Title

Investigations into the mycoheterotrophic symbiosis between Rhizopogon salebrosus and Pterospora andromedea and development of bioinformatic tools related to non-assembled fungal genomes

Author

Wong, Valerie

Publication Date

2011

Peer reviewed|Thesis/dissertation

Investigations into the mycoheterotrophic symbiosis between *Rhizopogon salebrosus* and *Pterospora andromedea* and development of bioinformatic tools related to nonassembled fungal genomes

By

Valerie L. Wong

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Plant Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Thomas D. Bruns, Chair
Professor N. Louise Glass
Professor Sandrine Dudoit

Spring 2011

Investigations into the mycoheterotrophic symbiosis between *Rhizopogon salebrosus* and *Pterospora andromedea* and development of bioinformatic tools related to nonassembled fungal genomes

By Valerie L. Wong

Abstract

Investigations into the mycoheterotrophic symbiosis between *Rhizopogon salebrosus* and *Pterospora andromedea* and development of bioinformatic tools related to non-assembled fungal genomes

by

Valerie L. Wong

Doctor of Philosophy in Plant Biology

University of California, Berkeley

Professor Thomas D. Bruns, Chair

This dissertation examines fungal genetics in two systems: *Penicillium chrysogenum* and *Rhizopogon salebrosus*. *Penicillium chrysogenum* holds historical and medical significance for the discovery and development of penicillin as an antibiotic. *Rhizopogon salebrosus* is an ecologically important mycorrhizal fungus that forms symbioses with both green plants and non-photosynthetic mycoheterotrophic plants. Mycoheterotrophs, plants that parasitize mycorrhizal fungi and cheat the classical mycorrhizal symbiosis, have been integral to the discovery and continuing study of mycorrhizal fungi.

The first chapter describes the utilization of assembly-free synteny analysis to identify rearrangements in wild *P. chrysogenum* strains. While traditional methods of mutagenesis and selection have been effective in improving production of compounds to commercial scale, the genetic changes behind the altered phenotypes have remained largely unclear. We utilized high-throughput Illumina short read sequencing of a wild *Penicillium chrysogenum* strain in order to make whole genome comparisons to a sequenced improved strain (WIS 54-1255). We developed an assembly-free method of identifying rearrangements and validated the *in silico* predictions with a PCR-based assay followed by Sanger sequencing. In addition to finding a previously published inversion in the penicillin biosynthesis cluster, we located several genes related to penicillin production associated with these rearrangements. By comparing the configuration of rearrangement events among several historically important strains known to be good penicillin producers to a collection of recently isolated wild strains, we suggest that wild strains with rearrangements similar to those in known good penicillin producers may be viable candidates for further improvement efforts.

The second chapter describes use of the phytohormone gibberellic acid (GA), a phytohormone which induces germination in many plants, to induce asymbiotic germination of two obligate mycoheterotrophic plants. These plants had previously only been seen to germinate in the presence of a symbiotic fungus, and such limitations prevented further study with these plants. We found that GA stimulates *Pterospora andromedea* germination at frequencies that can be far higher than that achieved with its fungal symbiont, *R. salebrosus*, alone. With constant exposure to 0.5M GA, we found nearly 80% germination, as opposed to only 15% with *R. salebrosus*. Even a short exposure to GA is sufficient for significantly enhanced germination. One day of exposure to 0.5mM GA induced ~55% stage 3 germination, and three days of exposure led to ~ 70% stage 3 germination. Exposing seeds to a lower dosage (0.1mM GA) for a longer time (2 weeks) produced a comparable germination rate. The closely related plant *Sarcodes sanguinea* was more resistant to GA stimulation. Exposure to GA is the only known method of inducing monotrope germination without the presence of a specific fungal symbiont. GA stimulation achieved workable germination rates for further studies, and GA may also be used as an improved assay for seed viability.

In the third chapter we utilize GA germination of *P. andromedea* seeds to examine transcriptional differences in *R. salebrosus* with partner and non-partner mycohterotroph seeds. The monotropes constitute a group of obligately dependent, non-photosynthetic mycoehterotrophs with extreme host specificity. Here we generate a *de novo* transcriptome assembly and examine differential expression when *R. salebrosus* interacts with seeds from partner (*R. salebrosus* type) and non-partner (*R. arctostaphyli* type) *P. andromedea*. Because *P. andromedea* seeds are very small, with few nutritional resources, germination in the presence of a suitable host fungus is a critical step in the establishment of symbiosis. We have identified putative genes in categories thought to be important for ectomycorrhizal symbiosis. These include stress response, host defense, cellular transport, secretion, carbohydrate metabolism and transport, and signal transduction. To our knowledge, this is the first study of gene expression with any mycorrhizal fungus involved with a mycoheterotrophic plant.

To my family

Who have been quietly behind all my academic endeavors

And

To the friends who are family

Acknowledgements

I would like to thank Tom Bruns and the rest of my committee members, Sandrine Dudoit and Louise Glass. Tom introduced me to the world of fungi and has been my guide through the underground movement.

I am indebted to the support of my cohort in graduate school as well as the inimitable members of the BTC. Jennifer Kerekes, Nhu Nguyen, and Tim Szaro have been great friends as well as field help. Chris Villalta has been a stalwart friend, and with Emily Whiston has made the short bus a less lonely ride. Chris Ellison has been a brilliant collaborator and an incredibly generous colleague. And without the insights of Shannon Schechter, the other half of the balcony, this would have been a dull tour of duty indeed.

This endeavor would not have been possible without the support of my family. My father taught me appreciation for all things created by humans and nature. My mother exemplifies the persistence and diligence required to accomplish anything worthwhile. Perhaps most importantly, they never told me that I ought to become a real doctor. My brother, Tim, has always been there to bring perspective. And finally, thank you to RZF, CQD, and MEF for being my family out here on the edge of the continent.

TABLE OF CONTENTS

INTRODUCTION

The course of this dissertation has navigated shifting technology in order to examine the genetics of *Rhizopogon salebrosus*, a fungus for which there were no tools for studying gene expression. Part of that process required training in genomics and bioinformatics. The first chapter presented here describes a project that grew from work begun in MCB 247, and the subsequent chapters are the result of applying these skills to a new system.

Penicillium chrysogenum

Modern life has been enabled by the fortuitous identification of organisms from the wild that produce compounds of interest. Traditional mutation, breeding, and artificial selection have been remarkably successful for improvement of wild type germplasm with regard to commercial production of important compounds (Parekh, Vinci, & Strobel, 2000). While the success of strain improvement is clear, historically the specific genetic changes leading to the enhancements have remained unknown.

The discovery of penicillin and development of its antibiotic properties by Fleming, Chain and Florey was a landmark in medicine and pharmacology (Chain et al., 1940; Fleming, 1929). However, Fleming's original strain produced only relatively small quantities of penicillin. The efforts to make antibiotics more available, particularly in response to great demand during World War II, entailed both a search for wild strains with enhanced production of penicillin and improvement of strains already in culture (Elander, 1967). Notably, Raper, Alexander, and Coghill (1944) cultivated and tested isolates from a variety of food products, spoiled produce, and soils. Nearly all of their *Penicillium* strains produced detectable levels of penicillin, but very few were comparable to the best industrially important strain of the time.

Many commercial strains used by pharmaceutical companies such as Lilly Industries and Wyeth Lab trace their ancestry back to a single wild strain (*P. chrysogenum* NRRL 1951) isolated from a moldy cantaloupe found in Peoria, Illinois (Demain & Elander, 1999; Elander, 1967; Muñiz, Zelaya, & Esquivel, 2007; Raper & Alexander, 1945). Despite decades of work on strain improvement, most of which involved multiple rounds of non-directed mutagenesis and selection, little is known about the indirect regulation of penicillin biosynthesis and how improvement occurred in this "Wisconsin family" of strains. We utilized an assembly-free method of identifying rearrangements to locate genes associated with rearrangements and potentially involved in penicillin production. This method may be utilized to identify wild strains with potential for improvement.

Mycorrhizal Symbiosis and Mycoheterotrophic Plants

Mycorrhizal symbioses between plants and fungi are nearly ubiquitous and typically required for competitive survival in the field. The conventional mycorrhizal symbiosis, where a plant photosynthesizes and trades a portion of its fixed carbon for water and mineral nutrients collected by the soil fungus, is mutually beneficial. However, obligate mycoheterotrophs are non-photosynthetic cheaters that form mycorrhizae with the fungus but do not provide nutrients in return. Mycoheterotrophs have been historically important to the discovery of mycorrhizae by A.B. Frank (Frank, 2004). Mycoheterotrophs are not only parasites on fungi but also indirect "epiparasites" on other plants (Bjorkman, 1960), potentially affecting species interactions and ecosystem processes (van der Heijden et al., 1998; van der Heijden et al., 2006; Robinson, 1999; Simard et al., 1997).

Mycoheterotrophs represent an extreme in the mutualism-parasitism continuum of plant-fungal interactions. Unlike green plants, mycoheterotrophs are connected to only one fungus, making this an excellent simplified model for studying the network of interactions between plants and fungi in the field. The ecologically important mycorrhizal fungus *Rhizopogon salebrosus* and the mycoheterotroph *Pterospora andromedea* (pine drops, Monotropoideae) together comprise a good model system for mycoheterotroph interactions.

Monotropes rely entirely on specific fungal partners for their entire life cycle, from seed germination onward. Thus, these plants require unique and highly specific conditions. Seed germination and seedling establishment are critical steps in mycoheterotroph development, because growth requires specific fungal species. Several closely related *Rhizopogon* species induce germination of *P. andromedea* seeds by a diffusible substance (Bruns & Read, 2000), but only the plants associated with the "correct" fungus continue to develop (Bidartondo & Bruns, 2005). Thus, after germination, plants with the "wrong" fungus either switch associates or die. It is known that transcriptional shifts occur in mycorrhizal fungi during host recognition and establishment of the ectomycorrhizal interaction (Duplessis, Courty, Tagu, & Martin, 2005; Duplessis, Encelot, Martin, & Voiblet, 2001; Johansson et al., 2004; Le Quéré et al., 2004; Martin et al., 2008). We describe an asymbiotic method of *P. andromedea* seed germination, and we explore gene expression in the initial stages of interaction between *R. salebrosus* and seeds from partner and non-partner plants.

LITERATURE CITED

Bidartondo, M., & Bruns, T. (2005). On the origins of extreme mycorrhizal specificity in the Monotropoideae (Ericaceae): performance trade-offs during seed germination and seedling development. *Molecular ecology*, *14*(5), 1549-60.

Bjorkman, E. (1960). Monotropa Hypopitys L. - an Epiparasite on Tree Roots. *Physiologia Plantarum*, *13*(2), 308-327.

Bruns, T. D., & Read, D. J. (2000). In vitro germination of nonphotosynthetic, myco-heterotrophic plants stimulated by fungi isolated from the adult plants. *New Phytologist*, *148*(2), 335-342.

Chain, E., Florey, H., Gardner, A., Heatley, N., Jennings, M., Orr-Ewing, J., et al. (1940). Symptoms relating. *The Lancet*, 226-228.

Demain, a L., & Elander, R P. (1999). The beta-lactam antibiotics: past, present, and future. *Antonie van Leeuwenhoek*, *75*(1-2), 5-19.

Duplessis, S., Courty, P.E., Tagu, D., & Martin, Francis. (2005). Transcript patterns associated with ectomycorrhiza development in Eucalyptus globulus and Pisolithus microcarpus. *The New phytologist*, *165*(2), 599-611.

Duplessis, Â., Encelot, N., Martin, Francis, & Voiblet, C. (2001). Identification of symbiosis-regulated genes in Eucalyptus globulus ± Pisolithus tinctorius ectomycorrhiza by differential hybridization of arrayed cDNAs. *Science*, *25*.

Elander, R. P. (1967). Enhanced Penicillin Biosyntheis in Mutant and Recombinant Strains of Penicillium chrysogenum. *Induzierte Mutationen un ihre Nutzung. Induced mutations and their utilization*, *2*, 403-423.

Fleming, A. (1929). On the antibacterial action of Penicillium, with special reference to their use in the isolation of B. influenzae. *British Journal of Experimental Pathology*, *10*, 185-194.

Frank, B. (2004). On the nutritional dependence of certain trees on root symbiosis with belowground fungi (an English translation of A . B . Frank s classic paper of 1885) Structure of the mycorrhiza Development of the mycorrhiza. *Mycorrhiza*, (May 1880), 1-9.

Heijden, M. G. A. van der, Klironomos, J. N., Ursic, M., Moutoglis, P., Streitwolf-engel, R., Boller, T., et al. (1998). Mycorrhizal fungal diversity determines plant biodiversity, ecosystem variability and productivity. *Nature*, *74*(1994), 69-72.

Heijden, M. G. a van der, Streitwolf-Engel, R., Riedl, R., Siegrist, S., Neudecker, A., Ineichen, K., et al. (2006). The mycorrhizal contribution to plant productivity, plant nutrition and soil structure in experimental grassland. *The New phytologist*, *172*(4), 739-52.

Johansson, T., Le Quéré, A., Ahren, D., Söderström, B., Erlandsson, R., Lundeberg, J., et al. (2004). Transcriptional responses of Paxillus involutus and Betula pendula during formation of ectomycorrhizal root tissue. *Molecular plant-microbe interactions : MPMI*, *17*(2), 202-15.

Le Quéré, A., Schützendübel, A., Rajashekar, B., Canbäck, B., Hedh, J., Erland, S., et al. (2004). Divergence in gene expression related to variation in host specificity of an ectomycorrhizal fungus. *Molecular ecology*, *13*(12), 3809-19.

Martin, F, Aerts, a, Ahrén, D., Brun, a, Danchin, E. G. J., Duchaussoy, F., et al. (2008). The genome of Laccaria bicolor provides insights into mycorrhizal symbiosis. *Nature*, *452*(7183), 88-92.

Muñiz, C., Zelaya, T., & Esquivel, G. (2007). Penicillin and cephalosporin production: A historical perspective. *Rev Latinoam*, *49*(3-4), 88-98.

Parekh, S., Vinci, V. a, & Strobel, R. J. (2000). Improvement of microbial strains and fermentation processes. *Applied microbiology and biotechnology*, *54*(3), 287-301.

Raper, K. B., & Alexander, D. F. (1945). Penicillin: V. Mycological aspects of penicillin production. *Journal of the Mitchell Society*, (August).

Raper, K. B., Alexander, D. F., & Coghill, R. D. (1944). Penicillin: II. Natural variation and penicillin production in Penicillium notatum and allied species. *Journal of Bacteriology*, *48*, 639-659.

Robinson, D. (1999). The magnitude and control of carbon transfer between plants linked by a common mycorrhizal network. *Journal of Experimental Botany*, *50*(330), 9-13.

Simard, S. W., Perry, D. A., Jones, M. D., Myrold, D. D., Durall, D. M., & Molinak, R. (1997). Net transfer of carbon between ectomycorrhizal tree species in the field. *Nature*, *388*(August), 579-582.

CHAPTER 1

Utilization of Assembly-Free Synteny Analysis to Identify Rearrangements in Wild *Penicillium chrysogenum* strains

Abstract

Strain acquisition and improvement have been important tools for human bioproduct utilization. While traditional methods of mutagenesis and selection have been effective in improving production of compounds to commercial scale, the genetic changes behind the altered phenotypes have remained largely unclear. We utilized high-throughput Illumina short read sequencing of a wild *Penicillium chrysogenum* strain in order to make whole genome comparisons to a sequenced improved strain (WIS 54-1255). We developed an assembly-free method of identifying rearrangements and validated the *in silico* predictions with a PCR-based assay followed by Sanger sequencing. In addition to finding a previously published inversion in the penicillin biosynthesis cluster, we located several genes related to penicillin production associated with these rearrangements. By comparing the configuration of rearrangement events among several historically important strains known to be good penicillin producers to a collection of recently isolated wild strains, we suggest that wild strains with rearrangements similar to those in known good penicillin producers may be viable candidates for further improvement efforts.

Introduction

Isolation of wild type strains with desired bioactivity, followed by traditional mutation, breeding, and artificial selection has been remarkably successful for improvement of wild type germplasm with regard to commercial production of important compounds (Parekh, Vinci, & Strobel, 2000). While the success of strain improvement is clear, historically the specific genetic changes leading to the enhancements have remained unknown. With the recent availability of high-throughput DNA sequencing, it is now possible to compare whole genomes of wild and industrial strains in order to identify genomic differences that may be responsible for the improved phenotype (Le Crom et al., 2009).

The discovery of penicillin and its antibiotic properties begun by Alexander Fleming and developed by Chain and Florey was a landmark in medicine and pharmacology (Chain et al., 1940; Fleming, 1929). However, Fleming's original strain produced only relatively small quantities of penicillin. The efforts to make antibiotics more available, particularly in response to great demand during World War II, entailed both a search for wild strains with enhanced production of penicillin and improvement of strains already in culture (R. P. Elander, 1967). Notably, Raper, Alexander, and Coghill (1944) cultivated and tested isolates from a variety of food products, spoiled produce, and soils. Nearly all of their *Penicillium* strains produced detectable levels of penicillin, but very few were comparable to the best industrially important strain of the time (Raper, Alexander, & Coghill, 1944). The new isolates formed a bimodal distribution of penicillin production, indicative of natural variation in the wild population and suggesting that some wild strains may be predisposed to be good penicillin producers and to give rise to good industrial strains.

Many commercial strains used by pharmaceutical companies such as Lilly Industries and Wyeth Lab trace their ancestry back to a single wild strain (*P. chrysogenum* NRRL 1951) isolated from a moldy cantaloupe found in Peoria, Illinois (Demain & Elander, 1999; Elander, 1967; Muñiz, Zelaya, & Esquivel, 2007; Raper & Alexander, 1945). Despite decades of work on strain improvement, most of which involved multiple rounds of non-directed mutagenesis and selection, little is known about the indirect regulation of penicillin biosynthesis and how improvement occurred in this "Wisconsin family" of strains.

Penicillins are a subset of β-lactam antibiotics, and each penicillin is named and defined by its specific hydrophobic or hydrophilic acyl side chain. The *P. chrysogenum* core genes for penicillin biosynthesis (*pcbAB, pcbC*, and *penDE*) are clustered together among other ORFs in a 56.8kb region (Díez et al., 1990; Fierro et al., 2006). Tandem duplications of this cluster can be found in *P. chrysogenum* strains that are high penicillin producers (Fierro et al., 1995). These three core enzymes enable the synthesis of penicillins from L-α-aminoadipic acid, L-cysteine, and L-valine. In *P. chrysogenum*, α-aminoadipic acid is an intermediate in lysine biosynthesis and can be catabolized from lysine, which itself inhibits the rate limiting first step in synthesis, ACV synthetase (Demain & R P Elander, 1999; Juan F. Martín, Ullán, & García-Estrada, 2010). Other enzymes outside the core cluster are required to activate the first step in the pathway as well as to activate the side chains (Juan F. Martín, Ullán, & García-Estrada, 2010). The last two steps in the penicillin biosynthesis pathway, the IPN acyltransferase coded by *penDE* and activation of the side chain, take place in the microbody (peroxisome), and strains with more microbodies produce more penicillin (J. K. W. Kiel et al., 2009). Side chain incorporation can be supported by adding precursors such as penylacetic acid (PAA) to the culture medium, pushing synthesis towards penicillin G, one of two main commercial penicillins (Brakhage 1999).

The penicillin biosynthesis cluster appears devoid of regulators specific to penicillin production (van Den Berg et al., 2008; Fierro et al., 2006), and regulation of the process seems to be controlled by unspecific regulators such as heterochromatin modification, as well as nitrogen regulation and pH-dependent carbon source regulation (Demain & Elander, 1999; Martín, Ullán, & García-Estrada, 2010). Other methods of increasing penicillin production include modifying the growth conditions and reducing sporulation and growth, which occur at the expense of secondary metabolite production (Elander and Espenshade 1976).

The recent sequencing of one improved Wisconsin family strain (WIS 54-1255) has produced insights into the genetics of penicillin biosynthesis (van Den Berg et al., 2008), but this information alone is insufficient to elucidate the genomic changes between wild strains, improved strains, and wild strains with enhanced penicillin production. The WIS 54-1255 strain (hereafter referred to as WI) was produced via multiple rounds of selection and mutagenesis, including ultraviolet radiation, X rays, and nitrogen mustard, which may have led to rearrangements (Elander, 1967; Muñiz et al.,

2007; Raper & Alexander, 1945). There is a previously identified rearrangement in the biosynthesis cluster (Fierro et al., 1995; 2006), presumably induced by strain improvement efforts. Identification of further rearrangement events may elucidate genes influential in improved penicillin biosynthesis.

We utilized Illumina short read sequencing of a wild *P. chrysogenum* strain (PC0184C, hereafter referred to as UCB) to develop an assembly-free computational pipeline using only mate-pair information to identify chromosomal rearrangements between this wild strain and an improved strain (WIS 54-1255). We further validated our *in silico* predictions with a PCR-based assay and screened additional wild and industrially important *P. chrysogenum* strains (described further in Materials and Methods), thereby assessing which, if any, genomic changes are specific to the industrially important strains and thus potentially contribute improvements in penicillin biosynthesis.

Materials and Methods

Cultures and Strains

The following *P. chrysogenum* strains were obtained from the ATCC culture collection: 28089 (designation WIS 54-1255 and hereafter referred to as WI), 9480 (NRRL 1951), 9179 (NRRL 832), and 9478 (NRRL 824, originally deposited as *P. notatum*). NRRL 1951 is the founding member of the industrially important Wisconsin family of improved *P. chrysogenum* strains (Raper & Alexander, 1945). The WI strain is descended from NRRL 1951. ATCC 9179 was identified as a strain particularly suited for penicillin production in submerged culture, which is preferable to surface culture for industrial production (Moyer & Coghill, 1945). ATCC 9478 is Fleming's original strain (Fleming, 1929).

