EVOLUTIONARY STABILITY OF FUNGAL-BACTERIAL ENDOSYMBIOSES

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Stephen James Mondo

August 2013

EVOLUTIONARY STABILITY OF FUNGAL-BACTERIAL ENDOSYMBIOSES

Stephen James Mondo, Ph. D.

Cornell University 2013

Many eukaryotes interact with heritable endobacteria to satisfy diverse metabolic needs. Of the characterized fungal-bacterial symbioses, endobacterial associations with the Gigasporaceae (Glomeromycota) and *Rhizopus microsporus* (Mucoromycotina) are the best described. Both fungal hosts associate with closely related bacterial endosymbionts from the *Burkholderia* lineage of β-proteobacteria. Through investigating patterns of co-divergence between partners, we have shown that the Glomeribacter-Glomeromycota symbiosis is at least 400 million years old, while still remaining non-essential for the host. To further explore what adaptations have taken place to allow for the persistence of this association, we created a computational pipeline which utilizes patterns of adaptation to infer microbial lifestyle. We show that this pipeline is effective at inferring microbial lifestyle, and that genes involved in DNA regulation, energy metabolism, and pathogenicity are likely important for survival of *Ca.* Glomeribacter within their fungal hosts. Additionally, we identified that non-essential endosymbionts are as effective at purging slightly deleterious mutations from their genomes as free-living organisms. Unlike Glomeribacter, *Burkholderia rhizoxinica*, the endosymbiont of *Rhizopus microsporus* is capable of free living yet is simultaneously of great importance to host survival. Our work has

revealed that endosymbionts are required for sexual reproduction of the fungal host. Through phenotypic observation and transcriptome profiling, we found that endosymbionts control fungal reproduction through hijacking of host reproductive machinery. Specifically, bacteria control expression levels of Ras2, a signaling protein important for reproductive development as well as filamentous growth. We also exploited endosymbiont control over reproduction to explore conservation of sexually relevant genes across Fungi, including the Mucoromycotina. This approach identified several genes that appear core to all fungal reproduction, as well as reproduction related genes which are specific to members of the Mucoromycotina. In particular, we found two candidate class C seven transmembrane G-protein coupled receptors (GPCRs), TriR1 and TriR2, which may be responsible for perception of trisporic acid during mating in Mucoromycotina. These receptors are closely related to the retinoic acid GPCRs present in animal systems.

BIOGRAPHICAL SKETCH

Stephen Mondo was born in 1985 in Larchmont, New York. In 2007, he received his Bachelors of Science in Ecology, Evolution and Behavioral Biology from the State University of New York at Binghamton. Here, he also received Honors in Biology for his work on Legume-Rhizobial interactions in Costa Rica. He then conducted research in the lab of Dr. Maria Harrison for one year studying plant genes involved in the arbuscular mycorrhizal symbiosis. In August, 2008, he started his graduate career in the department of Plant Pathology and Plant Microbe Biology and received his Ph.D. in August, 2013.

ACKNOWLEDGMENTS

I am grateful to all members of my committee, Esther Angert, Teresa Pawlowska, and Gillian Turgeon, for all of their help during my degree, as well as their useful comments on the chapters presented in this dissertation. I am especially thankful to my major advisor, Teresa Pawlowska, for her support and guidance on nearly all aspects of my dissertation research. I would also like to thank many other people for their assistance on various aspects of my research, including Kevin Toomer, Olga Lastovetsky, Catherine Fahey, Xiaotian Qin, Yan He, Minghui Wang, Choon-Lin Tiang, Qi Sun, Ylva Lekberg, Joseph Morton, Paola Bonfante, and Vivienne Gianinazzi-Pearson. I would also like to thank Haim Bar for assistance with statistics. Finally, I am extremely grateful to Colleen Cleere, as well as my parents, Peter and Heide Mondo, for motivating and supporting me throughout my Ph.D.

TABLE OF CONTENTS

doi:10.1111/j.1558-5646.2012.01611.x

EVOLUTIONARY STABILITY IN A 400-MILLION-YEAR-OLD HERITABLE FACULTATIVE MUTUALISM

Stephen J. Mondo,[1,2] Kevin H. Toomer,[1] Joseph B. Morton,[3] Ylva Lekberg,[4] and Teresa E. Pawlowska[1]

[1]Department of Plant Pathology & Plant Microbe-Biology, Cornell University, Ithaca, New York 14853–5904

[2]E-mail: tep8@cornell.edu

[3]Division of Plant and Soil Sciences, West Virginia University, Morgantown, West Virginia 26506–6108

[4]MPG Ranch, Missoula, Montana 59802 and Department of Ecosystem and Conservation Sciences, University of Montana, Missoula, Montana 59812

Received September 27, 2011
Accepted February 10, 2012

Many eukaryotes interact with heritable endobacteria to satisfy diverse metabolic needs. Some of these interactions are facultative symbioses, in which one partner is not essential to the other. Facultative symbioses are expected to be transitional stages along an evolutionary trajectory toward obligate relationships. We tested this evolutionary theory prediction in *Ca.* Glomeribacter gigasporarum, nonessential endosymbionts of arbuscular mycorrhizal fungi (Glomeromycota). We found that heritable facultative mutualisms can be both ancient and evolutionarily stable. We detected significant patterns of codivergence between the partners that we would only expect in obligate associations. Using codiverging partner pairs and the fungal fossil record, we established that the Glomeromycota–Glomeribacter symbiosis is at least 400 million years old. Despite clear signs of codivergence, we determined that the Glomeribacter endobacteria engage in recombination and host switching, which display patterns indicating that the association is not evolving toward reciprocal dependence. We postulate that low frequency of recombination in heritable endosymbionts together with host switching stabilize facultative mutualisms over extended evolutionary times.

KEY WORDS: **Arbuscular mycorrhizal fungi, *Ca.* Glomeribacter gigasporarum, cospeciation, endosymbiosis, host switching, recombination.**

Most eukaryotes satisfy their various metabolic needs through interactions with bacteria (Sachs et al. 2011). Evolutionary theory predicts that reciprocal selection and close spatial association of the partners will over time lead to coupling of partner metabolic and reproductive interests through endosymbiont vertical transmission, eventually producing a relationship in which the partners cannot live without each other (Yamamura 1993; Frank 1995; Law and Dieckmann 1998). This prediction implies that countless facultative symbioses with heritable bacteria, in which one of the partners is not essential to the other, represent short-lived transitory evolution stages along the trajectory leading to reciprocal partner dependence (Fig. 1A). Support for the transitory nature of facultative endosymbioses comes from two observations: (1) many obligate endosymbioses, including symbiotic organelles in eukaryotic cells (Margulis 1981) and endobacteria essential to nutrition of insects (Moran et al. 2008), are of great evolutionary age, whereas (2) examples of ancient facultative endosymbioses are few (Baumann 2005; Moran et al. 2008; Degnan et al. 2010). However, evolutionary histories of facultative endosymbioses are exceedingly difficult to reconstruct because phylogenies of the partners are often incongruent with each other (Bright and Bulgheresi 2010). Consequently, the apparent shortage of old facultative endosymbioses may be related to impediments in estimating the evolutionary history and the age of these associations rather than to their ephemeral nature. Our goal was to test the hypothesis of the transitory nature of heritable

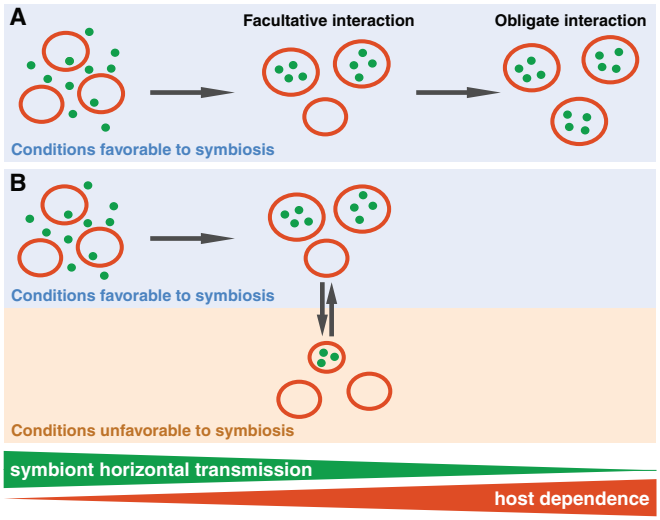

Figure 1. **Hypothetical evolutionary trajectories in heritable mutualisms. Hosts are depicted as red ovals; endosymbionts are shown as green dots. Relative host fitness is reflected by the size of ovals. (A) Evolutionary trajectory leading to obligate reciprocal partner dependence. (B) Shifting environmental conditions are expected to arrest an association at the facultative dependence stage. If conditions remain unfavorable for prolonged periods of time, host populations would be expected to completely lose endosymbionts.**

facultative endosymbioses. We identified several aspects of partner life histories and evolution that needed to be unraveled to achieve this goal, including (1) the extent of dependency between the partners, (2) the role of recombination and host switching in the history of the association, and (3) the time of evolutionary origin of the association. (1) The extent of the dependency between the partners defines whether the association is facultative or obligate. It can be uncovered by studying the distribution of symbionts among closely related hosts. (2) Recombination and host switching interfere with progressive coevolution between the partners (Frank 1994; Bright and Bulgheresi 2010). Evaluation of the role that these two processes played in the history of the association is critical to resolve whether the partners are in the process of evolving toward mutual dependence or whether the association is locked in a facultative mode. (3) The evolutionary origin of the association is needed to discern whether it is a recently formed relationship or an ancient symbiosis. The time of origin can be established based on fossil record. If fossil evidence is available for one partner only, a pattern of codivergence between the partners is required to infer the initial association event and date it.

Endosymbiont transmission mode is a key life-history feature that defines its population structure and evolutionary history. Strict vertical transmission of essential endosymbionts makes uncovering evolutionary origins and histories of their symbiotic associations relatively straightforward because hosts and symbionts develop clear patterns of codivergence (Bright and Bulgheresi 2010). In addition, vertical transmission has dramatic effects on genomic architecture of the endosymbiont, including increased fixation of slightly deleterious mutations that results in molecular evolution rate acceleration and genome size reduction relative to free-living taxa (Moran et al. 2008). These features can be attributed to the powerful role that genetic drift plays in endosymbiont populations (Moran 1996). Drift is enhanced in these populations because of their exceedingly small effective sizes brought about by clonal reproduction, population subdivision, and recurrent population bottlenecks. The patterns of molecular evolution associated with small effective sizes are witnessed nearly universally in essential endosymbionts of insects despite their divergent origins (Moran et al. 2008), revealing how strongly transmission mode influences evolution.

Although vertical transmission can be extensive in nonessential endobacteria, they do not share highly similar patterns of molecular evolution (Dale and Moran 2006; Moran et al. 2008). This is likely due to the fact that most nonessential endosymbionts maintain the capacity for both vertical and horizontal transmission as well as recombination (Toh et al. 2006; Degnan and Moran 2008; Degnan et al. 2010). A balance between these processes is expected to heavily contribute to the genomic variability observed across nonessential endosymbionts, as genetic drift is reduced through horizontal transmission and recombination, while enhanced during phases of vertical transmission. Through symbiont mixing and replacement, recombination and host switching also have the capacity to obscure codivergence patterns that could be used to make inferences about the age and history of an association.

To test the hypothesis that facultative endosymbioses are transitive, we studied an association of arbuscular mycorrhizal (AM) fungi from the phylum Glomeromycota (Smith and Read 2008) and *Ca*. Glomeribacter gigasporarum (Bianciotto et al. 1996; Bianciotto et al. 2003) representing the *Burkholderia* lineage of beta-proteobacteria. Among diverse heritable symbioses, the Glomeromycota–Glomeribacter association provides an ideal system for testing the transitory nature of heritable facultative associations. Glomeromycota are one of the ancient fungal lineages (Heckman et al. 2001) with a rich fossil record (Remy et al. 1994; Redecker et al. 2000; Dotzler et al. 2006) that can be used to date their evolutionary history. The entire phylum appears to be asexual (Smith and Read 2008) and therefore, the evolutionary history of these fungi is unlikely to be confounded by frequent gene exchanges (Croll and Sanders 2009; Rosendahl et al. 2009; den Bakker et al. 2010). Glomeromycota colonize roots of the vast majority of terrestrial plants and improve plant mineral nutrient uptake from the soil in exchange for plant-assimilated

carbon (Smith and Read 2008). This energy dependence on plant hosts may be a source of significant selective pressure on Glomeribacter endosymbionts, which, in turn, may contribute to enhanced resolution of coevolution patterns in the Glomeromycota–Glomeribacter association. The Glomeribacter endobacteria improve hyphal growth of their fungal host after spore germination and prior to plant root colonization (Lumini et al. 2007). This developmental effect is related to Glomeribacter's ability to modify metabolism of lipids that are contained inside the fungal spores and fuel spore germination and presymbiotic hyphal proliferation (Salvioli et al. 2010). As Glomeromycota are horizontally transmitted obligate biotrophs, their success in root colonization and securing plant hosts depends on their ability to proliferate presymbiotically (Smith and Read 2008). The extent of hyphal proliferation needed to secure a plant host is largely dependent on environmental factors (Klironomos and Moutoglis 1999).

Glomeribacter endobacteria have been detected so far only in the Gigasporaceae family of Glomeromycota (Bianciotto et al. 2003). They are vertically transmitted by their AM fungal hosts and can be cleared from fungal cells through serial subculturing in vitro (Lumini et al. 2007). Although the endobacteria are not cultivable, they are capable of surviving for several weeks outside of the fungal host, albeit without replication (Jargeat et al. 2004). The genome sequence of Glomeribacter revealed that these endobacteria are energy dependent on their fungal host (Ghignone et al. 2011). It is not surprising, therefore, that Glomeribacter exhibit evolutionary patterns typical for vertically transmitted endobacteria, including an accelerated rate of molecular evolution relative to free-living taxa (Castillo and Pawlowska 2010) and a rather small 1.72 Mb genome (Ghignone et al. 2011). However, rRNA and protein-coding genes of Glomeribacter show very little accumulation of deleterious mutations (Castillo and Pawlowska 2010) that would be expected in heritable endosymbionts (Lambert and Moran 1998). The pattern of acceleration of the molecular evolution rate, combined with limited accumulation of deleterious mutations relative to free-living taxa, suggest that, even though vertical transmission has played a large role in the evolutionary history of this lineage, Glomeribacter have the capacity to at least partially mitigate the negative effects of genetic drift associated with vertical transmission.

In the present study, we reconstructed the evolutionary history of the symbiosis between *Ca.* Glomeribacter gigasporarum and Glomeromycota to resolve whether these bacteria are relatively new associates of Glomeromycota on the trajectory toward becoming essential endosymbionts, or whether this association is ancient and permanently locked in its current facultative state. Using the host fossil record and codiverging partner pairs, we inferred that the symbiosis of Glomeromycota with Glomeribacter is at least 400 million years old. We show that unlike many facultative endosymbioses, which have rapidly evolved into obligate associations, the association between AM fungi and Glomeribacter is maintained in a facultative state through a balance of vertical transmission, recombination in endosymbiont populations, and likely host switching.

Materials and Methods

DIVERSITY OF GLOMERIBACTER IN AM FUNGI

Globally distributed populations representing the taxonomic diversity of the Gigasporaceae family were obtained from the International Culture Collection of Vesicular AM Fungi (INVAM), West Virginia University, Morgantown, West Virginia, the International Bank for the Glomeromycota (BEG), INRA, Dijon, France, and Ylva Lekberg collection (Table S1). Rather than a single-spore isolate, each of the accessions represents an experimental population of morphologically similar fungi collected at a specific sampling location and maintained as a live culture. To understand how endobacterial diversity varied within and among fungal populations, we sampled individual fungal spores (isolates) from each accession.

AM FUNGAL SPORE DECONTAMINATION

We developed a new spore surface decontamination technique. Spores were rinsed in 0.05% Tween 20 and subjected to two 15 min rinses in H_2O_2, the first at 1 mM concentration and the second at 50 mM (Imlay et al. 1988; Imlay and Linn 1988), followed by a 20 min soak in 4% chloramine T before final rinsing in nanopure water (three 20 min soaks). All spores were processed individually to minimize cross-contamination. Wash controls were set up for three spores per species by placing intact surface-decontaminated spores in illustra^{TM} GenomiPhi-V2 Whole Genome Amplification (WGA) sample buffer (GE Healthcare, Piscataway, NJ), vigorous vortexing, and removing the spores. WGA of wash controls followed by polymerase chain reaction (PCR) with universal bacterial 16S rRNA gene primers 704f and 1495r (Bianciotto et al. 1996) confirmed the efficacy of spore decontamination.

MOLECULAR MARKERS

Following surface decontamination, total DNA of individual spores was globally amplified using illustra^{TM}GenomiPhi-V2 kit (GE Healthcare). WGA products were diluted 1 to 20 in water for subsequent PCR reactions. Specific PCR primers were designed for bacterial loci encoding the 16S rRNA, 23S rRNA, FtsZ, and PstA (Tables 1 and 2). To ensure PCR amplification of divergent Glomeribacter genotypes, the 23S rRNA gene primers were designed and verified to successfully amplify not only

Table 1. Molecular markers sampled in Glomeromycota and in Glomeribacter.

Gene	Taxon	Fragment length (bp)	Annotation
18S rRNA	Glomeromycota	516	Small ribosomal subunit
28S rRNA	Glomeromycota	664	Large ribosomal subunit
Beta-tubulin	Glomeromycota	671	Microtubule structural component
16S rRNA	*Ca*. Glomeribacter gigasporarum	930	Small ribosomal subunit
23S rRNA	*Ca*. Glomeribacter gigasporarum	553	Large ribosomal subunit
ftsZ	*Ca*. Glomeribacter gigasporarum	483	Structural binary fission protein
pstA	*Ca*. Glomeribacter gigasporarum	695	Phosphate transport system permease protein

Glomeribacter, but their sister lineage *Burkholderia rhizoxinica* as well (data not shown). The 16S and 23S rRNA genes are present in one copy in the Glomeribacter genome (Ghignone et al. 2011). The rRNA, *ftsZ*, and *pstA* genes are all located in different contigs in the Glomeribacter draft genome assembly. To amplify fungal loci encoding 18S rRNA, 28S rRNA, and beta-tubulin, we used a combination of published and our own primers (Table 2). PCR amplifications targeting bacterial loci were carried out in 50-μl reactions containing 25 μl REDTaq PCR Reaction Mix (Sigma-Aldrich, St. Louis, MO), 21.8 μl water, 0.6 μl of each 10 μM primer, and 2 μl of diluted WGA product under the following cycling conditions: initial denaturation of 2 min at 94°C, 10 touchdown cycles of 30 sec at 94°C, 30 sec at 59–1°C per cycle, and 1 min at 72°C followed by 25 cycles of 30 sec at 94°C, 30 sec at 49°C, and 1 min at 72°C with final extension of 7 min at 72°C. For the fungal 18S rRNA and beta-tubulin genes, we conducted touchdown PCR with target annealing temp of 49 and 52°C, respectively. Conventional PCR was done for the fungal 28S rRNA genes at an annealing temp of 57°C. All other conditions were identical to those in bacterial PCR amplifications. Amplicons were purified using the QIAquick 96 PCR Purification Kit (Qiagen, Valencia, CA), cycle sequenced with BigDye version 3.1 (Applied Biosystems, Carlsbad, CA), purified using Performa Ultra 96-well plates (Edge BioSystems, Gaithersburg, MD), and sequenced on the Automated 3730 DNA Analyzer (Applied Biosystems). For the *pstA* gene in endobacteria from *S. pellucida*, amplicons were cloned using TOPO-TA Cloning Kit for Sequencing (Invitrogen, Carlsbad, CA). Recombinant plasmid DNAs were amplified with illustra^{TM}TempliPhi kit (GE Healthcare), and sequenced using the T3 and T7 primers.

Marker sequences were edited in Sequencher 4.9 (Gene Codes Corporation, Ann Arbor, MI) and aligned with MUSCLE

Table 2. Primers for PCR amplification of molecular markers in Glomeromycota and in Glomeribacter.

Gene	Taxon	Primer[1]	Oligonucleotide sequence	*Tm*[2]	Reference
18S rRNA	Glomeromycota	NS1	GTAGTCATATGCTTGTCTC	47.2	White et al. (1990)
18S rRNA	Glomeromycota	NS2	GGCTGCTGGCACCAGACTTGC	63.9	White et al. (1990)
28S rRNA	Glomeromycota	LR1	GCATATCAATAAGCGGAGGA	46.7	Van Tuinen et al. (1998)
28S rRNA	Glomeromycota	NDL22	TGGTCCGTGTTTCAAGACG	55.2	Van Tuinen et al. (1998)
Beta-tubulin	Glomeromycota	Btub16f	ATGATGTGCACGTTTTCTGT	52.9	This study
Beta-tubulin	Glomeromycota	2iR	GTGAAGACGTGGGAAAGGAAC	55.9	Msiska and Morton (2009)
16S rRNA	Glomeribacter	SSU534f	GGTAAATAATCGGAGTGGAT	49.2	This study
16S rRNA	Glomeribacter	SSU709f*	GCATACGTGACTGGCAGACT	57.2	This study
16S rRNA	Glomeribacter	SSU1520r*	TCTGGTAAAACTCACTCCC	51.3	This study
23S rRNA	Glomeribacter	GlomGiGf*	GGGTCCATTGCGGATTACTTC	55.9	Bianciotto et al. (2004)
23S rRNA	Glomeribacter	GlomGiGr	GGGACCAGGACTTCCATCCCCC	63.8	Bianciotto et al. (2004)
23S rRNA	Glomeribacter	LSU483r*	GGTGCAGGAATATTAACC	47.6	This study
ftsZ	Glomeribacter	FtsZ354f	GTGGTCTCAAAGCCGTTT	53.1	This study
ftsZ	Glomeribacter	FtsZ857r	CCATCGCATCATCGTAGA	51.5	This study
ftsZ	Glomeribacter	FtsZ16fPell*	CTTGAGGAGCATGTGGATTC	53.3	This study
ftsZ	Glomeribacter	FtsZ944rPell*	GTGTGGAGCAAGGTCATGG	56.1	This study
pstA	Glomeribacter	PstA13f	CTGCTATGGCTCGCGTGGAT	60.0	This study
pstA	Glomeribacter	PstA727r	GCGAGCGTATTCAGGCCCAG	60.9	This study

[1]In the case of the bacterial 16S rRNA, 23S rRNA, and *ftsZ* genes, primer combinations used to amplify *S. pellucida* endobacteria are denoted with an asterisk.
[2]*Tm* values in °C were determined by Integrated DNA Technologies, Inc. (Coralville, IA).

(Edgar 2004). Bayesian and maximum likelihood (ML) multigene phylogenies were constructed in MrBayes version 3.1.2 (Ronquist and Huelsenbeck 2003) and RAxML version 7.0.4 (Stamatakis 2006), respectively, using the Generalized Time Reversible (GTR) nucleotide substitution model (Tavaré 1986) plus invariant sites (I) and Γ rate heterogeneity. Genes were partitioned separately, allowing for independent models of sequence evolution across loci. The Markov chain Monte Carlo (MCMC) analysis was run for 10 million generations. A total of 1000 bootstrap replicates were performed for trees constructed with RAxML (Stamatakis 2006). Placement of the *Ca*. Glomeribacter gigasporarum lineage in the bacterial phylogeny was examined using reference organisms with complete genomes available at GenBank (accession numbers in Fig. 2). To root the fungal tree, we used *Glomus versiforme* BEG47 (AJ132666, FJ461852, FJ174286). Sequences generated in this study were deposited in GenBank under accession numbers JF816707–JF817217.

CODIVERGENCE ANALYSES

The Permutation test (Hommola et al. 2009) and the ParaFit test (Legendre et al. 2002) were used to determine the overall signal of codivergence within the dataset. Both these methods test the null hypothesis of independent host and symbiont evolution. The Permutation test is based on correlation between host and parasite genetic distances (Hommola et al. 2009). The ParaFit method uses a similar approach but transforms host and symbiont patristic distances derived from their phylogenetic trees into matrices of principal coordinates (Legendre et al. 2002). In addition to testing the global null hypothesis, ParaFit can be also used to examine the contribution of each individual host–symbiont pair to the overall codivergence signal. As these tests require that only partner pairs are included in the analysis, no outgroups or host fungi without endosymbionts were analyzed. To generate patristic distances for ParaFit, unrooted Bayesian multigene phylogenies were constructed for hosts and symbionts only (Fig. S1) using MrBayes version 3.1.2 (Ronquist and Huelsenbeck 2003). The MCMC analysis was run for 10 million generations using the GTR + I + Γ nucleotide substitution model. Principal coordinates were generated from patristic distances using the DistPcoA program (Legendre and Anderson 2002). At least 10,000 permutations were conducted for each codivergence test. For each host–endosymbiont pair, *P*-values were determined by ParaFitLink1 (Legendre et al. 2002). Tanglegrams in Figure S1 were generated with Dendroscope 3 (Huson et al. 2007). For consistency, we use the term "codivergence" to describe patterns of significant parallel diversification at the level of host and symbiont populations. "Cospeciation" is reserved for patterns detected within a single host genus, whereas "cophyly" refers to patterns apparent across genera.

DETECTION OF RECOMBINATION IN GLOMERIBACTER

The Genetic Algorithm for Recombination Detection (GARD) (Kosakovsky Pond et al. 2006) available through a web interface (Delport et al. 2010) was used to identify recombination breakpoints in both fungal and Glomeribacter genes. GARD is a likelihood-based model selection procedure that uses a genetic algorithm to search multiple sequence alignments for evidence of recombination breakpoints (Kosakovsky Pond et al. 2006). The alignments of concatenated sequences, including the rRNA genes, were searched for evidence of segment-specific phylogenies under the GTR+Γ nucleotide substitution model (Tavaré 1986). Goodness of fit was assessed using the small sample corrected AIC (AIC_C) derived from an ML model fit to each segment. The Kishino–Hasegawa (KH) method (Kishino and Hasegawa 1989) was applied to evaluate congruence of the segment-specific topologies.

To measure the extent of recombination in the Glomeribacter population, concatenated sequences of *ftsZ* and *pstA* genes were examined with ClonalFrame (Didelot and Falush 2007). ClonalFrame reconstructs the clonal genealogy of a sample of individuals, as well as the mutation and recombination events that took place on the branches of this genealogy, based on a coalescent model (Vos and Didelot 2009). Ribosomal RNA genes were excluded from this analysis because recombination signal could be confounded by nucleotide substitution rate heterogeneity often observed in rRNA gene evolution (Vos and Didelot 2009). Run specifications followed those of Vos and Didelot (2009) with the MCMC run of 200,000 generations after an initial 200,000 burn-in. Convergence was assessed using the Gelman–Rubin statistic (Gelman and Rubin 1992) with a cutoff of 1.1.

Bacterial individuals for which all genes could not be PCR-amplified were excluded from both GARD (Kosakovsky Pond et al. 2006) and ClonalFrame (Didelot and Falush 2007) analyses. Only one representative per accession was used for both tests as endobacterial diversity did not vary within an accession.

GLOMERIBACTER HETEROGENEITY IN AM FUNGAL INDIVIDUALS

To assess genetic heterogeneity of Glomeribacter in AM fungal individuals, the 23S rRNA gene fragments were PCR amplified from individual spores using primers GlomGiGf (Bianciotto et al. 2004) and LSUb-483r (Table 2) in 50-μl reactions with RED-Taq (Sigma) under cycling conditions of initial denaturation of 2 min at 94°C, 25 cycles of 30 sec at 94°C, 30 sec at 49°C, and 1 min at 72°C followed by final extension of 7 min at 72°C. Amplicons were cloned and sequenced as described above. At least 16 recombinant colonies were analyzed per spore.

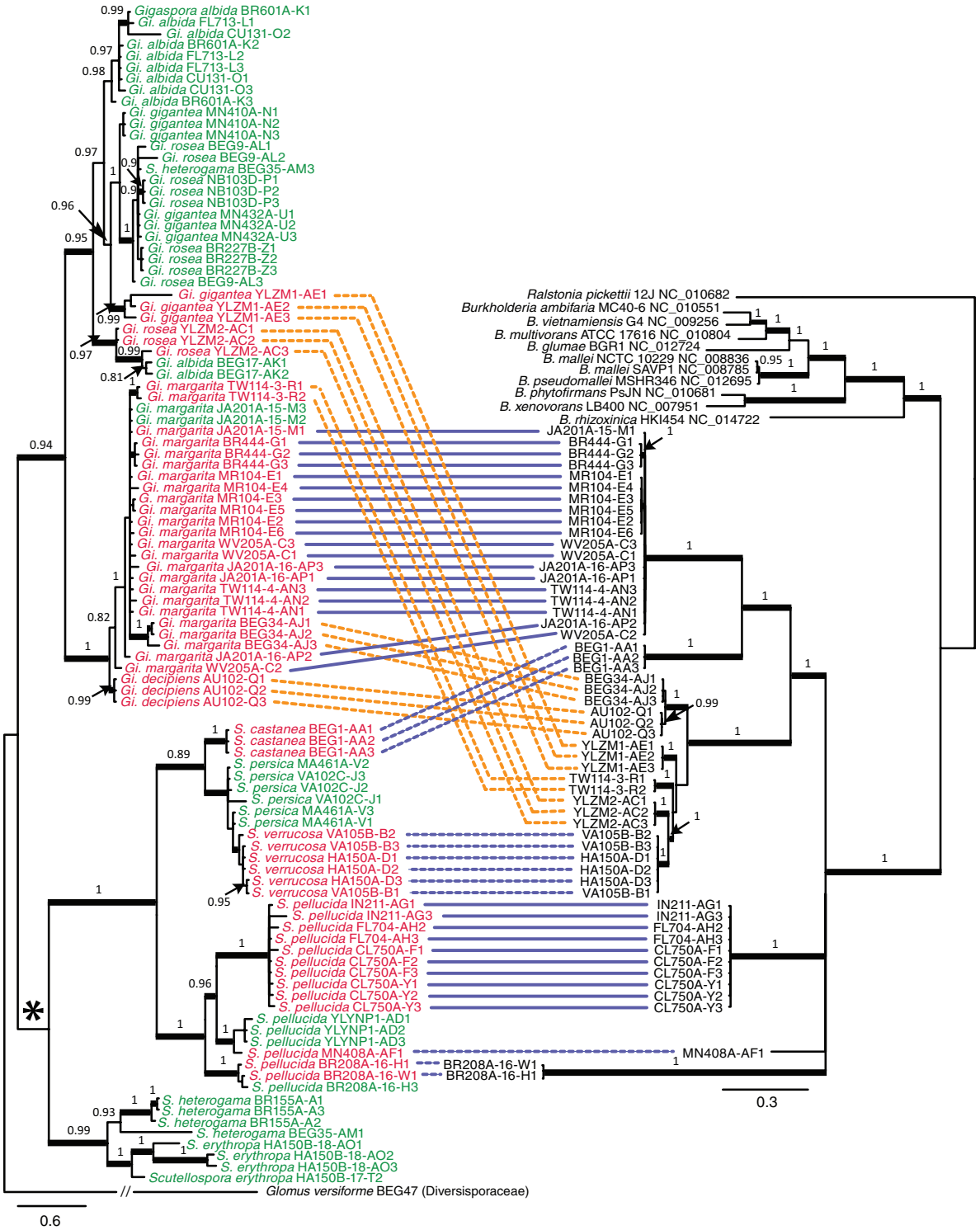

Figure 2. Patterns of coevolution between the Gigasporaceae fungal hosts (left) and the *Ca*. Glomeribacter gigasporarum bacterial endosymbionts (right). The fungal phylogeny was reconstructed using 18S rRNA, 28S rRNA, and beta-tubulin gene sequences; the bacterial phylogeny is based on 16S rRNA, 23S rRNA, *ftsZ*, and *pstA* genes. Bayesian posterior probabilities greater than 0.80 are shown above branches. Branches with ML bootstrap support over 70% are thickened. Solid blue lines connecting host and symbiont pairs indicate significant evidence of codivergence detected by ParaFit in concatenated datasets of fungi and bacteria (Tables 3 and S2; Fig. S1). Dashed blue lines link partners showing evidence of codivergence after recombination was accounted for in the bacterial dataset. Dashed orange lines indicate partner pairs with no evidence of codivergence, likely due to host switch events. Fungal isolates harboring endobacteria are colored red; individuals with no endobacteria are shown in green. The node likely associated with the *Scutellospora* Rhynie chert fossil record (Dotzler et al. 2006) is marked by an asterisk.

Results

DIVERSITY OF GLOMERIBACTER ENDOBACTERIA IN AM FUNGI

The association between Glomeromycota and *Ca*. Glomeribacter gigasporarum appears to be facultative for the fungus, as it can survive and reproduce after removal of the endobacteria under laboratory conditions (Lumini et al. 2007). However, it has been unclear whether the association between Glomeribacter and Glomeromycota is also nonessential for the fungus in nature, given the significance of presymbiotic growth to the fitness of Glomeromycota. To test the hypothesis that the endobacteria are nonessential for the fungus in nature, we examined the distribution of Glomeribacter among isolates of Glomeromycota. We surveyed 115 isolates from 34 globally distributed experimental populations representing the Gigasporaceae family of Glomeromycota (Table S1). Fungal individuals were characterized by PCR amplification and sequencing of the 18S rRNA, 28S rRNA, and beta-tubulin genes, whereas bacterial endosymbionts were identified by PCR amplification and sequencing of the 16S rRNA, 23S rRNA, *ftsZ*, and *pstA* genes (Tables 1 and 2). Only 18 of the surveyed populations harbored Glomeribacter (Fig. 2). Endobacteria presence varied even within closely related fungal groups, indicating that very similar fungal genotypes can survive with or without endobacteria (Fig. 2). For example, although most *Scutellospora pellucida* populations harbored Glomeribacter, endobacteria were not detected in any individual sampled from the YLYNP1 population. Moreover, presence of endobacteria was not consistent within populations. For instance, only two of the three spores sampled from *S. pellucida* BR208A contained Glomeribacter. Similarly, endosymbionts were only present in four of the six spores sampled from the JA201A population of *Gigaspora margarita*. Altogether, these data support the hypothesis that the Glomeribacter endobacteria are not essential for the survival of their fungal hosts.

To understand the evolution of the Glomeribacter bacteria, we reconstructed their phylogeny using 16S rRNA, 23S rRNA, *ftsZ*, and *pstA* gene sequences. We found that Glomeribacter form a monophyletic clade nested within the *Burkholderia* genus, with endobacteria associated with *S. pellucida* being the most similar to free-living *Burkholderia* (Fig. 2). This pattern suggests that the ability to associate with Glomeromycota evolved in the *Burkholderia* lineage only once.

GLOMERIBACTER CODIVERGENCE WITH GLOMEROMYCOTA

To explore the history of the association between Glomeromycota and the Glomeribacter endobacteria and examine how closely Glomeribacter have coevolved with their host, we measured the degree of phylogenetic codivergence of the partners using the Permutation test (Hommola et al. 2009) and the ParaFit test (Legendre et al. 2002) on sequences of concatenated fungal genes encoding 18S rRNA, 28S rRNA, and beta-tubulin and concatenated endobacterial genes encoding 16S rRNA, 23S rRNA, FtsZ, and PstA (Tables 1 and 2). Using these markers, the phylogenies of both host and endosymbiont were well resolved and strongly supported (Figs. 2 and S1), and thus the codivergence patterns are unlikely to be obscured by topological uncertainties. However, internal conflict generated by recombination (if present) may still affect codivergence patterns. Both tests rejected the null hypothesis of independent evolution of the Gigasporaceae hosts and the Glomeribacter endobacteria (the Permutation test $P < 0.0001$ and the ParaFit test $P = 0.0001$). The ParaFit test, which uses phylogenetic distances to detect the overall codivergence signal within the dataset, also uses these distances to identify which individual host–endosymbiont associations significantly contribute to the codivergence signal. In the case of Gigasporaceae and Glomeribacter, the test revealed that the global signal of codivergence was not due to uniformly codiverging host and endosymbiont populations, but instead was generated mostly by several significant contributions from relatively few lineages (*Gi. margarita* and *S. pellucida*; Tables 3 and S2; Figs. 2 and S1). This finding indicates that other processes, such as recombination and host switching, are likely interfering with codivergence between these partners.

GLOMEROMYCOTA ARE LOCKED IN FACULTATIVE DEPENDENCE ON GLOMERIBACTER

Recombination and host switching can interfere with progressive coevolution between the partners by disrupting reciprocally adapted suites of alleles (Frank 1994). Consequently, evaluating the contributions of recombination and host switching in the history of a symbiosis is likely to reveal whether the association is in the process of transitioning toward reciprocal dependence. These two processes introduce incongruence between host and symbiont phylogenies and disrupt patterns of codivergence.

To evaluate the contributions of recombination and host switching to partner coevolution, we searched for breakpoints in the fungal host and in the Glomeribacter sequence data and then reassessed the patterns of codivergence by comparing the evolutionary histories of partner gene sequences partitioned into segments defined by the recombination breakpoints. We used the GARD (Kosakovsky Pond et al. 2006), followed by the KH test (Kishino and Hasegawa 1989) to identify recombination breakpoints. No evidence of recombination was detected in concatenated 18S rRNA, 28S rRNA, and beta-tubulin gene sequences of fungal hosts harboring endosymbionts. This finding is consistent with the expectation that Gigasporaceae are asexual. In contrast, the GARD detected three possible recombination breakpoints in a set of concatenated bacterial 16S rRNA, 23S rRNA, *ftsZ*, and

Table 3. Population-level evidence of codivergence between Glomeromycota and Glomeribacter represented as a majority-rule consensus of the ParaFit test results for each fungal isolate (complete data are in Table S2). Patristic distances were based on phylogenetic trees generated from fungal and bacterial sequence data without partitions as well as bacterial data partitioned into segments defined by recombination breakpoints (Fig. S1). MS = marginally significant ($P \leq 0.07$); *$P \leq 0.05$; **$P \leq 0.01$; ***$P \leq 0.001$; **** $P \leq 0.0001$; — = not significant ($P > 0.07$).

Fungal population	Bacterial concatenated gene sequence alignment positions			
	All	1–1085	1086–1905	1906–2693
Gi. decipiens AU102-23	–	–	–	–
Gi. gigantea YLZM1	–	–	–	–
Gi. margarita BEG34	–	–	–	–
Gi. margarita BR444-22	***	***	**	*
Gi. margarita JA201A-15	***	***	**	MS
Gi. margarita JA201A-16	***	**	***	MS
Gi. margarita MR104-5	***	***	**	MS
Gi. margarita TW114-3	–	–	–	–
Gi. margarita TW114-4	***	**	**	MS
Gi. margarita WV205A-24	***	***	***	*
Gi. rosea YLZM2	–	–	–	–
S. castanea BEG1	–	***	–	–
S. pellucida BR208A-16-H	****	–	****	****
S. pellucida BR208A-16-W	****	–	****	***
S. pellucida CL750A-13-F	****	***	****	****
S. pellucida CL750A-13-Y	****	****	****	****
S. pellucida FL704-14	****	***	****	****
S. pellucida IN21-7	****	***	****	****
S. pellucida MN408A-16	****	–	****	****
S. verrucosa HA150A-16	–	–	*	–
S. verrucosa VA105B-16	–	–	*	–

pstA gene sequences at alignment positions 1085 (the 5′-end of the 23S rRNA gene), 1617 (the 5′-end of the *ftsZ* gene), and 1905 (the 3′-end of the *ftsZ* gene). The AIC_C score of 13,851.6 for the best-fitting model allowing for different topologies of the alignment segments defined by these breakpoints was lower than the AIC_C score of 13,937.5 for the model that assumed the same topology for all segments, indicating that a multiple tree model is preferable over a single tree model. Using the KH test, only the 1085 ($P < 0.05$) and 1905 ($P < 0.01$) breakpoints were identified as resulting in significant topological incongruence between segments.

The GARD results suggested that overall codivergence patterns based on the concatenated bacterial dataset may be obscured by multiple recombination events. Therefore, based on these recombination breakpoints, we partitioned the Glomeribacter dataset into three alignment segments (Fig. S1) and conducted the ParaFit test for each of these partitions separately. We found that the endobacteria were engaging in recombination and likely in host switching (Tables 3 and S2). Recombination was evidenced by disappearance of the codivergence signal in some of the partitions of data that showed codivergence without partitioning (*S. pellucida* BR208A and *S. pellucida* MN408A), and by appearance of codivergence signal in partitions of data that did not show codivergence without partitioning (*S. castanea* and *S. verrucosa*; Figs. 2 and S1). A lack of codivergence across all data partitions observed in *Gi. decipiens*, *Gi. gigantea*, *Gi. margarita* BEG34, *Gi. margarita* TW114–3, and *Gi. rosea* was interpreted as an indication of host switching. Although this is the most parsimonious interpretation (Brooks and McLennan 1991), it is also possible, although unlikely, that the lack of codivergence across all data partitions was caused exclusively by recombination. Interestingly, after recombination breakpoints were considered, it became apparent that all members of the *Scutellospora* genus that harbor endosymbionts (*S. pellucida*, *S. castanea*, and *S. verrucosa*) show evidence of cospeciation. However, patterns of recombination and host switching were evident even within *Gi. margarita* and *S. pellucida*, which, for the most part, harbored perfectly codiverging endosymbionts. The fact that recombination and host switching occur even in predominantly codiverging clades indicates that the partners still maintain the flexibility to associate with different genotypes of each other despite the establishment of long-term associations. Therefore, we conclude that

the Gigasporaceae-Glomeribacter association is not in transition toward reciprocal dependence.

The finding of recombination and host switching throughout the Glomeribacter population further implies that genetically different strains have similar effects on their hosts. Such functional interchangeability of Glomeribacter genotypes suggests that retaining the capacity for horizontal transmission in this highly coevolved association is likely to have an adaptive role. We speculate that if the host costs of maintaining the association outweigh the benefits of improved presymbiotic hyphal growth imparted by endobacteria, the endobacteria can be lost from the hosts. If conditions change, the endobacteria can be reacquired. Horizontal transmission offers a mechanism that could restore the association, stabilize it in fluctuating environments, and thus contribute to its evolutionary antiquity.

RECOMBINATION AND MIXING IN GLOMERIBACTER POPULATION

Genetic drift associated with small effective population sizes in vertically transmitted endobacteria is the underlying cause of their genomic degeneration (Moran et al. 2008, 2009). Recombination can mitigate the negative effects of vertical transmission. To determine whether the extent of recombination in the Glomeribacter population is sufficient to protect their genomes from accumulation of deleterious mutations, we examined concatenated sequences of *ftsZ* and *pstA* genes using the ClonalFrame software (Didelot and Falush 2007), which quantifies recombination frequency in a population. We modeled Glomeribacter genealogy and estimated the rate of recombination relative to the rate of mutation in the population (ρ/θ). We also computed the per-site effect of recombination relative to mutation, that is, the ratio of rates at which nucleotides become substituted as a result of recombination versus mutation (r/m). The values measured for Glomeribacter were $\rho/\theta = 0.1$ with a 95% confidence interval of 0.01–0.39 and $r/m = 2.5$ with a 95% confidence interval of 0.44–6.30. The ρ/θ value suggests that recombination events were less frequent than mutations in the history of Glomeribacter. However, because each recombination event is likely to introduce several nucleotide polymorphisms, the per-site effect of recombination r/m was on average over twofold larger than the effect of mutation. Such impact of recombination on nucleotide diversity is considered to be moderately high and is comparable to that found in many free-living bacteria (Vos and Didelot 2009). Therefore, it appears that the recombination frequency in the Glomeribacter population is sufficient to moderate accumulation of deleterious mutations in its genome (Castillo and Pawlowska 2010). This level of gene exchange is also consistent with the impact that recombination has on codivergence between Glomeribacter and Glomeromycota.

