (Table 23).

For the MaSURCA assembly of *L. amplifica*, round one resulted in a total 23,318 gene models, 17,686 of which were complete. The complete models had a BUSCO score of 49.7% (Table 24). Of these complete models, 13,727 were also canonical. These canonical, complete models had a BUSCO score of 35.2% (Table 24). The additions of non-canonical, complete gene models and partial gene models that have at least either a start or stop codon increased the BUSCO score to 35.4% (Table 24). Round two of MAKER resulted in 15,091 gene models, 6,472 of which were complete. The complete models had a BUSCO score of 30.9% (Table 24). Of these complete models, 4,271 were also canonical. These canonical, complete models had a BUSCO score of 20.7% (Table 24). The addition of non-canonical, complete gene models and partial gene models that had at least either a start or stop codon increased the BUSCO score to 40.9% (Table 24). Round three of MAKER resulted in 13,761 gene models, 4,749 of which were complete. The
complete models had a BUSCO score of 25.2% (Table 24). Of these complete models, 4,470 were also canonical. These canonical, complete models had a BUSCO score of 23.6% (Table 24). The addition of non-canonical, complete gene models and partial gene models that had at least either a start or stop codon increased the BUSCO score to 42.9% (Table 24). Since this assembly annotated much better than the SPAdes assembly the rest of the annotation process was only completed using the MaSURCA assembly for *L. amplifica*.

**Braker annotations**

Braker was run twice on both *L. aenigmaticum* and *L. daileyi*, once with only EST evidence and once with both EST and protein evidence. For *L. aenigmaticum*, the run with only EST evidence resulted in 14,039 total genes, 8,088 of which were complete and canonical. These gene models had a BUSCO score of 42.4% (Table 25). The additions of non-canonical, complete gene models, and partial gene models that had at least either a start or stop codon increased the BUSCO score to 44.8% (Table 25). The run with EST and protein evidence resulted in 14,425 total genes, 8,232 of which were complete and canonical. These gene models had a BUSCO score of 44.7% (Table 25). The additions of non-canonical, complete gene models, and partial gene models that had at least either a start or stop codon increased the BUSCO score to 47.7% (Table 25).

For *L. daileyi*, the run with EST evidence only resulted in 13,586 total genes, 7,860 of which were complete and canonical. These gene models had a BUSCO score of 42.3% (Table 26). The addition of non-canonical, complete gene models and partial gene models that had at least either a start or stop codon increased the BUSCO score to 51.6% (Table 26). The run with both EST and protein evidence resulted in 14,205 total genes, 8,428 of which were complete and canonical. These gene models had a BUSCO score of 47.4% (Table 26). The addition of non-
canonical, complete gene models and partial gene models that had at least either a start or stop codon increased the BUSCO score to 57.5% (Table 26).

**GMAP annotations**

Mapping the master transcriptome of *L. aenigmaticum* to its genome resulted in 18,541 total gene models, 8,791 of which were complete and canonical. These gene models had a BUSCO score of 55.4% (Table 27). The addition of non-canonical, complete gene models and partial gene models that had at least either a start or stop codon resulted in 14,784 gene models with a BUSCO score of 67.7% (Table 27).

Mapping the master transcriptome of *L. daileyi* to its genome resulted in 14,400 total gene models, 5,044 of which were complete and canonical. These gene models had a BUSCO score of 47.7% (Table 27). The addition of non-canonical, complete gene models and partial gene models that had at least either a start or stop codon resulted in 10,200 gene models with a BUSCO score of 61.8% (Table 27).

Mapping the master transcriptome of *L. aenigmaticum* to the genome of *L. amplifica* resulted in 57 total gene models, 20 of which were complete and canonical. These genes had a BUSCO score of 0.2% (Table 27). The addition of non-canonical, complete gene models and partial gene models that had at least either a start or stop codon resulted in 42 gene models with a BUSCO score of 0.2% (Table 27).

Mapping the master transcriptome of *L. daileyi* to the genome of *L. amplifica* resulted in 32 total gene models, 14 of which were complete and canonical. These genes had a BUSCO score of 0.1% (Table 27). The addition of non-canonical, complete gene models and partial gene models that had at least either a start or stop codon resulted in 26 gene models with a BUSCO score of 0.1% (Table 27).
Final Annotation Gene Models

For *L. aenigmaticum*, it was decided to move forward with the second round of MAKER, which resulted in 6,922 genes, because it yielded a lower number of monoexonic genes than round one and a higher BUSCO score than round three. It was also decided to move forward with the second round of Braker (which used both EST and protein evidence), which resulted in 8,866 genes, since it had a higher number of gene models and a higher BUSCO score than the round that used only the EST data. In addition, the GMAP annotation for this species resulted in 14,874 gene models. In all three cases, the gene models included both canonical and non-canonical complete genes and partials genes with at least a start or stop codon.

Removing the overlapping genes across the outputs of these three programs resulted in 5,608 MAKER gene models, 6,750 Braker gene models, and 8,712 GMAP gene models (Table 28). In total, 21,070 unique gene models were identified for *L. aenigmaticum*; these had a BUSCO score of 50.9% (Table 28). Of these, 17,663 were multiexonic genes (Table 29) and 3,407 were monoexonic genes (Table 30). The average overall gene size was 8,324 bp, the average overall coding sequence (CDS) size was 1,225 bp, and the average overall exon size was 224 bp (Table 31). The average size of the multiexonic genes was 9,784, with the range being 96–11,549 (Table 29). The average size of the multiexonic CDS was 1,318 bp, with a range of 63–29,283 bp. The average number of exons in a multiexonic gene was 6.3; these exons had an average size of 208 bp, with a range of 10–14,966 bp. The average number of introns per multiexonic gene was 5.3; these introns had an average size of 1,584 bp, with a range of 10–175,174 bp (Table 31). The average size of the monoexonic genes was 775 bp, with a range of 96–11,549 bp (Table 30). Overall, 14,096 (67%) of the final gene models were annotated with a protein domain.
For *L. daileyi*, it was decided to move forward with the second round of MAKER, which resulted in 10,011 genes, because it yielded a lower number of monoexonic genes than round one and a higher BUSCO score than round three. It was also decided to move forward with the second round of Braker (which used both EST and protein evidence), which resulted in 11,212 genes, since it had a higher number of gene models and a higher BUSCO score than the round that used only the EST data. In addition, the GMAP annotation for this species resulted in 10,200 gene models. In all three cases, the gene models for each of these programs include both canonical and non-canonical complete genes and partials genes with at least a start or stop codon.

Removing the overlapping genes between these three programs resulted in 6,625 MAKER gene models, 7,791 braker gene models, and 4,668 GMAP gene models. In total, there were 19,084 unique gene models for *L. daileyi* with a BUSCO score of 51.2% (Table 28). Of these, 15,759 were multiexonic genes (Table 29) and 3,325 were monoexonic genes (Table 30). The average overall gene size was 7,818 bp, the average overall CDS size was 1,181 bp, and the average overall exon size was 227 bp (Table 31). The average size of the multiexonic genes was 9,282 bp, with a range of 100–119,396 bp (Table 29). The average size of multiexonic CDS was 1,280 bp, with a range of 20–22,581 bp. The average number of exons in a multiexonic gene was 6; these exons had an average size of 210 bp and a range of 10–8,920 bp. The average number of introns per multiexonic gene was 5; these introns had an average size of 1,548 bp and a range of 10–99,372 bp (Table 31). The average size of the monoexonic genes was 879 bp, with a range of 51–11,505 bp (Table 30). Overall, 12,335 of these gene models were annotated with protein domains.

For *L. amplifica*, as stated previously, it was decided to move forward with the MaSURCA assembly since it annotated better with the MAKER pipeline. For the MAKER annotation, it was decided to move forward with the third round, which resulted in 8,319 genes, because it had a
lower number of monoexonic genes and had a higher BUSCO than round two. In addition, the GMAP annotation for this species resulted in 68 gene models. In all three cases, the gene models include both canonical and non-canonical complete genes, and partials genes with at least a start or stop codon.

Removing the overlapping genes between these programs resulted in 8,318 MAKER gene models and 42 GMAP gene models. In total, there were 8,358 unique gene models for *L. amplifica* with a BUSCO score of 42.9% (Table 28). Of these, 6,635 were multiexonic (Table 29) and 1,725 were monoexonic genes (Table 30). The average overall gene size was 6,452 bp, the average overall CDS size was 1,129 bp, and the average overall exon size was 272 bp (Table 31). The average size of the multiexonic genes was 7,923 bp, with a range of 233–112,622 bp (Table 29). The average size for multiexonic CDS was 1,239 bp, with a range of 162–24,507 bp. The average number of exons in a multiexonic gene was 5; these exons had an average size of 249 bp, with a range of 10–7,855 bp. The average number of introns per multiexonic gene was 4; these introns had an average size of 1,682 bp, with a range of 10–55,259 bp (Table 31). The average size of the monoexonic genes was 796 bp, with a range of 111–16,740 bp (Table 30). Overall, 6,691 of these gene models were annotated with protein domains.

**Differences between *L. aenigmaticum* and *L. daileyi***

*Orthofinder Gene and Species Trees*

The Orthofinder species tree, a majority consensus tree of 5,337 orthogroup gene trees, for the 14 individual transcriptomes of the *Litobothrium* species is shown in Figure 14. Whereas the multiple specimens of *L. nickoli* and *L. aenigmaticum* were found to comprise monophyletic assemblages, that was not the case for *L. daileyi*. Two of the three specimens of *L. daileyi* grouped more closely with the clade of *L. aenigmaticum* specimens than with the third specimen of *L.*
The clustering of these transcriptomes resulted in 718 single-copy orthogroups. Examining the gene trees for these 718 orthogroups revealed that the 5 specimens of *L. aenigmaticum* comprised a monophyletic group in only 9 of these trees. In the 709 other trees, the specimens of *L. aenigmaticum* either formed a paraphyletic group or were found to be identical to at least one or more of the specimens of *L. daileyi*. Examination of the gene trees of the 4,617 multicopy orthogroups showed that the 5 specimens of *L. aenigmaticum* formed a monophyletic group in 22 trees. In these 22 trees, the specimens of *L. daileyi* were monophyletic in only 4. In the other 4,595 trees, the specimens of *L. aenigmaticum* either formed a paraphyletic group and/or were found to be identical to at least one or more of the specimen of *L. daileyi*.

The Orthofinder species tree that resulted from clustering the genome annotation gene models for *L. aenigmaticum*, *L. daileyi*, and *L. amplifica* with the gene models for *Echinococcus multilocularis* from Wormbase ParaSite resulted in a tree in which *L. aenigmaticum* and *L. daileyi* were sister species and *L. amplifica* was sister to that group (Figure 15). This clustering resulted in 684 single-copy orthogroups. Examination of the gene trees for these orthogroups revealed that 263 trees displayed phylogenetic relationship that match our current hypothesis of the evolutionary history of this group (i.e., the same relationship displayed in the species tree). Furthermore, all 263 of these trees demonstrated phylogenetic distance between *L. aenigmaticum* and *L. daileyi*, indicating the two are separate species. Another 205 of the single-copy gene trees also showed phylogenetic distance between *L. aenigmaticum* and both other litobothriidean species but reflected interrelationships that differed from our current hypothesis of the phylogenetic relationships of this group.

*QUAST LG*
Comparison of the genome of *L. aenigmaticum* with that of *L. daileyi* using QUAST LG revealed multiple differences between the two genomes. In total, 22,100,993 bp of the *L. aenigmaticum* assembly could not be aligned to that of *L. daileyi*. Furthermore, 140,533,210 bp of the *L. aenigmaticum* assembly was reported to be misassembled when compared to that of *L. daileyi*. Those misassembled regions contained 2,810 contigs that consisted of 49 contig relocations, 2,116 contig translocations, and 39 contig inversions. There were also 214 possible transposable elements and 162,478 indels (with a total length of 3,590,728 bp) reported from *L. aenigmaticum* that were not found in the genome of *L. daileyi*.

**Mapping cyclophyllidean genes**

Tsai et al. (2013) found that there were three gene families, the laminin family, novel protocadherin family, and thrombospondin-containing family, that were expanded in Platyhelminthes overall. The laminin family (1 gene present in the cyclophyllideans) mapped to 1 unique location in *L. aenigmaticum*, 1 unique location in *L. daileyi*, and 4 unique locations in *L. amplifica*. The novel protocadherin family (3–6 genes present in the cyclophyllideans) mapped to 8 unique locations in *L. aenigmaticum*, 6 unique locations in *L. daileyi*, and 9 unique locations in *L. amplifica*. The thrombospondin-containing family (5–8 genes present in the cyclophyllideans) mapped to 49 unique locations in *L. aenigmaticum*, 46 unique locations in *L. daileyi*, and 61 unique locations in *L. amplifica* (Figure 16; Table 32).

Tsai et al. (2013) found two gene families, the LDL receptor family and galactosyl transferases family, that were expanded in only the digeneans and tapeworms. The LDL receptor (18–26 genes present in the cyclophyllideans) mapped to 2 unique locations in *L. aenigmaticum*, 2 unique locations in *L. daileyi*, and 4 unique locations in *L. amplifica*. The galactosyl transferase
family (15–20 gene present in the cyclophyllideans) mapped to 19 unique locations in *L. amplifica*, 15 unique locations in *L. daileyi*, and 27 unique locations in *L. amplifica* (Figure 17; Table 32).

Tsai et al. (2013) found two gene families, CD2 domain containing protein family and novel transmembrane family, that were tapeworm-specific. The CD2 domain containing protein family (1–3 genes present in the cyclophyllideans) mapped to 5 unique locations in *L. aenigmaticum*, 5 unique locations in *L. daileyi*, and 10 unique locations in *L. amplifica*. The novel transmembrane family (2–13 genes present in the cyclophyllideans) mapped to 1 unique location in *L. aenigmaticum*, it did not map to any locations in *L. daileyi*, and it mapped to 1 unique location in *L. amplifica* (Figure 18; Table 32).

Tsai et al. (2013) found 6 gene families, the diagnostic antigen 50, tegumental dynein light I antigen, BTB, BACK, Kelch protein family, ortoperin like genes, novel repeat domain family, and ubiquitin conjugating enzyme families, that had tapeworm-specific expansions. The diagnostic antigen 50 family (9–29 genes present in the cyclophyllideans) mapped to 1 unique location in *L. aenigmaticum*, 1 unique location in *L. daileyi*, and 2 unique locations in *L. amplifica*. The tegumental dynein light I antigen family (5–10 genes present in the cyclophyllideans) mapped to 2 unique locations in *L. aenigmaticum*, 2 unique locations in *L. daileyi*, and 2 unique locations in *L. amplifica*. The BTB, BACK, and Kelch protein family (3–24 genes present in the cyclophyllideans) mapped to 1 unique location in *L. aenigmaticum*, 1 unique location in *L. daileyi*, and it did not map at all to *L. amplifica*. The ortoperin-like family (4–9 genes present in cyclophyllideans) mapped to 2 unique locations in *L. aenigmaticum*, 2 unique locations in *L. daileyi*, and 2 unique locations in *L. amplifica*. The novel repeat domain family (1 gene present in the cyclophyllideans) mapped to 5 unique locations in *L. aenigmaticum*, 14 unique locations in *L. daileyi*, and 1 unique location in *L. amplifica*. The ubiquitin conjugating enzyme (5–11 genes
present in the cyclophyllideans) mapped to 1 unique location in *L. aenigmaticum*, did not map to *L. daileyi*, and mapped to 4 unique locations in *L. amplifica* (Figure 19; Table 32).

Tsai et al. (2013) found two gene families, the novel flatworm gene family (that is superficially similar to the Zona pellucida-like domain) and the novel taeniid protein, that were present in all cyclophyllideans, and that are especially expanded in the *Taenia* genomes. The novel flatworm family (1–2 genes present in the cyclophyllideans) mapped to 3 unique location in *L. aenigmaticum*, 5 unique locations in *L. daileyi*, and 6 unique locations in *L. amplifica*. The novel Taeniid protein did not map to any of the litobothriideans (Figure 20; Table 32).

Tsai et al. (2013) found 6 gene families, the novel *E. multilocularis* gene family (novel domains: novel_000011, novel_000049), novel *E. multilocularis* gene family (novel domains: novel_000051, novel_002642), novel *E. multilocularis* gene family (similar to Chromo-domain family), novel *H. microstoma* family (gag-pol transposable element), novel *E. multilocularis* gene family (similar to Chromo-domain family), novel *H. microstoma* family (transposable element), and protein kinase family, that have undergone species-specific expansions. The novel *E. multilocularis* family (with the protein domains: novel_000011 and novel_000049) did not map to any of the litobothriidean genomes. The novel *E. multilocularis* family (with protein domains: novel_000051, novel_002642; there are typically 6 genes in the cyclophyllideans but 46 genes in this species) mapped to 25 unique locations in *L. aenigmaticum*, 29 unique locations in *L. daileyi*, and 93 unique locations in *L. amplifica*. The novel *E. multilocularis* family (similar to Chromo-domain; 3 genes present in most cyclophyllideans but expanded to 32 genes in *E. multilocularis*) mapped to 1 unique location in *L. aenigmaticum*, and 1 unique location in *L. daileyi*, but it did not map to *L. amplifica*. The novel *H. microstoma* family (gag-pol TE; the family is expanded within *H. microstoma* to include 28 genes) mapped to 2 unique locations in *L. aenigmaticum*, 2 unique
locations in *L. daileyi*, and 2 unique locations in *L. amplifica*. The novel *H. microstoma* family (TE; the family is expanded within *H. microstoma* to include 47 genes) mapped to 5 unique locations in *L. aenigmaticum*, 4 unique locations in *L. daileyi*, and 4 unique locations in *L. amplifica*. The protein kinase family (6–16 genes present in cyclophyllideans) mapped to 2 unique locations in *L. aenigmaticum*, 2 unique locations in *L. daileyi*, and 2 unique locations in *L. amplifica* (Figure 21; Table 32).

**Identifying litobothriidean-specific gene families**

Parsing the output of clustering the genome annotation gene models for *L. aenigmaticum*, *L. daileyi*, and *L. amplifica* with the genome annotations of *E. multilocularis*, *H. microstoma*, *T. multiceps*, *S. solidus*, and *S. mansoni* from Wormbase ParaSite resulted in the identification of 123 orthogroups that appear to be unique to the Litobothriidea. The annotations for the diphyllobothriideans were not included since their genomes are too fragmented.

The functional enrichment analysis, with *L. aenigmaticum* as the background, indicated that the following biological process GO terms were enriched in these 123 orthogroups: system process, feeding behavior (single-organism behavior, locomotory behavior, reproductive behavior, multi-organism behavior), catabolic process, intercellular transport, sexual reproduction (multi-multicellular organism process), response to chemical, pigmentation, regulation of biological quality, methylation, protein folding, developmental growth, cell proliferation, single organism metabolic process, establishment of localization, hormone metabolic process, protein activation cascade, regulation of molecular function, immune response (immune effector process, production of molecular mediator of immune response), response to biotic stimulus, detection of stimulus, response to endogenous stimulus, response to abiotic stimulus, response to external stimulus, maintenance of location, muscle adaptation, and response to stress (Fig. 22A). When *L. daileyi*
was used as the background for the functional enrichment analysis it was found that the following biological GO terms were also enriched: multi-organism cellular process (interspecies interaction between organisms), organic substance metabolic process, nitrogen compound metabolic process, and cellular metabolic process were enriched with this background (Fig. 22B). Furthermore, when *L. amplifica* was used as the background the following additional biological process terms were found to be enriched: single-organism metabolic process, organic substance metabolic process, localization of cell, response to estrogen, regulation of molecular function, response to chemical, muscle adaptation, reproductive behavior (single-organism behavior, multi-organism behavior), and multi-multicellular organism process were enriched biological process GO terms (Fig. 22C)

The functional enrichment analysis, with *L. aenigmaticum* as the background, indicated that the following cellular component GO terms were enriched: cell-cell junction, extracellular organelle (extracellular region part), synaptic membrane (synapse part), outer membrane, and membrane (Fig. 23A). All the same terms that were enriched with the *L. aenigmaticum* background were also enriched with the *L. daileyi* background (Fig. 23B). When *L. amplifica* was used as the background, the following additional cellular component GO term was enriched: membrane-bounded organelle (Fig. 23C).

The functional enrichment analysis, with *L. aenigmaticum* as the background, indicated that the following molecular function GO terms were enriched: MAP kinase activity, receptor activity (signaling receptor activity), structural constituent of muscle, enzyme regulator activity (guanyl-nucleotide exchange factor activity), neurotransmitter binding, drug transporter activity, transferase activity, quaternary ammonium group binding, drug binding, amide binding, lipid binding, small molecule binding, oxidoreductase activity, hydrolase activity, cofactor binding, ion binding, toxin transporter activity, neurotransmitter transporter activity, cofactor transporter
activity, carbohydrate derivative transporter activity, carbohydrate transporter activity, transmembrane transporter activity (substrate-specific transporter activity), and channel regulator activity (Fig. 24A). With *L. daileyi* as the background the following additional GO terms were enriched: signal transducer activity downstream of receptor, binding bridging, sulfur compound binding, protein binding, carbohydrate derivative binding, heterocyclic compound binding, vitamin transporter activity, carbohydrate transporter activity, RNA polymerase II transcription factor activity ligand-activated sequence-specific DNA binding, and transmembrane transporter activity (substrate-specific transporter activity) (Fig. 24B). With *L. amplifica* as the background, the following additional GO terms were enriched: RNA polymerase II transcription factor activity ligand-activated sequence-specific DNA binding, xenobiotic transporter activity, and vitamin transporter activity (Fig. 24C).

Parsing the Orthofinder results revealed that 5 orthogroups were specific to *L. aenigmaticum*. The genes in these groups were not functionally annotated by EnTAP.

