Spring 6-5-2014

Blattabacterium Genome: Structure, Function, and Evolution

Austin Alleman

BLATTABACTERIUM GENOME: STRUCTURE, FUNCTION, AND EVOLUTION

by

AUSTIN ALLEMAN

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science
Department of Biology

Srini Kambhampati, Ph. D., Committee Chair

College of Arts and Sciences

The University of Texas at Tyler
May 2014

The University of Texas at Tyler
Tyler, Texas

This is to certify that the Master's Thesis of

AUSTIN ALLEMAN

has been approved for the thesis requirement on
April 16, 2014
for the Master degree

Approvals:

Thesis Chair: Srini Kambhampati, Ph.D.

Member: Blake Bextine, Ph.D.

Member: John S. Placyk, Jr., Ph.D.

Chair, Department of Biology

Dean, College of Arts and Sciences

Acknowledgements

I would like to thank the Sam A. Lindsey Endowment, without whose support this research would not have been possible. I would also like to thank my research advisor, Dr. Srini Kambhampati, for his knowledge, insight, and patience; and my committee, Dr. John S. Placyk, Jr. and Dr. Blake Bextine, for their assistance and feedback.

“It is the supreme art of the teacher to awaken joy in creative expression and knowledge.”

--Albert Einstein

Table of Contents

List of Tables

List of Figures

Chapter 2

Abstract

BLATTABACTERIUM GENOME: STRUCTURE, FUNCTION, AND EVOLUTION

Austin Alleman

Thesis Chair: Srinivas Kambhampati

University of Texas at Tyler
May 2014

Blattabacterium, an obligate bacterial endosymbiont, functions as a mechanism of nitrogen recycling and nutrient synthesis within the Order Blattaria (cockroaches). Through genome annotation and the application of bioinformatics, the function of *Blattabacterium* within the cockroach *Nauphoeta cinerea* was described. Results of analyses indicate that the *Blattabacterium* genome, comprised of ~620,000 base pairs and ~620 individual genes, is drastically reduced when compared *Flavobacterium*, *Blattabacterium*'s closest free-living relative. However, the *Blattabacterium* genome retained functionality vital to host survival and fecundity and functions as a source of additional nutrient biosynthesis within its host. Like other intracellular endosymbionts, the *Blattabacterium* genome has a G+C content of ~27%. Synteny within the *Blattabacterium* genome is well conserved. In addition, results of genetic drift analyses indicate that Blattabacterium is experiencing elevated rates of functional genome evolution, when compared to free-living bacterial relatives, resulting from the unique evolutionary constraints of an intracellular lifestyle.

Chapter One

GENOMIC ANALYSIS OF THE COCKROACH ENDOSYMBIONT *BLATTABACTERIUM*, AND THE ROLE OF SELECTION IN THE EVOLUTION OF THIS HOST-SYMBIONT SYSTEM: A GENERAL INTRODUCTION

INTRODUCTION

Class Insecta

Evolution, acting over the course of almost four billion years, has produced millions of unique species on planet Earth. Class Insecta itself is one of the most diverse groups of macro-organisms. Containing over one million described species - which is likely a fraction of the actual number of extant species - insects are among the most resilient, adaptable, and diverse organisms alive today. Insects are a class of invertebrate within the phylum Arthropoda, and are characterized by a chitinous exoskeleton, three-pairs of jointed legs, three main body segments, compound eyes, and antenna. Insects are present in nearly all environments, though only a handful of insects reside in the world's oceans. As a group, insects possess one of the most successful lifestyles on earth. Often, diversification of a group came as the result of differentiation of pre-existing species, though this certainly is not the only process responsible for the biodiversity in insects we see today. Interspecies relationships also promote biodiversity. Predator-prey systems, parasitism, and symbiosis are a few examples of interspecific interactions that shape the diversity of life. One of the determining factors of insect diversity and success is that they have developed an expansive range of diets. Bacterial endosymbionts harbored by insects are at least partially responsible for such nutritional flexibility within this group of organisms (Buchner, 1965). Endosymbionts are frequently observed within or near the digestive tracts of their insect host, and it is generally accepted that bacterial endosymbionts play crucial roles in the nutritional provisioning of their hosts. Today, it is known that bacterial symbionts lend a wide array of functionality to their respective insect hosts. From nutritional provisioning to defense against parasitoids, bacterial

endosymbionts contribute to much of the diversity of this already expansive and successful group of organisms.

The Cockroach

'Cockroach' is the common name for a diverse and resilient lineage of winged insects within the Order: Blattodea. Currently, the approximately 4,500 species are divided into six well-established families (Blaberidae, Blattellidae, Blattidae, Cryptoceridae, Nocticolidae, and Polyphagidae), and two relatively new families (Lamproblattidae and Tryonicidae) (Inward, et. al., 2007). In light of recent molecular and phylogenetic developments, an ongoing debate has developed as to whether or not termites, currently of the Order Isoptera, should be reclassified within Blattodea under the family Termitidae. If Blattodea were reorganized, the number of extant families within this order would be brought up to nine. Currently, some disagreement exists over whether or not Cryptoceridae is monophyletic with the rest of the cockroaches (Inward, et. al, 2007), though reclassification of Isoptera may resolve this issue.

A handful of cockroach species have adapted to live in association with humans, and as such, these species have been classified as pest organisms. These pest species are the only cockroaches that have been studied in any great detail. Those taxa occupying niches not associated with humans are, comparatively, poorly understood. The life-history, behavior, and feeding habits of these lesser-studied species remain largely unknown.

The Bacterial Endosymbiont

Up to 50% of insect species in several taxonomic orders have developed symbiotic relationships with obligate bacterial mutualists (Ruby, et. al., 2004). These primary bacterial symbionts exist within the cells of host insects, and are often required for the continued survival and reliable reproduction of the host organism (Buchner, 1965; Moran and Telang, 1998, Moran and Baumann, 2000; Zientz, et. al., 2001). Living inside of their host affords these bacteria relative safety from competition and exploitation, in exchange for increased ecological flexibility imparted onto the host species. At present, bacterial endosymbionts have been studied from a variety of insect orders. Full genomes from a number of these symbionts have been published, including many strains of *Buchnera aphidicola* (Shigenobu, et. al, 2000; etc.) aphids, *Wigglesworthia* (Akman, et. al., 2002; Rio, et. al., 2012), associated with the tsetse fly, *Blochmannia* (Gil, et. al., 2003; etc.) within carpenter ants, and *Blattabacterium* (Sabree, et. al., 2009; etc.), extracted from cockroaches. An intracellular lifestyle has profound effects upon the bacterial genome. Primary symbionts are inherited vertically from mother to offspring, undergoing sever population bottlenecks with each host generation resulting in a drastic reduction in effective population size. Accordingly, these bacterial mutualists are generally characterized by a radically reduced genome and a very low G+C nucleotide content, when compared to their free-living relatives, resulting from the accumulation of deleterious mutations through genetic drift. Since bacterial genomes contain few non-functional DNA sequences, this genome reduction has come at the cost of metabolic and physiological capabilities. As such, examining these endosymbionts outside of their respective hosts has proved to be incredibly difficult. However, through recent

developments in DNA-sequencing techniques and bioinformatics, our understanding of insect endosymbionts has improved tremendously. These mechanisms allow us to elucidate the genomic effects of an intracellular lifestyle when compared to free-living taxa.

Blattabacterium: The Cockroach Endosymbiont

With the exception of a single cave-dwelling genus, *Noticola* (Blattodea, Nocticolidae), all cockroach species contain endosymbiotic bacteria within their fat bodies (Buchner, 1965; Blochmann, 1887; Brooks, 1970; Douglas, 1989). These obligate endosymbionts belong to the genus *Blattabacterium* (Class Flavobacteria, Phylum Bacteriodetes) (Bandi et. al., 1994; Kambhampati, 2010). Analysis of evidence suggests that cockroaches acquired these endosymbionts in a single infection event, dating between 300 million years ago, the approximate age of the first fossil roaches from the Carboniferous, and 140 million years ago, when currently extant families last shared a common ancestor. Initially, the function of these endosymbionts was subject to speculation, owing to their recalcitrance to culture outside their host. Since their discovery, advanced DNA-sequencing techniques have allowed for the description and study of a number of *Blattabacterium* genomes. From these genomes it was discovered that the function of these endosymbionts is primarily the synthesis of amino acids and vitamins from the nitrogenous waste products of the host cockroach (Kambhampati, 2010). Cockroaches store excess nitrogen as uric acid within their fat body cells (Mullins and Cochran, 1975). The decaying plant matter on which cockroaches typically feed is poor in nitrogen content. Thus, a mechanism for recycling nitrogenous waste would be beneficial to any organism whose diet is nitrogen-deficient. Unlike most insects, which

excrete waste nitrogen as uric acid, cockroaches excrete ammonia instead. *Blattabacterium* are capable of utilizing both urea and ammonia because they contain an active urease as well as a functioning urea cycle that converts host urea to ammonia (Bandi et. al., 1995; Mullins and Cochran, 1976; O'Donnell, 2008). In addition, increases in dietary nitrogen intake by host cockroaches correlates with increases in uric acid buildup within that host's fat bodies (Mullins and Cochran, 1974; Mullins and Cochran, 1975).

Blattabacterium are excellent models for studying genome evolution in endosymbionts for a variety of reasons. Cockroaches represent an evolutionary lineage consisting of diverse and ancient taxa that have adapted to many habitats and exhibit broad nutritional ecology. To date, eight *Blattabacterium* genomes have been sequenced from the following cockroach host species: *Peirplaneta americana* (Sabree, et. al., 2009), *Blatta germanica* (Lopez-Sanchez, et. al., 2011), *Cryptocercus punctulatus* (Neef, et. al., 2011), *Blaberus giganteus* (Huang, et. al., 2012), *Blatta orientalis* (Patino-Navarette, et. al., 2013), *Panesthia angustipennis* (Tokuda, et. al., 2013), *Nauphoeta cinerea* (Kambhampati, et. al., 2013) and the termite, *Mastotermes darwiniensis* (Sabree, et. al., 2012). While these genomes share similar gene composition and genome architecture, each also displays unique capacities for metabolic and physiological function. Thus, while the results of phylogenetic analysis support the hypothesis of co-cladogenesis between the endosymbionts and hosts (Clark, et. al., 2001; Lo, et. al., 2003), gene composition of *Blattabacterium* is not congruent with host phylogeny; rather it varies likely as a function of host nutrition, its relative importance in the mutualism, and the interaction between selection and drift. *Blattabacterium*, like other endosymbionts that

are transmitted exclusively from mother to offspring (vertical transmission), are excellent models for examining in the interplay between random genetic drift and natural selection because they undergo a population bottleneck with each host generation, shedding light on the interesting question of the relative roles of drift and selection within evolution. Finally, estimating the ratio of synonymous and non-synonymous substitutions, a measure of negative or positive selection, is also likely to illuminate the patterns and processes of genome evolution in *Blattabacterium*. It is for these reasons that *Blattabacterium* are superb models for studying host-symbiont cospeciation as well as genome evolution within bacterial endosymbionts.

REFERENCES

Akman L, Yamashita A, Watanabe H, Oshima K, Shiba T, Hattori M, Aksoy S. (2002). Genome Sequence of the Endocellular Obligate Symbiont of Tsetse Flies, *Wigglesworthia glossinidia*. *Nature Genetics*, Vol. 32, pg. 402-407.

Aldrich BT and Kambhampati S. (2004). Microsatellite markers for two species of dampwood termite in the genus Zootermopsis (Isoptera: Termopsidae). *Mol. Ecol. Notes.* Vol 4, pg. 719-721.

Andersson SG, Kurland CG. (1998). Reductive evolution of resident genomes. *Trends Microbiol.* Vol. 6, pg. 263–268.

Bandi C, Damiani G, Magrassi L, Grigolo A, Fani R. (1994). Flavobacteria as intracellular symbionts in cockroaches. *Proc Biol Sci* Vol. 257, pg. 43–48.

Bandi C, et al. (1995). The establishment of intracellular symbiosis in an ancestor of cockroaches and termites. *Proc R Soc London Ser B.* Vol. 259, pg. 293–299.

Bell WJ, Roth LM, Nalepa CA (2007). Cockroaches: Ecology, Behavior, and Natural History. (Johns Hopkins Univ Press, Baltimore).

Bernardet JF, Bowman JP. (2010). Genus *Flavobacterium*. *Bergey's Manual of Systematic Bacteriology*. 2nd Ed., Vol. 4, pg. 112-154. Springer, New York.

Blochmann F. (1887). U¨ ber das regelma¨ ssige Vorkommen von backteriena¨ hnlichen Gebilden in den Geweben und Eiern versichiedener Insekten. *Z Biol.* 24: 6.

Brooks MA (1970) Comments on the classification of intracellular symbiotes of cockroaches and a description of the species. *J Invert Pathol*. Vol. 16, pg. 249–258.

Brynnel EU, Kurland CG, Moran NA, Andersson SG. (1998). Evolutionary rates for *tuf* genes in endosymbionts of aphids. *Mol. Biol. Evol.* Vol. 15, pg. 574–582.

Buchner P. (1965). *Endosymbiosis of Animals with Plant Microorganisms*. *New York: Interscience Publishers*.

Clark MA, Moran NA. Baumann P. (1999). Sequence evolution in bacterial endosymbionts having extreme base compositions. *Mol. Biol. Evol.* Vol. 16, pg. 1586–1598.

Clark JW, Hossain S, Burnside CA, Kambhampati S. (2001). Coevolution between a cockroach and its bacterial endosymbiont: a biogeographical perspective. *Proc Biol Sci*. Vol. 268, pg. 393–398.

Douglas AE. (1989). Mycetocyte symbiosis in insects. *Biol. Rev. Camb. Phil. Soc.* Vol. 64, pg. 409–434.

Gil R, et. al. (2003). The genome sequence of *Blochmannia floridanus*: comparative analysis of reduced genomes. *PNAS*, Vol. 100, No. 16, pg. 9388-9393.

Heddi A, Charles H, Khatchadourian C, Bonnot G, Nardon P. (1998). Molecular characterization of the principal symbiotic bacteria of the weevil *Sitophilus oryzae*: a peculiar G+C content of an endocytobiotic DNA. *J. Mol. Evol.* Vol. 47, pg. 52–61.

Heddi A, Grenier AM, Khatchadourian C, Charles H, Nardon P. (1999). Four intracellular genomes direct weevil biology: nuclear, mitochondrial, principal endosymbiont, and *Wolbachia. Proc. Natl Acad. Sci. USA* Vol. 96, pg. 6814–6819.

Huang CY, Sabree ZL, Moran NA. (2012). Genome Sequence of *Blattabacterium* sp. Strain BGIGA, Endosymbiont of the Blaberus giganteus Cockroach. *J. Bacteriol.* Vol. 194, No. 19, pg. 4450.

Inward D, Beccaloni G, Eggleton P. (2007). Death of an order: a comprehensive molecular phylogenetic study confirms that termites are eusocial cockroaches. *Biol. Lett*. (2007) 3, 331-335. Dio10.1098/rsbl.2007.0102.

Kambhampati S. (2010). Genus *Blattabacterium*. *Bergey's Manual of Systematic Bacteriology*. 2nd Ed., Vol. 4, pg. 315-321. Springer, New York.

Kambhampati S, Alleman A, Park Y. (2013). Complete Genome Sequence of the endosymbiont *Blattabacterium* from the Cockroach Nauphoeta cinerea (Blattodea: Blaberidae). *Genomics.* Vol. 102, No. 5-6, pg. 479-483.

Lawrence J, Roth J. (1999). in *Organization of the Prokaryotic Genome* (ed. Charlesbois, R.) pg. 263–289 (ASM Press, Washington, DC).

Lo N, Bandi C, Watanabe H, Nalepa C, Beninati T. (2003). Evidence for cocladogenesis between diverse dictyopteran lineages and their intracellular endosymbionts. *Mol Biol Evol.* Vol. 20, pg. 907–913.

Lopez-Sanchez MJ, Neef A, Pereto J, Patino-Navarette R, Pignatelli M, Latorre A, Moya A. (2009). Evolutionary convergence and nitrogen metabolism in *Blattabacterium* strain Bge, primary endosymbiont of the cockroach Blattella germanica. *PLoS Genet*. 5:e1000721.

Mira A, Ochman H, Moran NA. (2001). Deletional bias and the evolution of bacterial genomes. *Trends Genet.* Vol. 17, pg. 589–596.

Mira A, Moran NA. (2002). Estimating population size and transmission bottlenecks in maternally transmitted endosymbiotic bacteria. *Microb. Ecol.* Vol. 44, pg, 137–143.

Moran NA. (1996). Accelerated evolution and Muller's ratchet in endosymbiotic bacteria. *Proc. Natl Acad. Sci. USA* Vol. 93, pg. 2873–2878.