To assess whether endosymbiont mixing, suggested by the recombination and host switching data, can also be detected in the Glomeribacter populations contained within host individuals, we cloned and sequenced 23S rRNA gene fragments from endobacteria within single AM fungal spores representing populations CL750A-13, VA105B-16, WV205A-24, and YLZM2. Although several single nucleotide polymorphisms were present within all but the VA105B-16 dataset, the observed polymorphism level was minimal and not different from that generated by PCR error alone, and thus, it did not provide evidence to reject Glomeribacter clonality within any AM fungal individual. Consequently, we found that, despite the signature of recombination and evidence of host switching, Glomeribacter populations associated with individual hosts remain genetically uniform, which, in turn, suggests a powerful role of transmission bottlenecks in the life history of these endosymbionts. Population bottlenecks are responsible for reducing effective sizes of endosymbiont populations and exacerbating the effects of genetic drift, which are evidenced in Glomeribacter by the accelerated rate of molecular evolution relative to free-living taxa (Castillo and Pawlowska 2010).

THE ANTIQUITY OF ASSOCIATION BETWEEN GLOMERIBACTER AND GLOMEROMYCOTA

As fossil information is available for the Gigasporaceae lineage of Glomeromycota (Dotzler et al. 2006), evidence of codivergence can be used to estimate the age of the association between AM fungi and the Glomeribacter endobacteria. Because the ParaFit (Legendre et al. 2002) approach utilizes phylogenetic distances to assess whether the host and the symbiont are diverging from their relatives at similar rates, the codivergent host–symbiont pairs can be used to infer the phylogenetic node where the initial codivergence event must have occurred. When the confounding effects of recombination were accounted for (Fig. S1), the ParaFit test revealed that all members of the *Scutellospora* lineage harboring endosymbionts (*S. castanea*, *S. pellucida*, *S. verrucosa*) show evidence of cospeciation with their bacterial partners (Tables 3 and S2; Figs. 2 and S1). Moreover, cophyly with endobacteria is apparent in the *Scutellospora* and *Gigaspora* hosts, as several isolates of *Gi. margarita* show significant codivergence with Glomeribacter. This pattern places the first association event between AM fungi and Glomeribacter at the base of the Gigasporaceae family tree (Fig. 2), indicating that the Glomeribacter endobacteria were present in the Glomeromycota lineage prior to the Gigasporaceae speciation.

As no true cospeciation was detected within the *Gigaspora* genus (Tables 3 and S2; Figs. 2 and S1), the hypothesis that endobacteria codivergent with *Gi. margarita* are the result of a more recent host switch from *Scutellospora* is worth considering. However, we believe that this scenario is unlikely for several reasons. First, other host switches including those involving a few *Gi. margarita* isolates were readily detected within our dataset. Second, the pattern of cophyly that we observed indicates that

divergence levels across host genera are similar to those apparent across the endosymbiont phylogeny. Third, the endosymbiont evolutionary history, reconstructed using both concatenated data and data partitioned at recombination breakpoints, shows differentiation into few deeply divergent clades, including a clade associated with *Gi. margarita* (Figs. 2 and S1). Some of these deep clades expand into shallow terminal branches. Such a divergence pattern is expected in endosymbionts with a lengthy history of codivergence and occasional host switches, and would not be likely in a young symbiosis (Ronquist 2003). Instead, a recent association characterized by a series of host switches would be evidenced by a more contemporaneous proliferation of endosymbiont clades.

The origin of the Glomeromycota–Glomeribacter symbiosis prior to the Gigasporaceae radiation allows us to place a lower bound on the age of this association. The oldest Gigasporaceae fossils, which closely resemble present-day spore structures of the genus *Scutellospora*, were found in the Rhynie chert dated to 396 $\pm$ 12 Mya (Dotzler et al. 2006). In conjunction with our codivergence results, the existence of this fossil record establishes that the association between Gigasporaceae and Glomeribacter is at least 400 million years old. However, as the radiation of the Gigasporaceae family likely predates the fossil record, the association is expected to be even older. This ancient origin suggests that the symbiosis between Glomeromycota and Glomeribacter existed in the mycorrhizal associations that AM fungi formed with the Early Devonian land plants (Remy et al. 1994), which might have significantly contributed to plants' ability to colonize land.

Discussion

Our analyses revealed that Glomeromycota are facultatively associated with *Ca*. Glomeribacter gigasporarum endobacteria and have been locked in this state for at least 400 million years. Of the 34 fungal populations surveyed, only 18 harbored endobacteria and showed strong overall evidence of codivergence. This global signal was due to large contributions from two fungal species, *S. pellucida* and *Gi. margarita*. However, when recombination breakpoints in the endosymbiont sequences were considered in the analyses (Fig. S1), cospeciation signals were detected across all members of the *Scutellospora* genus harboring endosymbionts, and codivergence was consistently observed in one member of the *Gigaspora* (*Gi. margarita)*. These codivergence patterns indicate that the association between Glomeromycota and Glomeribacter is maintained through a balance between vertical transmission, recombination, and host switching. Given the great evolutionary age of this association, it is likely that the rate of horizontal transmission in Glomeribacter exceeds the critical level needed to preserve endosymbiont fitness over time (O'Fallon 2008). Even though the means of horizontal transmission in Glomeribacter are unknown, the partner codivergence patterns are not the only indication of its incidence. Recombination and host switching are also supported by the observation that, in a highly reduced genome, the Glomeribacter endobacteria retain and express genes encoding most of the recombination machinery as well as genes required for the type III secretion system, which in other bacteria is involved in invasion of and persistence in eukaryotic hosts (Ghignone et al. 2011). The patterns of symbiont mixing apparent in the Glomeribacter population combined with the antiquity of Glomeribacter's association with Glomeromycota provide evidence that not all facultative endosymbioses represent transitory stages along the trajectory toward obligate reciprocal dependence predicted by the evolutionary theory (Yamamura 1993; Frank 1995; Law and Dieckmann 1998). Instead, the Glomeromycota–Glomeribacter symbiosis shows that long-term associations can exist in the absence of strict vertical transmission and host dependence. This finding offers an incentive for identification and critical evaluation of conditions that keep facultative symbioses from evolving toward reciprocal partner dependence.

Unlike associations in which endosymbionts provision their hosts with essential nutrients required by the host for survival, such as amino acids needed by phloem sap-feeding insects (Baumann 2005), nonnutritional endosymbioses are often not consistently beneficial (Russell and Moran 2005). This is almost certainly true for facultative endosymbioses that insects form with *Ca*. Hamiltonella defensa and *Ca*. Regiella insecticola, in which endosymbionts protect their hosts from parasite pressure (Russell and Moran 2005). Given that the cost of supporting Glomeribacter may be a significant part of the fungal host carbon budget and that the extent of hyphal proliferation needed to secure a plant host is largely dependent on environmental factors (Klironomos and Moutoglis 1999), we speculate that Glomeribacter endobacteria resemble these protective insect endosymbionts by benefiting their fungal hosts only under conditions where extensive presymbiotic hyphal development is needed to colonize plant roots. In all these symbioses, the fluctuation of costs and benefits across environments would make obligate host dependence less likely (Genkai-Kato and Yamamura 1999). Nevertheless, long periods of vertical transmission and resulting genetic drift will still propel endosymbionts toward obligate dependence on their hosts, making these interactions superficially appear to be transitioning between associative states. However, with hosts occasionally losing their endosymbionts, the benefits gained by endobacteria that maintain the capacity for horizontal transmission and recombination would remain high. Thus, we speculate that under conditions of high environmental variability, facultative endosymbioses will be maintained in their present state and are less likely to evolve toward obligate reciprocal dependency (Fig. 1B).

Patterns of molecular evolution in *Ca*. Hamiltonella defensa and *Ca*. Regiella insecticola may support the hypothesis that environmental variability locks endosymbioses in a facultative state. Similar to Glomeribacter, Hamiltonella and Regiella harbor largely reduced genomes of 2.1 and 2.07 Mb, respectively (Degnan et al. 2009; Degnan et al. 2010). In addition, Hamiltonella shows evidence of recombination (Degnan and Moran 2008; Degnan et al. 2010). However, with the possible exception of a subclade of aphids where Hamiltonella appear universally (Degnan and Moran 2008), both systems show no cospeciation between partners (Russell et al. 2003). The overall lack of cospeciation could potentially be due to a recent origin of these associations, enhanced horizontal transmission rates, or high recombination. However, divergence levels between Hamiltonella and Regiella indicate that these lineages have likely been around for as long as *Buchnera aphidicola*, an essential endosymbiont of aphids that is at least 60–200 million years old (Degnan et al. 2010). Consequently, the low cospeciation levels in these associations are most likely due to horizontal transmission in Regiella and horizontal transmission combined with recombination in Hamiltonella (Degnan and Moran 2008). This observation suggests that the protective associations of insects formed with Hamiltonella and Regiella, in addition to being evolutionarily old (Degnan and Moran 2008; Degnan et al. 2010), may also be stable in their facultative state.

Although the majority of completely obligate interactions are of great age, such interactions may transition toward reciprocal dependence over relatively short time spans. Notably, *Ca*. Tremblaya princeps, an essential endosymbiont of mealybugs (Pseudococcidae) (Baumann 2005) and close relative of Glomeribacter, was postulated to be as young as 40 million years (Moran et al. 2008). Even though these two symbioses are markedly different, development and generation times are comparable between mealybugs (Chong et al. 2008) and AM fungi (Smith and Read 2008). It therefore seems that sufficient time must have elapsed for the Gigasporaceae-Glomeribacter association to become obligate, yet it still maintains its facultative nature. This observation is concordant with estimates of the relative rates of 16S rRNA gene evolution in Glomeribacter versus other *Burkholderia* (Castillo and Pawlowska 2010). Although evolving 1.5 times faster than free-living relatives, Glomeribacter is evolving 2.3 times slower than Tremblaya (Castillo and Pawlowska 2010). This disparity in evolution rates between Glomeribacter and other *Burkholderia* can now be explained by our finding of recombination and host switching within the Glomeribacter population. Both of these processes are expected to result in larger effective population sizes (N_e) than those in the essential endosymbionts such as Tremblaya, whereas prolonged periods of vertical transmission are expected to force N_e to remain smaller than in free-living bacteria.

Conclusion

Our analysis of the association between *Ca*. Glomeribacter gigasporarum and Glomeromycota revealed that not all facultative associations with heritable endobacteria represent transitory stages along the trajectory toward obligate reciprocal dependence predicted by evolutionary theory (Yamamura 1993; Frank 1995; Law and Dieckmann 1998). The patterns of molecular evolution in the Gigasporaceae-Glomeribacter association indicate that this facultative symbiosis has been maintained over 400 million years through a balance of vertical transmission, recombination, and host switching and it is not evolving toward a mutually obligate relationship. We speculate that associations in which host costs and benefits fluctuate with shifting environmental pressures are less likely to evolve toward reciprocal obligate dependence and endosymbiont strict vertical transmission in comparison to associations exposed to relatively stable forms of selection generated, for example, by specialized nutritional demands of the host (Fig. 1). Preserving the capacity for horizontal transmission in endosymbionts under variable selective conditions permits recolonization of host lineages that may have lost their partners due to environmental change. Similarly, retaining the ability to exchange genes is adaptive in such environments, as it protects endosymbionts from losing the capacity for horizontal transmission due to genomic degeneration caused by accumulation of deleterious mutations in populations of small effective sizes. Low frequency of recombination in heritable endosymbionts together with host switching may stabilize facultative mutualisms over extended evolutionary times.

ACKNOWLEDGMENTS

We thank V. Gianinazzi-Pearson for fungal isolates from the International Bank for the Glomeromycota, and E. Angert, P. Bonfante, and N. VanKuren for comments on the manuscript. We are grateful to two anonymous reviewers for their insightful suggestions. This work was supported by the National Science Foundation (DEB-0918880).

LITERATURE CITED

Baumann, P. 2005. Biology of bacteriocyte-associated endosymbionts of plant sap-sucking insects. Annu. Rev. Microbiol. 59:155–189.

Bianciotto, V., C. Bandi, D. Minerdi, M. Sironi, H. V. Tichy, and P. Bonfante. 1996. An obligately endosymbiotic mycorrhizal fungus itself harbors obligately intracellular bacteria. Appl. Environ. Microb. 62: 3005–3010.

Bianciotto, V., A. Genre, P. Jargeat, E. Lumini, G. Becard, and P. Bonfante. 2004. Vertical transmission of endobacteria in the arbuscular mycorrhizal fungus *Gigaspora margarita* through generation of vegetative spores. Appl. Environ. Microb. 70:3600–3608.

Bianciotto, V., E. Lumini, P. Bonfante, and P. Vandamme. 2003. '*Candidatus* Glomeribacter gigasporarum' gen. nov., sp nov., an endosymbiont of arbuscular mycorrhizal fungi. Int. J. Syst. Evol. Microbiol. 53:121–124.

Bright, M., and S. Bulgheresi. 2010. A complex journey: transmission of microbial symbionts. Nat. Rev. Microbiol. 8:218–230.

Brooks, D. R., and D. A. McLennan. 1991. Phylogeny, ecology, and behavior: a research program in comparative biology. Univ. of Chicago Press, Chicago, IL.

Castillo, D. M., and T. E. Pawlowska. 2010. Molecular evolution in bacterial endosymbionts of fungi. Mol. Biol. Evol. 27:622–636.

Chong, J. H., A. L. Roda, and C. M. Mannion. 2008. Life history of the mealybug, *Maconellicoccus hirsutus* (Hemiptera: Pseudococcidae), at constant temperatures. Environ. Entomol. 37:323–332.

Croll, D., and I. R. Sanders. 2009. Recombination in *Glomus intraradices*, a supposed ancient asexual arbuscular mycorrhizal fungus. BMC Evol. Biol. 9:13.

Dale, C., and N. A. Moran. 2006. Molecular interactions between bacterial symbionts and their hosts. Cell 126:453–465.

Degnan, P. H., T. E. Leonardo, B. N. Cass, B. Hurwitz, D. Stern, R. A. Gibbs, S. Richards, and N. A. Moran. 2010. Dynamics of genome evolution in facultative symbionts of aphids. Environ. Microbiol. 12:2060–2069.

Degnan, P. H., and N. A. Moran. 2008. Evolutionary genetics of a defensive facultative symbiont of insects: exchange of toxin-encoding bacteriophage. Mol. Ecol. 17:916–929.

Degnan, P. H., Y. Yu, N. Sisneros, R. A. Wing, and N. A. Moran. 2009. *Hamiltonella defensa*, genome evolution of protective bacterial endosymbiont from pathogenic ancestors. Proc. Natl. Acad. Sci. USA 106:9063–9068.

Delport, W., A. F. Y. Poon, S. D. W. Frost, and S. L. Kosakovsky Pond. 2010. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. Bioinformatics 26:2455–2457.

den Bakker, H. C., N. W. VanKuren, J. B. Morton, and T. E. Pawlowska. 2010. Clonality and recombination in the life history of an asexual arbuscular mycorrhizal fungus. Mol. Biol. Evol. 27:2474–2486.

Didelot, X., and D. Falush. 2007. Inference of bacterial microevolution using multilocus sequence data. Genetics 175:1251–1266.

Dotzler, N., M. Krings, T. N. Taylor, and R. Agerer. 2006. Germination shields in *Scutellospora* (Glomeromycota: Diversisporales, Gigasporaceae) from the 400 million-year-old Rhynie chert. Mycol. Prog. 5:178–184.

Edgar, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792–1797.

Frank, S. A. 1994. Genetics of mutualism: the evolution of altruism between species. J. Theor. Biol. 170:393–400.

———. 1995. The origin of synergistic symbiosis. J. Theor. Biol. 176:403–410.

Gelman, A., and D. B. Rubin. 1992. Inference from iterative simulation using multiple sequences. Stat. Sci. 7:457–511.

Genkai-Kato, M., and N. Yamamura. 1999. Evolution of mutualistic symbiosis without vertical transmission. Theor. Popul. Biol. 55:309–323.

Ghignone, S., A. Salvioli, I. Anca, E. Lumini, G. Ortu, L. Petiti, S. Cruveiller, V. Bianciotto, P. Piffanelli, L. Lanfranco, et al. 2011. The genome of the obligate endobacterium of an AM fungus reveals an interphylum network of nutritional interactions. ISME J. 6:136–145.

Heckman, D. S., D. M. Geiser, B. R. Eidell, R. L. Stauffer, N. L. Kardos, and S. B. Hedges. 2001. Molecular evidence for the early colonization of land by fungi and plants. Science 293:1129–1133.

Hommola, K., J. E. Smith, Y. Qiu, and W. R. Gilks. 2009. A permutation test of host-parasite cospeciation. Mol. Biol. Evol. 26:1457–1468.

Huson, D. H., D. C. Richter, C. Rausch, T. Dezulian, M. Franz, and R. Rupp. 2007. Dendroscope: An interactive viewer for large phylogenetic trees. BMC Bioinform. 8:460.

Imlay, J. A., S. M. Chin, and S. Linn. 1988. Toxic DNA damage by hydrogen-peroxide through the fenton reaction *in vivo* and *in vitro*. Science 240:640–642.

Imlay, J. A., and S. Linn. 1988. DNA damage and oxygen radical toxicity. Science 240:1302–1309.

Jargeat, P., C. Cosseau, B. Ola'h, A. Jauneau, P. Bonfante, J. Batut, and G. Bécard. 2004. Isolation, free-living capacities, and genome structure of "*Candidatus* Glomeribacter gigasporarum," the endocellular bacterium of the mycorrhizal fungus *Gigaspora margarita*. J. Bacteriol. 186:6876–6884.

Kishino, H., and M. Hasegawa. 1989. Evaluation of the maximum-likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. J. Mol. Evol. 29:170–179.

Klironomos, J. N., and P. Moutoglis. 1999. Colonization of nonmycorrhizal plants by mycorrhizal neighbours as influenced by the collembolan, *Folsomia candida*. Biol. Fertil. Soils. 29:277–281.

Kosakovsky Pond, S. L., D. Posada, M. B. Gravenor, C. H. Woelk, and S. D. W. Frost. 2006. Automated phylogenetic detection of recombination using a genetic algorithm. Mol. Biol. Evol. 23:1891–1901.

Lambert, J. D., and N. A. Moran. 1998. Deleterious mutations destabilize ribosomal RNA in endosymbiotic bacteria. Proc. Natl. Acad. Sci. USA 95:4458–4462.

Law, R., and U. Dieckmann. 1998. Symbiosis through exploitation and the merger of lineages in evolution. Proc. R. Soc. Lond. B 265:1245–1253.

Legendre, P., and M. J. Anderson. 2002. Program DistPCoA. Available at: http://www.bio.umontreal.ca/casgrain/en/labo/distpcoa.html (accessed December 22, 2009).

Legendre, P., Y. Desdevises, and E. Bazin. 2002. A statistical test for host-parasite coevolution. Syst. Biol. 51:217–234.

Lumini, E., V. Bianciotto, P. Jargeat, M. Novero, A. Salvioli, A. Faccio, G. Bécard, and P. Bonfante. 2007. Presymbiotic growth and sporal morphology are affected in the arbuscular mycorrhizal fungus *Gigaspora margarita* cured of its endobacteria. Cell. Microbiol. 9:1716–1729.

Margulis, L. 1981. Symbiosis in cell evolution. W. H. Freeman, San Francisco, CA.

Moran, N. A. 1996. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. Proc. Natl. Acad. Sci. USA 93:2873–2878.

Moran, N. A., J. P. McCutcheon, and A. Nakabachi. 2008. Genomics and evolution of heritable bacterial symbionts. Annu. Rev. Genet. 42:165–190.

Moran, N. A., H. J. McLaughlin, and R. Sorek. 2009. The dynamics and time scale of ongoing genomic erosion in symbiotic bacteria. Science 323:379–382.

Msiska, Z., and J. B. Morton. 2009. Phylogenetic analysis of the Glomeromycota by partial beta-tubulin gene sequences. Mycorrhiza 19:247–254.

O'Fallon, B. 2008. Population structure, levels of selection, and the evolution of intracellular symbionts. Evolution 62:361–373.

Redecker, D., R. Kodner, and L. E. Graham. 2000. Glomalean fungi from the Ordovician. Science 289:1920–1921.

Remy, W., T. N. Taylor, H. Hass, and H. Kerp. 1994. Four hundred-million-year-old vesicular arbuscular mycorrhizae. Proc. Natl. Acad. Sci. USA 91:11841–11843.

Ronquist, F. 2003. Parsimony analysis of coevolving species associations. Pp. 22–64 *in* R. D. M. Page, ed. Tangled trees: phylogeny, cospeciation, and coevolution. Univ. of Chicago Press, Chicago, IL.

Ronquist, F., and J. P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574.

Rosendahl, S., P. McGee, and J. B. Morton. 2009. Lack of global population genetic differentiation in the arbuscular mycorrhizal fungus *Glomus mosseae* suggests a recent range expansion which may have coincided with the spread of agriculture. Mol. Ecol. 18:4316–4329.

Russell, J. A., and N. A. Moran. 2005. Horizontal transfer of bacterial symbionts: heritability and fitness effects in a novel aphid host. Appl. Environ. Microb. 71:7987–7994.

Russell, J. A., A. Latorre, B. Sabater-Munoz, A. Moya, and N. A. Moran. 2003. Side-stepping secondary symbionts: widespread horizontal transfer across and beyond the Aphidoidea. Mol. Ecol. 12:1061–1075.

Sachs, J. L., R. G. Skophammer, and J. U. Regus. 2011. Evolutionary transitions in bacterial symbiosis. Proc. Natl. Acad. Sci. USA 108:10800–10807.

Salvioli, A., M. Chiapello, J. Fontaine, A. L. Hadj-Sahraoui, A. Grandmougin-Ferjani, L. Lanfranco, and P. Bonfante. 2010. Endobacteria affect the metabolic profile of their host *Gigaspora margarita*, an arbuscular mycorrhizal fungus. Environ. Microbiol. 12:2083–2095.

Smith, S. E., and D. J. Read. 2008. Mycorrhizal symbiosis. Academic Press, New York.

Stamatakis, A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–2690.

Tavaré, S. 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. Lect. Math. Life Sci. 17:57–86.

Toh, H., B. L. Weiss, S. A. H. Perkin, A. Yamashita, K. Oshima, M. Hattori, and S. Aksoy. 2006. Massive genome erosion and functional adaptations provide insights into the symbiotic lifestyle of *Sodalis glossinidius* in the tsetse host. Genome Res. 16:149–156.

Van Tuinen, D., E. Jacquot, B. Zhao, A. Gollotte, and V. Gianinazzi-Pearson. 1998. Characterization of root colonization profiles by a microcosm community of arbuscular mycorrhizal fungi using 25S rDNA-targeted nested PCR. Mol. Ecol. 7:879–887.

Vos, M., and X. Didelot. 2009. A comparison of homologous recombination rates in bacteria and archaea. ISME J. 3:199–208.

White, T. J., T. Bruns, S. Lee, and J. Taylor. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. Pp. 315–322 *in* M. A. Innis, D. H. Gelfand, J. J. Sninsky, and T. J. White, eds. PCR protocols: a guide to methods and applications. Academic Press, San Diego, CA.

Yamamura, N. 1993. Vertical transmission and evolution of mutualism from parasitism. Theor. Popul. Biol. 44:95–109.

Associate Editor: P. Turner

Supporting Information

The following supporting information is available for this article:

Table S1. Arbuscular mycorrhizal fungal isolates surveyed for association with Glomeribacter.

Table S2. Evidence of codivergence between Glomeromycota and Glomeribacter detected by the ParaFit test using a set of bacterial gene sequences with no partitions as well as bacterial sequence data partitioned into segments defined by recombination breakpoints.

Figure S1. Patterns of coevolution between the Gigasporaceae fungal hosts and the *Ca*. Glomeribacter gigasporarum bacterial endosymbionts.

Supporting Information may be found in the online version of this article.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

CHAPTER 2

VitaScope: A Computational Pipeline for Inferring Microbial Lifestyle

Abstract

Microbes play important roles in all environments they occupy. Thanks to next generation sequencing technologies, novel microbial lineages are being discovered at an increasingly rapid rate. However, our understanding of the lifestyles exhibited by many microbes is still limited. For this reason, we have created a computational pipeline, VitaScope, which utilizes patterns of adaptation to infer microbial lifestyle. We expect that this pipeline will be most useful as a hypothesis-generating first step when characterizing novel microbial lineages. We show that VitaScope is effective at inferring microbial lifestyle. Furthermore, this pipeline is phylogeny independent and robust to large levels of sequence divergence. Using *Ca.* Glomeribacter gigasporarum (a non-essential endosymbiont of Glomeromycota) as a focal example, we used VitaScope to identify genes of potential relevance for microbial interactions with their environment. To this end, we established that genes involved in DNA regulation, energy metabolism, and pathogenicity are likely important for survival of *Ca.* Glomeribacter within their fungal hosts. By investigating both genome-wide patterns of *Ks/Ka* as well as in 1000 base pair windows, we also identified that non-essential endosymbionts are just as effective at purging slightly deleterious mutations from their genomes as free-living organisms. This is distinctly different from what is found in essential endosymbionts and suggests that genome reduction in non-essential endobacteria is more a result of adaptation rather than genome degradation.

Introduction

Microbes inhabit nearly all environments on the planet and are often key players in ecosystem structure and functioning. Microbe relevance extends to higher eukaryotes, including humans, where they play important roles in host fitness (Sharon et al., 2010, Turnbaugh et al., 2007). As new microbes are being discovered at an increasing rate, there is a growing need to better understand microbial lifestyles, how specialized they are within their environments, and what roles they may play in ecosystem maintenance.

While exploring genome content is an effective method to understand metabolic potential of an organism (Ghignone et al., 2012, Dumas et al., 2006), uncovering genome-wide patterns of positive selection is a powerful approach to infer both lifestyle and the role microbes play in their environment (Stuckenbrock et al., 2011). A signature of positive, or directional, selection is often a detectable result of adaptation in response to environmental pressures. Consequently, investigating the genes and pathways experiencing positive selection can provide insights into what environmental pressures have shaped an organism's genome. In particular, an increased rate of nonsynonymous substitution (*Ka*) relative to the synonymous substitution rate (*Ks*) provides a sensitive measure of selection pressure at the gene level.

As organisms of similar lifestyles, i.e. non-essential endosymbionts, are likely adapting to their environments in similar ways, the expectation is that closely related genes and gene clusters will show adaptive signatures within lifestyle. For example, irrespective of host, many pathogenic bacteria show patterns of positive selection on effectors and genes involved in assembly of the type III secretion system (Guttman et al., 2006, Weber & Koebnick, 2006, Stavrinides et al., 2008). By comparing patterns of adaptation across lineages exhibiting different lifestyles, researchers may not only gather information about the ecological niche that lineage occupies, but also what types of genes are undergoing adaptive

evolution to survive within it. Signatures of directional selection have proven to be a valuable tool in many systems for these purposes. For example, Stuckenbrock et al. (2011) have used patterns of positive selection to predict a recent host specialization event of the fungal pathogen *Zymoseptoria tritici* on its host plant, wheat. Following this work, through exploiting selection signatures, Brunner et al. (2013) identified particular cell wall degrading enzymes of *Z. tritici* which are likely important for interactions with their new host.

Next generation sequencing technologies have greatly expanded our capacity to explore the role microbes play in ecosystem functioning. As sufficient data are now either already available or easily acquired for many microbial lineages, it is becoming worthwhile to pursue this question using genome-wide patterns of adaptation as a predictive tool. For example, we can now sequence and interrogate whole genomes of uncultivated microbes that exist inside eukaryotic hosts (Ghignone et al., 2012, Shigenobu et al., 2000), or even sequence single bacterial cells isolated from the environment (Zhang et al., 2006, Yoon et al., 2011). Through analysis of community composition, next generation sequencing has also dramatically enhanced our rate of discovery of new bacterial lineages existing across a multitude of different environments. Unfortunately, for many microbes, especially newly discovered lineages, we still have no clear picture of what type of lifestyle they exhibit or what role they might play in their environments.

For this reason, we have created a comparative computational pipeline, VitaScope, which exploits patterns of positive selection as a tool to infer microbial lifestyle and role in ecosystem functioning. Here, we define lifestyle as the way in which an organism exists and interacts with its environment, i.e. essential endosymbiont, gut-specific symbiont, non-essential endosymbiont, free-living organism, etc. Our goal is to create a computational approach that will act as a hypothesis generating first step to inform researchers about genomic features that may be worth pursuing in newly discovered or difficult to manipulate organisms. This approach was designed to be simple to execute,

requiring only a nucleic acid fasta file and a corresponding amino acid file, and can be run using coding sequences extracted from draft genome sequencing data. VitaScope interrogates patterns of adaptation across lineages to ask the main question: Are patterns of positive selection within a target lineage similar to those found in organisms of known lifestyles? As positive selection can only be analyzed at the genic level, after recovering positive selection results for each gene individually, we place these results within the framework of the Gene Ontology (GO) (Ashburner et al., 2000) so broader comparisons can be made. Rather than be limited to comparisons of single genes, the Gene Ontology provides a framework, which allows us to hierarchically cluster information so we can assess the pathways and functional categories of genes that are undergoing adaptive evolution. We chose this approach because we expect pathways and functional categories of genes are much more likely to be evolving in similar ways to adapt to an environment than individual genes alone. After lifestyle is inferred, VitaScope will attempt to isolate functional gene categories that differ substantially in selection patterns from what is observed in organisms sharing a similar lifestyle. Our expectation is that organisms exhibiting the same lifestyle will overall look most similar to one another, and therefore interrogating the differences between them may reveal genes important for adaptation to their specific environment.

When developing VitaScope, we focused our attention on non-essential endosymbionts, which exhibit a lifestyle of increasing significance and interest. Non-essential, or secondary, endosymbionts are an understudied group of bacteria that are harbored inside eukaryotic cells. Here, they are often beneficial for their host. Most non-essential endosymbionts currently described are found in insects, particularly aphids, and provide their hosts with protection against various parasite pressures (Moran et al., 2008). For example, sister taxa *Ca.* Hamiltonella defensa and *Ca.* Regiella insecticola protect their aphid hosts from parasitoid wasps and fungal disease, respectively (Degnan et al., 2009, Scarborough et al., 2005). In addition to insects, non-essential endosymbiont *Ca.* Glomeribacter gigasporarum is known to inhabit arbuscular mycorrhizal (AM) fungi (phylum: Glomeromycota), particularly members of the

Gigasporaceae family. While the role of *Glomeribacter* in defense against parasites is currently unknown, they have been shown to enhance presymbiotic hyphal proliferation of their fungal hosts (Lumini et al., 2007). The *Glomeribacter* hosts, AM fungi, are obligate plant biotrophs which provide plants with nutrients from the soil in return for photosynthate (Smith & Read, 2008). Unsurprisingly, all of these non-essential endosymbionts are expected to have variable impacts on host fitness, either being highly favorable in certain environments, or detrimental in others.

It was recently discovered that non-essential endosymbionts are capable of establishing long-term associations with their hosts, and that these interactions can be evolutionarily stable (Mondo et al., 2012). However, despite this knowledge, we still have little understanding of commonalities that exist amongst non-essential endosymbionts or environmental pressures that have shaped endosymbiont genomes. This is mostly due to the fact that genome architecture appears highly variable amongst these organisms, making comparative genomics difficult. These bacteria differ widely in not only genome size and GC content, but also in the abundance of pseudogenes, mobile and repetitive elements, as well as phage (Moran et al., 2008).

These features make non-essential endosymbionts an ideal group for exploration using VitaScope for several reasons. First, non-essential endosymbionts can exhibit an evolutionarily stable, highly specialized lifestyle, which should be predictable using patterns of adaptation. Second, endosymbionts of all kinds are being discovered at a rapid rate, and any insights into what commonalities distinguish different types of endosymbionts would greatly enhance our understanding of endosymbiont biology**.** Finally, in addition to being difficult to investigate using more traditional comparative genomic approaches, most non-essential endosymbionts are currently unable to be cultured, making experimental identification of important genes difficult. We expect that both of these problems will be partially alleviated using VitaScope.

Before patterns of selection can be examined in non-essential endosymbionts, we must address how effective these organisms are at purging slightly deleterious mutations. Genome degradation is a common phenomenon in endosymbionts due to their highly reduced effective population sizes (N_e) (Moran et al., 2008), leading to a drift-induced over-accumulation of slightly deleterious mutations across the genome (Herbeck et al., 2003). Although directional selection can likely still occur in these organisms (Moran et al., 2008), the enhanced rate of amino acid change may lead to false positives when surveying genes for positive selection (Herbeck et al., 2003) and may negatively impact lifestyle inferences. Nearly all of this work has focused on essential endosymbionts, while non-essential endosymbionts remain largely unexplored. All non-essential endosymbionts maintain the capacity for limited genetic exchange as well as host switching, making population dynamics in this group substantially different from what is observed in essential endobacteria. Our prediction is that these capabilities will result in an overall larger N_e when compared to essential endosymbionts and therefore more effective purging of deleterious alleles from their populations. Superficially, the ability to exchange genes and switch hosts may seem to be in conflict with the reduced genomes present in these organisms. However, reductive genome evolution is known to occur even in organisms with large effective population sizes and even be adaptive given the right environmental circumstances (Morris et al., 2012).

Here we show that non-essential endosymbionts are as effective at purging slightly deleterious mutations as free-living bacteria, and that this is distinctly different from what we observe for essential endosymbionts. Additionally, we show that VitaScope is capable of clustering microbial lineages by lifestyle and that this is both phylogeny-independent and robust to large levels of sequence divergence within lineage. In the present study, we focus on bacteria, although we expect VitaScope can be applied to other forms of life. Specifically, we utilize non-essential endosymbionts, particularly *Ca.* Glomeribacter, as a focal group to illustrate the utility of VitaScope in inferring lifestyle as well as genes

of potential importance for partner interactions. The lifestyle of *Glomeribacter* is that of an ancient non-essential endosymbiont (Mondo et al., 2012), yet we have little understanding of the genes involved in host-symbiont interactions. For this project, we generated several draft genome assemblies representing divergent individuals of the Glomeribacter lineage. We have been able to identify a set of functional gene categories, which share very similar patterns of positive selection amongst non-essential endosymbionts, including genes involved in DNA regulation and metabolism, as well as particular *Glomeribacter* genes involved in energy metabolism and pathogenesis which may have relevance for interactions of *Glomeribacter* with its host.

Material and Methods

Genome sequencing, assembly and annotation

For investigation of genes under positive selection in Glomeribacter, we used reference genome sequence from BEG34 (Ghignone et al., 2012) as well as 2 additional endosymbiont genomes sequenced during this project, including endobacteria of *S. castanea* BEG1 and *G. margarita* JA201A-16. Both accessions were acquired from the International Vesicular Arbuscular Mycorrhizal Culture Collection, Morgantown, WV.

For DNA preparation, as endobacteria are known to be identical within a host experimental population (Mondo et al., 2012), 50-150 spores of each fungus were surface sterilized following Mondo et al. (2012) crushed in 500 µL of water, and passed through a 2 µm filter. Nuclei of *Gigaspora* are approximately 5 µm in diameter (Jabaji-Hare et al., 1986, Sward, 1981), and passage of spore contents through a 2 µm filter revealed successful retention of bacterial DNA as well as elimination of fungal nuclei (Figure 2.1). Following filtration, spore contents were concentrated using the SpeedVac Concentrator (ThermoFisher Scientific, Waltham, MA) to a final volume of 50 µL. 2 µL of spore contents per reaction were then used for 5 independent 20 µL whole genome amplification (WGA) reactions using the Genomi-Phi V2 whole genome amplification kit (GE-healthcare, Piscataway, NJ). WGA products

were treated to reduce chimeric products following the debranching protocol described in Gilbert et al. (2010). Treated DNA products were quantified using the Nanodrop (ThermoFisher Scientific) and pooled using equal amounts of DNA from each independent reaction to a total of 1 µg. As whole genome amplification using the phi-29 enzyme is randomly biased across the genome (Zhang et al., 2006), pooling DNA products across independent reactions is expected to normalize products and enhance even-ness of coverage across the genome (Zhang et al., 2006). 300-400 bp libraries were then prepared at the Cornell Sequencing Facility and sequenced using the Illumina Hi-Seq paired-end 100 bp platform (Illumina, San Diego, CA).

Due to the large volume of sequence data recovered, raw reads were split into 10 100x coverage subsets and assembled independently using Velvet version 1.1 (Zerbino & Birney, 2008). Recovered contigs were then assembled together using Geneious version 5.4 (Biomatters). Final contigs were then surveyed for predicted coding sequences using AMIgenes (Bocs et al., 2003) under the available Candidatus *Glomeribacter gigasporarum* gene model (Ghignone et al., 2012).

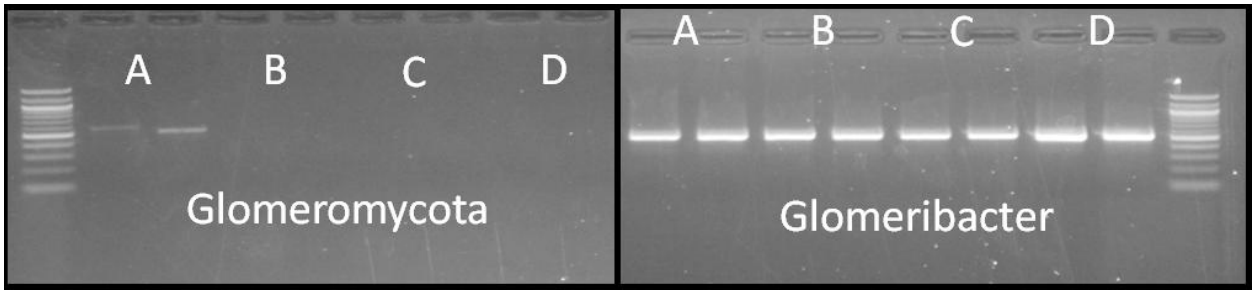

Figure 2.1. Passage of spore contents through a 2 µm filter reveals successful removal of fungal DNA and retention of bacterial DNA. Spore contents from two different fungi, *G. margarita* JA201A (lanes 2, 4, 6 and 8 in left panel and lanes 1, 3, 5, and 7 in right panel) and *S. pellucida* CL750A (lanes 1, 3, 5, and 7 in left panel and lanes 2, 4, 6, and 8 in right panel), were subjected to one of four different treatments (labeled A-D). A: untreated samples, B: spore contents filtered through a 2 µm filter, C: spore contents filtered through a 2 µm filter then treated with DNase and Proteinase K, D: spore contents filtered through a 2 µm filter then treated with DNase and Proteinase K and concentrated to 50 µL. Left panel: PCR amplification of fungal 28S rDNA using primers LR1 and NDL 22 (Van Tuinen et al., 1998). Right

panel: PCR amplification of Glomeribacter 23S rDNA using primers GlomGiGf (Bianciotto et al., 2004) and LSU483r (Mondo et al., 2012).

Computational pipeline to assess positive selection across lineages

We have developed a computational pipeline, VitaScope, to investigate patterns of selection within groups of organisms and compare these results across lineages (Figure 2.2). VitaScope is comprised of four mains steps for each lineage: (i) input of data, (ii) identification of single copy gene orthologs and positive selection analysis, (iii) Gene Ontology (GO) (Ashburner et al., 2000) annotation and identifying the adaptive change ratio (ACR) for each GO category, and (iv) comparison of ACRs for all lineages within each GO category. All scripts, aside from published programs, were written in Python v2.7.3.

Additionally, GO categories experiencing extreme levels of adaptation compared to other organisms of the same predicted lifestyle can be harvested. Our goal is to provide a tool to identify genes and pathways that may be unique and important for adaptation of an organism to their particular environment. For example, although *Ca.* Hamiltonella defensa, *Ca.* Regiella insecticola, and *Ca.* Glomeribacter gigasporarum are all non-essential endosymbionts, *Ca.* Glomeribacter exists inside a fungal environment and likely provides a very different benefit to its host than either *Ca.* Hamiltonella or *Ca.* Regiella, which may be reflected by the types of genes experiencing dramatically different amounts of adaptive change.

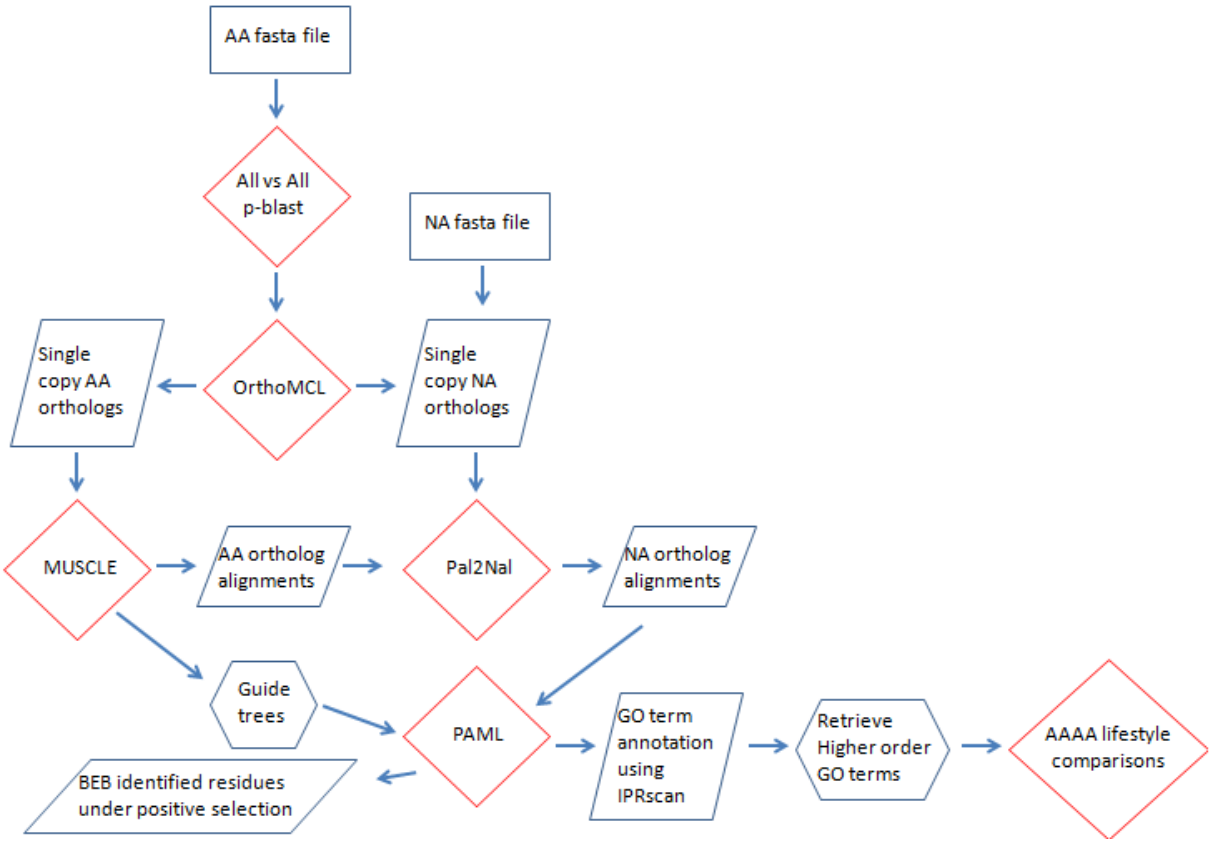

Figure 2.2. Flow chart depicting order of operations in the VitaScope pipeline prior to lifestyle comparisons. Legend: Blue rectangles = input files, red diamonds = computational operations, blue parallelograms = results that can be recovered by the user, blue hexagons = data files generated which are required for downstream steps (not recoverable by user).

i) Input of data

To run VitaScope, only two fasta formatted input files are needed: an amino acid (AA) sequence file containing all coding sequences within a lineage of interest and a corresponding nucleic acid (NA) sequence file. Coding sequences must share identical names across files.