Parsing the Orthofinder results revealed that four orthogroups were specific to *L. daileyi*. A functional enrichment analysis revealed that the enriched biological process GO terms were system process (single-multicellular organism process), cell communication, biosynthetic process, response to abiotic stimulus, macromolecule localization, cellular component biogenesis (cellular component organization), single-organism developmental process (anatomical structure development), regulation of biological quality, primary metabolic process, cellular metabolic process, organic substance metabolic process, single organism signaling, single-organism cellular process, cellular localization, response to external stimulus, response to stress, regulation of biological process, establishment of localization, response to chemical, and cellular response to stimulus (Fig. 25). The enrich cellular component GO terms were non-membrane-bounded
organelle, synapse part, and organelle part (Fig. 25). The enriched molecular function GO terms were protein binding, transferase activity, small molecule binding, ion binding, heterocyclic compound binding, and organic cyclic compound binding (Fig. 25).

Parsing the Orthofinder results revealed that 1 orthogroup was specific to *L. amplifica*. The genes in this orthogroup were not functionally annotated by EnTAP.

**Discussion**

**Transcriptome Assemblies**

*Litobothrium daileyi* and *L. nickoli* had similar transcriptome sizes at 9–21 Mb (13 Mb master transcriptome) for the former and 14–23 Mb (15 Mb master transcriptome) for the latter. The majority of the transcriptomes for *L. aenigmaticum* were larger, ranging from 24–37 Mb (with a 21 Mb master transcriptome). Within a species, transcriptome size is known to vary across tissues and developmental stages (Coate and Doyle, 2015). To control for these factors, transcriptomes were generated for whole worms, all of which were adult specimens. We expected some variation in transcriptome size across the three species but the substantially larger size of the transcriptome of *L. aenigmaticum* seems worth additional consideration. Possible explanations for larger transcriptome size include whole genome duplication, small-scale duplication of genes, larger cell size, and more diverse cell types (Coate and Doyle, 2015). Given the genome of *L. aenigmaticum* was estimated to be 320–355 Mb in size and those of the *L. daileyi* and *L. amplifica* were found to be 330–412 Mb and 358–473 Mb, respectively, whole genome duplication can be eliminated. The possibility that there have been small-scale gene duplications within the genome of *L. aenigmaticum* cannot be ruled out at this point. This possibility will be further explored in chapter three. Larger cells have been repeatedly shown to have larger transcriptomes in order to maintain their biomass and function (Marguerat and Bähler, 2012; Coate and Doyle, 2015). Furthermore,
species with larger cell sizes tend to be larger relative to species with smaller cell sizes (Marguerat and Bähler, 2012). *Litobothrium aenigmaticum* does conform to this trend in that it is easily five times the size of its congeners; the average *Litobothrium* species is 2–3 mm in length while *L. aenigmaticum* is approximately 1 cm in length. Furthermore, it does seem possible that *L. aenigmaticum* may have more diverse cell types since the scolex contains those 11 unique cell types. However, these final two explanation are impossible to explore in the absence of data on the cell sizes and types of the other litobothriideans, which are currently unavailable.

**Genome Assemblies & Annotations**

Differences in the performance of the assemblers used in this study were expected given they use different methods and make different assumptions regarding genome assembly. The assemblers that use a purely De Bruijn graph method, such as ABySS2 and SOAPdenovo2 (Ekblom and Wolf, 2014), performed poorly for all 3 *Litobothrium* species; the N50 scores resulting from these assemblers rarely were above 10,000. It is possible this is because appropriate k-mer sizes were not used given the k-mer size MaSURCA selected for the assembly of the litobothriidean genomes was 71, but ABySS2 and SOAPdenovo were run with k-mer sizes of 27, 31, and 35. If these programs were re-run with k-mer sizes closer to 71 it is possible these programs would perform better. Assemblers that use a combined or modified De Bruijn graph method performed much better for most of the litobothriidean species; these were MaSURCA, which uses a hybrid assembly approach that incorporates both De Bruijn and extension-based methods, and SPAdes, which uses a paired De Bruijn graph method (Bankevich et al., 2012; Ekblom and Wolf, 2014). It was somewhat surprising that SPAdes performed so well given it was designed for smaller genome sizes (it is typically the assembler of choice when working with bacteria). The litobothriidean genomes should have been at the higher end of the genome sizes this assembler is
able to effectively handle (Bankevich et al., 2012). However, SPAdes worked just as well, if not better than, MaSURCA; the assemblies from both assemblers were of very similar quality in the case of *L. daileyi* and *L. amplifica*. The exception was *L. aenigmaticum*. For an unknown reason MaSURCA did not perform well when assembling the genome of this species. Since SPAdes outperformed both of the other assemblers in the case of *L. aenigmaticum*, choosing this assembly to annotate was straightforward. In terms of *L. daileyi*, comparison of the N50, number of contigs, and BUSCO scores between the SPAdes and MaSURCA indicated that the former was the better assembly, and thus this assembly was chosen for annotation. With respect to *L. amplifica*, the statistics for the SPAdes and MaSURCA assemblies were very similar. As a consequence, both assemblies were annotated with MAKER in order to help decide which assembly was most appropriate for annotation. The MaSURCA assembly annotated substantially better than the SPAdes assembly; therefore, the MaSURCA assembly was used for the annotation process.

All three litobothriidean genomes are fragmented, draft assemblies. The average of the genome size estimates for *L. aenigmaticum* was 338 Mb, suggesting that the genome for *L. aenigmaticum* assembled approximately 87% of the gene space. The average of the genome size estimates for *L. daileyi* was 364 Mb, which indicates that the genome for *L. daileyi* assembled approximately 83% of the gene space. The average of the genome estimates for *L. amplifica* was 406 Mb, which indicates that the genome for *L. amplifica* assembled approximately 79% of the gene space. The BUSCO scores for the three litobothriidean genomes ranges from 54.5–60.2%.

As expected, the litobothriidean genomes were found to be larger than those of the cyclophyllideans (103–240 Mb) but smaller than those of the diphyllobothriideans (531 Mb–1.3 Gb) (Tsai et al., 2013; Zheng et al., 2013; Bennett et al., 2014; Maldonando et al., 2017; International Helminth Genomes Consortium, 2019). In terms of quality and completion, when
compared to the 15 published tapeworm genomes, the litobothriidean genomes are much more fragmented than those of *Echinococcus multilocularis*, *Echinococcus granulosus*, *Hymenolepis microstoma*, *Taenia asiatica*, *Taenia multiceps*, and *Taenia saginata*. These cyclophyllidean taxa have N50’s of 34.2–44.8Mb and BUSCO scores of 65.7–68.8%. The litobothriidean genomes are of similar quality to those of *Echinococcus canadensis*, *Hymenolepis diminuta*, and *Mesocestoides corti*, which have N50’s of 49.9–74.6 Kb and BUSCO scores of 67.3–69.6%. The litobothriidean genomes are more complete than those of the diphyllobothriideans *Diphyllobothrium latum*, *Schistocephalus solidus*, and *Spirometra erinaceieuropaei* and the cyclophyllideans *Hydatigera taeniaeformis* and *Hymenolepis nana*. These taxa have N50’s of 4.6–31.6 Kb and BUSCO scores of 19.1–67.5%.

The litobothriideans are similar to other cestodes with respect to some of the basic features of their genomes. All three litobothriidean genomes have similar GC content, approximately 43%, and have a low proportion of repeat regions, 27–28% in *L. aenigmaticum* and *L. daileyi* and 31% in *L. amplifica*. In comparison to the 12 published cyclophyllidean genomes and 3 diphyllobothriidean genomes, the litobothriideans have similar GC content (the cyclophyllideans have 35.2–43.8% GC content and the diphyllobothriideans have 43–44.9%). The litobothriidean genomes have similar proportion of repeat regions to the cyclophyllideans (7–23% repeat content); the diphyllobothriideans have much more repeat content (46–54%).

The genome annotations for *L. aenigmaticum* and *L. daileyi* are similar in both number of genes and quality. The final annotation set for *L. aenigmaticum* contains 21,070 genes with a BUSCO score of 59%, the final annotation set for *L. daileyi* contains 19,084 genes with a BUSCO score of 60%. In contrast, the final annotation set for *L. amplifica* contains both fewer genes, 8,358, and is of lower quality, with a BUSCO score of 54%. This is likely due to the lack of species-
specific evidence available for *L. amplifica* given transcriptomic data were not generated for this species. The availability of EST and protein evidence for *L. amplifica* would not only improve the predictions from MAKER but would also allow for annotation with the Braker pipeline and the GMAP aligner.

The number of predicted genes in the litobothriidean genomes (8,358–21,070) is more similar to number in the diphyllobothriideans (19,966–39,557) than the cyclophyllideans (10,614–13,777) (remember, 8,358 comes from *L. amplifica* and is likely an underestimate) (Tsai et al., 2013; International Helminth Genomes Consortium, 2019).

Overall, the gene statistics are very similar across the three litobothriidean species. Average gene size is 6,452–8,324, average CDS size is 1,181–1,225, and average exon size 224–272. *L. amplifica* differs in average size of multiexonics (7,923 versus 9,282–9,784) and had a somewhat smaller number of exons (5 vs 6) and introns (4 versus 5). The average exon size of the litobothriideans (224–272) is comparable to both the diphyllobothriideans (225–265) and cyclophyllideans (219–245). However, the average number of exons in a multiexonic gene for the litobothriideans (5–6.3) is more similar to that of the cyclophyllideans (3–5) than the diphyllobothriideans (2). This difference though could possibly be due to the fragmented nature of the diphyllobothriidean genomes (N50 scores range from 4.6–31 Kb) it is likely that improvement of these genomes will increase this number. Average intron size of the litobothriideans (1,548–1,682) is comparable to those of the diphyllobothriideans (1,062–2,109) but is larger than what is seen in the cyclophyllideans (457–860) (Tsai et al., 2013; International Helminth Genomes Consortium, 2019).

**Differences between *L. daileyi* and *L. aenigmaticum***
Examination of the gene trees for the single copy and multicopy transcripts for *L. aenigmaticum*, *L. daileyi*, and *L. nickoli* provided little insight into the question of the conspecificity of *L. aenigmaticum* and *L. daileyi*. In fact, the gene trees from the transcripts show little difference between *L. aenigmaticum* and *L. daileyi*. Examination of the single-copy ortholog gene trees for the genome annotation gene models revealed that 468 of these gene trees show differences between *L. aenigmaticum* and *L. daileyi*. Specifically, 263 trees reflected the current hypothesis regarding the phylogenetic relationships within this group and demonstrated phylogenetic distance between *L. aenigmaticum* and *L. daileyi*. Another 205 trees indicated that there is phylogenetic distance between *L. aenigmaticum* and the other litobothriideans but does not reflect the current phylogenetic hypothesis of the group.

To further examine this issue, the genomes of *L. aenigmaticum* and *L. daileyi* were also compared using QUAST LG. This analysis revealed that approximately one third of the genome for *L. aenigmaticum* is reported to be misassembled when compared to *L. daileyi*; this means that there are coding differences in these regions. Additionally, approximately 22 million bp of the *L. aenigmaticum* genome was not able to be aligned to the *L. daileyi* genome. This indicates there are numerous coding region differences between the genomes of these two species, providing further support for the ascertain that *L. aenigmaticum* and *L. daileyi* are separate species.

In combination, the differences in morphology, internal anatomy, reproductive strategies, and juvenile stages and the difference in the gene trees and genomes support the conclusion that *L. aenigmaticum* and *L. daileyi* are separate species.

**Testing the Conclusions of Tsai et al. (2013)**

Mapping the genes that Tsai et al. (2013) hypothesized to have novel protein domains or to be expanded, either within tapeworms or within certain cyclophyllidean groups, to the
litobothriidean genomes allowed the generality of their hypotheses to be examined. Each of their hypotheses is evaluated below.

Tsai et al. (2013) hypothesized that the laminin family, thrombospondin containing family, and novel protocadherin family were expanded in all Platyhelminthes. This hypothesis appears to be supported since all 3 of these gene families appear to have undergone expansions in the litobothriideans. In fact, *L. amplifica* appears to have undergone a further expansion in the laminin gene family, *L. aenigmaticum* and *L. amplifica* appear to have had a further expansion in the novel protocadherin family, and all 3 litobothriideans have a further expansion in the thrombospondin containing family than seen in the cyclophyllideans.

Tsai et al. (2013) also hypothesized that the LDL receptor family and galactosyl transferases family experienced expansions in all of the trematodes and tapeworms. This hypothesis is supported for the galactosyl transferases family since the litobothriideans had similar gene numbers to the cyclophyllideans. However, this hypothesis is rejected in regards to the LDL receptor since it appears none of the litobothriideans have undergone an expansion.

Although not expanded, Tsai et al. (2013) hypothesized that the CD2 domain-containing protein family and novel transmembrane family are tapeworm specific. This hypothesis is supported given that the genes from both of these families mapped to the 3 litobothriidean genomes. In fact, the litobothriideans appear to have experienced an expansion in the CD2 domain containing protein family since they seem to have more genes in this family than the cyclophyllideans.

The fourth hypothesis involved tapeworm-specific expansions of the following gene families: the diagnostic antigen 50, tegumental dynein light I antigen, BTB, BACK, Kelch protein family, ortoperin-like genes, novel repeat domain family, and ubiquitin conjugating enzyme
families. The novel repeat domain family was found to be expanded in all 3 litobothriideans species examined. In fact, this gene family appears to be more expanded in *L. aenigmaticum* and *L. daileyi* than in the cyclophyllideans. Therefore Tsai et al.’s (2013) hypothesis that this family is expanded in all tapeworms is supported. However, none of the other five gene families were found to be expanded in the litobothriideans. Therefore, the hypothesis that these 5 gene families are expanded in all tapeworms is rejected.

Tsai et al. (2013) hypothesized that the novel flatworm gene family (that is superficially similar to Zona pellucida-like domain) and a novel taeniid protein were present in all cyclophyllideans, but they are especially expanded in the *Taenia* genomes. The hypothesis is supported in regards to the novel taeniid protein since none of the genes in this family mapped to any of the litobothriidean genomes. The hypothesis is rejected in terms of the flatworm gene family (that is superficially similar to Zona pellucida-like domain) since it was found to be more expanded in the litobothriideans than the *Taenia* species.

Tsai et al. (2013) also hypothesized that the following gene families have undergone species-specific expansions: novel *E. multilocularis* gene family (novel domains: novel_000011, novel_000049), novel *E. multilocularis* gene family (novel domains: novel_000051, novel_002642), novel *E. multilocularis* gene family (similar to Chromo-domain family), novel *H. microstoma* family (gag-pol transposable element), novel *H. microstoma* family (transposable element), and protein kinase family. In general, the results support this hypothesis. The exception was that the novel *E. multilocularis* family (with the following novel protein domains: novel_000051, novel_002642) appears to also be expanded within all 3 *Litobothrium* species, particularly in *L. amplifica*. It is possible that there have been two independent expansions of this
gene family, however, since it was not expanded *E. multilocularis* alone this hypothesis is rejected for the novel *E. multilocularis* family.

**Identifying litobothriidean-specific gene families**

Clustering the litobothriidean genome annotation models with those of *E. multilocularis*, *H. microstoma*, *T. multiceps*, *S. solidus*, and *S. mansoni* resulted in the identification of 123 orthogroups that were specific to the litobothriideans. The functional enrichment analysis of these gene families resulted in a large number of enriched GO terms. Since it is not possible to discuss all of these terms, this section will focus primarily on the terms with a p-adjusted value of 0.1 (p<0.001).

Functional enrichment analyses of the litobothriidean specific orthogroups indicated that several of these orthogroups were associated with response to stimuli, specifically response to chemical, response to abiotic stimulus, and protein activation cascade. (Detection of stimulus, response to biotic stimulus, response to external stimulus, response to stress, and response to endogenous stimulus also were found to be significant, p<0.05, but did not meet the p<0.001 cut off.) There are many stimuli that tapeworms need to detect and to which they need to successfully respond in order to complete their life cycle. When a larval tapeworm enters a vertebrate host it needs to be able to detect environmental stimuli that indicate whether it is in the correct host. In the cases of the elasmobranch tapeworms it is believed that this stimulus is likely connected to the presence of urea (Hamilton and Byram, 1974; Cherry et al., 1991). Once the tapeworm has detected this stimulus, initiation of the developmental process during which the larval form transitions into the juvenile and then the adult form begins. Therefore, the genes associated with this process could potentially be classified as response to a chemical, external, and/or endogenous stimulus. Once the tapeworm has transitioned to the juvenile stage, it then needs to position and attach itself within
the spiral intestine of its host. Litobothriidean tapeworms are not found throughout the entirety of the spiral intestine of the pelagic thresher shark; instead they tend to aggregate in the anterior half of the spiral intestine (Gallagher, personal observation). This suggests that they have some site preference, perhaps based on the nutrition that is available in that portion of the spiral intestine. This means that tapeworms must be able to detect when they are in the preferred sites within the spiral intestine at which point they respond to that stimulus by attaching to the mucosa in that region. Once again, the genes associated with this process could potentially be classified as response to a chemical, external, and/or endogenous stimulus. Once a tapeworm has attached to the intestinal mucosa and has reached the adult stage it will begin the process of strobilization (i.e., producing proglottids). With the exception of *L. aenigmaticum*, in litobothriidean tapeworms, these proglottids drop from the strobila prior to fertilization. The free proglottids must find each other in order to exchange gametes (Dailey, 1969; Dailey, 1971; Kurochkin and Slankis, 1973; Caira and Runkle, 1993; Olson and Caira, 2001; Caira et al., 2014a). Thus, even free proglottids must be able to detect and respond to chemical, biotic, abiotic, and endogenous stimuli to successfully reproduce.

In general, all tapeworms need to be able detect and respond to both internal and external stimuli in order to successfully complete the tasks described above. The type of stimuli to which they respond likely differs depending on a variety of biological factors, such as their hosts, their reproductive strategy, etc. Since all the litobothriideans in this study are closely related and parasitize the same specimens of definitive host, it seems likely that some or all of the gene families associated with response to stimulus that have been identified as litobothriidean-specific may be employed for responding to host- or life strategy-specific stimuli. The exact stimuli to which such genes families may be responding is unclear at this time.
The function of some of the litobothriidean-specific orthogroups are related to the internal processes of these tapeworms. For example, several orthogroups were associated with the terms transmembrane transporter, establishment of localization, and intracellular transport. This implies that litobothriideans have specific gene families that are involved in the movement or tethering of substances. The transmission electron microscopy results from chapter one revealed a large number of secretory vesicles in the anterior region of the body of *L. aenigmaticum*. It was hypothesized that at least some of these secretory products are extruded in order to elicit an inflammatory host response (Gallagher et al., 2017). It is thus not surprising to find that gene families specific to the movement of substances within this species. However, no glands or other similar structures have been observed in any of the typical litobothriideans (Dailey, 1969; Dailey, 1971; Kurochkin and Slankis, 1973; Caira and Runkle, 1993; Olson and Caira, 2001). The presence of these orthogroups in the two typical species of *Litobothrium* is thus puzzling. It is not understood what specific function these gene families are serving in these species.

Some of the litobothriidean-specific orthogroups were associated with aspects of the nervous system since they were annotated with the GO terms neurotransmitter binding, neurotransmitter activity, and synaptic membrane. Tapeworms in general have fairly sophisticated nervous systems that consist of a complex of ganglia in the scolex and two longitudinal nerve cords that run throughout the full length of the body (Roberts and Janovy, 2009). The nervous system of litobothriideans has not been reported to differ morphologically from those found in other cestodes (Dailey, 1969; Dailey, 1971; Kurochkin and Slankis, 1973; Caira and Runkle, 1993; Olson and Caira, 2001; Caira et al., 2014a). Nonetheless, these results suggest that there may be some differences between litobothriideans and other cestodes groups in how they regulate their neurotransmissions.
Some of the litobothriidean-specific orthogroups are associated with drug binding and toxin transmembrane transport. These GO terms are applied to genes that interact with natural or synthetic products that can interfere with the function of the organism. Since the litobothriideans examined here were collected from wild-caught sharks, it is unlikely that these gene families are interacting with synthetic drugs. It seems much more likely that they are interacting with one or more aspects of the host immune system. In fact, the terms immune response and production of molecular mediator of immune response were also found in the enrichment analysis; however, they did not meet the p<0.001 cut off. All tapeworms must interact with, and counter, the immune system of their host. Some cestodes evade the host immune system by secreting products that allow the redirection the host immune system, others release proteases that can digest host proteins and protease inhibitors, and yet others release signaling peptides that act as messengers to the host, inhibiting host essential enzymes, and interfering with host signal transduction pathways (Zheng et al., 2013). Evasion strategies will likely differ depending on the host and cestode taxon. In the case of the litobothriideans, it is likely that the gene families involved in drug binding, toxin transmembrane transporter, and immune response specifically aimed at manipulating or evading specific aspects of the immune system of Alopias pelagicus.

Among the other GO terms that were found to highly enriched in the litobothriideans alone were enzyme regulator activity, quaternary ammonium group binding, guanyl-nucleotide exchange factor activity, system process, regulation of biological process, hormone metabolic process, small molecule binding, and ion binding. These are all very broad terms so it is difficult to identify how they may be affecting the structure or function of the litobothriideans.