The following wild strains were graciously provided by Daniel Henk (Imperial College, London): PC0184C, PC08-1B, PC08-3A, PC08-5A, PC08-8B, PC08-20A, PC08-97A, A1723, A2416, A24700, A3704, 1975, 3704, 3905, 20320, 32232, 924700, and 0887Ax. PC0184C is the strain we sequenced and was isolated from a hotel room in France. For brevity we refer to this strain as UCB. All *P. chrysogenum* strains were maintained on MEA (20g/L malt extract, 1.5% agar) at room temperature.

UCB Strain (PC0184C) Sequencing

The UCB strain was sequenced with three lanes of paired-end Illumina (San Diego, CA) 36bp reads, following standard Illumina protocols, at the Vincent J. Coates Genomic Sequencing Laboratory (UC Berkeley, Berkeley, CA). The Illumina standard pre-processing pipeline was used. Insert sizes were 300bp and 500bp.

Read Mapping

We mapped three paired-end Illumina libraries from the *P. chrysogenum* UCB strain onto the published *P. chrysogenum* WI genome assembly (Elander, 1967) using the read-mapping program Bowtie (Langmead, Trapnell, Pop, & Salzberg, 2009). If

chromosomal rearrangements have occurred between the strains, we would expect to find mate pairs from the UCB strain mapping to locations in the WI genome assembly that are much further apart than the 300-500bp fragment insert sizes used to prepare the library (Fig. 1). To ensure that mate pairs with larger than expected mapping distances were not discarded during the alignment process, we disregarded mate pair information while mapping. We controlled for artifactual rearrangement signatures from repetitive regions by requiring Bowtie to discard all alignments for reads that map equally well to multiple places in the genome.

Aberrant Mate-Pairs

Although reads were not mapped as mate pairs, we were able to identify the location of mate pairs in the WI genome assembly post-alignment because mate-pair information is encoded into the read IDs. After pairing mates together, the distance between their genomic locations was calculated, and pairs that had locations in the WI genome assembly further than 1,000bp apart or that mapped to two different WI contigs were identified as potentially spanning rearrangement breakpoints in the WI genome (Fig. 1).

Support for Rearrangements from High Coverage

Because the UCB strain genome was sequenced at ~30X coverage, each rearrangement breakpoint in the WI genome should be spanned by ~30 different mate pairs from the UCB reads. These pairs (presumably located within 500bp of each other in the UCB genome) should map to locations in the WI genome at distances equal to the size of the rearrangement event (Figs. 2, 3). Reads from each side of the event should also cluster together and should have mates that cluster together on the opposite side of the event. Genomic locations where reads from independent aberrant mate pairs cluster together thus provide a unique signature to identify the boundaries of putative rearrangement events in the WI genome. We required read clusters or “blocs” to be composed of at least 20 reads from aberrant mate pairs located no more than 100bp from each other. We also identified “mate blocs” by pairing up blocs that had at least 90% of the reads mate paired together. A pair of mate blocs therefore represents the boundaries of a single putative rearrangement event. Mate blocs were further manually curated to filter out putative transposons.

Using Strand Information to Classify Rearrangements

Strandedness of aberrant mate pairs that map to the same contig was used to classify rearrangements as either inversions or insertions/deletions. An insertion event in the WI strain or a deletion in the UCB strain would result in mate pairs mapping to opposite strands, preserving the mate pair directions. An inversion event would flip the strand of one read in each mate pair, resulting in mate pairs mapping to the same strand. Events classified as insertions were required to be supported by at least 90% of aberrant mate pairs mapping to opposite strands, while inversion events were required to be supported by at least 90% of aberrant mate pairs mapping to the same strand.

PCR Validation

In order to validate the rearrangement events predicted *in silico*, we designed primers to amplify across blocs as arranged in the WI strain (Fig. 3). If the predicted rearrangement were correct, using the primers in the same orientation with DNA from the UCB strain, no product should be detected. However, if PCR product amplified when we switched the primer pairs to be either both forward or both reverse primers from a mate bloc pair, this was consistent with an inversion event. Insertions were detected by priming across the insertion, using a forward primer for one bloc in a mate bloc pair and the reverse primer from its mate. PCR products for the WI and UCB strains were further validated by sequencing at the UC Berkeley Sequencing Facility.

DNA Extraction

Prior to extraction, all *P. chrysogenum* cultures were grown in flasks containing 100ml liquid malt extract medium (20g/L malt extract) for two days at 25°C with gentle shaking. Tissue was harvested by filtering through Miracloth (Calbiochem, Darmstadt, Germany) and rinsing with ~ 100mL sterile distilled water. Samples were frozen in liquid nitrogen and stored at -80°C prior to lyophilization for 2 days. DNA was extracted by beadbeating with 0.3g zirconia/silica beads (BioSpec Products, Bartlesville, OK) in a screwtop tube for 30 sec. Following addition of 0.5mL lysis buffer (50mM Tris-HCl, 50mM EDTA, 3% SDS, and 1% ß-mercaptoethanol), tubes were vortexed to resuspend all ground tissue and then incubated at 65°C for 45 min. Chloroform (0.5mL) was added, and tubes were vortexed and then spun at 1,320 rpm for 5 min. The aqueous phase (~350 μl) was transferred to a new tube with 35 μl Proteinase K and 350 μl Buffer AL from the DNeasy Blood and Tissue kit (Qiagen, Valencia, CA), and manufacturer's directions were subsequently followed.

Primer Design and PCR

Primers were designed with the PrimerQuest^{SM} tool by Integrated DNA Technologies (IDT) to amplify across each bloc in the WI-54-1255 genome (Table 1). Various primer pairs were selected to identify blocs in the WI arrangement, blocs that were inverted relative to WI, and blocs with insertions relative to WI. Each 25 μl reaction was made according to the following recipe: 1μl of DNA extract, 2.5 μl of 2mM dNTPs (Fermentas, Glen Burnie, MD), 2.5 μl buffer (0.5M KCl, 0.1M TrisHCl pH 8.3, 25mM $MgCl_2$, and 1mg/mL gelatin), 0.25 μl of each primer (50 μM, IDT, San Diego, CA), 0.5 μl Taq DNA polymerase (New England Biolabs, Ipswitch, MA), and 18 μl water. Reactions were run with the following PCR program:

1. 94°C for 1 min
2. 94°C for 1 min
3. 68°C for 1 min
4. 72°C for 1.5 min
5. Repeat steps 2 through 4 (30 times).
6. 72°C for 8 min

Penicillium chrysogenum WI-54-1255 Genes Associated with Rearrangements

Genes of interest were identified as those whose start was either within a rearrangement event or less then 500 bp outside it. The very large rearrangements were ignored for this purpose due to the sheer number of genes involved.

Results and Discussion

We utilized mate pair information to identify rearrangements. The UCB genomic libraries were constructed from 300bp and 500bp fragment insert libraries. However, mapping each of these paired-end reads as two single-end reads to the WI genome assembly uncovered 18,641 aberrant mate pairs on the same contigs with distances over 1,000bp. The distances between aberrant mates that mapped to the same contig were not evenly distributed, with most mate pairs being either less than 1Mb apart or just over 4Mb apart (Fig. 4).

These aberrant mate pairs formed 329 blocs of at least 20 reads with start positions no more than 100bp from each other. These blocs were further narrowed down to 175 pairs of mate blocs by taking the median mate pair distance for reads within a bloc and searching for other blocs within 1kb of that distance. These mate paired blocs were manually curated to correct for instances where one bloc had multiple hits.

We identified 51 insertion/deletion events where at least 90% of reads within mate pair blocs have mates that map to the opposite strand (as expected from Illumina paired-end library construction). Twenty-one inversion events were identified where at least 80% of reads within mate pair blocs have mates that map to the same strand (Table 2). After manual curation to eliminate probable transposons, we identified 10 insertion/deletion events and 4 inversions (Table 2).

In order to validate this *in silico* method, we developed a PCR based assay to amplify across blocs as arranged in the WI strain (Fig. 3). With the exception of event 231, PCR amplification of all blocs in the WI strain showed strong bands in the expected sized range. Event 231 showed an accessory band for WI bloc 231A and no product for WI bloc 231B (Fig. 5). However, using UCB DNA and the UCB primer configuration a clear insertion/deletion was detected, as predicted *in silico*. All loci identified *in silico* were validated through this PCR method (Table 3). With the exception of the blocs for event 231 in the WI strain, all events were also verified by sequencing.

Of the rearrangement events examined, only one (event 326) differed between the WI strain and its ancestor, NRRL 1951. Event 326 occurred in the penicillin biosynthesis cluster. This was first described as a “shift fragment” located at the right end border of the cluster, and it is bracketed by an internal set of the same hexanucleotide sequences that border the biosynthesis cluster itself (Fierro et al., 1995). The hexanucleotide sequences are part of the mechanism whereby the biosynthesis

cluster amplifies into tandem repeats. The 3.2kb shift fragment includes two ORFs: one has ankyrin repeat domains and may be involved in mediating interactions between proteins, and the other is a putative transmembrane sterol desaturase (Fierro et al., 2006).

If a rearrangement were found in the WI conformation in the known good penicillin producing strains but not all of the other wild strains, this would be consistent with the idea that the region around the rearrangement is involved in penicillin biosynthesis. We applied this primer set to the Fleming strain (NRRL 824) and NRRL 832 (identified as a good penicillin producer in submerged culture), both historically important wild strains from culture collections (Raper & Alexander, 1945), as well as two recently isolated wild strains (Henk PC08-3A and Henk A3704). As expected, the shift fragment inversion at event 326 was unique to the WI strain. The same ambiguous PCR pattern found for event 231 in the WI strain appeared in several others. With the exception of events 237 and 309, which showed the WI configuration for all strains except for UCB, the rearrangements were polymorphic across strains (Table 4). The WI configuration was found only in NRRL 1951 and WI for events 269, 270 312, and 318. If the NRRL 1951 strain were naturally predisposed to be a good penicillin producer, genes in these regions may be involved in regulation of β-lactam biosynthesis. Events 12, 17, and 287 showed the WI conformation in NRRL 1951 and NRRL 832, two strains that were used in surface culture, but not NRRL 824, the strain important in submerged culture. Genes associated with these events could help elucidate the manner by which the strains produce penicillin in different growth conditions.

Event 309 was associated with three predicted genes (Table 6). One (Pc13g11990) is a putative transcription factor. Pc13g11940 is predicted to be a glucan 1,4-alpha-glucosidase. Pc13g11930 was more highly expressed in the high penicillin G-producing strain DS17690 than in WI or the penicillin cluster deletion strain (DS50661) and only increased expression in response to PAA in DS17690 (van Den Berg et al., 2008; Harris et al., 2009). Pc13g11930 has strong similarity to acyl CoA dehydrogenase and was predicted *in silico* to localize to the microbody, which plays an important role in biosynthesis as the location of the last two enzymes of the core biosynthesis cluster.

Event 17 was an insertion/deletion of a gene with strong similarity to the sulfate transporter *sutB* (Pc12g01540) (Table 6). Transcript levels of *sutB* were elevated in high penicillin G-producing cultures compared to low- and non-penicillin G cultures (van Den Berg et al., 2008; Harris et al., 2009). Increased levels of transcripts involved in synthesizing sulfurous amino acids could signal elevated synthesis of amino acid β-lactam precursors. Transcription of *sutB* was the same in a strain without a functional penicillin biosynthesis cluster, regardless of the presence of PAA, as well as a penicillin G-producing strain grown with PAA. In the penicillin G-producing strain when grown without PAA, *sutB* levels were elevated.

Event 6 had the WI conformation in NRRL 1951 and NRRL 832 but also one wild strain. Event 6 was associated with several genes implicated in penicillin production (Table 6). Two transcripts (Pc20g13820 and Pc20g13860) were found at lower levels in the high penicillin G-producing strain DS17690 than WI, regardless of PAA (van Den Berg et al., 2008). Notably, Pc20g13880 had strong similarity to the *Aspergillus niger creA*, a well-known transcription factor involved in glucose repression and regulation of β-lactam biosynthesis, which showed decreasing amounts of mRNA as penicillin productivity increased (van Den Berg et al., 2008).

Possibly due to variation at priming sites, a single, distinct PCR product was not produced for every locus across a wider collection of recently cultivated strains. Events 6 and 312, both inversions, produced clean PCR product across many available strains and showed polymorphism across strains (Table 5). Event 6 appeared nearly equally in either the WI or UCB configuration. The WI configuration of event 312 was relatively rare, appearing in only two strains besides NRRL 1951 and WI. Given that genes associated with event 6 are implicated in penicillin production, wild strains with the WI configuration for this event may be good candidates for strain improvement.

Of the many wild strains that have been screened for penicillin production, NRRL 1951 was among a very few chosen for improvement. This strain founded a large collection of industrially successful strains and may have been predisposed to give rise to high penicillin producers. In identifying rearrangements between the recent wild collected UCB strain and the sequenced WI strain, we were able to screen for differences between NRRL 1951 and its descendant WI strain. Despite many rounds of mutagenesis and selection that occurred in the development of the WI strain from NRRL 1951, we did not find many new rearrangements in the WI strain. This could be due to our stringent *in silico* filtering of putative rearrangement events. We also compared rearrangements in a variety of wild strains. Genes associated with these rearrangement events, of which a number are linked with β-lactam biosynthesis, may be indicative of a strain's penicillin production capacity.

Acknowledgements

Chris Ellison contributed intellectual direction and computational expertise. Devin Scannell produced the genomic libraries for sequencing. Daniel Henk graciously provided strains. Rachel Brem, Michael Eisen, and Lior Pachter advised this project.

Literature Cited

Demain, a L., & Elander, R P. (1999). The beta-lactam antibiotics: past, present, and future. *Antonie van Leeuwenhoek*, *75*(1-2), 5-19.

Den Berg, M. A. van, Albang, R., Albermann, K., Badger, J. H., Daran, J.-M., Driessen, A. J. M., et al. (2008). Genome sequencing and analysis of the filamentous fungus Penicillium chrysogenum. *Nature Biotechnology*, *26*(10), 1161-8.

Díez, B., Gutiérrez, S., Barredo, J. L., Solingen, P. van, Voort, L. H. van der, & Martín, J F. (1990). The cluster of penicillin biosynthetic genes. Identification and characterization of the pcbAB gene encoding the alpha-aminoadipyl-cysteinyl-valine synthetase and linkage to the pcbC and penDE genes. *The Journal of biological chemistry*, *265*(27), 16358-65.

Elander, R. P. (1967). Enhanced Penicillin Biosyntheis in Mutant and Recombinant Strains of Penicillium chrysogenum. *Induzierte Mutationen un ihre Nutzung. Induced mutations and their utilization*, *2*, 403-423.

Fierro, F., Barredot, J. L., Dfezt, B., Gutierrez, S., Fernandez, F. J., & Martin, J. F. (1995). The penicillin gene cluster is amplified in tandem repeats linked by conserved hexanucleotide sequences. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(June), 6200-6204.

Fierro, F., García-Estrada, C., Castillo, N. I., Rodríguez, R., Velasco-Conde, T., & Martín, J.-F. (2006). Transcriptional and bioinformatic analysis of the 56.8 kb DNA region amplified in tandem repeats containing the penicillin gene cluster in Penicillium chrysogenum. *Fungal genetics and biology : FG & B*, *43*(9), 618-29.

Fleming, A. (1929). On the antibacterial action of Penicillium, with special reference to their use in the isolation of B. influenzae. *British Journal of Experimental Pathology*, *10*, 185-194.

Harris, D. M., Krogt, Z. a van der, Klaassen, P., Raamsdonk, L. M., Hage, S., Berg, M. a van den, et al. (2009). Exploring and dissecting genome-wide gene expression responses of Penicillium chrysogenum to phenylacetic acid consumption and penicillinG production. *BMC genomics*, *10*, 75.

Kiel, J. K. W., Den Berg, M. A. van, Fusetti, F., Poolman, B., Bovenberg, R. A. L., Veenhuis, M., et al. (2009). Matching the proteome to the genome: the microbody of penicillin-producing Penicillium chrysogenum cells. *Functional & integrative genomics*, *9*(2), 167-84.

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology*, *10*(3), R25.

Le Crom, S., Schackwitz, W., Pennacchio, L., Magnuson, J. K., Culley, D. E., Collett, J. R., et al. (2009). Tracking the roots of cellulase hyperproduction by the fungus Trichoderma reesei using massively parallel DNA sequencing. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(38), 16151-6.

Martín, Juan F., Ullán, R. V., & García-Estrada, C. (2010). Regulation and compartmentalization of β-lactam biosynthesis. *Microbial Biotechnology*, *3*(3), 285-299.

Moyer, A. J., & Coghill, R. D. (1945). Penicillin IX . The Laboratory scale production of penicillin in submerged cultures by Penicillium notatum Westling (NRRL 832). *Journal of Bacteriology*, *51*, 79-93.

Muñiz, C., Zelaya, T., & Esquivel, G. (2007). Penicillin and cephalosporin production: A historical perspective. *Rev Latinoam*, *49*(3-4), 88-98.

Parekh, S., Vinci, V. a, & Strobel, R. J. (2000). Improvement of microbial strains and fermentation processes. *Applied microbiology and biotechnology*, *54*(3), 287-301.

Raper, K. B., & Alexander, D. F. (1945). Penicillin: V. Mycological aspects of penicillin production. *Journal of the Mitchell Society*, (August).

Raper, K. B., Alexander, D. F., & Coghill, R. D. (1944). Penicillin: II. Natural variation and penicillin production in Penicillium notatum and allied species. *Journal of Bacteriology*, *48*, 639-659.

Primer Name	Sequence
Bloc_6A_F	CTCGCATGCATTGGGCTTATCGTT
Bloc_6A_R	AGTCTCCTCAACCAACACCAACCA
Bloc_6B_F	TGCCTACTTGTTGTGCCTAGGTGT
Bloc_6B_R	TTAACGCCTACCTCGCTGCTGTTA
Bloc_12A_F	TGCAGAGGGTGATGCAAGACTACA
Bloc_12A_R	AGATCCTCATATTTCGCACGCTCG
Bloc_12B_F	AAGCCAGGACATAGGAGAGAGACA
Bloc_12B_R	GCACAAGTCTTACATCACTGGGCT
Bloc_17A_F	ATCAACTTGGCCTGGTCTCTTCCA
Bloc_17A_R	GGGAAACTGAAGCCCACAAATGGT
Bloc_17B_F	TGATCGTTCCACCCAACACGATCT
Bloc_17B_R	GTTGTTGTCTCCCATGCGCTTGAA
Bloc_231A_F	AACAAGGAATCCACCGAGACTCCA
Bloc_231A_R	TAGTCCTCCTTCGTAGCCTCAACA
Bloc_231B_F	AGGCGGTAGGATTAGTGCCTGATT
Bloc_231B_R	AGGTACCTGTCTCTGGTAGCGAAT
Bloc_237A_F	AGTGATGTGGATGGATCGGAGCAT
Bloc_237A_R	AACTACACCAAGTCCAGCACCAGT
Bloc_237B_F	AGGCGGTACTCATTAAGGCATCCA
Bloc_237B_R	CTCACGTGGCTGACACAAAGTCAA
Bloc_267A_F	AGCTCCAAGTTACCGGTTTCTCCT
Bloc_267A_R	GAAGATTCGCCATTTCTCGCACCA
Bloc_267B_F	TGCAGGAAGCCAGGATATAGGAGA
Bloc_267B_R	TAGGCATCCTAGCGCTGTCGATTT
Bloc_269A_F	ACAACTTGTTCGTATGAAGCCGGG
Bloc_269A_R	GCATGCATTCGTATGGGAGTAAATGGC
Bloc_269B_F	AGAGAGCGGTCCTAATTCAGGCAT
Bloc_269B_R	TTGCCCTTGTCGCAGATCATCCTA
Bloc_270A_F	AACATGCCTTCAACAGTTCCACCG
Bloc_270A_R	TCGACATCCAGCTTATCGACAGCA
Bloc_270B_F	AAATCGGTTCGTTCTTTCAGCGGC
Bloc_270B_R	TCTCGAGAAACAAACCAGCATAGC
Bloc_287A_F	AGGGAGAGAAGAGCAGAAGAAGGT
Bloc_287A_R	CCCTACAAAGGCATGCTTCTACCTGA
Bloc_287B_F	AACCTGATGGCGGCAACGAGAATA
Bloc_287B_R	TATTCAAGCCTCTGGCACCAACCA
Bloc_309A_F	AGAGGATGCAAATCTCGGCCACTT
Bloc_309A_R	ATGCAGAGCCAGTGGTCACAATAC
Bloc_309B_F	TCTCGCGTAGGGCTGTTTGAAGAT
Bloc_309B_R	TGTGCGCATCCATTCGGAGTGTTA
Bloc_312A_F	AGACGATCTCGGAAACAACGCTGA
Bloc_312A_R	ACGCGACCAAGTACAAGAGAGCAA
Bloc_312B_F	AGTGAAGGGATCGAATGGAGGACA

Bloc_312B_R	GCTATCGCTTTGCAGCTTGCGTAT
Bloc_317A_F	TGTCTTTCCTAGCGCCATCGACAT
Bloc_317A_R	ACCGACATCACTCTGCTGGTGAAA
Bloc_317B_F	ACCAGCAGAGTGATGTCGGTTTCT
Bloc_317B_R	TGGACCTTCATGATGGGATGGGAA
Bloc_318A_F	GAAATGACAGCGGCACAACGGTTA
Bloc_318A_R	TTTGTACGGACAGCAACAATGCCG
Bloc_318B_F	AGTCTCGGTTTGCAGCAGGTTAGT
Bloc_318B_R	TTTATCCGCGCTTGCGATGACTTG
Bloc_326A_F	ACCCTTGTGAGATACCAAAGTGCG
Bloc_326A_R	ACTGCTGCGTTAAGCGGACATCTA
Bloc_326B_F	TTTCGTGATGGAGGACACTTGGCA
Bloc_326B_R	TTCCTTGTAACCCAGTGGAGAGAG

Table 1. Primer sequences designed to amplify across Wisconsin blocs. Each primer was named based on the target bloc and the direction of priming.