Moran NA, Telang A. (1998). Bacteriocyte-associated symbionts of insects. *Bioscience* Vol. 48, pg. 295–304.

Moran NA, Baumann P. (2000). Bacterial endosymbionts in animals. *Curr. Opin. Microbiol.* Vol. 3, pg. 270–275.

Moran NA, Wernegreen JJ. (2000). Lifestyle evolution in symbiotic bacteria: insights from genomics. *Trends Ecol. Evol.* Vol. 15, pg. 321–326.

Moran NA, Mira A. (2001). The process of genome shrinkage in the obligate symbiont *Buchnera aphidicola*. *Genome Biol.* Vol. 2.

Moran NA. (2002). Microbial minimalism: genome reduction in bacterial pathogens. *Cell* Vol. 108, pg. 583–586.

Muller HJ. (1964). The relation of recombination to mutational advance. *Mutat. Res.* Vol. 1, pg. 2–9.

Mullins DE, Cochran DG. (1974). Nitrogen metabolism in the American cockroach: An examination of whole body and fat body regulation of cations in response to nitrogen balance. *J Exp Biol.* Vol. 61, pg. 557–570.

Mullins DE, Cochran DG (1975). Nitrogen metabolism in American cockroach. I. An examination of positive nitrogen-balance with respect to uric-acid stores. *Comp Bio-chem Physiol*. Vol. 50 pg. 489 –500.

Mullins DE, Cochran DG. (1976). A comparative study of nitrogen excretion in 23 cockroach species. *Comp. Biochem. Physiol. A. Comp. Physiol.* Vol. 53 pg. 393-399.

Neef A, Latorre A, Pereto J, Silva FJ, Pignatelli M, Moya A. (2011). Genome Economization in the Endosymbiont of the Wood Roach Cryptocercus punctulatus Due to Drastic Loss of Amino Acid Synthesis Capabilities. *Genome Biol. Evol*. Vol. 3, pg. 1437–1448.

Ochman H, Moran NA. (2001). Genes lost and genes found: evolution of bacterial pathogenesis and symbiosis. *Science* Vol. 292, pg. 1096–1099.

O'Donnell M. (2008). Insect excretory mechanisms. *Advances in insect physiology*. Vol. 35, New York: Academic Press, pg. 1–122.

Ohta T. (1973). Slightly deleterious mutant substitutions in evolution. *Nature,* Vol. 246, pg. 96–98.

Patino-Navarette, R, Moya, A, Latorre, A, Pereto, J. (2013). Comparative genomics of *Blattabacterium cuenoti*: the frozen legacy of an ancient endosymbiont genome. *Genome Biol. Evol.* Vol. 5, pg. 351-361.

Rio VMR, Symula RE, Wang J. (2012). (2012). Insight into the Transmissional Biology and Species-Specific Functional Capabilities of Tsetse (Diptera: Glossinidae) Obligate Symbiont *Wigglesworthia*. *mBio*, Vol. 3, No. 1, e00240-11.

Rispe C, Moran NA. (2000). Accumulation of deleterious mutations in endosymbionts: Muller's ratchet with two levels of selection. *Am. Nat.* Vol. 156, pg. 424–441.

Ruby, E., Henderson, B., McFall-Ngai, M. (2004). We get by with a little help from our (little) friends. *Science*. Vol. 303, pg. 1305-1307.

Sabree ZL, Kambhampati S, Moran NA. (2009). Nitrogen recycling and nutritional provisioning by *Blattabacterium*, the cockroach endo-symbiont. *Proc Natl Acad Sci USA*. Vol. 106, pg. 19521–19526.

Sabree ZL, Huang CY, Arakawa G, Tokuda G, Lo N, Watanabe H, Moran NA. (2012). Genome shrinkage and loss of nutrient-providing potential in the obligate symbiont of the primitive termite Mastotermes darwiniensis. *Appl Environ Microbiol.* Vol. 78, pg. 204-210.

Shigenobu S, Watanabe H, Hattori M, Sakaki Y, Ishikawa H. (2000). Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. *Nature*, Vol. 407, pg. 81-86.

Tokuda G, Elbourne LDH, Kinjo Y, Saitoh S, Sabree Z, Hojo M, Yamada A, Hayashi Y, Shigenobu S, Bandi C, Paulsen IT, Watanabe H, Lo N. (2013). Maintenance of essential amino acid synthesis pathways in the *Blattabacterium cuenoti* symbionts of a wood-feeding cockroach. *Biol. Letters* 9, Mar. 20. doi: 10.1098/rsbl.2012.1153.

Wernegreen JJ, Moran NA. (1999). Evidence for genetic drift in endosymbionts (*Buchnera*): analyses of protein-coding genes. *Mol. Biol. Evol.* Vol. 16, pg. 83–97.

Wernegreen, J.J. (2002). Genome Evolution in Bacterial Endosymbionts of Insects. *Nat. Rev. Genet.* Vol. 3, No. 11, pg. 850-861.

Zientz E, Silva FJ, Gross R. (2001). Genome interdependence in insect–bacterium symbioses. *Genome Biol.* Vol. 2.

Chapter Two

COMPLETE GENOME SEQUENCE OF THE ENDOSYMBIONT *BLATTABACTERIUM* FROM THE COCKROACH *NAUPHOETA CINEREA* (BLATTODEA: BLABERIDAE)

Austin Alleman[1], Srinivas Kambhampati[1a], and Yonseong Park[2]

[1] Department of Biology, The University of Texas at Tyler, Tyler, TX 75799, USA
[2] Department of Entomology, Kansas State University, Manhattan, KS 66506 USA

This chapter was published as: Kambhampati S, Alleman A, Park Y. (2013). Complete Genome Sequence of the endosymbiont *Blattabacterium* from the Cockroach Nauphoeta cinerea (Blattodea: Blaberidae). *Genomics.* Vol. 102, No. 5-6, pg. 479-483.

[a] Corresponding author

INTRODUCTION

Cockroaches (Blattodea) represent an ancient and diverse lineage of winged insects (Bell, et. al., 2007). Largely tropical, a vast majority of the approximately 5,000 species inhabit forested areas and generally feed on decaying, nitrogen-poor organic matter (Bell, et. al., 2007). The handful of species that have adapted to an association with humans are the only ones that have been studied in detail, with incomplete information available for the biology of those taxa that live in the wild. Thus, the biology including life-history, feeding habits, and behavior remains unknown for many cockroach species.

Currently cockroaches are subdivided into eight families, with some debate over whether or not the monogeneric family Cryptocercidae is monophyletic with the rest of the cockroaches (Inward, et. al., 2007). The other seven families are Blattidae, Blattellidae, Polyphagidae, Blaberidae, Nocticolidae, Tryonicidae, and Lamproblattidae. Although there is no universal agreement on cockroach classification, Blaberidae is widely regarded as the most derived family (Bell, et. al., 2007; Kambhampati, 1995) with diverse biology, life history, and distribution.

All cockroaches, with the exception of *Noticola,* the cave dwelling genus, harbor within their fat bodies endosymbiotic bacteria (Buchner, 1965; Blochmann, 1887; Brooks, 1970) of the genus *Blattabacterium,* belonging to the class Flavobacteria and the phylum Bacteroidetes (Bandi, et. al., 1994; Kambhampati, 2010). The relationship between the host and the endosymbiont is an obligate one. Before the recent sequencing of the *Blattabacterium* genomes (see below), their function was subject to speculation

(Kambhampati, 2010); however, now it is clear that the endosymbionts are involved in amino acid and vitamin synthesis from nitrogenous waste products (Kambhampati 2010). Within cockroaches, surplus nitrogen, concentrated into uric acid for storage, is contained within fat body cells (Mullins and Cochran, 1975). Cockroaches excrete ammonia as the nitrogen waste product, unlike most insects, which excrete waste nitrogen as uric acid. It is likely that *Blattabacterium* are capable of utilizing both urea and ammonia because they contain an active urease as well as a urea cycle that converts host urea to ammonia (Bandi, 1995; O'Donnell, 2008). In addition, increase in dietary nitrogen consumption by the host cockroach correlates with increased uric acid buildup within fat bodies (Mullins and Cochran, 1974; Mullins and Cochran, 1975).

Eight *Blattabacterium* genomes have been sequenced to date from the following hosts: *Periplaneta americana* [BPLAN; (Sabree, et. al., 2009)], *Blatta germanica* [BBge; (Lopez-Sanchez, et. al., 2009)], *Cryptocercus punctulatus* [BCpu; (Neef, et. al., 2011)], *Blaberus giganteus* [BGIGA; (Huang, et. al., 2012)], *Blatta orientalis* [BBor; (Patino-Navarette, et. al., 2013)], *Panesthia angustipennis* [Pane (Tokuda, et. al., 2013)], *Nauphoeta cinerea* [BNCIN (Kambhampati, et. al., 2013)] and the termite, *Mastotermes darwiniensis* [MADAR; (Sabree, et. al., 2012)]. These genomes, while largely similar to one another, also exhibit differences in structure and function, indicating considerable independent evolution among lineages harbored by various host species (see Results and Discussion for details).

Blattabacterium are excellent models for studying genome evolution in endosymbionts for a number of reasons. Cockroaches represent an evolutionary lineage consisting of diverse and ancient taxa that have adapted to many habitats and exhibit

broad nutritional ecology. While phylogenetic analysis supports co-cladogenesis between cockroaches and *Blattabacterium* (Clark, et. al., 2001; Lo, et. al., 2003), the gene composition among *Blattabacterium* harbored by different host species varies suggesting selection that may be correlated with host nutritional ecology. On the other hand, *Blattabacterium*, like other vertically transmitted endosymbionts, are an interesting case study in the interplay between natural selection and genetic drift because they undergo a bottleneck with each host generation. The process by which *Blattabacterium* and other insect endosymbionts are passed on into the next generation of the host is presumably random and allows for only a subset of the bacteria to be passed on from the mother to her offspring (Wernegreen, 2011). Thus, both natural selection and random genetic drift likely play a role in genome evolution of *Blattabacterium* as was demonstrated for *Buchnera* (Herbeck, et. al., 2003) and *Blochmannia* (Wernegreen, 2011).

Here, the full genome sequence of *Blattabacterium* from the host cockroach, *Nauphoeta cinerea* (BNCIN), belonging to the family Blaberidae, is reported. *Nauphoeta cinerea* has been the subject of many behavioral and sexual selection studies in laboratory settings (e.g., Kou, et. al., 2009, Barrett, et. al., 2009); however, little is known about its natural history. The genome structure and function of *Blattabacterium* from *N. cinerea* to those from other *Blattabacterium* and elucidate the similarities and differences among them were compared.

MATERIALS AND METHODS

Insects

N. cinerea were obtained from a colony maintained at the Insect Zoo, Kansas State University.

DNA Isolation

DNA from host cockroaches was isolated as described in (Sabree, et. al., 2009). Briefly, freshly frozen cockroaches were dissected under a microscope to collect fat body tissue. Fat bodies were collected from 5 cockroaches directly into 200 µl of buffer ATL in the DNAeasy kit (Qiagen). The tissue was homogenized with a hand-held homogenizer and incubated at 55 ^{o}C for one hour after the addition of 20 µl of 50 µg/ml Proteinase-K. The homogenate was filtered through a 20 µm syringe filter (Millipore). Further steps followed the protocol recommended by the manufacturer for isolating DNA from tissue.

The amount of isolated DNA was estimated on a NanoDrop 2000. To estimate the relative amounts of host and bacterial DNA, PCR was set up using two genes: a 660 bp portion of the *Blattabacterium* 16S rRNA gene using primers from Clark and Kambhampati (Clark, et. al., 2001) and a 500 bp portion of the insect 18S rRNA gene using primers from Kambhampati and Aldrich (Kambhampati and Aldrich, 2006). The PCR protocol was: 95 ^{o}C for 3 minutes, followed by 35 cycles of 95 ^{o}C for 30 s, 50 ^{o}C for 45 s, and 72 ^{o}C for 45 s; the final extension step was 72 ^{o}C for 10 min. The PCR products were run on a 1% agarose gel stained with ethidium bromide to visualize the DNA fragments. The relative amounts of host and bacterial DNA were estimated qualitatively.

DNA Sequencing

The DNA was sent to the University of South Carolina Genomics Facility for sequencing using pyrosequencing (454 Life Sciences). The raw sequences were trimmed and assembled by Newbler software (454 Life Sciences); the assembled sequences and quality scores were electronically sent from the facility.

Genome Assembly

Trimming of the primers, adapters and polyA tails from the start or end of the sequences was conducted at the sequencing facility. Contigs were assembled using the default parameters in Newbler (454 Life Sciences). All contigs assembled by Newbler were submitted to the NCBI database using blastn and the search was restricted to *Blattabacterium*. Contigs with significant hits (e values of $\leq e^{-10}$) were identified. Each individual contig was then aligned using CLUSTAL (Larkin, 2007) to the full genome of *Blattabacterium* from the German cockroach, *B. germanica* (Lopez-Sanchez, et. al., 2009); GenBank Accession no. NC_013454.1, GI: 262340793) to determine the contig's position within the genome. The contigs were reverse complemented if necessary and placed in the order based on their position relative to the BBge genome.

Gene Annotation

Once the entire genome was assembled relative to the reference sequence, the sequence was submitted to GeneMark.hmm (Lukashin and Borodovsky, 1998) for prokaryotes to identify putative ORFs. "*Blattabacterium* sp.", the only *Blattabacterium* genome available in GeneMark at the time of the analysis, was selected as the reference sequence. The putative ORFs were subjected to a blastp search to determine the extent of

coverage of known BBge genes. Those predicted ORFs that consisted of only a portion of a gene (relative to the reference genome) were manually curated to identify the missing section of that gene (i.e., we manually searched for and annotated those genes that were not completely assembled by ORF prediction software). Using a combination of manual curation and blastp searches, we assembled a predicted set of genes contained within the genome along with their physical location on the chromosome. Transfer RNA (tRNA) and tmRNA genes were identified using tRNAScan (Lowe and Eddy, 1997).

Gene Ontology

To identify the putative function of each predicted gene, the curated set of genes was submitted to KEGG (Kanehisa, et. al., 2012) using BBge as the reference sequence. Each identified pathway was examined for completeness and the synthesis or lack thereof of each amino acid, vitamin, and other compounds was inferred.

Comparative Genomics

The assembled and annotated genome was compared to the six published *Blattabacterium* genomes using primarily Geneious (Drummond, et. al., 2004), Mauve (Darling, et. al., 2004), and NCBI's blast. Geneious was used to align the sequences and identify the indels. Mauve was used to examine the synteny between pairs of genomes. COGs for all genomes were identified using BASys, the Bacterial Annotation System (Van Domselaar, et. al., 2005). The genome sequence has been deposited in GenBank under accession numbers NC_013454.1 and GI: 262340793.

RESULTS AND DISCUSSION

One 454 sequencing run resulted in approximately 300 Mb of DNA sequence data. The quality of the reads was excellent, with a vast majority of the reads being assigned the highest quality scores. Newbler assembled the raw reads into 56,734 contigs of varying lengths (ranging from about 100 bp to greater than 250 Kb in length). Of the 56,734 contigs, 10 contigs belonged to *Blattabacterium* and the rest, presumably, to the host. The 10 contigs, made up of 155,529 reads, contained the entire genome of *Blattabacterium*. With an average read length of about 450 bp, the 155,529 reads totaled nearly 70 Mb of sequence data or an average *Blattabacterium* genome coverage of 112-fold.

The *Blattabacterium* genome from *N. cinerea* is 623,002 bp in length and has much in common with the other *Blattabacterium* genomes that have been sequenced to-date (Table 1). All *Blattabacterium* genomes are highly reduced compared to those of their free-living relatives (Lopez-Sanchez, et. al., 2008; Sabree, et. al., 2010). The BNCIN genome is about 98.5% identical in nucleotide sequence to that from *B. giganteus* (host family: Blaberidae). In addition, the endosymbiont included a 3,735 bp plasmid that contained four genes (ribonucleoside-diphosphate reductase subunit beta, two hypothetical proteins, and deoxyuridine 5'-triphosphate nucleotidohydrolase) and was therefore similar in gene composition to plasmids from the other *Blattabacterium* genomes. The base composition was AT-biased with a GC content of 26.2%, similar to other described *Blattabacterium* (Patino-Navarette, et. al., 2013). The average length of the open reading frames (ORF) was 1,005 bp.