For each lineage, VitaScope requires at least three genomes, which must satisfy a few requirements (see topic sections: "Deleterious mutation accumulation across lifestyle" and "Quality control of PAML inputs") in order to recover useful results. VitaScope uses three genomes within lineage rather than a single genome and several outgroups for three reasons. First, often the rates of molecular evolution vary dramatically between lineages (this is especially true for endosymbionts and their closest

free-living relatives), leading to unpredictable complications when identifying genes experiencing positive selection. These concerns are alleviated by working within a lineage, as rates of molecular evolution are often similar. Second, this approach allows us to observe what types of genes are experiencing ongoing adaptive change to life within the particular environment the organism was isolated from. Therefore, using three genomes within a lineage is expected to be more informative for dissecting which genes are important for persistence within this environment. Third, if outgroups are included, analysis of positive selection will result in mixed selection patterns, as there will be more than one lifestyle within the 'lineage' under inspection. Although this may still be useful, it is much less informative for VitaScope's purposes. Finally, we also found that rather than including many genomes within a lineage, VitaScope is most effective using only 3 representatives (see Results section: Data Input Considerations) due to enhanced recovery of single copy orthologs.

ii) Identification of single copy gene orthologs and positive selection analysis

Subsequent to upload, AA sequences are used for an All vs All blast-p (Altschul et al., 1990) with default parameters: e-value cutoff = $1x10^{-10}$, maximum matches = 500. All vs All blast comparisons were conducted at the Computational Biology Service Unit (CBSU) at Cornell University (Sun et al., 2010). Blast results are then used for identification of orthologs using OrthoMCL version 1.4 (Li et al., 2003). To ensure the removal of paralogs from OrthoMCL clusters and increase the number of single copy orthologs, we used stringent parameters for OrthoMCL, including: mode 3, pi_cutoff 40, pv_cutoff 1e30, and inflation 5. All identified single copy orthologs were recovered and folders containing both NA and AA fasta files were generated for each ortholog group. AA orthologs were then aligned using MUSCLE (Edgar, 2004), and both guide trees and alignments were retained for use in downstream steps. Protein alignments were then converted to phylip formatted codon alignments using Pal2Nal (Suyama et al., 2006). Poor alignments were removed from the dataset and alignments with frameshifts were

trimmed according to Orsi et al. (2008), then both curated codon alignments and guide trees for each gene were used in PAML analysis (Yang 2007) to identify positively selected genes and residues. For this work, we used the codeml algorithm implemented in the PAML software package (version 4.4d) with parameters: NSsites = 0 1 2 3 7 8, CodonFreq = 2, seqtype = 1, kappa = 0.3, omega = 1.3, ncatG = 10. To enhance confidence in our results, we chose the most conservative model (M7 vs M8) to test for genes experiencing positive selection and used the likelihood ratio test to determine significance (df = 2, χ^2 critical value = 6, equivalent to p = 0.05). All PAML runs were conducted on the BIOSIM cluster available at the CBSU. Specific AA residues under positive selection identified using the Bayes Empirical Bayes (BEB) method integrated into PAML (Yang, 2007) were also collected.

iii) GO annotation and identifying the ratio of adaptive change for each GO category

Following identification of genes under positive selection, we assigned GO terms (Ashburner et al., 2000) to all orthologs using the InterPro scan (Zdobnov & Apweiler, 2000), which allowed us to investigate multiple levels of organization of genetic information. A custom script was written to recover higher order GO terms associated with individual genes, which uses the complete GO.obo file available at http://www.geneontology.org/GO.downloads.ontology.shtml and accessed on 06/05/2012. As the gene ontology attempts to hierarchically cluster genes into functional categories, all of which are subsets of three major classes (biological process, molecular function, or cellular component), these umbrella terms were used as the terminal category assigned to each gene. For comparison across lineages, each GO category was investigated for the ratio of genes under positive selection compared to the total number of genes present (herein discussed as the Adaptive Change Ratio, or ACR). For example, if for lineage A, GO category X harbored 2 genes under positive selection and 10 total genes, the ACR in GO category X for organism A would be 2/10, or 0.2.

iv) Lifestyle inference through comparison of ACRs across all lineages within each GO category

For comparison of lifestyles, results for all lineages included in the study set were collected and organized in a single file (organized per GO category) containing ACRs as well as raw data regarding the total number of genes per category for all lineages and how many genes were identified to be under positive selection. Using the adaptive change ratios (ACRs) identified by VitaScope, we assessed whether ACRs were more similar between lineages known to maintain similar lifestyles (for example, non-essential endosymbionts), whether they were more similar between lineages from the same phylogenetic background (for example, within a genus such as *Burkholderia*), or whether no clustering occurred at all.

For this work, we conducted pairwise comparisons of ACRs for each lineage across the study set to identify which lineages are most similar. For example, when investigating lineage A, which exhibits an ACR of 0.25 in GO category X, we compare this value to those observed in all other lineages in the dataset, which in this example include lineages B, C and D, having ACRs of 0.30, 0.10 and 0.50, respectively. In the case of GO category X, lineage A is most similar to lineage B. For similarities to be identified, not only do ACRs have to match most closely to one another, but the magnitude of difference between them must be no greater than 0.1 to ensure that spurious matches are removed. This analysis is then repeated for all GO categories for which information is available for every lineage surveyed and number of best matches to each lineage are tallied. If ACR matched equally well to multiple lineages, these data points were removed. Although we expect this approach to come at the cost of some useful information, it effectively removes lower order GO categories with limited gene content (i.e. categories consisting of only one to two genes total), providing both a clearer and broader picture of adaptation.

VitaScope has two additional cutoff criteria that must be met for each lineage in order to approve a category for lifestyle comparisons: if the lineage being compared has either (1) no gene content, or (2) no genes under positive selection, there can only be one additional lineage with a 0 value, otherwise that category is skipped for this lineage. Both of these criteria are used to ensure that if

there is limited gene content in a particular GO category, spurious matches will not be made to multiple lineages lacking information. Subsequent to ACR clustering, values are averaged within and between known lifestyles. Significant differences in number of ACR matches between lineages across lifestyle are then examined using the Welch's T-test. Significance was determined after P-values were controlled for multiple comparisons using the Bonferroni correction.

While VitaScope infers lifestyle using pairwise ACR comparisons, it is not necessarily true that these patterns always represent important adaptations required for survival of all species existing in a given lifestyle. For example, *Hamiltonella*, *Regiella*, *Wolbachia* and *Glomeribacter* are all non-essential endosymbionts. However, *Glomeribacter* exists within a fungal host, while the others inhabit insects. Consequently, if adaptation patterns are similar within insect-infecting endosymbionts but differ in Glomeribacter, they likely represent important adaptation to life as a non-essential endosymbiont of insects, but are not representative of what kinds of adaptive change occur to survive as a non-essential endosymbiont in general. For identifying what types of genes are characteristic of all individuals of the same lifestyle, we identified GO categories which always display ACR patterns most similar to other individuals within lifestyle, herein discussed as 'core lifestyle' categories. As a result, any GO category that displayed ACR patterns that were not shared by all individuals of the same lifestyle were removed.

Deleterious mutation accumulation across lifestyle

It is thought that genome degradation can lead to difficulties in identification of residues under positive selection (Fares et al., 2002, Herbeck et al., 2003). In order to assess whether non-essential endosymbionts are experiencing genome degradation as opposed to positive selection, we investigated the ratio of neutral substitutions to amino acid substitutions (*Ks/Ka*) across the genomes of all lineages surveyed (see Lineages surveyed, below). We expect that lineages capable of purging slightly deleterious mutations will over-accumulate neutral substitutions over time rather than amino acid changes, leading

to larger *Ks/Ka* values. For genome-wide surveillance, all single copy orthologs identified by OrthoMCL (Li et al., 2003) were aligned using MUSCLE (Edgar, 2004), concatenated, and then analyzed using DnaSP v.5 (Librado & Rozas, 2009). *Ks* and *Ka* as well as the *Ks/Ka* ratio (Jukes-Cantor corrected) were calculated as described in equations 1-3 from Nei and Gojobori (1986). Results were calculated pairwise for all individuals within lineage, and then averaged to arrive at lineage-wide values .

Histograms of *Ka/Ks* across genomes were generated by investigating pairwise *Ka/Ks* in 1000 base pair windows across all concatenated orthologs using DnaSP (Librado & Rozas, 2009). Resulting values were then binned by frequency of occurrence using *Ka/Ks* increments of 0.005 for all comparisons.

Quality control of VitaScope inputs

As large levels of genome degradation can lead to uncertainty regarding accurate identification of genes under true positive selection (Fares et al., 2002, Herbeck et al., 2003), any individuals with *Ks/Ka* ratios near or below the empirically identified value of 4.0 were omitted from lifestyle comparisons. Additionally, in general, lineages were removed if uncorrected synonymous divergence (*Ks*) was greater than 0.75, indicating saturation of synonymous sites and an inability to accurately assess positive selection patterns. Although codeml is expected to handle large levels of divergence fairly well (Anisimova et al., 2001), we felt it was appropriate to make initial comparisons within traditionally acceptable boundaries. Finally, individuals were removed if within group they were too closely related to one another ($Ks < 0.05$), as likely insufficient time has elapsed to allow for a signature of selection to emerge. However, as limited genetic information is available for many lineages, we intentionally included two datasets of animal infecting mycoplasma, one at each *Ks* extreme, to assess how robust VitaScope is at dealing with large divergence between lineages of the same lifestyle.

Lineages surveyed

We investigated genomes of 47 microorganisms (33 of which were used in lifestyle comparisons) representing 5 different lifestyles (essential endosymbionts, non-essential endosymbionts, free-living bacteria, gut specific symbionts and animal infecting mycoplasmas) (Table 2.1). For essential endosymbionts, we included 3 genomes of *Buchnera* (*Ca.* Buchnera aphidicola [Host: *Cinara tujafilina*], *Ca.* Buchnera aphidicola str. LSR1 [Host: *Acyrthosiphon pisum*], and *Ca.* Buchnera aphidicola str. Bp [Host: *Baizongia pistaciae*]), 3 genomes of *Sulcia* (*Ca.* Sulcia muelleri DMIN, *Ca.* Sulcia muelleri SMDSEM, and *Ca.* Sulcia muelleri CARI), and 3 genomes of *Blochmannia* (*Ca.* Blochmannia vafer str. BVAF, *Ca.* Blochmannia pennsylvanicus str. BPEN, and *Ca.* Blochmannia floridanus). For non-essential endosymbionts, we included 3 genomes of *Ca.* Glomeribacter gigasporarum (*Ca.* Glomeribacter gigasporarum BEG34 [Host: *Gigaspora margarita*], *Ca.* Glomeribacter gigasporarum BEG1 [Host: *Scutellospora castanea*], and *Ca.* Glomeribacter gigasporarum JA201A-16 [Host: *Gigaspora margarita*]), and 8 genomes of *Wolbachia* (*Wolbachia pipentis* strain wMel, *Wolbachia* endosymbiont of *Drosophila simulans* strain wSim, *Wolbachia* endosymbiont of *Culex quinquefasciatus* strain wPip, *Wolbachia* endosymbiont of *Drosophila willistoni* TSC#14030-0811.24, *Wolbachia sp.* wRi, *Wolbachia* endosymbiont of *Diaphorina citri* FL, *Wolbachia pipientis* wAlbB, and *Wolbachia* endosymbiont of *Drosophila ananassae*). Additionally, as sufficient amounts of genomic data were not available for each individual species, we grouped two genomes of *Ca.* Hamiltonella defensa (strains MED and 5AT) with one genome of *Ca.* Regiella insecticola (Strain LSR1) to define the *Hamiltonella/Regiella* lineage. Both of these species are sister taxa, exist exclusively as non-essential endosymbionts of insects, and perform similar roles in protection against parasite pressures (Moran et al., 2008). For free-living bacteria, we included 3 closely related genomes of free-living *Burkholderia* (*Burkholderia ambifaria* MC40-6, *Burkholderia multivorans* ATCC 17616, and *Burkholderia vietnamiensis* G4), 3 closely related genomes of free-living enteric (*Escherichia coli* ATCC 8739, *Enterobacter cloacae subsp. cloacae* NCTC 9394, and *Salmonella enterica*

subsp. enterica serovar Choleraesuis str. SC-B67), 3 genomes of *Prochlorococcus* (*Prochlorococcus marinus* str. MIT 9215, *Prochlorococcus marinus subsp. pastoris* str. CCMP1986, and *Prochlorococcus marinus* str. MIT 9301), and 3 genomes of *Bradyrhizobium* (*Bradyrhizobium sp.* ORS278, *Bradyrhizobium sp.* BTAi1, and *Bradyrhizobium japonicum* USDA 110). For gut specific symbionts, we included 3 genomes of *Lactobacillus* (*Lactobacillus acidophilus* NCFM, *Lactobacillus johnsonii* NCC 533, and *Lactobacillus farciminis* KCTC 3681) and 3 genomes of *Bifidobacterium* (*Bifidobacterium animalis* subsp. *lactis* AD011, *Bifidobacterium longum* subsp. *longum* JCM 1217, and *Bifidobacterium catenulatum* DSM 16992). For animal infecting *Mycoplasma*, we included 3 genomes of *Mycoplasma canis* (*Mycoplasma canis* PG 14, *Mycoplasma canis* UF31, and *Mycoplasma canis* UF33) and 3 genomes of more distantly related *Mycoplasma* (*Mycoplasma pneumoniae* FH, *Mycoplasma agalactiae* PG2, and *Mycoplasma penetrans* HF-2).

Table 2.1. List of genomes used in the present study, the lineage they belong to, and the lifestyle they exhibit.

Organism	Lineage	Lifestyle
Candidatus Blochmannia vafer str. BVAF	*Blochmannia*	Essential Endosymbiont
Candidatus Blochmannia pennsylvanicus str. BPEN	*Blochmannia*	Essential Endosymbiont
Candidatus Blochmannia floridanus	*Blochmannia*	Essential Endosymbiont
Candidatus Buchnera aphidicola (Cinara tujafilina)	*Buchnera*	Essential endosymbiont
Candidatus Buchnera aphidicola str. LSR1 (Acyrthosiphon pisum)	*Buchnera*	Essential endosymbiont
Candidatus Buchnera aphidicola str. Bp (Baizongia pistaciae)	*Buchnera*	Essential endosymbiont
Candidatus Sulcia muelleri DMIN	*Sulcia*	Essential endosymbiont
Candidatus Sulcia muelleri SMDSEM	*Sulcia*	Essential endosymbiont
Candidatus Sulcia muelleri CARI	*Sulcia*	Essential endosymbiont
Candidatus Glomeribacter gigasporarum BEG34 (Gigaspora margarita)	*Glomeribacter*	Non-essential Endosymbiont
Candidatus Glomeribacter gigasporarum BEG1 (Scutellospora castanea)	*Glomeribacter*	Non-essential Endosymbiont
Candidatus Glomeribacter gigasporarum JA201A (Gigaspora margarita)	*Glomeribacter*	Non-essential Endosymbiont
Candidatus Hamiltonella defensa MED (Bemisia tabaci)	*Hamiltonella/Regiella*	Non-essential Endosymbiont
Candidatus Hamiltonella defensa 5AT (Acyrthosiphon pisum)	*Hamiltonella/Regiella*	Non-essential Endosymbiont
Candidatus Regiella insecticola LSR1 (Acyrthosiphon pisum)	*Hamiltonella/Regiella*	Non-essential Endosymbiont
Wolbachia pipentis strain wMel	*Wolbachia*	Non-essential Endosymbiont
Wolbachia endosymbiont of *Drosophila simulans* strain wSim	*Wolbachia*	Non-essential Endosymbiont
Wolbachia endosymbiont of *Culex quinquefasciatus* (mosquito) strain wPip	*Wolbachia*	Non-essential Endosymbiont
Wolbachia endosymbiont of *Drosophila willistoni* TSC#14030-0811.24	*Wolbachia*	Non-essential Endosymbiont
Wolbachia sp. wRi	*Wolbachia*	Non-essential Endosymbiont
Wolbachia endosymbiont of *Diaphorina citri* FL	*Wolbachia*	Non-essential Endosymbiont
Wolbachia pipientis wAlbB	*Wolbachia*	Non-essential Endosymbiont
Wolbachia endosymbiont of *Drosophila ananassae*	*Wolbachia*	Non-essential Endosymbiont
Bradyrhizobium sp. ORS278	*Bradyrhizobium*	Free-living
Bradyrhizobium sp. BTAi1	*Bradyrhizobium*	Free-living
Bradyrhizobium japonicum USDA 110	*Bradyrhizobium*	Free-living
Burkholderia ambifaria MC40-6	Free-living *Burkholderia*	Free-living
Burkholderia multivorans ATCC 17616 (Prj:19401)	Free-living *Burkholderia*	Free-living
Burkholderia vietnamiensis G4	Free-living *Burkholderia*	Free-living
Escherichia coli ATCC 8739	Free-living enterics	Free-living
Enterobacter cloacae subsp. *cloacae* NCTC 9394	Free-living enterics	Free-living
Salmonella enterica subsp. *enterica* serovar *Choleraesuis* str. SC-B67	Free-living enterics	Free-living
Prochlorococcus marinus str. MIT 9215	*Prochlorococcus*	Free-living
Prochlorococcus marinus subsp. *pastoris* str. CCMP1986	*Prochlorococcus*	Free-living

Prochlorococcus marinus str. MIT 9301	*Prochlorococcus*	Free-living
Bifidobacterium animalis subsp. *lactis* AD011	*Bifidobacterium*	Gut specific symbiont
Bifidobacterium longum subsp. *longum* JCM 1217	*Bifidobacterium*	Gut specific symbiont
Bifidobacterium catenulatum DSM 16992 = JCM 1194	*Bifidobacterium*	Gut specific symbiont
Lactobacillus acidophilus NCFM	*Lactobacillus*	Gut specific symbiont
Lactobacillus johnsonii NCC 533	*Lactobacillus*	Gut specific symbiont
Lactobacillus farciminis KCTC 3681	*Lactobacillus*	Gut specific symbiont
Mycoplasma canis PG 14	*Mycoplasma canis*	Animal Infecting Mycoplasma
Mycoplasma canis UF31	*Mycoplasma canis*	Animal Infecting Mycoplasma
Mycoplasma canis UF33	*Mycoplasma canis*	Animal Infecting Mycoplasma
Mycoplasma pneumoniae FH	*Mycoplasma*	Animal Infecting *Mycoplasma*
Mycoplasma agalactiae PG2	*Mycoplasma*	Animal Infecting *Mycoplasma*
Mycoplasma penetrans HF-2	*Mycoplasma*	Animal Infecting *Mycoplasma*

Results

Nonessential endosymbionts are capable of adaptive evolution

Typically endosymbionts exist in populations with highly reduced effective population sizes (Moran et al., 2008). As a result, genome degradation due to genetic drift is rampant in these organisms, leading to difficulties in differentiating between amino acid changes due to degeneration and amino acid changes due to positive selection. This feature of endosymbiont evolution has created controversy over the ability to accurately detect positive selection in these organisms (Fares et al., 2002, Herbeck et al., 2003). It has been alternatively proposed that genome degradation due to over-accumulation of slightly deleterious mutations predominates, and that host-level selection is largely responsible for maintenance of endosymbiont populations (Klasson & Andersson, 2004). However, most of this debate surrounds essential endosymbionts, for which no capacity for host switching or recombination exist, leaving non-essential endosymbionts largely unexplored.

To investigate whether non-essential endosymbionts are capable of experiencing positive selection, we explored genome wide ratios of neutral (synonymous) substitutions to amino acid changing (nonsynonymous) substitutions, or *Ks/Ka*. Here, we assume that if a lineage is capable of

purging slightly deleterious mutations, over time, an over-accumulation of neutral substitutions will be observed across the genome rather than amino acid changes, leading to larger *Ks/Ka* values. For example, if a *Ks/Ka* ratio of 5 is observed, this means that on average, for every 5 neutral substitutions per synonymous site there is only a single amino acid change per non-synonymous site. Alternatively, if a *Ks/Ka* of 2 is observed, this would be indicative of rampant amino acid change and likely over-accumulation of deleterious mutations due to genetic drift. For this work, all single copy, orthologous coding sequences were collected, aligned, concatenated, and then scanned using DnaSP (Librado & Rozas, 2009) for a genome-wide *Ks/Ka* ratio (Table 2.2) as well as *Ka/Ks* over 1000 bp windows (Figure 2.3) to identify how representative this mean value is at an approximately genic level.

Table 2.2. Average *Ks*, *Ka*, *Ks*/*Ka*, and *Ka*/*Ks* values for all lineages included in this study. Values were calculated by averaging pairwise values within each lineage together. Values with Jukes-Cantor correction for multiple hits are presented in columns labeled as 'Corrected'.

Taxa	**Uncorrected *Ks***	**Corrected *Ks***	**Uncorrected *Ka***	**Corrected *Ka***	**Uncorrected *Ks/Ka***	**Corrected *Ks/Ka***	**Uncorrected *Ka/Ks***	**Corrected *Ka/Ks***
Buchnera	0.56	1.01	0.28	0.35	1.99	2.90	0.50	0.34
Sulcia	0.24	0.30	0.12	0.13	1.99	2.22	0.50	0.45
Blochmannia	0.52	0.89	0.19	0.22	2.76	4.10	0.36	0.24
Wolbachia	0.26	0.34	0.06	0.06	4.58	5.80	0.22	0.17
Hamiltonella/Regiella	0.49	1.47	0.15	0.17	3.38	8.66	0.30	0.12
Glomeribacter	0.42	0.61	0.09	0.10	4.48	6.14	0.22	0.16
Prochlorococcus	0.47	0.84	0.09	0.10	5.25	8.67	0.19	0.12
Free living *Burkholderia*	0.25	0.31	0.05	0.05	4.79	5.64	0.21	0.18
Bradyrhizobium	0.39	0.57	0.12	0.13	3.40	4.41	0.29	0.23
Enterics	0.59	1.15	0.08	0.08	7.38	13.61	0.14	0.07
Lactobacillus	0.66	1.59	0.26	0.33	2.50	4.82	0.40	0.21
Bifidobacterium	0.59	1.15	0.20	0.23	2.92	4.91	0.34	0.20
Mycoplasma canis	0.06	0.06	0.01	0.01	9.89	10.25	0.10	0.10
Other animal infecting *Mycoplasma*	0.70	NA	0.37	0.52	1.87	NA	0.53	NA

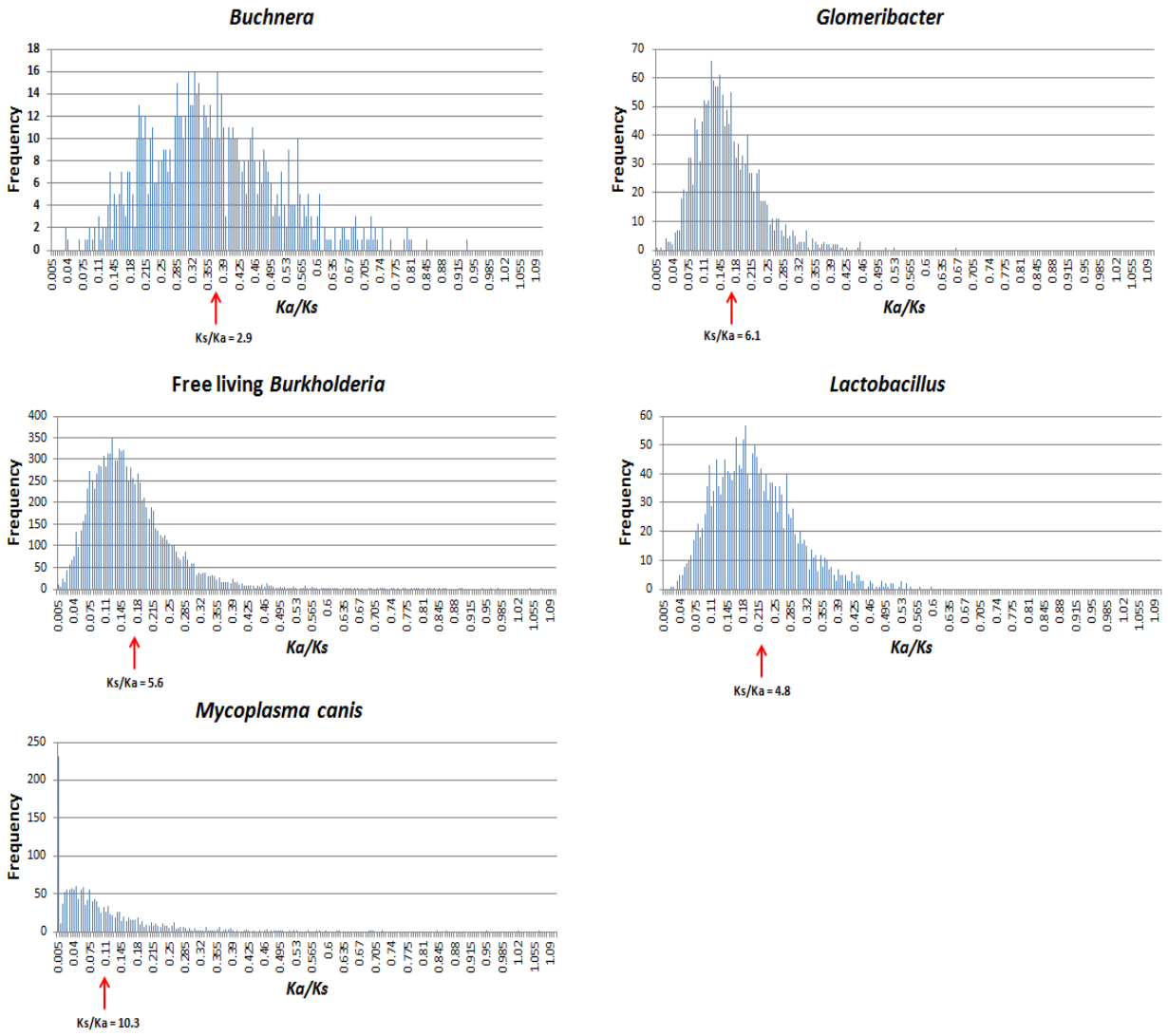

Figure 2.3. Distribution of *Ka*/*Ks* across the genome (1000 bp windows) for lineages representing different lifestyles. Top left – *Buchnera*: essential endosymbiont; top right – *Glomeribacter*: non-essential endosymbiont; middle left – free-living *Burkholderia*: free-living bacteria; middle right – *Lactobacillus*: Gut specific symbiont; bottom left – *Mycoplasma canis*: animal infecting *Mycoplasma*. *Mycoplasma canis* was intentionally included to assess the impact of recent divergence time on lifestyle predictions. Red arrows point to the genome-wide mean value and the corresponding *Ks/Ka* value.

We found that, consistent with previous studies (Herbeck et al., 2003), essential endosymbionts appear to be experiencing large levels of genome degradation, as witnessed by the strikingly low overall *Ks/Ka* ratios observed in *Buchnera* and *Sulcia* (*Ks/Ka* = 2.90 and 2.20, respectively), as well as a much wider distribution of *Ka/Ks* across the genome (Table 2.2, figure 2.3). Interestingly, *Blochmannia*, an essential endosymbiont of carpenter ants exhibited a comparatively high *Ks/Ka* of 4.09, although this value was still lower than that observed in any lineage of a different lifestyle. In stark contrast with essential endosymbionts, non-essential endosymbionts exhibited much larger *Ks/Ka* ratios. The lowest value was observed in *Wolbachia* (*Ks/Ka* = 5.80), while both *Ca.* Glomeribacter (*Ks/Ka* = 6.14) and the *Hamiltonella/Regiella* group (*Ks/Ka* = 8.66) maintain fairly high *Ks/Ka* ratios, indicating that these endosymbionts are capable of effectively purging slightly deleterious mutations. Two additional features further support this assertion: (i) observed *Ks/Ka* ratios of non-essential endosymbionts fall within those found in free-living organisms where positive selection is readily detected (lowest observed *Ks/Ka* = 4.41, highest = 13.61; Table 2.2), and (ii) the distribution of *Ka/Ks* across the genome is tightly clustered around the mean, similar to that seen in free-living lineages and distinctly different from the spread-out distribution observed in essential endosymbionts (Figure 2.3). *Ks/Ka* of gut symbionts *Lactobacillus* and *Bifidobacterium* (*Ks/Ka* = 4.82 and 4.91, respectively) also fell within the lower bound of values observed in free-living lineages. For further inspection of how large levels of sequence divergence impact lifestyle predictions (see section: VitaScope is robust to large levels of sequence divergence within lifestyle), we included two lineages of animal infecting *Mycoplasma,* which displayed either a saturation of *Ks* (*Ks/Ka* incalculable) or very low *Ks* (*Mycoplasma canis*, *Ks* = 0.06, *Ks/Ka* = 10.25).

Taken together, these results indicate that, unlike essential endosymbionts, non-essential endosymbionts are effective at purging slightly deleterious mutations from their genomes and are therefore likely experiencing little genome degradation. Consequently, we expect that investigating positive selection in non-essential endosymbionts will be a sensitive measure of adaptation to their

environment. This also implies that N_e is much larger for non-essential endosymbionts than essential endosymbionts, and that genome reduction may be due to adaptation rather than degradation (Morris et al., 2012). This raises interesting questions regarding how non-essential endosymbionts maintain this capacity, which likely has to do with their ability to recombine as well as switch hosts. In contrast, we believe that interrogating positive selection patterns in essential endosymbionts is largely futile, as many of the "positively selected" genes identified will be false positives due to the high levels of genome degradation these organisms experience. For this reason, we provide a suggested *Ks/Ka* ratio cutoff of 4.0 for lineages to be compared using VitaScope. We arrived at this value empirically by investigating *Ks/Ka* as well as lifestyle inferences resulting from inclusion of organisms below this cutoff. We found that if these lineages were included, although the overall trends did not change, the number of categories available for comparison were dramatically reduced, leading to less powerful lifestyle inferences (data not shown). Consequently, for this study we removed any lineages with *Ks/Ka* ratios below or near 4.

VitaScope goals and outputs

With the low cost and relative ease of current genome sequencing, the largest limitation in genomics is no longer data collection, but rather formulating questions that will use genomic data and provide the most insights into the biology of an organism, especially of newly discovered or difficult to study lineages. With this in mind, we designed VitaScope, a computational pipeline used for deciphering not only what genes may be worth investigating within a particular group of organisms, but also inferring their lifestyle. We expect that this tool will be of particular utility in endosymbiotic systems, as discovery rate of novel endosymbionts is increasing and these bacteria are often difficult to manipulate due to their resilience to cultivation under laboratory conditions.

VitaScope investigates patterns of adaptation within a lineage and compares them across clades. The goal is to assess what commonalities exist amongst lineages inhabiting similar environments

(for example, non-essential endosymbionts), as well as what is unique within a particular lineage of interest which may represent important adaptations to their specific environment. VitaScope uses PAML (Yang 2007) to identify genes experiencing positive selection and represents results within the Gene Ontology (GO) framework (Ashburner et al., 2000) so various levels of organization of genetic information can be compared. VitaScope produces an output file containing the adaptive change ratios (ACRs) exhibited by each lineage within particular GO categories. From this file, ACRs are compared to identify: (i) which lineages share the most similar adaptation signatures overall (Figures 2.4 and 2.5), (ii) how similar these are to organisms of known lifestyles, and (iii) what GO categories exhibit extreme patterns of selection.

While lifestyle inference is one of our main foci, in cases where this information is either uncertain or already known, we enabled VitaScope to provide users with information that may be useful for better understanding adaptation within the focal lineage. To this end, VitaScope will output several additional datasets aside from lifestyle comparisons, including: (i) folders containing all single copy orthologous genes within a lineage, identified by OrthoMCL (Li et al., 2003), one of NA sequences and one of corresponding AA sequences, (ii) protein and nucleotide alignments of all orthologs with AA sequences aligned with MUSCLE (Edgar, 2004) and NA alignments generated by converting AA alignments using Pal2Nal (Suyama et al., 2006)), (iii) a dataset of all genes identified by PAML (Yang, 2007) as experiencing significant positive selection, determined using the M7 vs M8 model; $p \leq 0.05$, and (iv) a dataset describing which specific residues are experiencing adaptive change within genes, identified using the Bayes Empirical Bayes (BEB) method integrated into PAML (Yang, 2007). We expect that BEB identified residues will be most useful in cases where organisms can be manipulated genetically, although if datasets are of sufficient size, these can also be used to predict whether genes are likely under balancing versus directional selection.

VitaScope inferences of lifestyle

To investigate VitaScope capability for lifestyle inferences, we collected genomic data from 11 bacterial lineages (33 genomes total; see Lineages surveyed), of known lifestyles, including *Glomeribacter*, and then assessed how efficient VitaScope was at clustering lineages by lifestyle using ACRs. For this work, ACRs were compared across lineages to identify lineages experiencing similar levels of adaptive change, herein discussed as 'best ACR matches'. Best matches were then tallied and lineages were clustered by their known lifestyle. We found that the number of best ACR matches was consistently higher within lifestyle than between (Figure 2.4; Figure 2.5). For example, non-essential endosymbionts shared on average 81 best matching GO categories, while on average, non-essential endosymbionts match best to free-living bacteria at 54 GO categories. Best ACR matches decreased to only 17 GO categories in non-essential endosymbiont comparisons to gut symbionts, and to 36 GO categories in comparisons to animal-infecting mycoplasmas. Similar results were found for free-living lineages, gut symbionts, and animal-infecting mycoplasmas, where the number of best ACR matches were always more than 2 times greater within lifestyle than between (Figure 2.4). Welch's T-tests after Bonferroni correction for multiple comparisons revealed in nearly all cases that the number of matches within lifestyle were significantly greater than those to the next best matching lifestyle (non-essential endosymbionts to free-living $P = 8.04\ 10^{-4}$, free-living to gut specific symbionts $P = 1.57\ 10^{-4}$, animal infecting mycoplasmas to non-essential endosymbionts $P = 2.72\ 10^{-5}$).

Interestingly, gut symbionts were only marginally different from free-living bacteria in their patterns of adaptive evolution (gut symbionts to free-living $P = 0.06$). When looking closer at gut symbionts, we found that ACRs for these lineages (*Lactobacillus* and *Bifidobacterium*) were on average larger than that observed in other lifestyles (data not shown), indicating consistently rapid change in a moderately reduced genome. However, more importantly, in comparisons to free-living bacteria, both

lineages shared many common ACR matches particularly to free-living *Burkholderia* (*Lactobacillus* to *Burkholderia*: 159 ACR matches, *Bifidobacterium* to *Burkholderia*: 110 ACR matches), all of which are opportunistic pathogens of humans (they are particularly problematic in individuals with cystic fibrosis) (Vandamme et al., 1997). *Lactobacillus* additionally showed many similarities to opportunistic mammalian gut associated enteric (92 ACR matches). We would predict that this closer similarity in ACR ratios to free-living bacteria is due to an overlap in the environments these organisms can occupy.

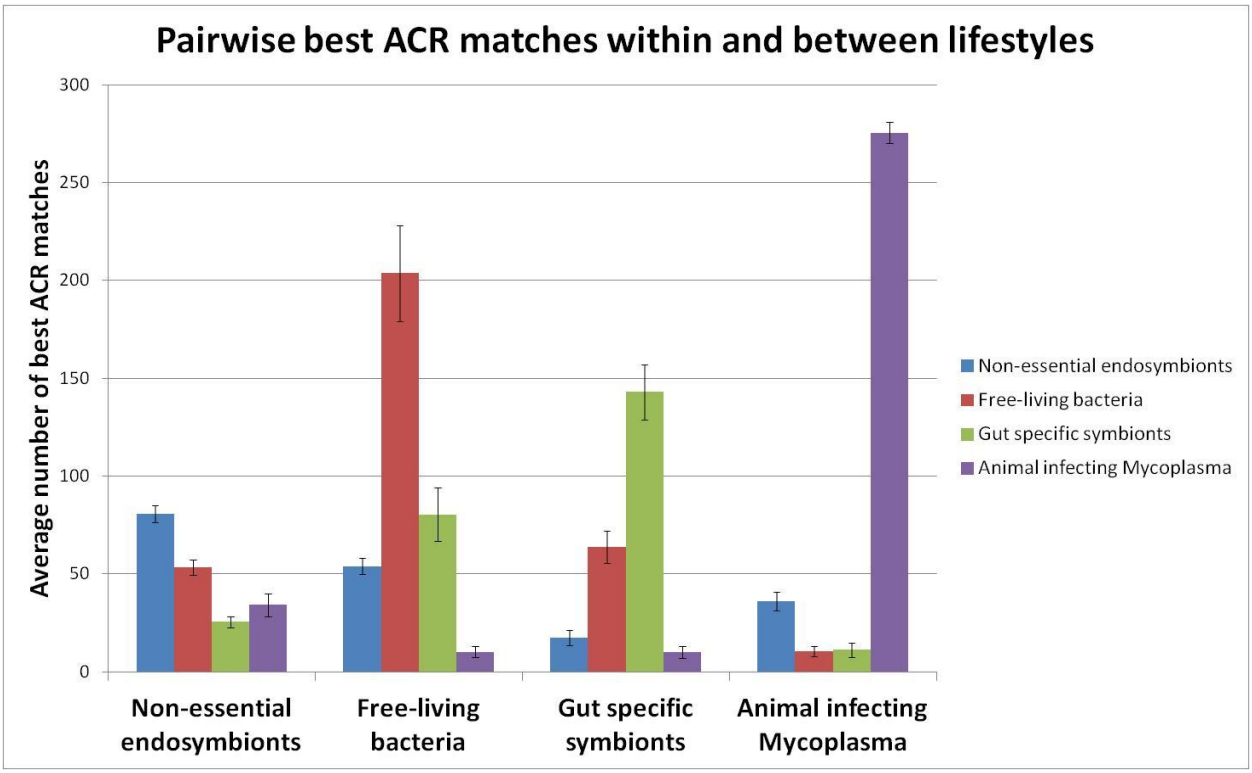

Figure 2.4. Pairwise best ACR matches averaged within and between lifestyles. Error bars represent standard error of the mean.

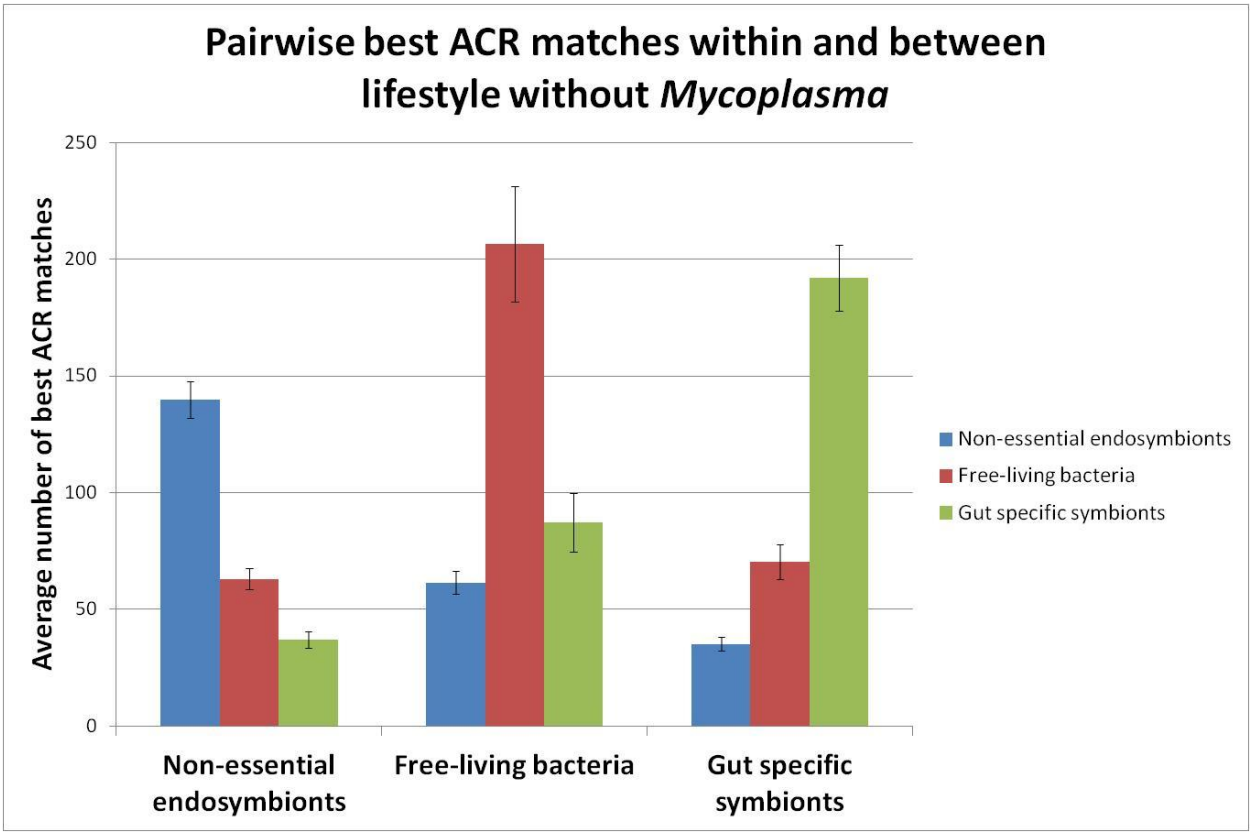

Figure 2.5. Pairwise best ACR matches within and between lifestyles with *Mycoplasma* excluded from the analysis. This comparison was done to show the impact of small, hyper-variable genomes on overall lifestyle predictions. Organisms of modest genome size, such as non-essential endosymbionts and gut specific symbionts, are particularly benefitted by the removal of *Mycoplasma*, while large genome organisms such as free-living bacteria show little change (for comparison, see figure 3). Error bars represent standard errors of the mean.

VitaScope is phylogeny independent

Although we expect patterns of adaptation to emerge mostly due to environmental pressures, we needed to ensure that these patterns were phylogeny independent if VitaScope is to be broadly applicable. To test whether lifestyle inferences are phylogeny-independent, we included in our study set several lineages that were closely related phylogenetically, but differed in lifestyle, including Glomeribacter (*Burkholderia* lineage, β-*proteobacteria*, non-essential endosymbiont) and free-living *Burkholderia* (β-*proteobacteria,* soil inhabiting, free-living), and the *Hamiltonella*/*Regiella* group (γ-

proteobacteria, non-essential endosymbionts) and free-living enteric (γ-*proteobacteria*; free-living, animal gut associated). At a larger phylogenetic distance, we also included *Wolbachia* (α-*proteobacteria*; non-essential endosymbiont/reproductive parasite) and *Bradyrhizobium* (α-*proteobacteria*; free-living rhizobia), as well as *Lactobacillus* (*Firmicutes*; gut specific) and animal infecting mycoplasmas (*Mollicutes*).

We found that in all cases, VitaScope identified more similarity in ACRs across organisms of similar lifestyle than those of similar phylogenetic origins (Figure 2.6). For example, *Glomeribacter* shared most ACRs in common with the *Hamiltonella*/*Regiella* group (90 matches) and *Wolbachia* (87 matches), while only half as many matches were observed to other free-living *Burkholderia* (49 matches). Additionally, the number of matches to free-living *Burkholderia* was no different than what was observed to remaining free-living lineages within the dataset, if not slightly lower (*Bradyrhizobium* = 47 matches, free-living enterics = 52 matches, *Prochlorococcus* = 59 matches).

While the *Hamiltonella/Regiella* lineage also matches best to *Glomeribacter* (91 matches) and fairly well to *Wolbachia* (78 matches), they initially showed high similarity to their closest phylogenetic relative, the free-living enterics (82 matches). However, we found that this closeness in ACR matches was due to a reduced number of GO categories to sample across due to inclusion of small genome organisms such as *Mycoplasma* (Figure 2.5) (see section: "Data input considerations"). When *Mycoplasma* were removed from the study set, number of best matches for *Hamiltonella*/*Regiella* to *Glomeribacter* and *Wolbachia* increased to 117 and 139, respectively, while number of matches to free-living enterics only increased to 84. Similar results were found for other closely related lineages of different lifestyle, and so we therefore concluded that VitaScope is truly phylogeny independent.