Event	Number	% of Aberrant Reads	Number after transposon filter
Total mate bloc pairs	175	64.4	29
Insertion/deletions	51	20.4	10
Inversions	21	10.5	4
Total well-supported rearrangement events	72		14

Table 2. Putative rearrangement events. Percent of aberrant reads indicate how many of the total aberrant reads were accounted for by these blocs. Some aberrant reads had mate pairs mapped to different contigs in the WI strain.

Event Type	Event	Size (bp)	WI DNA		UCB DNA	
			WI Configuration	UCB Configuration	WI Configuration	UCB Configuration
Inversion	6	37,858	Yes	No	No	Yes
Insertion/Deletion	12	1,906	Yes	No	No	Yes
Insertion/Deletion	17	2,957	Yes	No	No	Yes
Insertion/Deletion	231		Ambiguous	No	No	Yes
Insertion/Deletion	237	3,778	Yes	No	No	Yes
Insertion/Deletion	267	1,858	Yes	No	No	Yes
Insertion/Deletion	269	650	Yes	No	No	Yes
Insertion/Deletion	270	5,893	Yes	No	No	Yes
Insertion/Deletion	287	2,059	Yes	No	No	Yes
Inversion	309	23,748	Yes	No	No	Yes
Inversion	312	4,135,317	Yes	No	No	Yes
Insertion/Deletion	317	684	Yes	No	No	Yes
Insertion/Deletion	318	5,892	Yes	No	No	Yes
Inversion	326	2,275	Yes	No	No	Yes

Table 3. Summary of PCR Validations. PCR products for each primer configuration were scored as present (Yes), absent (No), or ambiguous based on their migration on a 1.5% agarose gel (Fig. 5). Present scores indicate very bright bands of the expected size for both of the blocs spanning each breakpoint of an event. Absent scores indicate the lack of a bright band of the expected size for one or both blocs. The ambiguous score indicates unclear results. Event sizes are the number of base pairs in the WI genome internal to two mate blocs.

Event	NRRL 1951	NRRL 832	NRRL 824	Henk PC08-3A	Henk A3704
6	WI	WI	UCB	UCB	WI
12	WI	WI	UCB	UCB	UCB
17	WI	WI	UCB	UCB	UCB
231	WI	WI	WI	WI	UCB
237	WI	WI	WI	WI	WI
267	WI	Ambiguous	WI	WI	UCB
269	WI	UCB	UCB	UCB	UCB
270	WI	UCB	UCB	UCB	UCB
287	WI	WI	UCB	Ambiguous	UCB
309	WI	WI	WI	WI	WI
312	WI	UCB	UCB	UCB	UCB
317	WI	WI	UCB	WI	WI
318	WI	UCB	UCB	UCB	UCB
326	UCB	Ambiguous	UCB	UCB	UCB

Table 4. PCR validation results for all rearrangement events across several strains. NRRL 1951 differed from its descendant, the WI strain, in the state of only event 326. The WI strain consistently showed the same ambiguous banding for event 231 (Fig.4), and strains labeled WI for that event were ambiguous in the same manner.

Strain	Event 6		Event 312	
	WI Configuration	UCB Configuration	WI Configuration	UCB Configuration
WI	Yes	No	Yes	No
NRRL 1951	Yes	No	Yes	No
NRRL 832	Yes	No	No	Yes
NRRL 824	Ambiguous	Yes	No	Yes
UCB (PC0184C)	No	Yes	No	Yes
Henk 1975	Yes	No	Yes	No
PC08-1B	Yes	No	No	Yes
PC08-3B		Yes	No	Yes
PC08-5A	Yes	No	No	Yes
PC08-8B		Yes	No	Yes
PC08-20A	No	Yes	No	Yes
PC08-97A	Yes	No	No	Yes
A1723	No	Yes	No	Yes
A2416	No	Yes	No	Yes
A24700	No	Yes	No	Yes
A3704	Yes	No	No	Yes
3704	No	No	No	Yes
3905	Yes	No	Yes	No
20320	Yes	No	No	Yes
32232	No	Yes	No	Yes
0887Ax	Yes	Yes	No	Yes

Table 5. PCR validation results for inversion events 6 and 312 across wild and culture collection strains. In the strains examined, the configuration of event 6 was evenly distributed. The WI state for event 312 was relatively rare.

A.

Name	Description	Rearrangement Event
Pc12g01540	Strong similarity to sulfate_permease_sutB	17
Pc13g11930	strong similarity to acyl CoA dehydrogenase aidB - *Escherichia coli*	309
Pc13g11940	Glucan 1,4-alpha-glucosidase glaA	309
Pc13g11990	Putative transcription factor CAF32051.1	309
Pc20g13820	Strong similarity to hypoProt_An02g03740	6
Pc20g13860	Strong similarity to hypoProt_An02g03790	6
Pc20g13880	Strong similarity to CreA; TF involved in β-lactam biosynthesis; Glucose repressor	6
Pc20g13890	Duplicate or misannotation of creA	6

PART OF THE PEN CLUSTER:

Pc21g21270	Strong similarity to hypothetical protein B9I2.20 - *Neurospora crassa*	326
Pc21g21280	Strong similarity to methyl sterol oxidase Erg25 - *Saccharomyces cerevisiae*	326
Pc21g21290	Strong similarity to hypothetical protein mg02069.1 - *Magnaporthe grisea*	326

B.

Name	Interpro Terms	Rearrangement Event
Pc13g11880	Tubulin/FtsZ, 2-layer sandwich domain;Tubulin, conserved site;Gamma tubulin; Tubulin/FtsZ, C-terminal; Tubulin/FtsZ, GTPase;Tubulin	309
Pc13g11890	Zinc finger, AN1-type	309
Pc13g11950	Phosphotyrosyl phosphatase activator, PTPA	309
Pc13g11970	Actin/actin-like	309
Pc13g11980	NAD(P)-binding;6-phosphogluconate dehydrogenase, NAD-binding;Dehydrogenase, multihelical;6-phosphogluconate dehydrogenase, C-terminal-like	309
Pc13g12000	GTP cyclohydrolase II	309
Pc21g13970	Fumarate reductase/succinate dehydrogenase flavoprotein, N-terminal	318
Pc21g13990	Monooxygenase, FAD-binding;Aromatic-ring hydroxylase-like	318
Pc20g13780	Histidine triad-like motif;Histidine triad motif	6
Pc20g13800	Mob1/phocein	6
Pc20g13820	Tyrosine protein kinase;Protein kinase, core	6
Pc20g13830	Peptidase S16, lon C-terminal;Peptidase S16, ATP-dependent protease La;Peptidase S16, active site;Peptidase S16, Lon protease, C-terminal region;Peptidase S16, lon N-terminal;ATPase, AAA+ type, core;ATPase, AAA-type, core	6

Pc20g13840	Chaperone DnaJ, C-terminal;Heat shock protein DnaJ, N-terminal;HSP40/DnaJ peptide-binding;Heat shock protein DnaJ;Heat shock protein DnaJ, cysteine-rich region;Heat shock protein DnaJ, conserved site	6
Pc20g13910	Armadillo-type fold	6
Pc12g01550	NA	17
Pc13g11870	NA	309
Pc13g11900	NA	309
Pc13g11910	NA	309
Pc13g11920	NA	309
Pc13g11960	NA	309
Pc20g13790	NA	6
Pc20g13810	NA	6
Pc20g13850	NA	6
Pc20g13870	NA	6
Pc20g13900	NA	6

C.

Event	**Contents**
12	MarinerTE
17	sulfate permease/sulfate transporter
237	conserved hypothetical protein repetitive within WI genome
267	Mariner/pogoTE
269	Repetitive Within WI genome; no other hits
270	Copia_like_retrotransposon
287	FlankedByMicrsatSeq_noMatchesInBLASTX_or_BLASTN
317	non-repetitive_noMatchesInBLASTX_or_BLASTN
318	CopiaTE

Table 6. Genes associated with validated rearrangement events. Genes of interest were identified as those whose start was either within a rearrangement event or less then 500 bp outside it. The very large rearrangements were ignored for this purpose due to the sheer number of genes involved. A. Genes found in literature (van Den Berg et al., 2008; Harris et al., 2009). B. Genes annotated by van Den Berg et al. but otherwise undescribed. C. Contents of insertion events as determined by BLAST restults.

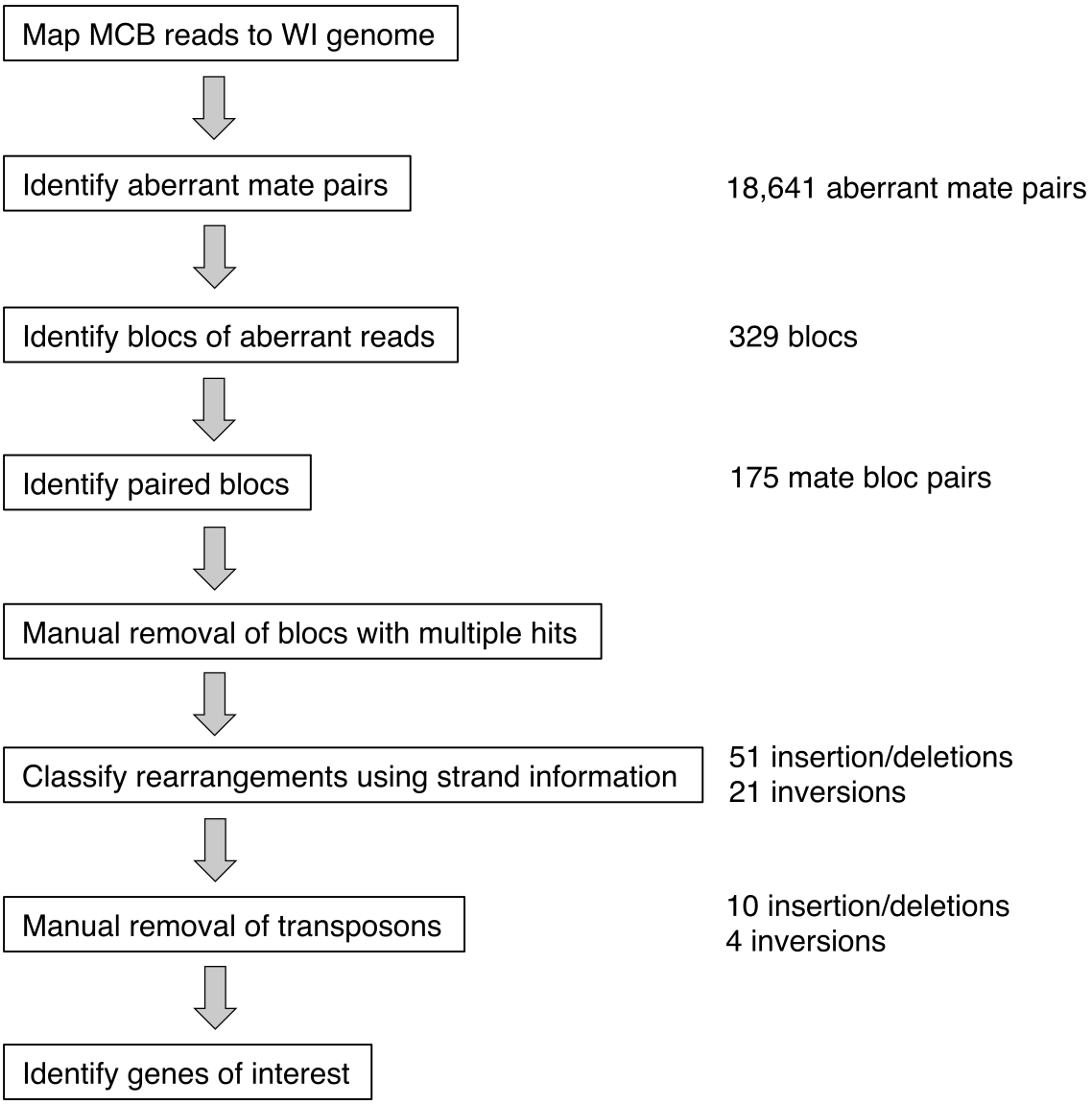


Figure 1. Flow chart for assembly-free rearrangement identification.

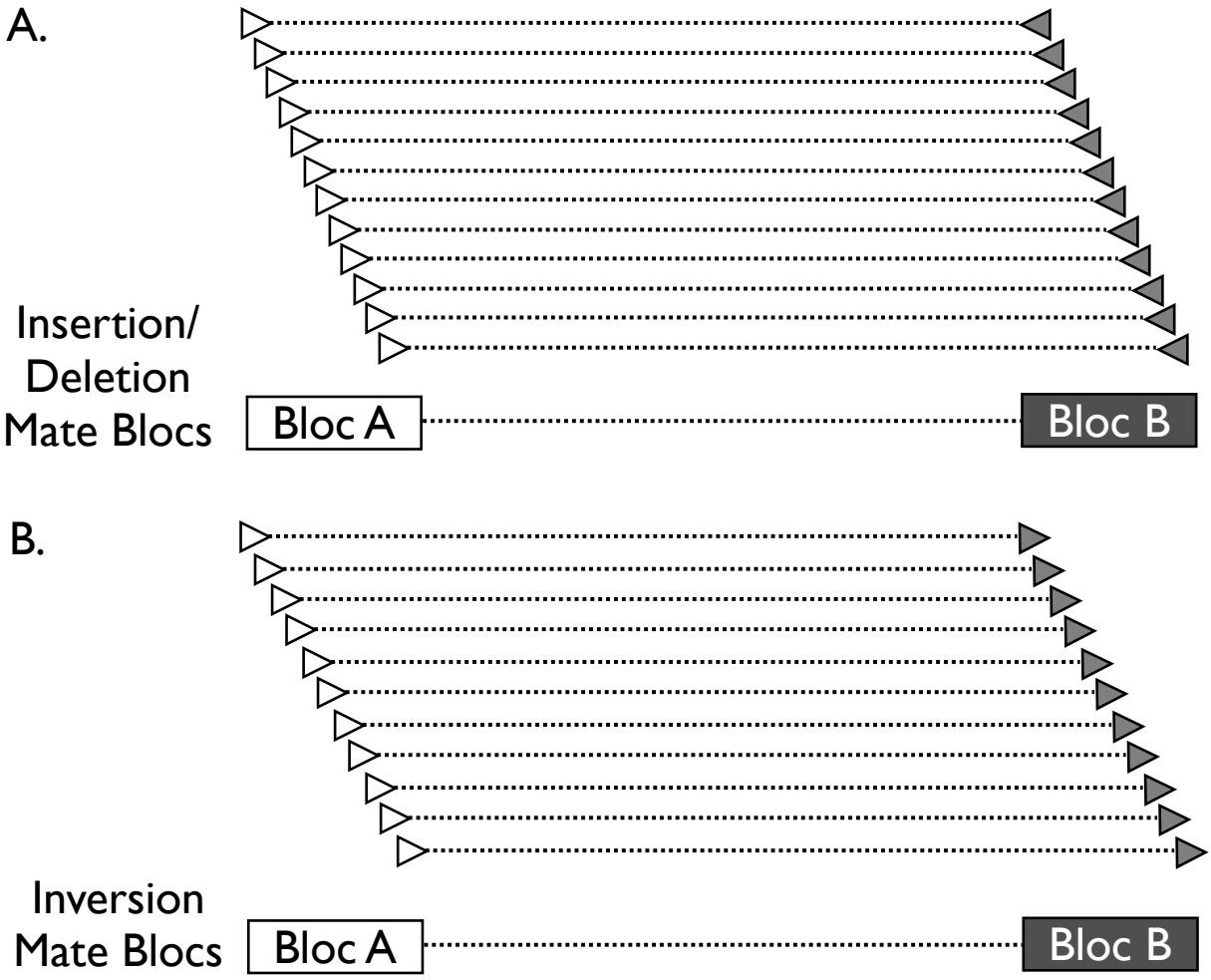


Figure 2. Diagram of read orientation for rearrangements. Read alignments are represented by triangles, with the orientation of the triangle indicating the strand. Mate pairs are connected by dotted lines. Rectangles represent mate blocs as defined by clustering of reads from aberrant mate pairs originating from the same event. Mate pairs on opposite strand indicate an insertion/deletion event (**A**), and mate pairs on the same strand indicate an inversion (**B**).

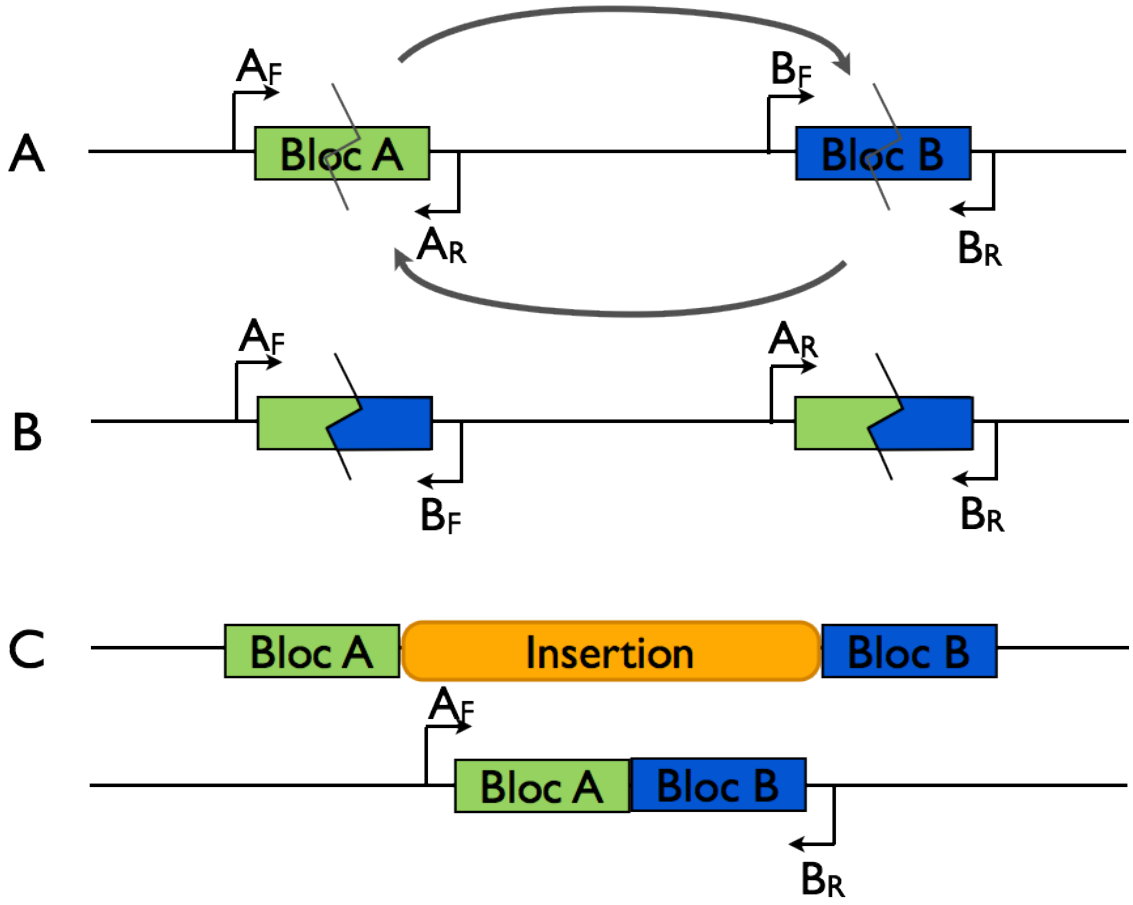


Figure 3. Primer configurations for PCR amplification across blocs. A. Small arrows indicate relative position and direction of forward (F) and reverse (R) primers for a set of paired mate blocs in the WI strain. Zigzag lines indicate breakpoints. Large arrows indicate direction of inversion. B. Primer arrangement to PCR amplify across blocs in a strain with a relative inversion. C. Primer arrangement to PCR amplify across blocs in a strain where the the WI strain has had an insertion relative to UCB.