All existing *Blattabacterium* genomes were aligned in Geneious to examine synteny. To a large extent, the genomes were similar in structure. As noted by others (Neef, et. al., 2011; Patino-Navarette, et. al., 2013; Sabree, et. al., 2012), BCpu and MADAR are most divergent in gene composition relative to the remainder of the sequenced *Blattabacterium* genomes. The BNCIN genome, not surprisingly, was most similar in structure to that from *B. giganteus*. However, relative to BBge, BNCIN exhibited 9 major deletions. The gaps between BNCIN and BBge because the indels cumulatively spanned a few thousand base pairs and encompassed several genes. These gaps were confirmed using PCR amplification using primers anchored in the flanking regions (primer sequences available on request). Each of the gaps included one or more genes present in BBge but not in BNCIN. The genes deleted in BNCIN relative to BBge are: CAAX amino acid terminal protease family protein, 4Fe-4S ferredoxin iron-sulfur binding protein, uracil DNA glycosylase, phosphoadenylyl-sulfate reductase, sulfate adenylyltransferase subunit 1, cysteine synthase A, Uroporphyrinogen-III C-methyltransferase, sulfite reductase (NADPH) hemoprotein subunits alpha and beta, hydroxymethylbilane synthase, K+ uptake transporter subunit KtrA, Uroporphyrinogen III synthase, hypothetical protein BLBBGE_594 and 595, and Glycoprotease M22 family domain containing protein. Many of the above deleted genes are involved in sulfur metabolism (Sekowska, et. al., 2000). Therefore, BNCIN is incapable of completely synthesizing methionine and retains the ability to synthesize cysteine, the two sulfur-containing amino acids. A comparison of gene composition among all genomes with the exception of BNCIN was provided in (Patino-Navarette, et. al., 2013).

Gene Composition

A total of 582 ORFs in the BNCIN genome, plus four more ORFs on BNCIN's plasmid were identified. These numbers are similar to ORF counts from other *Blattabacterium* genomes sequenced to-date (Table 1). Almost all the genes essential for DNA replication, RNA transcription, and mRNA translation machinery were inferred. Thirty-two tRNA genes, capable of transferring all amino acids and a single transfer-messenger RNA (tmRNA) gene were identified. Three ribosomal RNA genes (rRNA) were also present.

Table 1: Genome characteristics of BNCIN compared to all seven other published *Blattabacterium* species. Host species abbreviations are as follows: BNCIN*, N. cinerea*; BGIGA, *B. giganteus;* BBge, *B. germanica*; BPLAN, *P. americana; ;* BCpu, *C. punctulatus;* MADAR, *M. darwiniensis,* BBor, *B. orientalis*.

	BNCIN	BGIGA	BBge	BPLAN	BCpu	MADAR	BBor	BPane
Genome size (bp)	623,002	632,588	640,935	640,442	609,561	590,336	638,184	632,490
Plasmids	1	1	1	1	1	1	1	0
Plasmid size (bp)	3,735	3,423	4,085	3,448	3,816	3,088	3,735	N/A
Chromosome size (bp)	619,267	629,165	636,850	636,994	605,745	587,248	634,449	632,490
G+C content (%)	26.2 / 20.6	25.7/ 30.9	27.1 / 29.8	28.2 / 28.5	23.8 / 30.5	27.5/ 31.9	28.2/ 30.6	26.4/ N/A
Total number of genes	627	616	631	634	589	597	627	615
CDSs	582 + 4	573 + 4	586 + 4	587 + 4	545 + 3	544 + 4	572 + 7	575
rRNAs	3	3	3	3	3	3	3	3
tRNAs	32	34	34	33	32	34	33	34
Other RNAs: tmRNA, ffs, rnpB	1	1	3	1	3	3	3	3
Pseudogenes	5	1	1	6	3	9	9	5

The COG composition for all sequenced *Blattabacterium* genomes is shown in Figure 1. All the *Blattabacterium* genomes sequenced to-date appear to be highly similar across the COG categories. One possible exception is COG category E (amino acid transport and metabolism), in which both BCpu and MADAR seem to have fewer genes

than the other *Blattabacterium*. Not surprisingly, BNCIN, like all other strains of *Blattabacterium,* has lost virtually all genes coding for cell motility, a trait not needed within the controlled environment of an insect host. COGs responsible for secondary metabolite biosynthesis, transport, and catabolism, as well as signal-transduction mechanisms are also noticeably absent within the genome.

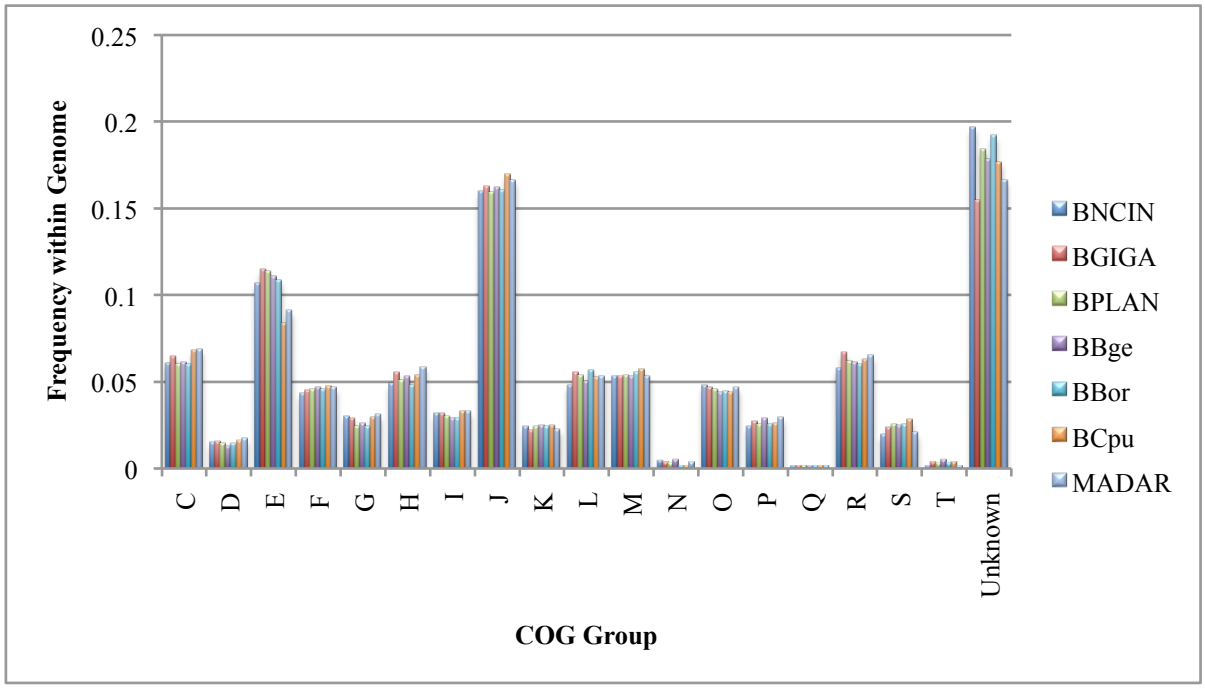

Figure 1: COG composition of BNCIN genome. Letters refer to COG functional categories as follows. C - Energy production and conversion; D - Cell division and chromosome partitioning; E - Amino acid transport and metabolism; F - Nucleotide transport and metabolism; G - Carbohydrate transport and metabolism; H - Coenzyme metabolism; I - Lipid metabolism; J - Translation, ribosomal structure and biogenesis; K - Transcription; L - DNA replication, recombination and repair; M - Cell envelope biogenesis, outer membrane; N – Cell motility; O - Posttranslational modification, protein turnover, chaperones; P - Inorganic ion transport and metabolism; Q - Secondary metabolites biosynthesis, transport, and catabolism; R - General function prediction only; S - COG of unknown function; T - Signal transduction mechanisms. Host species abbreviations are as follows: BNCIN, *N. cinerea*; BGIGA, *B. giganteus;* BBge, *B. germanica*; BPLAN, *P. americana;* BCpu, *C. punctulatus;* MADAR, *M. darwiniensis;* BBor, *B. orientalis*.

Amino Acid Biosynthesis

The putative metabolic pathways using KEGG's automatic annotation server, KASS, which revealed that BNCIN has the required genes for the biosynthesis of most, but not all, amino acids was reconstructed. Specifically, the BNCIN genome is lacking a full complement of genes for the stand-alone synthesis of methionine, asparagine, and glutamine (Table 2).

Table 2: A direct comparison between the eight sequenced *Blattabacterium* genomes with respect to their individual abilities to synthesize the 10 essential and 10 nonessential amino acids. S represents a complete metabolic pathway, and thus the ability to synthesize the given amino acid without intervention from the host organism; - represents an incomplete metabolic pathway where that specific amino acid cannot be produced by the endosymbiont; I represents a questionably incomplete metabolic pathway where some (usually terminal step) enzyme-coding genes are missing from the genome, but the endosymbiont is still able to produce that amino acid. Host species abbreviations are as follows: BNCIN*, N. cinerea*; BGIGA, *B. giganteus;* BBge, *B. germanica*; BPLAN, *P. americana;* BCpu, *C. punctulatus;* MADAR, *M. darwiniensis;* BBor, *B. orientalis*.

Essential	BNCIN	BGIGA	BBge	BPLAN	BBor	MADAR	BCpu	BPane
Histidine	S	S	S	S	S	S	S	S
Isoleucine	S	S	S	S	S	-	-	S
Leucine	S	S	S	S	S	-	-	S
Lysine	S	S	S	S	S	S	I	S
Methionine	-	-	S	-	-	-	-	-
Phenylalanine	S	S	S	S	S	S	S	S
Threonine	S	S	S	S	S	-	-	S
Tryptophan	S	S	S	S	S	-	-	S
Tyrosine	S	S	S	S	S	S	S	S
Valine	S	S	S	S	S	-	-	S
Nonessential								
Alanine	I	S	S	S	S	S	S	S
Arginine	S	S	S	S	S	S	I	S
Asparagine	-	-	-	-	-	-	S	-
Aspartate	S	S	S	S	S	S	S	S
Cysteine	S	S	S	S	S	S	-	S
Glutamate	S	S	S	S	S	S	S	S
Glutamine	-	-	-	-	-	-	S	-
Glycine	I	S	S	S	S	S	S	S
Proline	I	-	-	S	-	S	S	-
Serine	I	S	S	-	S	S	S	S

In addition, BNCIN possesses incomplete pathways (incomplete being defined as missing the terminal step for the synthesis of a given amino acid) for alanine, glycine, proline, and serine. By comparing the amino acid synthetic pathways of BNCIN with the six other described *Blattabacterium* genomes (Sabree et. al., 2009; Lopez-Sanchez, et. al., 2009; Neef, et. al., 2011; Huang, et. al., 2012; Patino-Navarette, et. al., 2013; Sabree, et. al., 2012), it is evident that none of the *Blattabacterium* species possess a full complement of genes for all of the amino acids. BNCIN's pathways for the metabolism of sulfur have been reduced when compared to the other *Blattabacterium,* although it remains more complete than the sulfur pathway possessed by BCpu, which has lost all genes needed for sulfur metabolism except for the gene encoding sulfite reductase (NADPH) flavoprotein alpha-component (Neef, et. al., 2011), which has been described to have additional functions such as electron relay (Siegel and Davis, 1974) and thus may have been retained within BCpu. BNCIN has lost the ability to reduce intracellular sulfate to sulfite. When compared to BBge and MADAR, BNCIN has lost the genes required for the coding of 3'-phosphoadenosine 5'-phosphosulfate synthase and phosphoadenosine phosphosulfate reductase, although both BBge and MADAR lack the gene necessary for the conversion of adenosine-5'-phosphosulfate (APS) into 3'-phospho adenosine-5'-phosphosulfate (PAPS). Not surprisingly, BNCIN lacks the gene for this step as well. However, the BNCIN genome does encode sulfite reductase (NADPH) flavoprotein alpha-component, cysJ, the function of which is to facilitate the reduction of sulfate into hydrogen sulfide (Siegel and Davis, 1974). This sulfide is then incorporated into the two sulfur-containing amino acids L-cysteine and L-methionine. This indicates that BNCIN receives sulfate from an outside source for the production of cysteine and methionine. In

BNCIN, cysteine is synthesized from L-serine via enzymes serine O-acetyltransferase and cysteine synthase A.

Like other sequenced *Blattabacterium* strains, BNCIN does not generate methionine by cysteine transsulfuration, as was previously suggested by experimental evidence (Block and Henry, 1961; Henry and Block, 1961). Genomic analysis revealed that this pathway is incomplete, lacking genes coding for homoserine O-acetyltransferase, metX, and cystathionine gamma-synthase, metB. Normally, homoserine O-acetyltransferase yields O-acetyl-L-homoserine, although in this alternate pathway, the need for acetlyhomoserine is bypassed by the production of homocysteine by cystathionine beta-lyase using O-phosphohomoserine and cysteine.

While *Blattabacterium* can make arginine from glutamate, this endosymbiont is lacking the final step for the production of proline on its own, suggesting an incomplete ornithine-urea cycle (Inokuchi, et. al., 1969; Reddy and Campbell, 1977). There are two similar, but separate, pathways for the production of proline from the urea cycle, and genomic analysis indicated that coding for the final enzyme of both of these pathways is missing in the BNCIN genome. While the *Blattabacterium* genome does code for the ornithine-urea cycle it does not possess the gene coding for ornithine cyclodeaminase, the enzyme responsible for the conversion of ornithine into proline. Therefore, *Blattabacterium* must either acquire ornithine cyclodeaminase from its host in order to produce proline from ornithine, or it must acquire proline in its entirety from the host. Urea and ammonia are both products of proline biosynthesis through arginine degradation. These products can be used by BNCIN-encoded glutamate dehydrogenase and urease to produce glutamate. This method of proline biosynthesis does conserve

nitrogen; thus, BNCIN, like other *Blattabacterium*, produces amino acids through the recycling of nitrogenous wastes.

Metabolic Pathways

BNCIN has a slightly reduced capacity for DNA repair compared to BBge and BPLAN, lacking the ability to produce uracil-DNA glycosylase, which eliminates uracil from DNA molecules in order to prevent mutagenesis (Lindhal, et. al., 1977). BNCIN also lacks coding for ATP-dependent DNA helicase (PcrA). This reduction in repair function, however, is not as severe as that displayed by MADAR, which possesses even fewer genes required for DNA repair (Sabree, et. al., 2012).

Few substrate-specific transporters, especially those involved in amino acid uptake or secretion, were identifiable in the BNCIN genome. This is consistent with other insect nutritional endosymbionts. Like other *Blattabacterium* species, BNCIN encodes the alternate sigma factor RpoN (RNA polymerase sigma-54 factor), which functions as a transcriptional regulator of genes involved in nitrogen assimilation. While RpoN is encoded by all *Blattabacterium* sequenced to-date, it is absent in other insect sequenced endosymbiont genomes (Sabree, et. al., 2009).

Genomic analysis of BNCIN also indicated the presence of a shortened glycolysis pathway. The genes coding for phosphofructokinase (PFK) (a metabolite needed at the beginning of the pathway to convert beta-D-fructose-6-phosphate into beta-D-fructose-1,6 biphosphate), and pyruvate kinase (PK) (required for the conversion of phosphoenolpyruvate into pyruvate) are both missing from the genome. Like other

Blattabacterium, BNCIN likely produce pyruvate via NADP+-dependent malate dehydrogenase (Lopez-Sanchez, et. al., 2009).

CONCLUSIONS

Blattabacterium within *N. cinerea* is a nutritional endosymbiont that is genetically and functionally similar to *Blattabacterium* harbored by other cockroach species, performing tasks such as nitrogen metabolism entailing the recycling of nitrogen from ammonia and urea as well as the provisioning of amino acids. *Blattabacterium* are unique among insect nutritional endosymbionts in that they have retained the transcriptional regulator RpoN, and in this respect BNCIN is no different from other *Blattabacterium*. BNCIN, like other *Blattabacterium* and insect symbionts, has a drastically reduced genome, which lacks almost all specific transporters and regulatory genes, an indication of the highly specialized and dependent nature of insect endosymbionts. The metabolic capacities provided to the cockroaches by *Blattabacterium* played an important role in the expansion of ecological niches for cockroaches allowing for the exploitation of nitrogen-poor or nitrogen-variable food sources. Like other cockroach endosymbionts, BNCIN lacks a uricase homologue, which is integral to the utilization of uric acid as a way to store nitrogen. This step, then, is not carried out by *Blattabacterium*, but either by the cockroach host itself, or by bacteria within the insect's gut (Sabree, Kambhampati, and Moran, 2009), as is the case with termites (Tokuda and Watanabe, 2007).

REFERENCES

Bandi C, Damiani G, Magrassi L, Grigolo A, Fani R, et al. (1994). Flavobacteria as intracellular symbionts in cockroaches. *Proc Biol Sci* Vol. 257, pg. 43–48.

Bandi C, et al. (1995). The establishment of intracellular symbiosis in an ancestor of cockroaches and termites. *Proc R Soc London Ser B.* Vol. 259, pg. 293–299.

Barrett EL, et. al. (2009). Separate and combined effects of nutrition during juvenile and sexual development on female life-history trajectories: the thrifty phenotype in a cockroach. *Proc Biol Sci.* Vol. 22, No. 276(1671), pg. 3257-3264.

