Original Contribution

ACTA ZOOLOGICA BULGARICA
Acta zool. bulg., Suppl. 6, 2014: 29-34

Investigation of Endosymbiotic Bacteria Associated with Scale Insects of the Family Putoidae (Hemiptera: Coccoidea)

Matthew E. Gruwell, Zachary Duda, Joshua MacCready

Penn State Erie, School of Science, P-1 Prischak Building 4205 College Drive, Erie, PA 16563 USA.

Abstract: One of the less diverse families of scale insects is the Putoidae. Here we present additional data towards the characterization of bacteria associated with insects in the genus *Puto*. We include two phylogenetic estimates (16S rRNA and a multilocus sequence analysis of four protein-coding genes) of the relationships of these endosymbionts with other bacteria, both free living and endosymbiotic. Using five genes we once again fail to recover monophyly in bacteria associated with *Puto*, though all sequences are closely related, all are sister to insect endosymbionts and all are part of the Gammaproteobacteria. In previous analyses and in this project, bacteria associated with *Puto barberi* are not sister to the remaining bacteria found in *Puto*.

Key words: genomics, phylogeny, endosymbiont

Introduction

The family Putoidae is represented by a single extant genus, *Puto* Signoret, which has a fewer than 50 species, and by one extinct genus with a single species. For many years, the genus *Puto* was thought to be in the Pseudococcidae, related to the tribe Phenacoccinini, due to morphological similarity of the adult females. However, the superfamily Coccoidea is divided into two informal groups, the archaeococcoids (essentially the more "primitive" families) and the neococcoids, which includes all of the more "advanced" or derived families. Recent phylogenetic analysis using 18S (Gullan, Cook 2007), 28S and Ef1-alpha (Hardy *et al* 2008) have shown that *Puto* is not part of the Pseudococcidae and is possibly not even part of the neococcoids. Rather, it appears to be part of archaeococcoids and potentially the sister taxon to either the Ortheziidae (molecular data) or Phenacoleachiidae (male morphological data (Hodgson, Hardy, 2013)). Morphological and cytogenetic work has also presented the same conclusion by placing *Puto* in its own family based on a series of characteristics, such as an XX-XO chromosome system, lack of paternal genome elimination, and some key male characters (Hodgson, Foldi 2006, Williams *et al.* 2011, Hodgson, Hardy, 2013).

Little research has been done on the endosymbionts of Putoidae and the primary endosymbionts of this group have not been characterized. Histological work on endosymbionts of *Puto antennatus* Signoret was done by Buchner (1965) in which he described the transmission of symbionts in this group as one of the most unique systems of the insects he studied. When symbionts are being transferred, the cells of the bacteriome divide into two groups. One group of cells retains their ability to divide and are reduced in size in order to crowd around the ovarioles. Symbionts of this cell type retain their rodlet shape. The other bacteriocytes stop dividing but continue to grow in size and

acquire odd-shaped nuclei. In these cells, the symbionts degenerate and supposedly die off. Bacterial transmission to the next generation occurs from the first cell type. Once the ovarioles are developed, bacteriocytes migrate into folds of the follicle and are directly transmitted into the oocytes. Buchner points out the very interesting case of somatic cells entering the next generation with the possibility of immortality typically only seen in the germ line (Buchner 1965).

Two studies have previously published molecular work on the symbionts of Putoidae, and both have shown a bacterial association between Putoidae and the Gammaproteobacteria according the 16S rRNA gene. The first was presented in the *Proceedings of the International Symposium on Scale Insect Studies* (2004) by a single sample of an unidentified species of *Puto* as sister to a *Sodalis*-like symbiont from the rice weevil *Sitophilus oryzae,* also known as SOPE (*S. oryzae* primary endosymbiont) (Gruwell *et al.* 2005). The second report of molecular endosymbiont data confirmed the earlier result by showing an association of three *Puto* species (*P. albicans* McKenzie, *P. yuccae* Ferris, and *P. barberi* Ferris) with bacteria from the Gammaproteobacteria. In this case monophyly was not established, but sister taxa to these samples were similar to those presented previously (Gruwell *et al.* 2010). In this study, *Candidatus* Moranella endobia (from the Gammaproteobacteria) was included and was somewhat closely related to samples from *Puto.* [Convention dictates that endosymbiotic bacteria which cannot be cultured are written in normal type, proceeded by italicized *Candidatus*. However, following contemporary trends, henceforth endosymbiotic bacteria will simply be named as italicized genera and species, see Moran *et al*. (2008).] *Moranella* is a co-primary endosymbiont of mealybugs which lives inside the other primary endosymbiont, *Tremblaya princeps*. The close relationship between the Putoidae associated bacteria and *Moranella* in combination with recent data of a bacterium associated with Coccidae from this group, suggests the possibility of an ancient bacterial association of Gammaproteobacteria in the scale insects.