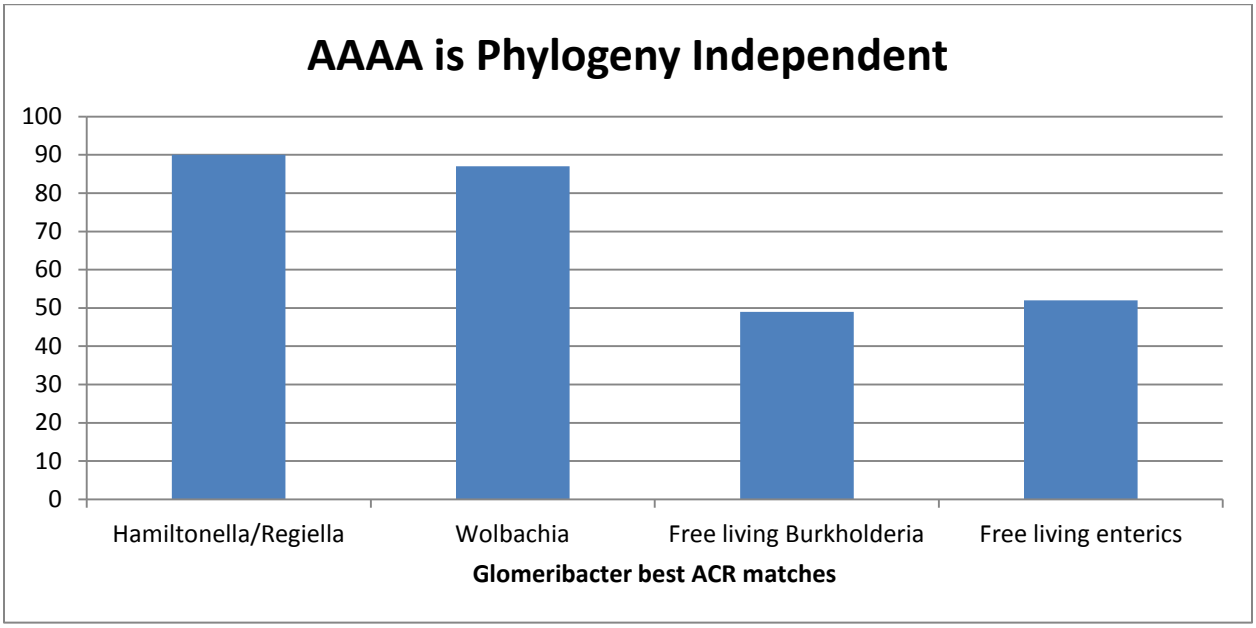

Figure 2.6. Comparison of the number of best ACR matches *Glomeribacter* has with other lineages in the dataset. ACRs identified by VitaScope appear mostly influenced by lifestyle rather than phylogeny, as the numbers of ACR matches to other non-essential endosymbionts are higher than to their closest free-living relative, the free-living *Burkholderia*.

VitaScope is robust to large levels of sequence divergence within lifestyle

Although current sequencing technologies make it easy to collect genomic information, it is still not always possible to work with taxa separated by equal divergence times. For this reason, we set out to test how robust VitaScope is at inferring lifestyle even when lineages harbor large levels of sequence divergence. To test the boundaries of VitaScope, we included two lineages of animal infecting *Mycoplasma* (Table 2.2), one with individuals which diverged relatively recently based on identified *Ks* values (average *Ks*: 0.06), and another with individuals which diverged long ago (saturated; average *Ks* > 0.75). We found that, despite this difference in divergence times, ACR patterns were much more similar between these animal infecting *Mycoplasma* lineages (average ACR best matches: 274) than compared

to any other lifestyle (average ACR matches to next most similar lifestyle: 34 - non-essential endosymbionts) (Figure 2.4).

Based on these results, we concluded that VitaScope is capable of handling large levels of sequence divergence when comparing lineages to infer lifestyle. We believe that this capacity is largely due to the use of the codeml module in the PAML software package for identification of genes under positive selection. Because codeml is a maximum likelihood based approach, many complexities of the data are considered which would not be in other approaches, including multiple substitutions per site and complex nucleotide substitution models. Although these features allow for robust predictions up to very large levels of sequence divergence (Yang, 1998), power is still reduced when the number of taxa per lineage is low (Anisimova et al., 2001). Consequently, we believe that our use of the most conservative model for predicting genes experiencing positive selection (the M7 vs M8 model) is a large part of why VitaScope retains the capacity to effectively infer lifestyle despite variance in divergence levels.

Data Input Considerations

Due to the nature of the evaluations being made, the strength of VitaScope's comparative approach depends on two major features of the input data. The first is the number of taxa included in each 'lineage'. Unsurprisingly, although the predictive power of PAML increases with increasing numbers of taxa in the dataset (Anisimova et al., 2001), the total number of single copy orthologs will decrease, especially if draft genomes are included. In effect, this means that the more taxa included in a group, the smaller the fraction of the genome that will be available for comparison. As VitaScope compares adaptive change ratios, results are effectively normalized before lifestyle comparisons begin, and therefore this difference in the total number of orthologs will be somewhat controlled for within a GO category. However, we do expect that this will have an overall effect on lifestyle inference by reducing the total number of available categories for comparison, as well as rendering many categories

no longer useful due to limited gene content. We predict that increasing number of taxa will force a trend towards increased numbers of matches to organisms with reduced genomes, as there are increased chances of categories with no gene content or with ACR values of 0.

To investigate the impact number of taxa per lineage has on lifestyle inference, we conducted several lifestyle comparisons using non-essential *Wolbachia* endosymbionts that varied in total number of taxa within this 'lineage', ranging from three to eight individuals. Although *Wolbachia* is often parasitic while other non-essential endosymbionts are beneficial, we chose *Wolbachia* for this analysis because it was the only lineage for which ample genomic data were available. As expected, we found that not only did total number of orthologs decrease dramatically with increasing number of taxa (487 orthologs when 3 *Wolbachia* taxa were included down to 196 with 8 taxa included), but there was also a tendency to have more matches to organisms with highly reduced genomes, in this case, animal infecting mycoplasmas (44 average ACR best matches to *Mycoplasma* with only 3 taxa, up to 79.5 when 8 taxa were included). We believe this tendency is exclusively due to the reduced number of orthologs present, as ACRs of GO categories with high gene content were still more similar to other non-essential endosymbionts than *Mycoplasma* (data not shown). Not surprisingly, the number of comparable categories for *Wolbachia* also decreased by 11% when 8 taxa were included as compared to 3 taxa. Lastly, we found that the average number of similarities of *Wolbachia* to other non-essential endosymbionts decreased slightly with increasing number of taxa, but this was not very dramatic (69.5 when 3 *Wolbachia* genomes included vs 55 when 8 *Wolbachia* genomes included). These results suggest that while ACRs of categories with sufficient gene content appear to give fairly consistent results regardless of number of taxa included, the reduction in information for lower order GO categories lead to artificial inflation of the number of matches to organisms with highly reduced genomes. Therefore, it appears that high numbers of gene orthologs is more desirable than high numbers of taxa for the purpose of lifestyle prediction.

The second data input consideration is the number of GO categories available for lifestyle comparisons. This consideration is related to several factors, including number of taxa per lineage, but most importantly the genome composition and size for lineages involved in lifestyle comparison. If genomes within a lineage are highly variable in composition, i.e. high number of dispensable genes, recent genome duplication events, etc, the number of identified single copy orthologs will decrease. Similarly, if exceedingly small genomes are included, there will be fewer GO categories to make comparisons across. Although this should not impact VitaScope's efficiency at clustering by lifestyle, the total number of matching categories will be decreased across the entire dataset.

To illustrate the impact of genome composition and reduced genome size on lifestyle comparisons, we investigated ACR patterns both with and without *Mycoplasma* (genome sizes: ≈ 1.0 Mb) lineages present. In addition to small genomes, mycoplasmas are notorious for harboring few core genes and many dispensable genes (Liu et al., 2012), thus further reducing the total number of single copy orthologs. We found that when mycoplasmas were included in comparisons, we were still able to predict lifestyle using pairwise best ACR matches (see VitaScope is predictive of lifestyle, above). However, if they were removed, the number of comparable GO categories went up considerably. Additionally, we found that this increase in comparable categories was not uniform across the dataset, which certainly has to do with the cutoff criteria used to approve a GO category for lifestyle comparisons (see Materials and Methods – Computational pipeline to assess positive selection across lineages: section iv). These criteria are set to restrict comparisons if limited genetic information is available. Therefore, for lineages of modest genome size, the effect of removing mycoplasmas was much greater than for lineages with large genomes (Figure 2.5).

Removal of mycoplasmas from our study set was particularly beneficial for non-essential endosymbionts, where we found that this increased the average number of pairwise best matches within lifestyle from 81 to 140. Furthermore, removal of mycoplasmas had little effect on the average

ACR matches of non-essential endosymbionts to other lifestyles, only increasing best matches to free-living bacteria from 54 to 61.5, and gut symbionts from 23 to 35.2. Similar results were found for gut symbionts, where pairwise best matches within lifestyle increased from 143.5 to 192. Conversely, removal of mycoplasmas had very little effect on lifestyle comparisons for free-living bacteria, either within lifestyle (average ACR within lifestyle with mycoplasma included: 204, without mycoplasma: 207) or between lifestyles. These results reveal that inclusion of lineages with small or hyper-variable genomes may strongly impact the power of lifestyle comparisons, especially for lineages of modest genome size, but this does not change the overall patterns observed. Thus, in order to discern potential lifestyle of novel organisms, if not immediately transparent, a two tiered approach may be appropriate.

Similarities amongst individuals of the same lifestyle

As VitaScope is capable of clustering lineages by lifestyle (see above), we became interested in addressing which types of genes (for example, genes involved in cell wall synthesis) are exhibiting patterns most characteristic of a given lifestyle, i.e., non-essential endosymbionts. We define these GO categories as 'core lifestyle' categories. For consideration as a 'core lifestyle' category, ACR patterns had to always be most similar to other lineages of the same lifestyle, regardless of how many lineages were included in that lifestyle. Interestingly, we found that the 'core lifestyle' categories consistently comprised a large proportion of the total average pairwise matches in every lifestyle analyzed (67**%** in free-living bacteria, 55**%** in non-essential endosymbionts, 68**%** in gut symbionts, and 47**%** in mycoplasmas).

Furthermore, we also found that for all except 'free-living bacteria', the most poorly defined of lifestyles in the dataset, these core categories included two to three of the highest order GO terms (biological process, molecular function and cellular component). Because these are umbrella terms which encompass nearly all genes in the genome, this finding indicates that the overall amount of

adaptation occurring for organisms inhabiting the same environments is very similar. This feature raises interesting questions regarding how organisms of similar lifestyle converge upon similar levels of adaptive change overall.

Aside from the highest order GO terms, few GO categories appeared to consistently be core lifestyle component for more than one of the lifestyles analyzed. One notable exception was GO:0004366 – Helicase activity. This category includes proteins involved in the unwinding of DNA or RNA helices and appeared to always show similar levels of adaptive change within lifestyle and no overlapping patterns between lifestyle. These exceptions aside, most 'core lifestyle' components seemed to be unique to the lifestyle being scrutinized, which makes sense considering that evolutionary pressures may vary dramatically between different ecological niches, leading to variable selection pressure across different suites of genes.

An exploration of commonalities amongst non-essential endosymbionts

Little is known regarding commonalities between non-essential endosymbionts. Currently, our best suggestion comes from the work of Brownlie et al., 2007. This study focused on identifying genes experiencing positive selection across *Wolbachia* that were either essential mutualists of nematodes or non-essential parasites of *Drosophila*. However, while this work highlights the types of genes under selection in this system, we do not know how common these observed patterns are across non-essential endosymbionts. Consequently, we assessed all core lifestyle GO categories for non-essential endosymbionts (44 categories). We found that many of these core categories were involved in DNA metabolism and regulation (9/44 categories). We also found several categories involved in cofactor binding and metabolism (4/44) as well as localization of cellular substances (3/44). Additional core features of non-essential endosymbionts appear to be genes involved in small molecule binding (excluding cofactors; 3/44), nucleotide binding (5/44) and nitrogen compound metabolic processes

(2/44). For many of these categories, there was a reduction in adaptive change for non-essential endosymbionts as compared to other lifestyles.

We had expected to observe core similarities in transport/secretion, cell wall synthesis, cell division and across non-essential endosymbionts. However, we found no evidence that this is the case. Looking closer at these categories, we found that all non-essential endosymbionts were experiencing different levels of change. In GO:0019867 (outer membrane), *Hamiltonella/Regiella* and *Glomeribacter* displayed high ACR ratios 0.27 and 0.38, respectively, while *Wolbachia* showed no evidence of positive selection. While cell envelope related features were reported to be under selection in *Wolbachia* previously (Brownlie et al., 2007), it is possible that this pattern was lost due to the inclusion of different taxa, looking exclusively within non-essential *Wolbachia* endosymbionts of *Drosophila*, or perhaps due to differences between studies in how genes were assigned to functional categories. Similar results are observed in GO:0051301 (cell division), where non-essential endosymbionts display exceptionally variable levels of positive selection, with *Wolbachia* showing the most adaptive change (ACR = 0.33) and *Hamiltonella/Regiella* showing a substantially lower ACR (0.08) than any other lineage in the dataset (next lowest ACR: 0.15 – *Bradyrhizobium*) aside from *Mycoplasma* lineages. While the localization GO category is more related to movement within cell, it appeared that transport/secretion was under exceptionally large amounts of positive selection in *Wolbachia* as opposed to other non-essential endosymbionts, as the 'transmembrane transport' GO category (GO:0055085) revealed 6/38 (ACR = 0.16) genes under positive selection in *Wolbachia*, while only 4/65 (ACR = 0.06) and 5/47 (ACR = 0.1) were experiencing selection in *Hamiltonella/Regiella* and *Glomeribacter*, respectively.

In general, these results show that several types of genes are experiencing similar levels of adaptive change across non-essential endosymbionts, the most striking of which are genes involved in DNA metabolism and regulation. For all categories involved in DNA regulation we observed very little adaptive change relative to that found for organisms of other lifestyles, suggesting that alterations in

these genes may come at a high cost for non-essential endosymbionts. Interestingly, other features which we expected were evolving similarly which have relevance to host-symbiont interactions, such as cell division and transmembrane transport, seem quite variable in the amount of change observed. We would predict that this may have to do with differences in the host's ability to tolerate endosymbiont change.

Outstanding features of *Ca.* Glomeribacter

Compared to *Hamiltonella*, *Regiella* and *Wolbachia*, we know relatively little about what *Ca.* Glomeribacter may provide to their fungal hosts and even less about the genes involved in host-symbiont interactions. This lack of information can in large part be attributed to the difficulty of manipulating these organisms experimentally. Ca. *Glomeribacter* is itself uncultured, and exists inside an obligate plant biotroph which produces little fungal biomass and has a long generation time (Smith & Read, 2008). For these reasons, comparative genomics appears the most attractive approach to pursue for unraveling this symbiosis. Unfortunately, the fungal genome has not yet been sequenced and is predicted to be > 500 Mb as well as have a low GC content (≈33%) (Van Buuren et al., 1999, Harrison, 1999), leaving only bacterial genomic sequence data to explore currently. However, given that *Glomeribacter*'s environment is that of its fungal host, we can still address questions regarding potential genes involved in partner interactions by using VitaScope to investigate what genes are experiencing large selective pressure in order to adapt to their hosts.

Using VitaScope we have already identified that *Glomeribacter* share similar ACR ratios with other non-essential endosymbionts for many functional gene categories. We would expect that these are functional categories of genes that are similarly constrained due to existence in a eukaryotic cell environment. Therefore, to identify what features of the *Glomeribacter* genome may be relevant for interactions with its particular host, we have identified functional gene categories for *Glomeribacter* which are experiencing extreme levels of adaptation (either dramatically more or less) compared to

other non-essential endosymbionts. We define categories as experiencing extreme levels of adaptation if ACRs are at least two fold different from the average ACR level observed within other organisms of the same lifestyle.

While *Glomeribacter* displayed unique ACR patterns in many functional gene categories, there were two particularly striking patterns that emerged. First, we found several categories relevant to energy metabolism (Table 2.3) wherein ACRs of *Glomeribacter* were abnormally high. These categories included genes present in the tricarboxylic acid (TCA) cycle (GO:0006099, *Glomeribacter* ACR: 0.14; average of other non-essential endosymbionts: 0.05), as well as cellular respiration (GO:0022904, *Glomeribacter* ACR: 0.30; average of other non-essential endosymbionts: 0.08). Second, we found that *Glomeribacter* had an extremely high ACR in the pathogenesis category (GO:0009405, *Glomeribacter* ACR: 0.75; average of other non-essential endosymbionts: 0.125). Moreover, *Glomeribacter* had the highest ACR for this category across all taxa analyzed as well as the most gene members (8 genes; next highest number of members: 4 genes - *Hamiltonella/Regiella*). Of the 6 genes in this category experiencing positive selection, 5 of these belonged to a group of cell surface proteins from the YadA family. YadA (short for Yersinia Adhesin A) proteins are involved in adhesion to and invasion of eukaryotic cells.

Table 2.3. Energy metabolism related functional GO categories showing extreme selective pressure in *Glomeribacter* as compared to other non-essential endosymbionts. Ca. *Glomeribacter* BEG34 was chosen for representing the exact genes experiencing positive selection. ETC = Electron Transport Chain, TCA = Tricarboxylic Acid cycle.

Go term ID	*Wolbachia* ACR	*Hamiltonella/Regiella* ACR	*Glomeribacter* ACR	Pathway
GO:0015992	0	0.111111	0.214286	ETC
GO:0045333	0.105263	0.0625	0.294118	ETC
GO:0042625		0	0.4	ETC
GO:0022904	0.153846	0	0.3	ETC
GO:0022900	0.153846	0	0.272727	ETC
GO:0009060	0	0.090909	0.222222	ETC
GO:0016469	0	0	0.333333	ETC
GO:0015662		0	1	ETC
GO:0015986	0	0	0.25	ETC
GO:0015985	0	0	0.25	ETC
GO:0033178		0	0.5	ETC
GO:0045259	0	0	0.285714	ETC
GO:0046961		0	0.5	ETC
GO:0019829		0	0.5	ETC
GO:0045261		0	0.5	ETC
GO:0015988	0	0	0.333333	ETC
GO:0006754	0	0	0.25	ETC, TCA
GO:0015980	0.105263	0.0625	0.294118	Cofactor
GO:0006084	0	0.1	0.285714	Cofactor
GO:0009109	0	0.090909	0.375	Cofactor
GO:0051187	0	0.090909	0.375	Cofactor
GO:0046356	0	0.1	0.285714	Cofactor
GO:0015936	0	0.25	0.5	Cofactor
GO:0006099	0	0.1	0.285714	TCA
GO:0004450	0	0	1	TCA
GO:0004448	0	0	1	TCA
GO:0045239		0	1	TCA

Overall, our findings allow us to make new predictions as to what types of genes may be important for interactions between *Glomeribacter* and its host. In particular, our results show that genes involved in energy metabolism and adhesion to eukaryotic cells are experiencing unusually large amounts of selective pressure in *Glomeribacter*, suggesting that these genes may play a role in partner interactions. Although the genes involved in adhesion to eukaryotic surfaces belong to the pathogenesis GO category (GO:0009405), we predict that because they are surface proteins involved in adhesion (Pizarro-Cerdá & Cossart, 2006), they are more likely important for endosymbiont recognition and potentially host specificity than they are for causing damage to the host. Similarly, the large amount of positive selection observed in energy metabolism genes in *Glomeribacter* may represent important adaptations for survival inside AM fungi. *Glomeribacter* exist within protein-dense vesicles inside their fungal host and are incapable of utilizing sugars for energy production (Ghignone et al., 2012), suggesting that perhaps these microbes have specialized in breaking down proteins for energy production. Based on our finding of positive selection in these pathways, we speculate that this has implications for host-symbiont interactions, and that conversion of protein into energy may be an important component of the benefit these endosymbionts provide their hosts.

Discussion

In this study we have developed VitaScope, a computational pipeline which explores patterns of adaptation as a comparative tool to infer microbial lifestyles. For this work, we selected 2-4 representative bacterial lineages for 4 different lifestyles (33 genomes total): non-essential endosymbionts, free-living bacteria, gut specific symbionts, and animal infecting mycoplasmas. We found that VitaScope was capable of effectively clustering all microbes by their known lifestyle. This tool has been shown useful not only for clustering microbes by lifestyle, but also at predicting functional suites of genes which may be relevant to survival in their environment. VitaScope has been shown to be phylogeny independent and robust to large levels of sequence divergence within lineage. We have used

this tool to explore a focal group of particular interest to us, non-essential endosymbionts. In addition to resolving what commonalities link these organisms together, we have also shown that unlike essential endosymbionts, non-essential endosymbionts are not experiencing large levels of genome degradation.

Non-essential endosymbionts are not subject to genome degradation

During this study, we found that non-essential endosymbionts are just as effective at purging slightly deleterious mutations from their genomes as free-living bacteria, and that this is distinctly different from what is observed in essential endosymbionts (Figure 2.3). Initially, this may seem counter-intuitive, as often both essential and non-essential endosymbionts are subject to large population bottlenecks every host generation, resulting in reductions in N_e and elevated rates of mutation accumulation. However, one key difference between these two modes of endosymbiosis is that non-essential endosymbionts maintain the capacity for exchanging genes and host switching (Mondo et al., 2012, Moran et al., 2008, Degnan & Moran, 2008), while nearly all essential endosymbionts lack these capacities and are maintained mostly by host level selection (Klasson & Andersson, 2004). These results suggest that the reduction in genome size for non-essential endosymbionts is more likely due to adaptation than genetic drift. Many free-living bacteria, such as oceanic bacteriplankton *Prochlorococcus* and *Ca.* Pelagibacter, have evolved reduced genomes and dependency on resources produced by other microbes as a result of natural selection, leading to the formulation of the Black Queen hypothesis of reductive genome evolution (Morris et al., 2012). The observed *Ks/Ka* of the *Prochlorococcus* strains used in this study supports this statement and is similar to what we observe for non-essential endosymbionts (Table 2.2). The Black Queen hypothesis posits that, as some genes perform functions that are costly, and therefore undesirable, and may be provided by other organisms in the environment, there is likely a selective advantage for individuals that lack these genes. For

endosymbionts, this would lead to gene loss and increased host dependence independent of genetic drift.

While low levels of recombination have been reported for non-essential endosymbionts previously (Mondo et al., 2012, Degnan & Moran, 2008), in light of our current results, it is possible that recombination is frequent within host, and perhaps the previously reported low levels only represent rare recombination events between endosymbionts from different hosts. This appears the most parsimonious explanation, as it explains both how non-essential endosymbionts are capable of purging deleterious mutations and how they maintain sufficient population sizes in the face of frequent bottlenecks. Additionally, many non-essential endosymbionts exist in genetically uniform populations within the same host for prolonged periods of time (Mondo et al., 2012, Moran et al., 2008). This results in long periods in which endosymbionts may adapt and specialize within their hosts before contacting novel endosymbiont lineages, leading to easily detected recombination events and large introductions of new DNA when it does occur. This process is apparent in *Ca.* Glomeribacter, as the rate of recombination in these organisms is low, however, the per site effect of recombination is high (Mondo et al., 2012). These features are so pronounced in *Glomeribacter* that different recombinant fragments from within a single lineage of Glomeribacter can show codivergence signals matching different fungal hosts (S.J. Mondo, unpublished data).

Commonalities within lifestyles extend to the highest order GO terms

When comparing ACRs across lineages, we found that the highest order GO terms (biological process, molecular function and cellular component) were consistently effective predictors of lifestyle. Additionally, because these categories encompass nearly all of the gene content present in a genome, ACRs from these categories may serve as useful estimates for the amount of genome-wide positive selection occurring within a lineage. Since we found that these high order terms are consistently

effective predictors of lifestyle, we hypothesize that lifestyle is one of the major features driving how much positive selection a lineage experiences. We would expect that this result would come about through one of several different means. First, this may be due to similarities in the efficacy of positive selection for organisms inhabiting the same environment. However, as we have shown, non-essential endosymbionts are just as effective at purging deleterious mutations from their genomes as free-living taxa, and therefore we expect that, aside from essential endosymbionts, all other lineages used in this study are equally capable of experiencing positive selection.

Second, it is possible that high order GO terms show consistent ACR patterns within lifestyle due to a stable level of adaptation required to exist within the current environment. Our expectation here would be that certain environments cannot tolerate large amounts of change. As VitaScope is particularly investigating patterns of ongoing adaptation within a lineage to their current environment, this scenario seems likely. Supporting this hypothesis, environments that can be dramatically impacted even from slight alterations in occupant genotypes, such as the cellular environment inhabited by non-essential endosymbionts and mycoplasmas, show an overall reduction in the amount of positive selection occurring. Similarly, environments that can tolerate more change, such as those inhabited by free-living organisms and gut symbionts, show higher amounts of positive selection.

Finally, in addition to environmental constraints, it is possible that the ability of highest GO categories to predict lifestyle could also be explained (or exaggerated) by limited environmental change over time. As mentioned above, VitaScope investigates patterns of ongoing adaptation within a lineage. Therefore, all the adaptation accrued to specialize initially within the environment would presumably be under purifying selection at present, and would go undetected by VitaScope. With VitaScope only identifying genes currently under adaptation, ACRs of higher order GO terms may be an indication of how dynamic an environment is at present. Thus, if an environment is rapidly changing, we would likely observe a correspondingly rapid overall rate of adaptive change occurring in its inhabitants. This appears

to be the case for gut symbionts, which exist in an environment which is highly dynamic not only in the types of organisms these symbionts must interact with over time, but also in resource availability. Conversely, although it may be very difficult to adapt to initially, life inside of a eukaryotic cell may be fairly stable over evolutionary time, and therefore would not require high amounts of adaptive change in order to maintain stable populations.

Commonalities amongst non-essential endosymbionts

In the work of Brownlie et al. (2007) several types of genes were identified in *Wolbachia* that were experiencing similar selective pressures regardless of whether they were contained in parasitic or mutualistic *Wolbachia*, including genes involved in DNA metabolism, cofactor biosynthesis and secretion. Although this work was done exclusively within *Wolbachia*, it provides predictions of what types of genes may be under similar evolutionary pressures for all endosymbionts. Consistent with these findings, we were able to identify that DNA regulation and metabolism as well as genes binding cofactors experience similar levels of positive selection across non-essential endosymbionts. It is possible that, as non-essential endosymbionts exist exclusively inside of eukaryotic cells, the types of genes expressed and the timing of their expression may be tightly constrained. One would expect that slight alterations in protein expression may have large impacts on the endosymbionts capacity to interact favorably with its host. Similarly, any such changes may even induce damage the host through expression of harmful genes. With this in mind, it is not surprising that we consistently see very little positive selection in these GO categories across non-essential endosymbionts. Interestingly, several types of genes reported by Brownlie et al. (2007) that we expected to show similar patterns of adaptation within non-essential endosymbionts, including genes involved in cell division, cell wall synthesis and transport/secretion, displayed variable ACRs across non-essential endosymbionts. Although genes involved in any of these processes certainly are under selective pressure in all of these

endosymbionts, perhaps the difference in the amount of positive selection occurring is due to how tolerant the particular host they inhabit is to change.

Glomeribacter specific adaptations relevant to host-symbiont interactions

We expect that VitaScope will be especially useful in situations when organisms are recalcitrant to cultivation or are difficult to manipulate experimentally. Here we have used *Ca.* Glomeribacter, an uncultured endosymbiont of Glomeromycota, as an example lineage to explore the types of genes VitaScope predicts are evolving to adapt to their particular environment. As a result, we identified two major functional categories of genes which are experiencing exceptionally large levels of adaptive change relative to other non-essential endosymbionts, including genes involved in primary energy metabolism, as well as pathogenesis related genes. The elevated levels of positive selection across all of energy metabolism in Glomeribacter is not unexpected. Previously, it was discovered that these endosymbionts lack several key enzymes in glycolysis, including both phosphofructokinase and pyruvate kinase, suggesting that these bacteria are unable to convert sugars into pyruvate (Ghignone et al., 2012), which is necessary to initiate the TCA cycle. These genes are also absent in the *Glomeribacter* genomes sequenced during this study. As the later steps of energy production are indispensible to aerobic life, it was alternatively proposed that these endosymbionts utilize amino acids as the starting source for energy, as they maintain many amino acid importers and are exclusively harbored inside of fungal vesicles rich in protein (Ghignone et al., 2012, Bonfante et al., 1994). Therefore, perhaps the large level of positive selection observed has to do with alterations in this pathway resulting from either changes in energy substrates in this organism or adaptation to better integrate host and endosymbiont systems (Brownlie et al., 2007).

Alternatively, it is possible that these changes in energy production are related to the benefit that endosymbionts provide to their hosts. As previous work has indicated, AM fungi harboring

endosymbionts are capable of extending presymbiotically over two times as far as fungi lacking endobacteria (Lumini et al., 2007). Importantly, their hosts are obligate plant biotrophs, and thus cannot acquire any energy prior to plant colonization, leaving only the energy present in spore reserves for use in presymbiotic growth. One may expect that if these endosymbionts had no role in energy metabolism of their hosts, they would come at a cost to fungal growth presymbiotically, as now energy must be spent not only on hyphal growth, but also on supporting a growing endosymbiont population. As we know that this expected scenario is not true, the hypothesis that endosymbionts impact energy metabolism of their hosts seems probable.

We were initially surprised to discover such a large proportion of genes (6 out of 8) in the pathogenesis GO category were experiencing positive selection in *Glomeribacter*, as this organism is a known mutualist of AM fungi. Upon closer inspection, we discovered that nearly all of these genes under positive selection (5 out of 6), belonged to a complex of cell surface proteins known as YadA. In *Yersinia*, where these proteins were originally discovered (Bialas et al., 2012), YadA proteins are known to form a fibrillar matrix on the bacterial cell surface which is involved in eukaryotic cell surface adhesion, and are required for successful invasion of host cells. Interestingly, fibrillar matrices surrounding *Ca.* Glomeribacter are often found in electron micrographs (Jargeat et al., 2004), suggesting that perhaps these proteins are present and play a role in partner interactions throughout the lifespan of the fungus. Often, eukaryotes exploit bacterial surface proteins as recognition motifs, important for deciding how to react to bacterial presence (Mengin-Lecreulx & Lemaitre, 2005). For example, peptidoglycan and flagellin are two bacterial cell surface molecules which are typically detected by plants and animals alike for use in activating innate immune responses (Mengin-Lecreulx & Lemaitre, 2005, Gómez- Gómez & Boller, 2002). Similarly, a galactofuranose lipopolysaccharide present on the cell surface of *Burkholderia rhizoxinica*, an endosymbiont of *Rhizopus microsporus*, is required for proper host recognition and intracellular survival of the endosymbiont (Lackner et al., 2010). Perhaps the YadA surface proteins in

Glomeribacter play a similar role in their interactions with AM fungi, implying a role for these proteins in host specificity as well as survival.

Conclusion

In conclusion, we have created a computational pipeline which is capable of inferring lifestyle through utilization of positive selection patterns. This pipeline, VitaScope, was shown to be phylogeny independent and robust to large levels of sequence divergence. We used *Glomeribacter* as a focal example to illustrate VitaScope's efficacy at inferring lifestyle as well as predicting genes relevant to survival in their environment. To this end, we have found that genes involved in DNA metabolism and regulation appear to experience consistently low levels of positive selection across non-essential endosymbionts, while genes involved in energy metabolism and adhesion to eukaryotic surfaces may be of particular importance for interactions between Glomeribacter and their fungal hosts.

In the process, we also determined that non-essential endosymbionts are just as capable of purging slightly deleterious mutations from their genomes as free-living organisms. These results suggest that, in contrast to essential endosymbionts, genome reduction in non-essential endobacteria may be the result of adaptation rather than genome degradation. This implies that non-essential endosymbionts are evolving towards reduced genomes and host dependence through an evolutionary process similar to that which is described by the Black Queen hypothesis.

REFERENCES

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. J Mol Biol 215:403–410.

Anisimova, M., J. P. Bielawski, and Z. Yang. 2001. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. Mol Biol Evol 18:1585–1592.

Ashburner, M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, and G. Sherlock. 2000. Gene Ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. 25:25–29.

Białas, N., K. Kasperkiewicz, J. Radziejewska-Lebrecht, and M. Skurnik. 2012. Bacterial cell surface structures in *Yersinia enterocolitica*. Arch. Immunol. Ther. Exp. 60:199–209.

Bianciotto, V., A. Genre, P. Jargeat, E. Lumini, G. Becard, and P. Bonfante. 2004. Vertical transmission of endobacteria in the arbuscular mycorrhizal fungus *Gigaspora margarita* through generation of vegetative spores. Appl Environ Microb 70:3600–3608.

Bocs, S., S. Cruveiller, D. Vallenet, G. Nuel, and C. Médigue. 2003. AMIGene: Annotation of MIcrobial Genes. Nucl. Acids Res. 31:3723–3726.

Bonfante, P., R. Balestrini, and K. Mend Gen. 1994. Storage and secretion processes in the spore of *Gigaspora margarita* Becker &Hall as revealed by high-pressure freezing and freeze substitution. New Phytol 128:93–101.

Brownlie, J. C., M. Adamski, B. Slatko, and E. A. McGraw. 2007. Diversifying selection and host adaptation in two endosymbiont genomes. BMC Evol Biol 7:68.

Brunner, P. C., S. F. F. Torriani, D. Croll, E. H. Stukenbrock, and B. A. McDonald. 2013. Coevolution and life cycle specialization of plant cell wall degrading enzymes in a hemibiotrophic pathogen. Mol Biol Evol 30:1337–1347.

Degnan, P. H., T. E. Leonardo, B. N. Cass, B. Hurwitz, D. Stern, R. A. Gibbs, S. Richards, and N. A. Moran. 2009. Dynamics of genome evolution in facultative symbionts of aphids. Environ. Microbiol, 12(8):2060-2069.

Degnan, P. H., and N. A. Moran. 2008. Evolutionary genetics of a defensive facultative symbiont of insects: exchange of toxin-encoding bacteriophage. Mol Ecol 17:916–929.

Dumas, M.-E., R. H. Barton, A. Toye, O. Cloarec, C. Blancher, A. Rothwell, J. Fearnside, R. Tatoud, V. Blanc, J. C. Lindon, S. C. Mitchell, E. Holmes, M. I. McCarthy, J. Scott, D. Gauguier, and J. K. Nicholson. 2006. Metabolic profiling reveals a contribution of gut microbiota to fatty liver phenotype in insulin-resistant mice. Proc Natl Acad Sci USA 103:12511–12516.

Edgar, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucl. Acids Res. 32:1792–1797.

Fares, M. A., E. Barrio, B. Sabater-Muñoz, and A. Moya. 2002. The evolution of the heat-shock protein GroEL from *Buchnera*, the primary endosymbiont of aphids, is governed by positive selection. Mol Biol Evol 19:1162–1170.

Ghignone, S., A. Salvioli, I. Anca, E. Lumini, G. Ortu, L. Petiti, S. Cruveiller, V. Bianciotto, P. Piffanelli, L. Lanfranco, and P. Bonfante. 2012. The genome of the obligate endobacterium of an AM fungus reveals an interphylum network of nutritional interactions. ISME J 6:136–145.

Gilbert, J. A., K. Zhang, and J. D. Neufeld. 2010. Multiple displacement amplification. Pp. 4255–4263 *in* K. N. Timmis, ed. Handbook of Hydrocarbon and Lipid Microbiology. Springer Berlin Heidelberg.

Gomez-Gomez, L., and T. Boller. 2002. Flagellin perception: a paradigm for innate immunity. Trends Plant Sci 7:251–256.

Guttman, D. S., S. J. Gropp, R. L. Morgan, and P. W. Wang. 2006. Diversifying selection drives the evolution of the type III secretion system pilus of *Pseudomonas syringae*. Mol Biol Evol 23:2342–2354.

Harrison, M. J. 1999. Molecular and cellular aspects of the arbuscular mycorrhizal symbiosis. Annu Rev Plant Phys 50:361–389.

Herbeck, J. T., D. J. Funk, P. H. Degnan, and J. J. Wernegreen. 2003. A conservative test of genetic drift in the endosymbiotic bacterium *Buchnera*: slightly deleterious mutations in the chaperonin groEL. Genetics 165:1651–1660.

Jabaji-Hare, S. H., Y. Piché, and J. A. Fortin. 1986. Isolation and structural characterization of soil-borne auxiliary cells of *Gigaspora Margarita* becker & hall, a vesicular-arbuscular mycorrhizal fungus. New Phytol103:777–784.

Jargeat, P., C. Cosseau, B. Ola'h, A. Jauneau, P. Bonfante, J. Batut, and G. Becard. 2004. Isolation, free-living capacities, and genome structure of *Candidatus* Glomeribacter gigasporarum, the endocellular bacterium of the mycorrhizal fungus *Gigaspora margarita*. J Bacteriol 186:6876.

Klasson, L., and S. G. Andersson. 2004. Evolution of minimal-gene-sets in host-dependent bacteria. Trends Microbiol 12:37–43.

Leone, M. R., G. Lackner, A. Silipo, R. Lanzetta, A. Molinaro, and C. Hertweck. 2010. An unusual galactofuranose lipopolysaccharide that ensures the intracellular survival of toxin-producing bacteria in their fungal host. Angew Chemie 122:7638–7642.

Li, L., C. J. Stoeckert, and D. S. Roos. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 13:2178–2189.

Librado, P., and J. Rozas. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–1452.

Liu, W., L. Fang, M. Li, S. Li, S. Guo, R. Luo, Z. Feng, B. Li, Z. Zhou, G. Shao, H. Chen, and S. Xiao. 2012. Comparative genomics of *Mycoplasma*: analysis of conserved essential genes and diversity of the pan-genome. PLoS ONE 7:e35698.

Lumini, E., V. Bianciotto, P. Jargeat, M. Novero, A. Salvioli, A. Faccio, G. Becard, and P. Bonfante. 2007. Presymbiotic growth and sporal morphology are affected in the arbuscular mycorrhizal fungus *Gigaspora margarita* cured of its endobacteria. Cell Microbiol 9:1716–1729.

Mengin-Lecreulx, D., and B. Lemaitre. 2005. Structure and metabolism of peptidoglycan and molecular requirements allowing its detection by the *Drosophila* innate immune system. J Endotoxin Res 11:105–111.

Mondo, S. J., K. H. Toomer, J. B. Morton, Y. Lekberg, and T. E. Pawlowska. 2012. Evolutionary stability in a 400-million-year-old heritable facultative mutualism. Evolution 66:2564–2576.

Moran, N. A., J. P. McCutcheon, and A. Nakabachi. 2008. Genomics and evolution of heritable bacterial symbionts. Annu Rev Genet 42:165–190.

Morris, J. J., R. E. Lenski, and E. R. Zinser. 2012. The Black Queen Hypothesis: evolution of dependencies through adaptive gene loss. mBio 3.2.

Nei, M., and T. Gojobori. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3:418–426.

Pizarro-Cerdá, J., and P. Cossart. 2006. Bacterial adhesion and entry into host cells. Cell 124:715–727.

Scarborough, C. L., J. Ferrari, and H. C. J. Godfray. 2005. Aphid protected from pathogen by endosymbiont. Science 310:1781–1781.

Sharon, G., D. Segal, J. M. Ringo, A. Hefetz, I. Zilber-Rosenberg, and E. Rosenberg. 2010. Commensal bacteria play a role in mating preference of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA 107:20051–20056.

Shigenobu, S., H. Watanabe, M. Hattori, Y. Sakaki, and H. Ishikawa. 2000. Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera sp.* APS. Nature 407:81–86.

Smith, S. E., and D. J. Read. 2010. Mycorrhizal symbiosis. Academic Press.

Stavrinides, J., H. C. McCann, and D. S. Guttman. 2008. Host-pathogen interplay and the evolution of bacterial effectors. Cell Microbiol 10:285–292.

Stukenbrock, E. H., T. Bataillon, J. Y. Dutheil, T. T. Hansen, R. Li, M. Zala, B. A. McDonald, J. Wang, and M. H. Schierup. 2011. The making of a new pathogen: insights from comparative population genomics of the domesticated wheat pathogen *Mycosphaerella graminicola* and its wild sister species. Genome Res. 21:2157–2166.

Sun, Q., R. Bukowski, C. Myers, L. Ponnala, S. Stefanov, M. Howard, A. Clark, and J. Pillardy. 2010. Computational Biology Service Unit: Cornell University core facility for computational biology. J Biomol Tech 21:S68.

Suyama, M., D. Torrents, and P. Bork. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucl. Acids Res. 34:W609–W612.

Sward, R. J. 1981. The structure of the spores of *Gigaspora margarita*. I. The dormant spore. New Phytol 87:761–768.

Turnbaugh, P. J., R. E. Ley, M. Hamady, C. M. Fraser-Liggett, R. Knight, and J. I. Gordon. 2007. The human microbiome project. Nature 449:804–810.

Van Buuren, M. L., L. Lanfranco, S. Longato, D. Minerdi, M. J. Harrison, and P. Bonfante. 1999. Construction and characterization of genomic libraries of two endomycorrhizal fungi: *Glomus versiforme* and *Gigaspora margarita*. Mycol Res 103:955–960.

Vandamme, P., B. Holmes, M. Vancanneyt, T. Coenye, B. Hoste, R. Coopman, H. Revets, S. Lauwers, M. Gillis, K. Kersters, and J. R. W. Govan. 1997. Occurrence of multiple genomovars of *Burkholderia cepacia*

in cystic fibrosis patients and proposal of *Burkholderia multivorans* sp. nov. Int J Syst Bacteriol 47:1188–1200.

Weber, E., and R. Koebnik. 2006. Positive selection of the Hrp pilin HrpE of the plant pathogen *Xanthomonas*. J. Bacteriol. 188:1405–1410.

Yang, Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate Lysozyme evolution. Mol Biol Evol 15:568–573.

Yang, Z. 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. Mol Biol Evol 24:1586–1591.

Yoon, H. S., D. C. Price, R. Stepanauskas, V. D. Rajah, M. E. Sieracki, W. H. Wilson, E. C. Yang, S. Duffy, and D. Bhattacharya. 2011. Single-cell genomics reveals organismal interactions in uncultivated marine protists. Science 332:714–717.

Zdobnov, E. M., and R. Apweiler. 2001. InterProScan – an integration platform for the signature-recognition methods in InterPro. Bioinformatics 17:847–848.

Zerbino, D. R., and E. Birney. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18:821–829.

Zhang, K., A. C. Martiny, N. B. Reppas, K. W. Barry, J. Malek, S. W. Chisholm, and G. M. Church. 2006. Sequencing genomes from single cells by polymerase cloning. Nat Biotech 24:680–686.

CHAPTER 3

Burkholderia Endosymbionts of *Rhizopus microsporus* Control Host Sexual Reproduction: A Window into Sexual Reproduction and Trisporic Acid Perception in Mucoromycotina

Abstract

Reproduction is an indispensable component of life. However, in many eukaryotic systems, reproduction can be manipulated by prokaryotes. Here, we show that sexual reproduction of *Rhizopus microsporus* is completely controlled by their bacterial endosymbiont *Burkholderia rhizoxinica*. Through phenotypic observation and transcriptome profiling, we have identified that endosymbionts control fungal reproduction through hijacking of host reproductive machinery. Specifically, bacteria control expression levels of Ras2, a signaling protein important for reproductive development as well as filamentous growth. We have also exploited endosymbiont control over reproduction to explore conservation of sexually relevant genes across Fungi, including the Mucoromycotina. We have identified several genes that appear core to all fungal reproduction, as well as reproduction related genes that are specific to members of the Mucoromycotina. In particular, we have found two candidate class C seven transmembrane G-protein coupled receptors (GPCRs), TriR1 and TriR2, which may be responsible for perception of trisporic acid during mating in Mucoromycotina. These receptors are closely related to the retinoic acid GPCRs present in animal systems.

Introduction

Reproduction is a basic requirement of life. In addition to affecting genetic diversity, reproduction plays an important role in dispersal and survival in harsh conditions. However, given its significance, it is surprising how often reproduction of one organism can be manipulated by other organisms, particularly by bacteria. For example, commensal gut symbionts of *Drosophila* can have long-term impacts on mating preference of their hosts by impacting pheromone production (Sharon et al., 2010). Similarly, *Wolbachia* endosymbionts are well known reproductive manipulators of many

arthropods, wherein they alter sex ratios, favoring females (Stouthamer et al., 1999, Jiggins et al., 2000). The ways in which bacteria accomplish reproductive manipulations are multifold, suggesting that many avenues to altering eukaryotic reproduction exist. In addition to affecting arthropods, endobacteria were recently shown to influence reproduction in fungi. In particular, *Burkholderia rhizoxinica* exert a complete control over the ability of their fungal host *Rhizopus microsporus* (Mucoromycotina) to reproduce asexually (Partida-Martinez et al., 2007c). In the absence of endosymbionts, these fungi can only grow in a vegetative state. However, if bacteria are reintroduced, normal sporangia and sporangiospores develop.