Many other GO terms were found to be enriched among the litobothriidean-specific orthogroups but fell above the p<0.001 significance level. Several of them that were of particular
interest to this study are discussed below. The first set of terms is related to reproductive function. Reproduction in the litobothriideans, diphyllobothriideans, and cyclophyllideans differs in two major respects. The first is that most cyclophyllideans undergo asexual reproduction as larvae (Roberts and Janovy, 2009) while that is not the case for litobothriidean or diphyllobothriideans, (Dailey, 1969; Dailey, 1971; Kurochkin and Slankis, 1973; Caira and Runkle, 1993; Olson and Caira, 2001; Roberts and Janovy, 2009; Caira et al., 2014a). The second major difference relates to the release of proglottids from the strobila. The cyclophyllideans do not release their proglottids until after fertilization has occurred (Roberts and Janovy, 2009; Cunningham and Olson, 2010). The diphyllobothriideans do not release their proglottids at all; instead they release their eggs through a uterine pore (Roberts and Janovy, 2009). The litobothriideans release their proglottids prior to fertilization (Dailey, 1969; Dailey, 1971; Kurochkin and Slankis, 1973; Caira and Runkle, 1993; Olson and Caira, 2001; Caira et al., 2014a). This means that free proglottids must locate one another within the spiral intestine in order to mate. So overall, the litobothriideans differ from the cyclophyllideans and diphyllobothriideans in both reproductive processes and behaviors. The litobothriidean-specific gene families associated with reproduction were annotated with the GO terms sexual reproduction and reproductive behavior. It is possible that these gene families are associated with the process of dropping proglottids prior to fertilization and the mechanisms that free proglottids use to locate one another. This behavior is not, however, unique to the litobothriideans. It is seen in several other elasmobranch tapeworm groups, such as members of Onchoproteocephalidea II, Phyllobothriidea, and Rhinobothriidea (Caira and Jensen, 2017). Thus, if these particular gene families are associated with these reproductive behaviors, they are likely not to be restricted to the litobothriideans.
Another GO term of interest was a litobothriidean-specific orthogroup associated with muscle adaptations. In *L. aenigmaticum* these muscle adaptation genes likely contribute to the unusually thick layer of muscular tissue below the tegument in the anterior region of its body (Gallagher et al., 2017). Throughout most of their body, other litobothriideans have a muscle layer that resembles what is typical for a tapeworm (Dailey, 1969; Dailey, 1971; Kurochkin and Slankis, 1973; Caira and Runkle, 1993; Olson and Caira, 2001). The only difference that may be considered a muscular adaptation is that the anterior, cruciform pseudosegments of these species contain musculature (Dailey, 1969; Dailey, 1971; Kurochkin and Slankis, 1973; Caira and Runkle, 1993; Olson and Caira, 2001).

Finally, some of the litobothriidean-specific orthogroups are associated with methylation. DNA methylation is an important mechanism for regulation of gene expression. Some researchers have suggested that the members of Platyheminthes may have lower methylation rates compared to other organisms (Zheng et al., 2013); however, methylation has been found to regulate oviposition in schistosomes (Geyer et al., 2011) and regulate gene expression during the cysticercus stage of *T. solium* (Shumin et al., 2018). Therefore, this implies that methylation is likely a mechanism that the litobothriideans use in order to control gene expression.

Species-specific orthogroups were also identified for *L. aenigmaticum*, *L. daileyi*, and *L. amplifica*. The five that were identified for *L. aenigmaticum* and the one identified for *L. amplifica* were not able to be functionally annotated with EnTAP, although protein domains were assigned to each of these genes. This indicates they are likely true genes but insufficient information is available for these gene families at this time.

The orthogroups specific to *L. daileyi*, on the other hand, were able to be functionally annotated. It appears that the four gene families specific to this species are associated with the
following functions: response to stimulus, biosynthesis, metabolism, regulation of external morphology and internal anatomy, and protein binding. However, since these functions are so similar to those of the litobothriidean-specific orthogroups it is difficult to speculate how they may be contributing the structure or behavior of this species.

It is important to note that all the comparative analyses were performed on taxa that represent only 3 of the 19 cestode orders. The inclusion of taxa that are more closely related to the litobothriideans would likely yield different results; therefore, the conclusions of this study should be re-assessed as other cestodes genomes become available.
Chapter 3: Determining the mechanisms behind the novelty of *L. aenigmaticum*

Introduction

The generation of genomic and transcriptomic resources have now made it possible to address some of the mechanisms that may account for the evolution of the unusual body form of *Litobothrium aenigmaticum* relative to its congeners. As stated in the introduction, there are two general types of changes that are thought to lead to the evolution of novelty, genetic and developmental (Müller and Wagner, 1991; Wagner and Lynch, 2010; Carroll et al., 2005). This chapter will focus on some of the genetic changes that may have occurred. Some of these possible changes are structural gene mutations, chromosomal rearrangements, genome size change, regulatory mutations, gene duplications, transposable element insertions, and the evolution of novel genes (Müller and Wagner, 1991; Lynch and Connery, 2000; Wagner and Lynch, 2010). This chapter will address the following 11 specific hypotheses in terms of *L. aenigmaticum* relative to its congeners: (1) There has been a genome duplication in *L. aenigmaticum*. (2) There have been chromosomal rearrangements in the genome of *L. aenigmaticum*. (3) There have been changes in the way *L. aenigmaticum* regulates its gene expression. (4) There have been expansions in gene families associated with anatomical structure in *L. aenigmaticum*. (5) There has been upregulation of genes associated with anatomical structure in *L. aenigmaticum*. (6) There have been expansions in gene families associated with reproduction in *L. aenigmaticum*. (7) There has been upregulation of genes associated with reproduction in *L. aenigmaticum*. (8) There have been expansions in gene families associated with the production of secretory products in *L. aenigmaticum*. (9) There has been up regulation of genes associated with the production of secretory products in *L. aenigmaticum*. (10) There have been expansions in gene families associated with the immune
response in *L. aenigmaticum*. (11) There has been up regulation of genes associated with immune response in *L. aenigmaticum*.

The first three hypotheses are based on theory regarding the origination of novelty. In regards to the first hypothesis (that there has been a genome duplication in *L. aenigmaticum*), it has been found that species with genome duplications are often larger than their congeners (Coate and Doyle, 2015). Therefore, it is plausible that a genome duplication could explain the size difference between *L. aenigmaticum* and typical litobothriideans. The second hypothesis (that there have been chromosomal rearrangements in *L. aenigmaticum*) is based on the assertion by Müller and Wagner (1991) that chromosomal rearrangements can lead to the origination of novelty. The third hypothesis, that there have been changes in the way that *L. aenigmaticum* regulates its gene expression compared to the other litobothriideans, is based on the argument made by Carroll et al. (2005) that radical changes in morphology are most likely the result of changes in regulatory circuits.

The remaining 8 hypotheses were generated based on the knowledge gained from work by Caira et al. (2014a) and chapter one. These results suggest that *L. aenigmaticum* has undergone a massive restructuring of its morphology and anatomy compared to its congeners. It is expected that this novelty is due to both genomic (hypothesis 4) and transcriptomic changes (hypothesis 5). Caira et al. (2014a) also demonstrated that *L. aenigmaticum* differs from its congeners in regards to reproduction; *L. aenigmaticum* appears to be hyperapolytic while typical litobothriideans are euapolytic (Dailey, 1969; Dailey, 1971; Kurochkin and Slankis, 1973; Caira and Runkle, 1993; Olson and Caira, 2001; Caira et al., 2014a). Therefore, it is thought that there have been either genomic expansions (hypothesis 6) or transcriptomic upregulation (hypothesis 7) in genes associated with reproduction (or potentially both). The transmission electron microscopy work
from chapter one revealed that the scolex proper and cephalic peduncle of *L. aenigmaticum* contains a complex aggregate of 11 novel cell types. Each of these cell types was found to contain secretory vesicles that stained positive with Periodic acid-Schiff (PAS). It was concluded that it is likely that these secretory vesicles contain glycoproteins and/or mucoproteins. It was further hypothesized that some of these products may be extruded from the anterior of the tapeworm in order to elicit an inflammatory host reaction (Gallagher et al., 2017). As a consequence, it is expected that there will be genomic expansions (hypothesis 8) and/or transcriptomic upregulation (hypothesis 9) in genes associated with the production of secretory products and also immune response (hypotheses 10 and 11).

Comparative genomic and transcriptomic analyses will be used to test the above hypotheses. More specifically, a synteny analysis will be used to identify chromosomal rearrangements, a gene family evolution analysis will be used to look for possible gene family expansions and contractions, and a differential gene expression analysis will be used to examine differences in expression and regulation.

**Materials and Methods**

**Transcriptomic Analyses**

**Differential expression analyses**

In order to evaluate whether gene expression differences are contributing to the novel morphology of *L. aenigmaticum* differential expression analysis was run on the transcriptomes of *L. aenigmaticum, L. daileyi*, and *L. nickoli*. The trimmed raw reads for each species were mapped to their respective master transcriptome using the program HISAT2 (Kim et al., 2015). The resulting SAM file was sorted and converted to a BAM file using the program Samtools (Li et al., 2009). The alignments were then processed with the program eXpress to extract read count data.
Count matrices were generated from the eXpress results using the Trinity \texttt{abundance_estimates_to_matrix.pl} script. The resulting transcripts per million (TPM) count matrices for each species were normalized for transcriptome size following Musser and Wagner (2015), and also for transcript length with the R package DESeq2 (Love et al., 2014; Musser and Wagner, 2015). Due to the difficulties of running differential expression analyses with multiple species, only those transcripts found to occur in all three species were used in the subsequent analyses. Single-copy transcripts were identified following clustering of the master transcriptomes for each species with Orthofinder (Emms and Kelly, 2015). A python script was used to identify additional multi-copy orthogroups present in all three species. A second python script was used to pull the longest representative transcript for each species from the multi-copy transcripts. These representative multicopy transcripts and the single copy transcripts were then extracted from their respective normalized count matrices to form a smaller count matrix for each species. These smaller matrices were then merged in R and analyzed with the R package DESeq2 for differential expression across species using a p-value cut off point of 0.001 to account for the false discovery rate. The resulting differentially expressed transcripts were used to generate a Euclidean distance matrix. A clustered heatmap was generated for the 1000 most expressed genes with DESeq2 using row scaling based on z scores.

Gene ontology enrichment analysis was performed using the R package GOSeq. This analysis was performed on clusters that showed a series of similar expression patterns, namely those in which \textit{L. aenigmaticum} and \textit{L. daileyi} were up regulated in comparison to \textit{L. nickoli}, those in which \textit{L. daileyi} and \textit{L. nickoli} were up regulated in comparison to \textit{L. aenigmaticum}, those in which \textit{L. aenigmaticum} and \textit{L. nickoli} were up regulated in comparison to \textit{L. daileyi}, and those in which \textit{L. nickoli} was up regulated in comparison to \textit{L. aenigmaticum} and \textit{L. nickoli. Three different
backgrounds, one for each species (containing only the multicopy and single copy transcripts), were used for each analysis. These backgrounds were generated by functionally annotating the transcripts for each species with EnTAP and obtaining the gene lengths for each of the transcripts for each species. These functional annotations and transcript lengths were then used in GOSeq to fit a probability weighting function and identify GO terms that were significantly under- or overrepresented. The results from the GOSeq analysis were then run through the application REVIGO (Supek et al., 2011) to remove redundant terms; the results were visualized with the treemap package (version 2.4–2) in R (Tenneks and Ellis, 2017).

**Genomic Analyses**

**Syntenic Analysis**

Syntenic analyses were performed to assess the degree of conservation of homologous genes and gene order. These analyses were performed using the SynMap application in the Comparative Genomics (CoGe) online platform. Prior to the analyses, the genome assemblies were filtered to include only scaffolds 10Kb or longer. The filtered genomes and their annotations were then uploaded to the CoGe platform. SynMap was run with the BlastN algorithm; DAGChanier maximum gene distance was set to 50, minimum gene distance was set to 3, and the Quota Align algorithm was selected. The numbers of synonymous and non-synonymous sites were assessed with the CodeML SynMap analysis option. The SynMap analysis was run three times, once to compare *L. aenigmaticum* to *L. daileyi*, once to compare *L. aenigmaticum* to *L. amplifica*, and once to compare *L. daileyi* to *L. amplifica*.

**Gene Family Evolution Analysis**
In order to determine if there has been gene family expansions or contractions in the litobothriideans, gene family evolution was assessed with the program CAFE (De Bie et al., 2006). To run this analysis, the genome annotation gene models of the species of interest were clustered in order to generate a gene count file and species tree. The final annotation gene models for each of the three species of litobothriideans (i.e., *L. aenigmaticum*, *L. daileyi*, and *L. amplifica*) were clustered with the gene models for *Echinococcus multilocularis* with the program Orthofinder (Emms and Kelly, 2015). Because CAFE models stochastic birth and death processes along a phylogeny, it requires an ultrametric tree. As a consequence, a new species tree was generated as follows. The Orthofinder species tree, which represented a consensus tree of the gene trees of all genes found in all four species (4,116 gene trees overall), was modified as follows. The Python module ETE3 Toolkit 3 (Huerta-Cepas et al., 2016) was used to extend all of the branches of this tree until all were of equal length. The branch lengths of this transformed tree were then multiplied by 100 because CAFE cannot process trees with short branch lengths. To run CAFE, the gene count csv file was modified to include a column labeled “description” and the column labeled “total” was deleted. The CAFE python script cafetutorial_clade_and_size_filter.py was then used to divide the original file into two files, one containing gene families with fewer than 100 genes, and one with gene families with more than 100 genes. CAFE was then run to identify the rapidly evolving gene families within the phylogeny. In order to account for genome assembly errors and incompleteness, the caferror.py script was used to generate 24 error models and choose the best models from those. CAFE was then run on the gene families with fewer than 100 genes with a single lambda, the best error model estimated, a p-value of 0.01, and the number of random samples set to 1000. This analysis estimated a lambda for the data set, the probability of both gene gain and loss per gene per unit of time in a phylogeny. This value was used to identify the rapidly
evolving gene families. CAFE was run a second time on the count file with families with more than 100 genes with the lambda that was calculated in the first run, the best error model estimated, a p-value of 0.01, and the number of random samples set to 1000. The results were then summarized with the python script cafetutorial_report_analysis.py which output four files: anc.txt, fams.txt, node.txt, and pub.txt. The numbers of significantly (p<0.01) rapidly evolving genes were visualized on the phylogeny using the cafetutorial_draw_tree.py script. The number of gene families that were contracted and expanded were also visualized with the above script. The significantly rapidly evolving gene families were analyzed for enriched GO terms with the R package GOSeq. Redundant terms were removed with REVIGO. The results were visualized with the treemap package in R.

Results

Transcriptome Analyses

Differential expression analyses

Clustering the individual transcriptomes with Orthofinder yielded 7,855 multi-copy transcripts. These transcripts were pulled from their respective transcript per million (TPM) normalized transcriptomes to generate a count matrix for these multi-copy transcripts alone. Differential expression analyses was performed on this matrix. This yielded 4,439 transcripts that were differentially expressed. The hierarchical clustering of the transcripts using a Euclidean distance matrix yielded a dendrogram in which the specimens of each of the three species formed mutually monophyletic groups (Fig. 26). The topology of this dendrogram indicates that the expression patterns of *L. daileyi* and *L. nickoli* are more similar to one another than either is to the expression pattern of *L. aenigmaticum*. Furthermore, in the cases of *L. aenigmaticum* and *L. daileyi*, specimens of each of the two host individuals clustered together (i.e., the clusters of
specimens of both species consisted of subclusters of specimens from each of the two host
individuals). In contrast, the specimens of *L. nickoli* collected from the two host individuals were
intermingled.

Analysis of variance using a principal components analysis revealed that 91% of the
variance could be explained by principal components one, two, and three. Principal component
one accounted for 64% of the variance, principal component two accounted for 23% of the
variance, and principal component three accounted for 3% of the variance. Plotting principal
components one and two shows that the three species are all well separated from each other (Fig.
27). Plotting principal components two and three shows that *L. aenigmaticum* and *L. nickoli*
tend to fall near one another at the positive end meanwhile *L. daileyi* falls toward the negative end (Fig.
28).

The 1,000 most expressed transcripts were hierarchically clustered to produce a heatmap
of gene expression (Fig. 29). The dendrogram indicates that the expression of *L. aenigmaticum*
and *L. daileyi* are more similar to one another than either is to that of *L. nickoli*. Clusters with
expression patterns of interest (marked on Fig. 29) were further analyzed with the R package
GOSeq, the results of which are discussed below.

**GOSeq Results**

*L. aenigmaticum* downregulated cluster

In total, the heatmap indicated that 126 transcripts were in the clusters in which *L.
aenigmaticum* was downregulated compared to *L. daileyi* and *L. nickoli*. The functional enrichment
analysis, with *L. aenigmaticum* as the background, indicated that the following biological process
GO terms were enriched: cellular metabolism (primary metabolism, organic substance
metabolism, single-organism cellular process, single-organism metabolism), biosynthesis, multi-
organism cellular process, macromolecule localization (macromolecule localization, cellular localization, establishment of localization, localization of cell), response to activity (response to stress, response to chemical, response to biotic stimulus, response to endogenous stimuli, cellular response to stimulus), catabolism, antigen processing and presentation, myeloid cell homeostasis, flight, regulation of biological process, single-multicellular organism process, hormone metabolism, nitrogen compound metabolism, locomotory behavior, anatomical structure development, methylation, and cellular component organization (Fig. 30A). With *L. daileyi* as the background, the following additional terms were enriched: intraspecies interaction between organisms, developmental growth, and autophagy (Fig. 30B). With *L. nickoli* as the background, the following additional terms were enriched: response to estrogen, anatomical structure development, localization of cell, and cell cycle phase (Fig. 30C).

For the cluster in which *L. aenigmaticum* is down regulated, with *L. aenigmaticum* as the background, the cellular component process tree indicated the following GO terms were enriched: organelle lumen (non-membrane bound organelle, organelle part, extracellular organelle) and cell part (Fig. 31A). With *L. daileyi* as the background, the following additional terms were enriched: cell part and extracellular region part (Fig. 31B). With *L. nickoli* as the background, following additional terms were enriched: membrane-bounded organelle (Fig. 31C).

For the cluster in which *L. aenigmaticum* is down regulated, with *L. aenigmaticum* as the background, the molecular function process tree indicated organic cyclic compound binding (heterocyclic compound binding, small molecule binding, cofactor binding), structural of constituent ribosome, oxidoreductase activity, peroxidase activity, oxygen binding, antigen binding, ligase activity, amide binding, modified amino acid binding, sulfur compound binding, drug binding, and carbohydrate binding were enriched GO terms (Fig. 32A). With *L. daileyi* as the
background, the same GO terms were enriched (Fig. 32B). With L. nickoli as the background, the following additional terms were enriched: transcription factor activity and sequence specific DNA binding (Fig. 32C).

**L. daileyi downregulated cluster**

In total, the heatmap indicated that 77 transcripts were down regulated in L. daileyi compared to L. aenigmaticum and L. nickoli. The functional enrichment analysis, with L. aenigmaticum as the background, indicated that following biological process GO terms were enriched: single-organism developmental process (single-organism metabolism, single-multicellular organism process, multi-multicellular organism process), protein folding, directional locomotion, catabolism, cellular metabolism, rhythmic behavior, cellular component biogenesis, reproductive process, autophagy, biosynthesis, nitrogen compound metabolism, pigmentation, intracellular transport (establishment of localization), and response to endogenous stimuli (immune response, response to chemical, response to biotic stimulus, response to stress, response to abiotic stimulus) (Fig. 33A). With L. daileyi as the background, the following additional terms were enriched: primary metabolism, organic substance metabolism, and rhythmic behavior (Fig. 33B). With L. nickoli as the background, the following additional terms were enriched: developmental process involved in reproduction, multi-organism cellular process, regulation of biological process, and system process (Fig. 33C).

The cellular component tree, with L. aenigmaticum as the background, indicated that the following GO terms were enriched: extracellular organelle (organelle part, membrane-bound organelle, protein complex, organelar ribosome), membrane part, outer membrane, and cell-cell junction (Fig. 34A). The same terms that were enriched with the L. daileyi background (Fig. 34B).
With *L. nickoli* as the background, the following additional terms were enriched: organelle lumen and non-membrane-bounded organelle (Fig. 34C).

The molecular function tree, with *L. aenigmaticum* as the background, indicated that the following GO terms were enriched: carbohydrate transporter activity (substrate-specific transporter activity), isomerase activity, oxidoreductase activity, channel regulator activity, ion binding, cofactor binding, deaminase activity, hydrolase activity, carbohydrate binding, modified amino acid binding, and lipid binding were (Fig. 35A). With *L. daileyi* as the background, the following additional terms were enriched: electron transporter (transferring electrons from CoQH2-cytochrome c reductase complex and cytochrome c oxidase complex activity) (Fig. 35B). With *L. nickoli* as the background, the following additional terms were enriched: heterocyclic compound binding, organic cyclic compound binding, protein binding, and ion binding (Fig. 35C).

*L. nickoli* downregulated cluster

In total, the heatmap indicated that 173 transcripts were down regulated in *L. nickoli* compared to *L. aenigmaticum* and *L. daileyi*. The functional enrichment analysis, with *L. aenigmaticum* as the background, indicated that following biological process GO terms were enriched: protein folding, catabolic process, response to biotic stimulus, asexual reproduction, dormancy process, pigmentation, cellular component biogenesis, cellular localization, regulation of biological quality, autophagy, cellular metabolic process, single-multicellular organism process, biosynthetic process, cell growth, nitrogen compound metabolic process, primary metabolic process, single-organism metabolic process, response to estrogen, response to external stimulus, macromolecule localization, response to endogenous stimulus, response to stress, and establishment of localization (Fig. 36A). With *L. daileyi* as the background, the following additional terms were enriched: immune effector process (immune response, production of
molecular mediator of immune response, leukocyte activation), organic substance metabolic process, response to abiotic stimulus, and single-organism developmental process (Fig. 36B). With the L. nickoli background, the following additional terms were enriched: system process, protein folding, single-organism developmental process, multi-organism behavior, reproductive behavior, primary metabolic process, organic substance metabolic process, and multi-multicellular organism (Fig. 36C).

The cellular component process tree, with L. aenigmaticum as the background, indicated that the following GO terms were enriched: protein complex, extracellular region part (extracellular matrix, extracellular organelle), organelle part, membrane part, and non-membrane-bound organelle (Fig. 37A). With L. daileyi as the background, the following additional terms were enriched: receptor complex, anchoring junction, and protein-DNA complex (Fig. 37B). No additional GO terms were enriched with L. nickoli as the background (Fig. 37C).