The purpose of this study was to improve upon the previous data, confirm monophyly among bacteria associated with *Puto* and establish a known association between a primary endosymbiont and the scale insect family Putoidae. Here we confirm a previous association between Putoidae and Gammaproteobacteria using additional species and data from five loci across the bacterial genome.

Materials and Methods

Taxon Sampling

New specimens were collected by Dug Miller into 100% ETOH and stored at -20°C until DNA was extracted. Other bacterial samples of endosymbionts and free-living taxa were included to provide reference for *Puto* species in multiple gene trees. Insect bacterial symbionts used in this analysis were selected based on their classification as symbiotic and the availability of a completely sequenced and annotated genome. In total, 14 completely sequenced symbiont genomes across three classes of bacteria were used for this study. The three classes represent the classes of bacteria from which scale insect endosymbionts are known. The class Gammaproteobacteria was the most represented with sequence data with 11 symbionts because of previous data linking this group to Putoidae symbionts. Bacteroidetes was represented by a single symbiont and Betaproteobacteria was represented by two samples of *Tremblaya princeps*.

A total of 19 completely sequenced and annotated genomes of free-living bacteria, each representing an order within the three bacterial classes, were used to infer the placement of the symbiont within the respective taxonomic order. This method allowed for a more accurate representation of insect symbiont taxonomy. The bacterium *Bacillus cereus B4264* (NC 011725.1) was used as an outgroup to anchor the completed trees. A complete list of all bacterial species used, corresponding abbreviations used in the trees, taxonomy and NCBI accession number is available in Table 1.

Gene Selection

Genes were selected based upon their presence in all 14 symbionts and ability to be sequenced consistently. A relatively low number of shared genes were identified across the chosen taxa due to the symbiont *Tremblaya princeps*, which has the smallest known bacterial genome (McCutcheon, von Dohlen, 2011). In total, 38 orthologs were identified, 36 protein coding and 2 rRNA. Genes used for the analysis include one molecular chaperone (groES), two ribosomal genes (rplK and rplB), one polymerase subunit (rpoB) and 16S rRNA (rrsA).

