Effects of Transgenic Hybrid Aspen Over-expressing Polyphenol Oxidase on the Diversity of Rhizosphere Bacteria and Fungi

by

Kathryn Oliver

B.Sc., University of Victoria, 1999

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Biology

Supervisory Committee

Effects of Transgenic Hybrid Aspen Over-expressing Polyphenol Oxidase on the Diversity of Rhizosphere Bacteria and Fungi

by

Kathryn Oliver
B.Sc., University of Victoria, 1999

Supervisory Committee

Dr. W.E. Hintz, (Department of Biology)
Supervisor

Dr. C.P. Constabel, (Department of Biology)
Departmental Member

Dr. R. Roy, (Department of Biology)
Departmental Member

Dr. S. Misra, (Department of Biochemistry and Microbiology)
Outside Member

Supervisory Committee:

Dr. W.E. Hintz, (Department of Biology)
Supervisor

Dr. C.P. Constabel, (Department of Biology)
Departmental Member

Dr. R. Roy, (Department of Biology)
Departmental Member

Dr. S. Misra, (Department of Biochemistry and Microbiology)
Outside Member

Abstract

A greenhouse experiment was carried out to screen for potential effects of transgenic aspen over-expressing a hybrid poplar leaf polyphenol oxidase gene on rhizosphere communities. Heterotrophic plate counts and cultivation-independent methods were used to compare bacterial and fungal populations associated with transgenic PPO over-expressing and unmodified control trees. Total community DNA extracted from rhizosphere soils was used to establish libraries containing partial gene sequences that were PCR-amplified from community members, and putative taxonomy was assigned to clones based on similarity to reference sequences.

Gene libraries for the bacterial component of the rhizosphere were established using partial 16S rRNA and chaperonin-60 gene sequences, and the fungal community was characterized based on partial 18S rRNA gene sequences. Phylogenetic analysis revealed that bacterial 16S gene libraries were dominated by Alphaproteobacterial sequences, and the CPN-60 gene libraries were dominated by members of the Bacteroidetes/Chlorobi group, illustrating the biases potentially incurred by using a single gene locus to profile microbial diversity. In both CPN-60 and 16S rRNA libraries, only minor components of the bacterial community differed between transgenic and unmodified trees. Comparisons based on library coverage indicated that changes in bacterial community structure between transgenic and unmodified trees were minor in comparison to differences observed between individual trees of the same type, and no significant differences in terms of bacterial species diversity were revealed by the calculated diversity, dominance and evenness indices. In comparison to the bacterial

gene libraries, higher coverage of the underlying population was achieved in the fungal 18S libraries. Members of the Zygomycota, Chytridiomycota, Ascomycota, and Basidiomycota were recovered from both libraries. Dominant groups of fungi associated with each tree type were highly similar, although there were some qualitative differences in the recovery of less abundant fungi as a result of the underlying heterogeneity of the fungal population. No clear differences in terms of fungal species richness were associated with transgenic or unmodified trees, although control libraries were characterized by a slightly higher level of dominance. In general, the methods employed revealed only minor differences between the bacterial and fungal communities associated with transgenic and unmodified trees, suggesting that impacts of the transgenic plants on the rhizosphere community were minimal.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ARDRA	Amplified ribosomal DNA restriction analysis
CFU	Colony forming unit
CPN-60	Chaperonin-60
DGGE	Denaturing gradient gel electrophoresis
DOPA	Dihydroxyphenylalanine
FAME	Fatty acid methyl ester
MEA	Malt extract agar
NA	Nutrient agar
OTU	Operational taxonomic unit
PCR	Polymerase chain reaction
PPO	Polyphenol oxidase
PVPP	Polyvinylpolypyrrolidone
RDP	Ribosomal Database Project
rRNA	Ribosomal RNA
SSCP	Single strand conformation polymorphism
T4-L	T4-lysozyme
TGGE	Temperature gradient gel electrophoresis
UT	Universal target

Acknowledgements

I would like to thank my supervisor, Dr. Will Hintz for the opportunity to pursue graduate research in his lab, as well as my committee members, for their guidance during my studies. I would also like to acknowledge Melissa Hills, Brad Binges, Lan Tran, Dmytro Yevtushenko, Glenn Cooper, and Mike Wilson for assisting with technical details, and Paul de la Bastide and Meghan Chabot for being stellar co-workers. This project was carried out as part of a research program aimed at monitoring biodiversity and horizontal gene transfer in GMO associated communities, which was supported by Canadian Biotech Strategic Funds awarded to R. C. Hamelin, CFS-Quebec, (Project leader); A. Séguin, CFS-Quebec, (GM-Trees); R. Brousseau, BRI-NRC, (Microarrays); W. Hintz, UVIC, (Sequence analysis); and P. Hébert, Guelph, (Insects). I extend great appreciation to Richard Hamelin for providing the opportunity for this collaborative research.

Finally, I would like to put forth a huge thank you to my friends and family, for helping me to keep my head on my shoulders, and my feet firmly planted on the ground for the last couple of years. KJ, Ritchie, Laurie, Keith, Trish, Josh, Matt, Alyson, Dad, Grandma, Dave, Kim, and Brad: Thank you all for your encouraging words, your laughter, patience and love.

Introduction

1.1 Status of transgenic trees

Transgenic crops are currently important to the world wide economy, and are being planted at an increasing rate. In 2005, the estimated global area for genetically modified crops was 90 million hectares, an 11% increase from 2004 (James, 2005). The ability to create transgenic tree species may likewise offer benefits to the forest sector. Numerous field trials testing the commercial potential of genetically modified trees have been initiated in the USA, New Zealand, and Europe (van Frankenhuyzen and Beardmore, 2004). In China, large scale plantations of transgenic insect-resistant poplar (*Populus nigra*) expressing a *Bacillus thuringiensis* toxin gene have been established for commercial purposes (F.A.O., 2004). Thus far, the only transgenic tree commercially released in the USA is the Papaya, (*Carica papaya*), which is resistant to the papaya ring spot virus (Luis *et al.*, 1997), and as of the fall of 2005, the Canadian Forest Service was the only organization in Canada to have field tests of transgenic trees.

Research in the field of forest biotechnology has been directed to the development of model systems for basic research, and the improvement of the commercial qualities (i.e. growth rate, wood quality, and resistance to pests) of economically important tree species. Ultimately, research efforts seek more efficient land use and improved yield (Tang and Newton, 2003; van Frankenhuyzen and Beardmore, 2004). The contribution of basic forest biotechnology research is of particular significance in Canada, as the Canadian forest industry currently produces close to 1 billion seedlings each year. Most of these seedlings are grown in large containerized nurseries, where soil-borne pathogens and insect predation are major contributors to losses (Nevill *et al.*, 1995). Plantations currently supply about 12% of the world's wood, and will likely play a major role in future global sustainability. However, these plantations, comprised of fast-growing elite genotypes may also be particularly susceptible to disease (Fenning and Gershenzon, 2002). Thus, trees engineered to produce anti-microbial or insecticidal peptides may

reduce bacterial and fungal diseases and losses to insect pests, providing alternatives to chemical pesticide application.

The engineering of genetically modified plants has great potential for future agriculture and forestry, but also asks for well defined risk assessment. Concerns surrounding the use of genetically modified plants include the potential development of resistance in target pest organisms, gene flow through pollen transfer, horizontal gene transfer with micro-organisms, and the risk of invasion and subsequent reduction of native flora and fauna (Wolfenbarger and Phifer, 2000; Dunfield and Germida, 2004). This study addresses the concern that non-target and beneficial organisms within the plant rhizosphere may be affected by the over-expression of a hybrid poplar polyphenol oxidase gene in transgenic hybrid aspen. Although many physiological functions have been proposed for polyphenol oxidases (Vaughn and Duke, 1984), there is strong evidence that these enzymes play a role in plant defence against pathogens (Li and Steffens, 2002) and polyphenol oxidase over-expression may confer increased resistance to insect herbivory in *Populus* species (Wang and Constabel, 2004).

1.2 The rhizosphere community as a target for impact assessment

The health and productivity of plants is greatly dependent on root-associated microbial communities, which are comprised of a range of fungi and bacteria whose interactions with each other and with the tree may be beneficial or detrimental to plant health. Referred to as the rhizosphere community, these organisms are heavily influenced by root growth and metabolism (Schröder and Hartmann, 2003), and include fungi and bacteria that suppress plant pathogens through antibiotic production, competition for nutrients and induced resistance (Horton and Bruns, 2001, Whipps, 2001). Furthermore, some root-associated organisms positively influence plant growth and can aid in the acquisition of essential mineral nutrients and water (Morgan *et al.*, 2005). Changes in rhizosphere communities induced by genetically modified plants therefore are of concern since they play key roles in maintaining plant health.

Most environmental risk assessment strategies have focused on aboveground impacts, due in part to inherent challenges associated with studying soil-borne microbes, but the

rhizosphere community is a compelling target for risk-assessment strategies for several reasons. First, as primary consumers of plant exudates and material (Grayston *et al.*, 1996), rhizosphere communities are likely to respond to changes in plant activity and metabolism. Secondly, soil microbial diversity may be indicative of overall ecosystem functioning, since soil microbes play essential roles in the decomposition of organic compounds, nutrient transformations, and the suppression of soil-borne diseases (Dunfield and Germida, 2004, Wolfenbarger and Pfifer, 2000). Recent advances in molecular genetic approaches in conjunction with well-represented databases of fungal and bacterial genes now permit a rapid assessment of diversity in the soil microbial community and provide tools for impact assessments.

1.3 Methods used to assess microbial diversity

For many years, the diversity of microbial communities has been investigated using methods based on isolating and culturing fungi and bacteria, but it has been shown that a large proportion of soil microbes cannot be cultured on synthetic media (Kent and Triplett, 2002). Molecular methods have improved the detection of organisms that cannot be cultured *in vitro*. Most of these methods are based on the amplification of specific targets from genomic DNA using Polymerase Chain Reaction (PCR) technology, followed by a step to separate PCR products amplified from separate organisms. Methods used to separate DNA fragments, including denaturing gradient gel electrophoresis (DGGE), temperature gradient gel electrophoresis (TGGE), and amplified ribosomal DNA restriction analysis (ARDRA), allow rapid comparative community analysis (Kent and Triplett, 2002, Ranjard *et al.*, 2000). A molecular cloning approach was considered most suitable for this study, as the cloning and sequencing of specific targets within microbial genomes allows identification of individual members of a community and offers the most concise resolution of rhizosphere community structure. The most commonly used gene target for studying broad diversity is the 16S ribosomal RNA (rRNA) gene for bacteria and the 18S rRNA gene for fungi (Ranjard *et al.*, 2000, Filion *et al.*, 2004, Kent and Triplett, 2002).

Methods based on ribosomal RNA genes:

Ribosomal RNA molecules are present in all organisms and have a high degree of structural and functional conservation. Portions of the gene sequence are highly conserved, with interspersed hypervariable regions. Many primers, ranging from highly specific to universal, have been developed that amplify portions of rDNA, and a constantly expanding library of over 260,000 16S rDNA sequences is stored in a public database, the Ribosomal Database Project (http://rdp.cme.msu.edu/) (Cole *et al.*, 2005). The ability to amplify rRNA genes from environmental samples has allowed researchers to generate molecular profiles of soil microbial communities and these techniques have been used in a variety of studies aimed at determining the effects of transgenic plants on soil microbial communities. The tools that have been developed to analyze 16S data, as well as the wealth of 16S sequence information makes it a compelling candidate target for impact assessment studies including the experiment detailed herein.

Variation within the 16S rRNA gene sequence generally allows clear resolution of prokaryotes above the taxonomic rank of species (Rosselló-Mora and Amann, 2001), and methods based on the amplification of portions of this target using PCR have greatly improved assessments of microbial diversity. These methods are not without limitations however, since each step involved in the molecular genetic analysis of an environmental sample, from initial sampling, to cell lysis and PCR amplification, is potentially a source of bias (Wintzingerode *et al.*, 1997). It is acknowledged that these tools cannot be used to accurately and exhaustively describe the great diversity of soil microbial communities, but they do allow comparison of relative diversity between different treatments and communities. It has been suggested that the biased view that results from using a single gene target in community and phylogenetic studies of microbial diversity may be somewhat ameliorated by using a second target, thereby achieving a more accurate overall assessment of the microbial community.

Methods based on the Chaperonin-60 gene:

The gene encoding the 60kDa protein subunit of Type I chaperonins has been used as an alternate target for microbial identification and phylogenetic studies. The chaperonin-60 (CPN-60) gene is found in most prokaryotes and eukaryotes (Hill *et al*, 2002), an essential feature of a phylogenetic marker intended for assessing the diversity of the microbial community. Evolutionary trees based on CPN-60 sequences have been found to be similar to 16S rDNA trees, and can provide resolution in some cases where relationships are not clearly defined by variation in 16S rDNA sequences (Viale *et al.*, 1994, Brousseau *et al.*, 2001). Type I Chaperonins appear to have numerous cellular functions, including the post-translational folding and assembly of protein complexes, and intercellular signalling (Maguire *et al.*, 2002, Saibil and Ranson, 2002). Since these proteins perform conserved housekeeping roles and are essential for cell functioning, there is less chance for random mutation or intraspecies sequence variation in the CPN-60 gene (Martin *et al.*, 1993). In contrast to 16S gene sequences, which may be present in multiple copies (Wintzingerode *et al.*, 1997), there is generally a single copy of the CPN-60 gene in prokaryotic genomes (Hill *et al.*, 2004). Using the CPN-60 gene target may reduce the overestimation of diversity, which may result if divergent sequences from single bacteria which possess multiple rRNA copy numbers are included in the analysis of microbial communities. Variable regions are interspersed with highly conserved sequence regions in 16S rRNA genes, resulting in a highly stable secondary structure that is suspected to reduce PCR amplification efficiency (Wintzingerode, 1997). In contrast, variation extends uniformly throughout the CPN-60 coding region, and the highly stable secondary structure associated with 16S rRNA is not present (Hill *et al.*, 2004). The full sequence encoding the Chaperonin-60 gene is approximately 1700 bp long (Hill *et al.*, 2004), and the "universal target" (UT) is a portion of this gene, ranging from 549-567 base pairs in length, which can be amplified using universal, degenerate primers corresponding to nucleotides 274-828 of the E. coli CPN-60 sequence (Goh *et al.*, 1996). This portion of the CPN-60 gene has been used in the identification of organisms and in a limited number of microbial community studies. The sequence variation within the UT region has been shown to reflect the variation in the entire coding sequence (Hill *et al.*,

2004). Prokaryotic and eukaryotic chaperonin-60 gene sequences encompassing the UT region are catalogued in the Chaperonin Database (http://cpndb.cbr.nrc.ca), which currently contains approximately 8,000 records derived from reference strains, clinical isolates, field isolates, and microbial population studies.