The molecular function process tree, with L. aenigmaticum as the background, indicated that the following GO terms were enriched: structural constituent of muscle, channel regulator activity, isomerase activity, sulfur compound binding, peroxidase activity (peroxiredoxin activity), quaternary ammonium group binding, binding bridging, lyase activity, ligase activity, modified amino acid binding, drug binding, amide binding, lipid binding, carbohydrate derivative binding, small molecule binding, cytochrome-c oxidase activity, extracellular matrix structural constituent, and structural constituent of cytoskeleton (Fig. 38A). With L. daileyi as the background, the following additional terms were enriched: carbohydrate binding, hormone binding, binding, oxidoreductase activity, cofactor binding, hydrolase activity, and ion binding (Fig. 38B). With L. nickoli as the background, the following additional terms were enriched: transmembrane
transporter activity (substrate-specific transporter activity), oxidoreductase activity, and hydrolase activity (Fig. 38C).

*L. nickoli* upregulated cluster

In total, the heatmap indicated that 204 transcripts were upregulated in *L. nickoli* compared to *L. aenigmaticum* and *L. daileyi*. The functional enrichment analysis, with *L. aenigmaticum* as the background, indicated that following biological process GO terms were enriched: immune response (immune effector process), catabolic process, cellular component organization, reproduction of a single-celled organism, macromolecule localization, interspecies interaction between organisms (modification of morphology or physiology of other organism, multi-organism cellular process), single-organism developmental process (anatomical structure development), biosynthetic process, primary metabolic process, nitrogen compound metabolic process, single-multicellular organism process, cellular metabolic process, organic substance metabolic process, single-organism metabolic process, behavioral defense response, rhythmic behavior (circadian rhythm), and developmental growth (Fig. 39A). With *L. daileyi* as the background, the following additional terms were enriched: cell growth, regulation of biological process, single-organism metabolic process, cellular localization, and establishment of localization (Fig. 39B). With *L. nickoli* as the background, the following additional terms were enriched: cell proliferation, response to endogenous stimulus, antigen processing and presentation, regulation of biological process, and cellular localization (Fig. 39C).

The cellular component tree, with *L. aenigmaticum* as the background, indicated that outer membrane, extracellular organelle (extracellular region part), organellar ribosome, organelle lumen, organelle part, and non-membrane-bounded organelle were enriched GO terms (Fig. 40A).
With *L. daileyi* as the background, the following additional terms were enriched: cell part and excitatory synapse (Fig. 40B). With *L. nickoli* as the background, there were no enriched GO terms.

The molecular function tree, with *L. aenigmaticum* as the background, indicated that structural constituent of ribosome, channel regulator activity, deaminase activity, metal cluster binding, lyase activity, ligase activity, sulfur compound binding, carbohydrate binding, hydrolase activity, oxidoreductase activity, carbohydrate derivative binding, transferase activity, cofactor binding, heterocyclic compound binding, and organic cyclic compound binding were enriched GO terms (Fig. 41A). With *L. daileyi* as the background, the following additional terms were enriched: ligase activity and heterocyclic compound binding (Fig. 41B). With *L. nickoli* as the background, the following additional terms were enriched: transmembrane transporter activity (substrate-specific transporter activity), carbohydrate binding, binding bridging, lipid binding, and transferase activity (Fig. 41C).

**Genome Analyses**

**Synteny Analysis**

The synteny analysis was performed in order to identify conserved, homologous regions among genomes of the 3 litobothriidean species. In order to perform the analysis, the genomes were filtered to remove contigs smaller than 10 Kb in length. This reduced the number of contigs from 385,341 to 6,401 for *L. aenigmaticum*, from 380,502 to 6,819 for *L. daileyi*, and from 35,172 to 8,307 for *L. aenigmaticum*. The dot plot resulting from the SynMap analysis comparing *L. aenigmaticum* and *L. daileyi* is shown in Figure 42. The dot plot resulting from the SynMap analysis comparing *L. aenigmaticum* and *L. amplifica* is shown in Figure 43. The dot plot resulting from the SynMap analysis comparing *L. daileyi* and *L. amplifica* is shown in Figure 44. Calculation of the substitution rates from the CodeML SynMap (Fig. 45) results yielded a median synonymous
substitution rate (ks) of 0.0081 and an average substitution rate of 3.94 between *L. aenigmaticum* and *L. daileyi*. The median nonsynonymous rate (kn) between these two species was 0.0014 and the average substitution rate was 0.36. The median nonsynonymous to synonymous ratio was 0.068 and the average ratio was 1.16. The genome assembly for *L. amplifica* was too fragmented for the substitution rates to be calculated between *L. amplifica* and *L. daileyi* and *L. amplifica* and *L. aenigmaticum*.

**Gene Family Evolution**

The original Orthofinder species tree and the modified version of that tree are provided in Figures 15 and 46 respectively. In total, 10,427 gene families had fewer than 100 genes; only 3 gene families had over 100 genes. The best global error estimate was 0.182376098633 with a score of 56982.24318. CAFE yielded a lambda estimate of 0.00660033 when run on the gene families with fewer than 100 genes.

CAFE identified 82 gene families that were significantly rapidly evolving in *L. aenigmaticum* alone (Fig. 47). Of these 82 gene families, 76 were expanded and 6 were contracted. Among the 76 gene families that were expanded, the functional enrichment analysis indicated that the following biological process GO terms were enriched: immune effector process, response to biotic stimulus, cell growth (developmental growth), and protein folding were enriched GO terms (Fig. 48A). The enriched cellular component GO terms were cell-cell junction (anchoring junction), extracellular matrix (extracellular organelle, extracellular region part), RNA cap binding complex, and membrane part (Fig. 48B). The enriched molecular function GO terms were transcription factor activity, extracellular matrix structural constituent, isomerase activity, enzyme regulator activity, translation repressor activity (translation regulator activity, nucleic acid binding), modified amino acid binding, drug transporter activity, drug binding, sulfur compound
binding, amide binding, cofactor transporter activity, and substrate-specific transporter activity (transmembrane transporter activity) (Fig. 48C).

For the 6 gene families that were significantly contracted in *L. aenigmaticum*, the functional enrichment analysis indicated that the following biological process GO terms were enriched: immune effector process (antigen processing and presentation, immune response), system process, response to stress, cell communication, cell proliferation, catabolic process, sexual reproduction (reproductive process, multicellular organism reproduction, developmental process involved in reproduction), cellular component biogenesis (cellular component organization), maintenance of location, nitrogen compound metabolic process, single organism metabolic process, single organism signaling, regulation of biological quality, macromolecule localization, cellular localization, and establishment of localization (Fig. 49A). The enriched cellular component GO terms were extracellular organelle (extracellular region part), protein complex, and non-membrane-bounded organelle (Fig. 49B). The enriched molecular function GO term was hydrolase activity (Fig. 49C).

CAFE identified 30 gene families that were significantly rapidly evolving in *L. daileyi* alone (Fig. 47). All 30 of these gene families were expanded in this species. The functional enrichment analysis indicated that the following biological process GO terms were enriched: cell adhesion, feeding behavior (single-organism behavior, locomotory behavior, reproductive behavior), reproductive process (multicellular organism reproduction, developmental process involved in reproduction), filamentous growth (cell growth), hormone metabolic process, response to estrogen, multi-multicellular organism process, localization of cell, methylation, biosynthetic process, cell proliferation, primary metabolic process, nitrogen compound metabolic process, single-organism developmental process (anatomical structure process), regulation of biological
quality, response to chemical, single-organism cellular process, response to stress, and single-multicellular organism process (system process) (Fig. 50A). The enriched cellular component GO terms were cell-cell junction (anchoring junction), outer membrane, extracellular matrix, membrane-bounded organelle, cell part, and membrane part (Fig. 50B). The enriched molecular function GO terms were transcription factor activity transcription factor binding, transcription factor activity sequence-specific DNA binding, oxidoreductase activity, xenobiotic transporter activity, sulfur compound binding, binding bridging, lipid binding, protein binding, hydrolase activity, carbohydrate derivative binding, ion binding, drug transporter activity, and transmembrane transporter activity (Fig. 50C).

CAFE identified 8 gene families that were significantly rapidly evolving in *L. amplifica* alone (Fig. 47). Of these, 5 were contracted and 3 were expanded. With respect to the expanded gene families, the functional enrichment analysis indicated that the following biological process GO terms were enriched: abiotic stimulus, cellular component organization, antigen processing and presentation (immune effector process), multicellular organism reproduction (sexual reproduction), single-organism behavior (locomotory behavior, reproductive behavior, multi-organism behavior), maintenance of location, single-organism cellular process, detection of stimulus, response to external stimulus, and system process (Fig. 51A). The enriched cellular component GO terms were extracellular matrix (extracellular organelle, extracellular region part) and anchoring junction (Fig. 51B). The enriched molecular function GO terms were extracellular matrix structural constituent, isomerase activity, substrate-specific transporter activity (transmembrane transporter activity), cofactor binding, oxidoreductase activity, and small molecule binding (Fig. 51C).
For the gene families that were significantly contracted in *L. amplifica*, the functional enrichment analysis indicated that the following biological process GO terms were enriched: developmental process involved in reproduction (single-organism developmental process, reproductive process, anatomical structure development, sexual reproduction, multicellular organism reproduction), response to endogenous stimulus, methylation, single-multicellular organism process, localization of cell, catabolic process, organic substance metabolic process, response to estrogen, response to biotic stimulus, response to chemical, and response to stress (Fig. 52A). The enriched cellular component GO term was organelle lumen (Fig. 52B). The enriched molecular function GO terms were oxidoreductase activity, cofactor binding, heterocyclic compound binding, small molecule binding, ion binding, and organic cyclic compound binding (Fig. 52C).

CAFE identified 21 gene families that were significantly rapidly evolving in both *L. aenigmaticum* and *L. daileyi* (Fig. 47). All 21 of these gene families were expanded. For these gene families, the functional enrichment analysis with *L. aenigmaticum* as the background indicated that the following biological process GO terms were enriched: protein folding, immune response, cell adhesion, hormone metabolic process, single-multicellular organism process, anatomical structure development (single-organism developmental process), catabolic process, autophagy, biosynthetic process, cellular component organization, cell proliferation, single-organism cellular process, response to external stimulus, response to abiotic stimulus, natural killer cell mediated cytotoxicity (leukocyte mediated cytotoxicity), and developmental growth (cell growth) (Fig. 53A). With *L. daileyi* as the background, the following additional GO terms were enriched: reproductive process, developmental process involved in reproduction, methylation,
response to estrogen, response to external stimulus, and response to endogenous stimulus (Fig. 53B).

In terms of gene families found to be expanded in *L. aenigmaticum* and *L. daileyi*, the functional enrichment analysis, with *L. aenigmaticum* as the background, indicated that the following chemical component GO terms were enriched: cell-cell junction, extracellular matrix (extracellular region part), non-membrane-bounded organelle, protein complex, cell part, organelle part, and apoplast (Fig. 54A). With *L. daileyi* as the background, the following additional GO terms were enriched: membrane part and membrane-bounded organelle (Fig. 54B).

In terms of gene families expanded in *L. aenigmaticum* and *L. daileyi*, the functional enrichment analysis, with *L. aenigmaticum* as the background, indicated that the following molecular function GO terms were enriched: structural constituent of cytoskeleton, oxidoreductase activity, sulfur compound binding, amide binding, lipid binding, protein binding, transferase activity, ion binding, carbohydrate derivative binding, and small molecule binding (Fig. 55A). With the *L. daileyi* background, the same GO terms were enriched (Fig. 55B).

CAFE identified 2 gene families that were significantly rapidly evolving in all three litobothriideans (Fig. 47). Both gene families were contracted. The functional enrichment analysis, with both *L. aenigmaticum* and *L. daileyi* as the background (*L. amplifica* had no genes in these families and therefore was not able to be used as a background), yielded no enriched GO terms. In order to understand the function of these litobothriidean-contracted gene families, the functional enrichment analysis was run on the *E. multilocularis* genes that belong to these gene families. The functional enrichment analysis indicated that the following biological process GO terms were enriched: cell adhesion, cell growth, localization of cell, and cellular component organization (Fig. 56A).
The enriched cellular component GO terms were extracellular matrix (extracellular region part) (Fig. 56B). The only enriched molecular function GO term was protein binding (Fig. 56C).

CAFE identified 13 gene families that were significantly rapidly evolving in *E. multilocularis* relative to the the litobothriideans (Fig. 47). Of these gene families, 10 were contracted and 3 were expanded. In terms of the expanded families, the functional enrichment analysis indicated that the following biological process GO terms were enriched: immune response, cell adhesion, cell growth, and localization of cell were enriched GO terms (Fig. 57A). The enriched cellular component GO term was extracellular matrix (extracellular region part) (Fig. 57B). There were no enriched GO terms for molecular function. In terms of the contracted gene families of *E. multilocularis*, the functional enrichment analysis did not find any enriched GO terms. In order to determine the function of these gene families, a functional enrichment analysis was run on the genes for *L. aenigmaticum*, *L. amplifica*, and *L. daileyi* that belong to these families. The functional enrichment analysis on the *L. aenigmaticum* genes indicated that the following biological process GO terms were enriched: production of molecular mediator of immune response (immune response, immune effector process), system process (single-multicellular organism process), cell proliferation, response to endogenous stimulus, hormone metabolic process, multi-organism behavior, biosynthetic process, single-organism metabolic process, regulation of biological quality, response to external stimulus, response to chemical, maintenance of location, and antigen processing and presentation (Fig. 58A). The functional enrichment analysis on the *L. daileyi* genes indicated that the following additional biological process GO terms were enriched: single-organism behavior and response to biotic stimulus (Fig. 58B). The functional enrichment analysis on the *L. amplifica* genes indicated that the following additional biological process GO terms were enriched: single-organism behavior and detection of stimulus (Fig. 58C).
The functional enrichment analysis on the *L. aenigmaticum* genes indicated that the following cellular component GO terms were enriched: extracellular region part (extracellular organelle), membrane part, and cell part (Fig. 59A). The functional enrichment analysis on the *L. daileyi* genes indicated the following additional chemical component GO terms were enriched: outer membrane and extracellular region part (Fig. 59B). The functional enrichment analysis on the *L. amplifica* genes indicated that the following additional terms were enriched: organelle part, anchoring junction, and membrane-bounded organelle (Fig. 59C).

The functional enrichment analysis on the *L. aenigmaticum* genes indicated that the following molecular function GO terms were enriched: lipid binding, transferase activity, and ion binding were enriched GO terms (Fig. 60A). The functional enrichment analysis on the *L. daileyi* genes found that the same GO terms were enriched (Fig. 60B). The functional enrichment analysis on the *L. amplifica* genes found that the following additional GO terms were enriched: isomerase activity and ligase activity (Fig. 60C).

**Discussion**

**Transcriptomic analyses**

Both the hierarchical correlation matrix (Fig. 26) and the PCA plots (Figs. 27, 28) revealed a strong species signal in the expression data. All 14 transcriptomes clustered according to species in both the correlation matrix and PCA plots. *Litobothrium aenigmaticum* and *L. daileyi* subclustered based on host individual in the correlation matrix. This was not, however, the case in the PCA plots for any of the three species. In combination, these results indicate that transcriptomes from specimens of the same species have more similar expression patterns than those from specimens from the same host individual. It is interesting to note though that the dendrogram for the correlation matrix differed from that of the clustered heatmap. The correlation
matrix dendrogram indicated that the expression of *L. daileyi* and *L. nickoli* were more similar to each other than either were to that of *L. aenigmaticum*. However, the dendrogram for the clustered heatmap indicated that the expression of *L. aenigmaticum* and *L. daileyi* were more similar to each other than to *L. nickoli*. These differences are due to the fact that the correlation matrix dendrogram was constructed based on the differences between all 7,855 transcripts while the heatmap dendrogram was constructed with only 1,000 most expressed transcripts.

The heatmap of the 1,000 most expressed genes (Fig. 29) was used to identify clusters with interesting expression patterns. Functional enrichment analysis of these clusters yielded a large number of enriched GO terms. For the purposes of this discussion, only those that were significantly enriched, i.e., p<0.05 (Fig. 61), will be addressed below.

Transcripts associated with biosynthetic processes were identified in clusters in which *L. aenigmaticum* was down regulated (Fig. 61). A biosynthetic process is typically an enzyme catalyzed process in which simple molecules are transformed into complex molecules. Some examples of biosynthesis in tapeworms are the production of lipids (Johnson and Cain, 1985), fatty acids (Jacobsen and Fairbairn, 1967; Tsai et al., 2013), serine, proline, and molybdopterin (Tsai et al., 2013). Since biosynthetic processes are used to form many different products the function of these transcripts cannot be determined without further information.

Transcripts related to interspecies interaction between organisms were identified in clusters in which *L. aenigmaticum* was downregulated (Fig. 61). There is currently no evidence that the tapeworms within pelagic thresher sharks interact with each other; therefore, it is more likely that the interspecies interaction is occurring between the tapeworm and its host. As mentioned in chapter two, all tapeworms must interact with and counter the immune system of their host. Some strategies that have been described are the secretion of products that redirect the host immune
system, release of proteases that can digest host proteins and protease inhibitors, release of signaling peptides that act as messengers to the host, inhibition of essential host enzymes, and interference with host signal transduction pathways (Zheng et al., 2013). Therefore, it seems likely that these gene families are involved with the tapeworms’ interaction with the host immune system. It is surprising though that these transcripts are down regulated \textit{L. aenigmaticum} since it is the only species that induces an immune response in the host. This seems to indicate that these transcripts must not be involved in inducing the inflammatory host response caused by this species.

Transcripts that were annotated with GO terms related to immune response and antigen processing and presentation were identified in clusters in which \textit{L. aenigmaticum} was down regulated (Fig. 61). Again, this was a surprising result since neither of these two species interacts with the host immune response as much as \textit{L. aenigmaticum} does. Therefore, it appears that these transcripts also do not contribute to the inflammatory response seen in association with \textit{L. aenigmaticum}. However, these transcripts are likely an important portion of the process in which \textit{L. daileyi} and \textit{L. nickoli} counter the host immune system.

Transcripts associated with response to biotic stimulus were identified in clusters in which \textit{L. aenigmaticum} and \textit{L. daileyi} were up regulated (Fig. 61). Additionally, transcripts associated with response to activity were identified in clusters in which \textit{L. aenigmaticum} was down regulated. As discussed in chapter two, tapeworms must be able to detect and respond to many biotic stimuli to complete their life cycle. Therefore, these transcripts could be involved in anything from the transition of life cycle stages to reproductive behavior. Further information is therefore necessary in order to understand how these transcripts contribute to the structure and function of these tapeworms.
Transcripts associated with reproduction were identified in clusters in which *L. aenigmaticum* and *L. daileyi* were down regulated and transcripts associated with developmental processes in reproduction were identified in clusters in which both *L. aenigmaticum* and *L. nickoli* were up regulated (Fig. 61). As mentioned previously, *L. daileyi* and *L. nickoli* use the same reproductive strategy (i.e., they are euapolytic; Olson and Caira, 2001) while *L. aenigmaticum* uses a completely different strategy (i.e., it is hyperapolytic; Caria et al., 2014a). Due to this, it was expected that *L. aenigmaticum* would differ in expression in reproductive genes from the other two. Therefore, it was surprising that instead *L. aenigmaticum* seems to have more similar expression of reproductive genes to *L. daileyi* than *L. daileyi* did to *L. nickoli*. On the other hand, up regulation of transcripts involved in developmental processes in reproduction in *L. aenigmaticum* is unsurprising since these worms bear only immature proglottids on their strobila. It is surprising, however, to find these transcripts upregulated in *L. daileyi*. This worm does bear immature proglottids on its strobila (i.e., proglottids undergoing development), but *L. nickoli* bears a similar number of immature proglottids on its strobila as well (Olson and Caira, 2001). Therefore, it is unclear why these transcripts are up regulated in *L. daileyi* and down regulated in *L. nickoli*.

A number of metabolic processes were associated with transcripts that were upregulated in all three litobothriidean species (Fig. 61). Specifically, primary metabolic process, organic substance metabolic process, and nitrogen compound metabolic process were up regulated in *L. nickoli* and *L. daileyi* while cellular metabolic process was up regulated in *L. aenigmaticum*, *L. daileyi*, and *L. nickoli*. Primary metabolism typically contributes to the growth and normal cellular functioning of an organism. Organic substance metabolic processes are those involved in the formation of any molecules that contain carbon. Meanwhile, nitrogen compound metabolic processes are those involved in the production of molecules that contain nitrogen. Finally, cellular
metabolism is typically involved in the production of enzymes, nucleic acids, and amino acids and the construction of complex molecules. Overall, these are broad terms that can be applied to many different functions and processes. Therefore, the specific function of these transcripts cannot be determined without further information.

Transcripts related to oxidoreductase activity, isomerase activity, and peroxidase activity were found to be up regulated in all three litobothriidean species (Fig. 61). All three of these terms refer to types of catalytic activities and all have been found to be expressed in the secretome of *E. multilocularis* (Wang et al., 2015). Certain isomerase activities have been tested as drug targets in schistosomes and cyclophyllidean tapeworms (Roberts et al., 1995; Khattab et al., 1999; McLauchlan et al., 2000; Colebrook et al., 2002). Meanwhile, peroxidase activities, which usually catalyze the reduction of hydrogen peroxide, have been reported from several helminths, including *Ascaris* species, *Fasciola hepatica*, and several cyclophyllidean tapeworms; this activity has been suggested to be involved in respiration and/or aerobic metabolism (Threadgold et al., 1968; Lumsden et al., 1969; Bogitsh, 1975). It seems that these activities are important in not just tapeworms but helminths overall; however, since these are such broad terms it is not possible to determine the specific functions for these transcripts in the litobothriidean tapeworms.

Transcripts associated with deaminase activity were identified in clusters in which *L. nickoli* was up regulated (Fig. 61). This type of activity has been noted in other tapeworms and is believed to contribute to energy metabolism (Bennet et al., 1990). Therefore, it is likely that these transcripts are involved in the breakdown of nutrients from the host intestines.