Bell WJ, Roth LM, Nalepa CA (2007). Cockroaches: Ecology, Behavior, and Natural History. (Johns Hopkins Univ Press, Baltimore).

Blochmann F. (1887). U¨ ber das regelma¨ ssige Vorkommen von backteriena¨ hnlichen Gebilden in den Geweben und Eiern versichiedener Insekten. *Z Biol.* 24: 6.

Block RJ, Henry SM. (1961). Metabolism of sulphur amino-acids and of sulphate in *Blattella germanica*. *Nature*. Vol. 191, pg. 392–393.

Brooks MA (1970) Comments on the classification of intracellular symbiotes of cockroaches and a description of the species. *J Invert Pathol*. Vol. 16, pg. 249–258.

Buchner P. (1965). *Endosymbiosis of Animals with Plant Microorganisms*. *New York: Interscience Publishers*.

Clark JW, Hossain S, Burnside CA, Kambhampati S. (2001). Coevolution between a cockroach and its bacterial endosymbiont: a biogeographical perspective. *Proc Biol Sci*. Vol. 268, pg. 393–398.

Darling ACE, Mau B, Blatter FB, and Perna NT. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Research*. Vol. 14, No. 7, pg. 1394-1403.

Drummond AJ, Ashton B, Buxton S, Cheung M, Cooper A, Duran C, Field M, Heled J, Kearse M, Markowitz S, Moir R, Stones-Havas S, Sturrock S, Thierer T, Wilson A. (2011). Geneious v5.4, Available from http://www.geneious.com/

Henry SM, Block RJ (1961). Sulfur metabolism of insects. 6. Metabolism of sulfur amino acids and related compounds in German cockroach, Blattella germanica (L.). *Contrib Boyce Thompson Inst*. Vol. 21, pg. 129 –145.

Herbeck TH, Funk DJ, Degnan PH, Wernegreen JJ. (2003). A conservative test of Genetic Drift in the Endosymbiotic Bacterium Buchnera: Slightly Deleterious Mutations in the Chaperonin *groEL. Genetics*. Vol. 165, pg. 1651-1660.

Huang CY, Sabree ZL, Moran NA. (2012). Genome sequence of *Blattabacterium* sp. strain BGIGA, endosymbiont of the *Blaberus giganteus* cockroach. *J. Bacteriol.* Vol. 194, No. 19, pg. 4450.

Inokuchi T, Horie Y, Ito T. (1969). Urea cycle in silkworm Bombyx mori. *Biochem Biophys Res Commun.* Vol. 35, pg. 783–787.

Inward D, Beccaloni G, Eggleton P. (2007). "Death of an order: a comprehensive molecular phylogenetic study confirms that termites are eusocial cockroaches. *Biol. Lett*. (2007) 3, 331-335. Dio10.1098/rsbl.2007.0102.

Kambhampati, S. (1995). Phylogeny of cockroaches and related insects based on DNA sequence of mitochondrial ribosomal RNA genes. *Proc. Natl. Acad. Sci. USA*. Vol. 92, pg. 2017-2020.

Kambhampati S, Aldrich BT. (2006). Inexpensive molecular weight markers. *J. Entomol. Sci.* Vol 41, pg. 412-415.

Kambhampati S. (2010). Genus Blattabacterium. *Bergey's Manual of Systematic Bacteriology*. 2nd Ed., Vol. 4, pg. 315-321. Springer, New York.

Kambhampati S, Alleman A, Park Y. (2013). Complete Genome Sequence of the endosymbiont *Blattabacterium* from the Cockroach Nauphoeta cinerea (Blattodea: Blaberidae). *Genomics.* Vol. 102, No. 5-6, pg. 479-483.

Kanehisa M, Goto S, Sato Y, Furumichi M, and Tanabe M. (2012). KEGG for integration and interpretation of large-scale molecular datasets. *Nucleic Acids Res*. Vol. 40, D109-D114.

Kou R, et. al. (2009). Juvenile hormone and the ontogeny of cockroach aggression. *Horm Behav*. Vol. 56, No. 3, pg. 332-338.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*. Vol. 23, pg. 2947-2948.

Lindahl T, Ljungquist S, Siegert W, Nyberg B, Sperens B (1977). DNA N-glycosidases: properties of uracil-DNA glycosidase from Escherichia coli. *J. Biol. Chem*. Vol. 252 No. 10, pg. 3286–3294.

Lo N, Bandi C, Watanabe H, Nalepa C, Beninati T. (2003). Evidence for cocladogenesis between diverse dictyopteran lineages and their intracellular endosymbionts. *Mol Biol Evol.* Vol. 20, pg. 907–913.

Lopez-Sanchez MJ, et al. (2008). *Blattabacteria*, the endosymbionts of cockroaches, have small genome sizes and high genome copy numbers. *Environ Microbiol*. Vol. 10, pg. 3417–3422.

Lopez-Sanchez MJ, et al. (2009). Evolutionary convergence and nitrogen metabolism in Blattabacterium strain Bge, primary endosymbiont of the cockroach Blattella germanica. *PLoS Genet*. 5:e1000721.

Lowe, TM and Eddy, SR (1997). tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res*. Vol. 25, pg. 955-964.

Lukashin A, and Borodovsky M. (1998). GeneMark.hmm: new solutions for gene finding. *NAR*. Vol. 26, No. 4, pg. 1107-1115.

Mullins DE, Cochran DG. (1974). Nitrogen metabolism in the American cockroach: An examination of whole body and fat body regulation of cations in response to nitrogen balance. *J Exp Biol.* Vol. 61, pg. 557–570.

Mullins DE, Cochran DG (1975). Nitrogen metabolism in American cockroach. I. An examination of positive nitrogen-balance with respect to uric-acid stores. *Comp Biochem Physiol*. Vol. 50 pg. 489 –500.

Neef A, Latorre A, Pereto J, Silva FJ, Pignatelli M, Moya A. (2011). Genome economization in the endosymbiont of the wood roach *Cryptocercus punctulatus* due to drastic loss of amino acid synthesis capabilities. *Genome Biol. Evol*. Vol. 3, pg. 1437–1448.

O'Donnell M. (2008). Insect excretory mechanisms. *Advances in insect physiology*. Vol. 35, New York: Academic Press, pg. 1–122.

Patino-Navarette, R, Moya, A, Latorre, A, Pereto, J. (2013). Comparative genomics of *Blattabacterium cuenoti*: the frozen legacy of an ancient endosymbiont genome. *Genome Biol. Evol.* Vol. 5, pg. 351-361.

Reddy SR, Campbell JW. (1977). Enzymic basis for the nutritional requirement of arginine in insects. *Experientia*. Vol. 33, pg. 160 –161.

Sabree ZL, Kambhampati S, Moran NA. (2009). Nitrogen recycling and nutritional provisioning by Blattabacterium, the cockroach endo-symbiont. *Proc Natl Acad Sci USA*. Vol. 106, pg. 19521–19526.

Sabree ZL, Degnan PH, Moran NA. (2010). Chromosome stability and gene loss in cockroach endosymbionts. *Appl Environ Microbiol*. Vol. 76, pg. 4076–4079.

Sabree ZL, Huang CY, Arakawa G, Tokuda G, Lo N, Watanabe H, Moran NA. (2012). Genome shrinkage and loss of nutrient-providing potential in the obligate symbiont of the primitive termite *Mastotermes darwiniensis*. *Appl Environ Microbiol.* Vol. 78, pg. 204-210.

Sekowska A, Kung HF, Danchin A. (2000). "Sulfur metabolism in *Escherchia coli* and related bacteria: facts and fiction. *J. Mol. Miocbiol. Biotecnol*. Vol 2, pg. 145-177.

Siegel LM, Davis PS. (1974). Reduced nicotinamide adenine dinucleotide phosphate-sulfite reductase of enterobacteria. IV. The Escherichia coli hemoflavoprotein: subunit structure and dissociation into hemoprotein and flavoprotein components. *J Biol Chem*. Vol. 249, pg. 1587–1598.

Tokuda G, Watanabe H. (2007). Hidden cellulases in termites: revision of an old hypothesis. *Biol. Lett*. Vol. 3, No. 3, pg. 336-339.

Tokuda G, Elbourne LDH, Kinjo Y, Saitoh S, Sabree Z, Hojo M, Yamada A, Hayashi Y, Shigenobu S, Bandi C, Paulsen IT, Watanabe H, Lo N. (2013). Maintenance of Essential Amino Acid Synthesis Pathways in the *Blattabacterium* cuenoti Symbiont of a Wood-Feeding Cockroach. *Biol. Lett.* Vol. 9, No. 3, 20121153.

Van Domselaar GH, Stothard P, Shrivastava S, Cruz JA, Guo A, Dong X, Lu P, Szafron D, Greiner R, Wishart DS. (2005). BASys: a web server for automated bacterial genome annotation. *Nucleic Acids Res*. Vol. 33(Web Server issue): W455 – 459.

Wernegreen JJ. (2011). Reduced selective constraint in endosymbionts: elevation in radical amino acid replacements occurs genome-wide. *PLoS One*. Vol. 6, No. 12, e28905. doi:10.1371/journal.pone.0028905.

Chapter Three

SELECTION VS. RANDOM GENETIC DRIFT: THE SHAPING OF A *BLATTABACTERIUM* GENOME

INTRODUCTION

Comprised of over 1,200,000 species, Class: *Insecta* is regarded as the largest group of macroorganisms. Within this expansive group, at least half of extant genera are estimated to harbor bacterial endosymbionts (Buchner, 1965; Ishikawa, 2003). While some intracellular bacteria can be harmful or even lethal to their insect host, many others play an important role in host survival and fecundity (Margulis and Fester, 1991; Ruby, et. al., 2004; Dasch, et. al., 1984; Douglas, 1989; Moran and Telang, 1998; Moran and Baumann, 2000). These obligate bacterial mutualists often function in the provisioning, recycling, or degradation of essential nutrients, and are vital to those insect species that depend entirely upon nutritionally narrow diets, for example those diets composed primarily of woody material, plant sap, blood, or decaying organic material (Bourtzis and Miller, 2003; Douglas, 1998; Moran and Baumann 2000).

Insect Endosymbionts

Through bioinformatics and molecular evolutionary analysis of endosymbiont genomes, it has demonstrated that the intracellular lifestyle strongly influences the rates and patterns of genome evolution. Within bacterial endosymbionts, genome evolution is characterized by a lack of genetic recombination through random genetic drift resulting from the fixation of deleterious mutations, bacterial asexuality, and elevated mutation rates and biases (Brynnel, et. al., 1998; Clark, et. al., 1999; Wernegreen and Moran, 1999). The effects of asexuality, random drift, and irreparable mutation are often interconnected and mutually reinforcing. Primary endosymbionts are passed on through host lineages via strict vertical transmission, from mother directly to offspring, as

evidenced by co-cladogenesis between host insects and their endosymbionts (e.g., Chen, et. al., 1999; Clark, et. al., 2000; Funk, et. al., 2000; Patino-Navarette, et. al., 2013; Lo, et. al., 2003). This method of transmission, resulting in successive bottlenecks throughout the evolution of the obligate endosymbiont, has a profound impact on endosymbiont population structure, severely limiting their effective population size (Mira and Moran, 2002). Compared to the population sizes of free-living bacteria species, population sizes for endosymbionts, which are determined almost entirely by individual hosts, are much smaller (Mira and Moran, 2002). Under the nearly-neutral model of evolution, a reduced effective population size elevates the rate of fixation of deleterious mutations through genetic drift (Ohta, 1973). Over evolutionary time, these deleterious mutations may accumulate up to a point where they have a negative effect upon the fitness of the bacterial endosymbiont, and as a result this, will negatively affect the host as well (Rispe and Moran, 2000).

Selection, as it is used here, may be defined simply a change in nucleotide frequency within the DNA sequence of a genome. The magnitude and direction of natural selection acting upon functional protein coding genes within a genome may be inferred by the ratio of non-synonymous substitutions per non-synonymous site (d_N) to the number of synonymous substitutions per synonymous site (d_S) (McDonald and Kreitman, 1991). A d_N/d_S ratio greater than one indicates a trend towards positive mutation events within the examined sequence; less than one implies a bias towards negative or stabilizing selection events. A ratio close to or equal to one indicates no (or neutral) selection. When compared to free-living bacteria, those adapted to intracellular lifestyles exhibit increased levels of mutation at synonymous and non-synonymous sites, as well as

higher d_N/d_S ratios, indicating an increase in positive selective pressures and rapid protein evolution (Yang and Bielawski, 2000). It is unlikely that this observed elevation of change in non-synonymous sites is attributable to mutational pressures alone. Instead, this pattern is best explained by the elevation of fixation of deleterious mutations through genetic drift (Brynnel, et. al., 1998; Clark, et. al., 1999), to which the endosymbionts are subjected to with each host generation due to the aforementioned mode of vertical transmission. This is in keeping with the observation that primary endosymbionts as well as many intracellular pathogens undergo extreme population bottlenecks with each new host generation (Funk, et. al., 2001).

Exacerbating the effect of small population sizes is the strict asexual mode of reproduction through which these bacteria replicate (Moran, 1996). Acting through Muller's Ratchet, asexual reproduction prevents the recovery of wild-type genotypes through recombination (Muller, 1964). Within primary endosymbionts, asexuality itself might have resulted from the loss of gene elements through containment within host cells, as well as the loss of recombination functions that typically mediate the acquisition of novel genes (Tamas, et. al., 2002). In addition to the combined effects of these evolutionary forces driving the accumulation of deleterious mutations through random genetic drift, primary endosymbionts also display a mutational bias towards A+T base pairs within the genome. This trend is common within intracellular mutualists and pathogens alike (Heddi, et. al., 1998; Heddi, et. al., 1999). Many common mutations that occur within the bacterial genome result in the shifting of sequence composition from G+C to an A+T base pair bias. This compositional bias within endosymbionts that have undergone a radical reduction in genome size is likely the result of a loss of general DNA

maintenance and repair genes (Lawrence and Roth, 1999; Moran and Baumann, 2000; Moran, 2002).

All of these factors contribute to the reduction in functional genome size within bacterial endosymbionts. Bacteria genomes contain primarily coding DNA, and as such these genome reductions result through the loss of metabolic and physiological capabilities (Andersson and Kurland, 1998; Ochman and Moran, 2001). The drastic loss of functionality resulting from genome reduction likely accounts for the inability to culture primary bacterial endosymbionts independently of their host. Not surprisingly, recent studies suggest an overall mutational bias towards deletions rather than insertions within these genomes. This reflects a bias in spontaneous mutation, rather than selection towards smaller genomes (Mira, et. al., 2001). Thus, selection here does not favor those individuals with smaller genomes; instead, reduced strength of selection increases the number of nucleotide sites that may be altered without significant fitness consequences, strengthening the impact of deletion biases. However, endosymbionts are under selective pressure to retain specific metabolic functions that are necessary for host survival and reproduction. After all, a decrease in host fitness reduces the fitness of the endosymbiont as well (Rispe and Moran, 2000). Thus, the endosymbiont genome is the result of a continuous interplay between random genetic drift and the reduction of genes through relaxed selection within large portions of the genome, and natural selection acting to preserve those genes vital to host survival and fecundity.

While the comparison of closely related taxa can highlight the evolutionary patterns operating on relatively short timescales, a somewhat different approach is needed to explain the long-term reduction of the bacterial genome. A somewhat recent approach

has elucidated the history of deletion within the primary aphid endosymbiont *Buchnera* (Moran and Mira, 2001). A maximum parsimony approach was used to reconstruct the ancestral genome of *Buchnera-Ap*. Results of analysis indicated that early gene loss in this primary endosymbiont was dominated by a small number of large deletions comprised of contiguous regions of the ancestral genome, often containing genes with no related functionality. This demonstrates that early genome reduction involved selection acting upon the combined fitness of large sets of genes, as well as a great deal of chance resulting from the location of these gene sets. Thus, significant deletion events early in the evolutionary history of a bacterial endosymbiont may alter the selective pressures of the loci that remain, and the subsequent erosion of individual genes. Many insect endosymbionts display genome stability since the initial loss of functionality immediately after their transition from free-living to an intracellular lifestyle.