The species that hosts these *Burkholderia* endosymbionts, *R. microsporus*, is a highly versatile group of fungi that can thrive in a multitude of different environments. These fungi exist as saprotrophs, which are commonly involved in food spoilage, or as animal pathogens, which are the causal agents of Mucoromycosis, a highly fatal disease of immune-compromised individuals (Verweij et al., 1997). Due to their rapid growth rates, isolates of *R. microsporus* are also used for diverse industrial purposes, including crude protein production (Jasti et al., 2008, Jasti et al., 2009), amylase synthesis (Piexoto et al., 2003), tempeh preparation (Wang et al., 1975), and brewing (Celestino et al., 2006). Fungi harboring endosymbionts have been able to expand their niche-space even further and are capable of causing disease on rice due to endosymbiont synthesis of rhizoxin, a potent antimitotic (Partida-Martinez & Hertweck, 2005). With endosymbionts playing important roles in both the lifecycle and lifestyle of their fungal hosts, it is clear that further exploration of this symbiotic system is required to more fully understand host biology.

Interestingly, within every single asexual fungal spore, there is a single bacterial cell, indicating that the reproductive interests of these partners are tightly coupled (Partida-Martinez et al., 2007c). However, although asexual reproduction is certainly an extremely important component of the host's lifecycle, we cannot be certain of whether *B. rhizoxinica* are actually essential to host survival in nature

until we investigate their role in the host sexual reproduction. If endosymbionts play no role in this process, sexual reproduction may offer an avenue of escape for the host from endobacteria. Although the fitness cost would be substantial, i.e. the loss of asexual reproduction and virulence on rice, the host may be able to survive in other niches fairly well without its partner as long as sexual reproduction is frequent. As both mating types of *R. microsporus* are found across the world (Zheng et al., 2007), it is not clear how dependent these fungi are on asexual reproduction (and thus also their endosymbiont) for survival and dispersal. Historically, many fungi have evolved to survive using only a single mode of reproduction. For example, some basidiomycetes, such as the smuts and many agarics, lack asexual reproduction completely. Conversely, asexuality has evolved repeatedly across the fungal kingdom, and is even so prevalent that an entire phylum, the Glomeromycota, is thought to be completely asexual (Jany & Pawlowska, 2010).

Even though the association of *R. microsporus* with *B. rhizoxinica* is of great importance, this interaction is considered facultative. The main reason for this is that the endosymbiont, *B. rhizoxinica* is capable of survival outside of the host and can even be cultivated on general microbial media (Partida-Martinez & Hertweck, 2005). However, we would predict that due to endosymbiont vertical transmission and the population bottlenecking that ensues, *B. rhizoxinica* will eventually drift to a state where they depend on their host for survival (Rispe & Moran, 2000). This process appears to have already begun, as the endosymbionts genome is fairly reduced (≈3.95 Mb; Lackner et al., 2011) compared to free living relatives (8-10Mb). From an experimental standpoint, these features make the *R. microsporus-B. rhizoxinica* symbiosis extremely valuable for addressing questions relevant to evolution of endosymbiotic systems. Currently, nearly all known endosymbionts, including those that are non-essential to host survival, are uncultivable (Moran et al., 2008, Ghignone et al., 2012). Similarly, if hosts are obligately dependent upon their endosymbionts, they too are uncultivable without endobacteria present. Not only can *R. microsporus* and *B. rhizoxinica* be cultivated independently, the

symbiosis is re-established with relative ease. The fungal host can also grow rapidly in a vegetative state even without the endosymbiont present (Partida-Martinez et al., 2007c). Furthermore, although growth is slow outside of their hosts, transformation methods are available for genetic manipulations of the endosymbiont (Partida-Martinez et al., 2007c, Lackner et al., 2011).

We anticipate that this symbiosis can be further exploited to enhance our understanding of the reproductive biology of the host. *R. microsporus* is a member of a fairly unexplored lineage of lower fungi, the Mucoromycotina. For investigation of genes involved in fungal reproduction, because nearly all Mucoromycotina are resilient to genetic modification, they were largely abandoned in favor of the more easily manipulated ascomycetes and basidiomycetes. Currently, only one lineage from the Mucoromycotina, *Mucor circinelloides*, can be genetically modified, and this technology only came about in 2011 (Gutierrez et al., 2011). Consequently, at present, we only know the structure of the mating type (*MAT*) locus (Idnurm et al., 2008) and several genes involved in trisporic acid production (Tagua et al., 2012). Trisporic acids are the product of a complex breakdown of β-carotene (vitamin A) involving both interacting mates and are used as sex pheromones (Schachtschabel et al., 2005). However, the genes involved in perception of trisporic acid, as well as nearly all other genes involved in sexual reproduction in these fungi are still unknown. Without more information from the Mucoromycotina, even though much is known regarding sexual and asexual reproduction of the dikarya (ascomycetes and basidiomycetes), our understanding of reproduction in kingdom fungi is incomplete.

In this study, we have addressed whether the *Burkholderia* endosymbionts are essential for host survival by exploring their role in sexual reproduction. We found that, as with asexual reproduction, the fungal host was completely unable to reproduce sexually without endosymbionts present, implying that hosts are unable to escape their endobacteria and that these bacteria truly are essential for long-term survival of the host lineage. Unlike other bacteria, which may alter mating preferences or sex ratios, *B. rhizoxinica* currently appear to be the only endosymbiont with the capacity to completely control host

reproduction. Additionally, because fungi do not fail to grow without endosymbionts due to dietary restrictions (as is the case for many other essential endosymbioses; Moran et al., 2008), it is likely that endosymbionts are actively hijacking host machinery to specifically control reproduction and force host dependence. In order to better understand how endosymbionts accomplish this, we have conducted multiple phenotyping experiments as well as a large-scale RNAseq analysis and revealed that endosymbionts are manipulating a signaling molecule required to activate a Mitogen Activated Protein (MAP) Kinase cascade core to reproduction across Fungi.

As the *Burkholderia* endosymbionts control reproductive mode in a lineage of fungi that is recalcitrant to genetic manipulation, we have exploited them to explore genes relevant to sexual reproduction across the Mucoromycotina. In the process, we have uncovered two candidates involved in perception of trisporic acid, which appear to be Class C G-protein coupled receptors (GPCRs) bearing resemblance to those used to sense retinoic acid (another β-carotene derivative) in animal systems. Finally, our symbiotic study system has also allowed us to explore conservation of sexually relevant genes across the fungal kingdom.

Materials and Methods

Strains, Culture Conditions, and Curing of Bacterial Endosymbionts

All four strains of *R. microsporus* used in this study were acquired from the American Type Culture Collection (ATCC), including *R. microsporus* var *microsporus* ATCC 52813, *R. microsporus* var *microsporus* ATCC 52814, *R. microsporus* var *chinensis* ATCC 52811, and *R. microsporus* var *chinensis* ATCC 62417. For all experiments, fungi were either cultivated on half-strength potato dextrose agar (PDA) containing 2 g L^{-1} potato extract, 10 g L^{-1} dextrose, and 15 g L^{-1} agar. All plates were sealed with Parafilm® M (Pechiney Plastic Packaging Company, Chicago, Il), unless otherwise noted. Fungi were cured of their endosymbionts following protocols described in Partida-Martinez & Hertweck (2005).

Absence of endosymbionts was confirmed by PCR using *Burkholderia* specific primers GlomGiGf (Bianciotto et al., 2004) and LSUb 483r (Mondo et al., 2012) and following the cycling conditions described in Mondo et al. (2012).

Extraction, Cultivation, and Reinfection of Bacterial Endosymbionts

We have developed a new method for extracting bacteria from their fungal hosts. Young fungal mycelium (1-2 days old) containing endosymbionts was collected and finely chopped in 500 µL Luria-Bertani (LB) broth. Mycelium was then pressed gently to ensure that cellular contents were extracted into the surrounding broth. The broth was then filtered using a 2 µm filter to remove fungal debris and varying amounts of filtrate were added to LB plates containing 10 mL L^{-1} glycerol and 100 µg mL^{-1} cycloheximide. Single colonies were isolated and either grown on LB plates at 30°C or in 5 mL liquid cultures at 30°C incubated in the I2400 New Brunswick Scientific incubator shaker at 250 rpm. For reinfection of fungal hosts with their endosymbionts, a plug of medium the size of a P-1000 pipet tip was removed from a half-strength PDA plate and replaced with a plug of LB agar. Bacterial cultures were then placed on the LB agar plug, and a plug of cured fungus was either placed directly on wounded mycelium, or placed somewhere nearby on the plate.

Mating experiment

Two pairs of *R. microsporus* opposite mates, ATCC 52813 sex+ and ATCC52814 sex- as well as ATCC 62417 sex+ and ATCC 52811 sex-, were identified and tested for success of mating with or without bacterial endosymbionts present. Half-strength PDA plates were used for all mating experiments and opposite mates were placed at the edges of the plate, allowing fungi to grow towards each other and develop an interaction zone in the center of the plate (Figure 3.1). Plates were incubated in the dark at 30°C. Conditions tested included: (i) both mates harboring endosymbionts, (ii) sex+ only harboring endosymbionts, (iii) sex- only harboring endosymbionts, and (iv) both mates lacking endosymbionts. As ATCC 62417 grows very poorly after removal of endosymbionts, mating pair ATCC 62417 and ATCC

52811 was only used to test conditions (i) and (iv). All conditions were replicated ten times and the entire experiment, including re-curing of endosymbionts, was repeated three times. Prior to investigation of the role endosymbionts play in sexual reproduction, we tested efficacy of fungal mating in both darkness and in ambient light. We found that, with respect to mating, *R. microsporus* had no light preference.

Attempts to Restore Mating in the Absence of Bacteria

For attempts to restore mating in the absence of bacteria, cured opposite mates were grown in half-strength PDA amended with various substrates (at least 5 replicates per condition), including: 3 g L^{-1} $NaNO_3$, 0.5 g L^{-1} $NaNO_2$, 1 mg L^{-1} thiamine, filtered extracts from successfully mating 25 mL liquid cultures (filtered through 0.22 micron filter, then discs were soaked in filtrate and placed on half-strength PDA), and 16.6 mg L^{-1} β-carotene. Additionally, aside from amending media with substrates, we attempted to restore mating through cold treatment of cured fungal lines (1-, 2-, 4-, and 10-week incubation at 4°C), mating partners in ambient light or complete darkness, or through reintroduction of bacterial endosymbionts. Finally, as moisture is known to impact the reproductive mode of *Rhizopus* (Blakeslee, 1904), we attempted to mate fungi grown on half-strength PDA at a high moisture level generated by adding 1 mL H_2O to each plate and sealing plates with Parafilm M, and at a low moisture level generated by sealing plates with Micropore™ surgical tape (3M Health Care, St. Paul, MN).

As environmental stressors including oxidative stress and osmotic stress are often involved in fungal reproduction, we tested their impacts on fungal growth with or without bacteria by either amending PDA with 0.3, 0.6, 1.2, 2.4 or 10 mM H_2O_2, or with 0.4, 0.8, or 1.6 M NaCl.

RNA-seq

Reproductive mode in lower fungi is almost completely unexplored. However, as we are able to cure fungi of their bacterial endosymbionts, and these endobacteria appear to control both modes of reproduction in their fungal host (Partida-Martinez et al., 2007c; see below “Bacterial Endosymbionts

Control Sexual Reproduction of their Hosts"), we have used these bacteria as a tool to explore the genes involved in reproduction of *Rhizopus*. Specifically, we conducted an RNAseq experiment comparing across six different conditions (all grown on half-strength PDA; each condition had two biological replicates), including: (1) *R. microsporus* ATCC 52813 growing alone with endosymbionts, (2) *R. microsporus* ATCC 52813 growing alone without endosymbionts, (3) *R. microsporus* ATCC 52814 growing alone with endosymbionts, (4) *R. microsporus* ATCC 52814 growing alone without endosymbionts, (5) both opposite mates growing together with bacteria, and (6) both opposite mates growing together without endosymbionts. For each condition, fungal plugs were placed at the edge of the plate and, after six days, equal amounts of mycelium from the center of the plate were harvested. We chose to harvest after six days because at this point opposite mates were actively undergoing sexual reproduction in the condition containing both fungi and their endosymbionts. Each biological replicate consisted of five culture plates, which were pooled prior to RNA extraction. Total RNA was then extracted using the ToTALLY RNA kit (Ambion®) in order to recover both fungal and bacterial transcripts. In all conditions, ribosomal RNA was removed for both the fungal host and their bacterial endosymbionts using two Ribo-Zero rRNA Removal Kits (Epicentre, Madison, WI): Human/Mouse/Rat for fungi, and Gram-Negative Bacteria for endosymbionts. After rRNA was removed, total RNA sequencing libraries were constructed and sequenced by the Cornell Sequencing Facility using the TruSeq RNA Sample Preparation Kit (Illumina, San Diego, CA). Samples were sequenced using the Illumina Hi-Seq 100 bp paired end platform.

Illumina data were quality controlled using the FASTX-Toolkit (Gordon & Hannon, 2010) and assembled using Trinity (Grabherr et al., 2011). Transcript abundances were measured using RSEM (Li & Dewey, 2011). FPKM values were then used for analysis of differential expression across conditions. Custom scripts were written to investigate dramatic fold changes (at least 1.5 fold different) between conditions. However, for the purposes of this work, fold changes needed to be consistent across all

conditions compared. For example, when identifying fungal gene expression impacted by bacterial presence/absence, gene expression for all conditions with fungi lacking endosymbionts (3 conditions, 6 total replicates) had to be at least 1.5 fold different from all conditions with fungi harboring endosymbionts (3 conditions, 6 total replicates). We expect that this approach may be overly conservative, as it only recovers genes that are differentially expressed in both fungal hosts and will not capture genes that may show differential expression only in one strain. However, since the same phenotype is observed in both hosts when endosymbionts are removed, our approach will successfully capture the fungal genes mostly responsible for this. Differentially expressed datasets regarding genes involved in sexual reproduction were similarly collected.

Parsing bacterial from fungal transcripts

After Trinity *de-novo* assembly (Grabherr et al., 2011), we parsed bacterial transcripts from the complete dataset using a BLAST based approach. The transcriptome was used to query a database containing reference genome sequences from *Burkholderia rhizoxinica* HKI 454 (Lackner et al., 2011) as well as *Rhizopus delemar* 99-880 (Ma et al., 2009). Any transcripts that matched better to the *B. rhizoxinica* genome were then separated into a bacteria-specific file and analyzed separately.

Identification of Sexual Reproduction Genes Across Fungi

To identify which genes may be core to reproduction across fungi, including Mucorales, we looked for presence/absence across the Fungi of genes proven to play a role in sexual for various fungal lineages. Fungal lineages surveyed include *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Neurospora crassa*, *Aspergillus nidulans*, *Cochliobolus heterostrophus*, *Cryptococcus neoformans*, *Ustilago maydis*, *Coprinopsis cinerea*, *Mucor circinelloides*, *Phycomyces blakesleeanus*, *Rhizopus delemar* and *Rhizopus microsporus* (transcripts only). Genes experimentally shown to have reproductive phenotypes in any of the following fungal lineages were queried for presence/absence across Fungi: *Saccharomyces cerevisiae* (≈20 genes), *Neurospora crassa* (≈200 genes), *Aspergillus nidulans* (≈70

genes), or *Ustilago maydis* (≈20 genes). *MAT* loci were also collected from sequences available in GenBank for *Phycomyces blakesleeanus*, *Rhizopus delemar*, and *Rhizopus oryzae*. Additionally, we included genes, which were differentially expressed during sexual reproduction of *R. microsporus* identified during this study.

Similar to adding tags to samples when multiplexing for sequencing, all sexually relevant genes were tagged with a 'Sex@' tag (for example, *Aspergillus nidulans* gene AN6230|*medA* becomes Sex@AN6230|*medA*) and used in a large-scale clustering of orthologous genes across Fungi. Coding sequences from all genomes were used in an All vs All Blast-p (Altschul et al., 1990) with default parameters: e-value cutoff = $1x10^{-3}$, maximum matches = 500. OrthoMCL (Li et al., 2003) was used to identify orthologs. Even though we expect some artificial clustering of paralogs, because the phylogenetic distance across which we are making phylogenetic comparisons is fairly large, we chose to use loose clustering criteria to define ortholog groups. OrthoMCL was run with parameters: mode 3, pi_cutoff 0, pv_cutoff 1e0, and inflation 0. Clusters containing sexually relevant genes were collected using a custom script written to extract any ortholog cluster containing a gene tagged with a 'Sex@' symbol.

For clarity, we have defined several levels of conservation of sexually relevant genes. If clusters containing sexually relevant genes were found to have orthologs across all fungi, they were considered 'core' sexual genes. Similarly, if we found genes that were present across all fungi, but missing from some taxa (for example, present in filamentous Ascomycota, Basidiomycota, and Mucorales, but missing from ascomycete yeasts), they were considered as 'ancestral' sexual genes. Genes present only in one lineage of fungi are discussed as either 'Ascomycota specific', 'Basidiomycota specific', 'Mucorales specific', or 'Dikarya specific'.

For phylogeny reconstruction, proteins were aligned using MUSCLE (Edgar, 2004) and trees were constructed with Mr.Bayes v3.1.2 (Huelsenbeck & Ronquist, 2003) using the MtRev substitution model.

The MCMC chain was allowed to run for 1.2 million generations and trees were sampled every 200 generations. Posterior probabilities were calculated after a burn-in of 300,000 generations.

Results

Bacterial Endosymbionts Control Sexual Reproduction of their Hosts

Previously, it was discovered that bacterial endosymbionts of *Rhizopus microsporus* completely control the capacity of their fungal hosts to reproduct asexually (Partida-Martinez et al., 2007c). Consequently, we were interested in investigating whether this control also extends to sexual reproduction. In order to test this, we manipulated bacterial presence/absence and conducted mating experiments using two different pairs of opposite mates, including pairs *R. microsporus* var *microsporus* ATCC 52813 (sex+) and ATCC 52814 (sex-) as well as *R. microsporus* var *chinensis* ATCC 62417 (sex+) and ATCC 52811 (sex-). We found that bacterial endosymbionts control the ability of their fungal hosts to reproduce sexually (Figure 3.1). Control appeared to be complete, as mating of fungal hosts lacking endosymbionts resulted in the inability to produce any sexual structures at all. Furthermore, sexual reproduction was arrested prior to initiation, as no observable trisporic acid was produced and no mating initials were found. All interacting partners were observed for at least one month, during which no sexual reproduction occurred without bacteria present. In contrast, wild-type controls, in which both partners harbored endosymbionts, produced zygospores and visible amounts of trisporic acid identifiable as a yellow pigmentation at the interaction zone as early as one day after mates contacted one another. While most of our work was done using *R. microsporus* var *microsporus* mates, to test the breadth of this phenotype, we also conducted mating experiments with *R. microsporus* var *chinensis* and observed the same results.

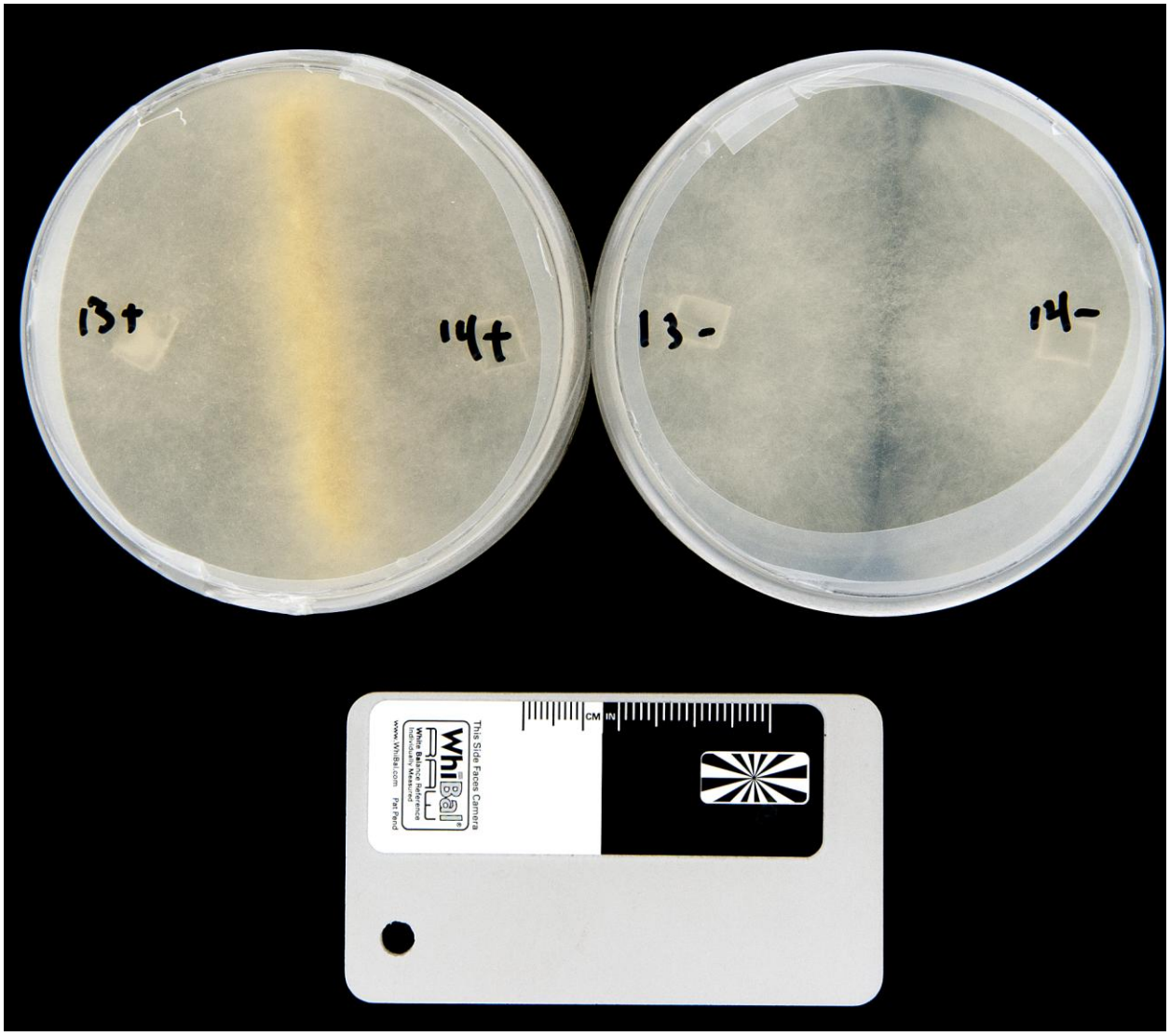

Figure 3.1. Sexual reproduction is endosymbiont dependent. Opposite mates ATCC 52813 (sex+) and ATCC 52814 (sex-) either growing together with bacterial endosymbionts present (left), or without endosymbionts (right). 13+: ATCC 52813 with endosymbionts, 14+: ATCC 52814 with endosymbionts, 13-: ATCC 52813 without endosymbionts, 14-: ATCC 52814 without endosymbionts. Yellow pigmentation at the interaction zone is indicative of trisporic acid production. Dark brown coloration at the interaction zone represents zygospores produced through successful mating between opposite mates.

Although we do not expect this is the case for *Rhizopus* (Blakeslee, 1904), it has been reported in other fungi that the sex+ mating type often loses fertility over multiple rounds of serial subcultivation under laboratory conditions (Blakeslee, 1904). As the approach to curing fungi of their endosymbionts

requires at least six rounds of subcultivation (Partida-Martinez & Hertweck, 2005), we generated "mock-cured" lines, which were wild-type lines subjected to the same number of subcultivation rounds as cured lines, except antibiotic was not added. We found that serial propagation had no effect on sexual reproduction, as the mock-cured lines were just as effective at mating as normal wild-type, while cured lines failed to reproduce sexually.

Interestingly, we also observed that very few bacteria are required to stimulate sexual reproduction. However, it appeared that when limited bacteria were present in the mycelium, very few zygospores were produced. We found that even if fungi failed to sporulate asexually and endosymbionts were present in amounts undetectable by PCR, sexual reproduction could still occur. Further curing of these fungal lines revealed no sexual reproduction, indicating that endobacteria must be completely removed to entirely abolish fungal sexual reproduction.

Overall, we conclude that fungal sexual reproduction is tied to endosymbiont presence. However, even minute amounts of endosymbionts can be sufficient to allow for sex. From these results, we hypothesized that the control of sexual reproduction by *B. rhizoxinica* endosymbionts could be through one of several means: (1) fungi require a particular endosymbiont-derived metabolite to initiate reproduction, (2) endosymbionts are necessary to sense the environmental cues which trigger fungal mating, or (3) fungal sexual reproductive pathways are manipulated by endosymbionts. We tested these hypotheses in a series of experiments.

Are endosymbiont-derived metabolites required for fungal mating?

To test the hypothesis that bacteria produce a metabolite necessary for fungal reproduction, we investigated whether we could restore mating in the absence of endosymbionts through addition of any particular compounds to the media. We searched the *B. rhizoxinica* genome for any metabolites made by endosymbionts that may have relevance to fungal mating, and found that these bacteria are able to synthesize thiamine (vitamin B_1), various forms of nitrogen as well as β-carotene intermediates. When

we added these compounds to half-strength PDA and attempted to mate compatible but bacteria-free partners, we were unable to restore sexual reproduction under any circumstance. Moreover, addition of these compounds did not affect mating of wild-type partners. Additionally, we tried: adding filtrates from successfully mating cultures to mating interactions of cured lines, conducting mating experiments in high or low moisture conditions, and mating partners in light or dark conditions. We also tried cold treating cured lines for 1, 2, 4, or 10 weeks; after 10 weeks of cold treatment, cured lines failed to grow. In all of these conditions, we were unsuccessful at restoring either sexual or asexual reproduction.

Finally, as fungal growth and development are often tied to stress response, we tested the impact of oxidative and osmotic stress on triggering asexual reproduction in cured lines. We found that neither type of stress could induce cured fungi to reproduce asexually, although if subjected to high osmotic stress, cured lines would overproduce a yellow pigment reminiscent of β-carotene or one of its derivatives. Based on these collective observations, we rejected the hypothesis that endosymbionts produce a particular metabolite that is required by fungi for sexual reproduction.

Response to Environmental Cues is Altered without Endosymbionts

Rhizopus and other members of the Mucorales are known to use moisture levels as an environmental cue to regulate asexual versus sexual reproduction (Blakeslee, 1904). The lower the moisture content, the higher the preference is for asexual reproduction and *vice versa*. Interestingly, in the course of our work, we discovered that fungi respond differently to the quality of the surrounding atmosphere depending on whether endosymbionts were present or not. When fungi harbored bacteria, as expected, they responded to openings in the Parafilm by overproducing asexual sporangia (Figure 3.2), whereas cured lines responded by overproducing a yellow pigment reminiscent of trisporic acid immediately below the opening (Figure 3.2). Based on this observation, we designed an experiment wherein we cultivated both *R. microsporus* var *microsporus* mating types either with or without endosymbionts (3 replicates each) on half-strength PDA and intentionally created an opening in the

Parafilm after 7 days of growth. We found that 4/6 plates of wild-type (WT) fungi overproduced sporangia at the opening site within a week, while 6/6 plates of fungi lacking endosymbionts overproduced yellow pigmentation right below the opening in the Parafilm after 21 days. No sexual structures were present in these regions. We additionally confirmed that yellowed sites were endosymbiont-free by DNA extraction and PCR using *Burkholderia*-specific primers (Mondo et al., 2012).

Although we have not confirmed whether the yellow pigmentation is indeed due to trisporic acid production or other β-carotene derivatives, these results indicate that fungi are cable of reacting to environmental cues without endosymbionts. However, the way in which they respond in the absence of bacteria seems to differ from the response in the endobacterial presence. Of note, the environmental cue of moisture level to is one of the most important environmental triggers relevant to reproductive mode (Blakeslee 1904), which lead us to reject the hypothesis that endosymbionts are necessary to sense the environmental cues which trigger fungal mating. In contrast, these results suggest that bacteria may be mediating fungal signaling in response to external cues.

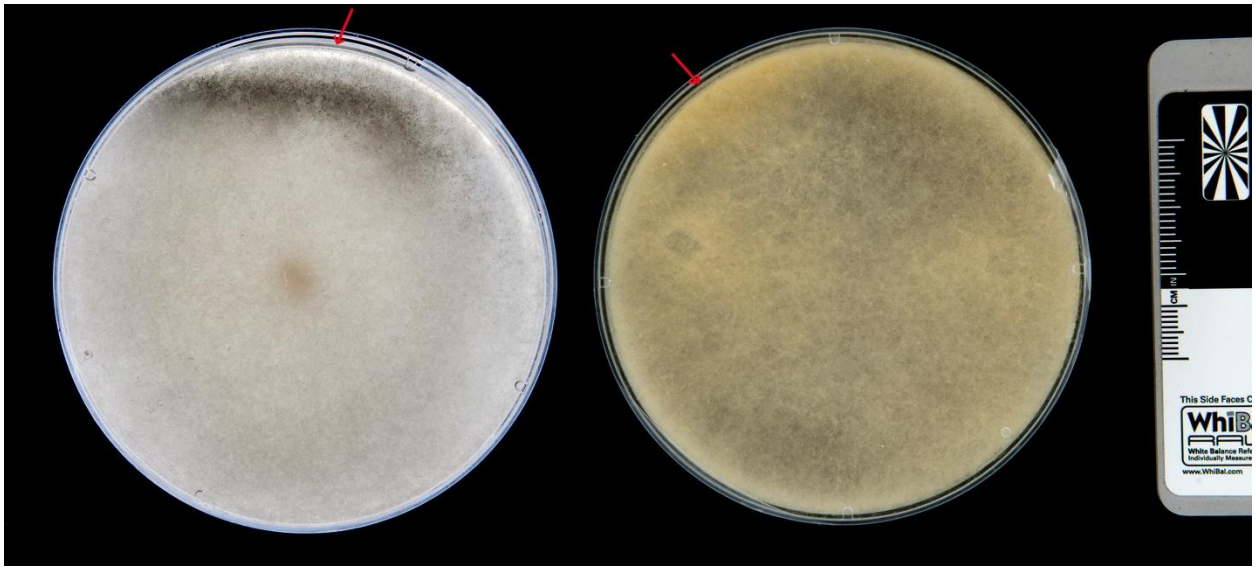

Figure 3.2. Fungi respond differently to environmental cues depending on endosymbiont presence. Red arrows point towards openings in the Parafilm sealing the plate. Left: ATCC 52814 with endosymbionts present. Black pigmentation results from production of asexual sporangia surrounding the Parafilm opening. Right: Opposite mates growing on the same plate without endosymbionts present. No yellowing occurs at the zone of interaction, however, a bright yellow pigment is deposited right below the Parafilm opening.

Reintroduction of Endosymbionts and its Impact on Fungal Reproduction

Previous work has shown that reintroduction of endobacteria restores asexual reproduction (Partida-Martinez et al., 2007c). However, the breadth of this complementation is uncertain, as restoration of host reproduction was only tested using endosymbionts isolated from the same host strain. In order to understand whether different genotypes of endosymbionts isolated from various *R. microsporus* hosts can restore asexual sporulation in any host, we attempted to reintroduce endosymbionts into their native and non-native fungal hosts and investigate how this impacts fungal reproduction. For this work, we extracted endosymbionts from ATCC 52813, ATCC 52814, and ATCC 62417 and reintroduced each of these bacteria into all cured host strains. We found that regardless of the mode of reinfection (either placing wounded fungi directly on bacterial colonies, or nearby), all three strains of bacteria could successfully reinfect and trigger asexual sporulation of ATCC 52813 and ATCC 52184. In contrast, we were only able to reinfect ATCC 62417 with endosymbionts from ATCC 52813

(this combination also triggered fungal sporulation) but not from ATCC 52814. We believe that this failure to reinfect with other bacterial strains is mostly due to the poor growth exhibited by this fungus after removal of endosymbionts, leading to experimental difficulties during reinfection. Nonetheless, these reinfection experiments reveal that, importantly, both the bacterial trait that activates sporulation of its host, as well as the corresponding fungal component, are conserved across all of partner lineages. This finding also reveals that this characteristic evolved prior to diversification of endosymbiont-harboring *R. microsporus*.

Additionally, we used reinfected lines to test whether reintroduction of endosymbionts can restore sexual reproduction. Much to our surprise, we found that although asexual reproduction was restored immediately upon reinfection, fungi were still unable to complete sexual reproduction even if paired with their native endosymbionts (0/10 mating interactions produced zygospores). However, many mycelial aggregates were present at the zone of interaction between bacteria-free mates, suggesting an initiation and subsequent failure of reproduction. Consequently, we speculated that an additional trigger not present in our experimental design was required to elicit sexual reproduction after reintroduction of endosymbionts. To address this hypothesis, we cold-treated the re-infected mates for 4 or 8 weeks to simulate an overwintering period, but were unable to restore mating following neither treatment.

RNA-seq: Main Goals and Approach

Through our phenotyping work, we found that the way in which endosymbionts are manipulating their hosts is conserved across all host-symbiont combinations. However, how the endobacteria are actually altering fungal reproduction is still unknown. These endosymbionts are likely not providing their hosts with a particular metabolite required for fungal reproduction, nor are they responsible for perceiving environmental cues required to activate reproduction. However, our results

suggest that endobacteria may impact fungal signaling in response to environmental cues. Consequently, it seems likely that endobacteria are actively manipulating fungal reproductive pathways.

Because *B. rhizoxinica* endosymbionts completely control the capacity of their host to reproduce, we have exploited this symbiosis to ask two main questions. First, we wanted to pinpoint the fungal reproductive defects resulting from the removal of endosymbionts. Second, we wanted to identify the genes involved in sexual reproduction in Mucoromycotina, which are largely unexplored in this respect. To address these two questions we conducted an RNA-seq experiment including six different conditions. Conditions included both opposite mates of *R. microsporus var microsporus*, ATCC 52813 (sex+) and ATCC 52814 (sex-), growing alone with or without endosymbionts, as well as both partners in mating interaction either with endosymbionts (successful mating), or without endosymbionts (failure to initiate mating). Total RNA was extracted and sequenced so we could recover both fungal and bacterial transcripts. Sequencing reads were quality controlled and then transcriptomes where assembled *de novo* using Trinity (Grabherr et al., 2011). For analysis of differential expression, we investigated fold changes across different samples. However, because we know that the same phenotype is observed in both fungal hosts (sex+ and sex- isolates), we chose to make conservative comparisons requiring both opposite mates to show similar expression patterns, essentially turning them into additional biological replicates. This approach enabled us to narrow our focus onto genes that are most relevant to our particular questions and effectively remove any strain-specific changes.

Goal 1: Differential Gene Expression due to Endosymbiont Presence

To visualize the types of genes, which are impacted by endobacterial presence, we interrogated our RNA-seq dataset and recovered any fungal gene, which was differentially expressed in response to bacterial presence. Here, transcripts needed to be at least 1.5 fold differentially expressed in all conditions where fungi lacked endosymbionts (3 conditions, 6 replicates) in comparison to all conditions where fungi harbored endobacteria (3 conditions, 6 replicates).

We found that, of the 64542 fungal transcripts recovered, only 153 appeared to be dramatically differentially expressed in both fungi due to endobacteria. A subset of these proteins, their protein annotations and their relative fold changes are shown in Table 3.1. For a full list, see Table A3.1. Interestingly, exactly half of these differentially expressed transcripts were under-expressed (77 transcripts) in endosymbiont free fungi, and half were over-expressed (77 transcripts). While these numbers may seem fairly low, we believe it is due to the very conservative cutoff criteria we selected for this analysis.

Out of the 77 under-expressed transcripts, 24 had no BLAST matches and only six were hypothetical proteins. At least four of these transcripts encoded predicted proteins involved in regulation in response to environmental signals, including: a blue light regulator similar to WctD (White-Collar family) from *Phycomyces blakesleeanus*, inactivation of a lowly expressed CMGC/CLK protein kinase related to LAMMER kinase Lkh1 (regulates cyclic adenosine monophosphate (cAMP) cycling in *Schizosaccharomyces pombe* (Yu et al., 2013)), Sho1, a protein involved in osmolarity sensing, which interacts with the HogA stress response signaling cascade as well the Cdc42 filamentous growth pathway (Tatebayashi et al., 2006, Cullen et al., 2004), and Prr2p, a pheromone response regulator characterized in *Saccharomyces cerevisiae* (Burchett et al., 2001). We additionally found down-regulation of one mucin, which may have implications in interactions with Rho-GTPases and MAPK cascades (Cullen et al., 2004). Two transcripts were identified which show similarity to proteins differentially expressed during reproduction in other fungi, including a conidiation specific protein present in *Sordaria macrospora* as well as a GPI-anchored protein up-regulated during meiosis in *Schizosaccharomyces pombe.* Finally, one small G-protein related to Ras2 from *Ustilago maydis* was found to be down-regulated in the absence of bacteria. Small Ras proteins are well known across fungi for their involvement in reproduction and cAMP cycling (Budovskaya et al., 2005, Dumaz & Marais,

2005). cAMP is often involved in cell cycle regulation and thus has implications in fungal growth and development.

Out of the 77 over-expressed transcripts, only one had no BLAST matches, and 13 were hypothetical proteins. We found that at least five of these transcripts encoded putative proteins involved in various aspects of cell wall regulation, including chitin synthases, chitinases, and β-glucan synthesis. We also found up-regulation of five genes involved in energy production and mitochondrial functioning. These transcripts include genes involved in assembly of the cytochrome c oxidase complex (cytochrome c oxidase subunit 1), two genes involved in acetyl-CoA metabolism, as well as several genes related to mitochondrial functioning, including uroporyphyrinogen-III synthase and a mitochondrial NADH dehydrogenase. Additionally, there was very weak activation of a surfeit-1-like protein (involved in cytochrome C oxidase complex construction) as well as down-regulation of a mitochondrial iron ion transporter in the absence of endosymbionts, altogether suggesting that mitochondrial functioning and energy production is impacted by the symbiosis. There was also upregulation of seven proteins relevant to regulation of DNA processing, two of which were involved in gene silencing (dicer-2 and rDNA silencing protein 4) and one involved in pre-mRNA splicing (Slu7).

Interestingly, we also found up-regulation of Cin1, a multi-modular protein characterized in *Cryptococcus neoformans* (Shen et al., 2010). This protein contains a RhoGEF domain and was shown to interact with Cdc42, a small Rho GTPase essential for cell cycle regulation, particularly cell polarity and organization of actin and septins (Shen et al., 2010). Cdc42 is conserved across both fungi and animals. Supporting a similar role of the Cin1-like protein in *Rhizopus*, we also observe down-regulation of Mad2, a mitotic spindle formation checkpoint component, as well as up-regulation of a kinesin-like protein, an LSB5-like protein, and a Sla1-like protein, all of which are involved in regulation cytoskeletal organization. Finally, we also observed up-regulation of a MARK1-like protein kinase, which plays a role in microtubule disruption (Drewes et al., 1997)

Altogether, differential gene expression in the absence of bacterial endosymbionts suggests that fungi are altered not only in how they respond to environmental cues, but also in mitochondrial functioning, cAMP regulation, and cell cycle regulation. The upregulation of genes involved in both mitochondrial functioning and fungal growth (*i.e*. chitin and β-glucan related genes) suggest that, in addition to reproduction, vegetative growth also is impacted by this symbiosis. We believe that the most dramatic impact exerted by endobacteria on their fungal hosts is through altering signal transduction, in particular in cell cycle regulation. However, even though we have identified which genes are differentially expressed due to endosymbiont presence, more information is needed to pinpoint which gene(s) is responsible for the inability of *R. microsporus* to reproduce in the absence of endosymbionts. Specifically, we must first explore the genes involved in sexual reproduction of Mucoromycotina and their conservation across fungi. With this information in hand, we will then be able to identify which differentially expressed transcripts due to endosymbiont presence are also important components of fungal reproductive pathways.

Table 3.1. Subset of differentially expressed transcripts due to endosymbiont presence. 13+: ATCC 52813 with endosymbionts, 14+: ATCC 52814 with endosymbionts, 13-: ATCC 52813 without endosymbionts, 14-: ATCC 52814 without endosymbionts, 13+x14+: ATCC 52813 with endosymbionts interacting with ATCC 52814 with endosymbionts, 13-x14-: ATCC 52813 without endosymbionts interacting with ATCC 52814 without endosymbionts. Fold changes are calculated based on expression patterns of ATCC 52813 with endosymbionts. >1.5 fold upregulation is highlighted in dark red, >3.0 fold upregulation is highlighted in bright red, >1.5 fold downregulation is highlighted in dark green, >3.0 fold downregulation is highlighted in bright green. Average fold change is calculated by converting 0 to 0.001.

Transcript	13+	14+	13-	14-	13+x14+	13-x14-	Average Fold Change	Putative protein Annotation	Physiological Process
173914_c0_seq17	10	4	0	0	17	2	>100	conidiation-specific expression protein	Reproduction
173429_c0_seq1	5	1	0	0	2	0	>100	Prr2 (pheromone response regulator 2: *Saccharomyces cerevisiae*)	Reproduction
160904_c0_seq3	23	33	7	10	62	4	7	Ras2	Reproduction and cAMP signaling
159019_c1_seq4	45	31	9	6	29	9	5	Mucin, putative	potential cAMP signaling
171838_c0_seq2	1	2	0	0	2	0	>100	High osmolarity signaling protein Sho1	cAMP, stress, and potentially reproduction
173026_c0_seq4	0.2	0.1	0	0	0.2	0	19	CMGC/CLK protein kinase (LAMMER)	cAMP regulation
172773_c2_seq9	1	0	12	3	0	5	>100	Cin1	cAMP and cytoskeleton regulation
161109_c0_seq1	21	19	8	10	18	8	2.37	mitotic spindle checkpoint component Mad2	Cytoskeleton
173248_c0_seq18	0	0	1	1	0	6	>100	PREDICTED: kinesin-like protein KIF21A	Cytoskeleton
98130_c0_seq1	0	0	2	3	1	3	92	related to LSB5-possible role in the regulation of actin cytoskeletal organization	Cytoskeleton
174104_c3_seq89	0	1	6	3	0	7	>100	related to SLA1-cytoskeleton assembly control protein	Cytoskeleton
170620_c0_seq5	0	0	1	0.1	0	1	61	MARK1 protein kinase	Cytoskeleton
174389_c1_seq2	4	23	0	1	12	2	>100	mitochondrial iron ion transporter (predicted)	Energy
171789_c0_seq11	0	0	0.01	0.01	0	0.01	10	Surfeit-1	Energy
195842_c0_seq1	1	1	3	3	1	4	3	uroporphyrinogen-III synthase	Energy
174852_c0_seq1	1351	2118	6533	8176	1684	7677	5	Cytochrome c oxidase subunit 1	Energy
158847_c0_seq1	14	13	66	45	9	40	4	NDI, mitochondrial NADH dehydrogenase	Energy
172145_c0_seq1	47	65	1243	655	30	757	21	acetyl-CoA synthetase-like protein	Energy
98048_c0_seq2	18	17	42	36	13	54	2	probable acyl-CoA oxidase	Energy
172538_c0_seq2	1	2	0	0	2	0	>100	Endo-1,3-beta-glucanase, family GH81	Cell wall
159745_c0_seq1	14	3	132	66	9	95	18	Cell wall beta-glucan synthesis	Cell wall
174890_c0_seq1	437	688	1357	1475	505	1170	2	chitinase C1	Cell wall
175461_c0_seq1	14	16	239	82	16	379	15	chitinase	Cell wall
173358_c0_seq1	0	1	17	5	1	7	20	chitin synthase	Cell wall

175217_c0_seq1	**45**	**39**	**109**	**95**	**46**	**104**	**2**	chitin deacetylase	Cell wall
169217_c0_seq3	**0**	**0**	**42**	**74**	**0**	**29**	**>100**	ERG4/ERG24 ergosterol biosynthesis protein	Cell wall
160780_c0_seq1	**3**	**0**	**8**	**6**	**3**	**9**	**11**	transforming growth factor beta regulator 1	Growth
171214_c0_seq10	**1**	**1**	**0**	**0**	**1**	**0**	**>100**	DEAD/DEAH box helicase	DNA regulation
189801_c0_seq1	**2**	**1**	**8**	**5**	**3**	**9**	**5**	ATP-dependent RNA helicase HAS1	DNA regulation
170867_c0_seq1	**31**	**18**	**121**	**107**	**39**	**121**	**4**	histone H2A	DNA regulation
162668_c0_seq4	**7**	**2**	**15**	**11**	**5**	**17**	**4**	arid/bright DNA binding domain-containing protein Swi1	DNA regulation
99222_c0_seq1	**2**	**1**	**5**	**4**	**2**	**6**	**3**	Swi1-interacting protein swi3	DNA regulation
97828_c0_seq1	**36**	**25**	**63**	**91**	**33**	**79**	**2**	Pth12p (homeobox protein)	DNA regulation
173320_c4_seq1	**21**	**28**	**99**	**66**	**23**	**47**	**2**	WD40 repeat-containing protein	DNA regulation
178252_c0_seq1	**2**	**3**	**45**	**72**	**5**	**45**	**25**	increased rDNA silencing protein 4	DNA regulation
174987_c2_seq1	**0**	**0**	**5**	**7**	**1**	**6**	**18**	transposase-like protein	DNA regulation
174598_c0_seq42	**0**	**0**	**0.5**	**0.5**	**0.01**	**0.2**	**28**	dicer-2 protein	DNA regulation
180109_c0_seq1	**7**	**7**	**19**	**16**	**7**	**19**	**2**	pre-mRNA-splicing factor Slu7	DNA regulation
175799_c0_seq1	**18**	**16**	**37**	**47**	**11**	**55**	**3**	small nuclear ribonucleoprotein E	DNA regulation
169061_c0_seq4	**19**	**22**	**4**	**8**	**20**	**7**	**3**	WctD (White collar: *Phycomyces blakesleeanus*)	Light response regulation

Goal 2: Mating Relevant Genes of *Rhizopus microsporus*

Genetic underpinnings of sexual reproduction in Mucoromycotina are poorly understood at present. Only the structure of the mating type (*MAT*) locus (Idnurm et al., 2008) and several genes involved in trisporic acid synthesis (Tagua et al., 2012) are known. Our RNA-seq dataset enabled us to uncover many genes relevant to this process. For this work, we used a two-tiered approach. First, we identified any genes that were differentially expressed during sexual reproduction when compared to either opposite mate growing separately with endosymbionts present. This was done to ensure that the observed differences were attributable to sexual reproduction rather than endosymbiont presence. Second, after recovering these transcripts, we further reduced our dataset by removing any transcripts with similar expression patterns to the condition where both fungi are interacting but lack endosymbionts. Our prediction is that any transcripts showing similar expression patterns between these conditions are likely involved in vegetative interactions between fungi and not in sexual reproduction, as cured lines growing together fail to initiate mating.