Table 1. List of bacteria and insect hosts of endosymbiont bacteria

Sample List	Species (or Host species)	Accession				
	New Data	16S	groES	rplB	rpoB	rplK
P. barberi	Undet hosted by *Puto barberi*	KJ857032	KJ857020	KJ857016	KJ857024	KJ857027
P. ech.	Undet hosted by *Puto echinocacti*	X	KJ857022	KJ857018	KJ857026	KJ857029
P. yuccae	Undet hosted by *Puto yuccae*	KH857031	KJ857021	KJ857017	KJ857025	KJ857028
P. sp	Undet hosted by *Puto species*	KJ857030	KJ857019	KJ857015	KJ857023	X
P. albicans	Undet hosted by *Puto albicans*	HM449988.1	X	X	X	X
	Sequences from Genome data	Class			Accession	
ROOT1	*Bacillus cereus*	Bacilli			NC_011725.1	
RO1	*Acidithiobacillus caldus*	Acidithiobacillia			NC_015850.1	
RO2	*Aeromonas hydrophila*	γ-Proteobacteria			NC_008570.1	
RO3	*Shewanella baltica*	γ-Proteobacteria			NC_009997.1	
RO4	*Dichelobacter nodosus*	γ-Proteobacteria			NC_009446.1	
RO5	*Thioalkalivibrio nitratireducens*	γ-Proteobacteria			NC_019902.1	
RO6	*Escherichia coli*	γ-Proteobacteria			NC_017626.1	
RO7	*Legionella pneumophila*	γ-Proteobacteria			NC_006368.1	
RO8	*Methylomicrobium alcaliphilum*	γ-Proteobacteria			NC_016112.1	
RO9	*Portiera aleyrodidarum*	γ-Proteobacteria			NC_018677.1	
RO10	*Haemophilus influenzae*	γ-Proteobacteria			NC_014922.1	
RO11	*Pseudomonas syringae*	γ-Proteobacteria			NC_007005.1	
RO12	*Francisella tularensis*	γ-Proteobacteria			NC_007880.1	
RO20	*Nitrosomonas eutropha*	β-Proteobacteria			NC_008344.1	
RO21	*Aromatoleum aromaticum*	β-Proteobacteria			NC_006513.1	
RO22	*Caulobacter segnis*	α-Proteobacteria			NC_014100.1	
RO23	*Magnetococcus marinus*	α-Proteobacteria			NC_008576.1	
RO32	*Ornithobacterium rhinotracheale*	Bacteroidetes			NC_018016.1	
RO33	*Saprospira grandis*	Bacteroidetes			NC_016940.1	
BA1	*Buchnera aphidicola* str *C. tujafilina*	γ-Proteobacteria			NC_015662.1	
BA2	*Buchnera aphidicola* BCc	γ-Proteobacteria			NC_008513.1	
CB1	*Blochmannia floridanus*	γ-Proteobacteria			NC_005061.1	
CB2	Blochmannia pennsylvanicus str. BPEN	γ-Proteobacteria			NC_007292.1	
CME	Moranella endobia PCIT	γ-Proteobacteria			NC_015735.1	
CP1	Portiera aleyrodidarum BT-B	γ-Proteobacteria			NC_018507.1	
CTP1	Tremblaya princeps PCIT	β-Proteobacteria			NC_015736.1	
CTP2	Tremblaya princeps PCVAL	β-Proteobacteria			NC_017293.1	
CUD	*Uzinura diaspidicola*	Bacteroidetes			NC_020135.1	
ECE	Secondary endosymbiont of *C. eucalypti*	γ-Proteobacteria			NC_018419.1	
EHC	Secondary endosymbiont of *H. cubana*	γ-Proteobacteria			NC_018420.1	
SG	Sodalis glossinidius str. *'morsitans'*	γ-Proteobacteria			NC_007712.1	
SS	Serratia symbiotica str. *'Cinara cedri'*	γ-Proteobacteria			NC_016632.1	
WG1	Wigglesworthia glossinidia str *G. brevipalpis*	γ-Proteobacteria			NC_004344.2	

Data Collection

DNA from new samples was extracted using a DNeasy extraction kit (Qiagen) following protocols established in the Normark Lab (Anderson *et al* 2010) in which a single specimen of each scale insect is punctured in a non-critical location to release DNA while preserving the exoskeleton for slide mounting. Genomic DNA was quantified on a Nanospec and stored at -20 C.