Transcripts associated with protein folding, cofactor binding, and ion binding were also identified in clusters in which *L. aenigmaticum* and *L. daileyi* were up regulated (Fig. 61). However, since these terms could be applied to transcripts involved in many different
pathways/processes it is not possible to speculate as to the specific functions of these transcripts in the litobothriideans.

**Genomic Analyses**

**Synteny Analyses**

Comparison of *L. aenigmaticum* to *L. daileyi* with the program SynMap revealed some regions of synteny between the two; however, the genomes for these two species are still too fragmented to perform a comprehensive syntenic analysis (Fig. 20). Comparison of *L. aenigmaticum* to *L. amplifica* (Fig. 21) and *L. daileyi* to *L. amplifica* (Fig. 22) revealed that the genome of *L. amplifica* is too fragmented to properly identify syntenic regions. In order to effectively perform these analyses the genomes will need to be improved. This could be done with the incorporation of long read sequence data in future work.

**Gene Family Evolution Analysis**

The results of the CAFE analyses, which included *L. aenigmaticum*, *L. daileyi*, and *L. amplifica* as well as *E. multilocularis*, revealed a total of 156 rapidly evolving families: 82 in *L. aenigmaticum* alone, 30 in *L. daileyi* alone, 8 in *L. amplifica* alone, 21 in both *L. aenigmaticum* and *L. daileyi*, 2 in all three litobothriideans, and 13 in *E. multilocularis* alone. Functional enrichment analyses of these rapidly evolving genes identified a large number of enriched GO terms. For the purposes of this discussion, only those that were significantly enriched, i.e. p<0.05, will be addressed here (Fig. 62).

Gene families related to modified amino acid binding were expanded in *L. aenigmaticum* (Fig. 62). Modified amino acids are amino acids which have been altered after translation, by processes such as phosphorylation, glycosation, and lipidation (Khoury et al., 2011). Since glycoproteins and mucoproteins are both types of modified amino acids, it is possible that these
gene families may be involved in the production of the secretory vesicles in the anterior region of *L. aenigmaticum*.

Gene families associated with regulation, specifically translation repressor activity and translation regulator activity (nucleic acid binding), were expanded in *L. aenigmaticum* (Fig. 62). Meanwhile gene families associated with methylation were contracted in *L. amplifica*. This indicates that *L. aenigmaticum* may be regulating its gene expression differently from the other litobothriideans. More specifically, the expansion of translation repressor genes implies that *L. aenigmaticum* is down regulating certain regions of its transcriptome. Further, the fact that methylation genes appear to be contracted in *L. amplifica* in comparison to *L. aenigmaticum* and *L. daileyi* indicates that the latter species are also utilizing methylation to regulate gene expression.

Gene families associated with maintenance of location were contracted in *L. aenigmaticum* (Fig. 62); these families are likely involved in tethering substances. It was hypothesized in chapter 1 that the products of the secretory vesicles found in the anterior region of *L. aenigmaticum* are moved and extruded, perhaps to assist with attachment. Thus, it would not be surprising that there have been expansions in gene families associated with the movement of products in *L. aenigmaticum*.

Gene families associated with drug binding and immune effector process was expanded in *L. aenigmaticum* and families associated with immune response and natural killer cell mediated cytotoxicity were expanded in both *L. aenigmaticum* and *L. daileyi* (Fig. 62). Meanwhile, gene families related to immune effector process, antigen processing and presentation, and immune response were contracted in *L. aenigmaticum* and production of molecular mediator of immune response, immune effector process, and immune response were contracted in *L. aenigmaticum* and *E. multilocularis*. Since *L. aenigmaticum* induces a severe inflammatory host reaction (Caira et al.,
2014a), it is unsurprising that gene families associated with immune response are expanded in this species. *L. daileyi*, however, does not induce a host immune response so it was surprising to find that some of the same immune response genes that were expanded in *L. aenigmaticum* were expanded in this species as well. It was also surprising that some aspects of immune response, such as antigen processing and molecular mediator processes, are contracted in *L. aenigmaticum*. This implies that only a subset of the gene families associated with immune response are involved with the inflammatory host reaction. Currently it is not possible to determine which of these genes may be involved with this function.

Gene families associated with anatomical structure, specifically structural constituent of cytoskeleton and anatomical structure development, were expanded in *L. aenigmaticum* and *L. daileyi* (Fig. 62). As described in the introduction and chapter 1, the morphology and internal anatomy of *L. aenigmaticum* greatly differs from that of its congeners, so it is not surprising that these gene families are expanded within this species. *L. daileyi*, on the other hand, does not differ greatly in morphology or internal anatomy from *L. amplifica*. Therefore, it is unclear as to why these gene families are expanded within this species.

Gene families associated with extracellular matrix structural constituent were expanded in *L. aenigmaticum* and *L. amplifica* (Fig. 62). According to the Gene Ontology Browser, this GO term may be applied to extracellular matrix glycoproteins; therefore, it is possible that this term is being applied to gene families that produce glycoproteins found in the anterior region of *L. aenigmaticum*. However, it is unclear why these gene families are expanded in *L. amplifica* since the presence of glycoproteins has not been reported in this species.

Gene families associated with growth, specifically developmental growth, were expanded in *L. aenigmaticum* and *L. daileyi* (Fig. 62). Those related to cell growth were expanded in *L.
aenigmaticum, L. daileyi, and E. multilocularis but also contracted in L. aenigmaticum. Due to the large size of L. aenigmaticum it is not surprising that gene families associated with growth are expanded in this species; however, it is not clear why these families are expanded in the other species since they all are approximately 2–3mm in length. It is puzzling though that the gene families associated with cell growth is contracted in L. aenigmaticum; it was expected that these families would be expanded in this species due to its size.

Gene families associated with cofactor transmembrane transporter activity and lipid binding were expanded in L. aenigmaticum and L. daileyi and contracted in E. multilocularis (Fig. 62). Since lipid binding proteins can facilitate the transport of lipids (Glatz, 2015), this indicates that these gene families, and those associated cofactor transmembrane transporter activity, may contribute to the movement of products and substances. As previously mentioned, the results of chapter 1 indicate that the products of the secretory vesicles of L. aenigmaticum are moved between cell types. So, it would not be surprising that there have been expansions in gene families associated with the movement of products in L. aenigmaticum. However, the types of products that are extruded from L. aenigmaticum are thought to be glycoproteins or mucoproteins, not cofactors or lipids. Further, there has been no documentation of lipid movement in L. daileyi. Therefore, it is unclear how these gene families are functioning in L. aenigmaticum and L. daileyi.

Gene families associated with response to stimulus, response to endogenous stimulus, response to estrogen, and response to external stimulus were contracted in L. amplifica and E. multilocularis (Fig. 62), meaning that L aenigmaticum and L. daileyi have more genes that are related to response to stimulus. As discussed in chapter two, tapeworms need to be able to detect and respond to stimuli for many reasons, such as transitioning between life stages, positioning themselves within a host, and undergoing reproduction. Response to estrogen implies that this term
may be related to reproductive behavior. The other terms, on the other hand, are very broad labels and therefore could apply to a number of functions. Additional work is necessary to understand how these gene families influence the structure or function of these tapeworms.

Gene families associated with reproduction and the development of reproductive structures were expanded in *L. daileyi*, specifically reproductive process and developmental process in reproduction (Fig. 62). Most of the litothriidean tapeworms have a euapolytic reproductive strategy in which they drop mature proglottids prior to fertilization (Dailey, 1969; Dailey, 1971; Kurochkin and Slankis, 1973; Caira and Runkle, 1993; Olson and Caira, 2001). The only species in this genus that deviates from this is *L. aenigmaticum* which is hyperapolytic, i.e., it drops its proglottids before they are mature (Caira et al., 2014a). Therefore, it is surprising to find that there have been expansions in reproductive gene families in *L. daileyi* alone. Since reproduction in this species does appear to differ from *L. amplifica* it would have been expected to see expansions of families in *L. amplifica* as well.

**Why is *L. aenigmaticum* so weird?**

The goal of this chapter was to look for genomic expansions and transcriptomic upregulations between *L. aenigmaticum* and the other, more typical litobothriideans in order to examine the mechanisms that might have led to the evolution of the bizarre morphology and anatomy seen in *L. aenigmaticum*. In regards to the overall structure of *L. aenigmaticum*, there is some evidence that there have been genomic changes that accompany the change in structure and size of *L. aenigmaticum* but there is currently no evidence that there have been changes in expression of these genes. The genomic analyses found that gene families associated with the structural constituent of cytoskeleton, anatomical structure development, developmental growth, cell growth, and cell proliferation were expanded in *L. aenigmaticum*. However, none of the
differentially expressed transcripts were found to be related to structural development and growth. The lack of transcriptomic evidence does not rule out the possibility of a difference in the expression of these genes. Because such genes contribute to the structure and growth of the tapeworm, they are unlikely to be highly expressed in the adult stage. The point in time when they are most likely to be expressed is when the tapeworm is transitioning from a juvenile to adult. As a consequence, in order to effectively evaluate the expression of such genes in the future, the generation of transcriptomic data from an earlier life cycle stage is necessary.

In the first chapter, a complex of 11 novel cell types was described from the anterior region of *L. aenigmaticum*. Essentially all of these cell types were found to contain secretory vesicles thought to contain glycoproteins or mucoproteins. This finding led to the hypothesis that *L. aenigmaticum* should differ from its congeners in regards to genes that are associated with the production of secretory products. The gene family evolution analyses indicate that a number of gene families associated with extracellular matrix and modified amino acids were expanded in *L. aenigmaticum*. Since both terms can be applied to genes that produce glycoproteins, these results support the idea that there have been genomic changes associated with the production of secretory products. Unfortunately, none of the differentially expressed transcripts were annotated with GO terms related to secretory products. Thus, currently, there is no evidence of changes in gene expression in genes associated with secretion.

*Litobothrium aenigmaticum* is also unusual in that it elicits a severe inflammatory host response at its site of attachment to the surface of the mucosa. No such reaction is seen with the other litobothriideans species. Some gene families associated with immune response, drug binding, and natural killer cell mediated cytotoxicity were found to be expanded in *L. aenigmaticum*. Other gene families associated with immune response and antigen processing and presentation were
found to be contracted in *L. aenigmaticum*. The latter contraction was reflected in the transcriptomes, in that transcripts associated with interspecies interactions, immune response, and antigen processing and presentation were found to be down regulated in *L. aenigmaticum* relative to its congeners. This would suggest that the inflammatory response associated with *L. aenigmaticum* is being induced by only a subset of the immune-related genes.

*Litobothrium aenigmaticum* also differs conspicuously from its congeners in reproductive strategy. Whereas all other litobothriideans are euaploytic, in that they drop mature proglottids from their strobila (Olson and Caira, 2001), *L. aenigmaticum* is hyperapolytic, meaning it drops its proglottids while they are still immature (Caira et al., 2014a). However, the only differences observed with regard to reproduction was within the transcriptomes; transcripts associated with developmental processes in reproduction were up regulated in *L. aenigmaticum*. This is not unexpected since the strobila of *L. aenigmaticum* bears only immature proglottids. It is somewhat surprising that no other differences were observed.

Finally, there was some genomic evidence that indicates that *L. aenigmaticum* may regulate its gene expression differently than the other litobothriideans. Specifically, it was found that gene families associated with translation repressor activity and translation regulator activity were expanded in *L. aenigmaticum*. It was also found that *L. aenigmaticum* and *L. daileyi* have more genes associated with methylation than *L. amplifica*. It is currently unknown how *L. aenigmaticum* uses these genes to regulate gene expression.

**Conclusions**

The goal of this chapter was to assess 11 hypotheses that might account for the bizarre nature of *L. aenigmaticum* using comparative genomic and transcriptomic analyses. The genome size estimates from chapter two rejects the hypothesis that there has been a genome duplication in
*L. aenigmaticum*. Genome size estimates for *L. aenigmaticum* ranged from 320–355 Mb, which are well within the range of estimates for *L. daileyi* at 330–412 Mb. Based on this similarity in sizes, it was concluded that it was unlikely that a genome duplication was involved in the evolution of *L. aenigmaticum*.

The hypothesis that there were chromosomal rearrangements in *L. aenigmaticum* was assessed using a synteny analysis. The results of this analysis indicated that the litobothriidean genome assemblies are still too fragmented to perform a full analysis. Therefore, hypothesis two could not be evaluated with available data.

The hypothesis that there have been changes in the way that *L. aenigmaticum* regulates its gene expression was assessed with a gene family evolution analysis. The results from this analysis indicated that there have been expansions in gene families associated with translation regulation in *L. aenigmaticum*, therefore supporting this hypothesis.

The hypothesis that there have been expansions in gene families associated with anatomical structure in *L. aenigmaticum* was also assessed with a gene family evolution analysis. It was found that there have been expansions in gene families associated with anatomical structure and growth in *L. aenigmaticum*, therefore supporting this hypothesis.

The hypothesis that there has been up regulation of genes associated with anatomical structure in *L. aenigmaticum* was assessed with a differential expression analysis. None of the significantly enriched GO terms from this analysis were associated with anatomical structure. It is possible that the genes related to anatomical structure are more highly expressed when the tapeworm first enters its host, in which case, the transcriptomes for the juvenile stages should be examined before these hypotheses are rejected. Therefore, this hypothesis requires additional investigation.
The hypothesis that there have been expansions in gene families associated with reproduction in *L. aenigmaticum* was assessed with a gene family evolution analysis. The results indicated that there have not been expansions in gene families related to reproduction; hence, this hypothesis is rejected.

The hypothesis that there has been up regulation of genes associated with reproduction in *L. aenigmaticum* was assessed with a differential expression analysis. The results revealed that there have been up regulations in several gene families associated with developmental processes in reproduction within *L. aenigmaticum*, therefore supporting this hypothesis.

The hypothesis that there have been expansions in gene families associated with the production of secretory products in *L. aenigmaticum* was assessed with a gene family evolution analysis. The results from this analysis indicated that there have been expansions in gene families associated with extracellular matrix and modified amino acid binding. Both are terms that can be applied to glycoproteins; thus, this hypothesis is supported.

The hypothesis that there has been upregulation of genes associated with the production of secretory products in *L. aenigmaticum* was assessed with a differential expression analysis. None of the significantly enriched GO terms from this analysis were associated with the production of secretory products. However, it is possible that the genes the production of secretory products are more highly expressed when the tapeworm first enters the host. Therefore, the transcriptomes for the juvenile stages should be examined before this hypothesis is rejected.

The hypothesis that there have been expansions in gene families associated with immune response in *L. aenigmaticum* was assessed with a gene family evolution analysis. This analysis revealed that there have been expansions in gene families associated with immune response, supporting this hypothesis.
The hypothesis that there has been upregulation of genes associated with immune response in *L. aenigmaticum* was assessed with a differential expression analysis. The results indicated that there has been downregulation of gene families associated with immune response in *L. aenigmaticum*. It is possible though that these genes are upregulated more highly when they first enter the host; therefore, in order to fully address this hypothesis the transcriptomes of the juvenile stages need to be examined.

All of the hypotheses regarding genomic changes predicted that gene families would be expanded in *L. aenigmaticum* compared to its congeners; however, the comparative genomic analyses identified a number of contracted gene families in *L. aenigmaticum*. The CAFE analysis indicated that there were 6 gene families that were contracted in *L. aenigmaticum*. These gene families were annotated with GO terms related to immune function, cell proliferation, sexual reproduction, and maintenance of location. In addition, CAFE identified two gene families that were contracted in all three litobothriideans. These families were associated with cell growth and cellular component organization. This indicates that the contraction, not just the expansion, of gene families may also be involved in the evolution of the unusual morphology of *L. aenigmaticum*.

It should also be noted that all the hypotheses regarding transcriptomic changes predicted that there would be up regulation of transcripts in *L. aenigmaticum* compared to its congeners. However, there were no clusters in which *L. aenigmaticum* was upregulated compared to the other two litobothriideans. Instead, there was a cluster of 126 transcripts in which *L. aenigmaticum* was down regulated compared to *L. daleyi* and *L. nickoli*. These transcripts were associated with biosynthesis, cellular localization, response to stimulus, antigen processing and presentation, and anatomical structure development. Additionally, this analysis revealed another cluster of 204 transcripts that were down regulated in *L. aenigmaticum* and *L. daleyi* compared to *L. nickoli*. 
These transcripts in this cluster were associated with immune response, interspecies interactions, developmental processes, biosynthetic processes, and behavioral defense responses. Overall, this indicates that down regulation of transcripts in *L. aenigmaticum* may have played a crucial role in the evolution of its unusual body form.
Overall Discussion

The genus *Litobothrium* is unusual in that the most recently described species within this group, *Litobothrium aenigmaticum*, does not even remotely resemble its congeners. Typical litobothriideans exhibit a scolex that consists of an apical sucker and 3–6 pseudosegments and a strobila with 50–70 proglottids (Olson and Caira, 2001; Caira et al., 2014a). *L. aenigmaticum* lacks all of these characteristics and instead exhibits a dome-shaped scolex, an elongate cephalic peduncle, and a reduced strobila that consists of only tiny immature proglottids. These overwhelming morphological differences were used to justify the description of *L. aenigmaticum* as a new species (Caira et al., 2014a). However, a more recent phylogenetic analysis revealed that *L. aenigmaticum* is identical to *Litobothrium daileyi* for the 28S rDNA (D1–D3) gene region (Caira et al., 2017). Despite this result, it was still thought that these two are separate species and it is believed that this molecular similarity indicates that they arose in a recent, rapid divergence event. Therefore, the primary aim of this dissertation was to examine possible genetic mechanisms that contributed to the evolution of the bizarre morphology and anatomy of *Litobothrium aenigmaticum*. In order to do this the ultrastructure of *L. aenigmaticum* was characterized with transmission electron microscopy in chapter one, genomes and transcriptomes were generated for a subset of the litobothriidean species in chapter two, and comparative genomic and transcriptomic analyses were performed using these genomic resources in chapter three. A secondary aim for this dissertation was to identify molecular differences between *L. aenigmaticum* and *L. daileyi* that would support our assertion that these two are separate species.

The results of chapter one revealed that the anatomy of *L. aenigmaticum* is much more complicated than originally thought. Transmission electron microscopy revealed a complex of 11 novel cell types in its cephalic peduncle, each of which contained secretory vesicles. Staining with
Periodic acid-Schiff indicated that the products in the secretory vesicles are mucoproteins and/or glycoproteins. It seems likely that at least some of these products are extruded from the tapeworm in order to elicit an inflammatory host response.

The work from chapter two resulted in the generation of annotated genome assemblies for *L. aenigmaticum*, *L. daileyi*, and *L. amplifica* and transcriptomes for *L. aenigmaticum*, *L. daileyi*, and *L. nickoli*. Approximately 81–89% of the gene space of the litobothriidean genomes were assembled. In comparison, these genomes are more fragmented than those of *Echinococcus multilocularis*, *Echinococcus granulosus*, *Hymenolepis microstoma*, *Taenia asiatica*, *Taenia multiceps*, and *Taenia saginata*; as fragmented as *Echinococcus canadensis*, *Hymenolepis diminuta*, and *Mesocestoides corti*; and more complete than those of *Diphyllobothrium latum*, *Schistocephalus solidus*, *Spirometra erinaceieuropaei*, *Hydatigera taeniaeformis*, and *Hymenolepis nana* (Tsai et al., 2013; Zheng et al., 2013; Bennett et al., 2014; Maldonando et al., 2017; International Helminth Genomes Consortium, 2019). At the outset of this study it was predicted that the litobothriideans would have a genome size between those of the cyclophyllideans and the diphyllobothriideans. It was also hypothesized that the litobothriideans would have between 10,000–20,000 genes, and would, like other cestodes, have reduced repeat content relative to other Platyhelminthes. All three of these hypotheses were supported by the results. Litobothriidean genome size estimates range from 320–470 Mb, making them larger than those of the cyclophyllideans and the diphyllobothriideans. They had a repeat content similar to the cyclophyllideans. The number of genes in the litobothriidean genomes ranged from 8,358–21,070; these numbers were more similar to those seen in the diphyllobothriideans than those seen in the cyclophyllideans. The annotated genomes were also used in order to look for molecular differences between *L. aenigmaticum* and *L. daileyi*. Comparison of the genomes of
these two species revealed that a total of 22,100,993 bp of the *L. aenigmaticum* genome could not be aligned to that of *L. daileyi*. Furthermore, clustering the genome annotation models of the litobothriideans with that of *Echinococcus multilocularis* resulted in 684 single copy orthogroups. Examination of the gene trees for these orthogroups revealed that 468 of those trees indicated that there was phylogenetic distance between *L. aenigmaticum* and the other litobothriideans. Overall, this study showed a number of molecular differences between *L. aenigmaticum* and *L. daileyi*, adding to the abundance of morphological evidence that indicates that these two are likely separate species.

The goal of third and final chapter was to assess mechanisms that might account for the morphology and anatomy of *L. aenigmaticum* using comparative genomic and transcriptomic analyses. These analyses indicated that several different mechanisms may account for the evolution of *L. aenigmaticum* relative to its congeners and other cestodes. These include gene family expansions, differential regulation of translation, up regulation of specific gene families, and coding region changes. It is important to note though, that there are aspects of the study that could be improved. For instance, the genome family evolution analysis was run using a non-ultrametric tree that was modified by extending the branch lengths using the Python module ETE3 Toolkits (Huerta-Cepas et al., 2016). In the future, a fossil-calibrated ultrametric tree will be generated for the litobothriidean tapeworms and the gene family evolution analysis will be re-run using this new phylogeny. Another aspect of this study that could be improved involves the litobothriidean genomes themselves. The current genomes are fragmented. They were assembled using short read libraries (350 and 550 bp) and, in the case of *L. aenigmaticum*, mate pair libraries (2.5 and 7.5 kb). Although the assemblies and their annotations were sufficiently complete to perform some comparative analyses, they were too fragmented to perform an adequate synteny
analysis. In the future, the generation of long-read sequence data using PacBio or Promethion could be used to scaffold the current litobothriidean assemblies to produce more contiguous genomes.