We found that of the 64542 fungal transcripts, 1643 were differentially expressed during sexual reproduction. 1338 of the differentially expressed genes were up-regulated, while only 305 were down-regulated compared to other conditions. Many of the transcripts up-regulated during sexual reproduction were not expressed in any other conditions analyzed and were often expressed at low levels. Rather than discussing the exact transcripts we identified as differentially expressed during sexual reproduction, we chose to discuss our observations in the context of how they compare to sexually relevant genes known from other fungi.

Conservation of Reproductively Relevant Genes across Fungi

We have generated a dataset of genes relevant to sexual reproduction in *Rhizopus microsporus*. To interpret these observations, we placed them within the context of what is known regarding reproductive biology of other fungi. To date, comparisons of reproductive pathways have been

restricted to the Dikarya (Ascomycota and Basidiomycota) (Dyer & O'Gorman, 2012, Müller et al., 2003, Xue et al., 2008). By including Mucoromycotina, we were able to infer key pathways involved in reproduction in all fungi. With this approach, we were also able to firmly anchor our results regarding endosymbionts' impact on sexual reproduction. Additionally, we were interested in understanding what genes are important for sexual reproduction specifically in the Mucoromycotina. For example, unlike other fungi, Mucoromycotina utilize trisporic acids as mating pheromones. However, the genes involved in perception of trisporic acid have not been known and we set out to identify them.

To assess conservation of reproductive genes, we conducted a large-scale orthologous gene clustering experiment. We used information on genes known through experimental manipulation to have reproductive phenotypes in various higher fungi (367 genes, including 129 well characterized genes) and included them as well as sexually relevant *R. microsporus* transcripts as bait (1489 genes) for clustering orthologs across 12 fungal genomes (five Ascomycota, including two yeast species, three Basidiomycota, including one yeast species, and four Mucoromycotina). In this dataset, we only included taxa in which sexual reproduction is known to occur. We recovered 1331 ortholog groups. Of these, we found 250 core genes present amongst all fungi, 318 ancestral genes and 546 Mucoromycotina-specific genes. We expect that the high amount of Mucoromycotina specific genes recovered is due to the large number of sexually relevant differentially expressed *R. microsporus* genes we used as bait. A summary of the results for all well described sexually relevant genes used in this study is presented in Table 3.2. Additionally, to further explore these results, we grouped mating relevant genes into four main categories and investigated gene conservation within them. These categories include: pheromone response pathways (Table A3.2), environmental response regulators (Table A3.3), transcription factors (Table A3.4), and late development (Table A3.5). These categories are not mutually exclusive. Many important genes involved in regulation of sexual development have already been characterized and summarized very recently in *A. nidulans* (Dyer & O'Gorman, 2012). For this reason, we focused on this

organism as the main reference for investigation of gene conservation across fungi. For consistency, we used the *A. nidulans* gene nomenclature unless otherwise noted.

Table 3.2. Conservation of sexually relevant genes across Fungi. Unless otherwise noted, *A. nidulans* gene nomenclature is used. Core: present across all fungi, Ancestral: present in Mucoromycotina and in at least one lineage of the Dikarya, Ancestral (colored green): present in all fungi except one or both ascomycete yeasts, Dikarya specific: present in both Ascomycota and Basidiomycota, but not in Mucoromycotina. *Differentially expressed during sexual reproduction in *R. microsporus*, [a] *S. cerevisiae* nomenclature, [b] *P. blakesleeanus* nomenclature, [c] *U. maydis* nomenclature.

Protein	Conservation	Role in Reproduction	Additional Information
CpcA	Ancestral	Amino acid sensing	Basidiomycota lost
CpcB	Core	Amino acid sensing	
Uac1[c]	Core	Adenylate cyclase	*
Uka1-1/1-2[c]	Ancestral	cAMP dependent protein kinase catalytic subunit	Ascomycota lost
Ubc1[c]	Core	cAMP dependent protein kinase regulatory subunit	
MutA	Dikarya specific	Carbohydrate catabolism, mutanase activity	
Cdc24p[a]	Ancestral	Cell polarity	Basidiomycota lost, only present in ascomycete: *S. cerevisiae*
Cdc42p[a]	Ancestral	Cell polarity	*
SamB	Core	Cell polarity	
Cdc28p[a]	Ancestral	Cyclin-dependent kinase catalytic subunit	*
Ste6p[a]	Ancestral	Export of pheromones	
Gpa3[c]	Core	G protein α subunit: signaling to pheromone MAP Kinase and cAMP cascades	*
FadA	Ancestral	G protein α subunit: Signaling to pheromone MAP Kinase cascade	
SfaD	Core	G protein β subunit: Signaling to pheromone MAP Kinase cascade	
TriR1	Mucoromycotina specific	GPCR: Trisporic acid receptor, putative	*
TriR2	Mucoromycotina specific	GPCR: Trisporic acid receptor, putative	*
HxtA	Ancestral	Hexose transport (upregulated during sexual reproduction)	Ascomycota yeast lost
MAT2	Core	HMG box transcription factor: regulator of sexual reproduction	*
SexP[b]	Ancestral	HMG box transcription factor: regulator of sexual reproduction	*, Only present in basidiomycete: *C. cinereus*
bE11[c]	U. maydis specific	Homeobox domain transcription factor: regulator of sexual reproduction	

Table 3.2 continued

bW1[c]	*U. maydis* specific	Homeobox domain transcription factor: regulator of sexual reproduction	
SidC	Dikarya specific	Iron regulation (NRPS)	
ImeB	Core	Light response, sexual development, secondary metabolism	
LreB	Ancestral	Light sensing	
CryA	Dikarya specific	Light sensing	
LreA	Ancestral	Light sensing	
FphA	Dikarya specific	Light sensing	Ascomycota yeast lost
VeA/VosA	Ancestral	Light/dark response regulator, secondary metabolite production	*
VelB	Ancestral	Light/dark response regulator, secondary metabolite production	
LaeA	Ancestral	Light/dark response regulator, secondary metabolite production	Basidiomycota lost, Ascomycota yeast lost
VelC	Ancestral	Light/dark response regulator, secondary metabolite production	Basidiomycota lost, only present in ascomycete: *A. nidulans*
MpkB	Core	Map kinase (pheromone cascade)	
Ste7	Core	MAP kinase, kinase (Pheromone cascade)	
SteC	Core	MAP kinase, kinase, kinase (pheromone cascade)	
Ste20	Core	Map kinase, kinase, kinase, kinase (pheromone cascade)	
GrrA	Ancestral	Meiotic structure development	
TubB	Core	Microtubule assembly	
FhbA	Dikarya specific	Nitrogen metabolism	Only present in basidiomycete: *C. neoformans*
Msn5p[a]	Ascomycete specific	Nuclear import and export of sexually-relevant transcription factors	
LsdA	*A. nidulans* specific	Osmotic stress sensing	
SteD	Ancestral	pheromone MAP Kinase cascade regulator	
Ste5p[a]	*S. cerevisiae* only	Pheromone MAP Kinase cascade scaffold	

Table 3.2 continued

Axl1p[a]	*S. cerevisiae* specific	Pheromone production	
Ram1p[a]	Core	Pheromone production	
Ste13p[a]	Core	Pheromone production	
Ste14p[a]	Core	Pheromone production	
Ste23p[a]	Core	Pheromone production	
Ste24p[a]	Core	Pheromone production	
PpgA	*A. nidulans* specific	Pheromone production	
PreA	Dikarya specific	Pheromone receptor	
PreB	Ascomycota specific	Pheromone receptor	
GprD	Dikarya specific	Pheromone receptor	
GprK	Ascomycota specific	Pheromone receptor	Ascomycota yeast lost
Far1p[a]	*S. cerevisiae* specific	Pheromone regulation	
Ste20p[a]	Core	Pheromone response	
Pho80	Ancestral	Phosphorous sensing cyclin	
PhoA	Ancestral	Phosphorous sensing, interacts with Pho80	
RasA	Ancestral	Ras signaling protein involved in asexual development	Ascomycota yeast lost
Ras2[c]	Ancestral	Ras signaling protein involved in reproductive development, cAMP regulation, and filamentous growth	*
TrxA	Ascomycota specific	Redox regulation	Ascomycota yeast lost
NoxA	Dikarya specific	Redox regulation	Ascomycota yeast lost
CandA-C	Ancestral	Regulate reproductive mode in response to light	
CandA-N	*A. nidulans* specific	Regulate reproductive mode in response to light	
Dig1p[a]	*S. cerevisiae* only	Regulation of mating-specific genes	
Dig2p[a]	*S. cerevisiae* only	Regulation of mating-specific genes	
Msg5p[a]	Core	Regulation of pheromone MAP Kinase cascade	
CsnD	Ancestral	Regulation of sexual vs asexual development	

Table 3.2 continued

CsnA	Ancestral	Regulation of sexual vs asexual development	
CsnE	Core	Regulation of sexual vs asexual development	
AcoB	Ancestral	Regulation of sexual vs asexual development	
CsnB	Ancestral	Regulation of sexual vs asexual development	
PhnA	Ancestral	Regulator of GpgA, SfaD and FlbA	
FlbA	Core	Regulator of G-protein signaling	
Sql2[c]	Ancestral	Regulator of Ras2	
StrA	Core	Sexual and asexual development: striatin scaffolding and Ca signaling	*
EsdC	Ascomycota specific	sexual development	Ascomycota yeast lost
RcoA	Core	sexual development and secondary metabolism	
PpoA/PpoC	Ancestral	Sexual vs asexual development, secondary metabolism (Oxylipin)	Ascomycota yeast lost
PpoB	*A. nidulans* specific	Sexual vs asexual development, secondary metabolism (Oxylipin)	
GpgA	Ancestral	Signaling to pheromone MAP Kinase cascade	
Sir3p[a]	*S. cerevisiae* only	Silencing MAT loci	
Sir4p[a]	*S. cerevisiae* only	Silencing MAT loci	
HogA	Core	Stress response	
AtfA	Ancestral	Stress response	
MsnA	Ancestral	Transcription factor	*
StuA	Ancestral	Transcription factor	* Basidiomycota lost
DopA	Core	Transcription factor	*
NsdD	Ancestral	Transcription factor	
MedA	Ancestral	Transcription factor	
FhpA	Ancestral	Transcription factor	Basidiomycota lost, only present in Mucoromycotina: *P. blakesleeanus*
DevR	Ancestral	Transcription factor	
NosA	Dikarya specific	Transcription factor	Ascomycota yeast lost

Table 3.2 continued

FlbE	Ascomycete specific	Transcription factor	Ascomycota yeast lost
FlbC	Ancestral	Transcription factor	
NsdC	Ancestral	Transcription factor	Basidiomycota lost, Ascomycota yeast lost
RosA	*A. nidulans* specific	Transcription factor	
SteA	Ancestral	Transcription factor activated by pheromone MAP Kinase cascade	
SilG	Ancestral	Transcription factor: light response	
SilA	Dikarya specific	Transcription factor: light response	Ascomycota yeast lost
MAT1	*A. nidulans* specific	Transcription factor: regulator of sexual reproduction	

Pheromone response pathway

The pheromone response pathway has been studied extensively in many fungi and is known to be highly conserved across the Dikarya in both gene membership and function. In addition to playing an essential role in sexual reproduction, several members of this pathway are also known to be important for filamentous growth (Müller et al., 2003). Therefore, we predicted that many of these genes are found in the 'core' and 'ancestral' gene groups. We found that of the 22 pheromone response genes, eight were present across all fungi. These included all of the main constituents of the MAP-kinase signaling cascade, *ste20*, *steC*, *ste7*, and *mpkB*, as well as *flbA*, *S. cerevisiae STE20* (corresponds to *Aspergillus nidulans* gene AN2067), and *S. cerevisiae MSG5*.

SfaD (β-subunit) was the only G protein involved in signaling to the pheromone pathway that was present across all fungi, although FadA (α-subunit) was found in all fungi except *S. pombe*. Similarly, GpgA (γ-subunit) was found in all fungi except *S. pombe* and *R. microsporus*. It is known that *S. pombe* utilizes a Ras protein (Ras1) and only one of these subunits (α-subunit) for interaction with the pheromone response pathway (Xu et al., 1994). Similarly, Ras signaling proteins are known to often play key upstream roles in the pheromone pathway of basidiomycetes and mediate cross-talk with cAMP pathways (Klosterman et al., 2007; Lengeler et al., 2000). Basidomycete Ras orthologs were found across all fungi except *S. pombe* and *S. cerevisiae*.

Overall, we had expected a different picture of gene conservation in the *S. pombe* pheromone pathway than we recovered. As *S. pombe* only uses the Gα-subunit in the pheromone response pathway (Xu et al., 1994), we had anticipated that both the α-subunit and *ras1* would be orthologous to those found in other fungi, and that the β-subunit would be absent. Looking into this further, we found that the identified *S. pombe* gene ortholog of the β-subunit encoded for Git5, a protein, which is involved in cAMP cycling (Landry & Hoffman, 2001). These results suggest that, subsequent to divergence from *S.*

cerevisiae, *S. pombe* underwent a dramatic reworking and redistribution of the genes involved in signaling upstream of the MAP kinase cascade.

All of the genes involved in pheromone production or direct perception (*ppgA* – only present in *A. nidulans*, *gprA*, *gprB*, *gprD*, *gprK*) were not found in Mucoromycotina. While *gprA* and *gprB* were known previously to be absent from *R. delemar* (Xue et al., 2008), the absence of this entire pathway across all Mucoromycotina indicates that this is a derived system, which evolved after the divergence of the Dikarya from lower fungi. Furthermore, the Dikarya have evolved to use small peptides as pheromone molecules rather than the trisporic acids utilized by Mucoromycotina. Based on our findings, it is clear that Mucoromycotina regulate perception of trisporic acid using recognition systems that are distinct from those of higher fungi. Interestingly, although the way in which the actual pheromone is perceived differs between these fungal lineages, the signaling cascade that they activate is highly conserved.

Environmental Response Regulators

Overall, there is less conservation across fungi of environmental response regulators as compared to the pheromone pathway. We collected information on 23 environmental regulators (mostly described in *A. nidulans* by Dyer & O'Gorman, 2012), including genes involved in perception of: light, glucose, salt, phosphorus, amino acids, and oxidative stress, all of which have implications for sexual reproduction. Additionally, we included members of the Velvet complex in this category, as they are important transcription factors guiding response to environmental cues. We found that only three of these genes were found across all fungi: *imeB* (involved in cross-talk between mating and cAMP pathways in response to environmental stimuli), *cpcB* (amino-acid sensing), and *hogA* (stress response MAP kinase). In addition to sexual reproduction, all of these proteins likely perform important roles in other processes, which may explain their presence across all fungi.

With respect to light response regulators, *fphA*, *silA*, and *cryA* were only found in the Dikarya. With the single exception of *Phycomyces blakesleeanus*, *silG* was also present only in the Dikarya. *lreA* and *lreB* were found in all fungi except the ascomycete yeasts. Interestingly, ascomycete yeasts appeared to have no sexually relevant light sensing proteins at all except CryA, which was only present in *S. cerevisiae*. Additionally, consistent with previous findings, the velvet complex is completely absent from both *S. cerevisiae* and *S. pombe* (Bayram & Braus, 2012). Aside from yeasts, most velvet members were present across all other fungi. One notable exception was *velC*, which was found in all Mucoromycotina, but from higher fungi was only found in *A. nidulans*. LaeA, a regulator of secondary metabolism and sexual development, which interacts with the velvet complex, was only present in filamentous ascomycetes and *Phycomyces blakesleeanus*.

Nearly all fungi retain sexually relevant phosphorous sensing proteins (PhoA and Pho80), as well as genes involved in stress response (*atfA*, *hogA*). Similarly, genes involved in cAMP cycling are conserved across most fungi (*S. cerevisiae CDC42*, *rasA*, *U. maydis ras2*). In contrast, the high salt sensing protein LsdA (only present in *A. nidulans*) and the glycogen sensing protein esdC are only present in filamentous ascomycetes.

Transcription Factors

Transcription factors are involved in nearly all steps of fungal reproduction. Many of these are also important for regulating developmental mode (sexual versus asexual reproduction). We found that of the 28 known reproductively relevant transcription factors included in this study, only four were found across all fungi. Those included *dopA*, *rcoA*, *csnE*, and *MAT2* (HMG box; corresponds to Sex minus from Mucoromycotina). However, 9 additional transcription factors were present in all fungi except one or both ascomycete yeasts. These included *medA*, *devR*, *csnA*, *csnB*, *csnD*, *acoB*, *steA*, *veA*, and *velB*. Similarly, *msnA* was found in all fungi, including ascomycete yeasts, but was missing from *Cryptococcus neoformans*. These results suggest that many of the transcription factors involved in reproductive mode

are essential for filamentous and dimorphic fungi, but not yeasts. We speculate that this may have to do with secondary metabolite synthesis, as many of the transcription factors missing from ascomycete yeasts are also tied to secondary metabolism. In *Saccharomyces*, practically no genes are devoted to secondary metabolism (Förster et al., 2003), making any gene regulating synthesis of these metabolites dispensable and perhaps even detrimental.

Of the 28 transcription factors investigated, seven were completely missing from Mucoromycotina. Of these, only *nosA* and *flbE* were present in more than one taxon. Transcription factors only present in single lineages include: *rosA* in *A. nidulans*, *SIR3* and *SIR4* in *S. cerevisiae* as well as *bW* and *bE* in *U. maydis*. Compared to other types of genes involved in sexual reproduction, the number of 'singleton' transcription factors is fairly high. We believe this could be either because: (i) many transcription factors, such as the homeodomain proteins bW and bE from *U. maydis*, are diverging rapidly, leading to difficulties in identifying ortholog groups, or (ii) lineage specific transcription factors arise fairly frequently. We found that eight of the 28 transcription factors were completely missing from basidiomycetes. These included *stuA*, *nsdC*, *rosA*, *flbE*, *fhpA*, *velC*, *SIR3* (*S. cerevisiae* nomenclature) and *SIR4* (*S. cerevisiae* nomenclature). At least three of these proteins (StuA, NsdC, and FlbE) are involved in cross-talk between sexual and asexual reproductive pathways.

Late Development

We defined genes involved in 'late development' as any genes that play an important role in maturation of sexual structures. All of the late sexual development genes used in this study were collected from *A. nidulans* (Dyer & O'Gorman, 2012). Here, many of these genes play roles in both sexual reproduction and conidiation. Of the 25 late development genes, five were found across all fungi, including *csnE*, *crk1*, *tubB*, *cpcB*, and *samB*. *candA-N*, involved in protein neddylation, was the only gene found only in a single genome (*A. nidulans*). Similar to other categories, eight additional genes were found in all fungi except ascomycete yeasts. These genes included *csnA*, *csnB*, *csnD*, *acoB*, *medA*, *candA-*

C, *vosA*, and *strA*. Many of these are genes, which impact transcription, were also discussed in the Transcription Factors section.

Six of these genes, *nosA*, *mutA*, *gprK*, *noxA*, *sidC*, and *candA-N*, were completely absent from Mucoromycotina. Two of these proteins, NoxA and SidC, are involved in reactive oxygen species generation (NoxA) and iron sequestration (SidC), suggesting differences in oxidative stress levels during reproduction between higher and lower fungi. Aside from *gprK* and *candA-N*, these proteins were found across both ascomycetes and basidiomycetes.

Investigating Mucoromycotina Specific Genes Reveals Two Candidate GPCRs Potentially Involved in Trisporic Acid Recognition

As mentioned previously, using the differentially expressed transcripts identified from our RNA-seq experiments, we found that of the 1331 ortholog groups, 546 were only found in Mucoromycotina (Table A3.6). Additionally, we were able to show that Mucoromycotina lack all of the pheromone production and perception genes present in higher fungi, suggesting that these organisms rely on novel mechanisms for triggering the pheromone cascade essential to initiate sexual reproduction. While we are already aware of several genes involved in production of trisporic acid (Tagua et al., 2012) (these genes are also up-regulated during sexual reproduction of *R. microsporus*), little is known of the genes involved in pheromone perception. By investigating the sexually relevant *R. microsporus* genes found only in Mucoromycotina,we were able to identify some of these pheromone sensing genes. Of the 546 Mucoromycotina-specific genes, 124 were found in all four lineages surveyed (*R. microsporus*, *R. delemar*, *M. circinelloides*, and *P. blakesleeanus*), 86 were found in three out of four lineages, 81 were present only in two lineages, and 252 were only found in *R. microsporus*. As expected, many of these genes had no homologs in higher fungi, leading to more 'hypothetical proteins' and genes bearing closer resemblance to those found in animal systems.

For uncovering the genes relevant to pheromone perception, we assumed that, as trisporic acids are used by all Mucoromycotina as mating pheromones, any gene relevant to this process would be present in most or all members of this lineage. Consequently, we interrogated any gene found in three or more representatives of Mucoromycotina. We believe that this approach is reasonable considering that in ascomycetes, the mating pheromone receptors, PreA and PreB, were present in nearly all taxa (4 out of 5). Specifically, we anticipated that, similar to the systems present in ascomycetes and basidiomycetes, G-protein coupled receptors (GPCRs) will be largely responsible for perception of mating pheromones.

We found that, of the 210 transcripts which were up-regulated during sexual reproduction of *R. microsporus* and present in at least 3 of the 4 Mucoromycotina surveyed, 14 unique predicted proteins were GPCRs or related proteins. While the original number of signaling-related proteins was higher (approximately 30), many of these were closely related isoforms of the same predicted gene. Several of these signaling-related proteins were potentially involved in cytoskeletal regulation and cAMP cycling, including a Ras-GEF related to Cdc25p, a Diaphanous RhoA GTPase effector (Sahai et al., 1998) and a frizzled-related GPCR (Winter et al., 2001). We also identified one Ras-GEF potentially involved in mitochondrial signaling and iron regulation.

Of particular interest, we found two Mucoromycotina-specific class C 7-transmembrane domain GPCRs. Class C GPCRs are known for their affinity for interacting with metabotropic glutamate, animal pheromones, and retinoic acid (Kniazeff et al., 2011). Retinoic acid, similar to trisporic acid, is a breakdown product of β-carotene, which is essential for growth and development of animals. To identify which type of class C GPCR these receptors were most closely related to, we selected full length protein sequence representatives of all major groups of class C GPCRs and reconstructed their phylogeny using Mr.Bayes v3.1.2 (Huelsenbeck & Ronquist, 2003). We found that one of these transcripts, m.92126, clustered along with an orphan GPCR as a sister lineage to the GPCRs responsible for sensing retinoic

acid in animal systems (Figure 3.3). Closely related, the other transcript, m.68840, clustered as sister to human γ-aminobutyric acid (GABA) receptors. We have tentatively named protein m.92126 Trisporic acid Receptor 1 (TriR1) and protein m.68840 Trisporic acid Receptor 2 (TriR2).

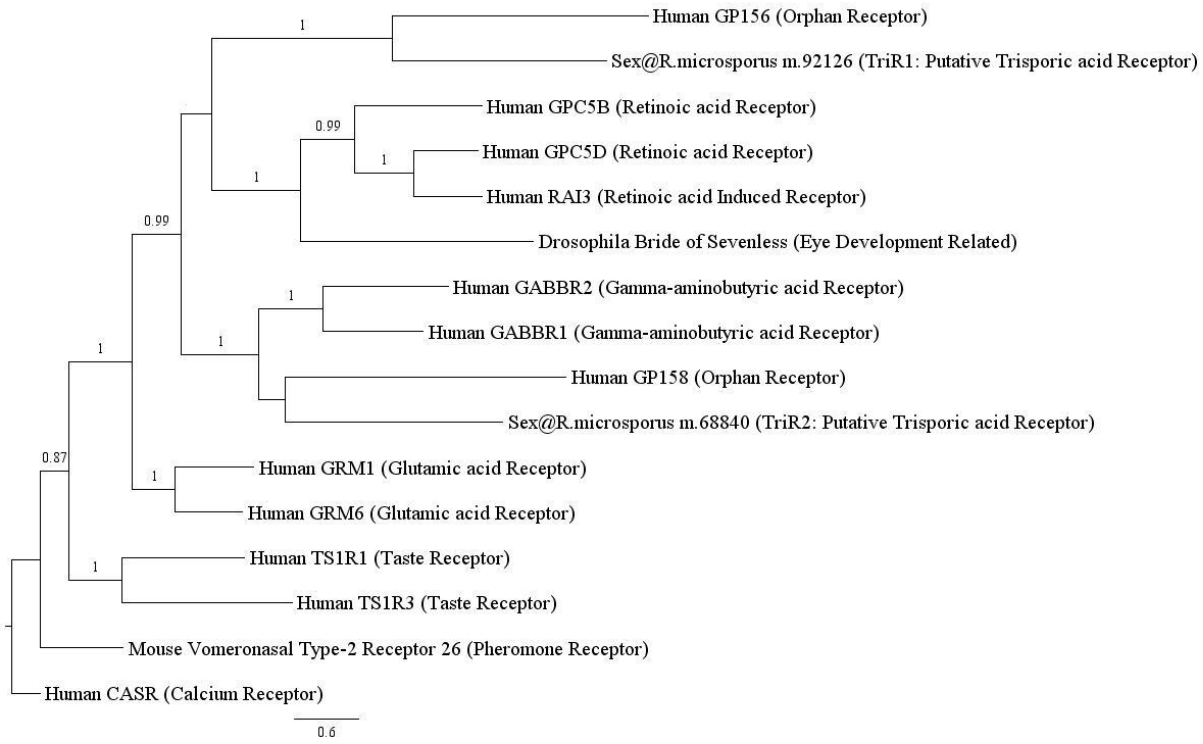

Figure 3.3. Bayesian reconstruction of representative class C GPCR proteins, including candidate trisporic acid receptors TriR1 (m.92126) and TriR2 (m.68840). Trees were constructed using Mr.Bayes v3.1.2 (Huelsenbeck & Ronquist, 2001) using the MtRev substitution model. The MCMC chain was allowed to run for 1.2 million generations and trees were sampled every 200th generation. Posterior probabilities were calculated after a burn-in of 300000 generations.

Our results provide three lines of evidence supporting the role of GPCRs in sensing trisporic acid. First, the GPCRs are present across all Mucoromycotina and not found in higher fungi, suggesting that they play a role in lower fungi, which is not conserved in the Dikarya. Second, they are up-regulated during sexual reproduction of *R. microsporus*, suggesting that the role they play is most important during sexual reproduction. Finally, they are most closely related to proteins known to bind β-carotene derivatives in other systems. Given these results, we believe these GPCRs are indeed sensing trisporic acid and we anticipate that they play important roles in stimulating the pheromone MAP Kinase cascade in these fungi.

We were also able to identify a GTPase activator protein that may also be involved in pheromone signaling in Mucoromycotina. While OrthoMCL failed to identify any actual orthologs in higher fungi, BLAST revealed similarity of a *R. microsporus* sexually relevant GTPase regulator to Syg1p, which binds to the Gβ subunit in *S. cerevisiae* and inhibits transduction of mating pheromone signals (Spain et al., 1995). Although the *R. microsporus* gene is likely paralogous to *SYG1*, it is possible that it also plays a role in pheromone signaling, although what role it may play is still uncertain. Finally, we found three signaling proteins, which we were unable to predict their function. These include a Rab3-like GTPase activator, which in animals is involved in vesicle trafficking in response to hormones and neurotransmitters, a RhoGAP domain containing protein, which may interact with Ras proteins, and a glutaredoxin family SH3 domain containing protein.

A Core Pathway Essential to Fungal Sexual Reproduction is Moderated by Endosymbionts

Through interrogating sexually relevant gene conservation amongst fungi we were able to predict which genes and pathways are present in the Mucoromycotina. Additionally, using what is already known about these genes in other organisms, we could predict their function in *R. microsporus*. With this information in hand, we re-investigated the genes that were differentially expressed due to endosymbiont presence. Our goal was to identify which of these genes is most likely responsible for the absence of reproduction when fungi lack their endosymbionts. We found that, of the 153 differentially expressed genes due to endosymbiont presence, seven were also present in the sexually relevant dataset. Interestingly, all of these transcripts were down-regulated in the absence of bacteria. While many of these transcripts were only found in the Mucoromycotina (5 out of 7), several of which had no putative domains or predicted functions (3 out of 7), we were capable of identifying dramatic down-regulation (17 fold) in the absence of endosymbionts of one gene that is required to activate the pheromone MAP Kinase signaling cascade in many fungi. This *R. microsporus* transcript is the ortholog

of the *U. maydis* G-protein Ras2. In *Ustilago*, Ras2 controls the initiation of the pheromone MAP Kinase cascade as well as morphogenesis and pathogenicity (Lee & Kronstad, 2002). This protein has been implicated in crosstalk between the cAMP signaling pathway and the pheromone MAP Kinase cascade, which may be responsible for the pleiotropic effects observed in *Ustilago* (Lee & Kronstad, 2002). However, upon further inspection of the OrthoMCL cluster containing Ras2, we found that it also contained the closely related *U. maydis* paralog Ras1. In order to discern which of the *U. maydis* Ras paralogs was most closely related to the *R. microsporus* Ras-like protein identified, we reconstructed the phylogeny of all genes present in this ortholog group using Mr.Bayes v3.1.2 (Huelsenbeck & Ronquist, 2003) and full length amino acid sequences (Figure 3.4). We found that indeed the *R. microsporus* Ras-like protein most closely resembles Ras2. Although we are unable to discern whether Mucoromycotina Ras2 is more closely related to ascomycetes or basidiomycetes, there is strong support that each of these fungal phyla harbor their own distinct but closely related copies of Ras2. In *Neurospora*, disruption of the identified Ras2 ortholog results in abnormal colony growth and a conidiation rate 0.005 that of wild-type (Kana-uchi et al., 1997). In *Rhizopus*, in the absence of endobacteria, we observe loss of both sexual and asexual reproduction. Given the impact of this protein on both reproductive pathways in other fungi as well as on cAMP cycling, we hypothesize that endobacteria are controlling both reproductive modes by up-regulating expression of Ras2 in their hosts.

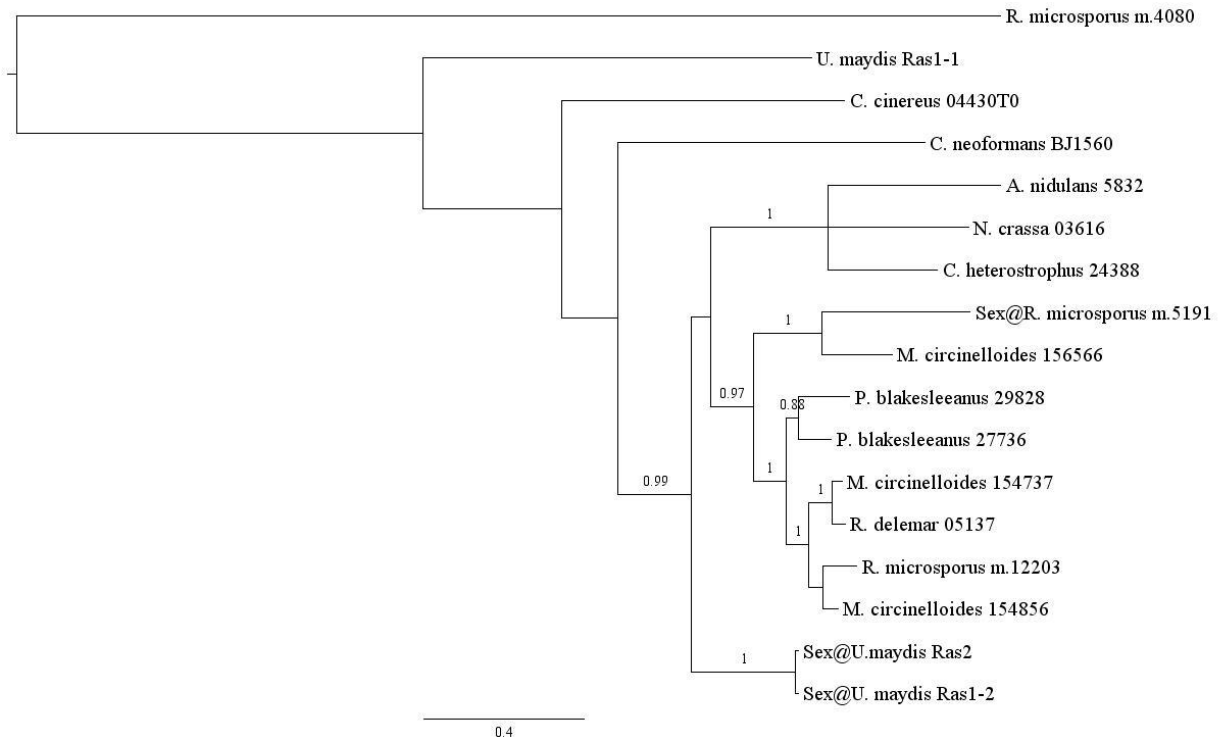

Figure 3.4. Bayesian reconstruction of all proteins present in the identified Ras2 ortholog group. Trees were constructed using Mr.Bayes v3.1.2 (Huelsenbeck & Ronquist, 2001) using the MtRev substitution model. The MCMC chain was allowed to run for 1.2 million generations and trees were sampled every 200th generation. Posterior probabilities were calculated after a burn-in of 300000 generations.

Bacterial Type III Secretion System is not Expressed During an Established Symbiosis

Based on our observation that bacteria are altering expression of host genes, we hypothesized that endosymbionts may be using their type III secretion system (T3SS) to release effectors relevant to fungal reproduction. T3SS is highly expressed prior to colonization and essential for *B. rhizoxinica* to successfully infect their fungal host (Lackner et al., 2011). However, it is unknown whether the T3SS plays any role during established interactions. To address this, we investigated expression of any genes involved in production of the T3SS machinery in our RNA-seq dataset. Interestingly, we found that this system is not assembled during established interactions. We could only recover transcripts from endosymbionts for four of the 23 genes involved in T3SS assembly (Lackner et al., 2011). Expressed

genes encoded the ATPase and its regulator (*sctN* and *sctL*) as well as two genes encoding proteins of unknown function, SctI and SctG. All of these genes belong to the same predicted operon. This operon also contained *sctJ*, the gene predicted to be responsible for production of the inner membrane ring, although we found no evidence of this gene being expressed, perhaps suggesting that this transcript is actively being degraded.

We speculate that, because the *Burkholderia* endosymbionts are not encapsulated within any host membrane (Partida-Martinez et al., 2007), post-infection pilus assembly is not necessary. However, because not even a complete T3SS pore is formed, we conclude that this system is likely not active once the symbiosis has been established. However, this does not exclude the possibility that effectors may still play a role in the symbiosis between *B. rhizoxinica* and *R. microsporus*, and that they may be released through an alternative system or perhaps through bacterial lysis.

Discussion

Our analyses revealed that *Rhizopus microsporus* is obligately dependent upon its endosymbiont, *Burkholderia rhizoxinica*, for survival. Although the fungus can grow vegetatively in the absence of endosymbionts, we have shown that, in addition to failing to reproduce asexually (Partida-Martinez et al., 2007c), it is also unable to reproduce sexually. Without any form of reproduction left, ultimate fitness of the fungal host without endosymbionts must be at or near 0. Interestingly, the *Burkholderia* endosymbionts appear essential to host survival not because they provide a product required by the host, but because they have hijacked an indispensable component of the hosts reproductive machinery. In particular, we have shown that these endosymbionts strongly impact expression of Ras2, a G-protein responsible for controlling reproductive mode in several higher fungi as well as regulation of cAMP levels (Lee & Kronstad, 2002, Kana-uchi et al., 1997).

The origin of endosymbiont control of the host reproduction is a puzzle. Certainly, prior to the establishment of symbiotic interactions, the fungi could reproduce independently of bacteria. Therefore,

somehow along the evolutionary timeline of the symbiosis bacteria must have taken control of the host reproductive processes. One can propose several hypotheses as to how such control might have evolved. For example, the establishment of the symbiosis could have been followed by host gene loss and compensation by endobacteria. This could perhaps be due to horizontal gene transfer of a reproductively relevant gene from host to endosymbiont. However, we were unable to detect any eukaryotic-like genes in the endosymbiont's genome that have any relevance to fungal reproduction. Furthermore, as endosymbionts can restore asexual reproduction in any host, bacterial control of fungal reproduction must have evolved prior to diversification in this group of fungi. Alternatively, it is possible that bacterial control of fungal reproduction evolved prior to symbiosis establishment as a host adaptation to protect against *B. rhizoxinica*. A pathogen to mutualist shift has already been proposed for these endosymbionts on the basis of their production of a general eukaryotic antimitotic and the evolution of resistant β-tubulin variants in the host (Schmitt et al., 2008). This antimitotic, rhizoxin, would very likely have lead to cell death in the host, suggesting that these endosymbionts were originally necrotrophic pathogens. As Ras2 impacts cAMP cycling, and therefore cell cycle regulation, perhaps bacterial targeting of this protein could also cause disease. Since overexpression of Ras proteins has strong effects on growth and can even lead to cell death (Gourlay & Ayscough, 2006, Colombo et al., 2004, Clark et al., 1996, Hilson et al., 1990), it is possible that fungi evolved to cope with endosymbiont-mediated overexpression by reducing their own control over Ras2. A side-effect of such defensive response would be an increased dependence of the host upon endosymbionts.

Targeting host signaling cascades is a common strategy in bacterial pathogens of eukaryotes. By the use of effectors, obligately biotrophic bacterial pathogens often target signaling pathways to evade or deactivate host defenses (Boller & He, 2009). However, as necrotrophs thrive by killing their hosts, effectors may also serve the purpose of targeting pathways involved in cell cycle regulation and cell death. We found that not only was Ras2 impacted by endosymbiont presence, but many other proteins

involved in regulation of cell division, cAMP cycling, and filamentous growth were similarly affected by endosymbiont presence. Interestingly, a common theme amongst many of these differentially expressed proteins (Cin1, Sho1, Mad2, and perhaps also a putative mucin) is that they either interact with or are regulated by Cdc42. In *S. cerevisiae*, Cdc42p is a small Rho GTPase, which activates Ste20p, a serine/threonine kinase involved in osmosensing, filamentous growth, and cell division (Raitt et al., 2000, Gancedo, 2001, Pruyne & Bretscher, 2000). Both *ste20* and *cdc42* are conserved across nearly all fungi. As Mucoromycotina utilize moisture as a trigger for reproduction (Blakeslee, 1904), it is possible that endosymbiont-mediated alterations of the Cdc42/Ste20 pathway are playing a role in the observed inability to reproduce when fungi lack endosymbionts. We speculate that these proteins are acting in concert with Ras2 to deactivate reproduction in *R. microsporus*. In *S. pombe* (Miller & Johnson, 1994) as well as perhaps in *U. maydis* (Müller et al., 2003), Ras2 acts upstream of both Cdc42 and the pheromone MAP Kinase cascade. Based on the expression patterns observed in *R. microsporus*, we predict that similar regulation is also occurring in *R. microsporus*. Consequently, we expect alterations in both filamentous growth as well as reproduction.

Although there appears to be no difference in growth rate for most fungi when they lack endosymbionts, it would be interesting to examine other features of filamentous growth in these organisms when endosymbionts are absent. Perhaps related to filamentous growth, we observed up-regulation of at least five genes involved in cell wall synthesis as well as genes involved in energy production (four genes - mostly related to mitochondrial functioning). Of note, one strain of *R. microsporus* var *chinensis* (ATCC 62417) exhibits dramatically different growth patterns when endosymbionts are absent. Not only is growth rate reduced, but colonies produce fewer aerial hyphae and display altered colony morphology. It would be interesting to investigate whether this can be linked to down-regulation of Ras2.

Although we can demonstrate that the bacteria are altering Ras2 expression in their hosts, we do not know the actual regulatory mechanism. Two hypotheses may explain the observed phenotype. First, microbe-associated-molecular-patterns (MAMPs) such as peptidoglycan or some cell surface component may be responsible for altering cAMP levels in the host. Many fungal adenylate cyclase genes retain leucine rich repeat (LRR) domains (Soanes & Talbot, 2010). In general, these LRR domains are involved in interactions with Ras proteins (Suzuki et al., 1990); however, in *Candida albicans*, they have been shown to interact directly with bacterial peptidoglycan (Xu et al., 2008). Furthermore, it was also shown that this interaction could promote hyphal growth in *C. albicans* (Xu et al., 2008). Although this is a tempting speculation, currently it does not explain the differential expression we observe in upstream proteins involved in cAMP cycling and cell cycle regulation (i.e. Cin1 and Ras2) when fungi lack bacteria. Alternatively, bacterial effectors may be responsible for activating fungal reproduction. Interestingly, we found that the T3SS of the endosymbionts, the most likely system to secrete effectors, was not assembled in established associations. However, as these bacteria are intracellular and appear to not be surrounded by a host membrane based on freeze-fracture imaging (Partida-Martinez et al., 2007), any molecule secreted through any system (or through bacterial lysis) will likely end up in the fungal cytoplasm. Consequently, many possibilities exist for how bacteria may transport effectors into their hosts, although whether they play any role in fungal reproduction remains to be elucidated.