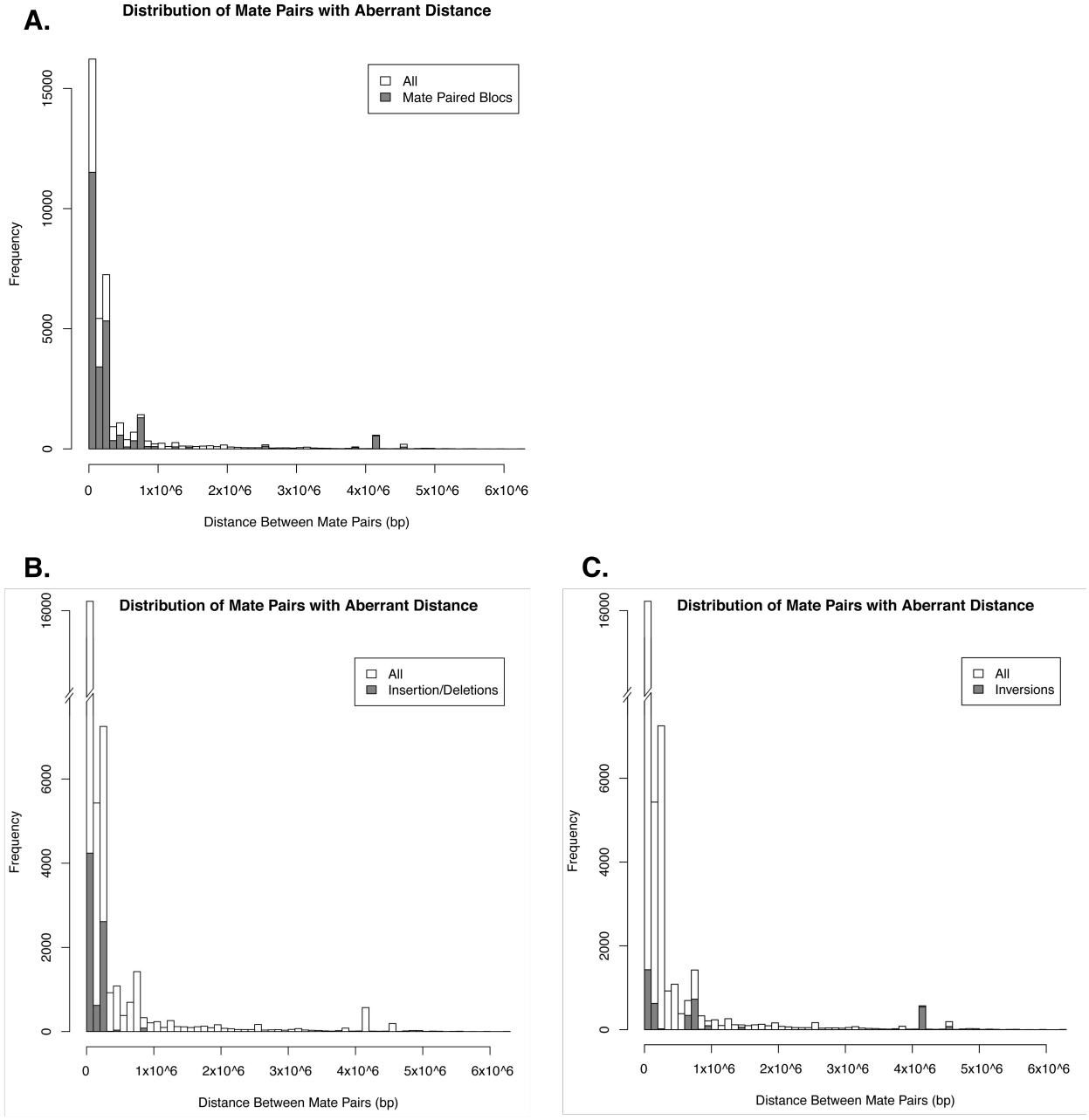


Figure 4. Distribution of aberrant mate pairs. Distribution of all aberrant mate pairs by distance (white) overlaid with mate pairs in all mated blocs (**A**), mate pairs in blocs defining putative insertion/deletions (**B**), and mate pairs in blocs defining putative inversions (**C**).

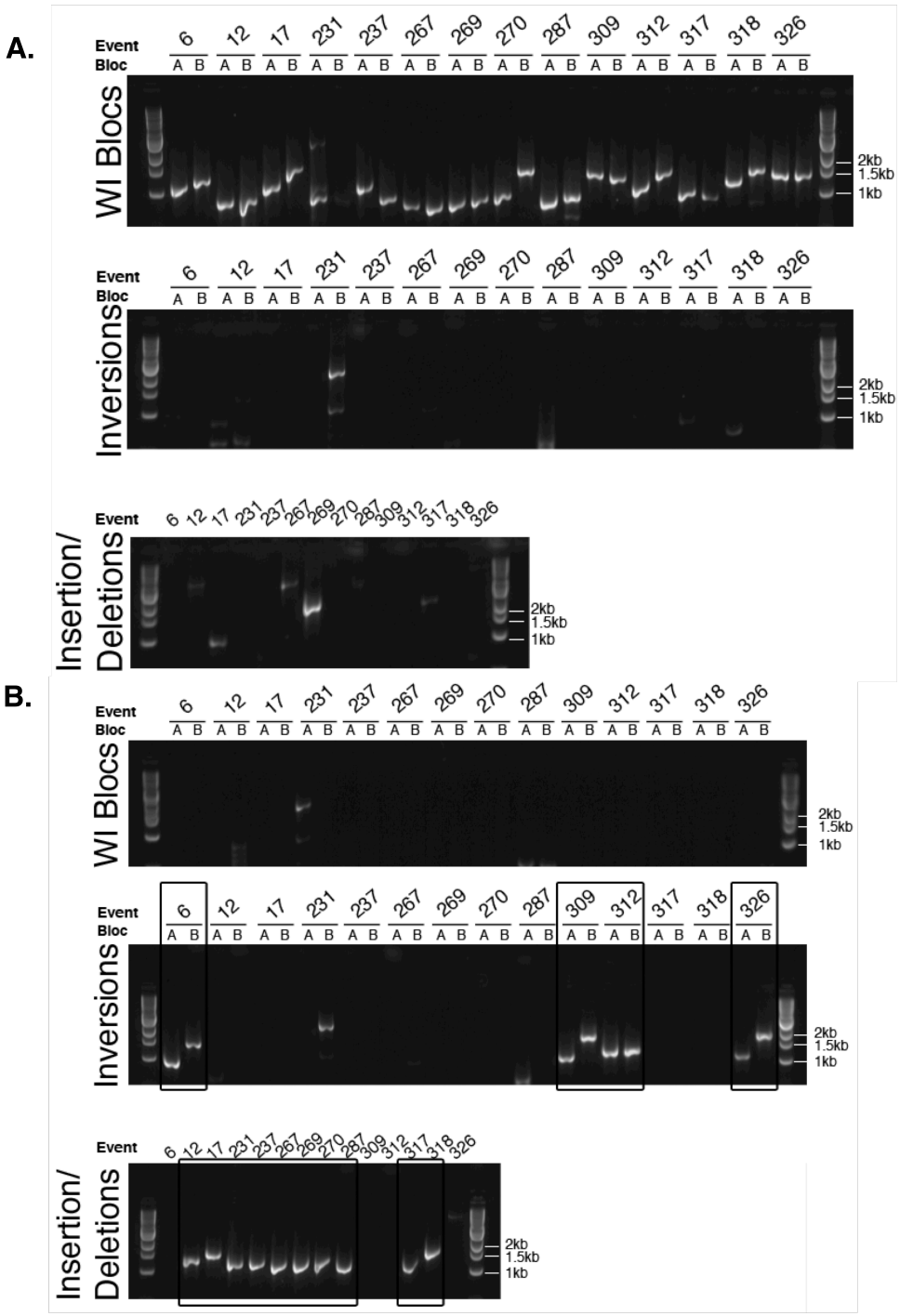


Figure 5. PCR validation of rearrangements identified *in silico*. PCR reactions used DNA from the **A.** WI strain and **B.** UCB strain. The top gel image in each figure used primers in the orientation to amplify across blocs in the WI strain. Middle images had primers to amplify across blocs in the UCB strain for a relative inversion. Bottom images show reactions priming across insertions in the WI strain relative to the UCB strain. All PCR products for each strain were run on a single gel. Each PCR reaction is labeled by event and bloc (A or B), and boxes highlight successful amplification of rearrangement events in the MCB strain.

CHAPTER 2

Gibberellic acid induces asymbiotic germination of the obligate mycoheterotroph *Pterospora andromedea*

Abstract

In order to overcome a very low germination rate in *Pterospora andromedea* seeds in lab settings, we developed an asymbiotic method to induce germination. Gibberellic acid (GA), a phytohormone which induces germination in many plants, stimulates *P. andromedea* germination at frequencies that can be far higher than that achieved with its fungal symbiont, *Rhizopogon salebrosus*, alone. With constant exposure to 0.5mM GA, we found nearly 80% germination, as opposed to only 15% with *R. salebrosus*. Even a short exposure to GA is sufficient for significantly enhanced germination. One day of exposure to 0.5mM GA induced ~55% stage 3 germination, and three days of exposure led to ~ 70% stage 3 germination. Exposing seeds to a lower dosage (0.1mM GA) for a longer time (2 weeks) produced a comparable germination rate. The closely related plant *Sarcodes sanguinea* was less sensitive to exogenous GA stimulation. *S. sanguinea* required a longer exposure to produce lower germination (~1% with 0.5mM or 1mM GA). Exposure to GA is the only known method of inducing monotrope germination without the presence of a specific fungal symbiont. These studies do not address whether GA is the signal produced by *Rhizopogon* or other mycoheterotroph symbionts, but tests for germination by volatiles suggest that it is diffusible rather than airborne. GA stimulation achieved workable germination rates for further studies, and GA may also be used as an improved assay for seed viability.

Introduction

In the classical ectomycorrhizal symbiosis, a photosynthetic plant fixes carbon from the atmosphere and trades a portion of its carbon to a fungus growing on its roots in exchange for mineral nutrients (Smith & Read 2008). Mycoheterotrophic plants parasitize mycorrhizal fungi and in doing so reverse the usual direction of net carbon flow in mycorrhizal interactions and apparently cheat one of the most widespread mutualisms in terrestrial ecosystems. Mycoheterotrophic interactions therefore represent extremes in a continuum of plant-fungal interactions within mycorrhizal symbioses. The monotropes (Ericaceae, Monotropoideae) have been a focus of scientific interest since the birth of mycorrhizal studies, and with the advent of molecular techniques, monotropes continue to inspire new lines of investigation (Frank 2004; Berch et al. 2005; Merckx et al. 2009). All members of this subfamily appear to be obligate, non-photosynthetic mycoheterotrophs that exhibit extreme specialization on mycorrhizal fungi (Bidartondo & Bruns 2001; Bidartondo & Bruns 2002; Bidartondo & Bruns 2005).

Pterospora andromedea (pine drops; Ericaceae, Monotropoideae), which can grow up to a meter tall, is one of the largest monotropes and has a range spanning the continental United States (Wallace 1975). However, even the largest specimens begin life as tiny dust seeds with little room for nutritional resources (Bakshi 1959). Such a seedling likely would be unable to survive for long without an appropriate host. Previous work indicates that *P. andromedea* germinates only in the presence of a specific fungal associate. Bakshi failed to germinate seeds under a wide variety of temperature,

storage, and nutrient conditions, including planting the seeds in soil collected around a *P. andromedea* root zone and placed under a *Picea pungens* tree (Bakshi 1959). Bakshi also used a tetrazolium method by Cottrell (1948) to determine viability, whereby he found no seeds older than nine weeks to be viable. Using fungal cultures, Bruns and Read (2000) discovered that a diffusible or volatile substance from either the specific fungal host or a very closely related fungus could induce a low level of germination. However, mature plant genotypes associate only with single host fungal species (Bidartondo & Bruns 2005).

The inability to grow *P. andromedea* from seed in the lab severely limits further work on this plant. Gibberellins (GAs), a class of phytohormone, are known to induce germination in many photosynthetic plants (Koornneef et al. 2002). The mechanism of GA action in seed germination has been most thoroughly studied in cereals (Bethke et al. 1997). The endosperm, seed tissue that supplies the nutrients for embryo development, includes starchy endosperm and an aleurone layer. Upon stimulation by GA, the aleurone layer synthesizes and secretes hydrolytic enzymes that break down the seed's starch reserves, making them available to the embryo, and then the aleurone layer cells undergo programmed cell death (Fath et al. 2000). Endogenous GA biosynthesis and activation are tightly regulated, but exogenous application of the hormone bypasses this step and activates a signal transduction cascade leading to developmental changes (Bethke & Jones 1998). Here we present the only known method for asymbiotic germination of *P. andromedea* seeds using gibberellic acid (GA).

Materials and Methods

Strains and Collections

All *P. andromedea* collections were made from Blodgett Forest, a University of California research station situated at ~4,000 ft in the Sierra Nevada foothills near Georgetown, CA. In order to type *P. andromedea* plants by fungal host, small root fragments were collected in midsummer, after emergence of inflorescences. Genomic DNA was extracted using the XNAP REDExtract-N-Amp™ Plant PCR Kit (Sigma-Aldrich) and PCR amplified using primers ITS1F and 4B (Gardes & Bruns 1993). PCR products were sequenced at the UC Berkeley DNA Sequencing Facility using ABI chemistry. Sequences were identified as either *R. salebrosus* or *R. arctostaphyli* by matching to sequences of known specimens in the GenBank database (Bidartondo & Bruns 2001; Bidartondo & Bruns 2002). Mature *P. andromedea* seeds from typed plants were collected after seed set in late summer or early autumn.

All experiments with *R. salebrosus* used strain TDB-379, itself isolated from *P. andromedea* roots, and grown on Modified Melin Norkrans (MMN) medium with 1.5% agar (Bruns & Read 2000).

Mature *Sarcodes sanguinea* seeds were collected from near the USDA Forest Service station near Dinkey Creek, CA in Sierra National Forest.

Gibberellic Acid Assays

Pterospora andromedea seeds less than one year old were de-winged and surface sterilized with gentle agitation for 20 min in saturated calcium hypochlorite with a drop of Tween 80, filtered on sterile Whatman paper, and rinsed twice with sterile water (Bruns & Read 2000). Seeds were plated onto 2% water agar and monitored for one week for contamination, which when found was excised. Seeds were manually transferred to 0.8% agarose with various concentrations of filter-sterilized GA or to agarose alone as a disruption control. Germination was assessed after two months unless otherwise noted, according to the stages outlined by Bruns and Read (2000) (Fig. 1). Experiments used at least 100 seeds of each treatment per biological replicate.

Volatile Seed Germination Assay

To test whether a volatile compound produced by *R. salebrosus* induces germination in *P. andromedea* seeds, we suspended seeds above growing *R. salebrosus* colonies. *Rhizopogon salebrosus* was grown on MMN, and a thin layer of MMN was also spread onto the inside surface of the petri dish lids. Surface sterilized *P. andromedea* seeds were thus adhered to the lids above the growing *R. salebrosus* colonies. All dishes were sealed with Parafilm.

Results

Continuous exposure to GA induced germination in *P. andromedea* (Fig. 2). Optimal total germination response of 75±10% germination occurred with 0.5mM GA, as compared with 0% germination on agarose alone, nearly zero germination with 0.01mM GA, and 21±12% germination with *R. salebrosus*. Stage 3 germination, indicated by radicle emergence, was highest with 23±14% germination on 0.1mM GA. Stage 3 germination induced by *R. salebrosus* was close to zero. A single replicate of *S. sanguinea* seeds exposed to GA failed to germinate but did have 42% germination with *R. salebrosus*.

A short exposure to GA was sufficient to induce *P. andromedea* germination and outperformed continuous exposure (Fig. 3). Three days of exposure to 0.5mM GA led to 90% total germination response and nearly 70% stage 3 germination. Two weeks of GA exposure produced nearly 100% stage 3 germination. Parallel treatments with 0.1mM GA showed less than 10% germination for 1 and 3 day exposures, but a two-week exposure produced 75% stage 3 and 80% total germination.

Sarcodes sanguinea seeds failed to respond to GA concentrations of 0.01, 0.1, 0.5, and 1 mM after two months. However, after three months low levels of germination occurred with 0.5mM and 1mM GA (1±1% and 1±0.7%, respectively). In the presence of *R. salebrosus*, 42±14% of the seeds germinated.

None of the *P. andromedea* seeds suspended above *R. salebrosus* germinated.

Discussion

We utilized *P. andromedea*'*s* germination in response to GA to optimize an improved method useful as a check for seed viability and to enable further studies using *P. andromedea* seeds. Continuous exposure to GA led to far higher germination response than with *R. salebrosus*, *P. andromedea*'s host fungus, alone (Fig. 2). Gibberellic acid-induced germination was also higher than the most *P. andromedea* germination observed by Bruns and Read (2000) of 26±21% (SD). The ability of GA to germinate seeds well beyond 9 weeks old indicate that it is a more accurate determination of seed viability than the methods applied by Bakshi (1959). However, while increasing concentrations of GA did lead to higher overall germination response, increased GA concentrations also had less stage 3 germination, suggesting either inhibition by or toxicity of GA.

This suppressive effect of higher GA concentrations was ameliorated by exposing seeds to GA for shorter periods and then transferring them to plain agarose. A low concentration of 0.1mM GA required 2 weeks of GA exposure to get more than 10% germination. Just one day with 0.5mM GA induced over 4.5 times as much total germination as interaction with *R. salebrosus*, and most of this germination was stage 3. Three days of 0.5mM GA produced nearly 90% total germination, and 2 weeks led to nearly all seeds reaching stage 3 germination.

Gibberellins were first isolated from a fungus (*Fusarium fujikuroi*) and have since been identified as a secondary metabolite in many other fungi (Bömke & Tudzynski 2009). However, in order for GA to be the diffusible substance posited by Bruns and Read (2000), one would expect GA production to be limited to *R. salebrosus* and *R. arctostaphyli* and not other fungi or bacteria *P. andromedea* is likely to encounter in the field. For this reason, it seems more likely that the chemical signal of Rhizopogon species either trigger the endogenous gibberellin pathway in *Pterospora* or work by some independent pathway. The failure of close proximity to the headspace above growing *R. salebrosus* colonies to germinate *P. andromedea* seeds suggest that the symbiont germination signal is diffusible rather than a volatile.

Sarcodes sanguinea low germination response to GA under these conditions. The seeds were viable, as demonstrated by their germination in the presence of *R. salebrosus* at a rate very close to the 46±22% found by Bruns and Read (2001). *Pterospora andromedea* seeds are far smaller and have more delicate seed coats than those of *S. sanguinea*. The low germination in *S. sanguinea* could be due to a different porosity or composition of the seed coats or an insensitivity to GA.

Acknowledgements

We would like to thank R. Jones and P. Bethke for advice on using gibberellic acid. Blodgett Forest Research Station for graciously allowing collections. This work

was supported by grants from the Mycological Society of San Francisco, the Mycological Society of America, the California Native Plant Society, and UC Berkeley.

Literature Cited

Bakshi, T., 1959. Ecology and morphology of Pterospora andromedea. *Botanical Gazette*, 120(4), pp.203-217.

Berch, S.M., Massicotte, H.B. & Tackaberry, L.E., 2005. Re-publication of a translation of “The vegetative organs of Monotropa hypopitys L.” published by F. Kamienski in 1882, with an update on Monotropa mycorrhizas. *Mycorrhiza*, 15(5), pp.323-32.

Bethke, P C & Jones, R L, 1998. Gibberellin signaling. *Current opinion in plant biology*, 1(5), pp.440-6.

Bethke, Paul C., Schuurink, R. & Jones, Russell L., 1997. Hormonal signalling in cereal aleurone. *Journal of Experimental Botany*, 48(7), pp.1337-1356.

Bidartondo, M I & Bruns, T D, 2001. Extreme specificity in epiparasitic Monotropoideae (Ericaceae): widespread phylogenetic and geographical structure. *Molecular ecology*, 10(9), pp.2285-95.

Bidartondo, M I & Bruns, T D, 2002. Fine-level mycorrhizal specificity in the Monotropoideae (Ericaceae): specificity for fungal species groups. *Molecular ecology*, 11(3), pp.557-69.

Bidartondo, M. & Bruns, T., 2005. On the origins of extreme mycorrhizal specificity in the Monotropoideae (Ericaceae): performance trade-offs during seed germination and seedling development. *Molecular ecology*, 14(5), pp.1549-60.

Bruns, Thomas D & Read, D.J., 2000. In vitro germination of nonphotosynthetic, myco-heterotrophic plants stimulated by fungi isolated from the adult plants. *New Phytologist*, 148(2), pp.335-342.

Bömke, C. & Tudzynski, B., 2009. Diversity, regulation, and evolution of the gibberellin biosynthetic pathway in fungi compared to plants and bacteria. *Phytochemistry*, 70(15-16), p.1876–1893.

Cottrell, H., 1948. Tetrazolium salt as a seed germination indicator. *Annals of Applied Biology*, pp.123-131.

Fath, A. et al., 2000. Programmed cell death in cereal aleurone. *Plant Molecular Biology*, (Figure 2), pp.255-266.

Frank, A.B., 2004. On the nutritional dependence of certain trees on root symbiosis with belowground fungi (an English translation of A. B. Frank’s classic paper of 1885) Structure of the mycorrhiza Development of the mycorrhiza. *Mycorrhiza*, (May 1880), pp.1-9.

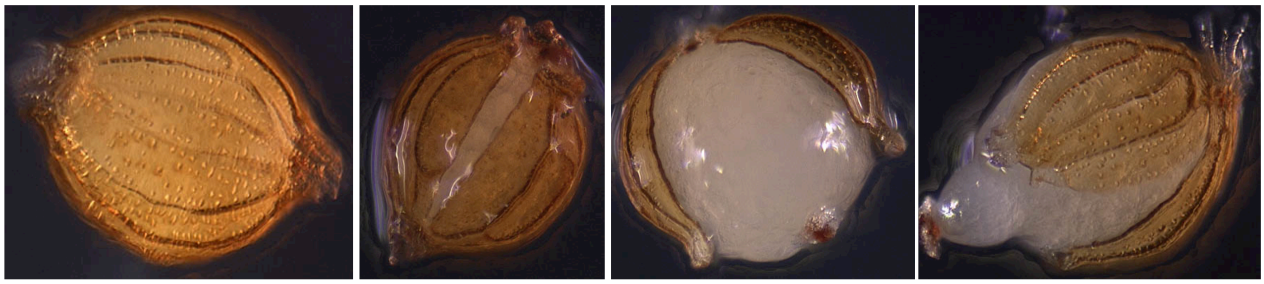
Gardes, M. & Bruns, T., 1993. ITS primers with enhanced specificity for basidiomycetes-application to the identification of mycorrhizae and rusts. *Molecular Ecology*, 2, p.113–118.

Koornneef, M., Bentsink, L. & Hilhorst, H., 2002. Seed dormancy and germination. *Current Opinion in Plant Biology*, 5, pp.33-36.

Merckx, V., Bidartondo, Martin I & Hynson, N.A., 2009. Myco-heterotrophy: when fungi host plants. *Annals of Botany*, 104(7), pp.1255-1261.

Smith, S.E. & Read, D.J., 2008. *Mycorrhizal Symbiosis* 3rd ed., San Francisco: Academic Press.