It should also be noted that this study only examined a subset of the possible genetic changes that could have led to the evolution of the unusual body form of *L. aenigmaticum*. Other mechanisms could be involved in this system. These possibilities include co-option and/or a number of developmental changes. In order to investigate co-option it will be necessary to isolate tissue-specific transcriptomes. This was not feasible in the present study due both to the small size of *L. daileyi* and *L. nickoli* and the difficulty of isolating individual tissues in acoelomates such as tapeworms – be they small or large. In the near future, the aim will be to use whole mount in situ hybridization (WISH) to examine whether co-option is involved. WISH has previously been used to identify tissue-specific expression in the digenean *Schistosoma mansoni* (Cogswell et al., 2011) and the tapeworms *Echinococcus multilocularis* and *Hymenolepis microstoma* (Koziol et al., 2014; Koziol et al., 2016). These studies have shown that WISH is an effective method for identifying the location of transcript expression in Platyhelminthes. As a consequence, this method appears to be a viable option for identifying the genes that may have been co-opted to form the novel structures seen in *L. aenigmaticum*. In order to determine whether developmental changes were involved in the evolution of *L. aenigmaticum*, it would be helpful to examine the transcriptomes of juvenile stages. However, this may prove to be especially difficult now that the pelagic thresher shark has been listed as a protected species.

Moving forward, it will also be interesting to determine if some of the mechanisms that appear to underlie the unusual morphology of *L. aenigmaticum* seen here, provide some insight into the mechanisms associated with instances of differences in morphology seen between other
closely related parasite species. The litobothriideans are just one of many examples of this phenomenon within the cestodes. For instance, Fyler (2009) identified three pairs of congeners, *Acanthobothrium zainali* and *Acanthobothrium sp. 23*, *Acanthobothrium masnihae* and *Acanthobothrium saliki*, and *Acanthobothrium zainali* and *Acanthobothrium saliki*, that differed greatly morphologically but were identical in the 28S rDNA (D1-D3) sequence data (Fyler, 2009). Another example is seen between the cestode species *Platybothrium auriculatum* Yamaguti, 1982 and *Prosobothrium armigerum* Cohn, 1902. These species are currently assigned to different genera based on their highly divergent scolex morphologies. However, phylogenetic analyses show *P. armigerum* to be nested amongst *Platybothrium* species as sister taxon to *Platybothrium auriculatum* (Caira et al., 2014b) – in this case both species parasitize the blue shark, *Prionace glauca* (Linnaeus, 1758). Another intriguing example is found within the cestode order Cathetocephalidea. In 2005, a new genus was erected for *Sanguilevator yearsleyi* Caira, Mega, and Ruhnke, 2005, an unusual tapeworm of the broadfin shark, *Lamiopsis tephrodes* White, Last, Naylor, and Harris, 2010. This bizarre tapeworm has a series of internal chambers within its scolex that appear to be used to store red and white blood cells of its host (Caira et al., 2005). These structures are completely lacking in all species of its sister genus *Cathetocephalus* (Caira et al., 2005). Since this is a repeated trend within the tapeworms, it is important to understand if the same mechanisms are driving these changes across all of the different orders. This study provided the baseline data needed in order to begin exploring the genomic causes of this morphological diversification in the cestodes. Future studies could use this information to inform the hypotheses addressed here in these other cestodes groups.
Literature Cited


Litobothrium aenigmaticum, sp. nov. (Platyhelminthes: Cestoda: Litobothriidea), from the pelagic thresher shark with comments on development of known Litobothrium species.

Invertebrate Systematics 28, 231–243.


Coate, J.E., Doyle, J.J., 2015. Variation in transcriptome size: are we getting the message?

Chromosoma 124, 27–43.


Maldonado, L.L., Assis, J., Araújo, F.M.G., Salim, A.C., Macchiaroli, N., Cucher, M., Camicia,


Freeze-fracture of the syncytial layer, microvilli and discoid bodies. Tissue and Cell 18, 869–885.

Shumin, S., Xiaolei, L., Guanyu, J., Xuelin, W., Junwen, W., Xue, B., Jing, X., Jianda, P., Yining, S., Xinrui, W., 2018. Regulation of DNA methylation on key parasitism genes of cysticercus cellulosae revealed by integrative epigenomic-transcriptomic analyses. bioRxiv, 353417.


Wilson, R., Barnes, P., 1974. The tegument of *Schistosoma mansoni*: observations on the


Table 1. Characteristics of the 11 new cell types in the cephalic peduncle of *Litobothrium aenigmaticum*.

<table>
<thead>
<tr>
<th>Cell Type</th>
<th>Cell Size</th>
<th>Nuclei Size</th>
<th>Cytoplasm : Nuclei</th>
<th>Avg. Nuclei per Frame</th>
<th>Mitochon.</th>
<th>Rough ER</th>
<th>Golgi</th>
<th>Free Ribosomes</th>
<th>ED Vesicles</th>
<th>EL Inclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>23.54 ± 4.6 µm</td>
<td>3.63 ± 1.06 µm</td>
<td>1–7.62 ± 2.26</td>
<td>1</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>+</td>
<td>–</td>
</tr>
<tr>
<td>B</td>
<td>11.78 ± 2.83 µm</td>
<td>3.2 ± 0.38 µm</td>
<td>1–3.65 ± 0.72</td>
<td>5</td>
<td>+</td>
<td>NS</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>–</td>
</tr>
<tr>
<td>C</td>
<td>10.05 ± 2.11 µm</td>
<td>3.24 ± 0.51 µm</td>
<td>1–3.1 ± 0.60</td>
<td>7</td>
<td>+</td>
<td>NS</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>D</td>
<td>9.32 ± 1.74 µm</td>
<td>3.22 ± 0.25 µm</td>
<td>1–2.89 ± 0.64</td>
<td>9</td>
<td>+</td>
<td>NS</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>–</td>
</tr>
<tr>
<td>E</td>
<td>–</td>
<td>2.14 ± 0.68 µm</td>
<td>–</td>
<td>4</td>
<td>NS</td>
<td>+</td>
<td>NS</td>
<td>NS</td>
<td>+</td>
<td>–</td>
</tr>
<tr>
<td>F</td>
<td>13.24 ± 2.20 µm</td>
<td>3.81 ± 0.78 µm</td>
<td>1–3.66 ± 1.04</td>
<td>6</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>–</td>
</tr>
<tr>
<td>G</td>
<td>11.85 ± 2.35 µm</td>
<td>3.89 ± 0.53 µm</td>
<td>1–3.09 ± 0.74</td>
<td>5</td>
<td>+</td>
<td>NS</td>
<td>NS</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>H</td>
<td>12.76 ± 1.87 µm</td>
<td>3.22 ± 0.32 µm</td>
<td>1–3.44 ± 0.43</td>
<td>6</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>–</td>
</tr>
<tr>
<td>I</td>
<td>12.28 ± 1.47 µm</td>
<td>2.94 ± 0.36 µm</td>
<td>1–4.23 ± 0.73</td>
<td>6</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td></td>
<td>+</td>
<td>–</td>
</tr>
<tr>
<td>J</td>
<td>9.35 ± 2.12 µm</td>
<td>2.82 ± 0.41 µm</td>
<td>1–3.38 ± 0.91</td>
<td>12</td>
<td>+</td>
<td>NS</td>
<td>NS</td>
<td>+</td>
<td>+</td>
<td>–</td>
</tr>
<tr>
<td>K</td>
<td>14.51 ± 3.79 µm</td>
<td>3.46 ± 0.59 µm</td>
<td>1–4.23 ± 0.95</td>
<td>3</td>
<td>+</td>
<td>NS</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>
Table 2. Statistics on transcriptomic raw reads.

<table>
<thead>
<tr>
<th></th>
<th>ID</th>
<th>Species</th>
<th>No. raw reads</th>
<th>No. reads post trimming</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>KJKG_2_R1</td>
<td>L. aenigmaticum</td>
<td>36,722,777</td>
<td>29,707,191</td>
</tr>
<tr>
<td></td>
<td>KJKG_3_R1</td>
<td>L. aenigmaticum</td>
<td>33,717,790</td>
<td>27,524,462</td>
</tr>
<tr>
<td></td>
<td>KJKG_9_R1</td>
<td>L. aenigmaticum</td>
<td>36,617,424</td>
<td>29,341,655</td>
</tr>
<tr>
<td></td>
<td>KJKG_10_R1</td>
<td>L. aenigmaticum</td>
<td>39,960,442</td>
<td>32,040,272</td>
</tr>
<tr>
<td></td>
<td>KJKG_11_R1</td>
<td>L. aenigmaticum</td>
<td>41,318,436</td>
<td>28,867,143</td>
</tr>
<tr>
<td></td>
<td>KJKG_4_R1</td>
<td>L. nickoli</td>
<td>38,629,495</td>
<td>31,653,002</td>
</tr>
<tr>
<td></td>
<td>KJKG_5_R1</td>
<td>L. nickoli</td>
<td>41,424,952</td>
<td>33,973,280</td>
</tr>
<tr>
<td></td>
<td>KJKG_6_R1</td>
<td>L. nickoli</td>
<td>30,350,303</td>
<td>24,662,602</td>
</tr>
<tr>
<td></td>
<td>KJKG_12_R1</td>
<td>L. nickoli</td>
<td>33,293,115</td>
<td>27,710,158</td>
</tr>
<tr>
<td></td>
<td>KJKG_13_R1</td>
<td>L. nickoli</td>
<td>37,953,201</td>
<td>31,201,485</td>
</tr>
<tr>
<td></td>
<td>KJKG_14_R1</td>
<td>L. nickoli</td>
<td>35,176,505</td>
<td>28,381,204</td>
</tr>
<tr>
<td></td>
<td>KJKG_7_R1</td>
<td>L. daileyi</td>
<td>28,580,831</td>
<td>23,010,291</td>
</tr>
<tr>
<td></td>
<td>KJKG_15_R1</td>
<td>L. daileyi</td>
<td>35,626,809</td>
<td>29,111,547</td>
</tr>
<tr>
<td></td>
<td>KJKG_16_R1</td>
<td>L. daileyi</td>
<td>29,531,552</td>
<td>23,572,666</td>
</tr>
</tbody>
</table>
Table 3. Statistics on the Trinity de novo assemblies for 14 individual litobothriidean transcriptomes.

<table>
<thead>
<tr>
<th>ID</th>
<th>Species</th>
<th>Min. scaffold length</th>
<th>Largest scaffold</th>
<th>N50</th>
<th>Total no. of scaffolds</th>
<th>Total genome size</th>
</tr>
</thead>
<tbody>
<tr>
<td>KJKG2</td>
<td>L. aenigmaticum</td>
<td>300</td>
<td>18,010</td>
<td>1,682</td>
<td>45,713</td>
<td>53,031,932</td>
</tr>
<tr>
<td>KJKG3</td>
<td>L. aenigmaticum</td>
<td>300</td>
<td>9,327</td>
<td>1,476</td>
<td>31,929</td>
<td>33,535,388</td>
</tr>
<tr>
<td>KJKG9</td>
<td>L. aenigmaticum</td>
<td>300</td>
<td>14,626</td>
<td>1,675</td>
<td>38,910</td>
<td>45,552,240</td>
</tr>
<tr>
<td>KJKG10</td>
<td>L. aenigmaticum</td>
<td>300</td>
<td>18,429</td>
<td>2,022</td>
<td>40,223</td>
<td>53,446,605</td>
</tr>
<tr>
<td>KJKG11</td>
<td>L. aenigmaticum</td>
<td>300</td>
<td>14,472</td>
<td>1,897</td>
<td>38,309</td>
<td>49,093,774</td>
</tr>
<tr>
<td>KJKG4</td>
<td>L. nickoli</td>
<td>300</td>
<td>9,268</td>
<td>1,645</td>
<td>28,292</td>
<td>33,377,702</td>
</tr>
<tr>
<td>KJKG5</td>
<td>L. nickoli</td>
<td>300</td>
<td>12,841</td>
<td>1,666</td>
<td>26,903</td>
<td>32,235,795</td>
</tr>
<tr>
<td>KJKG6</td>
<td>L. nickoli</td>
<td>300</td>
<td>7,965</td>
<td>1,230</td>
<td>19,019</td>
<td>17,888,152</td>
</tr>
<tr>
<td>KJKG12</td>
<td>L. nickoli</td>
<td>300</td>
<td>7,010</td>
<td>1,330</td>
<td>20,162</td>
<td>20,393,841</td>
</tr>
<tr>
<td>KJKG13</td>
<td>L. nickoli</td>
<td>300</td>
<td>7,941</td>
<td>1,414</td>
<td>23,764</td>
<td>24,894,484</td>
</tr>
<tr>
<td>KJKG14</td>
<td>L. nickoli</td>
<td>300</td>
<td>10,348</td>
<td>1,519</td>
<td>27,442</td>
<td>30,033,859</td>
</tr>
<tr>
<td>KJKG7</td>
<td>L. daileyi</td>
<td>300</td>
<td>5,273</td>
<td>925</td>
<td>15,404</td>
<td>11,737,573</td>
</tr>
<tr>
<td>KJKG15</td>
<td>L. daileyi</td>
<td>300</td>
<td>12,894</td>
<td>1,470</td>
<td>27,017</td>
<td>29,437,626</td>
</tr>
<tr>
<td>KJKG16</td>
<td>L. daileyi</td>
<td>300</td>
<td>13,919</td>
<td>1,524</td>
<td>26,408</td>
<td>30,015,264</td>
</tr>
</tbody>
</table>
**Table 4.** Statistics on the clustered Trinity *de novo* assemblies for 14 individual litobothriidean transcriptomes.

<table>
<thead>
<tr>
<th>ID</th>
<th>Species</th>
<th>Min. scaffold length</th>
<th>Largest scaffold</th>
<th>N50</th>
<th>Total no. scaffolds</th>
<th>Total genome size</th>
</tr>
</thead>
<tbody>
<tr>
<td>KJKG2</td>
<td><em>L. aenigmaticum</em></td>
<td>300</td>
<td>15,818</td>
<td>1,434</td>
<td>36,927</td>
<td>37,320,552</td>
</tr>
<tr>
<td>KJKG3</td>
<td><em>L. aenigmaticum</em></td>
<td>300</td>
<td>9,327</td>
<td>1,307</td>
<td>25,749</td>
<td>24,400,348</td>
</tr>
<tr>
<td>KJKG9</td>
<td><em>L. aenigmaticum</em></td>
<td>300</td>
<td>14,544</td>
<td>1,464</td>
<td>30,858</td>
<td>31,860,307</td>
</tr>
<tr>
<td>KJKG10</td>
<td><em>L. aenigmaticum</em></td>
<td>300</td>
<td>16,306</td>
<td>1,716</td>
<td>31,696</td>
<td>36,104,906</td>
</tr>
<tr>
<td>KJKG11</td>
<td><em>L. aenigmaticum</em></td>
<td>300</td>
<td>14,472</td>
<td>1,632</td>
<td>29,646</td>
<td>32,777,112</td>
</tr>
<tr>
<td>KJKG4</td>
<td><em>L. nickoli</em></td>
<td>300</td>
<td>9,253</td>
<td>1,469</td>
<td>22,413</td>
<td>23,651,108</td>
</tr>
<tr>
<td>KJKG5</td>
<td><em>L. nickoli</em></td>
<td>300</td>
<td>12,841</td>
<td>1,474</td>
<td>20,638</td>
<td>21,956,993</td>
</tr>
<tr>
<td>KJKG6</td>
<td><em>L. nickoli</em></td>
<td>300</td>
<td>7,965</td>
<td>1,147</td>
<td>15,976</td>
<td>14,155,628</td>
</tr>
<tr>
<td>KJKG12</td>
<td><em>L. nickoli</em></td>
<td>300</td>
<td>7,941</td>
<td>1,265</td>
<td>16,544</td>
<td>15,869,200</td>
</tr>
<tr>
<td>KJKG13</td>
<td><em>L. nickoli</em></td>
<td>300</td>
<td>10,348</td>
<td>1,369</td>
<td>21,078</td>
<td>20,942,829</td>
</tr>
<tr>
<td>KJKG7</td>
<td><em>L. daileyi</em></td>
<td>300</td>
<td>5,269</td>
<td>889</td>
<td>12,705</td>
<td>9,345,770</td>
</tr>
<tr>
<td>KJKG15</td>
<td><em>L. daileyi</em></td>
<td>300</td>
<td>12,894</td>
<td>1,338</td>
<td>20,755</td>
<td>20,736,307</td>
</tr>
<tr>
<td>KJKG16</td>
<td><em>L. daileyi</em></td>
<td>300</td>
<td>13,351</td>
<td>1,404</td>
<td>20,853</td>
<td>21,784,688</td>
</tr>
</tbody>
</table>
Table 5. BUSCO Completeness scores for the Trinity *de novo* assemblies.

<table>
<thead>
<tr>
<th>ID</th>
<th>Species</th>
<th>Trinity assemblies</th>
<th>Clumped Trinity assemblies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Complete</td>
<td>Compete &amp; single</td>
</tr>
<tr>
<td>KJKG2</td>
<td><em>L. aenigmaticum</em></td>
<td>75.0%</td>
<td>48.2%</td>
</tr>
<tr>
<td>KJKG3</td>
<td><em>L. aenigmaticum</em></td>
<td>75.7%</td>
<td>59.1%</td>
</tr>
<tr>
<td>KJKG9</td>
<td><em>L. aenigmaticum</em></td>
<td>76.0%</td>
<td>49.0%</td>
</tr>
<tr>
<td>KJKG10</td>
<td><em>L. aenigmaticum</em></td>
<td>75.3%</td>
<td>47.2%</td>
</tr>
<tr>
<td>KJKG11</td>
<td><em>L. aenigmaticum</em></td>
<td>76.2%</td>
<td>52.0%</td>
</tr>
<tr>
<td>KJKG4</td>
<td><em>L. nickoli</em></td>
<td>78.7%</td>
<td>56.5%</td>
</tr>
<tr>
<td>KJKG5</td>
<td><em>L. nickoli</em></td>
<td>62.1%</td>
<td>43.7%</td>
</tr>
<tr>
<td>KJKG6</td>
<td><em>L. nickoli</em></td>
<td>70.1%</td>
<td>56.5%</td>
</tr>
<tr>
<td>KJKG12</td>
<td><em>L. nickoli</em></td>
<td>72.3%</td>
<td>58.1%</td>
</tr>
<tr>
<td>KJKG13</td>
<td><em>L. nickoli</em></td>
<td>74.0%</td>
<td>53.7%</td>
</tr>
<tr>
<td>KJKG14</td>
<td><em>L. nickoli</em></td>
<td>78.0%</td>
<td>53.5%</td>
</tr>
<tr>
<td>KJKG7</td>
<td><em>L. daileyi</em></td>
<td>42.4%</td>
<td>35.7%</td>
</tr>
<tr>
<td>KJKG15</td>
<td><em>L. daileyi</em></td>
<td>72.3%</td>
<td>56.3%</td>
</tr>
<tr>
<td>KJKG16</td>
<td><em>L. daileyi</em></td>
<td>73.6%</td>
<td>58.7%</td>
</tr>
</tbody>
</table>
Table 6. The statistics on the clustered, master transcriptomes.

<table>
<thead>
<tr>
<th>Species</th>
<th>Min. scaffold length</th>
<th>Largest scaffold</th>
<th>N50</th>
<th>Total no. of scaffolds</th>
<th>Total genome size</th>
</tr>
</thead>
<tbody>
<tr>
<td>L. aenigmaticum</td>
<td>300</td>
<td>15,162</td>
<td>1,308</td>
<td>21,727</td>
<td>20,992,134</td>
</tr>
<tr>
<td>L. daileyi</td>
<td>300</td>
<td>12,696</td>
<td>1,122</td>
<td>16,039</td>
<td>13,950,882</td>
</tr>
<tr>
<td>L. nickoli</td>
<td>300</td>
<td>10,164</td>
<td>1,116</td>
<td>18,643</td>
<td>15,887,406</td>
</tr>
</tbody>
</table>

Table 7. The BUSCO completeness scores for the clustered, master transcriptomes.

<table>
<thead>
<tr>
<th>Species</th>
<th>Complete</th>
<th>Compete &amp; single copy</th>
<th>Fragmented</th>
<th>Missing</th>
</tr>
</thead>
<tbody>
<tr>
<td>L. aenigmaticum</td>
<td>66.3%</td>
<td>57.8%</td>
<td>6.0%</td>
<td>27.7%</td>
</tr>
<tr>
<td>L. daileyi</td>
<td>65.1%</td>
<td>61.5%</td>
<td>7.9%</td>
<td>27.0%</td>
</tr>
<tr>
<td>L. nickoli</td>
<td>67.1%</td>
<td>60.4%</td>
<td>7.1%</td>
<td>25.8%</td>
</tr>
</tbody>
</table>
**Table 8.** The statistics on the raw sequencing reads for each of the genomic libraries.