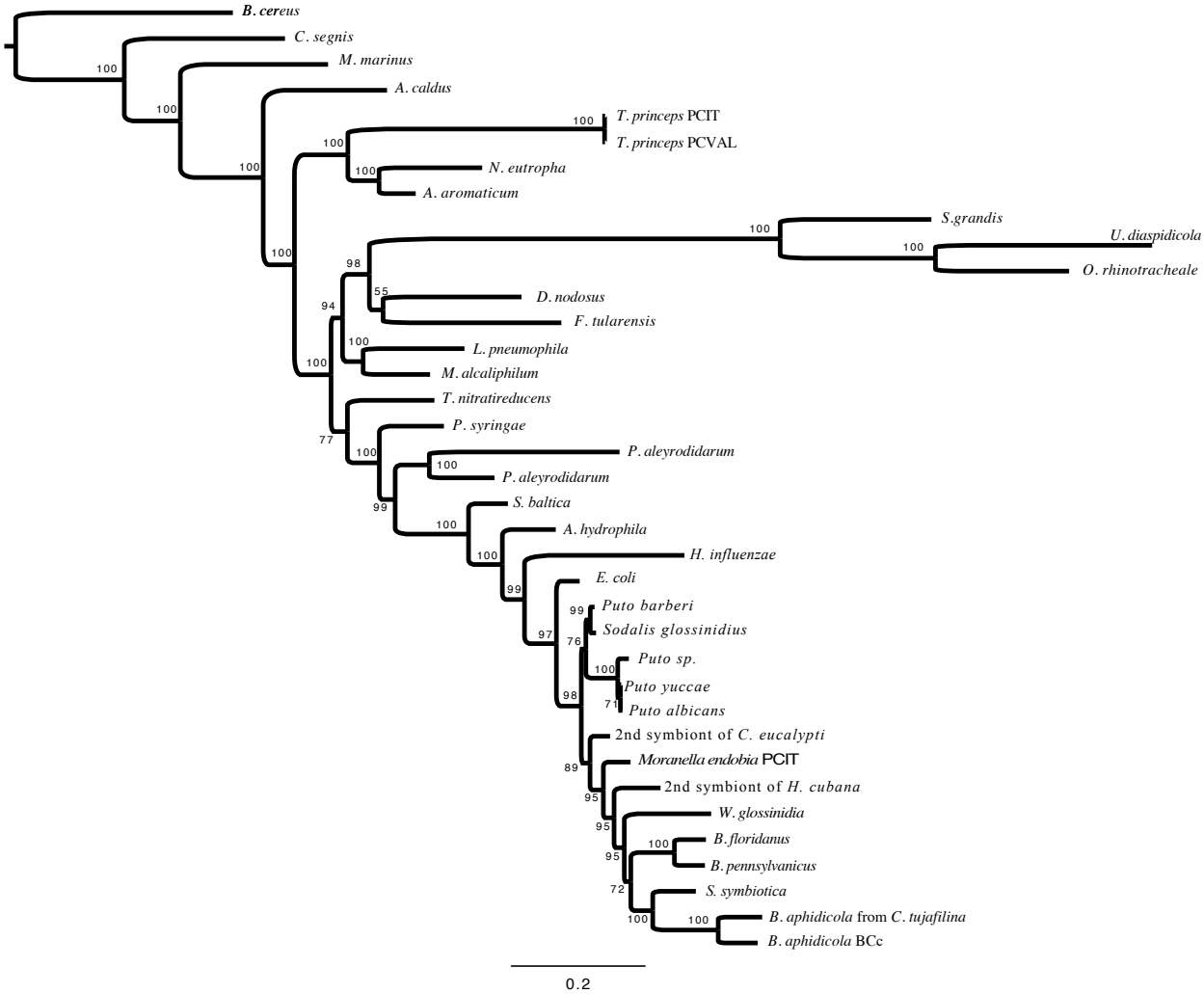

Fig. 1. Bayesian estimates of bacterial relationships using protein coding sequences of four genes, rplB rpoB, groES and rplK, from free living and endosymbiotic bacteria. Node numbers display Bayesian posterior probabilities

PCR was performed using AmpiTaq Gold Fast PCR Master Mix (Applied Biosystems). A typical 25 ul reaction consisted of 8-10 ul ultrapure H_2O, 12 ul PCR Master Mix, 1 ul each primer, 1-3 ul genomic DNA. Thermocycler conditions followed the suggested protocol utilizing very short cycle times: 95°C / 10min followed 35 cycles of 96° C / 3 sec, 45-60° C / 3 sec, 68° C / 5-30 sec depending on fragment length, followed by a final extension of 72° C / 10 sec and cooling to 10 C indefinitely. Primers for the individual genes were created new for the project with the exception of primers used to amply 16S rRNA which utilized general 16S primers provided by **Gruwell** et al (2010). Other genes were amplified using the following primer pairs: groES (groES_46_F GAAATCGAAGCCAAATCTGC and groES_272_R ATGTCGCTTTCGGACATGA), rplB (rplB_37_F CGTCGCCACGTTGTTAAAG and rplB_814_R TACGGCGACGTACGATGAAT), rpoB (rpoB_34_F CGTAAGGACTTCGGAAAACG and rpoB_1528_R GGGTATCCAGATCGCCTAAA), and rplK (rplK_48_F CATGGCAAACCCGAGTCC and rplK_381_R GCGAGACATCGCTTCCAC). For all primers, F denotes a forward primer and R denotes a reverse primer. Upon completion, PCR success was checked using agarose gel electrophoresis and successful reactions were sent to the High Throughput Genomics Center (Seattle, WA) for Sanger sequencing on an ABI 3730xl DNA Analyzer. All genes were sequenced twice utilizing each primer for double coverage to increase quality and lessen ambiguity.