Wallace, G.D., 1975. Studies of the Monotropoideae (Ericaceae): Taxonomy and Distribution. *The Wasmann Journal of Biology*, 33(1 and 2), pp.1-88.



Ungerminated Stage 1 Stage 2 Stage 3

Figure 1. Stages of *Pterospora andromedea* seed germination. All seeds were de-winged and placed on agarose with gibberellic acid. Ungerminated seeds have intact seed coats. Stage 1 seeds have cracked seed coats. Stage 2 seeds have imbibed and swelled. Stage 3 seeds have visible radicle emergence.

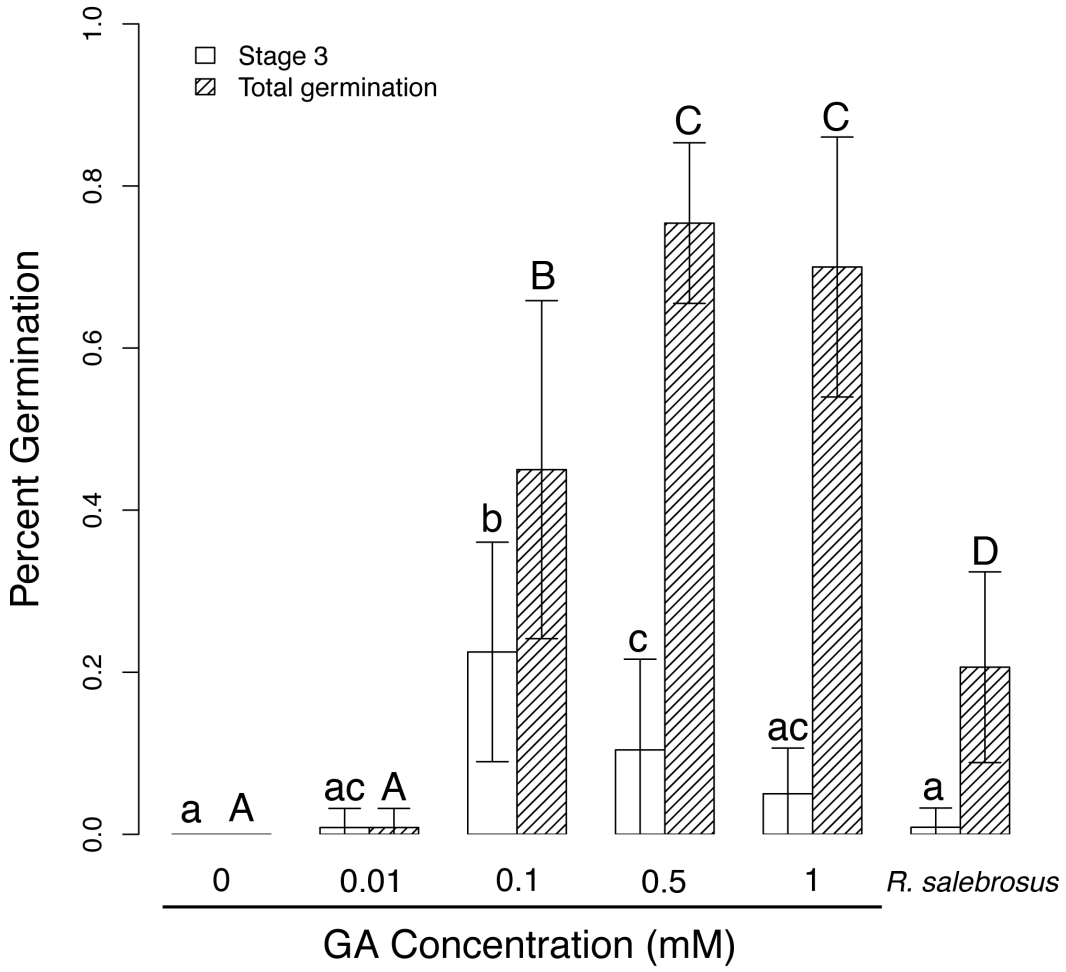


Figure 2. Gibberellic acid induces *Pterospora andromedea* seed germination. Error bars indicate standard deviations. Lower case and capital letters respectively indicate groups based on one-way ANOVA for stage 3 and total germination, with treatment condition as the independent variable and a 95% confidence.

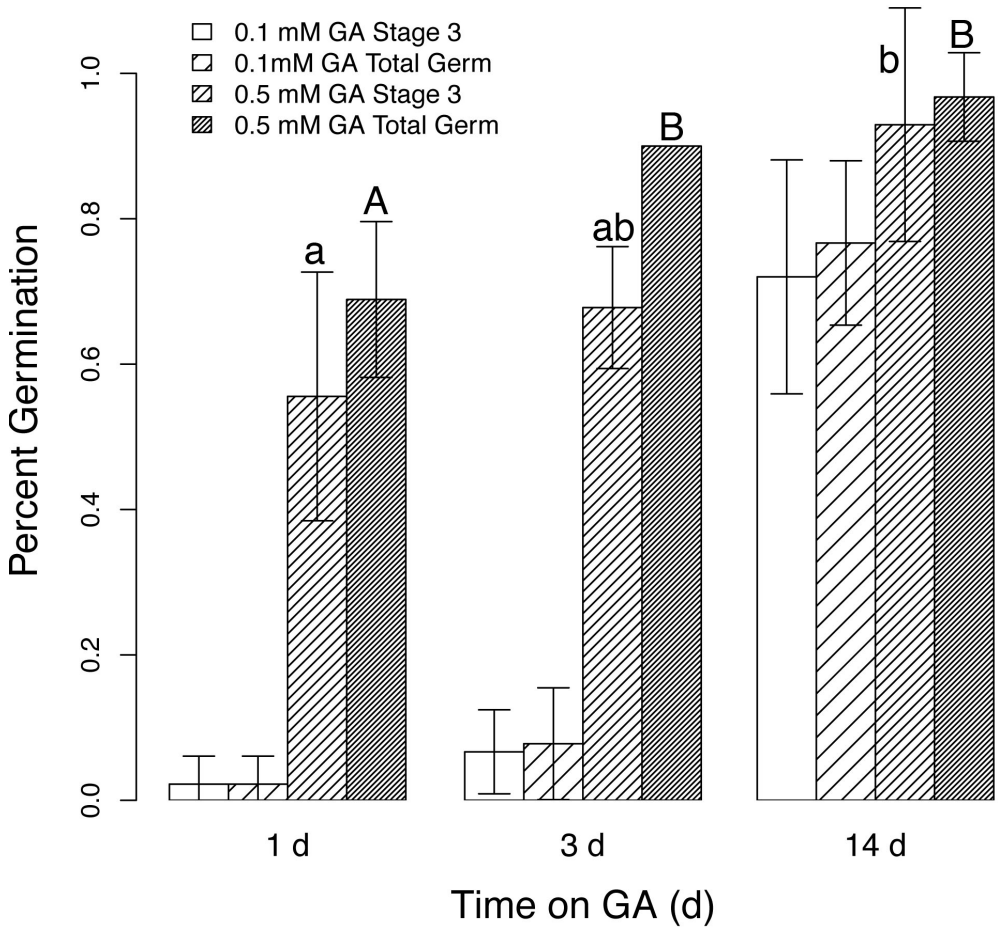


Figure 3. Short exposure to gibberellic acid induces *Pterospora andromedea* seed germination. Seeds were exposed to GA and then moved to plain agarose. Error bars indicate standard deviations. Lower case and capital letters respectively indicate groups based on one-way ANOVA for 0.5 mM GA stage 3 and total germination with 95% confidence. Time on GA was the independent variable.

CHAPTER 3

Transcriptional Differences in *Rhizopogon salebrosus* with Partner and Non-Partner Mycohterotroph Seeds

Abstract

Mycoheterotrophs, plants that parasitize mycorrhizal fungi and cheat the classical mycorrhizal symbiosis, have been integral to the discovery and continuing study of mycorrhizal fungi. The monotropes constitute a group of obligately dependent, non-photosynthetic mycoehterotrophs with extreme host specificity. Here we generate a *de novo* transcriptome assembly and examine differential expression when *R. salebrosus* interactions with seeds from partner (*R. salebrosus* type) and non-partner (*R. arctostaphyli* type) *Pterospora andromedea*. Because *P. andromedea* seeds are very small, germination in the presence of a suitable host fungus is a critical step in the establishment of symbiosis. We have identified putative genes in categories thought to be important for ectomycorrhizal symbiosis. These include stress response, host defense, cellular transport, secretion, carbohydrate metabolism and transport, and signal transduction. To our knowledge, this is the first study of gene expression with any mycorrhizal fungus involved with a mycoheterotrophic plant.

Introduction

In the classic ectomycorrhizal symbiosis, a photosynthetic plant fixes carbon from the atmosphere and trades a portion of its carbon to a fungus growing on its roots in exchange for mineral nutrients (Smith and Read 2008). Mycoheterotrophic plants, on the other hand, parasitize mycorrhizal fungi. In so doing, mycoheterotrophs reverse the usual direction of net carbon flow in mycorrhizal interactions and apparently cheat one of the most widespread mutualisms in terrestrial ecosystems. Mycoheterotrophs have also been termed epiparasites, because mycorrhizal fungi acquire carbon from photosynthetic plants, and the mycoheterotroph is thereby indirectly parasitizing another plant (Bjorkman 1960).

While a tripartite system is the minimum required for epiparasitism, in nature individual plants typically host many different fungi on their roots, and individual fungi can associate with multiple plants. Green plants may transfer carbon amongst each other through these common mycorrhizal networks (Simard et al. 1997; Simard and Durall 2004). Mycoheterotrophic interactions therefore represent extremes in a continuum of plant-fungal interactions within mycorrhizal symbioses. The monotropes (Ericaceae, Monotropoideae) have been a focus of scientific interest since the birth of mycorrhizal studies (Kamienski 1882; Berch, Massicotte, and Tackaberry 2005), and with the advent of molecular techniques, monotropes continue to inspire new lines of investigation (Berch et al., 2005; Frank, 2004; Merckx, Bidartondo, & Hynson, 2009). All members of this subfamily appear to be obligate, non-photosynthetic mycoheterotrophs that exhibit extreme specialization on mycorrhizal fungi (Bidartondo and Bruns 2001; Bidartondo and Bruns 2002; Bidartondo and Bruns 2005).

Pterospora andromedea (pine drops; Ericaceae, Monotropoideae), which can grow up to a meter tall, is one of the largest monotropes. It has a range spanning western North America and is disjunct in the lake states (Wallace 1975). However,

even the largest specimens begin life as tiny, dust-like seeds with little room for nutritional resources (Bakshi 1959). Such a seedling likely would be unable to survive for long without an appropriate host. *P. andromedea* seeds have been germinated only in the presence of its specific fungal associate, a closely related fungus, or asymbiotically when stimulated by gibberellic acid (GA) (see Chapter 2). Using fungal cultures, Bruns and Read (Bruns & Read, 2000) discovered that a diffusible or volatile substance from either the specific fungal host or very closely related fungi could induce a low level of germination. *Pterospora andromedea* contains two cryptic species that can be identified based on the fungal associate (either *R. salebrosus* or *R. arctostaphyli*) of mature plants, and plants only form mycorrhizae with the same fungal associate as their parent (Bidartondo and Bruns 2002; Bidartondo and Bruns 2005).

In this work, we focus on the mycorrhizal fungus partner. *Rhizopogon salebrosus* is the most genetically tractable player in the tripartite symbiosis, since it is the easiest to keep in culture in the lab and as a fungus likely has the smallest genome. By examining how the *R. salebrosus* responds to germination of seeds from partner and non-partner *P. andromedea* plants, we characterize a critical phase in the development of a previously poorly studied mycorrhizal symbiosis and elucidate host choice among distinct *P. andromedea* haplotypes. We used an RNA-Seq approach (Fig. 1) to identify significant transcriptional differences between *R. salebrosus* interacting with a partner *P. andromedea* and a non-partner *P. andromedea* haplotope associated with *R. arctostphyli*.

Materials and Methods

Strains and Collections

All *P. andromedea* collections were made from Blodgett Forest, a University of California research station situated at ~4,000 ft in the Sierra Nevada foothills near Georgetown, CA. In order to type *P. andromedea* plants by fungal host, small root fragments were collected in midsummer, after emergence of inflorescences. Genomic DNA was extracted using the XNAP REDExtract-N-Amp™ Plant PCR Kit (Sigma-Aldrich) and PCR amplified using primers ITS1F and 4B (Gardes and T Bruns 1993). PCR products were sequenced at the UC Berkeley DNA Sequencing Facility using ABI chemistry. Sequences were identified as either *R. salebrosus* or *R. arctostaphyli* by matching to sequences of known specimens in the GenBank database (Bidartondo and Bruns 2001; Bidartondo and Bruns 2002). Mature *P. andromedea* seeds from typed plants were collected after seed set in late summer or early autumn.

All experiments with *R. salebrosus* used strain TDB-379, itself isolated from *P. andromedea* roots, and grown on Modified Melin Norkrans (MMN) medium with 1.5% agar (Bruns and Read 2000).

Germination Conditions

Seed Preparation

Pterospora andromedea seeds less than one year old were physically de-winged and surface sterilized with gentle agitation for 20 min in saturated calcium hypochlorite with a drop of Tween 80, filtered on sterile Whatman paper, and rinsed twice with sterile water (Bruns and Read 2000). Seeds were plated onto 2% water agar and monitored for one week for contamination, which when found was excised. Seeds were manually transferred to 0.8% agarose with various concentrations of filter-sterilized GA (Sigma-Aldrich) or to agarose alone as a disruption control.

Heat-killed *P. andromedea* seeds were used as disturbance controls. Surface sterilized seeds were put into Eppendorf tubes with sterile distilled water and incubated in a 65°C water bath for 60 min (method modified from Egley, 1990; Gashaw & Michelsen, 2002; Granstrom & Schimmel, 1993). The heat killed seeds were visually indistinguishable from untreated seeds. To test the efficacy of this treatment, heat killed and untreated seeds were placed onto agarose plates infused with 0. 5 mM GA. Over 100 seeds for each treatment were used for this test. None of the heat killed seeds germinated after 4 weeks, while untreated seeds had 79% germination.

Symbiosis Sample Conditions

Hyphal plugs from growing fronts of *R. salebrosus* mycelium were used to inoculate 15 cm Petri dishes each containing 75mL of MMN with 1.5% agar and overlayed with sterilized wettable cellophane (Promega, Madison, WI). After four weeks, cultures were ~ 4-5 cm in diameter. Live and heat-killed *P. andromedea* seeds treated with GA were placed at 1-2 mm intervals just outside the growing front of the mycelium. Plates were kept in the dark at room temperature. Three biological replicates, each with at least 75 seeds, were used for each condition. At one and two week time points, the interaction zone, a 1 cm-wide circular strip of mycelium centered where the seeds were placed, was harvested (Fig. 1). Tissue was placed immediately into screwcap tubes containing 0.3 g of 0.5 mm zirconia/silica beads (Biospec Products, Bartlesville, OK) and 1 mL of TRIzol (Invitrogen Life Science Technologies, Carlsbad, CA), frozen in liquid nitrogen, and stored at -80°C.

Library Preparation and Sequencing

Total RNA extraction

Samples frozen in TRIzol were thawed on ice and then bead-beat at maximum speed for 30 sec twice. Following a gentle shake for 5 min at room temperature, 0.2 mL of chloroform was added. Tubes were mixed by bead-beating at maximum speed for 15 sec in a Mini-Beadbeater (BioSpec Products, Bartlesville, OK) and then spun for 15 min at 12,000 x g at 4°C. The supernatant was taken and added to 0.5 mL of isopropanol, then shaken for 10 min at room temperature. Samples were spun for 10 min at 12,000 x g at 4°C, and supernatants discarded. Pellets were washed with 1 mL 75% ethanol, dried, and resuspended in nuclease-free water. Aliquots were run on a 0.8% agarose gel to determine whether the RNA had degraded. Samples were stored at -80°C.

Library Preparation

For each library, 10 μg of total RNA was treated with rDNase (Ambion, Austin, TX). As with all other library preparation procedures unless otherwise noted, this was according to manufacturer's protocol. Messenger RNA was purified with dynal oligo (dT) beads (Invitrogen). The mRNA was fragmented with fragmentation buffer (Ambion), ethanol precipitated, and resuspended in 10.5μL of rnase-free water. For first strand cDNA synthesis, 1 μL of random hexamer primers (Invitrogen) were added and incubated at 65°C for 5 minutes before adding the following: 4μL first strand buffer (Invitrogen), 2μL of 100mM DTT (Invitrogen), 1μL of 10mM dNTPs, and 0.5μL of RNAseOUT (Invitrogen). Samples were incubated at 25°C for 2 min before addition of 1 μL Superscript II (Invitrogen) and undergoing the following program: 10 min at 25°C, 50 min at 42°C, and 15min at 70°C. For second strand cDNA synthesis, the tubes were cooled on ice prior to adding 51μL of water, 20 μL of second strand buffer (Invitrogen), and 3μL of 10mM dNTPs. After a 5 min incubation on ice, 1μL of RNaseH (Invitrogen) and 5 μL of DNA polymerase I (Invitrogen) were added, and tubes were incubated at 16°C for 2.5 hours. The cDNA was purified using 1.5 volumes of AMPure XP beads (Beckman Coulter).

Ends were repaired with the addition of the following, followed by incubation at 20°C for 30min and purification with AMPure XP beads as previous: 45 μL of water, 10 μL of T4 DNA ligase buffer with 10mM ATP (New England Biolabs), 4 μL of 10mM dNTPs, 5 μL of T4 DNA polymerase (New England Biolabs), 1 μL of Klenow DNA polymerase (New England Biolabs), and 5 μL of T4 PNK (New England Biolabs). Single A bases were added by adding the following, incubating at 37°C for 30 min, and purifying with AMPure XP beads as before: 5 μL of Klenow buffer (New England Biolabs), 10 μL of dATP, and 3 μL of Klenow 3' to 5' exo- (New England Biolabs). Adaptors were ligated by adding the following: 25 μL of 2x DNA ligase buffer (Enzymatics, Beverly, MA), 1 μL of adaptor oligo mix (Illumina), and 1 μL of DNA ligase (Enzymatics). The tubes were incubated at 25°C for 15 min, and then the cDNA was purified with 1 volume of AMPure XP beads to exclude small fragments and primer dimers and eluted to a volume of 30 μL.

For PCR enrichment, 10 μL of the purified, adaptor-ligated cDNA were added to the following: 5 μL of 10x Pfx amplification buffer (Invitrogen), 2 μL of 10mM dNTPs, 2 μL of 50 mM $MgSO_4$, 1 μL of each primer, 0.8 μL of Pfx DNA Polymerase (Invitrogen), and 28 μL of water. The reaction was run with the following program: 1. 98°C for 30 sec, 2. 98°C for 10 sec, 3. 65°C for 30 sec, 4. 68°C for 30 sec, 5. Go back to step 2 15 times, 6. 68°C for 5 min, 7. Hold at 4°C. DNA was purified with 1 volume of AMPure XP beads, and quality was assessed by Bioanalyzer (Agilent Technologies, Santa Clara, CA).

Sequencing

Each sample was sequenced with one lane of single-end Illumina (San Diego, CA) 76bp reads at the Vincent J. Coates Genomic Sequencing Laboratory (UC Berkeley, Berkeley, CA). The Illumina standard pre-processing pipeline was used. Three biological replicates for each condition were sequenced, each with one lane (Fig. 1). Samples were distributed across five flow cells. On average, approximately 28.5 million reads were produced per lane.

De novo Transcriptome Assembly

Because the genome for *R. salebrosus* has not yet been sequenced, there is no whole genome reference sequence against which to map reads for RNA-Seq. However, a large amount of transcript sequence is generated by RNA-Seq itself, and we took advantage of this to generate a hybrid *de novo* transcriptome assembly combining four lanes of Illumina sequence with ESTs generated by the Joint Genome Institute (Walnut Creek, CA) as part of the *R. salebrosus* genome project. To generate these ESTs, *R. salebrosus* was been grown on cellophane on top of MMN medium, and RNA was extracted as described above. Sequences were produced from a 454 run and assembled using Newbler (Roche, Branford, CT).

Reads from the four Illumina lanes were filtered prior to assembly. Mosaik (version 1.0.1388) (Hillier et al. 2008) was used to identify and remove reads that mapped to the Illumina primers or ribosomal sequence. These constituted 2.0% of the total reads. The remaining reads were trimmed to 55bp and assembled with Velvet (version 1.1.02) (Zerbino and Birney 2008) with a k-mer of 29. The Oases extension (version 0.1.20) for *de novo* transcriptome assembly with Velvet was then run on the output.

All of the isotigs from Oases were combined with the JGI ESTs, screened for redundancy, and clustered with a 99% seed identity threshold using the UCLUST function of USEARCH (version 4.0.43) (Edgar 2010). Each of these clusters were aligned with CAP3 (X Huang 1999), requiring overlaps of at least 25bp with at least 96% identity, and the resultant contigs were combined. Sequences less than 500bp were removed, and the remainder formed a working assembly.

The assembly was assessed based on contig length, comparison to *Laccaria bicolor* proteins, and predicted gene content. Each contig was compared to the *L. bicolor* peptide collection by translated BLAST search, and the top hit for each contig was taken to assess percent identity, e-values, and alignment length. Gene predictions on both strands were made with Augustus (version 2.5.5) (Stanke et al. 2008) trained on *L. bicolor*. For each sequence with more than one gene prediction, the model with the highest percentage of coding sequence was chosen for comparison.

Since there was plant material from *P. andromedea* seeds incorporated into the samples, we used tblastx to compare our transcriptome assembly against the EST

collection of *Vaccinium corymbosum* (highbush blueberry, Ericaceae) from the Genome Database for *Vaccinium* ("Genome Database for Vaccinium").