1.4 Effects of genetically modified plants on associated microbial communities

There are numerous ways that transgenic plants may affect soil microorganisms including both direct and indirect effects of the modified plant. As demonstrated by monitoring the fate of the toxin produced by transgenic corn cultivars expressing the *Bacillus thuringiensis* cry1Ab gene, transgenic proteins present in plant tissue may contact microbial organisms through the sloughing of root cells, through root exudation, or through decomposing leaf tissue (Saxena and Stotzky, 2000). In cases where transgene products have antimicrobial properties intended for disease resistance, rhizosphere communities may be influenced directly if any transgenic products are released into the soil. The extent of direct effects depends on the spectrum of activity of the specific protein (Oger *et al.*, 1997) and its levels of accumulation in the soil (Lui *et al.*, 2005). Transgenic plants may also exert effects indirectly through changes in normal plant protein and root exudate composition that may arise as a result of modifying metabolic pathways (Lui *et al.*, 2005).

A key greenhouse study by Oger *et al.* (1997) clearly demonstrated that transgenic plants can temporarily alter soil microbial diversity. A legume, *Lotus corniculatus*, was specifically engineered for this experiment to produce mannopine, a compound used by specific bacterial groups as an energy source. Culture based methods revealed that levels of mannopine-utilizers were 80 times greater when associated with mannopine producing plants, but that no difference was detected in numbers of total culturable bacteria and other screened groups. Following removal of the transgenic plants from the soil, the population and activity of mannopine-utilizers remained elevated for as long as four months, after which the soil system returned its original state. Since the changes were temporary, it was concluded that the effects caused by the plants on the

soil ecosystem were not significant. In this experiment, mannopine-utilizers were easily identifiable as target organisms for monitoring. Oger *et al.* (1997) suggested that the identification of groups likely to be affected is necessary for successful monitoring, and that the effects on microbiota are highly specific to each transgenic product. In practice however, many commercially relevant novel traits, such as herbicide tolerance or pathogen resistance, differ in the range of soil system organisms that they might affect, and the question remains as to whether these effects are detectable if target organisms are less well defined. When a broad effect of the transgenic product is predicted, accurate analysis demands that the effects of genetically modified plants are examined more generally (Kowalchuck *et al.*, 2003).

Impact assessments have been carried out for two main classes of engineered crop plants:

a) plants engineered to resist pathogens, and
b) plants engineered to tolerate herbicide application (Kowalchuck *et al.*, 2003)

Examples of these studies are summarized below in order to illustrate the range of approaches used and the results obtained.

Plants engineered to have pathogen resistance:

A series of studies targeting specific groups of micro-organisms and overall community structure have been based on transgenic T4-lysozyme (T4-L) producing potato plants. Plants were engineered to resist the pathogen *Erwinia carotovora* ssp. *altroseptica*, the causal agent of black-leg and soft rot of tubers. T4 lysozyme (T4-L) is a bacteriolytic enzyme that has been shown to affect a range of soil and plant associated bacteria (Lottmann, 1999, Arenholtz *et al.*, 2000) and the expression of this lysozyme was the first trait that was exploited to confer increased resistance to pathogens.

An initial study suggested minor effects on the relative abundances of cultivatable species in microbial communities on the leaf surfaces of T4-L producing plants but these differences were deemed insignificant compared to the natural variability observed in several samplings (Heuer and Smalla, 1999). In the same experiment, analysis of substrate utilization patterns were used to monitor for changes in groups of

microorganisms differing in their abilities to catabolize various carbon sources, but no differences were observed. Similarly, differences in community were not captured by DGGE analysis of 16S rDNA fragments, possibly because bands resulting from several species were clustered together, rendering the various species indistinguishable. A downfall of DGGE analysis is that profiles generated by highly complex microbial communities are often characterized a by a few dominating bands, and a diffuse background of unresolved fragments (Lukow *et al*, 2000).

Studies focusing on specific groups of root-associated micro-organisms have detected minor differences associated with T4-L transgenic plants. Lottmann *et al.* (1999) monitored potentially beneficial plant associated bacteria, and using culture-based methods, detected fewer antagonistic species associated with T4-L producing plants. These differences were minor relative to the natural variability observed during the two-year monitoring period. In a subsequent study two bacterial strains with antagonistic properties were introduced as biocontrol agents into the rhizosphere community, one strain being T4-L tolerant, and the other T4-L sensitive. Colony counts revealed an increased level of colonization by the lysozyme tolerant antagonistic bacteria associated with T4-L producing plants (Lottman *et al.*, 2000)

In a multi-tiered field study investigating the effect of field grown T4 producing potato plants on the diversity of the rhizosphere community, one of two transgenic lines tested was found to have differences in microbial community structure. These differences were detected by each method used: fatty acid analysis, substrate utilization profiles, DGGE analysis of 16S rDNA fragments, and by cloning and sequencing of specific organisms. The changes in microbial community however, were attributed to the growth characteristics of the transgenic line, which displayed a general weakness as a result of the transformation, and not to T4-L production, since T4-L levels were similar in both transgenic lines (Heuer *et al.*, 2002).

Crop plants engineered to have herbicide tolerance:

There are currently four main commercialized transgenic crops (soybean, maize, cotton, and canola), engineered to tolerate application of specific herbicides (James,

2005). Specific microbial groups likely to be affected by this trait are not easily identifiable, and the majority of studies have examined the effects on the entire soil microbial community.

Community analyses based on transgenic canola cultivars conferring either glufosinate or glyphosate tolerance detected a lower diversity of rhizosphere bacteria associated with glyphosate-tolerant varieties (Siciliano and Germida, 1999; Dunfield and Germida, 2001). These observations were based on fatty acid methyl ester (FAME) analysis and carbon utilization patterns. It is speculated that changes in root exudation of the glyphosate-tolerant plants resulted in the differences. A two year field study conducted over the course of a growing season indicated that changes to community structure were not permanent, since community structure was observed to return to its original state after winter (Dunfield and Germida, 2003).

A community-level analysis comparing glufosinate-tolerant maize to a non-transgenic cultivar having the parental genotype, did not reveal differences between profiles generated by PCR-SSCP (Single-strand conformation polymorphism) of amplified 16S rDNA. Sequencing of dominant SSCP bands indicated the presence of typical soil and rhizosphere bacteria (Schmalenberger and Tebbe, 2002). In contrast however, a study that compared DGGE profiles of eubacterial and *Pseudomonas* communities associated with the roots of glufosinate-tolerant oilseed rape did reveal that communities associated with the transgenic plants differed from the non-transgenic plants (Gyamfi *et al.*, 2000).

Impacts of transgenic plants on fungal communities:

Molecular studies investigating the impacts of transgenic plants have primarily focused on monitoring for changes in bacterial community composition (reviewed in Cartwright and Lilley, 2004), probably because bacteria are considered to be the primary consumers of the simple organic sugars that make up root exudates, whereas fungi generally decompose more complex substrates such as cellulose and lignin (de Boer *et al.*, 2005). Changes induced by altered root exudates are hypothesized to be pronounced in the bacterial community, but recent observations suggest that under certain conditions,

fungi may be better competitors for simple substrates than previously thought (Butler, 2003). Changes in root exudates may also be expected to induce changes in the fungal community since some released compounds act as host specific recognition signals for a range of symbiotic, associative, and pathogenic fungi (Grayston *et al.*, 1996).

Most reports of effects of transgenic plants on fungi have focused on mycorrhizal symbioses. Expression of antifungal pathogenesis related proteins in transgenic tobacco did not impact root colonization by the mycorrhizal fungus *Glomus mossae*; however reduced colonization of ß-1-3 glucanase over-expressing plants indicated that non-target effects on beneficial fungi do sometimes occur (Vierheilig *et al.*, 1995). Although challenging, it is also important to monitor for unexpected changes in fungal diversity, especially in view of the complex interactions that occur in the rhizosphere, and the apparent (albeit poorly understood) role of the rhizosphere in suppression of soil borne diseases and contribution to overall plant health. Studies initiated so far indicate that in some cases, fungal rhizosphere community structure may be influenced by transgenic plants, but the changes are often minor, and the functional consequences are unclear (Donegan *et al.*, 1995; Blackwood and Buyer, 2004; O'Callaghan *et al.*, 2004).

1.4.1 Conclusions drawn from research to date

A recent comprehensive review has been made that summarizes 25 peer-reviewed studies of the effects of various transgenic plants on soil systems, including the examples described above (Cartwright and Lilley, 2004) (http::://www.defra.gov.uk /environment/gm/research/epg-1-5-214.html). Effects on the soil community or soil system were noted in 16 of the cases, although the induced changes were generally transient. The reviewers concluded that while genetically modified crop plants may cause detectable changes in associated soil microbial communities, these changes are often minor relative to natural fluctuations in community composition. In many cases, abiotic and biotic factors, such as season, weather, plant genotype, and plant developmental stages were implicated as far more important drivers of microbial community structure in the rhizosphere than possible changes induced by genetically modified plants. As expected, baseline data of natural variation was identified as being

integral to assessing the significance of potential changes induced by genetically modified plants (Cartwright and Lilley, 2004). The wide range of results from independent assessments indicated that detection of changes in community composition depended greatly on the methods of detection used and which portions of the community were monitored, and this highlights the need for the continued development of these methods to improve current risk assessment strategies. To address this, Kowalchuck *et al.*, (2003) proposed that impacts of genetically modified plants be investigated on a case-by-case basis, within an investigative framework that monitors both potentially vulnerable targets as well as general community parameters. Although recent research has improved the understanding of potential effects induced by transgenic plants on soil microbial communities, this knowledge is far from complete. Assessment of the impact of transgenic plants on soil microbial communities will benefit from the continued development of techniques to study complex microbial communities, and from a more complete understanding of the dynamic nature of plant-microbe systems.

Long-term impact assessment strategies may be particularly important for transgenic forest tree species. It is possible that the effect of genetic modification would be more pronounced in communities associated with trees compared to those associated with crop plants. Agricultural practices, such as crop rotation and soil tillage disrupt established root-associated microbial communities, and in some cases, changes in microbial community structure associated with transgenic plants do not persist in to the following field season (Dunfield and Germida, 2003). In contrast, trees grow for time periods of several years to several decades prior to harvest, and novel transgenic products produced by genetically modified trees could potentially accumulate and exert effects on associated communities for extended periods of time. Many factors may influence the accumulation of transgenic proteins in the soil, including the amount that is present in plant tissues, the resistance of the protein to degradation, and the chemical characteristics of the soil (Liu *et al.*, 2005). During decomposition of plant litter, most transgenic proteins appear to be degraded, but some proteins can bind to mineral or organo-mineral particles in soils, thus reducing their susceptibility to biodegradation. For example, in soils characterized by low pH and high clay content, the presence of Cry1Ab Bt-toxin was detected 7 months after transgenic maize was harvested (Baumgarte and Tebbe,

2005). While there are numerous studies aimed at monitoring potential impacts of transgenic crop species on soil microbial communities (reviewed in Bruinsma *et al.*, 2003; Dunfield and Germida, 2004; Wolfenbarger and Pfifer, 2000), the present study is one of the few that seeks to monitor changes in rhizosphere communities associated with transgenic trees.

1.5 Objectives

The aim of this study was to characterize and compare the rhizosphere communities associated with transgenic hybrid aspen (*Populus tremula* X *P. alba* clone INRA 717I-B4) over-expressing a hybrid poplar polyphenol oxidase gene (PPO-1) and non-transgenic hybrid aspen grown in the greenhouse, using a culture independent approach.

Polyphenol oxidases (PPO, EC 1.10.3.1) are common enzymes in higher plants and fungi that catalyze the oxidation of o-diphenols to o-quinones using molecular oxygen (Mayer and Harel, 1979). The o-quinones are highly reactive molecules, which rapidly modify and crosslink proteins and cellular constituents. Enhanced resistance to herbivory by forest tent caterpillars was demonstrated in the transgenic trees used in the current study, providing evidence for a role of PPO-1 in plant defence against folivorous insects (Wang and Constabel, 2003). Since this PPO has a broad range of acceptable substrates (Wang and Constabel, 2003), it is possible that it is also involved in defence against fungal and microbial pathogens, a general role for PPOs that is supported by numerous examples in the literature. Li and Steffens (2002) demonstrated that tomato plants over-expressing a potato PPO gene had increased resistance to foliar infection by the bacterial pathogen *Pseudomonas syringae* pv tomato. Furthermore, comparison of PPO activity in resistant and susceptible pearl millet (*Penningsetum glaucum* (L) R.Br) cultivars demonstrated a role for PPO in induced resistance against the fungal pathogen, *Fusarium graminareum,* which causes head blight of wheat (Mohammadi and Kazemi, 2002). More recently, elevated levels of PPO activity were measured in pearl millet cultivars resistant to *Sclerospora graminicola*, the causative agent of downy mildew, and analysis of the induction and accumulation of PPO activity in this system demonstrated

that PPO is actively involved in plant defence (Raj *et al.*, 2006). Despite these correlations, the mechanisms linking PPO to disease resistance are unclear, making it difficult to predict potential non-target effects of PPO over-expression on the rhizosphere community. However, most biological effects of PPOs appear to be due to the high reactivity of the o-quinone products (Peter, 1989) suggesting that a broad range of microorganisms may be impacted. The persistence of polyphenol oxidases in the soil is unknown, although these enzymes are known to be highly stable. Poplar PPOs for example, are resistant to proteolytic enzymes in the caterpillar gut (Wang and Constabel, 2004), and latex PPOs from *Hevea brasiliensis* can withstand heat treatment (Wititsuwannakul *et al.*, 2002).