<table>
<thead>
<tr>
<th>Species</th>
<th>Target Insert Size</th>
<th>Avg. Length of Reads</th>
<th>Index Number</th>
<th>No. of read pairs</th>
<th>% of contamination</th>
</tr>
</thead>
<tbody>
<tr>
<td>L. aenigmaticum</td>
<td>350</td>
<td>516</td>
<td>32</td>
<td>81,426,186</td>
<td>1.10%</td>
</tr>
<tr>
<td>L. aenigmaticum</td>
<td>550</td>
<td>610</td>
<td>11</td>
<td>74,288,258</td>
<td>0.20%</td>
</tr>
<tr>
<td>L. aenigmaticum</td>
<td>2,500</td>
<td>2,395</td>
<td>32</td>
<td>21,227,661</td>
<td>0%</td>
</tr>
<tr>
<td>L. aenigmaticum</td>
<td>7,500</td>
<td>7,539</td>
<td>32</td>
<td>18,466,882</td>
<td>0%</td>
</tr>
<tr>
<td>L. daileyi</td>
<td>350</td>
<td>532</td>
<td>5</td>
<td>83,556,525</td>
<td>0.19%</td>
</tr>
<tr>
<td>L. daileyi</td>
<td>550</td>
<td>649</td>
<td>5</td>
<td>48,534,695</td>
<td>0.20%</td>
</tr>
<tr>
<td>L. amplifica</td>
<td>350</td>
<td>521</td>
<td>19</td>
<td>76,212,112</td>
<td>0.20%</td>
</tr>
<tr>
<td>L. amplifica</td>
<td>550</td>
<td>638</td>
<td>19</td>
<td>39,985,806</td>
<td>0.55%</td>
</tr>
</tbody>
</table>

**Table 9.** Statistics on the genomic reads after they have undergone trimming.

<table>
<thead>
<tr>
<th>Species</th>
<th>Target Insert Size</th>
<th># paired reads</th>
<th># single reads</th>
<th>Mate Pair Reads</th>
<th>Pair End Reads</th>
<th>Single End</th>
</tr>
</thead>
<tbody>
<tr>
<td>L. aenigmaticum</td>
<td>350</td>
<td>63,611,891</td>
<td>10,579,663</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>L. aenigmaticum</td>
<td>550</td>
<td>58,789,186</td>
<td>12,756,713</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>L. aenigmaticum</td>
<td>2,500</td>
<td>N/A</td>
<td>N/A</td>
<td>1,393,410</td>
<td>1,491,328</td>
<td>519,879</td>
</tr>
<tr>
<td>L. aenigmaticum</td>
<td>7,500</td>
<td>N/A</td>
<td>N/A</td>
<td>1,230,707</td>
<td>1,278,315</td>
<td>454,992</td>
</tr>
<tr>
<td>L. daileyi</td>
<td>350</td>
<td>64,131,707</td>
<td>9,383,706</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>L. daileyi</td>
<td>550</td>
<td>38,702,360</td>
<td>8,167,453</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>L. amplifica</td>
<td>350</td>
<td>58,048,780</td>
<td>14,803,398</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>L. amplifica</td>
<td>550</td>
<td>31,351,129</td>
<td>7,120,693</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>
### Table 10. Genome size estimation with the program Jellyfish.

<table>
<thead>
<tr>
<th>Species</th>
<th>Library Size</th>
<th>17mer</th>
<th>19mer</th>
<th>21mer</th>
<th>25mer</th>
<th>27mer</th>
</tr>
</thead>
<tbody>
<tr>
<td>L. aenigmaticum</td>
<td>350</td>
<td>322,141,361</td>
<td>322,467,931</td>
<td>323,516,240</td>
<td>344,885,045</td>
<td>349,324,545</td>
</tr>
<tr>
<td>L. aenigmaticum</td>
<td>550</td>
<td>N/A</td>
<td>319,829,228</td>
<td>321,947,948</td>
<td>349,532,904</td>
<td>349,400,317</td>
</tr>
<tr>
<td>L. daileyi</td>
<td>350</td>
<td>330,933,548</td>
<td>331,733,116</td>
<td>333,198,663</td>
<td>338,384,257</td>
<td>362,535,199</td>
</tr>
<tr>
<td>L. daileyi</td>
<td>550</td>
<td>350,834,937</td>
<td>362,553,958</td>
<td>376,523,289</td>
<td>379,048,204</td>
<td>398,519,922</td>
</tr>
<tr>
<td>L. amplifica</td>
<td>550</td>
<td>382,565,171</td>
<td>406,221,396</td>
<td>435,623,729</td>
<td>455,994,192</td>
<td>439,645,656</td>
</tr>
</tbody>
</table>

### Table 11. Genome size estimation with BBMap.

<table>
<thead>
<tr>
<th>Species</th>
<th>Library Size</th>
<th>17mer</th>
<th>19mer</th>
<th>21mer</th>
<th>25mer</th>
<th>27mer</th>
</tr>
</thead>
<tbody>
<tr>
<td>L. aenigmaticum</td>
<td>550</td>
<td>334,254,711</td>
<td>335,798,270</td>
<td>344,370,008</td>
<td>354,224,693</td>
<td>354,210,515</td>
</tr>
<tr>
<td>L. daileyi</td>
<td>350</td>
<td>347,223,206</td>
<td>348,446,970</td>
<td>357,347,686</td>
<td>366,933,118</td>
<td>378,132,989</td>
</tr>
<tr>
<td>L. daileyi</td>
<td>550</td>
<td>367,586,238</td>
<td>373,935,006</td>
<td>396,238,999</td>
<td>394,994,009</td>
<td>412,513,742</td>
</tr>
<tr>
<td>L. amplifica</td>
<td>350</td>
<td>376,058,867</td>
<td>383,986,681</td>
<td>390,970,225</td>
<td>390,662,470</td>
<td>398,537,714</td>
</tr>
<tr>
<td>L. amplifica</td>
<td>550</td>
<td>442,139,774</td>
<td>434,878,735</td>
<td>426,474,197</td>
<td>440,814,743</td>
<td>473,883,178</td>
</tr>
</tbody>
</table>
Table 12. SOAPdenovo assemblies without mate pair libraries.

<table>
<thead>
<tr>
<th>Species</th>
<th>k-mer size</th>
<th>Largest scaffold</th>
<th>N50</th>
<th>Total no. scaffolds</th>
<th>Total genome size</th>
<th>Largest scaffold</th>
<th>N50</th>
<th>Total no. scaffolds</th>
<th>Total genome size</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>L. aenigmaticum</em></td>
<td>27</td>
<td>180,810</td>
<td>7,447</td>
<td>102,713</td>
<td>300,629,526</td>
<td>184,527</td>
<td>11,965</td>
<td>72,663</td>
<td>331,857,039</td>
</tr>
<tr>
<td><em>L. aenigmaticum</em></td>
<td>31</td>
<td>179,496</td>
<td>6,273</td>
<td>118,943</td>
<td>308,445,946</td>
<td>215,935</td>
<td>10,773</td>
<td>79,954</td>
<td>342,101,097</td>
</tr>
<tr>
<td><em>L. aenigmaticum</em></td>
<td>35</td>
<td>171,719</td>
<td>5,347</td>
<td>133,570</td>
<td>319,337,072</td>
<td>286,207</td>
<td>9,900</td>
<td>86,158</td>
<td>352,959,590</td>
</tr>
<tr>
<td><em>L. daileyi</em></td>
<td>27</td>
<td>188,917</td>
<td>12,042</td>
<td>73,083</td>
<td>325,050,051</td>
<td>188,917</td>
<td>16,023</td>
<td>59,558</td>
<td>345,424,829</td>
</tr>
<tr>
<td><em>L. daileyi</em></td>
<td>31</td>
<td>165,897</td>
<td>10,366</td>
<td>83,490</td>
<td>321,594,163</td>
<td>186,526</td>
<td>14,333</td>
<td>65,618</td>
<td>346,420,159</td>
</tr>
<tr>
<td><em>L. daileyi</em></td>
<td>35</td>
<td>158,006</td>
<td>9,059</td>
<td>93,693</td>
<td>321,008,597</td>
<td>158,006</td>
<td>13,167</td>
<td>70,221</td>
<td>349,220,283</td>
</tr>
<tr>
<td><em>L. amplifica</em></td>
<td>27</td>
<td>198,570</td>
<td>15,803</td>
<td>57,323</td>
<td>319,365,064</td>
<td>198,570</td>
<td>20,948</td>
<td>44,500</td>
<td>344,855,725</td>
</tr>
<tr>
<td><em>L. amplifica</em></td>
<td>31</td>
<td>175,353</td>
<td>14,575</td>
<td>64,383</td>
<td>332,680,477</td>
<td>176,123</td>
<td>20,186</td>
<td>47,568</td>
<td>347,301,910</td>
</tr>
<tr>
<td><em>L. amplifica</em></td>
<td>35</td>
<td>186,652</td>
<td>14,178</td>
<td>68,526</td>
<td>334,544,798</td>
<td>246,492</td>
<td>20,118</td>
<td>48,902</td>
<td>350,171,298</td>
</tr>
</tbody>
</table>
**Table 13.** SOAPdenovo assemblies with mate-pair libraries.

<table>
<thead>
<tr>
<th>Species</th>
<th>k-mer size</th>
<th>Largest scaffold</th>
<th>N50</th>
<th>Total no. scaffolds</th>
<th>Total genome size</th>
<th>Largest scaffold</th>
<th>N50</th>
<th>Total no. scaffolds</th>
<th>Total genome size</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>L. aenigmaticum</em></td>
<td>27</td>
<td>180,780</td>
<td>7,511</td>
<td>98,632</td>
<td>296,778,647</td>
<td>184,661</td>
<td>12,520</td>
<td>69,050</td>
<td>339,774,740</td>
</tr>
<tr>
<td><em>L. aenigmaticum</em></td>
<td>31</td>
<td>179,556</td>
<td>6,298</td>
<td>115,764</td>
<td>305,528,263</td>
<td>216,316</td>
<td>11,548</td>
<td>75,965</td>
<td>352,544,282</td>
</tr>
<tr>
<td><em>L. aenigmaticum</em></td>
<td>35</td>
<td>171,919</td>
<td>5,367</td>
<td>130,891</td>
<td>317,074,558</td>
<td>226,193</td>
<td>10,854</td>
<td>81,614</td>
<td>364,915,107</td>
</tr>
<tr>
<td><em>L. daileyi</em></td>
<td>27</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td><em>L. daileyi</em></td>
<td>31</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td><em>L. daileyi</em></td>
<td>35</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td><em>L. amplifica</em></td>
<td>27</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td><em>L. amplifica</em></td>
<td>31</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td><em>L. amplifica</em></td>
<td>35</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>
Table 14. ABySS2 assemblies without mate pair libraries.

<table>
<thead>
<tr>
<th>Species</th>
<th>k-mer size</th>
<th>Largest scaffold</th>
<th>N50</th>
<th>Total no. scaffolds</th>
<th>Total genome size</th>
<th>Largest scaffold</th>
<th>N50</th>
<th>Total no. scaffolds</th>
<th>Total genome size</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>L. aenigmaticum</em></td>
<td>27</td>
<td>137,000</td>
<td>9,612</td>
<td>57,546</td>
<td>250,157,438</td>
<td>137,000</td>
<td>14,639</td>
<td>35,173</td>
<td>277,527,658</td>
</tr>
<tr>
<td><em>L. aenigmaticum</em></td>
<td>31</td>
<td>155,210</td>
<td>14,114</td>
<td>41,505</td>
<td>262,405,385</td>
<td>183,504</td>
<td>18,208</td>
<td>30,045</td>
<td>280,432,751</td>
</tr>
<tr>
<td><em>L. aenigmaticum</em></td>
<td>35</td>
<td>167,214</td>
<td>18,050</td>
<td>33,714</td>
<td>268,280,519</td>
<td>189,739</td>
<td>21,870</td>
<td>26,573</td>
<td>283,256,722</td>
</tr>
<tr>
<td><em>L. daileyi</em></td>
<td>27</td>
<td>95,613</td>
<td>7,915</td>
<td>65,874</td>
<td>224,128,984</td>
<td>120,178</td>
<td>12,593</td>
<td>39,162</td>
<td>274,533,532</td>
</tr>
<tr>
<td><em>L. daileyi</em></td>
<td>31</td>
<td>140,251</td>
<td>12,135</td>
<td>46,305</td>
<td>258,668,634</td>
<td>154,469</td>
<td>15,899</td>
<td>33,124</td>
<td>277,181,504</td>
</tr>
<tr>
<td><em>L. daileyi</em></td>
<td>35</td>
<td>142,877</td>
<td>15,244</td>
<td>37,910</td>
<td>265,010,818</td>
<td>181,113</td>
<td>18,341</td>
<td>29,679</td>
<td>279,899,663</td>
</tr>
<tr>
<td><em>L. amplifica</em></td>
<td>27</td>
<td>80,683</td>
<td>6,649</td>
<td>78,568</td>
<td>244,340,710</td>
<td>123,967</td>
<td>9,320</td>
<td>52,370</td>
<td>290,047,530</td>
</tr>
<tr>
<td><em>L. amplifica</em></td>
<td>31</td>
<td>143,416</td>
<td>9,562</td>
<td>59,294</td>
<td>268,993,684</td>
<td>144,462</td>
<td>12,398</td>
<td>42,540</td>
<td>298,829,428</td>
</tr>
<tr>
<td><em>L. amplifica</em></td>
<td>35</td>
<td>119,186</td>
<td>12,441</td>
<td>48,498</td>
<td>292,612,921</td>
<td>147,196</td>
<td>15,389</td>
<td>36,371</td>
<td>303,935,085</td>
</tr>
</tbody>
</table>
Table 15. ABySS2 assemblies with mate-pair libraries.

<table>
<thead>
<tr>
<th>Species</th>
<th>k-mer size</th>
<th>Largest scaffold</th>
<th>N50</th>
<th>Total number of scaffolds</th>
<th>Total genome size</th>
<th>Largest scaffold</th>
<th>N50</th>
<th>Total number of scaffolds</th>
<th>Total genome size</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>L. aenigmaticum</em></td>
<td>27</td>
<td>77,654</td>
<td>5,273</td>
<td>75,998</td>
<td>248,632,364</td>
<td>116,706</td>
<td>13,211</td>
<td>37,651</td>
<td>279,607,219</td>
</tr>
<tr>
<td><em>L. aenigmaticum</em></td>
<td>31</td>
<td>79,521</td>
<td>7,881</td>
<td>56,664</td>
<td>261,230,091</td>
<td>150,941</td>
<td>16,238</td>
<td>32,240</td>
<td>282,177,161</td>
</tr>
<tr>
<td><em>L. aenigmaticum</em></td>
<td>35</td>
<td>88,445</td>
<td>10,30</td>
<td>6</td>
<td>267,327,203</td>
<td>190,825</td>
<td>19,319</td>
<td>28,598</td>
<td>284,819,813</td>
</tr>
<tr>
<td><em>L. daileyi</em></td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td><em>L. daileyi</em></td>
<td>31</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td><em>L. daileyi</em></td>
<td>35</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td><em>L. amplifica</em></td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td><em>L. amplifica</em></td>
<td>31</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td><em>L. amplifica</em></td>
<td>35</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>
Table 16. SPAdes assembly statistics.

<table>
<thead>
<tr>
<th>Species</th>
<th>Largest scaffold</th>
<th>N50</th>
<th>Total no. scaffolds</th>
<th>Total genome size</th>
<th>SSPACE scaffolded assemblies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L. aenigmaticum</td>
<td>208,932</td>
<td>19,530</td>
<td>37,089</td>
<td>283,966,322</td>
<td>382,813</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>45,536</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>22,058</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>302,424,677</td>
</tr>
<tr>
<td>L. daileyi</td>
<td>157,398</td>
<td>16,852</td>
<td>45,351</td>
<td>287,655,135</td>
<td>452,856</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>37,572</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>23,081</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>296,102,134</td>
</tr>
<tr>
<td>L. amplifica</td>
<td>202,549</td>
<td>18,096</td>
<td>41,027</td>
<td>314,209,663</td>
<td>393,517</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>38,108</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>21,318</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>320,039,359</td>
</tr>
</tbody>
</table>

Table 17. SPAdes assemblies scaffolded with master transcriptomes using tranScaff.

<table>
<thead>
<tr>
<th>Species</th>
<th>Largest scaffold</th>
<th>N50</th>
<th>Total no. scaffolds</th>
<th>Total genome size</th>
</tr>
</thead>
<tbody>
<tr>
<td>L. aenigmaticum</td>
<td>446,001</td>
<td>58,137</td>
<td>20,131</td>
<td>302,510,833</td>
</tr>
<tr>
<td>L. daileyi</td>
<td>452,856</td>
<td>45,652</td>
<td>21,129</td>
<td>296,177,473</td>
</tr>
<tr>
<td>L. amplifica</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>


**Table 18.** MaSURCA assembly statistics.

<table>
<thead>
<tr>
<th>Species</th>
<th>Largest scaffold</th>
<th>(N_50)</th>
<th>Total no. scaffolds</th>
<th>Total genome size</th>
<th>Largest scaffold</th>
<th>(N_50)</th>
<th>Total no. scaffolds</th>
<th>Total genome size</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>L. aenigmaticum</em></td>
<td>76,510</td>
<td>4,533</td>
<td>120,419</td>
<td>329,296,342</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

**Table 19.** MaSURCA assemblies scaffolded with master transcriptomes using tranScaff.

<table>
<thead>
<tr>
<th>Species</th>
<th>Largest scaffold</th>
<th>(N_50)</th>
<th>Total no. scaffolds</th>
<th>Total genome size</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>L. aenigmaticum</em></td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td><em>L. daileyi</em></td>
<td>453,377</td>
<td>51,464</td>
<td>26,920</td>
<td>325,242,964</td>
</tr>
<tr>
<td><em>L. amplifica</em></td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>
### Table 20. The statistics for genome assemblies selected as final assemblies.

<table>
<thead>
<tr>
<th>Species</th>
<th>N50</th>
<th>Total no. scaffolds</th>
<th>Genome Size</th>
<th>Repeat content</th>
<th>GC content</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>L. aenigmaticum</em></td>
<td>58,137</td>
<td>20,131</td>
<td>302,510,833</td>
<td>27.40%</td>
<td>43.38%</td>
</tr>
<tr>
<td><em>L. daileyi</em></td>
<td>45,652</td>
<td>21,129</td>
<td>296,177,473</td>
<td>28.10%</td>
<td>43.50%</td>
</tr>
<tr>
<td><em>L. amplifica</em></td>
<td>42,370</td>
<td>28,835</td>
<td>355,593,269</td>
<td>31.35%</td>
<td>42.90%</td>
</tr>
</tbody>
</table>

### Table 21. MAKER Annotation statistics for *L. aenigmaticum*.

<table>
<thead>
<tr>
<th>Round</th>
<th>Total genes</th>
<th>BUSCO: All Genes</th>
<th>No. Complete Genes</th>
<th>BUSCO: Complete</th>
<th>Complete, Canonical</th>
<th>BUSCO: Complete, Canonical</th>
<th>No. Partial Genes</th>
<th>Usable Partial</th>
<th>Usable Models (Complete &amp; Partial)</th>
<th>BUSCO: Complete &amp; Partial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Round 1</td>
<td>14,886</td>
<td>61.2%</td>
<td>11,096</td>
<td>49.8%</td>
<td>7,618</td>
<td>35.3%</td>
<td>3,790</td>
<td>3,223</td>
<td>14,319</td>
<td>72.9%</td>
</tr>
<tr>
<td>Round 2</td>
<td>10,015</td>
<td>61.0%</td>
<td>4,221</td>
<td>32.8%</td>
<td>4,151</td>
<td>35.1%</td>
<td>5,794</td>
<td>2,701</td>
<td>6,922</td>
<td>56.2%</td>
</tr>
<tr>
<td>Round 3</td>
<td>10,196</td>
<td>74.90%</td>
<td>4,944</td>
<td>39.7%</td>
<td>3,758</td>
<td>32.2%</td>
<td>5,252</td>
<td>1,123</td>
<td>6,067</td>
<td>49.3%</td>
</tr>
</tbody>
</table>
### Table 22. MAKER Annotation statistics for *L. daileyi*.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Round 1</td>
<td>16,425</td>
<td>68.8%</td>
<td>12,234</td>
<td>50.6%</td>
<td>8,743</td>
<td>35.6%</td>
<td>3,562</td>
<td>72.1%</td>
</tr>
<tr>
<td>Round 2</td>
<td>11,965</td>
<td>76.9%</td>
<td>7,400</td>
<td>53.0%</td>
<td>6,878</td>
<td>35.7%</td>
<td>2,611</td>
<td>69.3%</td>
</tr>
<tr>
<td>Round 3</td>
<td>10,559</td>
<td>74.0%</td>
<td>4,557</td>
<td>33.9%</td>
<td>4,161</td>
<td>33.9%</td>
<td>2,600</td>
<td>64.3%</td>
</tr>
</tbody>
</table>

### Table 23. MAKER Annotation statistics for SPAdes assembly of *L. amplifica*.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Round 1</td>
<td>5,399</td>
<td>17.7%</td>
<td>4,183</td>
<td>10.9%</td>
<td>2,946</td>
<td>7.3%</td>
<td>1,065</td>
<td>5.2%</td>
</tr>
<tr>
<td>Round 2</td>
<td>3,802</td>
<td>17.7%</td>
<td>1,846</td>
<td>7.5%</td>
<td>1,329</td>
<td>6.1%</td>
<td>398</td>
<td>10.2%</td>
</tr>
<tr>
<td>Round 3</td>
<td>3,538</td>
<td>17.4%</td>
<td>1,401</td>
<td>5.9%</td>
<td>1,386</td>
<td>6.2%</td>
<td>912</td>
<td>11.1%</td>
</tr>
</tbody>
</table>
Table 24. MAKER Annotation statistics for MaSURCA assembly of *L. amplifica*.

<table>
<thead>
<tr>
<th>Round</th>
<th>Total genes</th>
<th>BUSCO: All Genes</th>
<th>Number Complete Genes</th>
<th>BUSCO: Complete</th>
<th>Complete &amp; Canonical</th>
<th>BUSCO: Complete &amp; Canonical</th>
<th>Number Partial Genes</th>
<th>Usable Partials</th>
<th>Usable Models (Complete &amp; Partial)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Round 1</td>
<td>23,318</td>
<td>67.1%</td>
<td>17,686</td>
<td>49.7%</td>
<td>13,727</td>
<td>35.2%</td>
<td>2,913</td>
<td>20,599</td>
<td>35.4%</td>
</tr>
<tr>
<td>Round 2</td>
<td>15,091</td>
<td>69.0%</td>
<td>6,472</td>
<td>30.9%</td>
<td>4,271</td>
<td>20.7%</td>
<td>1,767</td>
<td>8,239</td>
<td>40.9%</td>
</tr>
<tr>
<td>Round 3</td>
<td>13,761</td>
<td>68.9%</td>
<td>4,749</td>
<td>25.2%</td>
<td>4,470</td>
<td>23.6%</td>
<td>3,570</td>
<td>8,319</td>
<td>42.9%</td>
</tr>
</tbody>
</table>

Table 25. Braker annotation statistics for *L. aenigmaticum*.