Identification of Differentially Expressed Genes

Bowtie (version 0.12.7) (Langmead et al. 2009) was used to map reads to our transcriptome assembly. We used a seed size of 59 and allowed for up to three mismatches in the seed. As we expect the source organism to be a dikaryon as well as to have at least some alternative splicing, we did not discount reads that had more than one alignment. Requiring unique matches resulted in loss of 45% of the reads. Loss of data from poor mapping, especially of genes with multiple isoforms, could skew the results. Looking at the number of alignments per read, there did not appear to be a prevalence of numerous spurious matches for single reads.

DESeq (Anders and Huber 2010) was used to normalize read counts and to identify differentially expressed sequences. This statistical package uses a negative binomial model, which utilizes both mean and variance to model count data while also allowing application to data sets with relatively low replication. DESeq utilizes a more data-driven method of determining mean and variance than the edgeR package (Robinson, McCarthy, and Smyth 2010), which also uses a negative binomial model. The Poisson distribution has also been used to determine differential expression based on the mean of read counts for each gene and estimating variance from that (Marioni et al. 2008). However, this model can underestimate variation in biological replicates, leading to an increased false discovery (type I error) rate (Anders and Huber 2010; Turro et al. 2011). Given the long time (weeks) between our time points, we may expect substantial biological variation among replicates.

The top ten candidates with the lowest p-values under 0.01 were selected for further investigation. The sequences for these ESTs were used in a translated BLAST search of the *L. bicolor* peptide collection (v. 2.0, JGI). For the hit with the lowest e-value for each EST, functional annotation from GO terms, KEGG, KOG, InterPro, and SignalP for "best" filtered models were retrieved. All *L. bicolor* comparisons were done with genome version 2.0 from the Joint Genome Institute.

Results and Discussion

In order to work around not having a complete genome for *R. salebrosus*, we took advantage of the large amount of sequence generated by RNA-Seq itself. By combining these data with ESTs sequenced by JGI, we created a hybrid *de novo* transcriptome assembly. Excluding small contigs smaller than 500bp, the 35,249 contigs had a median length of 743 bp (Fig. 2). While many of these contigs are not large enough to span entire genes, they are of a size where one may expect to be able to identify portions of genes.

To determine the genic potential of the ESTs from our hybrid assembly, we compared them to the peptide collection of *L. bicolor*, which is currently the closest relative to *R. salebrosus* with a sequenced genome. The median percent identity of the top BLAST matches was 43.81% (Fig. 3), and the median e-value was less than 0.0001 (Fig. 4). Alignment lengths had a median of 133 (Fig.5). Given the distance between *R. salebrosus* and *L. bicolor*, it seemed reasonable that contigs in the assembly could represent ESTs. To determine how many of these putative ESTs resemble genes, we used Augustus to predict gene models. The percent coding sequence (CDS) was calculated for each gene model, and the gene model for each contig that had the highest percent CDS was considered the best. To be conservative, introns called by the program were not considered CDS, despite the fact that the sequences had originally come from mRNA. Of these 25,172 "best" gene models, 48.2% had both start and stop codons, and most of the 10,077 sequences that did not have any predictions were short sequences near 500 bp. The transcripts had a mean of 78.8% coding sequence (Fig. 6). As Augustus was trained on *L. bicolor*, these figures are probably an underestimation compared to if we were able to train the program on a closer relative to *R. salebrosus*.

As the samples collected from the interaction zone included both fungal and plant tissue, we looked for plant transcripts in the assembly. *V. corymbosum* is the plant closest to *P. andromedea* with a genome project initiative. A translated BLAST search of the hybrid assembly against the 3,703 available *V. corymbosum* ESTs resulted in many low e-value matches (median 0.12) and a median of 42% identity. However, these alignments were very short (median 30 bp). Since the median length of the *V. corymbosum* ESTs was 604 bp, this suggests that the amount of plant transcripts sequenced was far lower than fungal ones. While this is not a polished transcriptome, this hybrid assembly serves as a working reference for this first examination of *R. salebrosus* gene expression.

When mapping reads from each lane to the assembly, only 42% mapped uniquely. Since the assembly was constructed using all isoforms assembled by Oases, we expect some sequence overlap in the resulting ESTs. While many reads mapped twice, very few aligned more than three times. This was a reasonable number of alignments per read to account for isoforms, rather than many spurious constructs created in the assembly.

Interaction with the non-partner *R. arctostaphyli*-associated seeds induced very little difference from the control after one week (Fig. 8), with only three ESTs identified as differentially expressed with a p-value less than 0.01. Many more ESTs (425) were differentially expressed at two weeks (Fig. 9). The same pattern occurred with *R. salebrosus*-associated seeds, where 43 ESTs were differentially expressed after one week but 164 ESTs were identified after two weeks. This suggests that one week of interaction is early for *R. salebrosus* to react to the presence of living seeds, and that we were able to capture the initial stages of the symbiosis.

Previous studies with host specificity between *Paxillus involutus* and *Betula pendula* revealed a divergence in gene expression with different *P. involutus* strains as well as evolution in genes related to symbiosis (Le Quéré 2004). We also find differences in the expression profiles for *R. salebrosus* with partner and non-partner seeds. The comparisons between partner and non-partner seeds against controls for the first week had no shared differentially expressed ESTs. For the second week time point, there were 102 ESTs that were differentially expressed when *R. salebrosus* interacted with either partner or non-partner seeds.

To select ESTs to examine in greater detail, we filtered for p-values under 0.01 and then took the ten ESTs with the lowest p-values (Table 1). ESTs were found in several categories pertinent to establishment of ectomycorrhizal symbiosis: stress response, host defense, cellular transport, secretion, carbohydrate metabolism and transport, and signal transduction (Johansson et al. 2004; Morel et al. 2008). Mycorrhiza-induced small secreted proteins have been shown to be secreted into the apoplastic space within symbiotic root tips as well as being directed to the nucleus of host plant cells (Martin et al., 2008). The putative secreted proteins found in this study could be involved in the earliest interactions between the fungus and seeds. The putative sugar transporter that was upregulated in *R. salebrosus* with partner seeds is especially intriguing as mycoheterotrophs. Further, changes in carbohydrate metabolism have been implicated in involvement in ectomycorrhiza formation (Uwe Nehls et al. 2007). The trehalose phosphatase upregulated in week two with *R. arctostaphyli* seeds could be related to the increased trehalose levels found in ectomycorrhizae during symbiosis (Nehls et al. 2010).

Upregulation of genes involved in RNA processing in live seed conditions could indicate that the fungus is in the process of shifting its expression profile to deal with a new challenge. Changes in signal transduction likewise are expected, but unfortunately the putative protein kinases do not have detailed enough annotation to describe these pathways further. Genes involved in stress and defense have also been implicated in ectomycorrhizal symbiosis establishment (Johansson et al. 2004). Cerato-platanins, such as the one found in the interaction with non-partner seeds at week 1, compose a class of secreted fungal proteins implicated in plant pathogenesis as well as fungal growth and development (Zaparoli et al. 2009; Pazzagli et al. 2009).

Through this work we have produced a working *de novo* transcriptome assembly for further work on *R. salebrosus* gene expression. We have examined differential expression of *R. salebrosus* when it is interacting with seeds from partner (*R. salebrosus* type) and non-partner (*R. arctostaphyli* type) *P. andromedea*. We have identified putative genes in categories thought to be important for ectomycorrhizal symbiosis. To our knowledge, this is the first study looking at gene expression of any mycorrhizal fungus with a mycoheterotrophic plant.

Acknowledgements

We would like to thank C. Ellison, C. Villalta, and E. Purdom for advice on data analysis. Field collections were made at Blodgett Forest Research Station. This work was supported by grants from the Mycological Society of San Francisco, the Mycological Society of America, the California Native Plant Society, and UC Berkeley.

Literature Cited

Anders, Simon, and Wolfgang Huber. 2010. Differential expression analysis for sequence count data. *Genome biology* 11, no. 10 (January): R106.

Anon. Genome Database for Vaccinium. http://www.vaccinium.org/.

Bakshi, Ts. 1959. Ecology and morphology of Pterospora andromedea. *Botanical Gazette* 120, no. 4: 203-217.

Berch, S M, H B Massicotte, and L E Tackaberry. 2005. Re-publication of a translation of "The vegetative organs of Monotropa hypopitys L." published by F. Kamienski in 1882, with an update on Monotropa mycorrhizas. *Mycorrhiza* 15, no. 5 (July): 323-32.

Bidartondo, M I, and T D Bruns. 2001. Extreme specificity in epiparasitic Monotropoideae (Ericaceae): widespread phylogenetic and geographical structure. *Molecular ecology* 10, no. 9 (September): 2285-95.

---. 2002. Fine-level mycorrhizal specificity in the Monotropoideae (Ericaceae): specificity for fungal species groups. *Molecular ecology* 11, no. 3 (March): 557-69.

Bidartondo, M, and T Bruns. 2005. On the origins of extreme mycorrhizal specificity in the Monotropoideae (Ericaceae): performance trade-offs during seed germination and seedling development. *Molecular ecology* 14, no. 5 (April): 1549-60.

Bjorkman, Erik. 1960. Monotropa Hypopitys L. - an Epiparasite on Tree Roots. *Physiologia Plantarum* 13, no. 2 (April): 308-327.

Bruns, Thomas D, and David J Read. 2000. In vitro germination of nonphotosynthetic, myco-heterotrophic plants stimulated by fungi isolated from the adult plants. *New Phytologist* 148, no. 2 (November): 335-342.

Edgar, Robert C. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics (Oxford, England)* 26, no. 19 (October): 2460-1.

Egley, Grant H. 1990. High-Temperature Effects on Germination and Survival of Weed Seeds in Soil. *Weed Science* 38, no. 4/5: 429-435.

Frank, A B. 2004. On the nutritional dependence of certain trees on root symbiosis with belowground fungi (an English translation of A. B. Frank's classic paper of 1885) Structure of the mycorrhiza Development of the mycorrhiza. *Mycorrhiza*, no. May 1880: 1-9.

Gardes, M, and T Bruns. 1993. ITS primers with enhanced specificity for basidiomycetes-application to the identification of mycorrhizae and rusts. *Molecular Ecology* 2: 113–118.

Gashaw, Menassie, and Anders Michelsen. 2002. Influence of heat shock on seed germination of plants from regularly burnt savanna woodlands and grasslands in Ethiopia. *Plant Ecology*: 83-93.

Granstrom, A, and J Schimmel. 1993. Heat effects on seeds and rhizomes of a selection of boreal forest plants and potential reaction to fire. *Oecologia* 94 (May): 307-313.

Hillier, Ladeana W, Gabor T Marth, Aaron R Quinlan, David Dooling, Ginger Fewell, Derek Barnett, Paul Fox, et al. 2008. Whole-genome sequencing and variant discovery in C . elegans. *Nature Methods* 5, no. 2: 183-188.

Huang, X. 1999. CAP3: A DNA Sequence Assembly Program. *Genome Research* 9, no. 9 (September): 868-877.

Johansson, Tomas, Antoine Le Quéré, Dag Ahren, Bengt Söderström, Rikard Erlandsson, Joakim Lundeberg, Mathias Uhlén, and Anders Tunlid. 2004. Transcriptional responses of Paxillus involutus and Betula pendula during formation of ectomycorrhizal root tissue. *Molecular plant-microbe interactions : MPMI* 17, no. 2 (February): 202-15.

Kamienski, F. 1882. Les organs végétatifs du Monotropa hypopitys L. *Mémoires de la Societé Nationale des Sciences Naturelles et Mathématiques de Cherbourg* 24: 5-40.

Langmead, Ben, Cole Trapnell, Mihai Pop, and Steven L Salzberg. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology* 10, no. 3 (January): R25.

Le Quéré, Antoine, Andres Schützendübel, Balaji Rajashekar, Björn Canbäck, Jenny Hedh, Susanne Erland, Tomas Johansson, and Anders Tunlid. 2004. Divergence in gene expression related to variation in host specificity of an ectomycorrhizal fungus. *Molecular ecology* 13, no. 12 (December): 3809-19.

Marioni, John C, Christopher E Mason, Shrikant M Mane, Matthew Stephens, and Yoav Gilad. 2008. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome research* 18, no. 9 (September): 1509-17.

Martin, F, a Aerts, D Ahrén, a Brun, E G J Danchin, F Duchaussoy, J Gibon, et al. 2008. The genome of Laccaria bicolor provides insights into mycorrhizal symbiosis. *Nature* 452, no. 7183 (March): 88-92.

Merckx, Vincent, Martin I Bidartondo, and Nicole A Hynson. 2009. Myco-heterotrophy: when fungi host plants. *Annals of Botany* 104, no. 7: 1255-1261.

Morel, Mélanie, A Kohler, Francis Martin, Eric Gelhaye, and Nicolas Rouhier. 2008. Comparison of the thiol-dependent antioxidant systems in the ectomycorrhizal Laccaria bicolor and the saprotrophic Phanerochaete chrysosporium. *The New phytologist* 180, no. 2 (January): 391-407.

Nehls, U, F Göhringer, S Wittulsky, and S Dietz. 2010. Fungal carbohydrate support in the ectomycorrhizal symbiosis: a review. *Plant biology (Stuttgart, Germany)* 12, no. 2 (March): 292-301.

Nehls, Uwe, Nina Grunze, Martin Willmann, Marlis Reich, and Helge Küster. 2007. Sugar for my honey: carbohydrate partitioning in ectomycorrhizal symbiosis. *Phytochemistry* 68, no. 1 (January): 82-91.

Pazzagli, Luigia, Camilla Zoppi, Lara Carresi, Bruno Tiribilli, Francesca Sbrana, Silvia Schiff, Thelma a Pertinhez, Aniello Scala, and Gianni Cappugi. 2009. Characterization of ordered aggregates of cerato-platanin and their involvement in fungus-host interactions. *Biochimica et biophysica acta* 1790, no. 10 (October): 1334-44.

Robinson, Mark D, Davis J McCarthy, and Gordon K Smyth. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)* 26, no. 1 (January 1): 139-40.

Simard, Suzanne W, and Daniel M Durall. 2004. Mycorrhizal networks : a review of their extent , function , and importance 1: 1140-1165.

Simard, Suzanne W, David A Perry, Melanie D Jones, David D Myrold, Daniel M Durall, and Randy Molinak. 1997. Net transfer of carbon between ectomycorrhizal tree species in the field. *Nature* 388, no. August: 579-582.

Smith, Sally E, and David J Read. 2008. *Mycorrhizal Symbiosis*. 3rd ed. San Francisco: Academic Press.

Stanke, Mario, Mark Diekhans, Robert Baertsch, and David Haussler. 2008. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics (Oxford, England)* 24, no. 5 (March): 637-44.

Turro, Ernest, Shu-Yi Su, Angela Goncalves, Lachlan J M Coin, Sylvia Richardson, and Alex Lewin. 2011. Haplotype and isoform specific expression estimation using multi-mapping RNA-seq reads. *Genome Biology* 12, no. 2: R13.

Wallace, Gary D. 1975. Studies of the Monotropoideae (Ericaceae): Taxonomy and Distribution. *The Wasmann Journal of Biology* 33, no. 1 and 2: 1-88.

Zaparoli, Gustavo, Odalys García Cabrera, Francisco Javier Medrano, Ricardo Tiburcio, Gustavo Lacerda, and Gonçalo Guimarães Pereira. 2009. Identification of a second family of genes in Moniliophthora perniciosa, the causal agent of witches' broom disease in cacao, encoding necrosis-inducing proteins similar to cerato-platanins. *Mycological research* 113, no. Pt 1 (January): 61-72.

Zerbino, Daniel R, and Ewan Birney. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome research* 18, no. 5 (May): 821-9.

	DESeq		Best BLASTX Hit to *L. bicolor* Peptides				*L. bicolor* Functional Annotation	
EST Code	**log2 Fold Change**	**P-Value**	**Gene Code**	**% identity**	**Alignment Length (aa)**	**e-value**	**Annotation Code**	**Description**
A vs X, week 1								
hybrid_assembly_clust29963.0	-2.031	2.30E-04	172945	43.1	58	3.00E-05	IPR010829	Cerato-platanin
hybrid_assembly_clust4370.0	-1.822	1.55E-03	399525	63.95	147	2.00E-58	KOG0276 IPR011048	Vesicle coat complex COPI, beta' subunit Cytochrome cd1-nitrite reductase-like, C-terminal haem d1
hybrid_assembly_clust21296	-1.978	7.23E-03	691244	57.75	142	2.00E-39	SigP	hmm_signalpep_probability = 1
A vs X, week 2								
hybrid_assembly_clust12483.0	-1.375	1.05E-04	575112	27.92	265	2.00E-19	KOG1187	Serine/threonine protein kinase
hybrid_assembly_clust7707	-1.379	1.05E-04	690284	38	50	1.2	KOG2146	Splicing coactivator SRm160/300, subunit SRm160 (contains PWI domain); RNA processing
hybrid_assembly_clust30152	-1.357	1.07E-04	648053	28.43	102	0.005	None	
hybrid_assembly_clust1724	-1.409	1.12E-04	452239	31.02	374	2.00E-45	KEGG 2.7.7.7 KOG1060	DNA-directed DNA polymerase. Vesicle coat complex AP-3, beta subunit
hybrid_assembly_clust29786	-1.326	1.12E-04	629832	85.29	34	1.00E-11	IPR002680 GO3932	Alternative oxidase mitochondrial envelope
hybrid_assembly_clust643.0	-1.414	1.12E-04	306794	65.85	975	0	IPR003337	Trehalose-phosphatase
hybrid_assembly_clust8777.0	-1.338	1.12E-04	639901	38.44	372	1.00E-65	IPR011009	Protein kinase-like
hybrid_assembly_clust14646	-1.375	1.12E-04	648053	35.83	240	4.00E-34	no matches	
hybrid_assembly_clust20547	1.545	1.12E-04	385655	32.86	213	4.00E-28	IPR001128	Cytochrome P450
hybrid_assembly_clust9884	-1.356	1.14E-04	578170	22.42	165	0.26	KOG2071	mRNA cleavage and polyadenylation factor I/II complex, subunit Pcf11
S vs X, week 1								
hybrid_assembly_clust30870	-2.046	1.09E-04	252456	65.82	79	1.00E-24	KOG2144	Tyrosyl-tRNA synthetase, cytoplasmic
hybrid_assembly_clust11904	-1.872	1.91E-04	480240	46.57	277	7.00E-27	KOG2992	Nucleolar GTPase/ATPase p130
hybrid_assembly_clust18201	-2.443	2.86E-04	702552	43.41	205	6.00E-37	IPR003663 (GO) 6356	Sugar transporter L-arabinose isomerase activity

hybrid_assembly_clust27482	-1.970	3.39E-04	184382	74.58	59	2.00E-20	KOG1402	Ornithine aminotransferase
hybrid_assembly_clust34059	-1.747	6.78E-04	180140	69.44	108	2.00E-40	KOG1558	Fe2+/Zn2+ regulated transporter
hybrid_assembly_clust10823.0	-1.798	7.16E-04	440291	38.9	365	1.00E-66	KOG1579	Homocysteine S-methyltransferase
hybrid_assembly_clust17428	-1.702	7.95E-04	704679	71.62	222	6.00E-78	IPR001395	Aldo/keto reductase
							KOG1575	Voltage-gated shaker-like K+ channel, subunit beta/KCNAB
hybrid_assembly_clust19256	2.091	7.95E-04	486785	36.49	74	1.1	KOG0691	Molecular chaperone (DnaJ superfamily)
hybrid_assembly_clust18017.0	-2.323	1.41E-03	451323	43.41	129	3.00E-24	None	
hybrid_assembly_clust17340.0	-1.718	2.49E-03	678492	33.33	150	2.00E-13	None	
S vs X, week 2								
hybrid_assembly_clust19921	-1.330	1.04E-04	303116	32.43	37	1	KOG3511	Sortilin and related receptors
hybrid_assembly_clust31275	-1.643	1.08E-04	680613	73.24	71	2.00E-45	KOG1051	Chaperone HSP104 and related ATP-dependent Clp proteases
hybrid_assembly_clust28178	-1.296	1.27E-04	452239	34.33	67	0.14	KOG1060	Vesicle coat complex AP-3, beta subunit
							IPR011009	Protein kinase-like
hybrid_assembly_clust32624	1.535	1.27E-04	188777	39.44	71	1.00E-11	IPR007568	RTA1 like protein
							SignalP	Probability = 0.996
hybrid_assembly_clust15170	1.412	1.27E-04	694100	51.52	33	0.009	KOG0564	5,10-methylenetetrahydrofolate reductase
hybrid_assembly_clust35652	1.431	1.43E-04	669174	26	100	0.63	IPR002885	Pentatricopeptide repeat
hybrid_assembly_clust12153	-1.257	1.47E-04	300941	29.57	115	7.00E-06	None	
hybrid_assembly_clust17225.0	-1.231	2.04E-04	317898	25.64	78	0.43	None	
hybrid_assembly_clust20574	1.334	2.04E-04	396290	31.21	173	7.00E-20	KOG0158	Cytochrome P450 CYP3/CYP5/ CYP6/CYP9 subfamilies
							SignalP	Probability = 0.935
hybrid_assembly_clust16944	1.422	2.08E-04	620894	35.18	253	2.00E-32	None	

Table 1. Statistics and Functional Annotation for Differentially Expressed ESTs. The ten ESTs with the lowest p-values under 0.01 from DESeq are shown for each comparison of treatment with *R. arctostaphyli*-type seeds (A) or *R. salebrosus* type seeds (S) to heat-killed seeds after 1 or 2 weeks. The log2 fold change and p-values for DESeq comparisons are of normalized read counts. Functional annotations are from SignalP, InterPro (IPR), GO, KOG, and KEGG on "best" filtered models. SignalP probabilities are HMM signal peptide probabilities. For all comparisons to *L. bicolor*, genome version 2.0 from the Joint Genome Institute was used.