Since polyphenol oxidase over-expressing transgenic plants have previously been shown to have increased resistance to plant pathogens, and rhizosphere organisms are primary consumers of plant produced carbon and nutrients, it may be hypothesized that release of transgenic products from sloughed or damaged root cells, or an otherwise altered root exudation may influence the structure of bacterial and fungal communities in the vicinity of the roots. Members of the *Populus* genus are logical targets for impact assessment strategies because they are commonly used as model systems in the development of economically important transgenic conifer species (Bradshaw *et al.*, 2000), and baseline data regarding community diversity and methodology developed may be applied to future studies.

Through cloning and sequencing of specific gene loci, catalogues of sequence diversity associated with each tree type were generated, and community diversity was compared, as affected by the presence or absence of the introduced PPO gene. The main strengths of a sequencing approach were that the data were additive, retrievable and provided high levels of resolution. Since current understanding of the structure and function of rhizosphere communities is limited, it is challenging to evaluate changes in diversity (Bruinsma *et al.*, 2003). This lack of knowledge, as well as limitations associated with interpreting data from studies of microbial communities highlights the value of sequence data. Retrievable sequence datasets contribute to baseline knowledge regarding initial community composition and natural fluctuations, and can also be used to develop comparative methods that may assist in detecting changes in community

composition. For example, the 16S rRNA gene libraries of McCaig *et al.* (1999), derived from grassland soils have been extensively studied by microbiologists developing methodology to study microbial ecology (Hughes *et al.*, 2001; Martin, 2002; Schloss *et al.*, 2004; Schloss and Handelsman, 2005; Schloss and Handelsman, 2006).

In the present study, the activity of PPO in the root tissue of each tree genotype was compared, and a variety of approaches were used to profile and compare rhizosphere communities. For the bacterial component of the rhizosphere community, profiles were generated based on partial bacterial 16S rRNA gene fragments amplified from the rhizospheres of transgenic and non-transgenic hybrid aspen. To reduce biases associated with using a single genetic target to describe microbial community composition, to increase confidence in conclusions drawn from the 16S libraries, and to evaluate the CPN-60 gene sequence as an alternative phylogenetic anchor for use in impact assessment studies, community profiles were also compared based on the amplification of the CPN-60 "universal target". To screen for potential changes in the diversity of the fungal community, profiles were compared based on the amplification of a portion of the 18S rRNA coding region. For each primer set, the phylogenetic distribution of clones in each library was compared, and diversity was compared based on differences in species richness and levels of evenness within each library. Species richness refers to the number of species in a community, and evenness describes the proportional abundance of each species (Krebs, 1999). To screen for rudimentary differences in bacterial and fungal cell densities harboured by transgenic and unmodified trees, the portion of the community amenable to *in vitro* culture was also enumerated and compared.

With an overall goal of screening for general effects of greenhouse-grown PPO over-expressing hybrid aspen on the rhizosphere community, the specific objectives of this study were to:

1) test for rudimentary differences in the enumeration of the viable, culturable portion of the bacterial and fungal populations associated with each tree genotype;

2) characterize and compare the broad phylogenetic composition of bacterial taxa inhabiting the rhizospheres of greenhouse-grown PPO over-expressing hybrid aspen and parental genotypes based on 16S and CPN-60 gene libraries;

3) compare community structure and the relative diversity of bacteria recovered from rhizospheres of each tree genotype;

4) characterize and compare the taxa present in the fungal component of the rhizosphere communities of each tree genotype, based on 18S gene libraries;

5) compare community structure and the relative diversity of the fungal component of the rhizospheres of each tree genotype;

6) assess the significance of any differences detected, relative to natural variation in the community composition in the soils used in the study;

7) evaluate the utility of the CPN-60 gene target in assessing the impact of transgenic plants on soil microbial communities.

Materials and Methods

2.1 Plant material and soil sampling

Production of transgenic PPO over-expressing hybrid aspen:

Transgenic hybrid aspen (*Populus tremula* X *P. alba* clone INRA 717I-B4) over-expressing a hybrid poplar (*Populus trichocarpa* X *P. deltoides*) *PtdPPO1* gene were constructed, as described in Wang and Constabel (2004), and were graciously provided for this study by Dr. Peter Constabel, at the University of Victoria. Unmodified control plants having the parental genotype were also propagated. After in-vitro grown plantlets had rooted and had reached a height of 5-10 cm, they were planted in 4" pots containing Sunshine Mix #4 (Sungro, Seba Beach, AB, Canada) that had been thoroughly homogenized, and were transferred to a mist chamber for acclimation in the greenhouse at the Centre for Forest Biology at the University of Victoria. Following acclimation, plantlets were transplanted in to 11 litre pots.

Potting media used:

The potting media in the larger pots contained 2 parts premium sterilized potting soil (Professional Gardener Series, Island's Finest, Cinnibar Valley Farms Ltd., Nanaimo, B.C.), 1 part peat moss (Sunshine, Sun Gro Horticulture, Canada Ltd.), 0.5 part premium grade vermiculite (Ultra Tech, Richmond, CA., USA), and 0.5 part perlite (Dutch Treat Products, Surrey B.C.) A viable plate count on nutrient broth agar and malt extract agar confirmed that the potting soil harboured a diverse array of culturable bacteria and fungi. This potting media was chosen because it contained soil as opposed to completely synthetic media commonly used in nurseries.

Homogenization of potting media:

It was important that the potting media be homogenized thoroughly prior to distribution among pots. To achieve this, one batch of Sunshine Mix #4 was thoroughly mixed and stored for the initial transplantation of plantlets. In order to ensure media homogeneity among the 11 litre pots, sufficient quantities of each component were purchased for the entire experiment. Each component was thoroughly homogenized, and then a cement mixer was used to mix individual components together. The potting media was distributed immediately among pots, and these were stored on a greenhouse bench until trees were planted. All plants were subsequently maintained at approximately 18°C in the greenhouse at the Centre for Forest Biology at the University of Victoria under ambient light conditions. Plants were watered uniformly each day, and old leaves were pruned so that they did not contact the soil. No fertilizer was added during the experiment.

Sampling of plant material and soil

Sampling design:

After 5 months growth, 4 trees were harvested for the experiments detailed herein. Root tissue and rhizosphere soil was collected from two individual plants representing one transgenic line (previously designated as line 19 in Wang and Constabel, 2004) and two individual control plants (line 717). Each tree was also given an identifying number, and was defined as a replicate. Thus the two transgenic trees were identified as 19-4 (Rep A) and 19-9 (Rep B). The two control trees were named 717-10 (Rep A) and 717-11 (Rep B). Soil collected from each tree was used to create a replicate gene library as described below. All tissue and soil samples were stored at -80°C.

Collection of root tissue to assay for PPO activity:

After removing the adhering bulk soil, 3 entire roots, including attached lateral roots, were collected from each tree. The root tissue was then soaked in sterile water (approximately 200 mL) to remove any remaining soil, and was blotted dry. A sterile razor blade was used to slice the roots into 1 inch pieces, which were randomly distributed into sterile microfuge tubes, each tube receiving 0.25-0.30 grams of tissue. Dry ice was used to cool the samples prior to transfer to -80°C storage, which were subsequently assayed for PPO activity.

Definition of rhizosphere soil:

The rhizosphere soil was defined as that which was collected by soaking roots after the bulk soil had been rigorously shaken from it. For each tree, 50 grams of full length roots with adhering soil were soaked in 0.1% sodium pyrophosphate dilutent (200 mL) for 20 minutes, with occasional agitation. The soil was then separated from the dilutent by centrifugation (2900 rpm, 10 min).

A portion of the rhizosphere soil was used within 24 hrs of collection, to enumerate viable culturable fungi and bacteria, using the dilution plate count method. The remainder was lyophilized, prior to –80°C storage and subsequent DNA extraction for molecular analysis. Lyophilization had the added benefit of increasing the homogeneity of the sample. Once freeze dried, the soil was easily ground to a fine powder which could then be mixed by shaking.

2.2 Enumeration of viable culturable bacteria and fungi

For each tree, eight to ten grams of soil was divided evenly among three dilution bottles containing 0.1% sodium pyrophosphate dilutent (90 mL) and forty 3 mm glass beads (Fisher Scientific, Ottawa, Canada). Each dilution bottle was defined as a repetition. Dilution bottles were shaken vigorously by hand for 30-60 seconds, and then were shaken horizontally, at 190 rpm, for 30 minutes. The 10^{-1} dilution was removed

from the shaker, was allowed to stand for 30 seconds, and 1 mL was removed from the middle region of the suspension and was added to 9 mL sodium pyrophosphate (0.1%) to achieve a 10^{-2} dilution. Samples (100 uL) from 10^{-4} and 10^{-5} dilutions were plated on 1/4 strength Nutrient Agar (Sigma), to enumerate viable culturable bacteria. To enumerate viable culturable fungi, 100 uL samples from 10^{-2} and 10^{-3} dilutions were plated on 0.7% (w/v) malt extract agar (MEA) (DIFCO). This media was supplemented with 5 PPM methyl benzmidazole carbamate phosphate to limit the growth of fast growing fungal species, 100 PPM neomycin sulfate, and 500 PPM streptomycin sulfate to minimize bacterial growth (Johnson, 1995). Nutrient poor medium was used for dilution plating because this generally gives rise to larger numbers of bacteria and fungi than nutrient rich medium (Zuberer, 1994).

Plates were incubated at 20 °C, and bacterial colony forming units (CFU) on NA plates were counted after 24 hours, 48 hours, and 5 days. Fungal CFUs were enumerated on MEA plates after 48 hours and 5 days. Soil moisture content was calculated, and results were converted to cfu/dry gram soil.

Analysis of viable plate count data:

After 48 hours of growth at room temperature, bacterial colony counts on the NA plates from the 10^{-5} dilution ranged from 30-300, and these data were used to calculate and compare the number of colony forming units/dry gram of soil from each rhizosphere sample.

For fungi, the number of colony forming units/ dry gram of soil was calculated from the plates from the 10^{-2} dilution, after 48 hours of growth. There were generally fewer than 30 colony forming units on the plates, but these counts were used in the calculations because numerous plates from subsequent counts were overgrown.

For both fungi and bacteria, the reported means were calculated using the average colony counts of three replicate dilutions consisting of 3 plates each, and the standard error of the mean was reported. All statistical comparisons were performed on untransformed data because Levene's test for equality of variance indicated that the variances were homogenous, and the Kolmogorov-Smirnov test confirmed a normal

distribution of data. A one way analysis of variance (ANOVA) was performed followed by a Student-Newman-Keuls test for differences between the mean cfu/dry gram of soil for each rhizosphere sample (Sokal and Rolf, 1995). Data was considered to be statistically different if $p<0.05$.

2.3 Assays for PPO activity:

Frozen root tissue (0.3 grams) was ground with sand and polyvinylpolypyrrolidone (PVPP) in 300 µL extraction buffer (100 mM NaPO4, pH 7.0, 0.1% (v/v) Triton X-100), was clarified by centrifugation, and the supernatant was immediately assayed for PPO activity. The conversion of DL-DOPA (dihydroxyphenylalanine) to dopaquinone at 490 nm was monitored spectrophotometrically (Sherman *et al.*, 1991). The DOPA (25 mM) was dissolved in 100 mM citrate phosphate buffer (pH 5.0) with 0.15% SDS and 470 units/mL catalase, and the assay buffer was aerated for 5 minutes prior to the assay. Protein concentration was determined with the Bradford reagent using bovine serum albumim as a standard (Bradford, 1976). The reaction was inhibited by tropolone, confirming that DOPA oxidation was due to a PPO, rather than other oxidative enzymes. Assays were performed in duplicate, and PPO activity (units/mg protein) was calculated for each sample.

Statistical analysis of PPO data

For each tree, the value for PPO activity/mg protein reported was the mean enzyme activity of 5 root samples. Prior to statistical analysis, data were log transformed because Levene's test for equality of variance indicated that the variances were not homogenous, and the Kolmogorov-Smirnov test was used to confirm a normal distribution. A one way analysis of variance (ANOVA) was performed on log transformed data, followed by a Student-Newman-Keuls test for differences between the mean PPO activities for each tree (Sokal and Rolf, 1995). Differences between each tree

were considered to be statistically significant if $p \leq 0.05$. Activity values were then back-transformed and were reported as geometric means, with a 95% confidence interval.

2.4 Extraction of bacterial and fungal DNA from soil and PCR amplification

Total genomic DNA was extracted from 150 mg of a lyophilized soil sub-sample (1 DNA extraction/tree) using a bead-beating method. In 2 mL screw cap microcentrifuge tubes, soil was mixed with 0.75 mL 0.1 M sodium phosphate buffer pH 8 and 2g 0.1 mm zirconia/silica beads (Fisher Scientific). To aid in cell lysis, 0.37 mL of 100 mM NaCl /500 mM Tris-HCl pH 8/ 10% SDS was added, and the tubes were shaken with at high speed for 5 minutes on a mini-bead beater. After centrifuging for 3 minutes (11,200 rpm), the supernatant was chilled for 10 minutes on ice with 0.4 volume of 7.5 M ammonium acetate. The samples were then centrifuged for 3 minutes (11,200 rpm). To purify the crude DNA extract and reduce humic acid contamination, aliquots of the supernatant were filtered through spin columns packed with polyvinylpolypyrrolidone (PVPP) (Berthelet *et al.*, 1996). Concentration was determined spectophotometrically, and the purified DNA was stored in water at –20°C prior to PCR amplification.