Through our phenotyping work, we were able to show that the *Burkholderia* endosymbionts completely control reproductive mode in their hosts. We exploited this observation, to reconstruct key reproductive pathways across the fungal kingdom. Specifically, we examined how conserved sexually relevant genes from different lineages of ascomycetes and basidiomycetes were across all fungi, including the Mucoromycotina. As many of these genes were experimentally studied in higher fungi, we were able to augment our results with information on how conserved they were in their function. If proteins were present across fungi and also conserved in function, our expectation was that they play a

similar role in Mucoromycotina. From these comparisons, we inferred that many transcription factors as well as the pheromone MAP Kinase signaling cascade are conserved across all fungi. Conservation of the pheromone signaling cascade itself is not very surprising, as orthologs of these genes are also present in animal systems (Waskiewicz & Cooper, 1995). However, this means that we can be fairly certain that this pathway is playing the same role in Mucoromycotina. Similarly, many of the transcription factors involved in cross-talk between secondary metabolism and reproductive pathways were conserved across all fungi except the ascomycete yeasts, where secondary metabolites are rarely produced. This observation implies that sexual reproduction and secondary metabolism have been tied together since before most of Fungi diverged. To investigate how far back this coupling goes, data from other lower fungal phyla such as the Chytridiomycota would be necessary. Unfortunately, more genomic data from chytrids would be necessary to fully address this question, as currently only one chytrid (*Batrachochytrium dendrobatidis*) genome has been sequenced (*Batrachochytrium dendrobatidis* Sequencing Project, Broad Institute of Harvard and MIT (http://www.broadinstitute.org/)).

Using endosymbiont control over host reproduction we were also able to investigate genes likely involved in trisporic acid perception. Typically in higher fungi, class D GPCRs are responsible for pheromone sensing (Xue et al., 2008). Interestingly, in *R. microsporus*, we found no class D GPCRs which were up-regulated during sexual reproduction, and no genes bearing any resemblance to ascomycete or basidiomycete class D pheromone sensing GPCRs in any member of the Mucorales. We were, however, able to identify two class C GPCRs, which we believe are involved in perception of trisporic acid. These are conserved across all Mucoromycotina, up-regulated during sexual reproduction of *R. microsporus*, absent in higher fungi, and most closely resemble retinoic acid sensing GPCRs in animal systems. As mentioned earlier, retinoic acid and trisporic acid are both breakdown products of β-carotene. In both animal systems and the Mucoromycotina, β-carotene is broken down into retinal,which is then used to produce either retinoic acid (animal systems), or trisporic acid (Mucoromycotina) (Duester, 2008,

Schachtschabel et al., 2005). In animals, perception of retinoic acid is important for organogenesis as well as eye development (Duester, 2008). Similarly, in Mucoromycotina perception of trisporic acid also results in organogenesis, as it initiates sexual reproduction and zygospore formation.

As production of the final form of trisporic acid is a complex interplay between both opposite mates, we would not be surprised if more class C GPCRs are uncovered, which play a role in trisporic acid sensing. In our transcript data, we found that there were many isoforms of these GPCRs. However, because genomic data are not available currently, we are unable to reliably identify whether these represent distinct genes, splice variants, or errors due to inclusion of two opposite mates in transcriptome assembly. For this reason, we were only able to conservatively identify the two transcripts reported here. As *R. microsporus* is currently untransformable, it would be interesting to dirupt the orthologs of these genes in *M. circinelloides* and investigate how trisporic acid sensing is altered.

In conclusion, we have been able to demonstrate that *R. microsporus* is dependent upon its bacterial endosymbiont, *B. rhizoxinica* for sexual reproduction. Furthermore, as fungi depend on endosymbionts for both modes of reproduction, these bacteria represent essential endosymbionts of *R. microsporus*. However, endobacteria do not provision their hosts with nutrients required for reproduction, and appear to control reproduction through hijacking host reproductive machinery. In particular, endosymbionts alter expression levels of Ras2, a G-protein required for cAMP regulation and activation of reproduction. Endosymbiont control of reproduction has also allowed us to explore genes relevant to sexual reproduction across all fungi, including Mucoromycotina. In addition to identification of many conserved elements of fungal reproduction, we have also identified two candidate genes involved in trisporic acid perception, *triR1* and *triR2*. These candidates are class C seven-transmembrane GPCRs related to those used for sensing retinoic acid in animal systems and are upregulated during sexual reproduction of *R. microsporus*.

REFERENCES

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. J Mol Biol 215:403–410.

Bayram, Ö., and G. H. Braus. 2012. Coordination of secondary metabolism and development in fungi: the velvet family of regulatory proteins. FEMS Microbiol Rev 36:1–24.

Bianciotto, V., A. Genre, P. Jargeat, E. Lumini, G. Becard, and P. Bonfante. 2004. Vertical transmission of endobacteria in the arbuscular mycorrhizal fungus *Gigaspora margarita* through generation of vegetative spores. Appl Environ Microb 70:3600–3608.

Blakeslee, A. F. 1904. Sexual reproduction in the Mucorineae. P Am Acad Arts Sci 40:205–319.

Boller, T., and S. Y. He. 2009. Innate immunity in plants: an arms race between pattern recognition receptors in plants and effectors in microbial pathogens. Science 324:742–744.

Budovskaya, Y. V., J. S. Stephan, S. J. Deminoff, and P. K. Herman. 2005. An evolutionary proteomics approach identifies substrates of the cAMP-dependent protein kinase. Proc. Natl. Acad. Sci. USA 102:13933–13938.

Burchett, S. A., A. Scott, B. Errede, and H. G. Dohlman. 2001. Identification of novel pheromone-response regulators through systematic overexpression of 120 protein kinases in Yeast. J. Biol. Chem. 276:26472–26478.

Celestino, K. R., R. B. Cunha, and C. R. Felix. 2006. Characterization of a β-glucanase produced by *Rhizopus microsporus* var. *microsporus*, and its potential for application in the brewing industry. BMC Biochem 7:23.

Colombo, S., D. Ronchetti, J. M. Thevelein, J. Winderickx, and E. Martegani. 2004. Activation state of the Ras2 protein and glucose-induced signaling in *Saccharomyces cerevisiae*. J. Biol. Chem. 279:46715–46722.

Cullen, P. J., W. Sabbagh, E. Graham, M. M. Irick, E. K. van Olden, C. Neal, J. Delrow, L. Bardwell, and G. F. Sprague. 2004. A signaling mucin at the head of the Cdc42- and MAPK-Dependent filamentous growth pathway in yeast. Genes Dev. 18:1695–1708.

Drewes, G., A. Ebneth, U. Preuss, E.-M. Mandelkow, and E. Mandelkow. 1997. MARK, a novel family of protein kinases that phosphorylate microtubule-associated proteins and trigger microtubule disruption. Cell 89:297–308.

Duester, G. 2008. Retinoic acid synthesis and signaling during early organogenesis. Cell 134:921–931.

Dumaz, N., and R. Marais. 2005. Integrating signals between cAMP and the RAS/RAF/MEK/ERK signaling pathways. FEBS J 272:3491–3504.

Dyer, P. S., and C. M. O'Gorman. 2012. Sexual development and cryptic sexuality in fungi: insights from *Aspergillus* species. FEMS Microbiol Rev 36:165–192.

Edgar, R. C. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucl. Acids Res. 32:1792–1797.

Förster, J., I. Famili, P. Fu, B. Ø. Palsson, and J. Nielsen. 2003. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. Genome Res. 13:244–253.

Gancedo, J. M. 2001. Control of pseudohyphae formation in *Saccharomyces cerevisiae*. FEMS Microbiol Rev 25:107–123.

Ghignone, S., A. Salvioli, I. Anca, E. Lumini, G. Ortu, L. Petiti, S. Cruveiller, V. Bianciotto, P. Piffanelli, L. Lanfranco, and P. Bonfante. 2012. The genome of the obligate endobacterium of an AM fungus reveals an interphylum network of nutritional interactions. ISME J 6:136–145.

Clark, G. J., M. S. Kinch, T. M. Gilmer, K. Burridge, and C. J. Der. 1996. Overexpression of the Ras-related TC21/R-Ras2 protein may contribute to the development of human breast cancers. Oncogene 12:169–176.

Gordon, A., Hannon, G.J. 2010. Fastx-Toolkit. *FASTQ/A short-reads pre-processing tools (unpublished) http://hannonlab.cshl.edu/fastx_toolkit*.

Gourlay, C. W., and K. R. Ayscough. 2006. Actin-induced hyperactivation of the Ras signaling pathway leads to apoptosis in *Saccharomyces cerevisiae*. Mol. Cell. Biol. 26:6487–6501.

Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B. W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, and A. Regev. 2011. Full-length transcriptome assembly from RNA-seq data without a reference genome. Nat Biotech 29:644–652.

Gutiérrez, A., S. López-García, and V. Garre. 2011. High reliability transformation of the basal fungus *Mucor circinelloides* by electroporation. Journal Microbiol Meth 84:442–446.

Hilson, P., J. Dewulf, F. Delporte, P. Installé, J.-M. Jacquemin, M. Jacobs, and I. Negrutiu. 1990. Yeast RAS2 affects cell viability, mitotic division and transient gene expression in *Nicotiana* species. Plant Mol Biol 14:669–685.

Idnurm, A., F. J. Walton, A. Floyd, and J. Heitman. 2008. Identification of the sex genes in an early diverged fungus. Nature 451:193–196.

Jany, J., and T. E. Pawlowska. 2010. Multinucleate spores contribute to evolutionary longevity of asexual Glomeromycota. Am Nat175:424–435.

Jasti, N., S. K. Khanal, A. L. Pometto, and J. (Hans) van Leeuwen. 2008. Converting corn wet-milling effluent into high-value fungal biomass in a biofilm reactor. Biotechnol Bioeng 101:1223–1233.

Jasti, N., M. Rasmussen, S. Khanal, A. Pometto, and J. van Leeuwen. 2009. Influence of selected operating parameters on fungal biomass production in corn-ethanol wastewater. J Environ Engin135:1106–1114.

Jiggins, F. M., G. D. D. Hurst, and M. E. N. Majerus. 2000. Sex-ratio-distorting *Wolbachia* causes sex-role reversal in its butterfly host. Proc. R. Soc. Lond. B 267:69–73.

Kana-uchi, A., C. T. Yamashiro, S. Tanabe, and T. Murayama. 1997. A Ras homologue of *Neurospora crassa* regulates morphology. Mol Gen Genet 254:427–432.

Klosterman, S. J., M. H. Perlin, M. Garcia-Pedrajas, S. F. Covert, and S. E. Gold. 2007. Genetics of morphogenesis and pathogenic development of *Ustilago maydis*. Adv. Genet. 57:1–47.

Lackner, G., N. Moebius, L. Partida-Martinez, and C. Hertweck. 2011. Complete genome sequence of *Burkholderia rhizoxinica*, an endosymbiont of *Rhizopus microsporus*. J. Bacteriol. 193:783–784.

Landry, S., and C. S. Hoffman. 2001. The Git5 Gβ and Git11 Gγ form an atypical Gβγ dimer acting in the fission yeast glucose/cAMP pathway. Genetics 157:1159–1168.

Lee, N., and J. W. Kronstad. 2002. Ras2 controls morphogenesis, pheromone response, and pathogenicity in the fungal pathogen *Ustilago maydis*. Eukaryot Cell 1:954–966.

Lengeler, K. B., R. C. Davidson, C. D'souza, T. Harashima, W.-C. Shen, P. Wang, X. Pan, M. Waugh, and J. Heitman. 2000. Signal transduction cascades regulating fungal development and virulence. Microbiol. Mol. Biol. Rev. 64:746–785.

Li, B., and C. N. Dewey. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12:323.

Li, L., C. J. Stoeckert, and D. S. Roos. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 13:2178–2189.

Ma, L.-J., A. S. Ibrahim, C. Skory, M. G. Grabherr, G. Burger, M. Butler, M. Elias, A. Idnurm, B. F. Lang, T. Sone, A. Abe, S. E. Calvo, L. M. Corrochano, R. Engels, J. Fu, W. Hansberg, J.-M. Kim, C. D. Kodira, M. J. Koehrsen, B. Liu, D. Miranda-Saavedra, S. O'Leary, L. Ortiz-Castellanos, R. Poulter, J. Rodriguez-Romero, J. Ruiz-Herrera, Y.-Q. Shen, Q. Zeng, J. Galagan, B. W. Birren, C. A. Cuomo, and B. L. Wickes. 2009. Genomic analysis of the basal lineage fungus *Rhizopus oryzae* reveals a whole-genome duplication. PLoS Genet 5:e1000549.

Miller, P. J., and D. I. Johnson. 1994. Cdc42p GTPase is involved in controlling polarized cell growth in *Schizosaccharomyces pombe*. Mol. Cell. Biol. 14:1075–1083.

Mondo, S. J., K. H. Toomer, J. B. Morton, Y. Lekberg, and T. E. Pawlowska. 2012. Evolutionary stability in a 400-million-year-old heritable facultative mutualism. Evolution 66:2564–2576.

Moran, N. A., J. P. McCutcheon, and A. Nakabachi. 2008. Genomics and evolution of heritable bacterial symbionts. Ann Rev Genet 42:165–190.

Müller, P., J. D. Katzenberger, G. Loubradou, and R. Kahmann. 2003. Guanyl nucleotide exchange factor Sql2 and Ras2 regulate filamentous growth in *Ustilago maydis*. Eukaryot Cell 2:609–617.

Partida-Martinez, L. P., and C. Hertweck. 2005. Pathogenic fungus harbours endosymbiotic bacteria for toxin production. Nature 437:884–888.

Partida-Martinez, L. P., S. Monajembashi, K. O. Greulich, and C. Hertweck. 2007. Endosymbiont-dependent host reproduction maintains bacterial-fungal mutualism. Curr Biol 17:773–777.

Peixoto, S. C., J. A. Jorge, H. F. Terenzi, and M. de L. T. M. Polizeli. 2003. *Rhizopus microsporus* var. *rhizopodiformis*: a thermotolerant fungus with potential for [roduction of thermostable amylases. Int Microbiol 6:269–273.

Pruyne, D., and A. Bretscher. 2000. Polarization of cell growth in yeast. I. Establishment and maintenance of polarity states. J Cell Sci 113:365–375.

Raitt, D. C., F. Posas, and H. Saito. 2000. Yeast Cdc42 GTPase and Ste20 PAK-like kinase regulate Sho1-dependent activation of the Hog1 MAPK pathway. EMBO J 19:4623–4631.

Rispe, C., and N. A. Moran. 2000. Accumulation of deleterious mutations in endosymbionts: Muller's Ratchet with two levels of selection. Am Nat 156:425–441.

Ronquist, F., and J. P. Huelsenbeck. 2003. MrBayes 3: bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572 –1574.

Sahai, E., A. S. Alberts, and R. Treisman. 1998. RhoA effector mutants reveal distinct effector pathways for cytoskeletal reorganization, SRF activation and transformation. EMBO J 17:1350–1361.

Schachtschabel, D., C. Schimek, J. Wöstemeyer, and W. Boland. 2005. Biological activity of trisporoids and trisporoid analogues in *Mucor mucedo* (−). Phytochemistry 66:1358–1365.

Schmitt, I., L. P. Partida-Martinez, R. Winkler, K. Voigt, E. Einax, F. Dölz, S. Telle, J. Wöstemeyer, and C. Hertweck. 2008. Evolution of host resistance in a toxin-producing bacterial–fungal alliance. ISME J 2:632–641.

Sharon, G., D. Segal, J. M. Ringo, A. Hefetz, I. Zilber-Rosenberg, and E. Rosenberg. 2010. Commensal bacteria play a role in mating preference of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA 107:20051–20056.

Shen, G., A. Whittington, K. Song, and P. Wang. 2010. Pleiotropic function of intersectin homologue Cin1 in *Cryptococcus neoformans*. Mol Microbiol 76:662–676.

Soanes, D. M., and N. J. Talbot. 2010. Comparative genome analysis reveals an absence of Leucine-rich repeat pattern-recognition receptor proteins in the kingdom Fungi. PLoS ONE 5:e12725.

Spain, B. H., D. Koo, M. Ramakrishnan, B. Dzudzor, and J. Colicelli. 1995. Truncated forms of a novel yeast protein suppress the lethality of a G protein alpha subunit deficiency by interacting with the beta subunit. J. Biol. Chem. 270:25435–25444.

Stouthamer, R., J. A. J. Breeuwer, and G. D. D. Hurst. 1999. *Wolbachia pipientis*: microbial manipulator of arthropod reproduction. Ann Rev Microbiol 53:71–102.

Suzuki, N., H. R. Choe, Y. Nishida, Y. Yamawaki-Kataoka, S. Ohnishi, T. Tamaoki, and T. Kataoka. 1990. Leucine-rich repeats and carboxyl terminus are required for interaction of yeast adenylate cyclase with RAS proteins. Proc. Natl. Acad. Sci. USA 87:8711–8715.

Tagua, V. G., H. R. Medina, R. Martín-Domínguez, A. P. Eslava, L. M. Corrochano, E. Cerdá-Olmedo, and A. Idnurm. 2012. A gene for carotene cleavage required for pheromone biosynthesis and carotene regulation in the fungus *Phycomyces blakesleeanus*. Fungal Genet Biol 49:398–404.

Tatebayashi, K., K. Yamamoto, K. Tanaka, T. Tomida, T. Maruoka, E. Kasukawa, and H. Saito. 2006. Adaptor functions of Cdc42, Ste50, and Sho1 in the yeast osmoregulatory HOG MAPK pathway. EMBO J 25:3033–3044.

Verweij, P. E., A. Voss, J. P. Donnelly, B. E. de Pauw, and J. F. Meis. 1997. Wooden sticks as the source of a pseudoepidemic of infection with *Rhizopus microsporus* var. *rhizopodiformis* among immunocompromised patients. J. Clin. Microbiol. 35:2422–2423.

Wang, H. L., E. W. Swain, and C. W. H. Tine. 1975. Mass production of *Rhizopus oligosporus* spores and their application in tempeh fermentation. J Food Sci 40:168–170.

Waskiewicz, A. J., and J. A. Cooper. 1995. Mitogen and stress response pathways: MAP kinase cascades and phosphatase regulation in mammals and yeast. Curr Opin Cell Biol 7:798–805.

Winter, C. G., B. Wang, A. Ballew, A. Royou, R. Karess, J. D. Axelrod, and L. Luo. 2001. *Drosophila* Rho-associated kinase (Drok) links frizzled-mediated planar cell polarity signaling to the actin cytoskeleton. Cell 105:81–91.

Xu, H. P., M. White, S. Marcus, and M. Wigler. 1994. Concerted action of RAS and G proteins in the sexual response pathways of *Schizosaccharomyces pombe*. Mol. Cell. Biol. 14:50–58.

Xu, X.-L., R. T. H. Lee, H.-M. Fang, Y.-M. Wang, R. Li, H. Zou, Y. Zhu, and Y. Wang. 2008. Bacterial peptidoglycan triggers *Candida albicans* hyphal growth by directly activating the adenylyl cyclase Cyr1p. Cell Host Microbe 4:28–39.

Xue, C., Y.-P. Hsueh, and J. Heitman. 2008. Magnificent seven: roles of G protein-coupled receptors in extracellular sensing in Fungi. FEMS Microbiol Rev 32:1010–1032.

Yu, E.-Y., J.-H. Lee, W.-H. Kang, Y.-H. Park, L. Kim, and H.-M. Park. 2013. Fission yeast LAMMER kinase Lkh1 regulates the cell cycle by phosphorylating the CDK-inhibitor Rum1. Biochem Bioph Res Co 432:80–85.

Zheng, R.Y., Chen, G.Q., Huang, H., Liu, X.Y. 2007. A monograph of *Rhizopus*. Sydowia 59(2):273-372

APPENDIX

Table A3.1. Complete list of differentially expressed transcripts due to endosymbiont presence, their average fold change, and their putative annotations.

Transcript	Average Fold Change	Putative Protein Annotation
comp100584_c0_seq1	>100	NA
comp160904_c0_seq3	7	Ras2
comp179716_c0_seq1	47	NA
comp174746_c0_seq37	>100	NA
comp174039_c0_seq5	68	ATP dependent RNA helicase
comp161109_c0_seq1	2.37	mitotic spindle checkpoint component Mad2
comp159019_c1_seq4	5	Mucin, putative
comp174013_c0_seq1	>100	GPI anchored cell surface protein
comp169791_c0_seq2	>100	NA
comp174719_c2_seq19	>100	NA
comp173623_c0_seq18	>100	NA
comp171888_c0_seq1	36	NA
comp169900_c0_seq2	>100	NA
comp174390_c1_seq4	50	NA
comp174016_c0_seq1	>100	NA
comp174834_c0_seq86	>100	NA
comp169576_c1_seq1	>100	NA
comp169061_c0_seq4	3	WctD (White collar: *Phycomyces blakesleeanus*)
comp174066_c4_seq10	>100	60S Ribosomal protein L17
comp174296_c3_seq9	>100	NA
comp106586_c0_seq1	64	NA
comp180249_c0_seq1	>100	Hypothetical protein
comp170425_c0_seq1	50	NA
comp174374_c1_seq2	36	NA
comp174389_c1_seq2	>100	mitochondrial iron ion transporter (predicted)
comp171279_c1_seq1	51	NA
comp168747_c0_seq2	>100	Hypothetical protein
comp96826_c0_seq1	68	NA
comp173429_c0_seq1	>100	Prr2p (pheromone response regulator 2: *Saccharomyces cerevisiae*)

Table A3.1 continued

comp174390_c1_seq12	58	Hypothetical protein
comp174222_c0_seq71	18	Ubiquitin ligase E3A isoform 1
comp166563_c0_seq1	>100	vacuole protein
comp105492_c0_seq1	2	NA
comp159933_c0_seq1	>100	vacuolar protease A
comp159355_c0_seq4	>100	NA
comp97638_c0_seq1	6	NA
comp172737_c0_seq7	>100	NA
comp172357_c0_seq8	26	family 4 carbohydrate esterase
comp172538_c0_seq2	>100	Endo-1,3-beta-glucanase, family GH81
comp182846_c0_seq1	>100	NA
comp171397_c0_seq4	32	MFS multidrug resistance transporter, putative
comp173026_c0_seq4	19	CMGC/CLK protein kinase (LAMMER)
comp173130_c0_seq2	30	NA
comp177023_c0_seq1	>100	peroxisomal copper amine oxidase
comp96685_c0_seq2	61	Hypothetical protein
comp173888_c0_seq2	87	NA
comp176958_c0_seq1	>100	NA
comp140930_c0_seq1	86	NA
comp160552_c0_seq1	>100	NA
comp174759_c0_seq29	>100	NA
comp173887_c0_seq5	>100	NA
comp171214_c0_seq10	>100	DEAD/DEAH box helicase
comp172604_c1_seq9	>100	similar to ubiquitin carboxyl-terminal hydrolase
comp174746_c0_seq34	>100	NA
comp172312_c0_seq4	>100	Calcium transporting P-type ATPase, putative
comp168166_c0_seq1	>100	NA
comp171838_c0_seq2	>100	High osmolarity signaling protein Sho1
comp181527_c0_seq1	>100	NA

Table A3.1 continued

comp138360_c0_seq2	55	NA
comp159259_c0_seq1	3	NA
comp174390_c1_seq5	98	NA
comp108120_c0_seq1	>100	NA
comp174683_c0_seq32	>100	NA
comp173914_c0_seq17	>100	conidiation-specific expression protein
comp182442_c0_seq1	>100	NA
comp181365_c0_seq1	>100	NA
comp172767_c0_seq1	>100	NA
comp193958_c0_seq1	>100	NA
comp104314_c0_seq1	78	NA
comp177932_c0_seq1	25	sucrase/ferredoxin domain-containing protein
comp174763_c0_seq32	>100	NA
comp174792_c0_seq9	>100	NA
comp170337_c0_seq2	>100	NA
comp96757_c0_seq2	60	NA
comp171267_c0_seq1	>100	proteinase-K
comp178775_c0_seq1	4	NA
comp161551_c0_seq1	>100	glycosyltransferase family 15 protein
comp175072_c0_seq1	4	NA
comp97388_c0_seq1	12	NA
comp173106_c0_seq3	>100	Hypothetical protein
comp175799_c0_seq1	3	small nuclear ribonucleoprotein E
comp171037_c0_seq1	6	rhizopuspepsin 6 precursor
comp168347_c1_seq4	>100	Hypothetical protein
comp175426_c0_seq1	7	MFS peptide transporter, putative
comp170867_c0_seq1	4	histone H2A
comp170721_c0_seq3	>100	drug:h+ antiporter
comp174890_c0_seq1	2	chitinase C1

Table A3.1 continued

comp174753_c1_seq2	6	NA
comp97784_c0_seq2	10	thiamine biosynthesis protein
comp170620_c0_seq5	61	MARK1 protein kinase
comp171789_c0_seq11	10	Surfeit-1
comp171619_c1_seq1	>100	Hypothetical protein
comp173248_c0_seq18	>100	PREDICTED: kinesin-like protein KIF21A
comp166829_c0_seq3	4	sphingolipid long chain base-responsive protein PIL1
comp97780_c0_seq1	3	diadenosine hexaphosphate hydrolase
comp97651_c0_seq1	51	Hypothetical protein
comp175461_c0_seq1	15	chitinase
comp174851_c0_seq1	48	NA
comp195842_c0_seq1	3	uroporphyrinogen-III synthase
comp172773_c2_seq9	>100	Cin1
comp160780_c0_seq1	11	transforming growth factor beta regulator 1
comp173851_c0_seq36	>100	nuclear serine protease HtrA2/Nma111, putative
comp173358_c0_seq1	20	chitin synthase
comp162668_c0_seq4	4	arid/bright DNA binding domain-containing protein
comp175010_c0_seq1	12	NA
comp175043_c0_seq1	6	NA
comp139392_c0_seq1	8	NA
comp97816_c1_seq1	4	Hsp71-like protein
comp99222_c0_seq1	3	Swi1-interacting protein swi3
comp97828_c0_seq1	2	Pth12p (homeobox protein)
comp173320_c4_seq1	2	WD40 repeat-containing protein
comp97582_c0_seq1	>100	HCP-like protein
comp175513_c0_seq1	3	NA
comp170662_c0_seq1	6	Hypothetical protein
comp178252_c0_seq1	25	increased rDNA silencing protein 4
comp180986_c0_seq1	3	Hypothetical protein

Table A3.1 continued

comp97373_c0_seq1	6	Hypothetical protein
comp174555_c0_seq3	72	Hypothetical protein
comp172187_c1_seq2	>100	NA
comp175064_c0_seq1	8	2-haloalkanoic acid dehalogenase, putative
comp97772_c0_seq1	2	Hypothetical protein
comp172187_c1_seq3	20	NA
comp174987_c2_seq1	18	transposase-like protein
comp178828_c0_seq1	4	ribose-5-phosphate isomerase
comp172002_c0_seq1	11	Hypothetical protein
comp171789_c0_seq7	4	Hypothetical protein
comp98130_c0_seq1	92	related to LSB5-possible role in the regulation of actin cytoskeletal organization
comp168209_c0_seq4	2	related to sulphate transporter proteins
comp97582_c0_seq2	4	related to SKT5-protoplast regeneration and killer toxin resistance protein
comp174880_c0_seq1	48	NA
comp171597_c0_seq18	88	Hypothetical protein
comp175985_c0_seq1	22	NA
comp158847_c0_seq1	4	NDI, mitochondrial NADH dehydrogenase
comp175217_c0_seq1	2	chitin deacetylase
comp174598_c0_seq42	28	dicer-2 protein
comp97411_c0_seq2	5	Hypothetical protein
comp159547_c0_seq3	3	MFS general substrate transporter
comp162721_c0_seq4	>100	Hypothetical protein
comp175286_c0_seq1	5	NA
comp106197_c0_seq1	7	Hypothetical protein
comp173106_c0_seq2	>100	NA
comp169217_c0_seq3	>100	ERG4/ERG24 ergosterol biosynthesis protein
comp173465_c0_seq9	>100	something that sticks like glue (*Drosophila melanogaster*), isoform A
comp159745_c0_seq1	18	Cell wall beta-glucan synthesis
comp172640_c1_seq7	79	glycerol kinase 2

Table A3.1 continued

comp174104_c3_seq89	>100	related to Sla1-cytoskeleton assembly control protein
comp172145_c0_seq1	21	acetyl-CoA synthetase-like protein
comp98048_c0_seq2	2	probable acyl-CoA oxidase
comp173799_c0_seq8	3	epoxide hydrolase
comp189801_c0_seq1	5	ATP-dependent RNA helicase Has1
comp180109_c0_seq1	2	pre-mRNA-splicing factor Slu7
comp172187_c1_seq1	63	NA
comp98769_c0_seq1	3	NA
comp177274_c0_seq1	6	NA
comp174852_c0_seq1	5	Cytochrome c oxidase subunit 1

Table A3.2. Conservation of mating relevant genes in the pheromone response pathway across fungi. [a]*S. cerevisiae* nomenclature, [b]*P. blakesleeanus* nomenclature, [c]*U. maydis* nomenclature

Cluster	Sex Genes Recovered	Mucorales	Ascomycota	Basidiomycota	Ascomycota yeast	All yeast
ORTHOMCL1052	*GPA1*[a], *fadA*	4/4	4/5	3/3	1/2	2/3
ORTHOMCL1228	*bpp1*[c], *ste4*, *sfaD*	4/4	5/5	3/3	2/2	3/3
ORTHOMCL2883	*STE18*[a], *gpgA*	3/4	4/5	3/3	1/2	2/3
ORTHOMCL2418	*ste20*	4/4	5/5	3/3	2/2	3/3
ORTHOMCL2199	*ubc4*[c], *ste11*[a], *steC*	4/4	5/5	3/3	2/2	3/3
ORTHOMCL1426	*fuz7*[c], *STE7*[a], *ste7*	4/4	5/5	3/3	2/2	3/3
ORTHOMCL1423	*kpp2*[c], *FUS3*[a], *mpkB*	4/4	5/5	3/3	2/2	3/3
ORTHOMCL19033	*ppgA*	0/4	1/5	0/3	0/2	0/3
ORTHOMCL2362	*pra1*[c], *STE3*[a], *preA*	0/4	4/5	3/3	2/2	3/3
ORTHOMCL8434	*STE2*[a], *preB*	0/4	4/5	0/3	2/2	2/3
ORTHOMCL2374	*ubc2*[c], *STE50*[a], *steD*	4/4	4/5	3/3	1/2	2/3
ORTHOMCL1625	*phnA*	4/4	3/5	3/3	0/2	1/3
ORTHOMCL1695	*CDC42*[a]	4/4	3/5	2/3	0/2	1/3
ORTHOMCL15650	*STE5*[a]	0/4	1/5	0/3	1/2	1/3
ORTHOMCL1582	*STE20*[a]	4/4	5/5	3/3	2/2	3/3
ORTHOMCL1539	*ras1-2*[c], *ras2*[c]	4/4	3/5	3/3	0/2	1/3
ORTHOMCL4925	*sql2*[c]	4/4	3/5	3/3	0/2	1/3
ORTHOMCL5397	*gprD*	0/4	2/5	1/3	0/2	0/3
ORTHOMCL11314	*gprK*	0/4	3/5	0/3	0/2	0/3
ORTHOMCL2146	*SST2*[a], *flbA*	4/4	5/5	3/3	2/2	3/3
ORTHOMCL1625	*phnA*	4/4	3/5	3/3	0/2	1/3
ORTHOMCL1555	*MSG5*[a]	4/4	5/5	3/3	2/2	3/3

Table A3.3. Conservation of mating relevant genes in environmental sensing across fungi. [a]*S. cerevisiae* nomenclature, [b]*P. blakesleeanus* nomenclature, [c]*U. maydis* nomenclature

Cluster	Sex Genes Recovered	Mucorales	Ascomycota	Basidiomycota	Ascomycota yeast	All yeast
ORTHOMCL6805	*fhpA*	0/4	3/5	3/3	0/2	1/3
ORTHOMCL3339	*lreA*	4/4	3/5	3/3	0/2	1/3
ORTHOMCL1029	*lreB*	4/4	3/5	3/3	0/2	1/3
ORTHOMCL7629	*cryA*	0/4	4/5	2/3	1/2	1/3
ORTHOMCL729	*veA, vosA*	4/4	3/5	3/3	0/2	1/3
ORTHOMCL643	*velB*	4/4	3/5	3/3	0/2	1/3
ORTHOMCL8445	*velC*	4/4	1/5	0/3	0/2	0/3
ORTHOMCL5878	*laeA*	1/4	3/5	0/3	0/2	0/3
ORTHOMCL2400	*crk1*[c], *imeB*	4/4	5/5	3/3	2/2	3/3
ORTHOMCL9631	*silA*	0/4	3/5	1/3	1/2	1/3
ORTHOMCL6352	*silG*	1/4	5/5	2/3	2/2	3/3
ORTHOMCL7553	*cpcA*	2/4	4/5	0/3	1/2	1/3
ORTHOMCL2456	*cpcB*	4/4	5/5	3/3	2/2	3/3
ORTHOMCL19276	*lsdA*	0/4	1/5	0/3	0/2	0/3
ORTHOMCL2208	*phoA*	4/4	5/5	2/3	2/2	3/3
ORTHOMCL4435	*pho80*	3/4	4/5	3/3	1/2	2/3
ORTHOMCL11261	*esdC*	0/4	3/5	0/3	0/2	0/3
ORTHOMCL6874	*fhbA*	0/4	5/5	1/3	2/2	3/3
ORTHOMCL1790	*hogA*	4/4	5/5	3/3	2/2	3/3
ORTHOMCL1572	*atfA*	4/4	4/5	2/3	1/2	2/3
ORTHOMCL1695	*CDC42*[a]	4/4	3/5	2/3	0/2	1/3
ORTHOMCL2031	*ras1-1*[c], *rasA*	4/4	3/5	2/3	0/2	1/3
ORTHOMCL1539	*ras1-2*[c], *ras2*[c]	4/4	3/5	3/3	0/2	1/3

Table A3.4. Conservation of mating relevant transcription factors across fungi. [a]*S. cerevisiae* nomenclature, [b]*P. blakesleeanus* nomenclature, [c]*U. maydis* nomenclature

Cluster	Sex Genes Recovered	Mucorales	Ascomycota	Basidiomycota	Ascomycota yeast	All yeast
ORTHOMCL1174	*stuA*	4/4	4/5	0/3	1/2	1/3
ORTHOMCL2402	*medA*	4/4	3/5	3/3	0/2	1/3
ORTHOMCL2368	*devR*	4/4	4/5	3/3	1/2	2/3
ORTHOMCL33	*dopA*	4/4	5/5	3/3	2/2	3/3
ORTHOMCL1735	*nsdC*	4/4	3/5	0/3	0/2	0/3
ORTHOMCL2183	*nsdD*	4/4	2/5	3/3	0/2	1/3
ORTHOMCL5421	*nosA*	0/4	3/5	3/3	0/2	1/3
ORTHOMCL19082	*rosA*	0/4	1/5	0/3	0/2	0/3
ORTHOMCL1307	*flbC*	3/4	4/5	1/3	1/2	1/3
ORTHOMCL11634	*flbE*	0/4	3/5	0/3	0/2	0/3
ORTHOMCL7534	*fhpA*	1/4	4/5	0/3	1/2	1/3
ORTHOMCL777	*msnA*	4/4	5/5	2/3	2/2	2/3
ORTHOMCL1280	*rcoA*	4/4	5/5	3/3	2/2	3/3
ORTHOMCL1445	*csnA*	4/4	4/5	3/3	1/2	2/3
ORTHOMCL2432	*csnB*	4/4	4/5	3/3	1/2	2/3
ORTHOMCL3426	*csnD*	4/4	4/5	3/3	1/2	2/3
ORTHOMCL3394	*csnE*	4/4	5/5	3/3	2/2	3/3
ORTHOMCL1959	*acoB*	4/4	4/5	3/3	1/2	2/3
ORTHOMCL1438	*STE12*[a], *steA*	4/4	4/5	2/3	1/2	2/3
ORTHOMCL729	*veA, vosA*	4/4	3/5	3/3	0/2	1/3
ORTHOMCL643	*velB*	4/4	3/5	3/3	0/2	1/3
ORTHOMCL8445	*velC*	4/4	1/5	0/3	0/2	0/3
ORTHOMCL15644	*SIR4*[a]	0/4	1/5	0/3	1/2	1/3
ORTHOMCL15583	*SIR3*[a]	0/4	1/5	0/3	1/2	1/3
ORTHOMCL207	*SexM*[b], *prf1*[c], *rop1*[c], *MAT2*	4/4	5/5	3/3	2/2	3/3
ORTHOMCL15543	*bW1*[c]	0/4	0/5	1/3	0/2	0/3
ORTHOMCL15544	*bE11*[c]	0/4	0/5	1/3	0/2	0/3

ORTHOMCL557	*SexP*[b]	4/4	0/5	1/3	0/2	0/3

Table A3.5. Conservation of genes involved in late sexual development across fungi. [a]*S. cerevisiae* nomenclature, [b]*P. blakesleeanus* nomenclature, [c]*U. maydis* nomenclature

Cluster	Sex Genes Recovered	Mucorales	Ascomycota	Basidiomycota	Ascomycota yeast	All yeast
ORTHOMCL1445	*csnA*	4/4	4/5	3/3	1/2	2/3
ORTHOMCL2432	*csnB*	4/4	4/5	3/3	1/2	2/3
ORTHOMCL3426	*csnD*	4/4	4/5	3/3	1/2	2/3
ORTHOMCL3394	*csnE*	4/4	5/5	3/3	2/2	3/3
ORTHOMCL1959	*acoB*	4/4	4/5	3/3	1/2	2/3
ORTHOMCL4966	*hxtA*	4/4	3/5	1/3	0/2	0/3
ORTHOMCL5421	*nosA*	0/4	3/5	3/3	0/2	1/3
ORTHOMCL2400	*crk1*[c]*, imeB*	4/4	5/5	3/3	2/2	3/3
ORTHOMCL8445	*velC*	4/4	1/5	0/3	0/2	0/3
ORTHOMCL2402	*medA*	4/4	3/5	3/3	0/2	1/3
ORTHOMCL3471	*mutA*	0/4	4/5	2/3	1/2	2/3
ORTHOMCL11314	*gprK*	0/4	3/5	0/3	0/2	0/3
ORTHOMCL7514	*noxA*	0/4	3/5	1/3	0/2	0/3
ORTHOMCL738	*sidC*	0/4	4/5	2/3	1/2	1/3
ORTHOMCL7534	*fhpA*	1/4	4/5	0/3	1/2	1/3
ORTHOMCL19480	*candA-N*	0/4	1/5	0/3	0/2	0/3
ORTHOMCL2887	*candA-C*	4/4	4/5	3/3	1/2	2/3
ORTHOMCL2561	*grrA*	3/4	5/5	3/3	2/2	3/3
ORTHOMCL856	*tubA*	4/4	5/5	3/3	2/2	3/3
ORTHOMCL2264	*samB*	4/4	5/5	3/3	2/2	3/3
ORTHOMCL729	*veA, vosA*	4/4	3/5	3/3	0/2	1/3
ORTHOMCL504	*strA*	4/4	4/5	3/3	1/2	2/3
ORTHOMCL7553	*cpcA*	2/4	4/5	0/3	1/2	1/3
ORTHOMCL2456	*cpcB*	4/4	5/5	3/3	2/2	3/3
ORTHOMCL5878	*laeA*	1/4	3/5	0/3	0/2	0/3

Table A3.6. Sexually relevant Mucoromycotina specific genes.