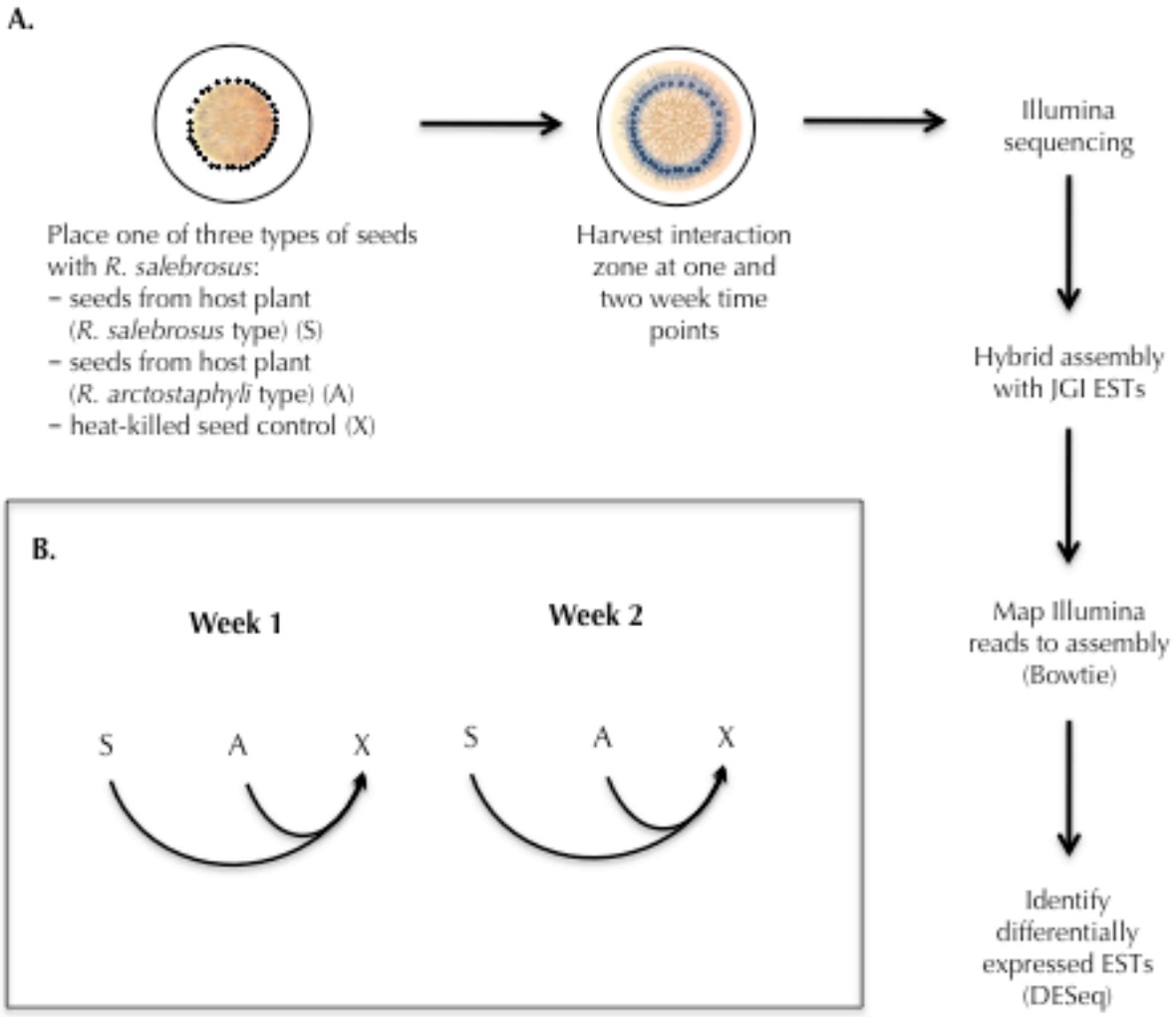


Figure 1. Experimental Set-up. A. Flow chart summarizing biological conditions, sample collection, sequencing, and data analysis. B. Three biological replicates from each experimental condition (*R. arctostaphyli* type seeds (A) or *R. salebrosus* type seeds (S)) were compared to heat-killed seed (X)) controls.

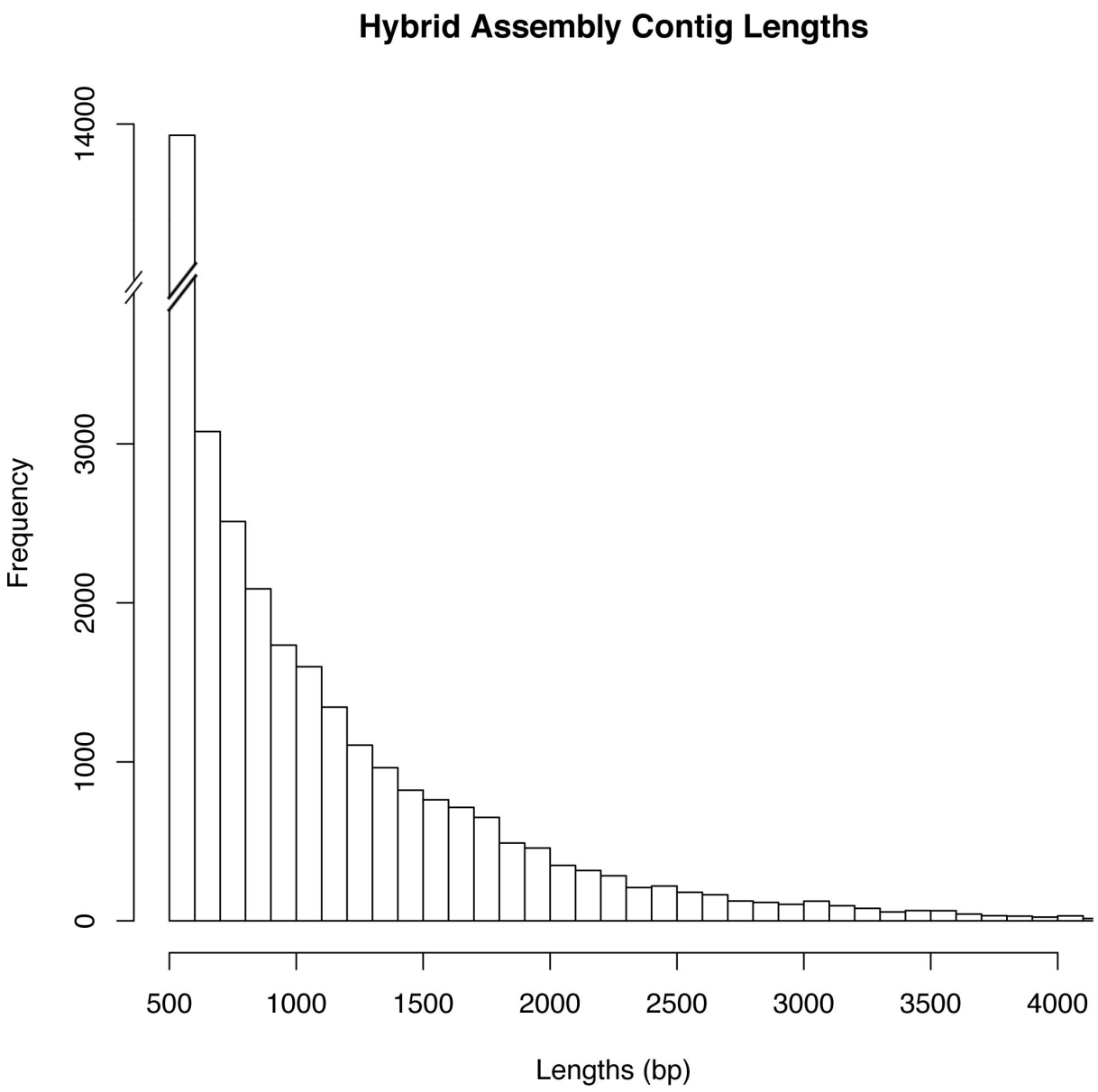


Figure 2. Contig Lengths for *De Novo* Hybrid Transcriptome Assembly. Seeds from *R. arctostaphyli* (A) type and *R. salebrosus* (S) type *P. andromedea* plants were placed with *R. salebrosus*. Four lanes of 76bp RNA-Seq Illumina from these conditions were assembled with Velvet/Oases, then clustered with UCLUST. The contigs resulting from each cluster were combined with ESTs sequenced by JGI and assembled with Newbler. The pooled sequences were assembled with CAP3. Contigs below 500 bp were discarded. There were 321 contigs over 4 kb (0.91% of contigs) not shown.

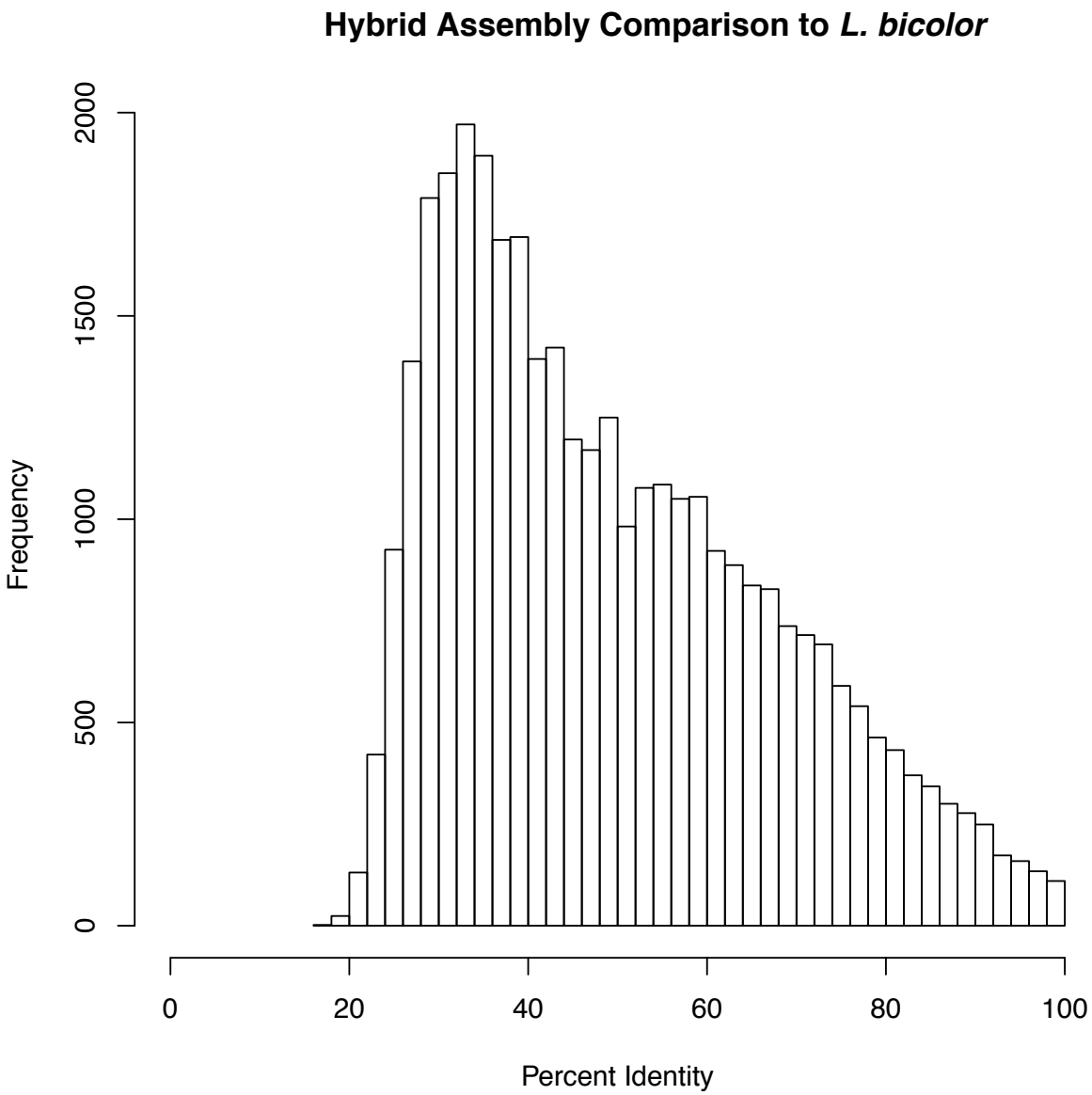


Figure 3. Percent Identity to *L. bicolor*. The hybrid assembly was compared to *L. bicolor* proteins using a translated BLAST search. The percent identity of the top hit for each contig was used.

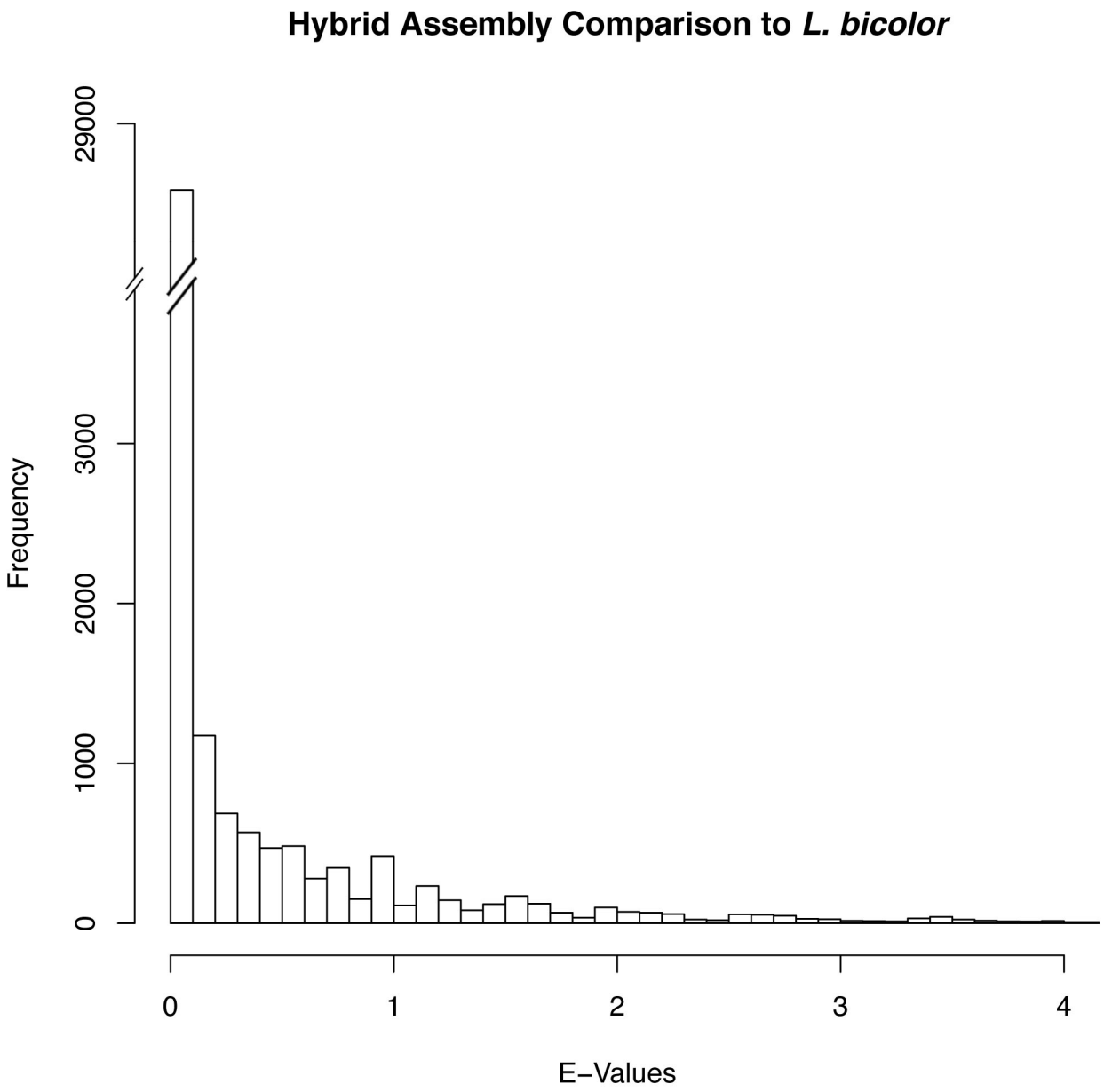


Figure 4. E-Values of Top BLAST Hits to *L. bicolor*. The hybrid assembly was compared to *L. bicolor* proteins using a translated BLAST search. The e-value of the top hit for each contig was used. There were 218 (0.62%) contigs with BLAST hit e-values over 4 that are not shown.

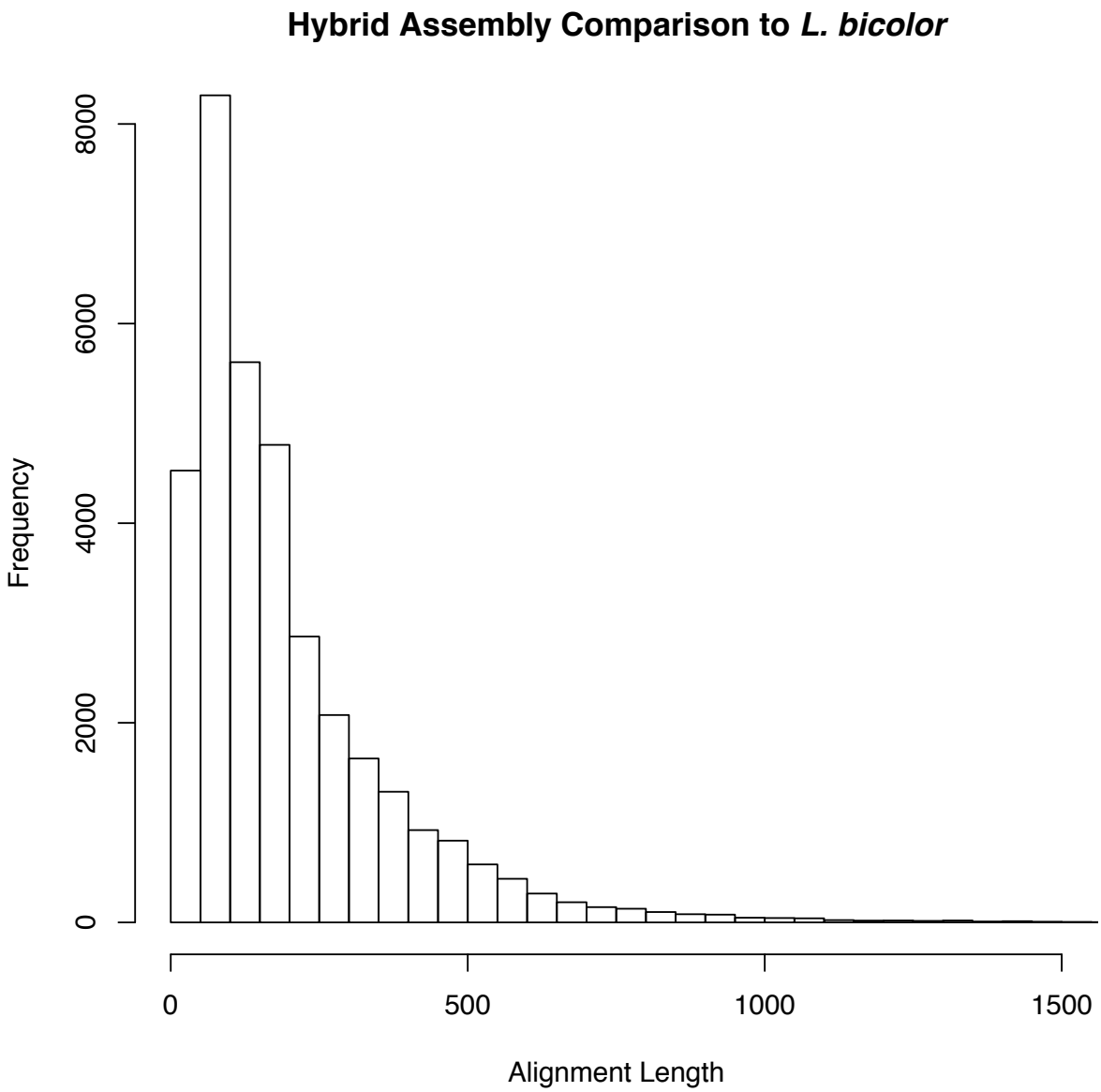


Figure 5. Alignment Lengths of Top BLAST Hits to *L. bicolor*. The hybrid assembly was compared to *L. bicolor* proteins using a translated BLAST search. The alignment length of the top hit for each contig was used. There were 44 contigs (0.12%) with BLAST hits with alignment lengths over 1.5kb that are not shown.