PCR amplification of community DNA

Table 1. Primers used to generate gene libraries

Target for PCR amplification	Primer names	Primer sequence 5'-3'	Original reference
16S	968F	AACGCGAAGAACCTTAC	Nubel *et al.*, 1996
	1401R	CGGTGTGTACAAGACCC	
CPN-60	H279	GAIIIIGCIGGIGA(T/C)GGIACIACIAC	Goh *et al.*, 1996
	H280	(T/C)(T/G)I(T/C)(T/G)ITCICC(AG)AAICCIGGIGC (T/C)TT	
Fungal 18S	EF4	GGAAGGG(G/A)TGTATTTATTAG	Smit *et al.*, 1999
	Fung5	GTAAAAGTCCTGGTTCCCC	

Bacterial 16S rRNA gene amplification:

Amplification of a 433 bp portion of the bacterial 16S gene using PCR was achieved using primers 968f and 1401r (Table 1). These primers encompass the variable regions V6-V8 of eubacterial 16S rDNA corresponding to positions 968 to 1401 in *Escherichia coli* (Nubel *et al.*, 1996). This fragment of the 16S gene was targeted since its length could be captured by sequencing in one direction, thus minimizing costs. Five microlitres of 10 X PCR buffer (Invitrogen, Burlington, Ontario, Canada), 2.5 μL of 50 mM $MgCl_2$, 50 pmol of each primer, 1 μL of a 10 mM concentration of each deoxyribonucleoside triphosphate, 1 unit of Taq polymerase (Invitrogen), 36 μL of 0.1 micron filtered Gibco ultra-pure water (Burlington, Ontario, Canada) and 1 ng template DNA were combined in a final volume of 50 μL, and were overlayed with filter sterilized mineral oil. Reactions included a negative control containing no template DNA to ensure that no contaminating template was present in the reactions. Amplification was carried out using a Perkin Elmer thermal cycler, according to the following parameters: (i) an initial denaturation step (3 min 94°C) (ii) 28 cycles, with each cycle consisting of: denaturation (1 min 94°C), annealing (1 min 61°C), extension (1 min 72°C), followed by a (iii) final extension step (4 minutes, 72°C). Twenty-eight cycles were used for amplification, since this resulted in a strong signal from rhizosphere community DNA, but eliminated minor amplification from the negative control sample, which was observed when more PCR cycles were used.

CPN-60 gene amplification:

For amplification of the 'universal target' within the chaperonin-60 gene, primers H279 and H280 (Table 1) were used (modified from Goh *et al.*, 1996). Inosine (I) was used to reduce the degeneracy of the sequences (Ohtsuka *et al.*, 1995). These primers were designed to amplify the region between codons 92 and 277 based on the *Escherichia coli* CPN-60 sequence (Accession number X07850). Reaction components were as described for 16S amplification and also contained 1 ng of template DNA. The PCR cycle consisted of an initial denaturation step of 2 minutes at 72°C, followed by 35

cycles (94°C 30seconds, 42°C 30 seconds, 72°C 30 seconds), and a final 2 minute extension at 72°C.

Fungal 18S rRNA gene amplification:

Primers EF4 and Fung5 (Table 1) were used to amplify a 550 bp fragment within the fungal 18S coding region (Smit *et al.*, 1999). Reaction components were as described for 16S amplification, except that 10 ng template DNA was used. The PCR cycle consisted of an initial denaturation step of 2 min 94°C, followed by 35 cycles (1 min 94°C, 1 min 56°C, 2 min 72°C) and a final 10 min extension at 72°C.

Following amplification, PCR products were separated using 1.5% gel electrophoresis in 1X Tris-Acetate-EDTA (TAE) buffer, and were visualized by staining with ethidium bromide and UV transillumination. Bands were excised using a sterile scalpel, and DNA was purified from gel slices by using the QIAquick gel extraction kit (Qiagen, Mississauga, Ontario, Canada).

2.5 Construction of gene libraries and sequencing

For each tree (2 individual transgenic plants and 2 individual control plants), one soil sample was used to create a gene library. The resulting libraries were named according to the line, plant identifying number, and primers used for PCR amplification.

Purified PCR products were ligated into pGem-T cloning vectors (Promega, Napeen, Ontario, Canada), following the manufacturer's protocol, and were transformed into electrocompetent DH5-α *Escherichia coli* (Invitrogen). One hundred ninety-two positive clones per sample were picked and were used to inoculate 96-well plates with 200 µL 2YT with ampicillin (50µg/mL) and 15% glycerol in each well. Culture plates were sealed and incubated overnight at 37°C prior to –80°C storage. Prior to sequencing, the presence of inserts was confirmed in a small number of clones (5-10 per 96 well plate). To achieve this, selected clones were used to inoculate 5 mL LB/ampicillin (50 µg/mL), which was grown overnight, shaking (250 rpm) at 37°C. Plasmid DNA

extraction was performed using the QIAprep Spin Miniprep kit (Quiagen), and insert presence and size was determined by SacI/SacII restriction digests of plasmid DNA, followed by visualization with similar gel electrophoresis methods.

Clone storage

Clones containing 18S, 16S, and CPN-60 inserts originally amplified from the rhizospheres of transgenic and control plants were stored at -80°C. For each tree, 192 clones for each of the three primer sets (16S, 18S, and CPN-60) were sequenced (2304 total).

Sequencing

All preparation of plasmid DNA and sequencing was carried out by the Centre for Biomedical Research, at the University of Victoria. Plasmid DNA for sequencing was isolated from cultures that had been grown overnight in 96 channel microtitre plates (VWR, Nepean, Ontario, Canada) containing 1 mL LB/well supplemented with 100 μg/mL ampicillin. Cells were pelleted with centrifugation (1800 g, 10 minutes), the supernatant was discarded, and pellets were resuspended (100 μL of 0.05 M Tris/HCl, 0.01 M EDTA, pH 7.5 and 50 μg/mL RNase A). Cells were lysed by adding 100 uL (0.2 N NaOH, 1% SDS) to each well and were vortexed for 2-5 minutes. Lysates were vortexed again for 2 minutes with potassium acetate (100 μL/well 3M KOAc pH 5.5). Lysates were then clarified by filtering through clarification plates (VWR) (2000 rpm, 3 min), into DNA Binding/Recovery plates (VWR) containing 225 μL isopropanol. After mixing, the DNA was pelleted by centrifugation (2254 g, 30 min). Once pellets had dried, they were washed with 80% ethanol, were dried and were resuspended in 50 μL elution buffer.

Sequencing reactions were conducted in 96-well microtiter plate format (VWR) using 200-300 ng of template, in a 5 μL Big Dye Terminator Sequencing reaction (ABI). Reactions were assembled and thermocycled according to the manufacturer's recommended protocol. Sequence extension reaction products were purifed using

ethanol precipitation, and pellets were re-suspended in 10 μL nuclease-free water. Completed reactions were sequenced in a single direction, on an ABI 3730 DNA analyzer. Raw sequence data were processed using Phred software (Ewing *et al.*, 1998), which assigns quality values to the bases and trims poor quality regions. Raw sequences were deposited in the password-protected Koop lab sequence data repository (http://woodstock.ceh.uvic.ca).

2.6 Sequence analysis and assignment of clones to phylogenetic groups

16S Libraries:

A total of 768 clones containing 16S inserts were submitted for sequencing. Of this total, 626 of the resulting sequences were high quality and encompassed the 400 bp region of interest. Using Orientation Checker, a program for checking and altering the orientation of 16S rRNA sequences (http://www.cf.ac.uk/biosi/research/biosoft/Squirrel/index.html), sequences were oriented in the same direction, were aligned using Mega 3.1 software (http://megasoftware.net/) (Kumar *et al.*, 2004) and primer regions were trimmed. The sequences were then screened for chimeric properties using Mallard software (http://www.cf.ac.uk/biosi/research/biosoft/Mallard/index.html) (Ashelford *et al.*, 2006). One sequence was identified as a potential chimera, and was excluded from subsequent analysis. Chimeric sequences are PCR artefacts that potentially occur when multiple DNA templates are present in the PCR mixture, as is the case when DNA is extracted from environmental samples. Chimeric sequences are thought to form during PCR amplification when sequence synthesis starts from one template and is interrupted, then continues from another template sharing a degree of localized homology with the original (Ashelford *et al.*, 2005). Each replicate library derived from a single tree was reduced to 130 randomly selected sequences so that each contained the same number (Table 2), and clone sequences were grouped into "operational taxonomic units" (OTUs) sharing 97% sequence similarity using the furthest neighbour clustering algorithm in the computer program, DOTUR (Schloss and Handelsman, 2005). Input files for DOTUR were constructed by aligning sequences in ClustalW using default parameters, and

converting this to a distance matrix using a Jukes-Cantor Model in the DNADist program of the Phylip analysis package (Felsenstein, 1989). The 97% similarity OTU definition is most commonly used to define a bacterial species. This value corresponds well with the 70% genomic DNA hybridization threshold, classically used to define bacterial species (Roselló-Mora and Amann, 2001).

Table 2. Bacterial 16S rRNA gene library construction and grouping of data

Tree type	Replicate tree	Number of clones sequenced	Number of high quality sequences obtained	Number of sequences analyzed/ replicate library	Number of sequences analyzed/ Pooled dataset
Transgenic	19-4	192	163	130	260
	19-9	192	175	130	
Control	717-10	192	158	130	260
	717-11	192	130	130	

CPN-60 Libraries:

A total of 768 clones containing CPN-60 inserts were submitted for sequencing. Four hundred seventy one of the resulting sequences were of high quality and encompassed the 550 bp region of interest. These sequences were aligned using Mega 3.1 software, and primer regions were trimmed. Transgenic and control libraries were compared only after pooling datasets from the two replicate trees, for a total of 215 sequences per library, since the replicate libraries contained significantly different numbers of high quality sequences (Table 3).

Table 3. Chaperonin-60 gene library construction and grouping of data

Tree type	Replicate tree	No. clones sequenced	Number of high quality sequences obtained	Number of sequences analyzed/ replicate library	Number of sequences analyzed/ pooled dataset
Transgenic	19-4	192	58	N/A	215
	19-9	192	157	N/A	
Control	717-10	192	139	N/A	215
	717-11	192	117	N/A	

18S Libraries:

A total of 768 clones containing 18S inserts were submitted for sequencing. High quality sequence data that encompassed the 500 bp region of interest was attained for 601 of the clones. These sequences were aligned using Mega 3.1 software, and primer regions were trimmed. Sequences were screened for chimeric properties, using the Ribosomal Database Chimera Check program, and those suspected of being chimeric were excluded from further analysis. Each replicate library derived from a single tree was reduced to 140 sequences, so that each contained the same number (Table 4). Clone sequences were grouped into OTUs sharing 99% sequence similarity using the furthest neighbour clustering algorithm in DOTUR (Schloss, 2005) because of the high level of conservation in the fungal 18S rRNA coding region. The same OTU definition was adopted by Andersonn *et al.* (2003). Input files for DOTUR were constructed as described above.

Table 4. Fungal 18S rRNA gene library construction and grouping of data

Tree type	Replicate tree	Total number clones sequenced	Number of high quality sequences obtained	Number of sequences analyzed/ replicate library	Number of sequences analyzed/ Pooled dataset
Transgenic	19-4	192	158	140	280
	19-9	192	140	140	
Control	717-10	192	151	140	280
	717-11	192	152	140	

Assignment of bacterial and fungal clones to phylogenetic groups:

Bacterial (16S) clones:

Once sequences were grouped into OTUs sharing 97% sequence similarity, the Library Compare Tool on the Ribosomal Database Project (RDP) website (Cole *et al.*, 2005) (http://rdp.cme.msu.edu/) was used to assign library sequences to bacterial taxa, and to compare taxonomic composition of replicate libraries. Replicate libraries from the same tree genotype were then pooled, and the comparison was repeated. The Library Compare tool used the RDP naïve Bayesian classifier to rapidly classify library sequences into the bacterial taxonomy. The classifier was trained on known type strain rRNA sequences as well as a small number of other sequences representing regions of bacterial diversity with few cultured organisms. Each library sequence was assigned to a set of hierarchical taxa (as proposed in Bergey's Manual of Systematic Bacteriology, 2004), along with a bootstrap confidence estimate for each rank assignment.

CPN-60 clones:

Unique CPN-60 nucleotide and peptide sequences were compared to approximately 8000 sequences in the Chaperonin-60 database using FASTA (Pearson and Lipman, 1988) and Blast-P search algorithms (Altschul *et al.*, 1990) and were assigned to major phylogenetic groups based on the most similar reference sequences (Hill *et al.*, 2004). To further illustrate the phylogenetic relationships between library nucleotide sequences and their nearest neighbours in the database, clones were grouped

into OTUs sharing 80% sequence similarity, and were aligned based on amino acid sequence using a PAM matrix. This OTU definition was chosen as bacterial strains within the same species have previously been shown to share this level of similarity (Brousseau *et al.*, 2001). A neighbour-joining tree was then constructed using a Jukes Cantor model of nucleotide substitution (Mega 3.1 software), and genetic distances were calculated using the first two nucleotide positions of each codon. The bootstrap consensus tree was rooted with *Aquifex aeolicus*, and where possible, clusters of similar sequences within the tree were named according to the identity of their nearest database neighbours.

Fungal (18S) clones:

Once fungal clones were grouped into OTUs, a Blast-n search (Altschul, *et al.*, 1990) was carried out for one representative sequence from each group. The most homologous sequences found in the Genbank database were used to create a multiple-sequence alignment. A neighbour-joining analysis was performed, using the Jukes-Cantor model of nucleotide substitution, and the phylogenetic tree was edited with 1,000 bootstrap replicates using Mega 3.1 software.

Statistical analysis:

Taxonomic composition of libraries was compared by performing a χ^2 test of homogeneity on major phylogenetic groups. Comparisons between control and transgenic libraries were made using pooled data sets. Variability between replicate libraries derived from individual trees was also assessed for 16S and 18S libraries. For χ^2 analysis, 2 x 2 contingency tables were constructed for the phylogenetic group being compared using frequency data, and Yates' correction for continuity was applied. In cases where the frequency of recovery was too low to allow meaningful χ^2 analysis (if any observed or expected frequency in the contingency table was <5), Fisher's exact test (2 sided) was used to test for statistical significance (Sokal and Rolf, 1995). P-values $\leq$ 0.05 were considered statistically significant, and were reported. Calculations for χ^2 and Fisher's exact test were performed online using S.I.S.A. (Simple Interactive Statistical Analysis) at http://home.clara.net/sisa/.