To date, the rate of genome evolution within insect endosymbionts has been the topic of a number of studies. Brynnel et. al. (2003) examined the rates of evolution acting upon the *tuf* gene within the aphid primary endosymbiont *Buchnera*. Brynnel's findings suggest that, at both non-synonymous and synonymous mutation sites, the nucleotide sequence for this gene in *Buchnera* is evolving at a 10-fold higher rate than rates observed in *Escherichia coli* and *Salmonella typhimurium*, and up to a 40-fold increase in synonymous codon mutation rates (Brynnel, et. al., 2003). Similarly, (Herbeck et. al., 2003) also examined rates of evolution within *Buchnera.* It was found through testing for genetic drift in the chaperonin *groEL*, that this gene is evolving at a 10 to 25-fold slower rate when compared to a handful of other examined *Buchnera* genes (Herbeck et. al., 2003), though the observed rate of polymorphism was still elevated when compared to

free-living taxa. This reduced rate of substitutions is likely due to strong purifying selection within *groEL*, which is over-expressed within *Buchnera*. Additionally, the tsetse fly endosymbiont *Wigglesworthia* also displays notable patterns of genome evolution, exhibiting codon usage biases promoted by the unique patterns of nucleotide usage (strong bias towards A+T base pairs) within bacterial endosymbionts (Herbeck, et. al., 2003).

The above studies confirmed previously established theoretical expectations that bacterial endosymbionts experience similarly elevated rates of sequence evolution when compared to free-living bacteria taxa. Reduced effective population sizes and random genetic drift characterize the evolutionary history of endosymbiotic bacteria, leading to the accumulation of slightly deleterious mutations within the genome. However, these previous studies are somewhat limited in scope, comparing relatively small sets of genes and genomes. The availability of eight separate, fully sequenced *Blattabacterium* genomes, in addition to the availability of free-living *Flavobacterium* genomes for comparison, offers an unparalleled opportunity to elucidate the patterns of genetic drift within primary insect endosymbionts across a number of genes and genomes that, until recently, has been inconceivable.

Blattabacterium: the Cockroach Endosymbiont

Eight *Blattabacterium* genomes have been sequenced to date from the following hosts: *Periplaneta americana* (Sabree, et. al., 2009), *Blatta germanica* (Lopez-Sanchez, et. al., 2009), *Cryptocercus punctulatus* (Neef, et. al., 2011)*, Blaberus giganteus* (Huang, et. al., 2012), *Blatta orientalis* (Patino-Navarette, et. al., 2013), *Nauphoeta cinerea*

(Kambhampati, et. al., 2013), *Panesthia angustipennis spadica* (Tokuda, et. al., 2013), and the termite, *Mastotermes darwiniensis* (Sabree, et. al., 2012). These genomes, while largely similar to one another, also exhibit some differences in structure and function, indicating considerable independent evolutionary pressures among separate lineages harbored by the various host species. Like many other obligate mutualist endosymbionts, *Blattabacterium* is transmitted vertically from mother to offspring with each new host generation.

In addition to divergence within *Blattabacterium*, of particular interest are the ways in which bacterial endosymbiont genomes diverge from those in free-living genera. The bacterial genus *Flavobacterium* is the closest extant free-living relative to the cockroach endosymbiont *Blattabacterium* (Bandi, et. al., 1994; Kambhampati, 2010). Selection pressures differ radically between free-living and endosymbiotic taxa. These pressures likely contribute to the patterns of genome evolution observed in insect bacterial endosymbionts such as a reduced genome size and high A+T content; patterns, which are, not surprisingly, present within all sequenced *Blattabacterium* genomes (Kambhampati, et. al., 2013; Patino-Navarette, et. al., 2013). Members of the genus *Flavobacterium* tend to occur in temperate or polar environments, typically within soil, freshwater, or marine habitats (Kambhampati, 2010). Only one *Flavobacterium* species, *Flavobacterium indicum*, has been isolated from a warm habitat. Many *Flavobacterium* species act as opportunistic pathogens in freshwater fish, resulting in considerable economic losses worldwide (Bernardet and Bowman, 2010).

It is likely that *Blattabacterium* had already broken away from *Flavobacterium* and been established as an endosymbiont by the time modern cockroaches and termites

shared a common ancestor. Subsequently, most termites have lost this endosymbiont - a notable exception being *Mastotermes darwiniensis* (Sabree, et. al., 2012) - while the *Blattabacterium* symbiosis has been retained by cockroaches. The complete annotation of eight separate *Blattabacterium* strains presents us with a unique data set with which to examine genome evolution and the direction of natural selection within this group. Utilizing information obtained in previous studies, and the data gathered here, we seek to elucidate the mechanisms affecting the evolutionary history of *Blattabacterium*, and explore the magnitude and direction of selection within this primary cockroach endosymbiont. We will calculate the positive and negative selection events occurring within the genomes of all sequenced *Blattabacterium* strains, and compare those events to those present within the closely related but free-living *Flavobacterium* species (*Flavobacterium indicum*, *Flavobacterium johnsoniae*, and *Flavobacterium psychrophilim*), and examine the similarities and differences between these two evolutionarily divergent groups. Furthermore, we will compare our results with those reported previously for *Buchnera* and *Wigglesworthia*. It is hypothesized that patterns of selection acting upon the *Blattabacterium* genome will be similar to those acting upon other insect endosymbionts (Wernegreen, 2002; Herbeck, et. al., 2003, Herbeck et. al., 2003, Brynnel, et. al., 2003). Specifically, an elevation in both non-synonymous and synonymous mutation events is predicted, as well as a higher d_N/d_S ratio at sites under significant levels of selection than the free-living *Flavobacterium;* indicating increased positive selection pressures and an elevated rate of protein evolution. Additionally, it is expected that those genes under significant positive selection within *Blattabacterium*, if any, will primarily be those relating to host nutrient supplementation.

MATERIALS AND METHODS

Sequences

Gene homolog files were manually compiled from sequences previously uploaded into GenBank [Blattabacterium: NC_017924.1, NC_020195.1, NC_013454.1, NC_016146.1, NC_022550.1, NC_020510.1, NC_013418.2, NC_016621.1; Flavobacterium: NC_017025.1, NC_009441.1, NC_009613.3). Homolog sets where all eight *Blattabacterium* and all three *Flavobacterium* were not present were excluded from further analysis. Three-hundred and four gene sets remained after exclusions.

Alignment

Complete homolog files were then aligned using the Edit/Build Alignment function within MEGA 5.2.2 (Tamura, et. al., 2011). Sequences were aligned by MUSCLE for Codons (Edgar, 2004), using standard parameters. Gaps were removed before alignment.

Tree Files

16S rRNA nucleotide sequences were manually compiled into a single text file for all eight *Blattabacterium* and all three *Flavobacterium*. 16S rRNA sequences were then aligned using the Edit/Build Alignment function within MEGA 5.2.2. Sequences were aligned by MUSCLE for DNA. A maximum likelihood tree was then produced using the aligned 16S rRNA sequences, utilizing the Construct/Test Maximum Likelihood Tree function within MEGA 5.2.2. This tree file was then exported as a Newick (.nwk) file. This file was manually edited to remove the branch lengths so that just the data for the

basic tree shape remained, as required by HyPhy's (Kosakovsky et. al., 2005) Quick Selection Detection function (Section 2.4).

Preparation of Aligned Sequences and Analysis

Within each aligned sequence file, a trimmed phylogenetic tree tag was inserted following the last aligned sequence. Following the insertion of this tag, each individual aligned file was analyzed using Hypothesis Testing Using Phylogenies (HyPhy) (Kosakovsky et. al., 2005). Using the Quick Selection Detection function (Standard Analysis > Positive Selection > QuickSelectionDetection.bf , default settings of Quick Selection Detection were used), each alignment file was analyzed for significantly positive or negative selection events. Multiple analyses were carried out, with modification of the significance level for each analysis set. Statistical significance level, as chosen by the user at the end of a Quick Selection Detection analysis, was varied from $p = 0.05$ to $p = 0.20$, in 0.05 increments. The results of each of these analyses are saved independently for later review and analysis.

Selection Data Analysis

Using Microsoft Excel for Mac 2008, a table was constructed containing the information gathered from the output of the selection analyses. Information listed includes gene position within the *Mastotermes darwiniensis* genome (used as the reference genome for this analysis), locus tag, gene length, COG grouping, number of positive selection sites within that gene across all available species, number of negative selection sites within that gene, and the total number of selection sites within that gene. With this information, additional statistical analyses could be performed.

Phylogenetic Distance Analysis

Sequence similarity and phylogenetic distance analysis performed using Geneious R6 6.1.7 for Mac (Drummond, et. al., 2011).

RESULTS AND DISCUSSION

Tree Files

The phylogenetic relationship of these groups through the creation of Newick-formatted tree files based upon complete 16S rRNA nucleotide sequences from all species analyzed was inferred. *Flavobacterium indicum*, the one described *Flavobacterium* species to have been extracted from a warm habitat, has the most similar 16S rRNA sequence when BLASTed against described *Blattabacterium* strains.

Selection

Initial analysis of *Blattabacterium* homolog sets was carried out across all eight of the fully sequenced strains, using a significance level of $p = 0.05$. At this significance level, *Blattabacterium* displays a strong negative mutational bias, with a ratio of sites under negative to sites under positive selection of 11:1 across 304 genes (Table 1). This result is in line with the patterns of strong negative selection observed in other insect endosymbionts (Wernegreen, 2002). At the same level of significance, $p = 0.05$, the *Flavobacterium* homolog sets display very little directional selection, with a slight but clear bias towards negative selection events. Compared to its closest free-living relative, *Blattabacterium* displays increased rates of polymorphism at both non-synonymous (d_N) and synonymous sites (d_S). Thus, in terms of over all selection events, *Blattabacterium*

Table 1. Comparison of selection events within the endosymbiont *Blattabacterium* and its free-living relative *Flavobacterium*. At a significance level of p = 0.05, *Flavobacterium* displays no nucleotide sites under significant positive selection. By expanding the level of significance within this group to p = 0.15, a better understanding of the ratio of positive to negative mutation events could be gained. However, even at p = 0.15, *Flavobacterium* displays only a single location under significant positive selection, indicating an extreme trend towards negative mutation events within the examined *Flavobacterium* gene set, when compared to the ratios observed in *Blattabacterium*.

Taxonomic Group	P value	Number of sites under positive selection	Number of sites under negative selection	Ratio of neg. selected sites to pos. selected sites
Blattabacterium	0.05	180	1993	11.07
Flavobacterium	0.05	0	207	N/A
Flavobacterium	0.15	1	2576	2576

displays a 10-fold increase in significantly selected sites when compared to *Flavobacterium*. This is strikingly similar to the findings of Brynnel et. al. (2003), who found that the nucleotide sequence of the *tuf* gene within *Buchnera* is evolving more than 10 times as quickly than the same sequence in the free living *E. coli* and *S. typhimurium*. In addition to an overall elevation in selection events, *Blattabacterium* also displays an increased ratio of synonymous mutations to non-synonymous, when compared to *Flavobacterium*, which displayed almost no synonymous polymorphisms. This trend is mirrored in the *Buchnera* chaperonin *groEL*, which displays a 5-fold increase in non-synonymous mutations, and a 10-fold increase in synonymous mutations, when compared to *E. coli* (Herbeck, et. al., 2003). Mutational pressure alone likely does not account for the magnitude of these d_N/d_S rate elevations. Within *Buchnera*, it has been suggested that this elevation of fixation occurs through random genetic drift resulting from the continual reduction of effective endosymbiont population size with each transmission from host parent to offspring (Brynnel, et. al., 1998; Clark, et. al., 1999; Funk, et. al., 2001). Given that this same elevation of polymorphisms is observed within *Blattabacterium* and that

Blattabacterium also undergoes similar population bottlenecks with each host generation, it is conceivable that similar mechanisms are shaping these two independent lineages.

Flavobacterium genes was also analyzed at a significance level of $p = 0.15$. Since selection events are comparatively rare within *Flavobacterium* genes at a significance level of $p = 0.05$, loosening the significance level to $p = 0.15$ displayed more selection events, in both directions. This expands the usable dataset, encompassing events that would otherwise have been overlooked as non-significant, and affords us a better understanding of the ratio of selection within this group, and how it compares to *Blattabacterium*. *Flavobacterium* genes displayed almost no positive selection across examined gene homologs, indicating an extreme negative selection bias at this significance level. This also parallels the findings of (Brynnel et. al., 2003), who suggested that the rate of synonymous codon substitution within *Buchnera* can be as much as 40 times higher than free-living relatives.

Blattabacterium displays elevated levels of both positive and negative selection events at a significance level of $p = 0.05$ when compared to free-living *Flavobacterium*, indicating an increase in mutation rates across the examined genes. In order to ensure that these patterns are not the result of sequences displaying radically different divergence times, we performed a simple phylogenetic analysis to elucidate the sequence similarity within each examined group. Phylogenetic analysis of both the *Blattabacterium* group (Table 2) and *Flavobacterium* group (Table 3) indicate similar levels of phylogenetic divergence between the individuals of each.

Table 2. Sequence divergence in the 16S rRNA gene of *Blattabacterium*. A phylogenetic tree was created using the 16S rRNA gene from each sequenced *Blattabacterium* species. From this tree, phylogenetic distances were calculated in order to estimate sequence similarity and divergence. Host species abbreviations are as follows: BNCIN*, N. cinerea*; BGIGA, *B. giganteus;* BBge, *B. germanica*; BPLAN, *P. americana;* BCpu, *C. punctulatus;* MADAR, *M. darwiniensis,* BBor, *B. orientalis*; BPane, *P. angustipennis spadica.*

	BPLAN	**BCpu**	**BBge**	**BGIGA**	**MADAR**	**BNCIN**	**BBor**	**BPane**
Amer		0.048	0.037	0.044	0.056	0.04	0.015	0.04
Crypt	0.048		0.038	0.043	0.048	0.043	0.043	0.045
Germ	0.037	0.038		0.024	0.045	0.026	0.038	0.021
Giga	0.044	0.043	0.024		0.059	0.026	0.044	0.021
Madar	0.056	0.048	0.045	0.059		0.049	0.051	0.056
Naup	0.04	0.043	0.026	0.026	0.049		0.039	0.021
Orien	0.015	0.043	0.038	0.044	0.051	0.039		0.036
Pane	0.04	0.045	0.021	0.021	0.056	0.021	0.036	

Table 3. Sequence divergence in the 16S rRNA gene of *Flavobacterium*. A phylogenetic tree was created using the 16S rRNA gene from each *Flavobacterium* species used in this study. From this tree, phylogenetic distances were calculated in order to estimate sequence similarity and divergence. Species abbreviations: Findic, *Flavobacterium indicum*; Fjohn, *Flavobacterium johnsoniae*; Fpsych, *Flavobacterium psychrophilim*.

	Findic	**Fjohn**	**Fpsych**
Findic		0.074	0.08
Fjohn	0.074		0.055
Fpsych	0.08	0.055	

In addition, each group displays comparable percentages of identical sites (Blatt: 89.5%, Flav 89.9%) as well as similar pairwise percent identities (Blatt: 95.8%, Flav: 92.8%) when aligning the ribosomal 16S rRNA gene. Combined, these statistics indicate fairly similar divergence times between these two groups. Thus, modern *Blattabacterium* display signs of elevated rates of genome evolution in the form of increased levels of selection events. The increase in the number of sites experiencing negative or positive selection when compared to the free-living *Flavobacterium* indicates elevated levels of functional protein evolution.

Using the results gathered here in conjunction with patterns of genome evolution published in previous studies (Kambhampati, et. al., 2013; Table 4), one might draw some conclusions about the mechanisms of sequence evolution within *Blattabacterium*. Firstly, like other intracellular mutualists, *Blattabacterium* are passed through host generations via vertical transmission, resulting in successive bottlenecks at each new host generation. These bottlenecks have a profound effect upon *Blattabacterium* population structure and severely limit the effective population size. In accordance with the nearly-neutral model of evolution (Ohta, 1973), a reduction in effective population size elevates the rate of fixation of slightly deleterious mutations through genetic drift. This is evidenced by the observation that *Blattabacterium* displays a nucleotide bias towards A+T base pairs within its genome (Table 2). Similar biases away from G+C nucleotide composition are also evident within *Wigglesworthia* (~23.7%, Wernegreen and Moran, 1999; Herbeck, et. al., 2003) from the tsetse fly, and *Buchnera* (28-31%, Brynnel, et. al., 1998) from the aphid. Since many common mutations within the bacterial genome result in an alteration of nucleotide composition towards A+T base pairs (Wernegreen, 2002; Herbeck, et. al., 2003, Herbeck et. al., 2003, Brynnel, et. al., 2003), it is likely that the long-term accumulation of these mutations within the *Blattabacterium* genome is accountable for this nucleotide bias. In addition, within *Wigglesworthia*, evidence suggests that the predominant bias towards A+T base pairs drives codon usage trends within the bacterial endosymbiont genome (Wernegreen and Moran, 1999; Herbeck, et. al., 2003).