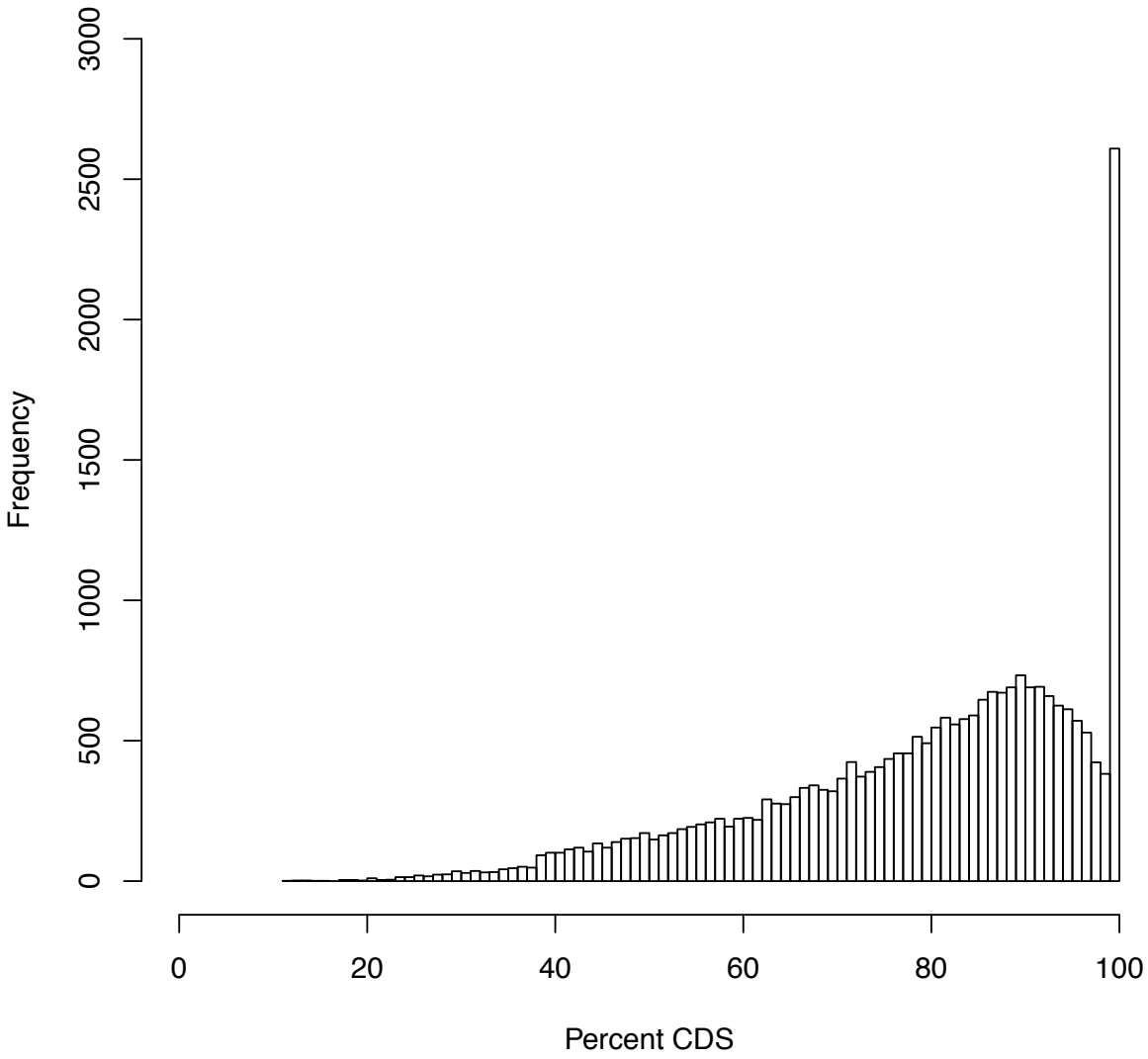


Figure 6. Predicted CDS of Augustus Gene Models. Gene predictions were made for the hybrid assembly using Augustus, and the percent of predicted coding sequence (CDS) was calculated for each contig.

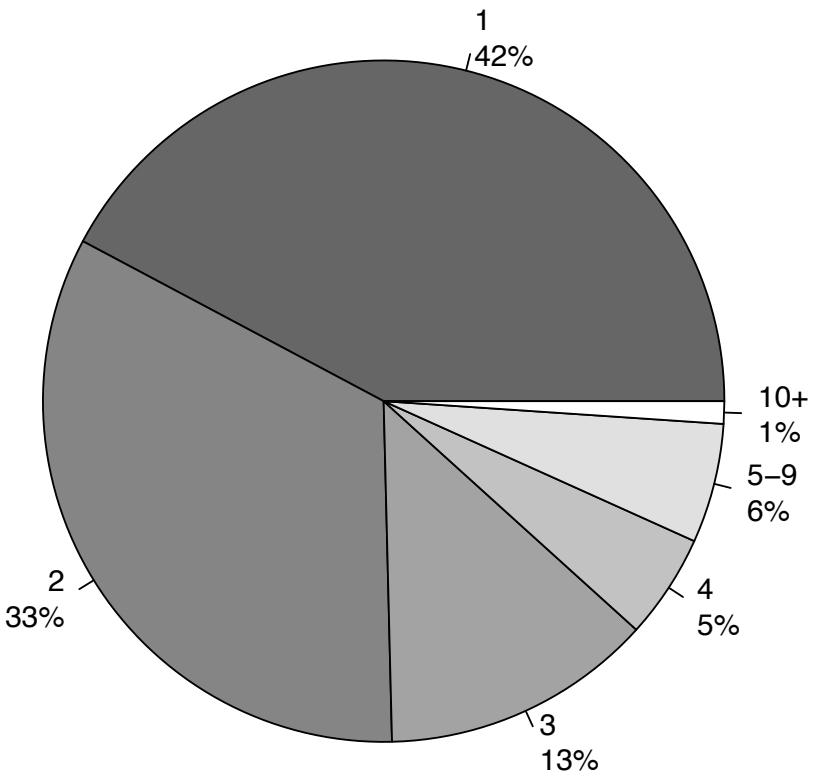


Figure 7. Alignments Mapped Per Read. Reads from each Illumina RNA-Seq lane were mapped to the hybrid assembly using Bowtie. Up to 50 alignments were allowed, but the maximum returned was 47. Labels indicate the number of times the reads in that slice mapped, as well as the percent of the pie. There were 546,366,512 total reads mapped.

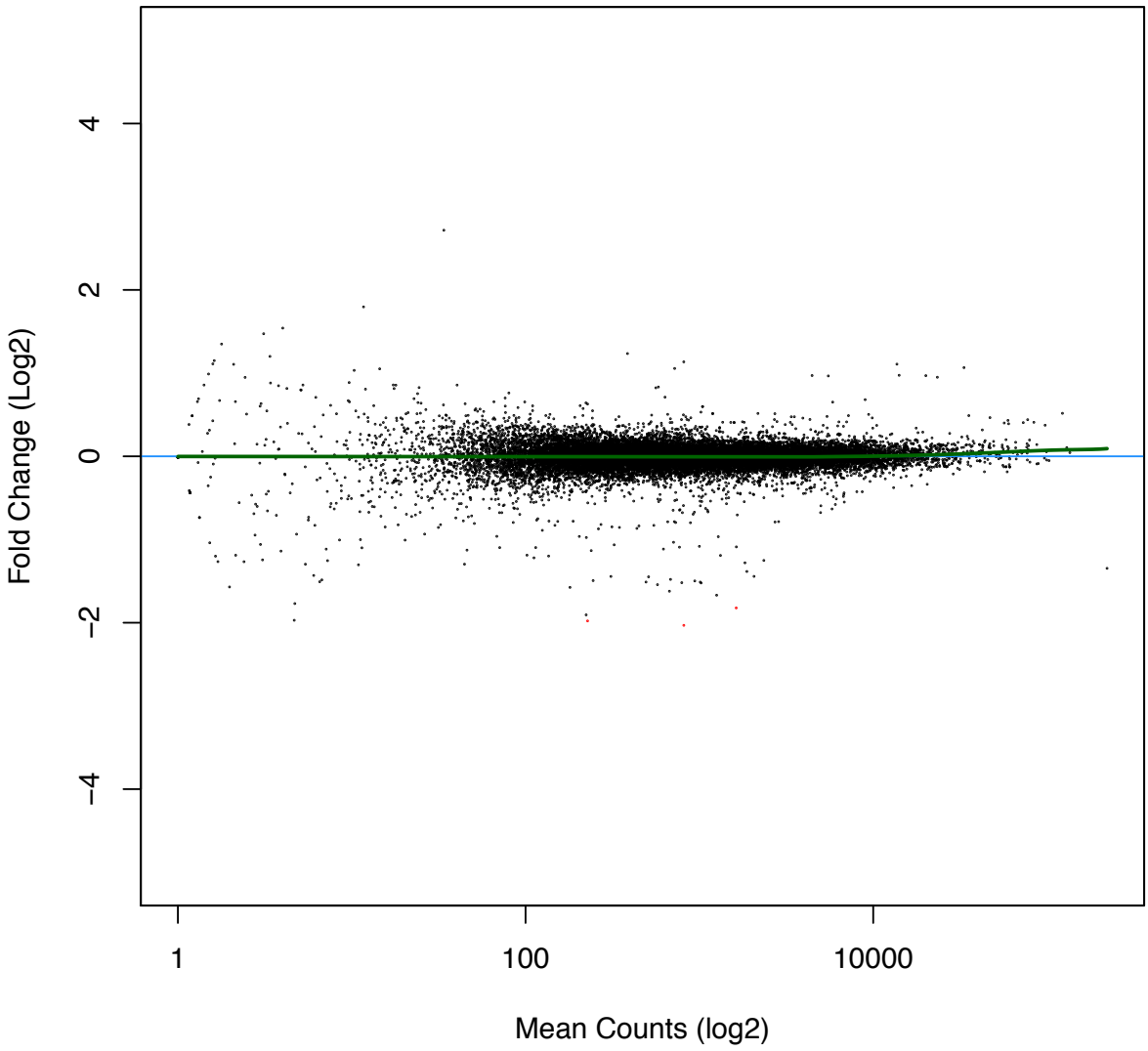


Figure 8. MA Plot of *R. arctostaphyli*-associated seeds (A) vs. Heat-Killed Control (X) After 1 Week. The interaction zone for three bioreps of each condition were used. Red points indicate differential expression with an adjusted p-value of less than 0.01. The green curve is a loess line. The blue horizontal line marks zero change in expression.

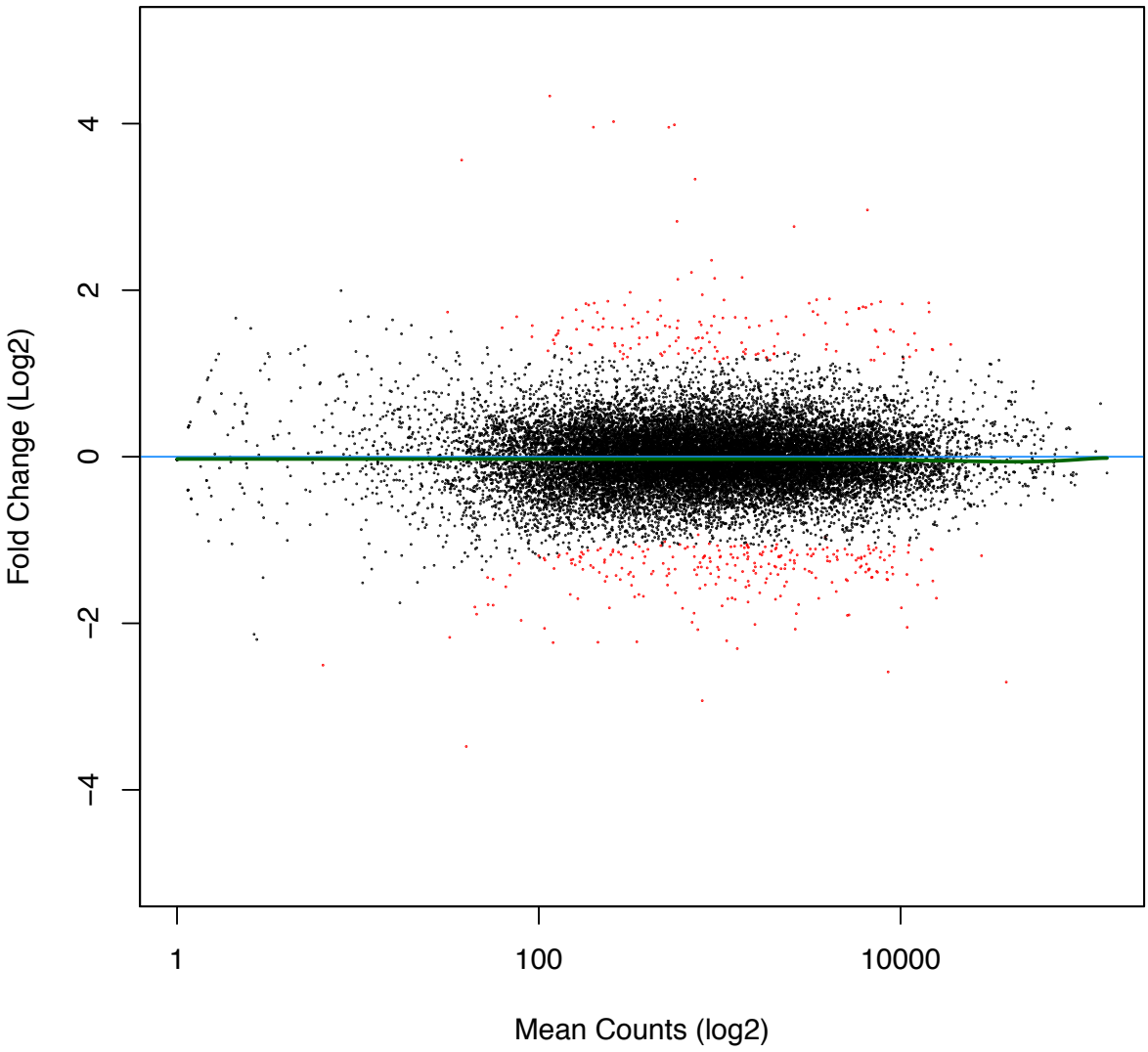


Figure 9. MA Plot of *R. arctostaphyli*-associated seeds (A) vs. Heat-Killed Control (X) After 2 Weeks. The interaction zone for three bioreps of each condition were used. Red points indicate differential expression with an adjusted p-value of less than 0.01. The green curve is a loess line. The blue horizontal line marks zero change in expression.

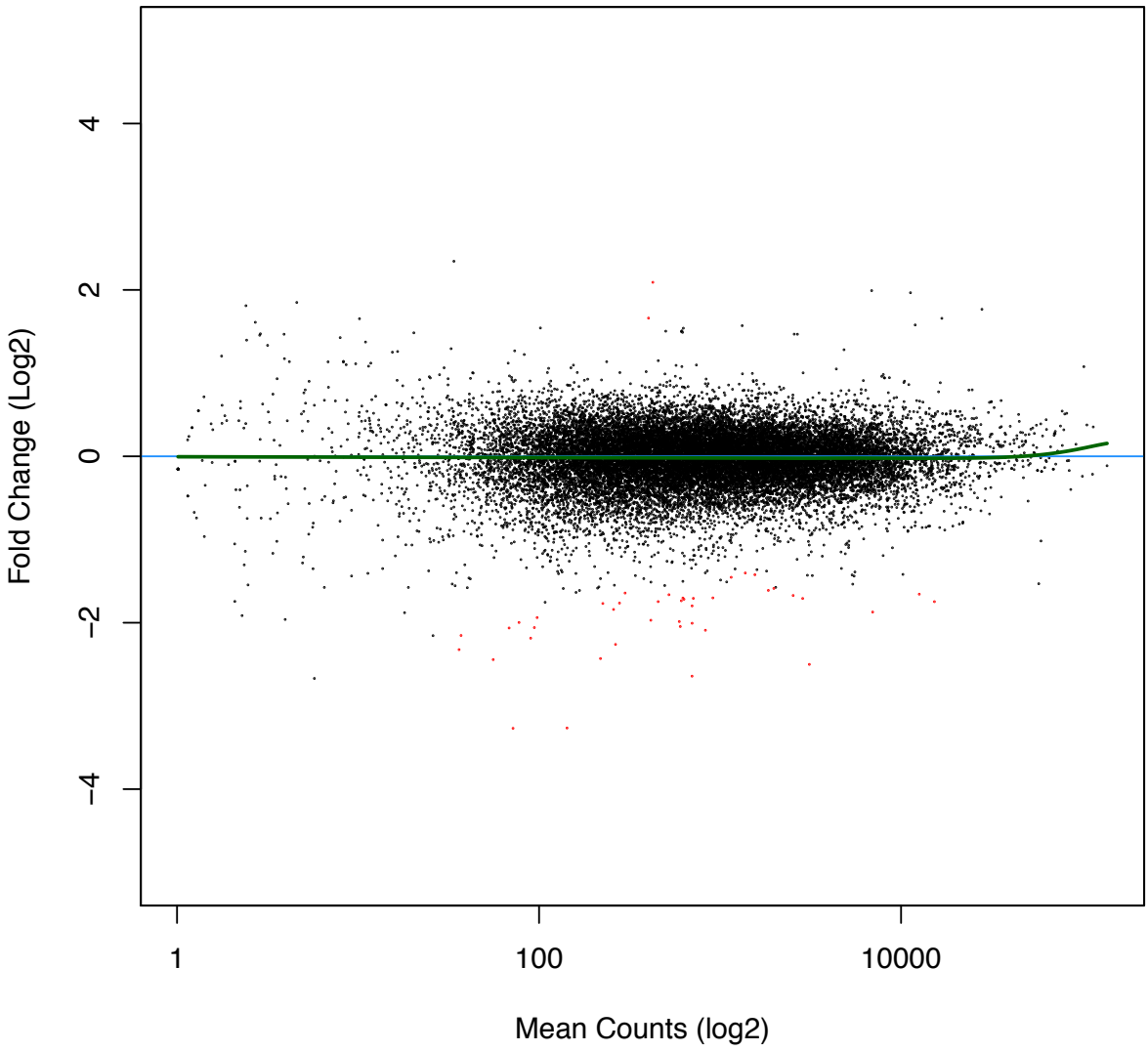


Figure 10. MA Plot of *R. salebrosus*-associated seeds (A) vs. Heat-Killed Control (X) After 1 Week. The interaction zone for three bioreps of each condition were used. Red points indicate differential expression with an adjusted p-value of less than 0.05. The green curve is a loess line. The blue horizontal line marks zero change in expression.

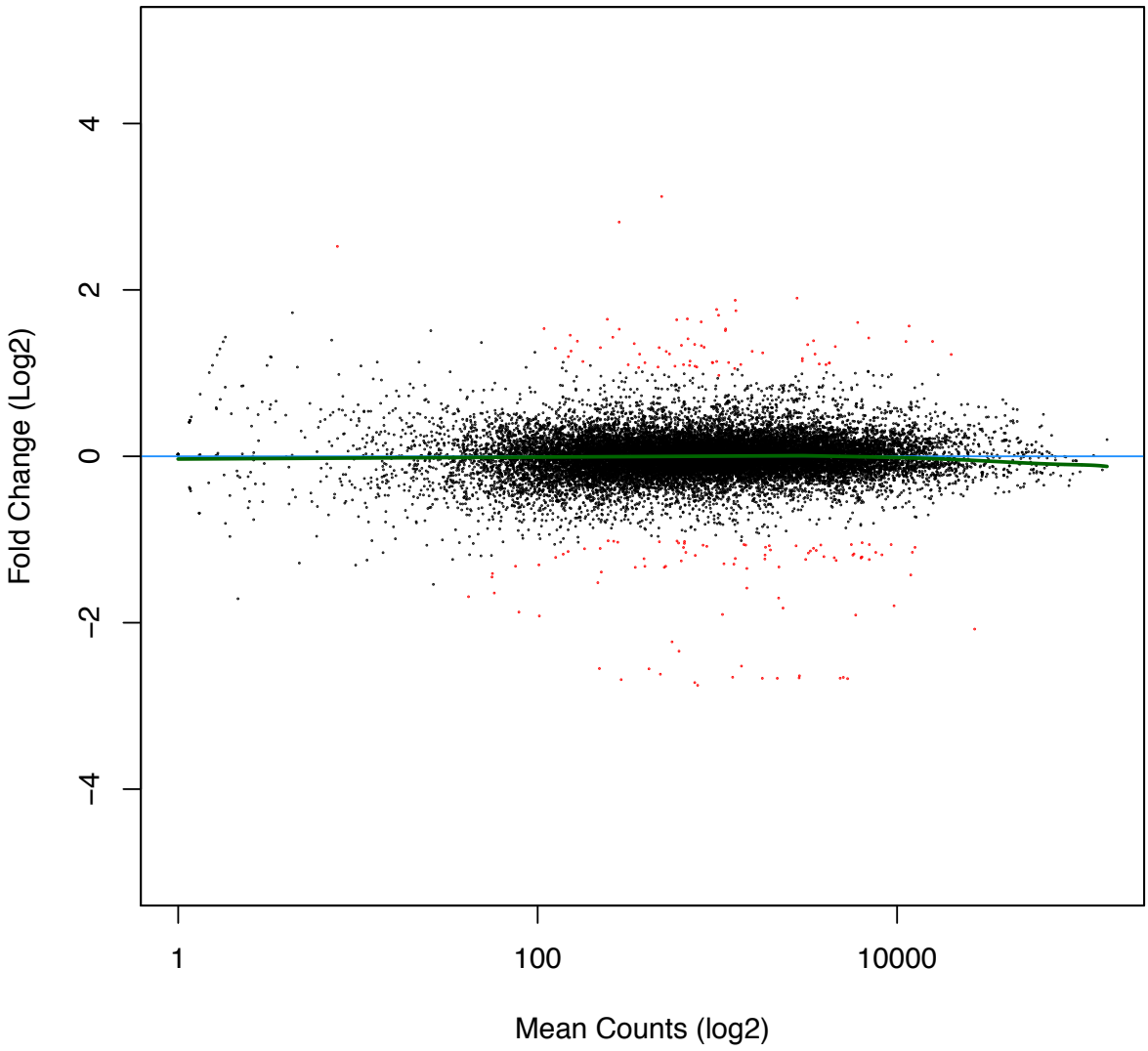


Figure 11. MA Plot of *R. salebrosus*-associated seeds (A) vs. Control (X) After 2 Weeks. Seeds of each type were placed just outside the margin of a *R. salebrosus* colony grown on a Petri dish. The interaction zone for three bioreps of each condition were used. Red points indicate differential expression with an adjusted p-value of less than 0.05. The green curve is a loess line. The blue horizontal line marks zero change in expression.