2.7 Estimation of species richness and evenness

Once bacterial (16S and CPN-60) and fungal (18S) sequences were assigned to OTUs, the frequency data was used to calculate a number of diversity indices to describe and compare the richness, defined as the number of OTUs present, and level of evenness (the relative abundance of OTUs) within each pooled dataset. The indices and 95% confidence intervals were calculated using DOTUR, which performed a random sampling without replacement procedure. Sample calculations are provided in the manual on the DOTUR website. The chance of drawing a representative sequence from each OTU was the number of times the OTU was observed divided by the total number of sequences in the library (Schloss and Handelsman, 2005). The equations used to calculate each diversity index and richness estimator are detailed in Appendix 2.

The richness indicators used were:

(i) the total number of OTUs recovered from each library;

(ii) the shannon diversity index (H'), a general index that considers both species richness and evenness (Magurran, 1988);

(iii) the Chao richness estimator, a non-parametric estimator derived from mark-recapture methods that estimates the true number of species (or OTUs) in an assemblage based on the number of rare OTUs in a sample. It takes into account the number of singletons (OTUs captured once) and doubletons (OTUs captured twice) in a sample (Chao, 1984);

(iv) the abundance based coverage estimator (ACE), which estimates species richness by incorporating data from all species with fewer than 10 individuals, rather than just singletons and doubletons (Chao and Lee, 1992).

The indices used to describe the degree of evenness or dominance were:

(i) the shannon evenness index, which is the ratio of the shannon diversity index to the maximum possible value that could theoretically be obtained with the observed number of OTUs (Magurran, 1988). Index values range from 0 to 1, where a value of 1 indicates that all OTUs are equally abundant;

(ii) the simpson's index of diversity, which gives the probability that two clones chosen at random will be from the same OTU and is strongly influenced by dominant species (Magurran, 1988). Index values range from 1 to 0, where a value of 0 indicates that all OTUs are equally abundant;

(iii) the berger-parker index, which is the relative abundance of the most common OTU (Magurran, 1988). Index values range from 0 to 1, where a value of 1 indicates that all OTUs are equally abundant.

2.8 Statistical comparison of gene libraries based on coverage

In 2001, a computer program (Libshuff) was implemented to determine the significance of differences in composition between clonal libraries of environmental rRNA gene sequences (Singleton *et al.*, 2001). The program has since been applied to a wide range of studies aimed at comparing gene libraries (e.g. Wang *et al.*, 2005; Stach *et al.*, 2003; Filion *et al.*, 2004), and a recently released version, ∫-Libshuff, detects differences with greater sensitivity (Schloss, 2004). To determine whether the gene libraries were drawn from the same population, the integral form of the Cramer-von Mises test statistic was calculated and compared using a Monte Carlo test procedure, using ∫-Libshuff (Schloss, 2004). For multiple comparisons of libraries, the Bonferroni correction was used to calculate the critical p-value. The 2 libraries being compared were considered to be significantly different for each other if the lower of the two p-values was

below or equal to the critical p-value. Input files were constructed by aligning the sequences in the libraries being compared (ClustalW, using default parameters), and by converting this to a distance matrix as described above.

Since the protein coding CPN-60 sequences were much more divergent than the rRNA sequences, two different methods of alignment were used to construct input files for ∫-Libshuff analysis, and similar results were obtained for each. For the first alignment, nucleotide sequences were aligned as described above. For the second alignment, sequences were aligned in Mega 3.1 (using a PAM matrix) using translated amino acid sequences, which was less divergent than the corresponding nucleotide sequences. Using Mega 3.1 software, it was possible to toggle back and forth between amino acid and nucleotide sequences. A distance matrix was then calculated using the nucleotide data and a Jukes-Cantor substitution model. Transgenic and control libraries were compared after pooling datasets, for a total of 215 sequences/library. Comparisons were also made between replicate libraries, but each contained a different number of sequences, as described above.

2.9 Similarity coefficients and collector's curves

For the rRNA libraries, similarity coefficients, which reflected the proportions of shared OTUs, were calculated by perfoming pairwise comparisons of replicate libraries (McCaig *et al.*, 1999). Similarity coefficients were not calculated for CPN-60 replicate libraries since each contained a significantly different number of sequences. Finally, collector's curves (the number of OTUs detected plotted versus the number of clones analyzed) were constructed to compare sequence diversity between pooled transgenic and control libraries.

Results

3.1 Polyphenol oxidase activity in root tissue

To characterize the transgenic plants, and to quantify levels of PPO activity in the root tissue of individual plants, spectrophotometric assays were performed to measure the conversion of dihydroxyphenylalanine (DOPA) to dopaquinone by polyphenol oxidase (PPO).

Elevated levels of PPO activity were observed in the root tissue of transgenic PPO over-expressing trees, relative to control trees (Figure 1). Although activity levels were highly variable between root samples from the transgenic plants, the average PPO activity in each of the PPO over-expressing transgenics was at least 10 fold greater than that measured in control trees. Analysis of variance suggested that there were significantly different levels of PPO activity in all the individual trees tested.

Rhizosphere soil adhered to the entire root length was used to profile the associated communities; therefore, PPO activity levels were also measured in entire root lengths. It is possible that this sampling method was partially responsible for the high variation observed in PPO activity in the transgenic roots, since different portions of the root may differentially express the transgene.

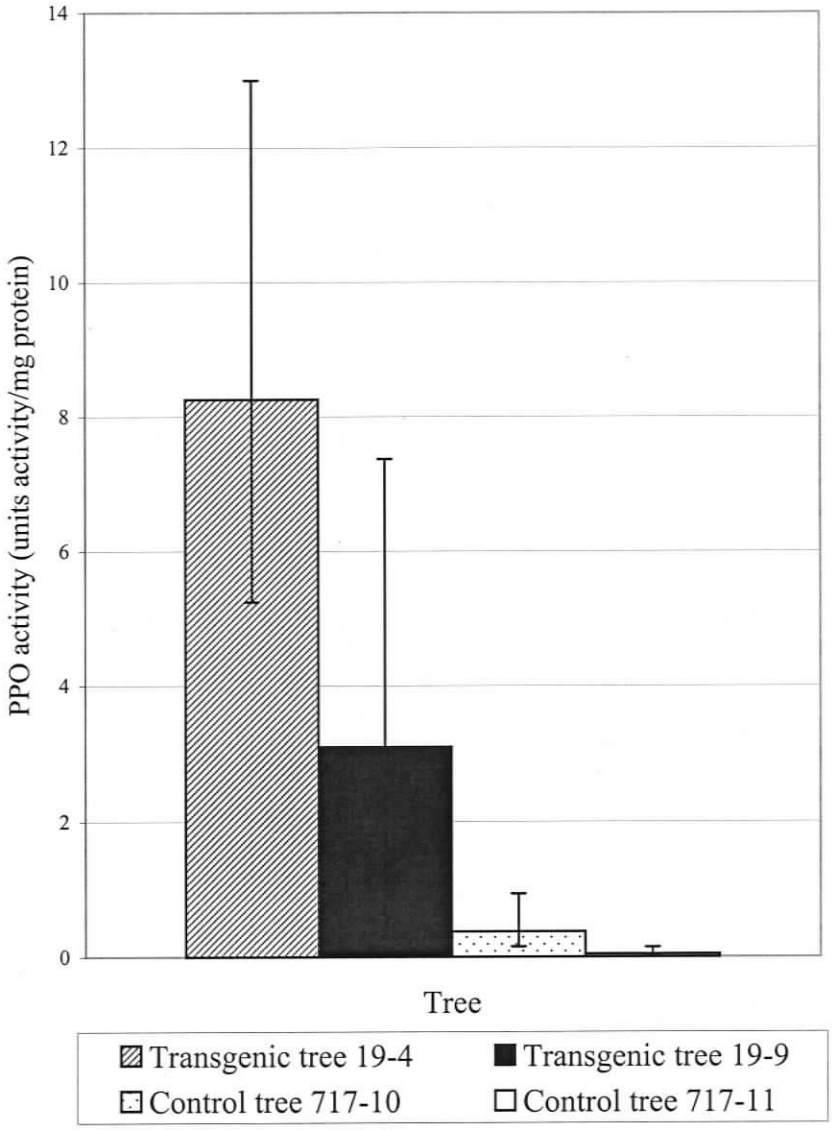


Figure 1. PPO activity in transgenic PPO over-expressing and non-transgenic hybrid aspen root tissue. Geometric means are reported with 95% confidence intervals, and are based on 5 root samples/tree. ANOVA analysis on log transformed data followed by Student-Newman-Keuls test for pairwise comparisons indicated that PPO activity in each tree was significantly different ($P<0.05$)

3.2 Enumeration of viable bacteria amenable to *in vitro* culture

Enumeration of the viable portion of the rhizosphere communities amenable to *in vitro* culture confirmed the presence of high numbers of bacteria associated with each tree type (Table 5). Plate counts ranged from 1.05×10^7 to 5.37×10^7 cfu/dry gram soil and there was variation in bacterial densities harboured by individual trees. Significantly lower plate counts from the transgenic trees compared to control samples suggested rudimentary differences between the two communities, but since plate counts for the two control replicates were also dissimilar, the observed difference is probably not a result of the transgenic nature of the plants, and likely reflected the underlying heterogeneity of the population and the biases inherent to a small sample size.

Table 5. Enumeration of viable culturable bacteria associated with the rhizospheres of PPO over-expressing transgenic and control hybrid aspen and results of ANOVA ($F=175$, $p<0.001$), followed by a Student-Newman-Keuls test for differences between the mean bacterial densities from each rhizosphere sample

Tree	Mean CFU/dry gram soil (± SE)	Significantly different from
Transgenic Tree 19-9 (a)	1.09×10^7 ($\pm 3.70 \times 10^5$)	c,d
Transgenic Tree 19-4 (b)	1.05×10^7 ($\pm 1.07 \times 10^6$)	c,d
Control tree 717-10 (c)	5.37×10^7 ($\pm 4.50 \times 10^6$)	a,b,d
Control tree 717-11 (d)	1.74×10^7 ($\pm 1.29 \times 10^6$)	a,b,c

Colony forming units on ¼ strength NA were counted after 48 hrs growth at 20°C. Differences in mean CFU/dry gram soil were considered significant if $p<0.05$. Means ± SE are the averages of 3 replicate 10^{-5} dilutions consisting of 3 plates each. Each series of replicate plates was generated from one soil sample derived from each tree.

3.3 Comparison of bacterial communities based on 16S gene libraries

3.3.1 Taxonomic composition of bacterial 16S rRNA gene libraries

Bacterial 16S rRNA gene sequences (260 from transgenic plant rhizospheres, and 260 from control plant rhizospheres) were grouped into operational taxonomic units (OTUs) sharing 97% sequence similarity, and were assigned to phylogenetic groups, using the Ribosomal Database Naïve Bayesian Classifier. The assignment of OTUs to hierarchical taxa and the bootstrap confidence for each rank assignment are summarized in Table A1, which is in Appendix 1. Classification of sequences using this method produced results similar to a Blast-n search of the genbank database (Altschul *et al.*, 1990), but was more efficient and practical. Pairwise identities obtained from Blast-n searches ranged from 91-100%. A large proportion (67%) of the library sequences were most similar to uncultured bacteria of unknown identity in the genbank database that were derived from environmental studies.

Overall, sequences most similar to members of the Proteobacteria were most abundant, comprising 72% of both the transgenic and control rhizosphere libraries, after datasets had been pooled (Table 6, Figure 2). This proportion was also reflected by sequences in the replicate library. The most abundant group that was not assigned to the Proteobacteria was the Actinobacteria, which comprised 7% of the control library and 9% of the transgenic library (Table 6).

Within the Proteobacteria, the most abundant group represented was the Alphaproteobacteria subdivision (comprising 36% of the transgenic and 37% of the control libraries after the replicate datasets had been pooled), followed by the Betaproteobacteria (comprising 16% of the transgenic and 15% of the control libraries), and the Gammaproteobacteria (comprising 12% and 14% of transgenic and control pooled datasets). These proportions are shown graphically in Figure 3.

Several other phylogenetic groups were present in low abundance, differing between transgenic and control libraries by 2% or less (Table 6). Within the less abundant groups, a statistically significant difference between the transgenic and control

libraries was observed in the recovery of clones assigned to the Bacteroidetes ($p \leq 0.05$, using χ^2 analysis). Two percent of the transgenic clones were assigned to Bacteroidetes, compared to 7% of the control clone sequences. The difference in the abundance of Bacteroidetes was attributable to the recovery of a significantly higher proportion of Flavobacteria from control soils (Table 6).

Major taxonomic groups were recovered with similar frequencies from replicate libraries, with the exception of Actinobacterial sequences, which were recovered differentially from control soils (Table 6).

Table 6. Relative abundance of 16S clones with respect to different bacterial taxa from PPO over-expressing and non-transgenic hybrid aspen rhizospheres

Phylogenetic group	Relative bacterial clone abundance (%)					
	rRNA gene library from transgenic rhizosphere			rRNA gene library from control rhizosphere		
	Rep A	Rep B	**Pooled**	Rep A	Rep B	**Pooled**
Genera incertae sedis BRC1	0	1	**<1**	0	0	**0**
Thermomicrobia	1	0	**<1**	0	0	**0**
Genera incertae sedis TM7	2	0	**1**	1	0	**<1**
Spirochaetes	0	0	**0**	1	0	**<1**
Chlamydiae	0	0	**0**	0	2	**1**
Cyanobacteria	0	2	**1**	1	2	**1**
Acidobacteria	2	2	**2**	2	1	**2**
Verrucomicrobia	5	2	**3**	2	2	**2**
Firmicutes	3	2	**3**	2	2	**2**
Chloroflexi	2	1	**1**	1	5	**3**
Gemmatimonadetes	4	8	**6**	3	3	**3**
Actinobacteria	9	8	**9**	12 *	2	**7**
Bacteroidetes	0	3	**2**	8	6	**7 ****
Flavobacteria	*0*	*2*	***1***	*6*	*5*	***6 *****
Sphingobacteria	*0*	*1*	***<1***	*2*	*1*	***2***
Proteobacteria	73	71	**72**	68	76	**72**
Alphaproteobacteria	*35*	*38*	***36***	*34*	*40*	***37***
Betaproteobacteria	*11*	*22*	***16***	*13*	*17*	***15***
Deltaproteobacteria	*12*	*4*	***8***	*6*	*5*	***6***
Gammaproteobacteria	*15*	*8*	***12***	*15*	*14*	***14***

Clone sequences (130/replicate library) were classified using the Ribosomal Database Naïve Bayesian Classifier. ** indicates a statistically significant difference ($p \leq 0.05$) in clone abundance, using pooled datasets. P-values were based on a χ^2 test with Yates Correction, or Fisher's Exact test. Statistically significant differences for comparisons of replicates within each library type are flagged with *.