Analysis

Sequence data was imported into Geneious v6.1 (Biomatters) where gene segments were assembled

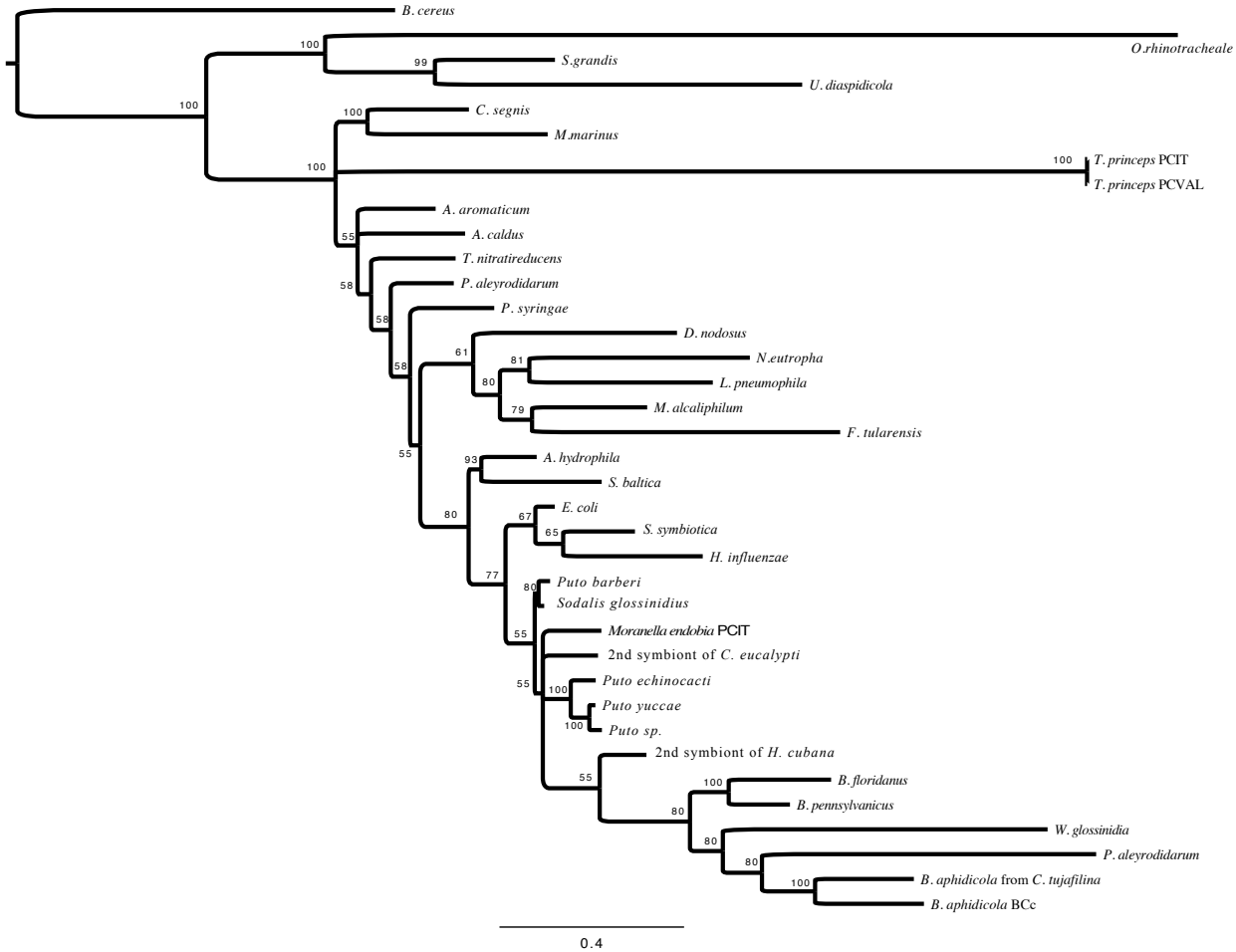

Fig. 2. Figure 1: Bayesian estimates of bacterial relationships using 16s rRNA sequences from free living and endosymbiotic bacteria. Node numbers display Bayesian posterior probabilities

and edited for accuracy. Each individual gene data set was aligned using the ClustalW tool in Geneious v6.1 with a gap open cost of 15 and gap extend cost 6.66 (default parameters). The resulting alignments were trimmed and edited. The protein coding genes data block consisted of 3014 aligned bp of DNA for 37 taxa, of which four taxa represent new data. The individual genes varied in length with rpoB being the longest at 1525 bp followed by rplB (857 bp), rplK (378 bp) and groES (295). The aligned 16s rRNA data set included 1717 aligned bp of DNA data. All protein coding genes had the same representation of *Puto* samples, but the 16S data set did not. Therefore all protein-coding genes were concatenated (groeES, rpoB, rplK and rplB) and two data sets were phylogenetically analyzed. Each gene alignment was independently run through jModelTest v.2.1.3 to statistically determine the best-fit model of nucleotide substitution. Settings for jModelTest were as follows: NumModels = 88, Base frequencies +F, Rate variation +I +G, Base tree for likelihood calculations = ML optimized and Tree topology search operation = BEST. For both data sets the optimized model was General Time Reversible (GTR) estimated by the log-likelihood ratio test.

Phylogenetic analysis for each dataset was independently performed through a parallel build of MrBayes v.3.2.1 (Ronquist, Hulsenbeck 2003) with the Markov chain Monte Carlo (MCMC) search technique for 10 million generations. A total of 4 chains (3 heated and one cold) were used and the analysis was stopped when the standard