Cluster	Sex-Related Transcripts Recovered	Mucorales	Ascomycota	Basidiomycota	Ascomycota yeast	All yeast
ORTHOMCL3	m.91485, m.91486, m.91581, m.91582, m.91650, m.91651, m.91722, m.91723, m.91813, m.91814	4/4	0/5	0/3	0/2	0/3
ORTHOMCL10	m.92126, m.92472	4/4	0/5	0/3	0/2	0/3
ORTHOMCL17	m.95606, m.95775	4/4	0/5	0/3	0/2	0/3
ORTHOMCL48	m.68772, m.68812, m.68814, m.68820, m.68821, m.68834, m.68836, m.68840, m.68900, m.68912, m.69004	4/4	0/5	0/3	0/2	0/3
ORTHOMCL61	m.18640	4/4	0/5	0/3	0/2	0/3
ORTHOMCL77	m.77970, m.78126	4/4	0/5	0/3	0/2	0/3
ORTHOMCL88	m.72363	4/4	0/5	0/3	0/2	0/3
ORTHOMCL116	m.85961	4/4	0/5	0/3	0/2	0/3
ORTHOMCL122	m.70591	4/4	0/5	0/3	0/2	0/3
ORTHOMCL140	m.71557, m.71813, m.71845	4/4	0/5	0/3	0/2	0/3
ORTHOMCL145	m.71561, m.71814, m.71846	4/4	0/5	0/3	0/2	0/3
ORTHOMCL146	m.68167, m.68168	4/4	0/5	0/3	0/2	0/3
ORTHOMCL205	m.70592	4/4	0/5	0/3	0/2	0/3
ORTHOMCL213	m.75796	4/4	0/5	0/3	0/2	0/3
ORTHOMCL230	m.72514	4/4	0/5	0/3	0/2	0/3
ORTHOMCL240	m.47457	4/4	0/5	0/3	0/2	0/3
ORTHOMCL258	m.49752	4/4	0/5	0/3	0/2	0/3
ORTHOMCL259	m.93482	4/4	0/5	0/3	0/2	0/3
ORTHOMCL266	m.56044, m.56045	4/4	0/5	0/3	0/2	0/3
ORTHOMCL318	m.58211, m.58212	4/4	0/5	0/3	0/2	0/3
ORTHOMCL426	m.90234	4/4	0/5	0/3	0/2	0/3
ORTHOMCL428	m.43474	4/4	0/5	0/3	0/2	0/3
ORTHOMCL495	m.75489	4/4	0/5	0/3	0/2	0/3
ORTHOMCL496	m.61575	4/4	0/5	0/3	0/2	0/3

Table A3.6 continued

ORTHOMCL614	m.58688, m.58734, m.58798	4/4	0/5	0/3	0/2	0/3
ORTHOMCL655	m.26801	4/4	0/5	0/3	0/2	0/3
ORTHOMCL694	m.92940	4/4	0/5	0/3	0/2	0/3
ORTHOMCL698	m.22979	4/4	0/5	0/3	0/2	0/3
ORTHOMCL701	m.84072	4/4	0/5	0/3	0/2	0/3
ORTHOMCL702	m.99280, m.99287	4/4	0/5	0/3	0/2	0/3
ORTHOMCL746	m.39180	4/4	0/5	0/3	0/2	0/3
ORTHOMCL803	m.61137	4/4	0/5	0/3	0/2	0/3
ORTHOMCL873	m.58478, m.58485	4/4	0/5	0/3	0/2	0/3
ORTHOMCL924	m.57956	4/4	0/5	0/3	0/2	0/3
ORTHOMCL926	m.64120, m.64121	4/4	0/5	0/3	0/2	0/3
ORTHOMCL929	m.82244	4/4	0/5	0/3	0/2	0/3
ORTHOMCL990	m.72516	4/4	0/5	0/3	0/2	0/3
ORTHOMCL1066	m.41256	4/4	0/5	0/3	0/2	0/3
ORTHOMCL1145	m.35609	4/4	0/5	0/3	0/2	0/3
ORTHOMCL1149	m.33115	4/4	0/5	0/3	0/2	0/3
ORTHOMCL1264	m.67222	4/4	0/5	0/3	0/2	0/3
ORTHOMCL1361	m.59282	4/4	0/5	0/3	0/2	0/3
ORTHOMCL1379	m.18183	4/4	0/5	0/3	0/2	0/3
ORTHOMCL1493	m.41257	4/4	0/5	0/3	0/2	0/3
ORTHOMCL1649	m.49753	4/4	0/5	0/3	0/2	0/3
ORTHOMCL1839	m.60083, m.60089	4/4	0/5	0/3	0/2	0/3
ORTHOMCL2057	m.38287, m.38326	4/4	0/5	0/3	0/2	0/3
ORTHOMCL2059	m.70151	4/4	0/5	0/3	0/2	0/3
ORTHOMCL2065	m.79829	4/4	0/5	0/3	0/2	0/3
ORTHOMCL2069	m.99188, m.99197	4/4	0/5	0/3	0/2	0/3
ORTHOMCL2086	m.35608	4/4	0/5	0/3	0/2	0/3
ORTHOMCL2087	m.25453, m.25458	4/4	0/5	0/3	0/2	0/3
ORTHOMCL2337	m.98431	4/4	0/5	0/3	0/2	0/3

Table A3.6 continued

ORTHOMCL2656	m.21208	4/4	0/5	0/3	0/2	0/3
ORTHOMCL2685	m.43027	4/4	0/5	0/3	0/2	0/3
ORTHOMCL3059	m.1409	4/4	0/5	0/3	0/2	0/3
ORTHOMCL3061	m.50014, m.50056	4/4	0/5	0/3	0/2	0/3
ORTHOMCL3063	m.59281	4/4	0/5	0/3	0/2	0/3
ORTHOMCL3642	m.25220	4/4	0/5	0/3	0/2	0/3
ORTHOMCL3682	m.7211, m.7212	4/4	0/5	0/3	0/2	0/3
ORTHOMCL4252	m.94573	4/4	0/5	0/3	0/2	0/3
ORTHOMCL4257	m.18184	4/4	0/5	0/3	0/2	0/3
ORTHOMCL4258	m.64311	4/4	0/5	0/3	0/2	0/3
ORTHOMCL4751	m.78455	4/4	0/5	0/3	0/2	0/3
ORTHOMCL4760	m.32929	4/4	0/5	0/3	0/2	0/3
ORTHOMCL4772	m.16857	4/4	0/5	0/3	0/2	0/3
ORTHOMCL4816	m.6319	4/4	0/5	0/3	0/2	0/3
ORTHOMCL5239	m.30131, m.30139	4/4	0/5	0/3	0/2	0/3
ORTHOMCL5311	m.22917	4/4	0/5	0/3	0/2	0/3
ORTHOMCL5592	m.15669	4/4	0/5	0/3	0/2	0/3
ORTHOMCL5615	m.15691	4/4	0/5	0/3	0/2	0/3
ORTHOMCL5623	m.13755	4/4	0/5	0/3	0/2	0/3
ORTHOMCL5691	m.4980, m.4991	4/4	0/5	0/3	0/2	0/3
ORTHOMCL5704	m.75035	4/4	0/5	0/3	0/2	0/3
ORTHOMCL6036	m.43609, m.43613, m.43615	4/4	0/5	0/3	0/2	0/3
ORTHOMCL6050	m.17215	4/4	0/5	0/3	0/2	0/3
ORTHOMCL6554	m.3668	4/4	0/5	0/3	0/2	0/3
ORTHOMCL6561	m.11116	4/4	0/5	0/3	0/2	0/3
ORTHOMCL6562	m.143454	4/4	0/5	0/3	0/2	0/3
ORTHOMCL6564	m.32109	4/4	0/5	0/3	0/2	0/3
ORTHOMCL6570	m.41344	4/4	0/5	0/3	0/2	0/3
ORTHOMCL6576	m.18235	4/4	0/5	0/3	0/2	0/3

Table A3.6 continued

ORTHOMCL6589	m.3667	4/4	0/5	0/3	0/2	0/3
ORTHOMCL6592	m.30812	4/4	0/5	0/3	0/2	0/3
ORTHOMCL6602	m.84898	4/4	0/5	0/3	0/2	0/3
ORTHOMCL7094	m.1997	4/4	0/5	0/3	0/2	0/3
ORTHOMCL7102	m.60299	4/4	0/5	0/3	0/2	0/3
ORTHOMCL7105	m.4978	4/4	0/5	0/3	0/2	0/3
ORTHOMCL7148	m.15587	4/4	0/5	0/3	0/2	0/3
ORTHOMCL7150	m.2704	4/4	0/5	0/3	0/2	0/3
ORTHOMCL7193	m.155	4/4	0/5	0/3	0/2	0/3
ORTHOMCL7217	m.8927	4/4	0/5	0/3	0/2	0/3
ORTHOMCL7392	m.2242	4/4	0/5	0/3	0/2	0/3
ORTHOMCL7431	m.19863	4/4	0/5	0/3	0/2	0/3
ORTHOMCL7451	m.142416	4/4	0/5	0/3	0/2	0/3
ORTHOMCL7903	m.3534	4/4	0/5	0/3	0/2	0/3
ORTHOMCL7906	m.144125	4/4	0/5	0/3	0/2	0/3
ORTHOMCL7915	m.6968	4/4	0/5	0/3	0/2	0/3
ORTHOMCL7922	m.109	4/4	0/5	0/3	0/2	0/3
ORTHOMCL7929	m.90127	4/4	0/5	0/3	0/2	0/3
ORTHOMCL7958	m.4018	4/4	0/5	0/3	0/2	0/3
ORTHOMCL7978	m.17057	4/4	0/5	0/3	0/2	0/3
ORTHOMCL7990	m.143172	4/4	0/5	0/3	0/2	0/3
ORTHOMCL8000	m.72263	4/4	0/5	0/3	0/2	0/3
ORTHOMCL8007	m.23431	4/4	0/5	0/3	0/2	0/3
ORTHOMCL8009	m.21651	4/4	0/5	0/3	0/2	0/3
ORTHOMCL8042	m.4250	4/4	0/5	0/3	0/2	0/3
ORTHOMCL8074	m.16101	4/4	0/5	0/3	0/2	0/3
ORTHOMCL8126	m.11246	4/4	0/5	0/3	0/2	0/3
ORTHOMCL8263	m.1856	4/4	0/5	0/3	0/2	0/3
ORTHOMCL8862	m.2765	4/4	0/5	0/3	0/2	0/3

Table A3.6 continued

ORTHOMCL8872	m.22748	4/4	0/5	0/3	0/2	0/3
ORTHOMCL8873	m.143991	4/4	0/5	0/3	0/2	0/3
ORTHOMCL8884	m.11131	4/4	0/5	0/3	0/2	0/3
ORTHOMCL8938	m.2013	4/4	0/5	0/3	0/2	0/3
ORTHOMCL8960	m.40832	4/4	0/5	0/3	0/2	0/3
ORTHOMCL8970	m.2091	4/4	0/5	0/3	0/2	0/3
ORTHOMCL8984	m.12296	4/4	0/5	0/3	0/2	0/3
ORTHOMCL8996	m.31116	4/4	0/5	0/3	0/2	0/3
ORTHOMCL9016	m.143110	4/4	0/5	0/3	0/2	0/3
ORTHOMCL9073	m.5150	4/4	0/5	0/3	0/2	0/3
ORTHOMCL9122	m.143814	4/4	0/5	0/3	0/2	0/3
ORTHOMCL9141	m.20146	4/4	0/5	0/3	0/2	0/3
ORTHOMCL9150	m.15620	4/4	0/5	0/3	0/2	0/3
ORTHOMCL78	m.77751, m.77773	3/4	0/5	0/3	0/2	0/3
ORTHOMCL103	m.87580	3/4	0/5	0/3	0/2	0/3
ORTHOMCL157	m.82245	3/4	0/5	0/3	0/2	0/3
ORTHOMCL195	m.92129	3/4	0/5	0/3	0/2	0/3
ORTHOMCL245	m.84562	3/4	0/5	0/3	0/2	0/3
ORTHOMCL278	m.49757	3/4	0/5	0/3	0/2	0/3
ORTHOMCL315	m.86213, m.86247	3/4	0/5	0/3	0/2	0/3
ORTHOMCL319	m.66080	3/4	0/5	0/3	0/2	0/3
ORTHOMCL379	m.101504	3/4	0/5	0/3	0/2	0/3
ORTHOMCL493	m.39181	3/4	0/5	0/3	0/2	0/3
ORTHOMCL497	m.90454, m.90558	3/4	0/5	0/3	0/2	0/3
ORTHOMCL553	m.34044, m.34045	3/4	0/5	0/3	0/2	0/3
ORTHOMCL556	m.20242	3/4	0/5	0/3	0/2	0/3
ORTHOMCL581	m.73558	3/4	0/5	0/3	0/2	0/3
ORTHOMCL749	m.86390, m.86399, m.86401, m.86407, m.86416	3/4	0/5	0/3	0/2	0/3

Table A3.6 continued

ORTHOMCL1063	m.72515	3/4	0/5	0/3	0/2	0/3
ORTHOMCL1064	m.55829	3/4	0/5	0/3	0/2	0/3
ORTHOMCL1068	m.57957	3/4	0/5	0/3	0/2	0/3
ORTHOMCL1077	m.66421	3/4	0/5	0/3	0/2	0/3
ORTHOMCL1144	m.68384, m.68392	3/4	0/5	0/3	0/2	0/3
ORTHOMCL1151	m.62095, m.62103	3/4	0/5	0/3	0/2	0/3
ORTHOMCL1267	m.84078	3/4	0/5	0/3	0/2	0/3
ORTHOMCL1356	m.66690, m.66715	3/4	0/5	0/3	0/2	0/3
ORTHOMCL1362	m.86106, m.86125	3/4	0/5	0/3	0/2	0/3
ORTHOMCL1484	m.51069	3/4	0/5	0/3	0/2	0/3
ORTHOMCL1487	m.15053	3/4	0/5	0/3	0/2	0/3
ORTHOMCL1491	m.57824, m.57827	3/4	0/5	0/3	0/2	0/3
ORTHOMCL1648	m.62042, m.62047	3/4	0/5	0/3	0/2	0/3
ORTHOMCL1670	m.26521	3/4	0/5	0/3	0/2	0/3
ORTHOMCL1830	m.30519	3/4	0/5	0/3	0/2	0/3
ORTHOMCL1836	m.73559	3/4	0/5	0/3	0/2	0/3
ORTHOMCL2089	m.13754	3/4	0/5	0/3	0/2	0/3
ORTHOMCL2302	m.15052	3/4	0/5	0/3	0/2	0/3
ORTHOMCL2305	m.17089	3/4	0/5	0/3	0/2	0/3
ORTHOMCL2651	m.25285	3/4	0/5	0/3	0/2	0/3
ORTHOMCL2654	m.43136	3/4	0/5	0/3	0/2	0/3
ORTHOMCL3077	m.22302	3/4	0/5	0/3	0/2	0/3
ORTHOMCL3621	m.91199	3/4	0/5	0/3	0/2	0/3
ORTHOMCL3636	m.78363	3/4	0/5	0/3	0/2	0/3
ORTHOMCL4744	m.85621	3/4	0/5	0/3	0/2	0/3
ORTHOMCL4843	m.44872	3/4	0/5	0/3	0/2	0/3
ORTHOMCL5196	m.11593	3/4	0/5	0/3	0/2	0/3
ORTHOMCL5213	m.36843	3/4	0/5	0/3	0/2	0/3
ORTHOMCL5227	m.72217	3/4	0/5	0/3	0/2	0/3

Table A3.6 continued

ORTHOMCL5233	m.80140	3/4	0/5	0/3	0/2	0/3
ORTHOMCL5236	m.11592	3/4	0/5	0/3	0/2	0/3
ORTHOMCL5565	m.5947	3/4	0/5	0/3	0/2	0/3
ORTHOMCL5993	m.40995	3/4	0/5	0/3	0/2	0/3
ORTHOMCL6002	m.42071	3/4	0/5	0/3	0/2	0/3
ORTHOMCL6014	m.44243	3/4	0/5	0/3	0/2	0/3
ORTHOMCL6587	m.75034	3/4	0/5	0/3	0/2	0/3
ORTHOMCL6594	m.30257, m.30266	3/4	0/5	0/3	0/2	0/3
ORTHOMCL6599	m.94702	3/4	0/5	0/3	0/2	0/3
ORTHOMCL6627	m.9712	3/4	0/5	0/3	0/2	0/3
ORTHOMCL7074	m.32379	3/4	0/5	0/3	0/2	0/3
ORTHOMCL7078	m.20738	3/4	0/5	0/3	0/2	0/3
ORTHOMCL7097	m.23910	3/4	0/5	0/3	0/2	0/3
ORTHOMCL7138	m.66079	3/4	0/5	0/3	0/2	0/3
ORTHOMCL7176	m.30811	3/4	0/5	0/3	0/2	0/3
ORTHOMCL7870	m.3052	3/4	0/5	0/3	0/2	0/3
ORTHOMCL7898	m.44153, m.44156	3/4	0/5	0/3	0/2	0/3
ORTHOMCL7961	m.1001, m.1007	3/4	0/5	0/3	0/2	0/3
ORTHOMCL8031	m.22145	3/4	0/5	0/3	0/2	0/3
ORTHOMCL8818	m.16904	3/4	0/5	0/3	0/2	0/3
ORTHOMCL8875	m.16023	3/4	0/5	0/3	0/2	0/3
ORTHOMCL8949	m.3917	3/4	0/5	0/3	0/2	0/3
ORTHOMCL8951	m.55738	3/4	0/5	0/3	0/2	0/3
ORTHOMCL8979	m.867	3/4	0/5	0/3	0/2	0/3
ORTHOMCL9075	m.60300	3/4	0/5	0/3	0/2	0/3
ORTHOMCL9105	m.14231	3/4	0/5	0/3	0/2	0/3
ORTHOMCL9137	m.142750	3/4	0/5	0/3	0/2	0/3
ORTHOMCL9149	m.5618	3/4	0/5	0/3	0/2	0/3
ORTHOMCL9446	m.12163	3/4	0/5	0/3	0/2	0/3

Table A3.6 continued

ORTHOMCL10286	m.6706	3/4	0/5	0/3	0/2	0/3
ORTHOMCL10321	m.59530	3/4	0/5	0/3	0/2	0/3
ORTHOMCL10325	m.144632	3/4	0/5	0/3	0/2	0/3
ORTHOMCL10425	m.362	3/4	0/5	0/3	0/2	0/3
ORTHOMCL10430	m.1025	3/4	0/5	0/3	0/2	0/3
ORTHOMCL10471	m.1764	3/4	0/5	0/3	0/2	0/3
ORTHOMCL10473	m.143622	3/4	0/5	0/3	0/2	0/3
ORTHOMCL10497	m.2534	3/4	0/5	0/3	0/2	0/3
ORTHOMCL10529	m.6181	3/4	0/5	0/3	0/2	0/3
ORTHOMCL10606	m.2460	3/4	0/5	0/3	0/2	0/3
ORTHOMCL10681	m.12582	3/4	0/5	0/3	0/2	0/3
ORTHOMCL10716	m.91418	3/4	0/5	0/3	0/2	0/3
ORTHOMCL10739	m.143836	3/4	0/5	0/3	0/2	0/3
ORTHOMCL10800	m.13119	3/4	0/5	0/3	0/2	0/3
ORTHOMCL46	m.93479	2/4	0/5	0/3	0/2	0/3
ORTHOMCL69	m.81113	2/4	0/5	0/3	0/2	0/3
ORTHOMCL132	m.93053, m.93054	2/4	0/5	0/3	0/2	0/3
ORTHOMCL144	m.83551, m.83563	2/4	0/5	0/3	0/2	0/3
ORTHOMCL162	m.93052	2/4	0/5	0/3	0/2	0/3
ORTHOMCL200	m.85965	2/4	0/5	0/3	0/2	0/3
ORTHOMCL303	m.52032	2/4	0/5	0/3	0/2	0/3
ORTHOMCL317	m.60659	2/4	0/5	0/3	0/2	0/3
ORTHOMCL502	m.75488	2/4	0/5	0/3	0/2	0/3
ORTHOMCL688	m.76668	2/4	0/5	0/3	0/2	0/3
ORTHOMCL689	m.77753, m.77774	2/4	0/5	0/3	0/2	0/3
ORTHOMCL690	m.85156	2/4	0/5	0/3	0/2	0/3
ORTHOMCL1138	m.87579	2/4	0/5	0/3	0/2	0/3
ORTHOMCL1146	m.32559	2/4	0/5	0/3	0/2	0/3
ORTHOMCL1247	m.55256	2/4	0/5	0/3	0/2	0/3

Table A3.6 continued

ORTHOMCL1490	m.76424, m.76449, m.76463, m.76485	2/4	0/5	0/3	0/2	0/3
ORTHOMCL1818	m.34092, m.34093	2/4	0/5	0/3	0/2	0/3
ORTHOMCL1864	m.114103, m.114143	2/4	0/5	0/3	0/2	0/3
ORTHOMCL2066	m.34933	2/4	0/5	0/3	0/2	0/3
ORTHOMCL2295	m.62732	2/4	0/5	0/3	0/2	0/3
ORTHOMCL2297	m.78128	2/4	0/5	0/3	0/2	0/3
ORTHOMCL2646	m.32658, m.32661, m.32667	2/4	0/5	0/3	0/2	0/3
ORTHOMCL3051	m.47174	2/4	0/5	0/3	0/2	0/3
ORTHOMCL3607	m.73468	2/4	0/5	0/3	0/2	0/3
ORTHOMCL3614	m.78692	2/4	0/5	0/3	0/2	0/3
ORTHOMCL3633	m.65247	2/4	0/5	0/3	0/2	0/3
ORTHOMCL3645	m.64965, m.64969, m.64973, m.64977	2/4	0/5	0/3	0/2	0/3
ORTHOMCL4229	m.92474	2/4	0/5	0/3	0/2	0/3
ORTHOMCL4265	m.57510	2/4	0/5	0/3	0/2	0/3
ORTHOMCL4742	m.90132, m.90135, m.90136	2/4	0/5	0/3	0/2	0/3
ORTHOMCL5179	m.37154	2/4	0/5	0/3	0/2	0/3
ORTHOMCL5180	m.37151	2/4	0/5	0/3	0/2	0/3
ORTHOMCL5183	m.81838	2/4	0/5	0/3	0/2	0/3
ORTHOMCL5559	m.52547	2/4	0/5	0/3	0/2	0/3
ORTHOMCL5560	m.19716	2/4	0/5	0/3	0/2	0/3
ORTHOMCL5581	m.54153	2/4	0/5	0/3	0/2	0/3
ORTHOMCL5964	m.32498, m.32501	2/4	0/5	0/3	0/2	0/3
ORTHOMCL5967	m.78690	2/4	0/5	0/3	0/2	0/3
ORTHOMCL7062	m.23909	2/4	0/5	0/3	0/2	0/3
ORTHOMCL7064	m.57072	2/4	0/5	0/3	0/2	0/3
ORTHOMCL7079	m.9714	2/4	0/5	0/3	0/2	0/3
ORTHOMCL7166	m.23044, m.23050, m.23057	2/4	0/5	0/3	0/2	0/3

Table A3.6 continued

ORTHOMCL7436	m.16914	2/4	0/5	0/3	0/2	0/3
ORTHOMCL7832	m.85235	2/4	0/5	0/3	0/2	0/3
ORTHOMCL7837	m.40818	2/4	0/5	0/3	0/2	0/3
ORTHOMCL7847	m.22746	2/4	0/5	0/3	0/2	0/3
ORTHOMCL7873	m.81536	2/4	0/5	0/3	0/2	0/3
ORTHOMCL7913	m.20374, m.20380	2/4	0/5	0/3	0/2	0/3
ORTHOMCL7941	m.20375	2/4	0/5	0/3	0/2	0/3
ORTHOMCL7993	m.34430	2/4	0/5	0/3	0/2	0/3
ORTHOMCL8292	m.6461	2/4	0/5	0/3	0/2	0/3
ORTHOMCL8747	m.16839, m.16841	2/4	0/5	0/3	0/2	0/3
ORTHOMCL8781	m.5155, m.5156	2/4	0/5	0/3	0/2	0/3
ORTHOMCL8841	m.14697	2/4	0/5	0/3	0/2	0/3
ORTHOMCL9429	m.571	2/4	0/5	0/3	0/2	0/3
ORTHOMCL10229	m.143469	2/4	0/5	0/3	0/2	0/3
ORTHOMCL10246	m.53836	2/4	0/5	0/3	0/2	0/3
ORTHOMCL10247	m.11910	2/4	0/5	0/3	0/2	0/3
ORTHOMCL10251	m.15994	2/4	0/5	0/3	0/2	0/3
ORTHOMCL10259	m.81564	2/4	0/5	0/3	0/2	0/3
ORTHOMCL10268	m.87545	2/4	0/5	0/3	0/2	0/3
ORTHOMCL10312	m.100539	2/4	0/5	0/3	0/2	0/3
ORTHOMCL10687	m.15870	2/4	0/5	0/3	0/2	0/3
ORTHOMCL10806	m.6666	2/4	0/5	0/3	0/2	0/3
ORTHOMCL10815	m.50899	2/4	0/5	0/3	0/2	0/3
ORTHOMCL12044	m.16646	2/4	0/5	0/3	0/2	0/3
ORTHOMCL12053	m.142	2/4	0/5	0/3	0/2	0/3
ORTHOMCL12056	m.144564	2/4	0/5	0/3	0/2	0/3
ORTHOMCL12060	m.143104	2/4	0/5	0/3	0/2	0/3
ORTHOMCL12066	m.7562	2/4	0/5	0/3	0/2	0/3
ORTHOMCL12077	m.144126	2/4	0/5	0/3	0/2	0/3

Table A3.6 continued

ORTHOMCL12078	m.5141	2/4	0/5	0/3	0/2	0/3
ORTHOMCL12094	m.4094	2/4	0/5	0/3	0/2	0/3
ORTHOMCL12100	m.8990	2/4	0/5	0/3	0/2	0/3
ORTHOMCL12119	m.6524	2/4	0/5	0/3	0/2	0/3
ORTHOMCL12150	m.144135	2/4	0/5	0/3	0/2	0/3
ORTHOMCL12229	m.144121	2/4	0/5	0/3	0/2	0/3
ORTHOMCL12243	m.227	2/4	0/5	0/3	0/2	0/3
ORTHOMCL12359	m.4097	2/4	0/5	0/3	0/2	0/3
ORTHOMCL12374	m.11903	2/4	0/5	0/3	0/2	0/3
ORTHOMCL12434	m.143526	2/4	0/5	0/3	0/2	0/3
ORTHOMCL12507	m.143895	2/4	0/5	0/3	0/2	0/3
ORTHOMCL28	m.80207, m.80371, m.80470	1/4	0/5	0/3	0/2	0/3
ORTHOMCL60	m.85967	1/4	0/5	0/3	0/2	0/3
ORTHOMCL75	m.81115	1/4	0/5	0/3	0/2	0/3
ORTHOMCL108	m.49407, m.49488	1/4	0/5	0/3	0/2	0/3
ORTHOMCL114	m.81114	1/4	0/5	0/3	0/2	0/3
ORTHOMCL123	m.63392, m.63421, m.63540	1/4	0/5	0/3	0/2	0/3
ORTHOMCL128	m.79067, m.79075	1/4	0/5	0/3	0/2	0/3
ORTHOMCL160	m.26633, m.26634, m.26735, m.26736	1/4	0/5	0/3	0/2	0/3
ORTHOMCL161	m.26631, m.26632, m.26734	1/4	0/5	0/3	0/2	0/3
ORTHOMCL229	m.84573	1/4	0/5	0/3	0/2	0/3
ORTHOMCL288	m.129944, m.129946	1/4	0/5	0/3	0/2	0/3
ORTHOMCL331	m.90262	1/4	0/5	0/3	0/2	0/3
ORTHOMCL392	m.80206, m.80370, m.80469	1/4	0/5	0/3	0/2	0/3
ORTHOMCL424	m.129994, m.129996	1/4	0/5	0/3	0/2	0/3
ORTHOMCL448	m.79195	1/4	0/5	0/3	0/2	0/3
ORTHOMCL449	m.72953	1/4	0/5	0/3	0/2	0/3
ORTHOMCL580	m.92473	1/4	0/5	0/3	0/2	0/3

Table A3.6 continued

ORTHOMCL685	m.68775, m.68974	1/4	0/5	0/3	0/2	0/3
ORTHOMCL686	m.68774, m.68973	1/4	0/5	0/3	0/2	0/3
ORTHOMCL687	m.54320, m.54344, m.54348, m.54360	1/4	0/5	0/3	0/2	0/3
ORTHOMCL790	m.69650	1/4	0/5	0/3	0/2	0/3
ORTHOMCL791	m.66422	1/4	0/5	0/3	0/2	0/3
ORTHOMCL863	m.129939	1/4	0/5	0/3	0/2	0/3
ORTHOMCL864	m.114115, m.114116, m.114155, m.114156	1/4	0/5	0/3	0/2	0/3
ORTHOMCL915	m.34934	1/4	0/5	0/3	0/2	0/3
ORTHOMCL979	m.72518	1/4	0/5	0/3	0/2	0/3
ORTHOMCL1055	m.44221, m.44244, m.44252	1/4	0/5	0/3	0/2	0/3
ORTHOMCL1137	m.68414, m.68439	1/4	0/5	0/3	0/2	0/3
ORTHOMCL1245	m.84564	1/4	0/5	0/3	0/2	0/3
ORTHOMCL1246	m.77972	1/4	0/5	0/3	0/2	0/3
ORTHOMCL1370	m.79068	1/4	0/5	0/3	0/2	0/3
ORTHOMCL1479	m.70691	1/4	0/5	0/3	0/2	0/3
ORTHOMCL1480	m.55830, m.55930	1/4	0/5	0/3	0/2	0/3
ORTHOMCL1661	m.82870	1/4	0/5	0/3	0/2	0/3
ORTHOMCL1816	m.96607	1/4	0/5	0/3	0/2	0/3
ORTHOMCL1849	m.71034	1/4	0/5	0/3	0/2	0/3
ORTHOMCL2046	m.56314	1/4	0/5	0/3	0/2	0/3
ORTHOMCL2047	m.42968, m.42982	1/4	0/5	0/3	0/2	0/3
ORTHOMCL2071	m.85009	1/4	0/5	0/3	0/2	0/3
ORTHOMCL2277	m.96279	1/4	0/5	0/3	0/2	0/3
ORTHOMCL2278	m.90263	1/4	0/5	0/3	0/2	0/3
ORTHOMCL2279	m.75581	1/4	0/5	0/3	0/2	0/3
ORTHOMCL2280	m.72955	1/4	0/5	0/3	0/2	0/3
ORTHOMCL2281	m.72954	1/4	0/5	0/3	0/2	0/3

Table A3.6 continued

ORTHOMCL2282	m.47460	1/4	0/5	0/3	0/2	0/3
ORTHOMCL2315	m.60661	1/4	0/5	0/3	0/2	0/3
ORTHOMCL2641	m.74779	1/4	0/5	0/3	0/2	0/3
ORTHOMCL2642	m.71562	1/4	0/5	0/3	0/2	0/3
ORTHOMCL2643	m.42020	1/4	0/5	0/3	0/2	0/3
ORTHOMCL3045	m.71815	1/4	0/5	0/3	0/2	0/3
ORTHOMCL3046	m.68972, m.69005	1/4	0/5	0/3	0/2	0/3
ORTHOMCL3605	m.58692, m.58736, m.58799	1/4	0/5	0/3	0/2	0/3
ORTHOMCL4222	m.85153	1/4	0/5	0/3	0/2	0/3
ORTHOMCL4223	m.84569	1/4	0/5	0/3	0/2	0/3
ORTHOMCL4224	m.68903, m.68915	1/4	0/5	0/3	0/2	0/3
ORTHOMCL4225	m.60132	1/4	0/5	0/3	0/2	0/3
ORTHOMCL4226	m.26116, m.26126	1/4	0/5	0/3	0/2	0/3
ORTHOMCL4227	m.24858	1/4	0/5	0/3	0/2	0/3
ORTHOMCL4228	m.17090	1/4	0/5	0/3	0/2	0/3
ORTHOMCL4721	m.99198	1/4	0/5	0/3	0/2	0/3
ORTHOMCL4722	m.98372	1/4	0/5	0/3	0/2	0/3
ORTHOMCL4723	m.87300, m.87317, m.87318, m.87321	1/4	0/5	0/3	0/2	0/3
ORTHOMCL4724	m.85966	1/4	0/5	0/3	0/2	0/3
ORTHOMCL4725	m.71848	1/4	0/5	0/3	0/2	0/3
ORTHOMCL4726	m.60662	1/4	0/5	0/3	0/2	0/3
ORTHOMCL4727	m.59149	1/4	0/5	0/3	0/2	0/3
ORTHOMCL4728	m.29831	1/4	0/5	0/3	0/2	0/3
ORTHOMCL4729	m.27549	1/4	0/5	0/3	0/2	0/3
ORTHOMCL4730	m.18642	1/4	0/5	0/3	0/2	0/3
ORTHOMCL5173	m.85466	1/4	0/5	0/3	0/2	0/3
ORTHOMCL5174	m.74796	1/4	0/5	0/3	0/2	0/3
ORTHOMCL5175	m.73467	1/4	0/5	0/3	0/2	0/3

Table A3.6 continued

ORTHOMCL5176	m.65300	1/4	0/5	0/3	0/2	0/3
ORTHOMCL5177	m.37361	1/4	0/5	0/3	0/2	0/3
ORTHOMCL5178	m.32629	1/4	0/5	0/3	0/2	0/3
ORTHOMCL5548	m.85234	1/4	0/5	0/3	0/2	0/3
ORTHOMCL5549	m.37153	1/4	0/5	0/3	0/2	0/3
ORTHOMCL5550	m.33824, m.33825	1/4	0/5	0/3	0/2	0/3
ORTHOMCL5551	m.31156	1/4	0/5	0/3	0/2	0/3
ORTHOMCL5552	m.24445, m.24463	1/4	0/5	0/3	0/2	0/3
ORTHOMCL5553	m.17251, m.17260	1/4	0/5	0/3	0/2	0/3
ORTHOMCL5554	m.128917, m.128935, m.129235, m.129244	1/4	0/5	0/3	0/2	0/3
ORTHOMCL5941	m.82248	1/4	0/5	0/3	0/2	0/3
ORTHOMCL5942	m.47205	1/4	0/5	0/3	0/2	0/3
ORTHOMCL5943	m.42876	1/4	0/5	0/3	0/2	0/3
ORTHOMCL5944	m.40996	1/4	0/5	0/3	0/2	0/3
ORTHOMCL5945	m.25845	1/4	0/5	0/3	0/2	0/3
ORTHOMCL5946	m.24194	1/4	0/5	0/3	0/2	0/3
ORTHOMCL5947	m.22303	1/4	0/5	0/3	0/2	0/3
ORTHOMCL5948	m.19088	1/4	0/5	0/3	0/2	0/3
ORTHOMCL5949	m.18982	1/4	0/5	0/3	0/2	0/3
ORTHOMCL5950	m.17188	1/4	0/5	0/3	0/2	0/3
ORTHOMCL6509	m.78355, m.78357, m.78359	1/4	0/5	0/3	0/2	0/3
ORTHOMCL6510	m.49489	1/4	0/5	0/3	0/2	0/3
ORTHOMCL6511	m.44253	1/4	0/5	0/3	0/2	0/3
ORTHOMCL6512	m.38288	1/4	0/5	0/3	0/2	0/3
ORTHOMCL6513	m.30119, m.30135	1/4	0/5	0/3	0/2	0/3
ORTHOMCL6514	m.25689	1/4	0/5	0/3	0/2	0/3
ORTHOMCL6515	m.14870, m.14872, m.14877, m.14880	1/4	0/5	0/3	0/2	0/3

Table A3.6 continued

ORTHOMCL7039	m.90453, m.90557	1/4	0/5	0/3	0/2	0/3
ORTHOMCL7040	m.87299, m.87316, m.87320	1/4	0/5	0/3	0/2	0/3
ORTHOMCL7041	m.85112, m.85125, m.85155	1/4	0/5	0/3	0/2	0/3
ORTHOMCL7042	m.52203	1/4	0/5	0/3	0/2	0/3
ORTHOMCL7043	m.46289	1/4	0/5	0/3	0/2	0/3
ORTHOMCL7044	m.44220, m.44251	1/4	0/5	0/3	0/2	0/3
ORTHOMCL7045	m.40237	1/4	0/5	0/3	0/2	0/3
ORTHOMCL7046	m.39722	1/4	0/5	0/3	0/2	0/3
ORTHOMCL7047	m.28136	1/4	0/5	0/3	0/2	0/3
ORTHOMCL7048	m.23041, m.23049, m.23060	1/4	0/5	0/3	0/2	0/3
ORTHOMCL7049	m.119691	1/4	0/5	0/3	0/2	0/3
ORTHOMCL7235	m.45804	1/4	0/5	0/3	0/2	0/3
ORTHOMCL7819	m.86291	1/4	0/5	0/3	0/2	0/3
ORTHOMCL7820	m.84077	1/4	0/5	0/3	0/2	0/3
ORTHOMCL7821	m.74594	1/4	0/5	0/3	0/2	0/3
ORTHOMCL7822	m.62478	1/4	0/5	0/3	0/2	0/3
ORTHOMCL7823	m.45270, m.45279	1/4	0/5	0/3	0/2	0/3
ORTHOMCL7824	m.40046	1/4	0/5	0/3	0/2	0/3
ORTHOMCL7825	m.30256, m.30259	1/4	0/5	0/3	0/2	0/3
ORTHOMCL7826	m.17250, m.17259	1/4	0/5	0/3	0/2	0/3
ORTHOMCL8115	m.76670	1/4	0/5	0/3	0/2	0/3
ORTHOMCL8138	m.49708	1/4	0/5	0/3	0/2	0/3
ORTHOMCL8272	m.122985	1/4	0/5	0/3	0/2	0/3
ORTHOMCL8703	m.91420, m.91421	1/4	0/5	0/3	0/2	0/3
ORTHOMCL8704	m.87319	1/4	0/5	0/3	0/2	0/3
ORTHOMCL8705	m.86400	1/4	0/5	0/3	0/2	0/3
ORTHOMCL8706	m.86391, m.86408	1/4	0/5	0/3	0/2	0/3
ORTHOMCL8707	m.83553	1/4	0/5	0/3	0/2	0/3
ORTHOMCL8708	m.79076	1/4	0/5	0/3	0/2	0/3

Table A3.6 continued

ORTHOMCL8709	m.78730, m.78731	1/4	0/5	0/3	0/2	0/3
ORTHOMCL8710	m.76425	1/4	0/5	0/3	0/2	0/3
ORTHOMCL8711	m.68383, m.68393	1/4	0/5	0/3	0/2	0/3
ORTHOMCL8712	m.65478	1/4	0/5	0/3	0/2	0/3
ORTHOMCL8713	m.62096, m.62104	1/4	0/5	0/3	0/2	0/3
ORTHOMCL8714	m.61034, m.61035	1/4	0/5	0/3	0/2	0/3
ORTHOMCL8715	m.60012	1/4	0/5	0/3	0/2	0/3
ORTHOMCL8716	m.53785	1/4	0/5	0/3	0/2	0/3
ORTHOMCL8717	m.53461	1/4	0/5	0/3	0/2	0/3
ORTHOMCL8718	m.49707	1/4	0/5	0/3	0/2	0/3
ORTHOMCL8719	m.44935	1/4	0/5	0/3	0/2	0/3
ORTHOMCL8720	m.39723	1/4	0/5	0/3	0/2	0/3
ORTHOMCL8721	m.32928	1/4	0/5	0/3	0/2	0/3
ORTHOMCL8722	m.32927	1/4	0/5	0/3	0/2	0/3
ORTHOMCL8723	m.31158	1/4	0/5	0/3	0/2	0/3
ORTHOMCL8724	m.27533	1/4	0/5	0/3	0/2	0/3
ORTHOMCL8725	m.25520	1/4	0/5	0/3	0/2	0/3
ORTHOMCL8726	m.17249	1/4	0/5	0/3	0/2	0/3
ORTHOMCL8727	m.17214	1/4	0/5	0/3	0/2	0/3
ORTHOMCL8728	m.16079, m.16080	1/4	0/5	0/3	0/2	0/3
ORTHOMCL8729	m.15124	1/4	0/5	0/3	0/2	0/3
ORTHOMCL8730	m.14989	1/4	0/5	0/3	0/2	0/3
ORTHOMCL8731	m.13200	1/4	0/5	0/3	0/2	0/3
ORTHOMCL8732	m.122987	1/4	0/5	0/3	0/2	0/3
ORTHOMCL8733	m.12096	1/4	0/5	0/3	0/2	0/3
ORTHOMCL9171	m.14990	1/4	0/5	0/3	0/2	0/3
ORTHOMCL9222	m.83484	1/4	0/5	0/3	0/2	0/3
ORTHOMCL10199	m.90128	1/4	0/5	0/3	0/2	0/3
ORTHOMCL10200	m.81515	1/4	0/5	0/3	0/2	0/3

Table A3.6 continued

ORTHOMCL10201	m.65939	1/4	0/5	0/3	0/2	0/3
ORTHOMCL10202	m.44873	1/4	0/5	0/3	0/2	0/3
ORTHOMCL10203	m.35627, m.35633	1/4	0/5	0/3	0/2	0/3
ORTHOMCL10204	m.26802	1/4	0/5	0/3	0/2	0/3
ORTHOMCL10205	m.25286	1/4	0/5	0/3	0/2	0/3
ORTHOMCL10206	m.23048, m.23056	1/4	0/5	0/3	0/2	0/3
ORTHOMCL10207	m.22749	1/4	0/5	0/3	0/2	0/3
ORTHOMCL10208	m.22147	1/4	0/5	0/3	0/2	0/3
ORTHOMCL10209	m.20407	1/4	0/5	0/3	0/2	0/3
ORTHOMCL10210	m.16083	1/4	0/5	0/3	0/2	0/3
ORTHOMCL10211	m.100545	1/4	0/5	0/3	0/2	0/3
ORTHOMCL11997	m.9274	1/4	0/5	0/3	0/2	0/3
ORTHOMCL11998	m.7318	1/4	0/5	0/3	0/2	0/3
ORTHOMCL11999	m.72262	1/4	0/5	0/3	0/2	0/3
ORTHOMCL12000	m.68708	1/4	0/5	0/3	0/2	0/3
ORTHOMCL12001	m.65455	1/4	0/5	0/3	0/2	0/3
ORTHOMCL12002	m.62490	1/4	0/5	0/3	0/2	0/3
ORTHOMCL12003	m.5948	1/4	0/5	0/3	0/2	0/3
ORTHOMCL12004	m.55739	1/4	0/5	0/3	0/2	0/3
ORTHOMCL12005	m.55737	1/4	0/5	0/3	0/2	0/3
ORTHOMCL12006	m.52345	1/4	0/5	0/3	0/2	0/3
ORTHOMCL12007	m.50895	1/4	0/5	0/3	0/2	0/3
ORTHOMCL12008	m.43137	1/4	0/5	0/3	0/2	0/3
ORTHOMCL12009	m.31115	1/4	0/5	0/3	0/2	0/3
ORTHOMCL12010	m.30518	1/4	0/5	0/3	0/2	0/3
ORTHOMCL12011	m.29017	1/4	0/5	0/3	0/2	0/3
ORTHOMCL12012	m.26520	1/4	0/5	0/3	0/2	0/3
ORTHOMCL12013	m.25439	1/4	0/5	0/3	0/2	0/3
ORTHOMCL12014	m.25421	1/4	0/5	0/3	0/2	0/3

Table A3.6 continued

ORTHOMCL12015	m.24443	1/4	0/5	0/3	0/2	0/3
ORTHOMCL12016	m.24234	1/4	0/5	0/3	0/2	0/3
ORTHOMCL12017	m.20379	1/4	0/5	0/3	0/2	0/3
ORTHOMCL12018	m.19865	1/4	0/5	0/3	0/2	0/3
ORTHOMCL12019	m.19661	1/4	0/5	0/3	0/2	0/3
ORTHOMCL12020	m.17547	1/4	0/5	0/3	0/2	0/3
ORTHOMCL12021	m.13396	1/4	0/5	0/3	0/2	0/3
ORTHOMCL12022	m.12463	1/4	0/5	0/3	0/2	0/3
ORTHOMCL15485	m.98443	1/4	0/5	0/3	0/2	0/3
ORTHOMCL15486	m.94227	1/4	0/5	0/3	0/2	0/3
ORTHOMCL15487	m.94091	1/4	0/5	0/3	0/2	0/3
ORTHOMCL15488	m.9314	1/4	0/5	0/3	0/2	0/3
ORTHOMCL15489	m.85623	1/4	0/5	0/3	0/2	0/3
ORTHOMCL15490	m.80775	1/4	0/5	0/3	0/2	0/3
ORTHOMCL15491	m.78376	1/4	0/5	0/3	0/2	0/3
ORTHOMCL15492	m.78354	1/4	0/5	0/3	0/2	0/3
ORTHOMCL15493	m.7610	1/4	0/5	0/3	0/2	0/3
ORTHOMCL15494	m.75795	1/4	0/5	0/3	0/2	0/3
ORTHOMCL15495	m.71847	1/4	0/5	0/3	0/2	0/3
ORTHOMCL15496	m.69302	1/4	0/5	0/3	0/2	0/3
ORTHOMCL15497	m.6864	1/4	0/5	0/3	0/2	0/3
ORTHOMCL15498	m.6778	1/4	0/5	0/3	0/2	0/3
ORTHOMCL15499	m.57085	1/4	0/5	0/3	0/2	0/3
ORTHOMCL15500	m.54821	1/4	0/5	0/3	0/2	0/3
ORTHOMCL15501	m.5407	1/4	0/5	0/3	0/2	0/3
ORTHOMCL15502	m.52344	1/4	0/5	0/3	0/2	0/3
ORTHOMCL15503	m.52204	1/4	0/5	0/3	0/2	0/3
ORTHOMCL15504	m.4896	1/4	0/5	0/3	0/2	0/3
ORTHOMCL15505	m.4452	1/4	0/5	0/3	0/2	0/3

Table A3.6 continued

ORTHOMCL15506	m.4451	1/4	0/5	0/3	0/2	0/3
ORTHOMCL15507	m.42077	1/4	0/5	0/3	0/2	0/3
ORTHOMCL15508	m.41272	1/4	0/5	0/3	0/2	0/3
ORTHOMCL15509	m.40404	1/4	0/5	0/3	0/2	0/3
ORTHOMCL15510	m.3806	1/4	0/5	0/3	0/2	0/3
ORTHOMCL15511	m.34429	1/4	0/5	0/3	0/2	0/3
ORTHOMCL15512	m.3181	1/4	0/5	0/3	0/2	0/3
ORTHOMCL15513	m.3126	1/4	0/5	0/3	0/2	0/3
ORTHOMCL15514	m.3072	1/4	0/5	0/3	0/2	0/3
ORTHOMCL15515	m.26542	1/4	0/5	0/3	0/2	0/3
ORTHOMCL15516	m.26000	1/4	0/5	0/3	0/2	0/3
ORTHOMCL15517	m.25424	1/4	0/5	0/3	0/2	0/3
ORTHOMCL15518	m.24992	1/4	0/5	0/3	0/2	0/3
ORTHOMCL15519	m.2483	1/4	0/5	0/3	0/2	0/3
ORTHOMCL15520	m.24227	1/4	0/5	0/3	0/2	0/3
ORTHOMCL15521	m.2092	1/4	0/5	0/3	0/2	0/3
ORTHOMCL15522	m.20408	1/4	0/5	0/3	0/2	0/3
ORTHOMCL15523	m.20403	1/4	0/5	0/3	0/2	0/3
ORTHOMCL15524	m.18464	1/4	0/5	0/3	0/2	0/3
ORTHOMCL15525	m.17797	1/4	0/5	0/3	0/2	0/3
ORTHOMCL15526	m.17733	1/4	0/5	0/3	0/2	0/3
ORTHOMCL15527	m.16422	1/4	0/5	0/3	0/2	0/3
ORTHOMCL15528	m.15869	1/4	0/5	0/3	0/2	0/3
ORTHOMCL15529	m.14928	1/4	0/5	0/3	0/2	0/3
ORTHOMCL15530	m.14881	1/4	0/5	0/3	0/2	0/3
ORTHOMCL15531	m.14873	1/4	0/5	0/3	0/2	0/3
ORTHOMCL15532	m.144610	1/4	0/5	0/3	0/2	0/3
ORTHOMCL15533	m.144283	1/4	0/5	0/3	0/2	0/3
ORTHOMCL15534	m.143990	1/4	0/5	0/3	0/2	0/3

Table A3.6 continued

ORTHOMCL15535	m.143960	1/4	0/5	0/3	0/2	0/3
ORTHOMCL15536	m.143421	1/4	0/5	0/3	0/2	0/3
ORTHOMCL15537	m.143356	1/4	0/5	0/3	0/2	0/3
ORTHOMCL15538	m.142459	1/4	0/5	0/3	0/2	0/3
ORTHOMCL15539	m.13348	1/4	0/5	0/3	0/2	0/3
ORTHOMCL15540	m.12969	1/4	0/5	0/3	0/2	0/3
ORTHOMCL15541	m.12560	1/4	0/5	0/3	0/2	0/3
ORTHOMCL15542	m.10603	1/4	0/5	0/3	0/2	0/3
ORTHOMCL15842	m.28137	1/4	0/5	0/3	0/2	0/3