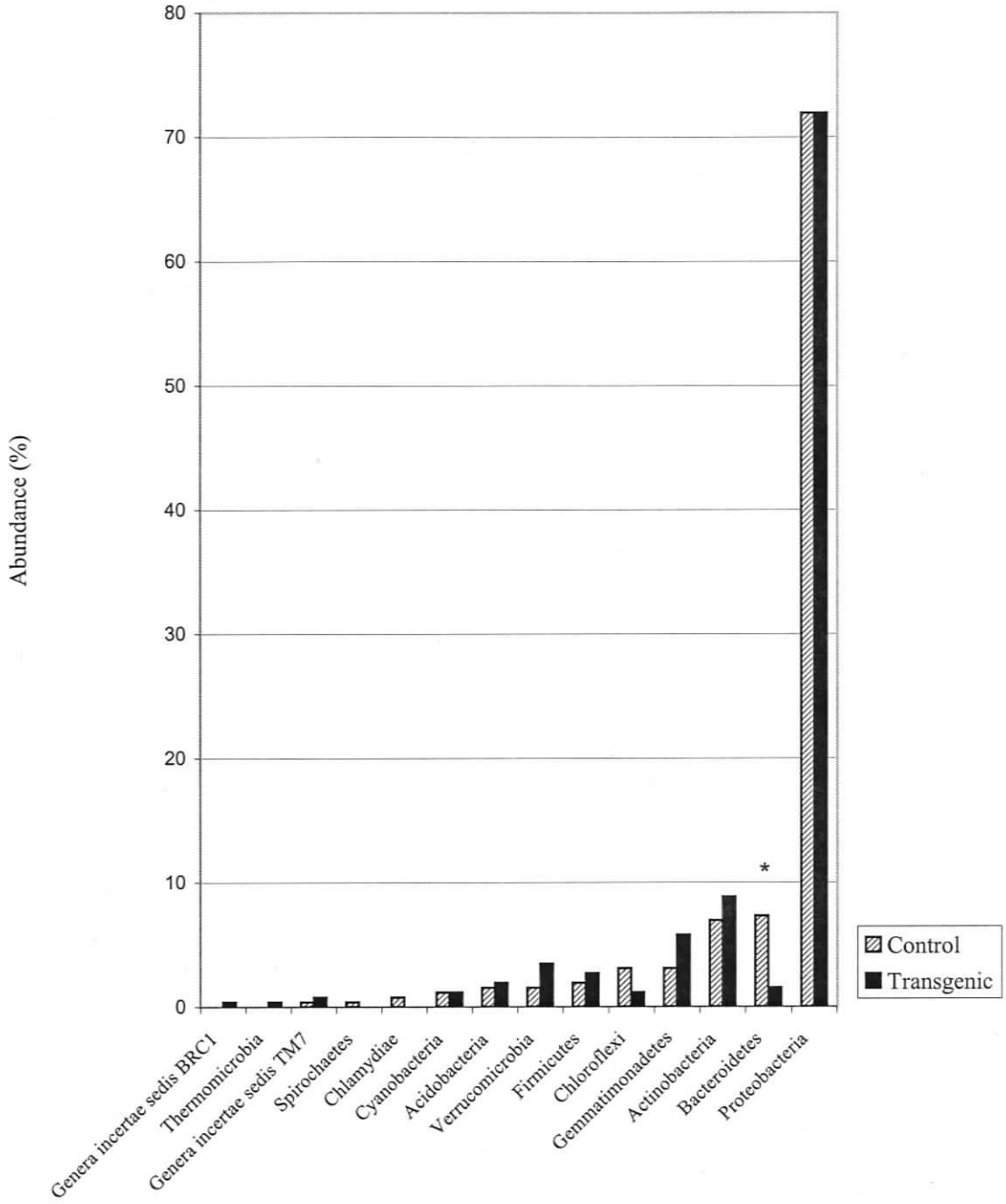


Figure 2. Taxonomic composition of 16S rRNA gene libraries obtained from PPO over-expressing and non-transgenic hybrid aspen rhizospheres. Replicate libraries were pooled, for a total of 260 clones/library. Clone sequences were assigned to taxonomic groups using the Naïve Bayesian rRNA Classifier on the Ribosomal Database website. "*" indicates a significant difference in clone abundance between control and transgenic libraries ($\chi 2$ test, $p \leq 0.05$).

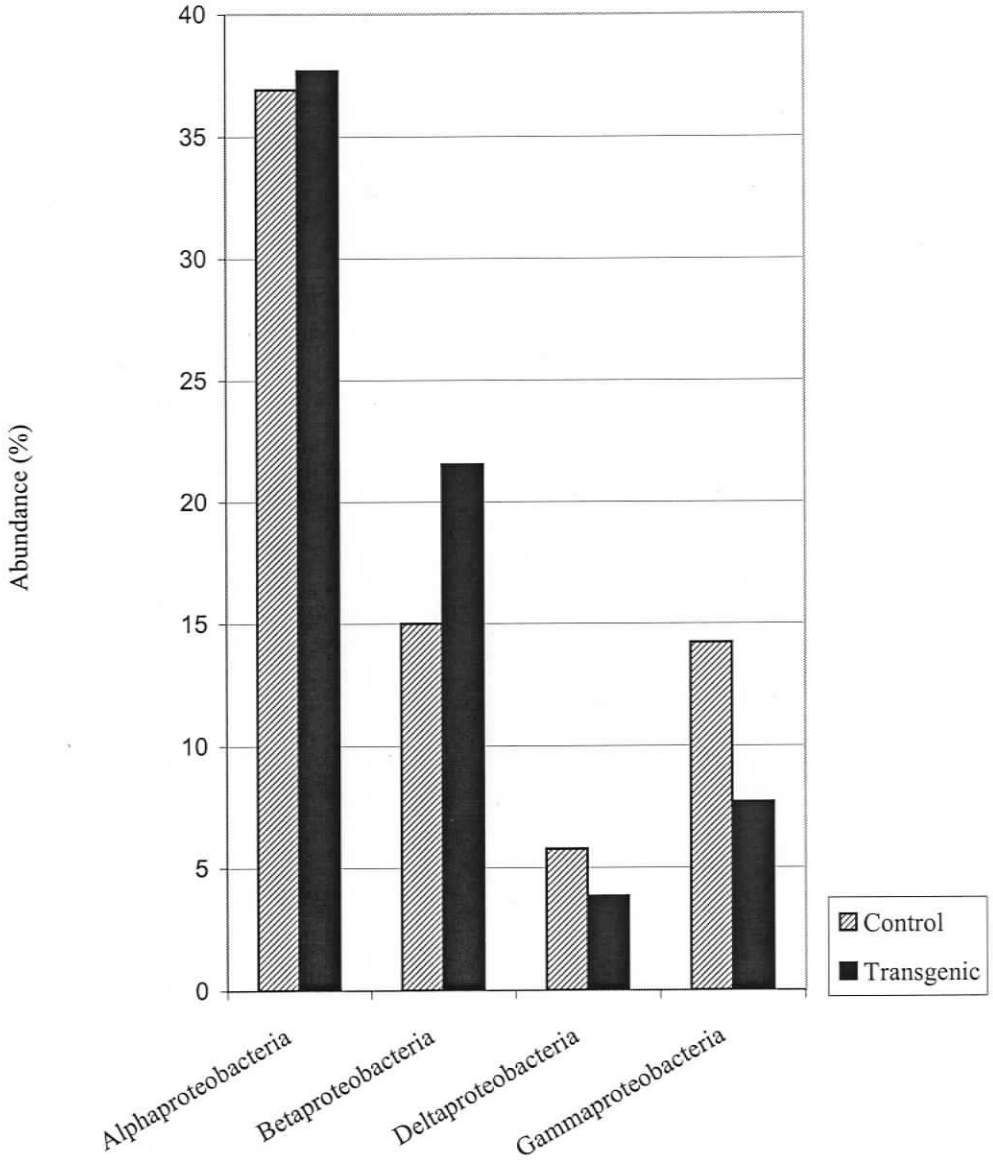


Figure 3. Relative abundance of proteobacterial groups, based on 16S rRNA gene sequences derived from the rhizospheres of PPO over-expressing and non-transgenic hybrid aspen. No significant differences in clone abundance were detected using $\chi 2$ analysis of pooled datasets.

3.3.2 Bacterial diversity in 16S rRNA gene libraries

In order to describe and compare the relative diversity of the bacterial rhizosphere communities, 16S rRNA gene sequences sharing 97% sequence similarity were grouped into operational taxonomic units (OTUs), and several diversity parameters were calculated using pooled datasets.

Both transgenic and control libraries were characterized by high levels of species richness, defined as the number of different OTUs in the sample (Table 7). This was reflected by the recovery of high numbers of OTUs from each library, as well as high shannon diversity values. The shannon diversity values, which essentially described the underlying complexity of the communities, were 4.87 and 4.73 for the transgenic and control libraries. For reference, if each OTU were recovered only once from each library, the maximum possible shannon diversity values would have been 5.09 and 5.01, for the two libraries respectively. The Chao and ACE richness estimators, which were used to estimate and compare the number of OTUs in the underlying communities, also supported the observation that there were very high levels of richness in the rhizospheres. After sequencing 260 clones, a slightly higher richness was found in the transgenic, compared to the control library, as reflected by the richness indices. However, it was not possible to state with confidence that the richness in the two soils differed, since the values for each richness index calculated to describe the transgenic library fell within the 95% confidence interval for the same estimate for the control library.

The second aspect of diversity that was compared was the level of evenness within the libraries. Evenness or dominance indices describe the relative abundance of OTUs or species. Dominance in both the libraries was low and levels of evenness were high (Table 7), as reflected by the shannon evenness and simpson's indices. The control library was characterized by a slightly higher level of dominance than the library derived from the rhizospheres of the transgenic trees. This was also reflected by the berger parker index, which showed that the most dominant OTU was recovered infrequently from both libraries, but was slightly more abundant in the control library.

Table 7. Diversity indices for 16S bacterial OTUs associated with PPO over-expressing and non-transgenic hybrid aspen rhizospheres.

Diversity parameter	Diversity Index/ Richness estimator	Rhizosphere soil type	
		Transgenic	Control
Species richness (95% CI indicated in brackets)	Number of OTUs	163	151
	Shannon (H')	4.87 (4.75-4.99)	4.73 (4.61-4.85)
	Chao Richness	451 (331-654)	369 (277-529)
	Ace Richness	502 (377-700)	447 (336-626)
Evenness or Dominance	Shannon Evenness (E)	0.956	0.943
	Simpson's (D)	0.007	0.009
	Berger Parker (d)	0.031	0.046

Replicate libraries were pooled, for a total of 260 clones/library. Bacterial OTUs were defined as sequences as having at least 97% sequence similarity. Where appropriate, 95% confidence intervals are indicated in brackets.

Ninety-eight OTUs were recovered only from control rhizosphere libraries (Table 8), 58 were recovered from both the transgenic and control libraries (Table 9), and 108 were recovered only from the transgenic libraries (Table 10). In many cases, each OTU contained only one sequence, and comprised less than 1% of the clone library. Clones assigned to OTU 90 were recovered with the highest frequency from both libraries, accounting for 5% of the clones recovered from control soil and 3% of the transgenic soil clones. This OTU was most closely related to the genus *Burkholderia,* and this taxonomic assignment was supported with a 100% bootstrap confidence estimate (Table 9).

Table 8. Taxonomic Assignment of 16S OTUs unique to control rhizosphere libraries, classified using the RDP Naïve Bayesian Classifier. The numbers following the Genus classification are bootstrap confidence estimates (%). Each pooled dataset consisted of 260 sequences, derived from the rhizosphere soils of two control hybrid aspen.