deviation of split frequencies converged to the accepted value of 0.01 (MrBayes Manual, page 9). For the likelihood model parameters (lset) of both datasets, the number of substitution types (nst) was set at 6 and the among-site variation (rates) was specified as invagamma. A relative burn-in with a fraction of 0.25 was used, discarding the first 25% of sampled trees.

Results

The resulting phylogenies for each data set are presented in Figure 1 (protein coding genes) and Figure 2 (16S). Bayesian Posterior Probabilities are reported on the branch nodes of both figures. The resulting phylogenies show that all Putoidae sequences are nested within Gammaproteobacteria, the closest relation being with *Sodalis glossinidius* from Tsetse flies, *Moranella endobia* from mealybugs and an unnamed secondary endosymbiont of psyllids. Interestingly, however, both data sets show a lack of monophyly for symbionts of Putoidae (Figure 1 and 2). In the phylogenies, three sequences form a *Puto* clade and the remaining sequence, from the species *Puto barberi,* is consistently sister to *Sodalis*, a closely related endosymbiont. This trend is seen in both phylogenies (Figure 1 and 2) and it is confirmed in every gene tested individually except groES. This gene had the shortest sequence length and grouped all *Puto* sequences as a monophyletic clade (data not shown).

Discussion

This represents the third study that has consistently amplified and sequenced 16S rRNA from a bacterium associated with Putoidae which has a phylogenetic affinity to Gammaproteobacteria (Gruwell *et al.* 2005, 2010). This result is confirmed with 4 additional protein-coding genes from the same bacterium showing phylogenetic placement in Gammaproteobacteria closely related to other endosymbiotic bacteria. These data were generated while failing to amplify or sequence any other bacterial sequences from genomic *Puto* DNA.