<table>
<thead>
<tr>
<th></th>
<th>Total genes</th>
<th>BUSCO: Total Genes</th>
<th>Canonical Genes</th>
<th>Canonical &amp; Complete Genes</th>
<th>BUSCO: Complete &amp; Canonical</th>
<th>Usable Partials</th>
<th>Total Models: Complete &amp; Partial</th>
<th>BUSCO: Complete &amp; Partial</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNA only</td>
<td>14,039</td>
<td>64.9%</td>
<td>11,383</td>
<td>8,088</td>
<td>42.4%</td>
<td>505</td>
<td>8,593</td>
<td>44.8%</td>
</tr>
<tr>
<td>RNA &amp; Protein</td>
<td>14,425</td>
<td>67.6%</td>
<td>11,659</td>
<td>8,232</td>
<td>44.7%</td>
<td>634</td>
<td>8,866</td>
<td>47.7%</td>
</tr>
</tbody>
</table>
Table 26. Braker annotation statistics for *L. daileyi*.

<table>
<thead>
<tr>
<th></th>
<th>Total genes</th>
<th>BUSCO - Total Genes</th>
<th>Canonical Genes</th>
<th>Canonical &amp; Complete Genes</th>
<th>BUSCO - Complete &amp; Canonical</th>
<th>Usable Partials</th>
<th>Total Models: Complete &amp; Partial</th>
<th>BUSCO - Complete &amp; Partial</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNA evidence</td>
<td>13,586</td>
<td>65.2%</td>
<td>11,105</td>
<td>7,860</td>
<td>42.3%</td>
<td>2,617</td>
<td>10,477</td>
<td>51.6%</td>
</tr>
<tr>
<td>RNA &amp; Protein evidence</td>
<td>14,205</td>
<td>70.3%</td>
<td>11,557</td>
<td>8,428</td>
<td>47.4%</td>
<td>2,784</td>
<td>11,212</td>
<td>57.5%</td>
</tr>
</tbody>
</table>

Table 27. GMAP annotation statistics. *L. amplica* (A) = *L. aenigmaticum* transcripts mapped to *L. amplifica*; *L. amplica* (B) = *L. daileyi* transcripts mapped to *L. amplifica*.

<table>
<thead>
<tr>
<th></th>
<th>Total genes</th>
<th>Canonical Genes</th>
<th>Canonical &amp; Complete Genes</th>
<th>BUSCO - Complete &amp; Canonical</th>
<th>Usable Partials</th>
<th>Total Models: Complete &amp; Partial</th>
<th>BUSCO - Complete &amp; Partial</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>L. aenigmaticum</em></td>
<td>18,541</td>
<td>17,037</td>
<td>8,791</td>
<td>55.4%</td>
<td>6,083</td>
<td>14,874</td>
<td>67.7%</td>
</tr>
<tr>
<td><em>L. daileyi</em></td>
<td>14,400</td>
<td>13,506</td>
<td>5,044</td>
<td>47.7%</td>
<td>5,156</td>
<td>10,200</td>
<td>61.8%</td>
</tr>
<tr>
<td><em>L. amplifica</em> (A)</td>
<td>57</td>
<td>55</td>
<td>20</td>
<td>0.2%</td>
<td>34</td>
<td>42</td>
<td>0.2%</td>
</tr>
<tr>
<td><em>L. amplifica</em> (B)</td>
<td>32</td>
<td>31</td>
<td>14</td>
<td>0.1%</td>
<td>12</td>
<td>26</td>
<td>0.1%</td>
</tr>
</tbody>
</table>
Table 28. Final annotation gene model statistics.

<table>
<thead>
<tr>
<th>Species</th>
<th>MAKER Models</th>
<th>Braker Models</th>
<th>GMAP Models</th>
<th>Total Models</th>
<th>Complete BUSCO</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>L. aenigmaticum</em></td>
<td>5,608</td>
<td>6,750</td>
<td>8,712</td>
<td>21,070</td>
<td>50.9%</td>
</tr>
<tr>
<td><em>L. amplifica</em></td>
<td>8,318</td>
<td>N/A</td>
<td>42</td>
<td>8,358</td>
<td>42.9%</td>
</tr>
<tr>
<td><em>L. daileyi</em></td>
<td>6,625</td>
<td>7,791</td>
<td>4,668</td>
<td>19,084</td>
<td>51.2%</td>
</tr>
</tbody>
</table>

Table 29. Genome annotation gene model statistics for multiexonic genes.

<table>
<thead>
<tr>
<th>Species</th>
<th>Multiexonic Genes</th>
<th>Avg. Size Multiexonic</th>
<th>Size Range Multiexonic</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>L. aenigmaticum</em></td>
<td>17,663</td>
<td>9,784</td>
<td>219–184,172</td>
</tr>
<tr>
<td><em>L. amplifica</em></td>
<td>6,635</td>
<td>7,923</td>
<td>233–112,622</td>
</tr>
<tr>
<td><em>L. daileyi</em></td>
<td>15,759</td>
<td>9,282</td>
<td>100–119,396</td>
</tr>
</tbody>
</table>
Table 30. Genome annotation gene model statistics for monoexonic genes.

<table>
<thead>
<tr>
<th>Species</th>
<th>Monoexonic Genes</th>
<th>Avg. Size Monoexonic</th>
<th>Size Range Monoexonic</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>L. aenigmaticum</em></td>
<td>3,407</td>
<td>775</td>
<td>96–11,549</td>
</tr>
<tr>
<td><em>L. amplifica</em></td>
<td>1,725</td>
<td>796</td>
<td>111–16,740</td>
</tr>
<tr>
<td><em>L. daileyi</em></td>
<td>3,325</td>
<td>879</td>
<td>51–11,505</td>
</tr>
</tbody>
</table>

Table 31. Genome annotation gene model statistics.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><em>L. aenigmaticum</em></td>
<td>8,324</td>
<td>1,225</td>
<td>224</td>
<td>6.3</td>
<td>208</td>
<td>5.3</td>
<td>1,584</td>
</tr>
<tr>
<td><em>L. amplifica</em></td>
<td>6,452</td>
<td>1,129</td>
<td>272</td>
<td>5</td>
<td>249</td>
<td>4</td>
<td>1,682</td>
</tr>
<tr>
<td><em>L. daileyi</em></td>
<td>7,818</td>
<td>1,181</td>
<td>227</td>
<td>6.0</td>
<td>210</td>
<td>5.0</td>
<td>1,548</td>
</tr>
</tbody>
</table>
Table 32. Number of genes in gene families associated with hypotheses of Tsai et al. (2013).

<table>
<thead>
<tr>
<th>Novel/Expanded Tapeworm Genes</th>
<th>No. in Cyclophyllideans</th>
<th>No. in L. aenigmaticum</th>
<th>No. in L. daileyi</th>
<th>No. in L. amplifica</th>
</tr>
</thead>
<tbody>
<tr>
<td>Laminin family</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Novel protocadherin family</td>
<td>3–6</td>
<td>8</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>Thrombospondin-containing</td>
<td>5–8</td>
<td>49</td>
<td>46</td>
<td>61</td>
</tr>
<tr>
<td>family</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LDL receptor</td>
<td>18–26</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Galactosyl transferases</td>
<td>5–20</td>
<td>19</td>
<td>15</td>
<td>27</td>
</tr>
<tr>
<td>CD2 domain containing</td>
<td>1–3</td>
<td>5</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>protein family</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Novel transmembrane family</td>
<td>2–13</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Diagnostic antigen 50</td>
<td>9–29</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>(classical)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tegumental antigen</td>
<td>5–10</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>BTB, BACK, Kelch protein</td>
<td>3–24</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>family</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ubiquitin conjugating enzyme</td>
<td>5–11</td>
<td>1</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Ortopetrin-like</td>
<td>4–9</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Novel repeat domain family</td>
<td>1</td>
<td>5</td>
<td>14</td>
<td>1</td>
</tr>
<tr>
<td>Fibronectin type III domain</td>
<td>1–10</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>containing protein</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Novel flatworm family</td>
<td>1–2</td>
<td>3</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Novel Taeniid protein</td>
<td>6–25</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Genes with species specific</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>expansions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Novel <em>E. multilocularis</em></td>
<td>0–13</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>family, unknown function</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Novel <em>E. multilocularis</em></td>
<td>6–46</td>
<td>25</td>
<td>29</td>
<td>93</td>
</tr>
<tr>
<td>family</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Novel <em>E. multilocularis</em></td>
<td>3–32</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>family, similar to Chromo-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>domain</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Novel <em>H. microstoma</em> family</td>
<td>28</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>gag-pol TE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Novel <em>H. microstoma</em> family</td>
<td>47</td>
<td>5</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>TE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein kinase</td>
<td>6–16</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>
Figure 1. Basic body of a typical tapeworm illustrated on *Echinococcus granulosus*. Drawing based on the illustration by Xiao et al., 2005. Sc=scolex; St=strobila.
Figure 2. Basic body plan of a typical litobothriidean, *L. nickoli* (A), and the unusual *L. aenigmaticum* (B). A. Sc indicates scolex, St indicates strobili. B. SP indicates scolex proper, CP indicates cephalic peduncle, PR indicates immature proglottids.
Figure 3. Phylogenetic tree of litobothriidean relationships from Caira et al. (2014a).
Figure 4. Comparison of larval and adult forms of *Litobothrium daileyi* (A) and *Litobothrium aenigmaticum* (B). In each case, scanning electron micrograph of the larval form to the left and line drawing of the adult to the right. Images from Olson and Caira (2001) and Caira et al. (2014a; 2017).
**Figure 5.** Phylogenetic tree of litobothriidean relationships from Caira et al. (2017).
Figure 6. Schematic of the four “tissue” types described by Caira et al. (2014a). Modified from Caira et al. (2014a).
Figure 7. Schematic of the location and associations of each of the 11 cell types in a lateral section of the scolex of *L. aenigmaticum*. The six cross sections illustrate the associations among cell types.
Figure 8. Transmission electron micrographs of cross sections of cell type A, B, and C of Litobothrium aenigmaticum. A–C. Transmission electron micrographs of cell type A. A. Basic characteristics of cell type A. B. Details of electron dense vesicles (V) and micro-vesicles (MV). C. Higher magnification of vesicles and micro-vesicles. D–G. Transmission electron micrographs of cell type B. D. Basic characteristics of cell type B. E. Details of nucleus (N), mitochondria (M), vesicles (V), and golgi apparatus (Go). F. Higher magnification of vesicles (V). G. Electron lucent vesicles (LV). H–J. Transmission electron micrographs of cell type C. H. Basic characteristics of cell type C. I. Details of mitochondria, Golgi apparatus, and vesicles. J. Details of electron lucent vesicles. Scale bars: A,D,G,H=2 µm; B,E,F,I,J=500 nm; C=100 nm.
Figure 9. Transmission electron micrographs of cross sections of cell types D, E, and F of *Litobothrium aenigmaticum*. A–C. Transmission electron micrographs of cell type D. A. Basic characteristics of cell type D. B. Details of mitochondria and free ribosomes (Ri). C. Details of Golgi apparati and vesicles. D–F. Transmission electron micrographs of cell type E. D. Basic characteristics of cell type E. E. Details of vesicles, microtubules (arrows), and membranes (Me). F. Details of RER (R) and membranes. G–J. Transmission electron micrographs of cell type F. G. Basic characteristics of cell type F. H. Details of mitochondria. I. Details of Golgi apparatus and vesicles. J. Details of RER and nucleus. Scale bars: A,D,G=2 µm; B,C,F,H,I,J = 500 nm; E=100 nm.
Figure 11. Transmission electron micrographs of cross sections of cell types J and K and the excretory system of *Litobothrium aenigmaticum*. A–B. Transmission electron micrographs of cell type J. A. Basic characteristics of cell type J. B. Details of the nucleus, mitochondria, vesicles, and free ribosomes. C–E. Transmission electron micrographs of cell type K. C. Basic characteristics of cell type K. D. Details of mitochondria, electron dense and lucent vesicles, and Golgi apparatus. E. Details of electron lucent vesicles. F–I. Transmission electron micrographs of sublateral pairs of ducts. F. One of the paired excretory ducts (Ex). G. Small collecting ducts (C) and flame cells (F) associated with the protonephridia. H. Small collecting duct connected to a flame cell. I. Cross section of a flame cell. Scale bars: A,C,G,H=2 μm; B,D,E,I=500 nm; F=10 μm.
Figure 13. Periodic acid-Schiff (PAS) stained sections of different levels (A–C) of the cephalic peduncle of *L. aenigmaticum*; note the variability in PAS-positive results across sections. Cell types present within these cross sections are cell type A (A), cell type B (B), and cell type E (arrow head). Scale bars=250 µm.
**Figure 14.** Species tree that resulted from clustering the 14 individual litobothriidean transcriptomes with Orthofinder.
Figure 15. Species tree from clustering the litobothriidean and *E. multilocularis* genome annotation gene models in Orthofinder.
**Figure 16.** Comparison of number of genes in gene families that were hypothesized to have flatworm specific expansions by Tsai et al. (2013). Red = *E. multilocularis*, Orange = *E. granulosus*, Pink = *H. microstoma*, Blue = *L. aenigmaticum*, Green = *L. daileyi*, Purple = *L. amplifica*. 
Figure 17. Comparison of number of genes in gene families that were hypothesized to have expansions in trematodes and cestodes by Tsai et al. (2013). Red = *E. multilocularis*, Orange = *E. granulosus*, Pink = *H. microstoma*, Blue = *L. aenigmaticum*, Green = *L. daileyi*, Purple = *L. amplifica*. 
Figure 18. Comparison of number of genes in gene families hypothesized to be specific to tapeworms by Tsai et al. (2013). Red = *E. multilocularis*, Orange = *E. granulosus*, Pink = *H. microstoma*, Blue = *L. aenigmaticum*, Green = *L. daileyi*, Purple = *L. amplifica*. 
Figure 19. Comparison of number of genes in gene families hypothesized to have tapeworm-specific expansions by Tsai et al. (2013). Red = *E. multilocularis*, Orange = *E. granulosus*, Pink = *H. microstoma*, Blue = *L. aenigmaticum*, Green = *L. daileyi*, Purple = *L. amplifica*. 
Figure 20. Comparison of number of genes in gene families hypothesized to be expanded in the *Taenia* genomes by Tsai et al. (2013). Red = *E. multilocularis*, Orange = *E. granulosus*, Pink = *H. microstoma*, Blue = *L. aenigmaticum*, Green = *L. daileyi*, Purple = *L. amplifica*. 
Figure 21. Comparison of number of genes in gene families hypothesized to have species-specific expansions by Tsai et al. (2013). Red = *E. multilocularis*, Orange = *E. granulosus*, Pink = *H. microstoma*, Blue = *L. aenigmaticum*, Green = *L. daileyi*, Purple = *L. amplifica*. 
A. *Litobothrium aenigmaticum*

B. *Litobothrium daileyi*

C. *Litobothrium amplifica*

**Figure 22.** Revigo biological process treemap for litobothriidean-specific orthogroups with three different species backgrounds.
Figure 23. Revigo cellular component treemap for litobothriidean-specific orthogroups with three different species background.
Figure 24. Revigo molecular function treemap for litobothriidean-specific orthogroups with three different species backgrounds.
Figure 25. Revigo treemaps for *Litobothrium daileyi*-specific orthogroups with three different species background.
Figure 26. Hierarchical correlation matrix of 14 litobothriidean transcriptomes from DESeq2.
Figure 27. Plot of first and second principal components from principal components analysis of 14 litobothriidean transcriptomes.
Figure 28. Plot of second and third principal components from principal components analysis of 14 litobothriidean transcriptomes.
Figure 29. Clustering of 1,000 most expressed genes from DESeq2 analysis of 14 litobothriidean transcriptomes. Clusters of interest that were examined are marked with boxes. Symbol key: star = L. nickoli upregulated clusters; circle = L. aenigmaticum downregulated clusters; triangle = L. daileyi downregulated clusters; square = L. aenigmaticum upregulated cluster; 7 pointed star = L. nickoli downregulated clusters.
Figure 30. Biological process treemap for transcripts down regulated in *L. aenigmaticum* with three different species backgrounds.
Figure 31. Cellular component treemap for transcripts down regulated in *L. aenigmaticum* with three different species backgrounds.
**Figure 32.** Molecular function treemap for transcripts down regulated in *L. aenigmaticum* with three different species backgrounds.
Figure 33. Biological process treemap for transcripts downregulated in *L. daileyi* with three different species backgrounds.
Figure 34. Cellular component treemap for transcripts downregulated in *L. daileyi* with three different species backgrounds.
Figure 35. Molecular function treemap for transcripts downregulated in *L. daileyi* with three different species backgrounds.
Figure 36. Biological process treemap for transcripts downregulated in *L. nickoli* with three different species backgrounds.
Figure 37. Chemical component treemap for transcripts downregulated in *L. nickoli* with three different species backgrounds.
Figure 38. Molecular function treemap for transcripts downregulated in *L. nickoli* with three different species backgrounds.
Figure 39. Biological process treemap for transcripts upregulated in *L. nickoli* with three different species backgrounds.
Figure 40. Cellular component treemap for transcripts upregulated in *L. nickoli* with *L. aenigmaticum* and *L. daileyi* backgrounds. No GO terms were significantly enriched with the *L. nickoli* background.
Figure 41. Molecular component treemap for transcripts upregulated in *L. nickoli* with three different species backgrounds.
Figure 42. Dot plot resulting from SynMap analysis comparing *L. aenigmaticum* and *L. daileyi*. 
Figure 43. Dot plot resulting from SynMap analysis comparing *L. aenigmaticum* and *L. amplifica*.
Figure 44. Dot plot resulting from SynMap analysis comparing *L. amplifica* and *L. daileyi*.
Figure 45. Histograms of synonymous substitution rates, nonsynonymous substitution rates, and nonsynonymous to synonymous substitution rates ratio. A–B. Nonsynonymous rates. B is detail of area indicates with rectangle in A. C–D. Synonymous rates. D is detail of area indicates with rectangle in C. E–F. Nonsynonymous to synonymous substitution rates ratio. F is detail of area indicates with rectangle in E.
Figure 46. Transformed species tree from ETE3 Toolkit used for CAFE analysis.
Figure 47. Phylogeny resulting from CAFE analysis. Numbers of rapidly evolving gene families are indicated next to species names or nodes; numbers in parentheses represent number of families that have been expanded and the number that have been contracted, respectively.
**Figure 48.** Biological process, cellular component, and molecular function GO terms associated with gene families CAFE found to be expanded in *Litobothrium aenigmaticum*. 
Figure 49. Biological process, cellular component, and molecular function GO terms associated with gene families CAFE found to be contracted in *Litobothrium aenigmaticum*. 
Figure 50. Biological process, cellular component, and molecular function GO terms associated with gene families CAFE found to be expanded in *Litobothrium daileyi*.
Figure 51. Biological process, cellular component, and molecular function treemaps for gene families CAFE analysis found to be rapidly expanding in *L. amplifica*. 
Biological process, cellular component, and molecular function treemaps for gene families CAFE analysis found to be rapidly contracted in *L. amplifica*.
Figure 53. Biological process tree map for gene families CAFE analysis identified as significantly rapidly expanding in both *L. aenigmaticum* and *L. daileyi.*
Figure 54. Cellular component treemap for gene families CAFE analysis identified as significantly rapidly expanding in both *L. aenigmaticum* and *L. daileyi*.
Figure 55. Molecular function treemap for gene families CAFE analysis identified as significantly rapidly expanding in both *L. aenigmaticum* and *L. daileyi*.
Figure 56. Biological process, cellular component, and molecular function treemaps for gene families CAFE analysis found to be contracted in all three litobothriideans.
Figure 57. Biological process and cellular component treemaps for gene families CAFE analysis found to be expanded in *E. multilocularis*. No molecular function GO terms were found to be significantly enriched.
Figure 58. Biological process treemaps for gene families CAFE analysis found to be contracted in *E. multilocularis*. 
Figure 59. Cellular component treemaps for gene families CAFE analysis found to be contracted in *E. multilocularis*. 
Figure 60. Molecular function treemaps for gene families CAFE analysis found to be contracted in *E. multilocularis*. 
<table>
<thead>
<tr>
<th>L. aerogenes</th>
<th>L. diaei</th>
<th>L. amylacea</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary metabolism</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organic substance Metabolism</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nitrogen compound metabolism</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biosynthetic process</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Response to biotic stimulus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Response to activity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interspecies interaction between organisms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Immune response</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antigen processing and presentation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reproduction of single cell organism</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Developmental process involved in reproduction</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deaminase activity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oxidoreductase activity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Isomerase activity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peroxidase activity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein folding</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cofactor binding</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ion binding</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 61.** Significantly enriched GO terms from DESeq2 analysis. Red cells indicate there have been expansions; blue cells indicate contractions.
<table>
<thead>
<tr>
<th>L. anguillarum</th>
<th>L. dorotheae</th>
<th>L. megistica</th>
<th>E. meliloti</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structural constituent of cytoskeleton</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anatomical structure development</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extracellular matrix structural constituent</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cofactor transmembrane transporter activity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lipid binding</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maintenance of location</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Response to stimulus</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Response to endogenous stimulus</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Response to estrogen</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Response to external stimulus</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Developmental growth</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cell growth</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cell proliferation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Immune response</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Natural killer cell mediated cytotoxicity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drug binding</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Xenobiotic transmembrane transporter activity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antigen processing and presentation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Molecular mediator of immune response</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Immune effector process</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reproductive process</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Developmental process in reproduction</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Translation repressor activity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Translation regulator activity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Methylation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Modified amino acid binding</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 62.** Significantly enriched GO terms from CAFE analysis. Red cells indicate expansions; blue cells indicate contractions; purple cells indicate both expansions and contractions.