Given that bacterial genomes are primarily functional DNA, the drastic genome reduction observed within *Blattabacterium* and other insect endosymbionts has come at

the cost of physiological functionality, including the loss of many genes coding for DNA maintenance and repair (Ochman and Moran, 2001; Sabree, et. al., 2009; Kambhampati, et. al., 2013; Patino-Navarette, et. al., 2013). Thus, because of this loss in functionality, the ancestral *Blattabacterium* was likely unable to completely repair these mutations, resulting in an overall mutational bias towards deletions rather than insertions. This indicates a bias within *Blattabacterium* towards spontaneous mutation rather than selection towards a smaller genome size. Thus, it is found that reduced strength of selection increases the number of nucleotide sites that may be altered without consequences in fitness, strengthening the impact of deletion biases. In conclusion, nucleotide composition biases observed in *Blattabacterium* today are largely due to the compounding effects of random genetic drift and the subsequent fixation of slightly deleterious mutations. Accordingly, this trend towards the accumulation of deleterious mutations is in agreement with Muller's Ratchet (Muller, 1964).

Gene-Level Patterns

Most genes within *Blattabacterium* and *Flavobacterium* displayed a bias towards negative selection. However, a few genes in *Blattabacterium* demonstrated neutral or even slightly positive selection (Table 4). As indicated by the COG groupings, genes *ppa* and *sdhB* are involved in energy production, *hisC* and *speB* in amino acid metabolism and transport, *folE* in coenzyme metabolism, and *rplL* and *pth* in translation, ribosome biogenesis and structure. That the vast majority of genes within the *Blattabacterium* genome are experiencing neutral or negative selection is in keeping with observations made by others (Patino-Navarette, et. al., 2013; Wernegreen, 2002) indicating conserved genome architecture within established endosymbiont lineages.

Table 4. Locus within *Blattabacterium* experiencing neutral or slightly positive selection. Genes classified here as 'slightly positive' are those with an equal or higher number of sites under positive selection than negative selection within the examined gene. 'Position' indicates that genes position within the *Mastotermes darwineinsis* genome, the model *Blattabacterium* genome used here.

Position	Locus	Avg. Length (n)	COG	Pos. sites	Neg. Sites	Ratio
4	*ppa*	657	Energy production and conservation	1	1	1
64	*ruvA*	781	DNA replication and repair	1	0	N/A
121	*hisC*	1412	Amino acid transport and metabolism	4	4	1
224	*mutL*	2303	DNA replication and repair	4	3	1.33
239	*folE*	861	Coenzyme metabolism	2	2	1
372	*rplL*	487	Translation, ribosomal structure and biogenesis	1	0	N/A
388	*sdhB*	985	Energy production and conservation	4	4	1
466	*speB*	1185	Amino acid transport and metabolism	1	1	1
483	*pth*	772	Translation, ribosomal structure and biogenesis	1	1	1

However, specific genes maintained by selective pressures will vary from endosymbiont lineage to endosymbiont lineage, depending upon the metabolic and physiological requirements of the host species. Thus, while random genetic drift plays a large role in shaping the *Blattabacterium* genome, selection plays a small, yet significant role in maintaining *Blattabacterium*'s functionality as a primary nutritional endosymbiont across the cockroach lineage. While similar mechanisms shape the genomes of *Wigglesworthia* and *Buchnera*, the gene composition and functionality of these genomes differs greatly due to the specific physiological requirements of the host organism; tsetse flies and aphids, respectively (Rita, et. al., 2012; Herbeck, et. al., 2003; Brynnel, et. al., 1998). Accordingly, the *Blattabacterium* genomes we examined here are the result of continual interplay between random genetic drift and the fixation of slightly deleterious mutations, and selection promoting the maintenance of cockroach-required metabolic functionality.

CONCLUSIONS

These findings indicate that the *Blattabacterium* genome is experiencing elevated rates of both positive and negative selection when compared to its free-living relative *Flavobacterium*, possessing a 10-fold increase in selection rate at the significance level p = .05 across 304 individual genes. The vast majority of observed selection events in the *Blattabacterium* genome were negative in direction, with only a handful of genes displaying selection in the positive direction. Combined with previous publications elucidating the evolutionary patterns within other insect endosymbionts, we may conclude that the *Blattabacterium* genome is shaped by similar evolutionary mechanisms. Previous studies have outlined the current state of the *Blattabacterium* genome, which is drastically reduced from its ancestral state and possesses a very strong bias towards A+T nucleotide base pairs. Analysis of these trends indicate that the *Blattabacterium* is experiencing an accumulation of slightly deleterious mutations through the continued effects of random genetic drift resulting from consecutive population bottlenecks throughout *Blattabacterium*'s evolutionary history. Additionally, *Blattabacterium* has lost many of its DNA repair functionality, likely though similar mechanisms discussed here, thus exacerbating this evolutionary bias towards slightly deleterious mutations. That these mutations cannot be repaired increases functional protein evolution rates within this endosymbiont. The patterns discussed here are highly similar to those evolutionary and genomic trends observed in other intracellular insect endosymbionts (Moran, 1996; Clark, 1999; Moran, 2002; Wernegreen, 2002).

The analysis presented here could be augmented through a robust analysis of genome reduction within *Blattabacterium*. Using a parsimony approach, the ancestral

genome of another primary insect endosymbiont, *Buchnera-Ap*, was reconstructed by Moran and Mira (Moran and Mira, 2001). The results of Moran and Mira's analysis indicated that much of the ancestral *Bucnhera* genome was lost during a relatively small number of large deletion events shortly after this bacteria's transition to an intracellular lifestyle. While it is likely that that *Blattabacterium*'s genome was reduced through similar mechanisms, a similar reconstruction within this group would offer us a more complete picture of the evolutionary origins of this unique cockroach endosymbiont.

REFERENCES

Andersson SG, Kurland CG. (1998). Reductive evolution of resident genomes. *Trends Microbiol.* Vol. 6, pg. 263–268.

Bandi C, Damiani G, Magrassi L, Grigolo A, Fani R. (1994). Flavobacteria as intracellular symbionts in cockroaches. *Proc Biol Sci* Vol. 257, pg. 43–48.

Bernardet JF, Bowman JP. (2010). Genus *Flavobacterium*. *Bergey's Manual of Systematic Bacteriology*. 2nd Ed., Vol. 4, pg. 112-154. Springer, New York.

Bourtzis K, Miller TA. (2003). *Insect Symbiosis*. CRC Press, Boca Raton, FL, USA.

Brynnel EU, Kurland CG, Moran NA, Andersson SG. (1998). Evolutionary rates for *tuf* genes in endosymbionts of aphids. *Mol. Biol. Evol.* Vol. 15, pg. 574–582.

Buchner P. (1965). *Endosymbiosis of Animals with Plant Microorganisms*. New York: Interscience Publishers.

Chen X, Li S, Aksoy S. (1999). Concordant evolution of a symbiont with its host insect species: molecular phylogeny of genus *Glossina* and its bacteriome-associated endosymbiont, *Wigglesworthia glossinidia*. *J. Mol. Evol.* Vol. 48, pg. 49–58.

Clark MA, Moran NA. Baumann P. (1999). Sequence evolution in bacterial endosymbionts having extreme base compositions. *Mol. Biol. Evol.* Vol. 16, pg. 1586–1598.

Clark MA, Moran NA, Baumann P, Wernegreen JJ. (2000). Cospeciation between bacterial endosymbionts (*Buchnera*) and a recent radiation of aphids (*Uroleucon*) and pitfalls of testing for phylogenetic congruence. *Evolution*. Vol. 54, pg. 517-525.

Dasch G, Weiss E, Chang K. (1984). *Bergy's Manual of Systematic Bacteriology* Vol. 1, pg. 811–833.

Douglas AE. (1989). Mycetocyte symbiosis in insects. *Biol. Rev. Camb. Phil. Soc.* Vol 64, pg. 409–434.

Douglas AE. (1998). Nutritional Interactions in insect-microbial symbioses: Aphids and their symbiotic bacteria Buchnera. *Annu. Rev. Entomol*. Vol 43, pg. 17-37.

Drummond AJ, Ashton B, Buxton S, Cheung M, Cooper A, Duran C, Field M, Heled J,Kearse M, Markowitz S, Moir R, Stones-Havas S, Sturrock S, Thierer T, Wilson A. (2011). Geneious v5.4, Available from http://www.geneious.com/

Edgar RC. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research.* Vol. 32, No. 5, pg. 1792-179.

Funk DJ, Helbling L, Wernegreen JJ, Moran NA. (2000). Intraspecific phylogenetic congruence among multiple symbiont genomes. *Proc. R. Soc. Lond. B. Biol. Sci*. Vol. 267, pg. 2517-2521.

Funk DJ, Wernegreen JJ, Moran NA. (2001). Intraspecific variation in symbiont genomes: bottlenecks and the aphid-Buchnera association. *Genetics*. Vol. 157, pg. 477-489.

Heddi A, Charles H, Khatchadourian C, Bonnot G, Nardon P. (1998). Molecular characterization of the principal symbiotic bacteria of the weevil *Sitophilus oryzae*: a peculiar G+C content of an endocytobiotic DNA. *J. Mol. Evol.* Vol. 47, pg. 52–61.

Heddi A, Grenier AM, Khatchadourian C, Charles H, Nardon P. (1999). Four intracellular genomes direct weevil biology: nuclear, mitochondrial, principal endosymbiont, and *Wolbachia. Proc. Natl Acad. Sci. USA* Vol. 96, pg. 6814–6819.

Herbeck TH, Funk DJ, Degnan PH, Wernegreen JJ. (2003). A conservative test of Genetic Drift in the Endosymbiotic Bacterium Buchnera: Slightly Deleterious Mutations in the Chaperonin *groEL. Genetics*. Vol. 165, pg. 1651-1660.

Herbeck JT, Wall DP, Wernegreen JJ. (2003). Gene expression level influences amino acid usage, but not codon usage, in the tsetse fly endosymbiont *Wigglesworthia*. *Microbiol.* Vol. 149, pg. 2585-2596.

Huang CY, Sabree ZL, Moran NA. (2012). Genome sequence of *Blattabacterium* sp. strain BGIGA, endosymbiont of the *Blaberus giganteus* cockroach. *J. Bacteriol.* Vol. 194, No. 19, pg. 4450.

Ishikawa H. (2003). Insect Symbiosis: an Introduction. P. 1-21. In K. Bourtzis and T.A. Miller (ed), Insect Symbiosis. CRC Press, Boca Raton, Fl, USA.

Kambhampati S. (2010). Genus *Blattabacterium*. *Bergey's Manual of Systematic Bacteriology*. 2nd Ed., Vol. 4, pg. 315-321. Springer, New York.

Kambhampati S, Alleman A, Park Y. (2013). Complete Genome Sequence of the endosymbiont *Blattabacterium* from the Cockroach Nauphoeta cinerea (Blattodea: Blaberidae). *Genomics.* Vol. 102, No. 5-6, pg. 479-483.

Kosakovsky Pond, SL, Frost SDW, Muse SV. (2005). HyPhy: hypothesis testing using phylogenies. *Bioinformatics*. Vol. 21, pg. 676-679

Lawrence J, Roth J. (1999). in *Organization of the Prokaryotic Genome* (ed. Charlesbois, R.) pg. 263–289 (ASM Press, Washington, DC).

Lo N, Bandi C, Watanabe H, Nalepa C, Beninati T. (2003). Evidence for cocladogenesis between diverse dictyopteran lineages and their intracellular endosymbionts. *Mol Biol Evol.* Vol. 20, pg. 907–913.

Lopez-Sanchez MJ, Neef A, Pereto J, Patino-Navarette R, Pignatelli M, Latorre A, Moya A. (2009). Evolutionary convergence and nitrogen metabolism in *Blattabacterium* strain Bge, primary endosymbiont of the cockroach *Blattella germanica*. *PLoS Genet*. 5:e1000721.

Margulis L, Fester R. (1991). *Symbiosis as a Source of Evolutionary Innovation*. MIT Press, Cambridge, MA, USA.

McDonald JH, Kreitman M. (1991). Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature*. Vol. 351, pg. 652-654.

Mira A, Ochman H, Moran NA. (2001). Deletional bias and the evolution of bacterial genomes. *Trends Genet.* Vol. 17, pg. 589–596.

Mira A, Moran NA. (2002). Estimating population size and transmission bottlenecks in maternally transmitted endosymbiotic bacteria. *Microb. Ecol.* Vol. 44, pg, 137–143.

Moran NA. (1996). Accelerated evolution and Muller's ratchet in endosymbiotic bacteria. *Proc. Natl Acad. Sci. USA* Vol. 93, pg. 2873–2878.

Moran NA, Telang A. (1998). Bacteriocyte-associated symbionts of insects. *Bioscience.* Vol. 48, pg. 295–304.

Moran NA, Baumann P. (2000). Bacterial endosymbionts in animals. *Curr. Opin. Microbiol.* Vol. 3, pg. 270–275.

Moran NA, Mira A. (2001). The process of genome shrinkage in the obligate symbiont *Buchnera aphidicola*. *Genome Biol.* Vol. 2.

Moran NA. (2002). Microbial minimalism: genome reduction in bacterial pathogens. *Cell* Vol. 108, pg. 583–586.

Muller HJ. (1964). The relation of recombination to mutational advance. *Mutat. Res.* Vol. 1, pg. 2–9.

Neef A, Latorre A, Pereto J, Silva FJ, Pignatelli M, Moya A. (2011). Genome economization in the endosymbiont of the wood roach *Cryptocercus punctulatus* due to drastic loss of amino acid synthesis capabilities. *Genome Biol. Evol*. Vol. 3, pg. 1437.

Ochman H, Moran NA. (2001). Genes lost and genes found: evolution of bacterial pathogenesis and symbiosis. *Science* Vol. 292, pg. 1096–1099.

Ohta T. (1973). Slightly deleterious mutant substitutions in evolution. *Nature,* Vol. 246, pg. 96–98.

Patino-Navarette, R, Moya, A, Latorre, A, Pereto, J. (2013). Comparative genomics of *Blattabacterium cuenoti*: the frozen legacy of an ancient endosymbiont genome. *Genome Biol. Evol.* Vol. 5, pg. 351-361.

Rispe C, Moran NA. (2000). Accumulation of deleterious mutations in endosymbionts: Muller's ratchet with two levels of selection. *Am. Nat.* Vol. 156, pg. 424–441.

Rio VMR, Symula RE, Wang J. (2012). Insight into the Transmission Biology and Species-Specific Functional Capabilities of Tsetse (Diptera: Glossinidae) Obligate Symbiont *Wigglesworthia*. *mBio*. Vol. 3, No. 1, e00240-11.

Ruby, E., Henderson, B., McFall-Ngai, M. (2004). We get by with a little help from our (little) friends. *Science*. Vol. 303, pg. 1305-1307.

Sabree ZL, Huang CY, Arakawa G, Tokuda G, Lo N, Watanabe H, Moran NA. (2012). Genome shrinkage and loss of nutrient-providing potential in the obligate symbiont of the primitive termite *Mastotermes darwiniensis*. *Appl Environ Microbiol.* Vol. 78, pg. 204-210.

Sabree ZL, Kambhampati S, Moran NA. (2009). Nitrogen recycling and nutritional provisioning by *Blattabacterium*, the cockroach endo-symbiont. *Proc Natl Acad Sci USA*. Vol. 106, pg. 19521–19526.

Tamas I, Klasson L, Canback B, Naslund AK, Eriksson AS, Wernegreen JJ, Sandstrom JP, Moran NA, Andresson SGE. (2002). 50 million years of genomic stasis in endosymbiotic bacteria. *Science* Vol. 296, pg. 2376–2379.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, and Kumar S. (2011). MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol Biol Evo*. Vol. 28, pg. 2731-2739.

Tokuda G, Elbourne LDH, Kinjo Y, Saitoh S, Sabree Z, Hojo M, Yamada A, Hayashi Y, Shigenobu S, Bandi C, Paulsen IT, Watanabe H, Lo N. (2013). Maintenance of Essential Amino Acid Synthesis Pathways in the *Blattabacterium* cuenoti Symbiont of a Wood-Feeding Cockroach. *Biol. Lett.* Vol. 9, No. 3, 20121153.

Wernegreen JJ, Moran NA. (1999). Evidence for genetic drift in endosymbionts (*Buchnera*): analyses of protein-coding genes. *Mol. Biol. Evol.* Vol. 16, pg. 83–97.

Wernegreen, J.J. (2002). Genome Evolution in Bacterial Endosymbionts of Insects. *Nat. Rev. Genet.* Vol. 3, No. 11, pp. 850-861.