The sequences from bacteria hosted by *Puto barberi* remain something of a mystery and require more work in order to untangle their evolutionary origin. Phylogenetic and sequence data clearly demonstrate an affinity to *Sodalis glossinidius,* not shown in the other bacteria studied. Unlike most scale insect endosymbionts, *Sodalis* is not obligately intracellular, but it is nonetheless an endosymbiotic bacterium, and sequences from bacteria associated with *P. barberi* appear to be endosymbiotic, perhaps of a more distant evolutionary origin with a chance of horizontal transfer playing a role somewhere in this complicated puzzle. Additional samples of *Puto* and genome sequence of the *P. barberi* associate will prove useful in untangling this problem. With those data we will be able to definitively characterize the primary endosymbionts for insects in the genus *Puto*.

Acknowledgements: We are grateful to Dug Miller for providing specimens and help in beginning this project, and to Alexander Campbell and Kristi Gdanetz-MacCready for providing help in reading the manuscript and providing comments. We would also like to thank Chris Hodgson and an anonymous reviewer for helpful comments and improvements to the manuscript. M.E.G. would like to thank Ben Normark, Penny Gullan and Nate Hardy and the whole scale insect biologist community for providing continued encouragement, support and teaching M.E.G would also like to thank S. Gruwell for support. This research was funded by Bob Light and Undergraduate Grants from Penn State Erie, the Behrend College.

Refernces

Andersen, J. C., J. Wu, M. E. Gruwell, R. Gwiazdowski, S. E. Santana, N. Feliciano, G. E. Morse, B. B. Normark 2010. A phylogenetic analysis of armored scale insects (Hemiptera: Diaspididae), based upon nuclear, mitochondrial, and endosymbiont gene sequences. – *Molecular phylogenetics and evolution*, **57** (3): 992-1003.

Buchner, P. 1965. Endosymbiosis of Animals with Plant Microorganisms. Wiley, New York.

Gruwell, M. E., C. D. Von Dohlen, K. Patch, B. B. Normark 2005. Preliminary PCR survey of bacteria associated with scale insects (Hemiptera: Coccoidea). – In Proceedings of the Xth International Symposium on Scale Insect Studies, 19th, (101-116).

Gruwell M. E., N. B. Hardy, P. J. Gullan, K. Dittmar 2010 Evolutionary Relationships among primary endosymbionts of the mealybug subfamily Phenacoccinae (Hemiptera: Coccoidea: Pseudococcidae). – *Applied and Environmental Microbiology*, **76** (22): 7521-7525.

Gullan P. J., L. G. Cook 2007. Phylogeny and higher classification of the scale insects (Hemiptera: Sternorrhyncha: Coccoidea). – *Zootaxa*, **1668**: 413-425.

Hardy N. B. 2008. Systematic Studies of Scale Insects (Hemiptera: Coccoidea). ProQuest.

Hodgson C., I. Foldi 2006. A review of the Margarodidae *sensu* Morrison (Hemiptera: Coccoidea) and some related taxa based on the morphology of adult males. – *Zootaxa*, **1263:** 1-250.

Hodgson C. J., N. Hardy 2013 The phylogeny of the superfamily Coccoidea (Hemiptera: Sternorrhyncha) based on the morphology of extant and extinct macropterous males. – *Systematic Entomology*, **38**: 794-804.

McCutcheon, J. P., C. D. von Dohlen 2011. An interdependent metabolic patchwork in the nested symbiosis of mealybugs. – *Current Biology*, **21** (16): 1366-1372.

Moran N. A., J. P McCutcheon, A. Nakabachi, 2008. Genomics and evolution of heritable bacterial symbionts. – *Annual review of genetics*, **42**: 165-190.

Williams D. J., P. J. Gullan, D.R. Miller, D. Matile-Ferrero, and S. I. Han 2011. A study of the scale insect genera *Puto* Signoret (Hemiptera: Sternorrhyncha: Coccoidea: Putoidae) and *Ceroputo* Šulc (Pseudococcidae) with a comparison to Phenacoccus Cockerell (Pseudococcidae). – *Zootaxa*, **2802** (1): 22.

Ronquist, F., J. P. Huelsenbeck 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. – *Bioinformatics*, **19** (12): 1572-1574.