% Clones in control library	OTU	Genus	Bootstrap estimate	Major Group
<1	247	*Acidobacterium*	43%	Acidobacteria
<1	192	*Friedmanniella*	25%	Actinobacteria
<1	205	*Arthrobacter*	100%	
<1	206	*Sphaerobacter*	74%	
<1	232	*Mycobacterium*	100%	
<1	245	*Nocardioides*	42%	
<1	182	*Actinopolymorpha*	15%	
<1	194	*Conexibacter*	76%	
1	175	*Flavobacterium*	74%	Bacteroidetes
<1	222	*Weeksella*	25%	
<1	201	*Arcicella*	11%	
<1	214	*Pedobacter*	57%	
<1	253	*Microscilla*	19%	
<1	262	*Pedobacter*	93%	
<1	213	*Parachlamydia*	76%	Chlamydiae
<1	216	*Rhabdochlamydia*	35%	
<1	174	*Anaerolinea*	90%	Chloroflexi
<1	215	*Anaerolinea*	29%	
1	220	*Anaerolinea*	94%	
<1	241	*Anaerolinea*	77%	
<1	252	*Anaerolinea*	48%	
<1	256	*Anaerolinea*	97%	
1	227	*Gloeocapsa*	34%	Cyanobacteria
<1	197	*Bacillus*	65%	Firmicutes
<1	254	*Saccharococcus*	9%	
<1	185	*Thermoterrabacterium*	18%	
1	190	*Natroniella*	8%	
<1	188	*Gemmatimonas*	44%	Gemmatimonadetes
<1	207	*Gemmatimonas*	69%	
<1	218	*Gemmatimonas*	80%	
<1	178			Genera incertae sedis TM7
<1	176	*Spirochaeta*	97%	Spirochaetes
1	169	*Rhizobium*	72%	Alphaproteobacteria
1	171	*Agrobacterium*	76%	
1	173	*Rhizobium*	66%	
1	177	*Brucella*	62%	
<1	179	*Bosea*	70%	
<1	181	*Asticcacaulis*	67%	
<1	183	*Xanthobacter*	22%	
<1	187	*Paracraurococcus*	13%	

Table 8 continued

% Clones in control library	OTU	Genus	Bootstrap estimate	Major Group
<1	193	*Parvibaculum*	26%	
1	198	*Stella*	17%	
1	202	*Angulomicrobium*	43%	
<1	203	*Devosia*	51%	
<1	210	*Stella*	31%	
<1	211	*Angulomicrobium*	22%	
<1	217	*Caulobacter*	94%	
<1	219	*Stella*	24%	
<1	223	*Saccharibacter*	14%	
<1	225	*Inquilinus*	17%	
<1	231	*Phenylobacterium*	98%	
<1	234	*Rhodoplanes*	59%	
<1	235	*Sinorhizobium*	40%	
<1	243	*Stella*	34%	
<1	248	*Acidisphaera*	29%	
<1	249	*Mycoplana*	30%	
<1	250	*Stella*	80%	
<1	258	*Kozakia*	26%	
<1	260	*Kozakia*	16%	
<1	261	*Rhodoplanes*	40%	
<1	264	*Acetobacter*	53%	
1	172	*Duganella*	80%	Betaproteobacteria
<1	189	*Duganella*	28%	
1	208	*Methylobacillus*	70%	
<1	212	*Thiobacter*	55%	
<1	228	*Thiobacter*	37%	
<1	233	*Pandoraea*	100%	
<1	236	*Burkholderia*	100%	
<1	238	*Nitrosospira*	100%	
<1	239	*Thiobacter*	41%	
1	244	*Methylophilus*	89%	
<1	257	*Burkholderia*	100%	
<1	263	*Burkholderia*	46%	
<1	168	*Anaeromyxobacter*	9%	Deltaproteobacteria
<1	170	*Desulfobacca*	19%	
1	186	*Desulfobacca*	6%	
<1	196	*Chondromyces*	64%	
<1	199	*Chondromyces*	25%	
<1	200	*Bdellovibrio*	100%	
<1	221	*Haliangium*	63%	
<1	226	*Anaeromyxobacter*	19%	
<1	229	*Desulforhabdus*	10%	
<1	242	*Anaeromyxobacter*	8%	
<1	246	*Anaeromyxobacter*	23%	
<1	259	*Haliangium*	66%	
<1	167	*Chromatium*	24%	Gammaproteobacteria
<1	180	*Aquicella*	49%	

Table 8 continued

% Clones in control library	OTU	Genus	Bootstrap estimate	Major Group
1	184	*Isochromatium*	13%	Gammaproteobacteria
<1	191	*Rhodanobacter*	75%	
<1	195	*Aquicella*	100%	
<1	204	*Pseudomonas*	80%	
<1	209	*Luteimonas*	76%	
1	224	*Fulvimonas*	29%	
<1	230	*Legionella*	100%	
<1	237	*Rickettsiella*	25%	
<1	240	*Silanimonas*	32%	
<1	251	*Alkalilimnicola*	17%	
<1	255	*Aquicella*	69%	

Table 9. Taxonomic Assignment of 16S OTUs shared by transgenic and control rhizosphere libraries based on classification using the Ribosomal Database Naïve Bayesian Classifier. The numbers following the Genus classification are bootstrap confidence estimates (%). For transgenic and control trees, pooled datasets derived from rhizosphere soils consisted of 260 sequences.

% Clones					
Control	Transgenic	OTU	Genus	Bootstrap estimate	Major Group
1	<1	17	*Acidobacterium*	100%	Acidobacteria
<1	<1	28	*Acidobacterium*	100%	
2	1	23	*Microbacterium*	99%	Actinobacteria
1	<1	59	*Nocardioides*	90%	
<1	<1	94	*Mycobacterium*	99%	
<1	<1	4	*Agreia*	42%	
<1	<1	18	*Conexibacter*	48%	
<1	1	120	*Pedobacter*	57%	Bacteroidetes
4	<1	160	*Pedobacter*	89%	
<1	<1	130	*Dactylococcopsis*	16%	Cyanobacteria
1	1	87	*Gemmatimonas*	92%	Gemmatimonadetes
1	<1	113	*Gemmatimonas*	59%	
<1	<1	2	*Balneimonas*	35%	Alphaproteobacteria
1	2	11	*Devosia*	93%	
2	1	12	*Devosia*	85%	
1	1	26	*Bradyrhizobium*	100%	
1	<1	33	*Kozakia*	11%	
3	2	39	*Rhodoplanes*	43%	
1	1	42	*Rhizobium*	80%	
<1	1	46	*Angulomicrobium*	37%	
<1	<1	47	*Brucella*	78%	
1	1	50	*Inquilinus*	53%	
<1	<1	52	*Caulobacter*	79%	
1	<1	55	*Kozakia*	49%	
2	2	60	*Mesorhizobium*	99%	
<1	<1	77	*Caulobacter*	100%	
<1	<1	78	*Rhizobium*	100%	
<1	1	80	*Devosia*	50%	
1	<1	93	*Saccharibacter*	14%	
<1	<1	134	*Rhodoplanes*	66%	
<1	<1	135	*Ancylobacter*	15%	
2	1	137	*Rubritepida*	20%	
<1	1	143	*Inquilinus*	40%	
<1	<1	144	*Asticcacaulis*	80%	
<1	<1	145	*Rhizobium*	93%	
<1	<1	148	*Brevundimonas*	100%	
1	<1	158	*Balneimonas*	15%	
<1	1	40	*Thiobacter*	35%	Betaproteobacteria

Table 9 continued

% Clones					
Control	Transgenic	OTU	Genus	Bootstrap estimate	Major Group
1	1	48	*Rhodocyclus*	57%	
1	1	53	*Variovorax*	52%	
1	3	64	*Leptothrix*	82%	
5	3	90	*Burkholderia*	100%	
<1	<1	97	*Telluria*	76%	
1	1	104	*Wautersia*	100%	
<1	<1	133	*Thiomonas*	31%	
<1	<1	154	*Thiobacter*	52%	
<1	<1	6	*Chondromyces*	51%	Deltaproteobacteria
<1	<1	98	*Melittangium*	4%	
2	1	3	*Thiorhodococcus*	14%	Gammaproteobacteria
1	1	14	*Rhodanobacter*	94%	
3	2	29	*Lysobacter*	13%	
<1	1	43	*Chromatium*	11%	
2	1	109	*Pseudomonas*	100%	
<1	<1	125	*Hydrocarboniphaga*	63%	
<1	<1	165	*Nevskia*	69%	
<1	1	24	*Opitutus*	100%	Verrucomicrobia
<1	1	30	*Opitutus*	99%	
1	<1	89	*Opitutus*	100%	

Table10. Taxonomic Assignment of 16S OTUs unique to transgenic rhizosphere libraries based on classification using the Ribosomal Database Naïve Bayesian Classifier. The numbers following the Genus classification are bootstrap confidence estimates (%). Each pooled dataset consisted of 260 sequences, derived from the rhizosphere soils of two transgenic hybrid aspen.

% Clones in transgenic library	OTU	Genus	Bootstrap estimate	Major Group
<1	31	*Acidobacterium*	100%	Acidobacteria
<1	117	*Acidobacterium*	100%	
<1	118	*Acidobacterium*	32%	
<1	10	*Conexibacter*	67%	Actinobacteria
<1	22	*Sporichthya*	42%	
<1	36	*Arthrobacter*	98%	
<1	56	*Actinopolymorpha*	34%	
1	75	*Nocardioides*	95%	
<1	84	*Ornithinicoccus*	69%	
1	92	*Conexibacter*	96%	
<1	107	*Sporichthya*	30%	
<1	110	*Cryptosporangium*	14%	
<1	116	*Nocardioides*	27%	
<1	142	*Nocardioides*	24%	
<1	151	*Longispora*	58%	
<1	166	*Arthrobacter*	98%	
<1	131	*Actinopolymorpha*	15%	
<1	122	*Conexibacter*	77%	
<1	153	*Flavobacterium*	100%	Bacteroidetes
<1	1	*Anaerolinea*	76%	Chloroflexi
<1	99	*Anaerolinea*	69%	
<1	114	*Anaerolinea*	78%	
<1	111	*Microcystis*	23%	Cyanobacteria
<1	127	*Microcystis*	8%	
<1	15	*Ureibacillus*	20%	Firmicutes
<1	123	*Paenibacillus*	38%	
<1	19	*Sulfuricurvum*	52%	
1	86	*Thermohalobacter*	14%	
1	112	*Anaeromusa*	1%	
1	27	*Gemmatimonas*	67%	Gemmatimonadetes
<1	54	*Gemmatimonas*	75%	
<1	57	*Gemmatimonas*	82%	
<1	102	*Gemmatimonas*	70%	
1	139	*Gemmatimonas*	66%	
<1	146	*Gemmatimonas*	99%	
<1	155	*Gemmatimonas*	96%	
<1	163	*Gemmatimonas*	91%	
<1	49	*Thermomicrobium*	37%	Thermomicrobia
1	5	*Verrucomicrobium*	52%	Verrucomicrobia

Table 10 continued

% Clones in transgenic library	OTU	Genus	Bootstrap estimate	Major Group
<1	69	*Verrucomicrobium*	45%	
<1	149			Genera incertae sedis BRC1
<1	41			Genera incertae sedis TM7
<1	63			
<1	7	*Devosia*	18%	Alphaproteobacteria
<1	20	*Phaeospirillum*	39%	
<1	34	*Pedomicrobium*	73%	
<1	37	*Stella*	29%	
<1	38	*Methylocapsa*	48%	
<1	44	*Angulomicrobium*	37%	
1	51	*Caulobacter*	79%	
<1	58	*Brevundimonas*	43%	
<1	61	*Rhodoplanes*	45%	
<1	62	*Acidiphilium*	19%	
<1	66	*Rhodocista*	29%	
<1	71	*Rubritepida*	22%	
<1	74	*Acidisphaera*	17%	
<1	76	*Devosia*	74%	
<1	82	*Angulomicrobium*	49%	
2	83	*Acidisphaera*	13%	
1	85	*Xanthobacter*	86%	
<1	91	*Phenylobacterium*	100%	
<1	95	*Rhodoplanes*	36%	
<1	100	*Rhizobium*	35%	
<1	101	*Rhodoplanes*	26%	
<1	105	*Phaeospirillum*	43%	
<1	106	*Acidocella*	22%	
1	108	*Hyphomicrobium*	100%	
<1	115	*Ensifer*	58%	
<1	121	*Stella*	77%	
<1	124	*Methylobacterium*	96%	
<1	126	*Caedibacter*	7%	
<1	129	*Angulomicrobium*	15%	
<1	138	*Devosia*	93%	
<1	141	*Swaminathania*	36%	
<1	150	*Stella*	68%	
<1	156	*Stella*	52%	
<1	157	*Inquilinus*	87%	
<1	162	*Pseudaminobacter*	52%	
<1	164	*Phaeospirillum*	40%	
1	21	*Thiobacter*	12%	Betaproteobacteria
<1	35	*Xenophilus*	21%	
2	45	*Ramlibacter*	60%	
<1	67	*Comamonas*	26%	
<1	68	*Caenibacterium*	16%	

Table 10 continued

% Clones in transgenic library	OTU	Genus	Bootstrap estimate	Major Group
<1	119	*Schlegelella*	34%	
<1	136	*Leptothrix*	29%	
<1	152	*Gallionella*	31%	
<1	159	*Burkholderia*	100%	
<1	9	*Anaeromyxobacter*	6%	Deltaproteobacteria
1	13	*Stigmatella*	10%	
<1	16	*Chondromyces*	99%	
2	32	*Corallococcus*	21%	
<1	65	*Chondromyces*	98%	
<1	70	*Plesiocystis*	26%	
<1	72	*Malonomonas*	36%	
1	79	*Anaeromyxobacter*	30%	
<1	88	*Chondromyces*	100%	
<1	103	*Thermodesulforhabdus*	8%	
<1	128	*Desulforhabdus*	19%	
<1	132	*Nannocystis*	29%	
<1	147	*Desulfovirga*	13%	
3	8	*Cellvibrio*	100%	Gammaproteobacteria
<1	25	*Nevskia*	62%	
1	73	*Nevskia*	51%	
<1	81	*Chromohalobacter*	46%	
<1	96	*Thiococcus*	6%	
<1	140	*Fulvimonas*	13%	
<1	161	*Legionella*	18%	

3.3.3 Collector's curves for 16S rRNA gene libraries

To evaluate sampling progress, and to compare the relative diversity in the control and transgenic libraries, collector's curves were constructed. Once the diversity within a community has been sampled to its full extent, a collector's curve would be expected to plateau. The collector's curves for both libraries in this study were nearly linear, indicating that only a fragment of the population had been sampled, and that unique sequences would continue to be discovered if more clones were sequenced. As suggested by the diversity indices (Table 7), comparison of the collector's curves for the pooled and transgenic libraries indicated that a greater richness of OTUs were recovered from the transgenic library after sampling 260 sequences, relative to the control library (Figure 4).