Chapter Four

GENERAL DISCUSSION AND CONCLUSIONS: GENOME FUNCTION AND EVOLUTION OF THE COCKROACH SYMBIONT *BLATTABACTERIUM*

METABOLIC FUNCTION OF *BLATTABACTERIUM* WITHIN ITS HOST

From the genome analyzed here, and others previously described, we may elucidate the function of *Blattabacterium* as a nutritional symbiont within cockroaches. It functions primarily to synthesize amino acids and vitamins from the nitrogenous waste products of the host through the utilization of uric acid as a nitrogen reserve. This is beneficial to the cockroach since its primary diet of decaying organic matter tends to be nitrogen-deficient. Accordingly, and unlike many insects, cockroaches excrete waste nitrogen in the form of ammonia instead of uric acid. The ecological effects upon the host are profound, allowing cockroaches to move into and exploit ecological niches that were previously unavailable. While cockroaches most certainly benefit from this symbiosis, living inside of their host affords these bacteria relative safety from competition and exploitation. Thus, this insect-host symbiosis is a mutual one. However, as has been shown in other insect symbionts, studying *Blattabacterium* outside of its host is exceedingly difficult. Recent advances in DNA-sequencing techniques and bioinformatics have allowed for the elucidation of genome composition and function within *Blattabacterium*, and other endosymbionts as well, to a degree of accuracy that has never before been attainable. As such, *Blattabacterium* has been extracted from eight host species to date: *Peirplaneta americana* (Sabree, et. al., 2009), *Blatta germanica* (Lopez-Sanchez, et. al., 2009), *Cryptocercus punctulatus* (Neef, et. al., 2011), *Blaberus giganteus* (Huang, et. al., 2012), *Blatta orientalis* (Patino-Navarette, et. al., 2013), *Panesthia angustipennis* (Tokuda, et. al., 2013), *Nauphoeta cinerea* (Kambhampati, et. al., 2013) and the termite, *Mastotermes darwiniensis* (Sabree, et. al., 2012); and, with the

exception of the cave-dwelling genus Bocticola, it has been hypothesized that this symbiont is present within all cockroach genuses.

BLATTABACTERIUM GENOMICS

The *Blattabacterium* genome, comprised of ~620,000 base pairs and ~620 individual genes, is drastically reduced when compared to *Flavobacterium*, *Blattabacterium*'s closest free-living relative. Their genome is comprised of one chromosome and one plasmid of approximately 3,500 base pairs; a notable exception to this being the *Blattabacterium* species from *Panesthia angustipennis*, which lacks a plasmid altogether. Like other intracellular symbionts, the *Blattabacterium* genome has a G+C content of ~27%. The low occurrence of guanine and cytosine is common within bacterial symbionts with radically small genome sizes. Synteny within the *Blattabacterium* genome is high, with one large 242,000 base pair inversion occurring in *Mastotermes darwiniensis*, and two smaller inversions within Blattidae. Gene content within *Blattabacterium* has also been conserved (Patino-Navarette, et. al., 2013). This information suggests that the *Blattabacterium* genome underwent massive gene loss immediately after the transition from a free-living to intracellular lifestyle. The *Blattabacterium* genome lacks almost all specific transporters and regulatory genes, an indication of the highly specialized and dependent nature of insect endosymbionts. Interestingly, *Blattabacterium* lacks genes coding for uricase, or a uricase homolog, which is integral to the utilization of uric acid as a way to store nitrogen. This step, then, is not carried out by *Blattabacterium*, but either by the cockroach host itself, or by bacteria within the insect's gut (Sabree, 2009), as is the case with termites (Tokuda and

Watanabe, 2007). Seeing as how the *Flavobacterium* genome lacks genes coding for uricase as well, this absence in *Blattabacterium* is likely an evolutionary artifact.

EFFECTS OF SELETION UPON THE *BLATTABACTERIUM* GENOME

Our findings indicate that the *Blattabacterium* genome is experiencing elevated rates of selection when compared to its free-living relative *Flavobacterium*, possessing a 10-fold increase in selection rate at significance level .05 across 304 individual genes. This is in keeping with the rates of selection observed in other bacterial mutualists (Brynnel, et. al., 2003; Herbeck et. al., 2003; Herbeck, et. al., 2003). The vast majority of observed selection events in the *Blattabacterium* genome were negative in direction, with only a handful of genes displaying selection in the positive direction. However, what few genes were found to be under positive selection are critical to its role as a nutritional aid to its host. This is also in line with our original hypothesis.

Like other insect bacterial endosymbionts, *Blattabacterium* displays a drastically reduced genome, with much of its functionality lost when compared to free-living relatives such as *Flavobacterium*. Recent studies (Moran and Mira, 2001; Patino-Navarette, et. al., 2013) have suggested that the genome of *Blattabacterium*, and other insect symbionts, undergoes a drastic and rapid reduction very soon after entering the intracellular lifestyle. Results of analysis indicated that early gene loss in *Blattabacterium* was dominated by a small number of large deletions comprised of contiguous regions of the ancestral genome, often containing genes with no related functionality, a pattern of genome reduction that has been suggested in other symbionts as well (Patino-Navarette, et. al., 2013). This demonstrates that early genome reduction involved selection acting

upon the combined fitness of large sets of genes, as well as a great deal of chance resulting from the location of these gene sets.

However, current *Blattabacterium* genomes display extremely stable genome architecture, as well as elevated rates of spontaneous mutations - rather than direct selection towards a decreased genome size - resulting in continued and elevated rates of genome reduction when compared to free-living bacteria (Patino-Navarette, et. al., 2013; Mira, et. al., 2001). Since bacterial genomes contain mostly coding DNA, this genome reduction in *Blattabacterium* results in the continued loss of metabolic and physiological functionality. However, while much of the *Blattabacterium* is influenced by random genetic drift, selection also plays a role in maintaining those genes that are important to host survival. Thus, the *Blattabacterium* is the result of a continuous interplay between random genetic drift – the reduction of genes through relaxed selection within large portions of the genome, and natural selection acting to preserve those genes vital to host survival and fecundity.

Evidence of the combined effects of these evolutionary forces driving the accumulation of deleterious mutations through random genetic drift may be observed in the form of a mutational bias towards A+T base pairs within the *Blattabacterium* genome. This bias is shared with many other intracellular symbionts and pathogens alike (Heddi, et. al., 1998; Heddi, et. al., 1999). Many common mutations that occur within the bacterial genome result in the shifting of sequence composition from G+C to an A+T base pair bias. This compositional bias within symbionts that have undergone a radical reduction in genome size is likely the result of a loss of general DNA maintenance and repair genes (Moran and Wernegreen, 2000; Moran, 2002).

FUTURE DIRECTIONS

So far, the genome description of *Blattabacterium* and others offer insight into the loss of metabolic and physiological diversity within independent bacterial mutualist lineages, due to their intracellular associations. Their remaining metabolic abilities are likely influenced by a number of different factors, including host physiology, symbiont transmission mechanisms, and random aspects of deletion events that occurred early in the symbiont's evolutionary history. Current genomic data are becoming more comprehensive in encompassing the diversity of symbionts within various insect orders. However, there remain areas where further information is needed. For example, *Blattabacterium* genomes within the Blattodea family Polyphagidae have, as of the time of this writing, not yet been extracted and annotated. Comparing the gene inventories of an array of new intracellular endosymbionts will allow us to test the generality of the genomic patterns observed in *Blattabacterium* and other recently-described insect symbionts. The annotation of symbiont genomes and identification of genes and pathways will assist in elucidating the specific function of each symbiont with respect to its insect host. In addition, with the rise in number of described symbiont genomes, selection pressure analysis, like the one we have performed here, should become increasingly comprehensive as related genomes for analysis become more available. By looking at the selective pressures acting upon insect endosymbionts, we may better understand the developmental history, and well as the evolutionary trajectory, of this group of diverse and interesting bacteria.

REFERENCES

Brynnel EU, Kurland CG, Moran NA, Andersson SG. (1998). Evolutionary rates for *tuf* genes in endosymbionts of aphids. *Mol. Biol. Evol.* Vol. 15, pg. 574–582.

Heddi A, Charles H, Khatchadourian C, Bonnot G, Nardon P. (1998). Molecular characterization of the principal symbiotic bacteria of the weevil *Sitophilus oryzae*: a peculiar G+C content of an endocytobiotic DNA. *J. Mol. Evol.* Vol. 47, pg. 52–61.

Heddi A, Grenier AM, Khatchadourian C, Charles H, Nardon P. (1999). Four intracellular genomes direct weevil biology: nuclear, mitochondrial, principal endosymbiont, and *Wolbachia. Proc. Natl Acad. Sci. USA* Vol. 96, pg. 6814–6819.

Herbeck TH, Funk DJ, Degnan PH, Wernegreen JJ. (2003). A conservative test of Genetic Drift in the Endosymbiotic Bacterium Buchnera: Slightly Deleterious Mutations in the Chaperonin *groEL. Genetics*. Vol. 165, pg. 1651-1660.

Herbeck JT, Wall DP, Wernegreen JJ. (2003). Gene expression level influences amino acid usage, but not codon usage, in the tsetse fly endosymbiont *Wigglesworthia*. *Microbiol.* Vol. 149, pg. 2585-2596.

Huang CY, Sabree ZL, Moran NA. (2012). Genome Sequence of *Blattabacterium* sp. Strain BGIGA, Endosymbiont of the Blaberus giganteus Cockroach. *J. Bacteriol.* Vol. 194, No. 19, pg. 4450.

Kambhampati S, Alleman A, Park Y. (2013). Complete Genome Sequence of the endosymbiont *Blattabacterium* from the Cockroach Nauphoeta cinerea (Blattodea: Blaberidae). *Genomics.* Vol. 102, No. 5-6, pp. 479-483.

Lopez-Sanchez MJ, et al. (2009). Evolutionary convergence and nitrogen metabolism in *Blattabacterium* strain Bge, primary endosymbiont of the cockroach Blattella germanica. *PLoS Genet*. 5:e1000721.

Mira A, Ochman H, Moran NA. (2001). Deletional bias and the evolution of bacterial genomes. *Trends Genet.* Vol. 17, pg. 589–596.

Moran NA, Wernegreen JJ. (2000). Lifestyle evolution in symbiotic bacteria: insights from genomics. *Trends Ecol. Evol.* Vol. 15, pg. 321–326.

Moran NA, Mira A. (2001). The process of genome shrinkage in the obligate symbiont *Buchnera aphidicola*. *Genome Biol.* Vol. 2.

Moran NA. (2002). Microbial minimalism: genome reduction in bacterial pathogens. *Cell* Vol. 108, pg. 583–586.

Neef A, Latorre A, Pereto J, Silva FJ, Pignatelli M, Moya A. (2011). Genome economization in the endosymbiont of the wood roach *Cryptocercus punctulatus* due to drastic loss of amino acid synthesis capabilities. *Genome Biol. Evol*. Vol. 3, pg. 1437–1448.

Patino-Navarette, R, Moya, A, Latorre, A, Pereto, J. (2013). Comparative genomics of *Blattabacterium cuenoti*: the frozen legacy of an ancient endosymbiont genome. *Genome Biol. Evol.* Vol. 5, pg. 351-361.

Sabree ZL, Kambhampati S, Moran NA. (2009). Nitrogen recycling and nutritional provisioning by *Blattabacterium*, the cockroach endo-symbiont. *Proc Natl Acad Sci USA*. Vol. 106, pg. 19521–19526.

Sabree ZL, Huang CY, Arakawa G, Tokuda G, Lo N, Watanabe H, Moran NA. (2012). Genome shrinkage and loss of nutrient-providing potential in the obligate symbiont of the primitive termite *Mastotermes darwiniensis*. *Appl Environ Microbiol.* Vol. 78, pg. 204-210.

Tokuda G, Watanabe H. (2007). "Hidden cellulases in termites: revision of an old hypothesis." *Biol. Lett.* Vol. 3, No. 3, pg. 336-339.

Tokuda G, Elbourne LDH, Kinjo Y, Saitoh S, Sabree Z, Hojo M, Yamada A, Hayashi Y, Shigenobu S, Bandi C, Paulsen IT, Watanabe H, Lo N. (2013). Maintenance of Essential Amino Acid Synthesis Pathways in the *Blattabacterium* cuenoti Symbiont of a Wood-Feeding Cockroach. *Biol. Lett.* Vol. 9, No. 3, 20121153.

Appendix A: Raw Selection Data Across All Examined Genes and Significance Levels

				BLAT	p-value = 0.05		FLAV	p-value= 0.15		FLAV	p-value= 0.05	
Position	**Locus**	**Avg. Length (n)**	**COG**	**Pos. sites**	**Neg. Sites**	**Total**	**Pos. sites**	**Neg. sites**	**Total**	**Pos. sites**	**Neg. sites**	**Total**
1	nusA	1566	K	0	12	12	0	11	11	0	0	0
2	infB	3493	J	1	46	47	0	8	8	0	0	0
3	aspS	2241	J	1	23	24	0	6	6	0	1	1
4	ppa	657	C	1	1	2	0	0	0	0	0	0
5	sucB	1578	C	0	4	4	0	24	24	0	0	0
8	mreC	1067	M	2	7	9	0	2	2	0	1	1
9	mreB	1311	D	2	8	10	0	5	5	0	0	0
11	argD	1499	E	0	6	6	0	7	7	0	1	1
12	truA	1001	J	1	2	3	0	5	5	0	1	1
13	msbA	2296	V	0	5	5	0	15	15	0	1	1
14	asnS	1851	R	1	14	15	0	5	5	0	1	1
15	mdh	1215	C	1	6	7	0	19	19	0	4	4
16	gcvP	3726	E	2	16	18	0	2	2	0	0	0
18	clpP	896	OU	0	1	1	0	1	1	0	0	0
21	wzxC	1898	R	0	2	2	0	14	14	0	0	0
23	recA	1302	L	0	5	5	0	1	1	0	0	0
26	alaS	3496	J	1	24	25	0	4	4	0	0	0
27	sucA	3535	C	1	13	14	0	20	20	0	1	1
28	phosphol	885	R	1	4	5			0			0
29	sucC	1564	C	0	3	3	0	58	58	0	0	0
30	dnaX	1894	L	0	8	8	0	3	3	0	0	0
32	trxA	419	O	1	2	3	0	1	1	0	0	0
34	polA	1685	L	2	7	9	0	58	58	0	0	0

Continued on next page

Appendix A (Continued)

35	metG	2310	J	0	8	8	0	15	15	0	0	0
37	ligA	2639	L	0	20	20	0	9	9	0	1	1
38	trmE	1821	R	2	7	9	0	3	3	0	1	1
39	gltX	1966	J	0	19	19	0	3	3	0	1	1
40	rpsR	407	J	0	1	1	0	1	1	0	0	0
41	rplI	589	J	0	5	5	0	5	5	0	0	0
46	phospho-2	1407	E	4	7	11	0	1	1	0	0	0
48	fabI	1031	I	1	4	5			0			0
50	dnaG	2445	L	1	6	7	0	54	54	0	0	0
52	miaB	1763	J	0	9	9	0	10	10	0	0	0
59	rpsP	511	J	0	2	2	0	7	7	0	2	2
60	bacA	1066	V	1	10	11	0	1	1	0	0	0
61	gyrB	2513	L	7	15	22	0	3	3	0	0	0
62	trmD	874	J	0	8	8	0	0	0	0	0	0
64	ruvA	781	L	1	0	1	0	3	3	0	0	0
69	pheS	1293	J	0	6	6	0	3	3	0	0	0
75	acpT	827	H	0	2	2	0	3	3	0	1	1
76	pnuC	818	H	0	4	4	0	1	1	0	0	0
77	hisS	1833	J	0	6	6	0	16	16	0	1	1
78	ftsZ	1681	D	0	7	7	0	6	6	0	1	1
79	ftsA	1785	D	0	4	4	0	7	7	0	1	1
81	murC	1794	M	1	13	14	0	25	25	0	1	1
83	ftsW	1599	D	2	12	14	0	1	1	0	0	0
85	mraY	1545	M	0	7	7	0	23	23	0	7	7
86	murE	1920	M	0	10	10	0	5	5	0	0	0
93	dapF	1043	E	1	6	7	0	10	10	0	1	1
95	ATP	1353	J	3	7	10	0	6	6	0	1	1
96	tatC	1045	U	2	3	5	0	2	2	0	0	0
98	smpB	590	O	0	1	1	0	9	9	0	3	3
99	tktB	1121	G	0	5	5	0	2	2	0	0	0
100	dxs	1477	G	0	5	5	0	56	56	0	0	0
102	aroA	1619	E	0	14	14	0	20	20	0	4	4
103	rpsT	317	J	0	0	0	0	2	2	0	1	1

104	murI	1059	M	0	4	4	0	3	3	0	0	0
109	ccoN	2848	O	0	25	25	0	6	6	0	0	0
112	gcvH	523	E	0	2	2	0	2	2	0	0	0
114	yggS	870	R	1	5	6	0	0	0	0	0	0
115	serC	1392	E	0	6	6	0	4	4	0	0	0
117	aroK	674	E	0	0	0	0	0	0	0	0	0
118	xthA	999	L	0	5	5	0	2	2	0	0	0
119	hisG	1117	E	1	5	6	0	3	3	0	0	0
120	hisD	1694	E	0	7	7	0	10	10	0	0	0
121	hisC	1412	E	4	4	8	0	13	13	0	0	0
122	hisB	1497	E	1	4	5	0	5	5	0	0	0
124	hisA	951	E	0	3	3	0	2	2	0	0	0
125	hisF	1002	E	0	3	3	0	2	2	0	0	0
128	folD	1142	H	0	1	1	0	3	3	0	0	0
129	rsmA	1014	J	2	5	7	0	8	8	0	6	6
130	serS	1662	J	0	8	8	0	1	1	0	0	0
132	pdhB	1278	C	0	7	7	0	1	1	0	0	0
134	folA	653	H	0	6	6	0	0	0	0	0	0
135	valS	3385	J	1	8	9	0	23	23	0	0	0
138	ndk	578	F	0	2	2	0	4	4	0	1	1
139	sucA	3293	C	0	3	3	0	31	31	0	0	0
140	ilvE	1362	H	1	11	12	0	15	15	0	0	0
141	argS	2294	J	5	20	25	0	2	2	0	0	0
142	ffh	1726	U	0	4	4	0	6	6	0	0	0
146	thrS	1814	J	0	2	2	0	6	6	0	2	2
147	infC	612	J	0	2	2	0	0	0	0	0	0
148	rpmI	249	N/A	0	0	0	0	0	0	0	0	0
149	rplT	456	J	0	0	0	0	6	6	0	6	6
150	holiday	544	L	1	7	8	0	6	6	0	0	0
151	dapD	1072	E	0	5	5	0	15	15	0	2	2
153	cca	1791	J	0	8	8	0	5	5	0	0	0
160	fbaA	1400	G	0	3	3	0	4	4	0	0	0
161	accD	1098	I	0	3	3	0	9	9	0	0	0

Appendix A (Continued)

162	guaA	2008	F	1	5	6	0	0	0	0	0	0
163	purD	1644	F	1	3	4	0	0	0	0	0	0
165	purN	739	F	1	11	12	0	5	5	0	0	0
166	purM	1335	F	0	16	16	0	0	0	0	0	0
167	purF	1997	F	0	4	4	0	3	3	0	0	0
168	purC	1004	F	1	3	4	0	24	24	0	2	2
170	purL	4806	F	1	12	13	0	3	3	0	0	0
171	purB	1843	F	5	6	11	0	2	2	0	0	0
172	purA	1684	F	0	4	4	0	0	0	0	0	0
173	ruvB	1273	L	0	11	11	0	6	6	0	3	3
176	gyrA	3159	L	0	10	10	1	11	12	0	0	0
177	aroB	1379	E	0	4	4	0	1	1	0	0	0
179	tatA	576	J	0	2	2	0	4	4	0	0	0
180	rplY	805	J	1	6	7	0	1	1	0	1	1
181	prsA	1227	E	2	7	9	0	2	2	0	1	1
182	thyA	1057	F	2	3	5	0	5	5	0	1	1
188	rnhA	607	L	0	0	0	0	1	1	0	0	0
190	fbp	1305	G	0	3	3	0	32	32	0	0	0
191	marC	742	U	0	1	1	0	2	2	0	0	0
193	pyrD	1269	F	0	2	2	0	0	0	0	0	0
196	accB	626	I	0	0	0	0	4	4	0	0	0
197	accC	1768	I	0	12	12	0	25	25	0	1	1
198	prfA	1424	J	0	16	16	0	4	4	0	0	0
200	rho	2113	K	1	8	9	0	7	7	0	0	0
203	rpmB	309	J	0	2	2	0	0	0	0	0	0
204	rpmG	239	J	0	1	1	0	0	0	0	0	0
207	htpG	2420	O	5	20	25	0	4	4	0	0	0
208	evoX	1009	L	0	2	2	0	1	1	0	0	0
214	tatD	1003	L	3	5	8	0	12	12	0	0	0
215	fabD	1135	I	1	8	9	0	6	6	0	3	3
219	binding	392	L	0	4	4	0	1	1	0	0	0
220	mutY	1356	L	0	5	5	0	1	1	0	0	0
224	mutL	2303	L	4	3	7	0	12	12	0	3	3

Appendix A (Continued)

226	signal	1949	U	2	7	9	0	23	23	0	4	4
227	dapB	924	E	0	4	4	0	9	9	0	2	2
229	dnaJ	1457	O	0	8	8	0	15	15	0	4	4
230	grpE	761	O	0	6	6	0	6	6	0	0	0
231	trpS	1285	J	0	2	2	0	14	14	0	0	0
232	atpG	1136	C	0	6	6	0	8	8	0	1	1
233	atpA	2057	C	0	8	8	0	9	9	0	3	3
234	atpH	709	C	2	3	5	0	7	7	0	1	1
235	atpF	647	C	0	6	6	0	5	5	0	1	1
237	atpB	1359	C	0	2	2	0	20	20	0	0	0
238	cysS	1917	J	2	9	11	0	0	0	0	0	0
239	folE	861	H	2	2	4	0	6	6	0	1	1
240	nadE	1028	H	0	0	0	0	3	3	0	0	0
241	m22	843	O	0	6	6	0	11	11	0	0	0
245	hemD	944	H	0	1	1	0	38	38	0	0	0
251	cytoC	1754	C	0	12	12			0			0
252	carB	4101	F	1	29	30	0	38	38	0	7	7
253	carA	1428	F	4	6	10	0	4	4	0	0	0
254	argD	1499	E	0	6	6	0	7	7	0	1	1
257	polC	5224	L	0	28	28			0			0
258	rpsA	2326	J	1	19	20	0	5	5	0	1	1
264	asd	1293	E	0	11	11	0	16	16	0	1	1
265	gmk	751	F	0	3	3	0	12	12	0	2	2
266	rpiB	606	G	0	0	0	0	9	9	0	0	0
267	pgk	1588	G	0	9	9	0	0	0	0	0	0
268	sodA	803	P	0	4	4	0	2	2	0	0	0
269	folB	464	H	0	2	2	0	1	1	0	0	0
270	glmS	2407	M	0	26	26	0	7	7	0	1	1
271	glgA	1068	G	1	5	6	0	11	11	0	3	3
272	rnr	2542	K	6	11	17	0	11	11	0	4	4
273	coaD	614	H	0	3	3	0	1	1	0	0	0
276	serA	1246	HE	0	8	8	0	12	12	0	1	1
288	betalact	906	R	0	2	2			0			0

Appendix A (Continued)

289	menA	1186	H	0	3	3	0	4	4	0	0	0
290	menB	1079	H	0	14	14	0	3	3	0	1	1
293	eno	1682	G	0	11	11	0	11	11	0	4	4
294	rplQ	589	J	0	4	4	0	3	3	0	0	0
295	rpoA	1303	K	0	7	7	0	1	1	0	0	0
296	rpsD	803	J	0	1	1	0	9	9	0	1	1
297	rpsK	500	J	0	1	1	0	2	2	0	0	0
298	rpsM	495	J	0	1	1	0	2	2	0	2	2
300	infA	281	J	0	2	2	0	0	0	0	0	0
301	rplO	601	J	0	1	1	0	0	0	0	0	0
302	rpsE	671	J	0	11	11	0	1	1	0	0	0
304	rplF	712	J	0	8	8	0	3	3	0	0	0
305	rpsH	508	J	0	4	4	0	2	2	0	0	0
306	rpsN	351	J	0	4	4	0	1	1	0	1	1
307	rplE	726	J	0	7	7	0	0	0	0	0	0
308	rplX	324	J	1	2	3	0	1	1	0	1	1
310	rpsQ	366	J	0	1	1	0	1	1	0	0	0
312	rplP	541	J	1	3	4	0	7	7	0	0	0
313	rpsC	930	J	1	5	6	0	3	3	0	1	1
314	rplV	490	J	0	3	3	0	2	2	0	0	0
315	rpsS	371	J	0	1	1	0	0	0	0	0	0
316	rplB	1078	J	0	8	8	0	4	4	0	2	2
317	rplW	378	J	0	2	2	0	0	0	0	0	0
318	rplD	821	J	1	4	5	0	7	7	0	1	1
319	rplC	821	J	1	4	5	0	0	0	0	0	0
320	rpsJ	400	J	0	6	6	0	6	6	0	1	1
321	fusA	2770	J	1	20	21	0	19	19	0	0	0
322	rpsG	611	J	0	0	0	0	0	0	0	0	0
323	rpsL	499	J	0	2	2	0	0	0	0	0	0
324	map	1071	J	1	10	11	0	5	5	0	0	0
325	pgm	2008	G	1	11	12	0	2	2	0	0	0
326	tsf	1085	J	0	3	3	0	3	3	0	0	0
327	rpsB	954	J	0	5	5	0	7	7	0	0	0

Appendix A (Continued)

328	rpsI	504	J	0	4	4	0	4	4	0	1	1
329	rplM	587	J	0	0	0	0	3	3	0	0	0
330	dnaK	2476	O	1	6	7	0	13	13	0	0	0
331	psd	859	I	1	6	7	0	3	3	0	0	0
332	cdsA	1035	R	0	1	1	0	3	3	0	0	0
333	ftsH	2545	O	1	10	11	0	94	94	0	0	0
336	pyrG	2117	F	0	9	9	0	42	42	0	7	7
337	obgE	1286	R	0	3	3			0			0
338	adk	761	F	0	3	3	0	2	2	0	0	0
339	lpd	1823	C	0	4	4	0	7	7	0	0	0
343	dapA	1150	EM	0	10	10	0	3	3	0	2	2
346	miaB	1800	J	1	11	12	0	10	10	0	0	0
347	ntrC	1649	T	0	4	4	0	20	20	0	5	5
349	groES	372	O	0	3	3	0	10	10	0	0	0
350	groL	2133	O	2	9	11	0	2	2	0	0	0
351	hydrolase	668	R	0	1	1	0	4	4	0	1	1
355	yajC	415	U	0	3	3	0	10	10	0	0	0
356	coaE	804	H	0	7	7	0	6	6	0	2	2
359	entC	1374	Q	0	1	1	0	1	1	0	0	0
361	idi	698	I	0	1	1	0	12	12	0	3	3
362	sscR	538	H	0	5	5	0	0	0	0	0	0
364	matE	1699	V	0	6	6			0			0
365	recQ	2201	LKJ	0	7	7	0	5	5	0	0	0
366	aroC	1409	E	0	12	12	0	5	5	0	0	0
368	folC	1590	H	0	9	9	0	0	0	0	0	0
369	glnS	2178	J	2	11	13	0	8	8	0	3	3
370	rpoC	5515	K	0	33	33	0	14	14	0	0	0
371	rpoB	4955	K	0	19	19	0	3	3	0	1	1
372	rplL	487	J	1	0	1	0	1	1	0	0	0
373	rplJ	695	J	2	7	9	0	0	0	0	0	0
374	rplA	914	J	0	6	6	0	5	5	0	0	0
375	rplK	569	J	1	2	3	0	0	0	0	0	0
376	nusG	721	K	0	1	1	0	19	19	0	0	0

378	tuf	1547	J	0	7	7	0	19	19	0	1	1
379	lspA	794	N/A	0	4	4	0	5	5	0	1	1
380	ileS	4476	J	9	28	37	0	5	5	0	0	0
381	mutS	3298	L	0	1	1	0	7	7	0	0	0
384	rplS	451	J	0	4	4	0	2	2	0	0	0
385	phosphog	1807	G	0	4	4	0	2	2	0	0	0
386	dnaB	2008	L	1	12	13	0	2	2	0	0	0
387	accA	1232	I	0	6	6	0	17	17	0	2	2
388	sdhB	985	C	4	4	8	0	4	4	0	0	0
389	sdhA	2631	C	1	34	35	0	11	11	0	2	2
390	sdhC	863	N/A	0	4	4	0	10	10	0	4	4
391	iscS	1508	E	0	5	5	0	1	1	0	0	0
392	phosphod	1985	R	3	8	11	0	5	5	0	2	2
393	deoxy	1266	O	0	3	3	0	33	33	0	5	5
394	lpdA	1847	C	0	4	4	0	4	4	0	0	0
395	Fe4	444	N/A	0	3	3			0			0
397	aroE	998	E	0	2	2	0	3	3	0	0	0
398	rpsU	267	N/A	0	0	0	0	2	2	0	0	0
399	recG	2764	LK	1	8	9	0	9	9	0	0	0
400	uvrD	2824	L	2	21	23	0	42	42	0	0	0
403	murA	1706	M	0	18	18	0	8	8	0	0	0
405	hinT	536	FGR	1	2	3	0	1	1	0	0	0
406	yigI	998	K	0	10	10			0			0
410	feoB	2688	P	1	9	10	0	32	32	0	7	7
413	rluD	1326	J	1	2	3	0	5	5	0	0	0
415	leuS	3682	J	0	6	6	0	60	60	0	0	0
416	gdhA	1839	E	0	3	3	0	5	5	0	0	0
417	pyrH	928	F	0	4	4	0	6	6	0	0	0
418	frr	724	J	1	2	3	0	1	1	0	0	0
419	asnS	1851	J	1	14	15	0	5	5	0	1	1
420	rpoN	1909	K	0	9	9	0	7	7	0	1	1
422	sufE	567	R	0	4	4	0	3	3	0	0	0
425	rpmE	335	J	1	2	3	0	1	1	0	0	0

Appendix A (Continued)

426	fabH	1441	I	0	10	10	0	6	6	0	1	1
427	auxin	1970	N/A	2	11	13	0	22	22	0	3	3
428	rpsO	347	J	0	0	0	0	2	2	0	1	1
429	pnp	2793	J	1	23	24	0	94	94	0	1	1
430	rpoD	1127	K	3	9	12	0	1	1	0	0	0
431	tpiA	990	G	0	1	1	0	19	19	0	1	1
433	folP	1075	H	0	8	8	0	3	3	0	0	0
436	acoA	1307	C	0	7	7	0	1	1	0	0	0
437	aceF	1536	C	0	7	7	0	17	17	0	0	0
438	integral	995	P	0	2	2	0	2	2	0	0	0
446	upps	961	I	4	9	13	0	2	2	0	0	0
447	ppnK	1155	G	0	5	5	0	15	15	0	0	0
448	HPmadar528	938	J	0	2	2	0	4	4	0	1	1
451	alphabeta	1016	R	0	8	8	0	9	9	0	2	2
452	rnc	968	K	0	6	6	0	1	1	0	0	0
453	fabF	1643	IQ	1	10	11	0	46	46	0	0	0
454	acpP	340	IQ	0	2	2	0	8	8	0	0	0
456	ribC	779	H	0	5	5	0	3	3	0	2	2
457	pdxA	1362	H	0	2	2	0	9	9	0	0	0
458	atpC	320	C	0	2	2	0	0	0	0	0	0
459	atpD	1973	C	0	6	6	0	6	6	0	1	1
460	ribF	1207	H	0	5	5	0	22	22	0	0	0
462	tyrS	1648	J	0	6	6	0	2	2	0	0	0
464	prfB	1359	J	0	6	6	0	15	15	0	2	2
465	speA	1822	E	0	3	3	0	29	29	0	4	4
466	speB	1185	E	1	1	2	0	9	9	0	0	0
467	cmk	914	F	0	2	2	0	5	5	0	1	1
468	fabG	960	IQR	0	3	3	0	9	9	0	0	0
471	secA	4293	U	2	17	19	0	14	14	0	1	1
473	fmt	1244	J	1	6	7	0	8	8	0	0	0
475	pdxH	852	H	0	4	4	0	28	28	0	0	0
476	dnaN	1482	L	0	3	3	0	5	5	0	0	0
477	pheT	2771	J	1	14	15	0	8	8	0	0	0

Appendix A (Continued)

478	glyS	1910	J	2	6	8	0	2	2	0	0	0
479	trxB	1234	O	0	1	1	0	17	17	0	1	1
481	gapA	1318	G	0	9	9	0	26	26	0	1	1
482	lepA	2350	M	0	7	7	0	28	28	0	6	6
483	pth	772	J	1	1	2	0	1	1	0	0	0
484	sucD	1153	C	1	4	5	0	8	8	0	1	1
487	rpmH	197	N/A	0	2	2	0	0	0	0	0	0
490	rbn	1197	R	1	8	9	0	25	25	0	0	0
491	nfsA	442	N/A	0	0	0	0	4	4	0	3	3
492	rpmA	340	J	0	0	0	0	3	3	0	0	0
493	rplU	459	J	0	2	2	0	3	3	0	0	0
494	sufS	1593	E	1	5	6	0	3	3	0	0	0
496	sufC	984	O	0	4	4	0	7	7	0	0	0
498	sufA	431	S	0	5	5	0	7	7	0	0	0
500	lgt	1155	M	0	4	4	0	2	2	0	0	0
504	pdxJ	955	H	0	2	2	0	14	14	0	0	0
505	mscS	1639	M	0	15	15						0