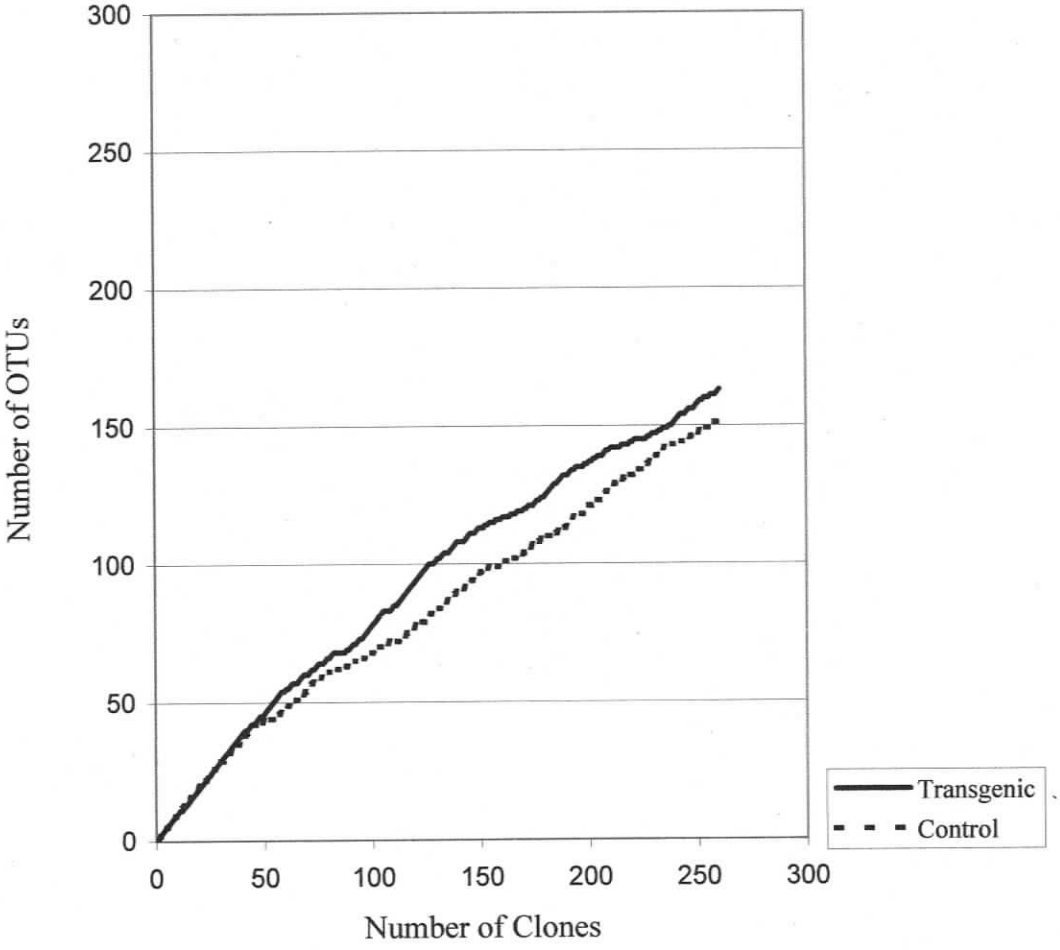


Figure 4. Collector's curves for 16S rRNA gene libraries derived from the rhizospheres of transgenic PPO over-expressing and control aspen (replicates pooled, for a total of 260 clones/library). Sequences sharing 97% similarity were grouped in to OTUs

3.3.4 Significance of difference between coverage curves for 16S libraries

In contrast to library comparisons based on differences in phylogenetic distributions and diversity indices, analysis using the ∫-Libshuff computer program does not rely on the definition of an OTU (Schloss, 2005). The program calculates a Cramer-von Mises-type test statistic by comparing differences between homologous and heterologous coverage curves for two libraries over a range of evolutionary distances, and a Monte Carlo resampling procedure is used to determine whether the differences observed are significant. Traditionally, the Cramer-Von Mises statistic is used to test the quality of a curve fit, but when applied to 16S libraries, the statistic measures the number of sequences that are unique to a library when two are compared.

A gene library represents a sample of a larger population, and the coverage of the population by a sample can be calculated by applying the formula of Good (1953), $C_x = (1- (N_x/n))$, where N_x in this case = the number of unique sequences in the library, and n=the total number of sequences in the library. Since it is possible to define operational taxonomic units using a range of evolutionary distances, a homologous coverage curve can be generated by plotting coverage vs evolutionary distance. The homologous coverage curve describes how well the sample represents the entire library using a range of evolutionary distances to define uniqueness. A heterologous coverage curve can be generated in the same manner, using the formula $C_{xy}=1-N_{xy}/n$, where N_{xy} is the number of sequences in a sample of X that are not found in a sample of Y, and n is the number of sequences in the sample of X (Singleton *et al.*, 2001).

Replicate 16S rRNA gene libraries, derived from the two individual trees of the same type were pooled, for a total of 260 clones/library. Analysis of the coverage curves using ∫-Libshuff detected significant differences between the bacterial lineages associated with the rhizospheres of the control and transgenic plants. Moreover, it suggested that the sequences in the control library were a subset of those in the transgenic library. The Cramer-von Mises statistic (dCxy) is essentially a sum of the differences between the homologous and heterologous coverage curves over the entire range of evolutionary distances tested. Comparison of the libraries when the transgenic library was treated as the homologous library resulted in a statistically significant dCxy value (34.7, p=0.001)

(Figure 5a) indicating that the transgenic library contained sequences of one or more taxa not found in the control library. When the control library was the homologous library, the calculated dCxy value was not statistically significant (11.7, p= 0.278) (Figure 5b) suggesting that all of the taxa found in this library were also found in the transgenic library.

In order to determine whether this difference was reflected by the replicate libraries derived from individual trees, each combination was compared. All comparisons suggested that each replicate library had been drawn from the same microbial population, except for the comparison between the two control libraries. The results suggested that the library sequences derived from one control tree (Tree 717-11) were a subset of those in the other (Tree 717-10). When tree 717-10 was treated as the homologous library, the dCxy (55.4) was significant (p=0.0001), indicating that it contained sequences not present in the second library. When the 717-11 library was treated as the homologous library, the dCxy value was insignificant, suggesting that all of the sequences in this library were also present in the 717-10 dataset. Although differences between all other replicate libraries were statistically insignificant (p>0.0043), the lowest p-values were obtained when the library derived from tree 717-11 was treated as the heterologous library, suggesting that this library was the most dissimilar.

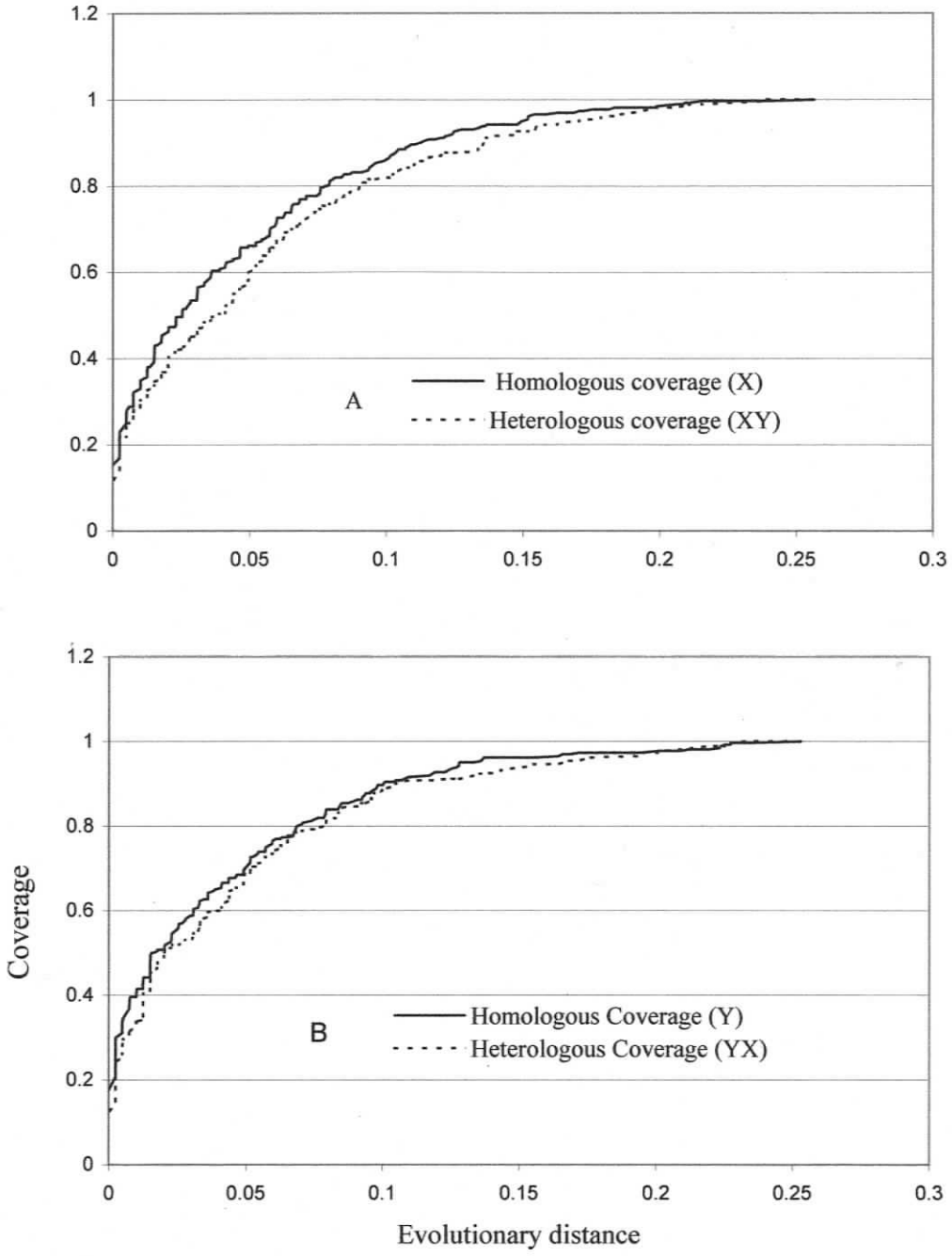


Figure 5. Comparison of 16S rRNA gene libraries from transgenic PPO over-expressing and unmodified hybrid aspen rhizospheres based on coverage curves. (A) Transgenic (X) compared to control (Y) libraries (dCxy = 34.7, p = 0.012 *) (B) Control (Y) compared to transgenic (X) libraries (dCyx = 11.7, p=0.278). Replicate libraries were pooled, for a total of 260 clones/library. The critical p-value for comparing 2 libraries was 0.025.

3.3.5 Similarity coefficients for 16S rRNA gene libraries

Similarity coefficients, which reflect the proportions of shared OTUs, were calculated by performing pairwise comparisons of replicate libraries. The lowest similarity coefficients (0.221-0.288) were obtained when replicate libraries were compared with library 717-11 (Table 11). This supported the ∫-Libshuff analysis, which indicated that library 717-11 was the most dissimilar.

Table 11. Pairwise comparisons of bacterial 16S rRNA gene libraries obtained rhizosphere soil samples associated with PPO over-expressing and control hybrid aspen.

		Similarity coefficient		
Rhizosphere sample		Transgenic	Control	
		Tree 19-9	Tree 717-10	Tree 717 – 11
Transgenic	Tree 19-4	0.289	0.296	0.228
	Tree 19-9		0.260	0.221
Control	Tree 717-10			0.227

Clones having 97% sequence similarity were grouped into operational taxonomic units (OTUs). The similarity coefficients reflect proportions of shared OTUs and were calculated by the method of McCaig *et al.*, 1999: $S = 2C/(A+B)$
where C= # shared OTUs, A= # OTUs in Library A, B= # OTUs in Library B.

3.4 Comparison of bacterial communities based on CPN-60 gene libraries

3.4.1 Taxonomic composition of CPN-60 libraries

Clones containing CPN-60 inserts were assigned to phylogenetic groups by comparing nucleotide and translated amino acid sequences to the CPN-60 database using

FASTA and Blast-p search algorithms, respectively. The reference nucleotide sequences most homologous to library sequences are summarized in Table 12, and are arranged by the frequencies of recovery of major phylogenetic groups from transgenic and control libraries. Analysis of 430 sequences (215 from the rhizospheres of control plants, and 215 from the rhizospheres of transgenic plants) revealed the presence of 337 unique nucleotide sequences, which were 62-90% identical to reference nucleotide sequences in the database. Interestingly, all library sequences were most closely associated with bacterial reference sequences, with the exception of one, whose nearest database neighbour was fungal in origin. This sequence was recovered from the control library, and shared low identity (67%) with *Candida albicans*. Clustering the sequences based on their nearest neighbours in the database revealed that groups recovered only from the control library were present in low abundance, each accounting for less than 2% of the total library contents. The same was true for sequence groups recovered solely from the transgenic soils. The groups most frequently recovered were present in both control and transgenic libraries, and these sequences were most closely associated with the Bacteroidetes. One hundred fifty two sequences unique to the study were most closely associated with *Flavobacterium ferrugineum*, accounting for 48% and 54% of the clones recovered from control and transgenic libraries, respectively, with DNA identities ranging from 71-85%. Forty nine sequences unique to the study were most closely associated with *Sphingobacterium multivorum*, accounting for 18% and 12% of the clones recovered from control and transgenic libraries respectively, with DNA identities ranging from 72-81% (Table 12).

The relative frequencies of recovery of DNA sequences associated with major phylogenetic groups from control and transgenic soils are shown in Table 13. In contrast to the 16S libraries, the majority of clone sequences recovered were most closely associated with the Bacteroidetes/Chlorobi group, accounting for 79% and 83% of the transgenic and control rhizosphere libraries respectively. Nucleotide sequences most closely associated with the Proteobacteria accounted for 11% and 6% of the transgenic and control libraries, and several other phylogenetic groups were recovered with frequencies of 5% or less.

Table 12. Taxonomic assignment of CPN-60 sequences, showing occurrence of each in control and transgenic libraries. Classification was based on the nearest DNA sequence neighbor in the CPN60 database, obtained using a FASTA search.

Phylogenetic group		Nearest CPN-60 database neighbor	Genbank Nucleotide Accession number	Number of unique sequences	Clones in control library (%)	Clones in transgenic library (%)	Nucleotide sequence identity (%)
Chlamydiae		Chlamydia muridarum MoPn ATCCVR-123	NC_002620	1	<1		69
		Chlamydophila pneumoniae CWL029	NC_000922	2	1		70-73
Bacteroidetes	Bacteroides	Prevotella heparinolytica ATCC35895	AY691283	1	<1		71
	Flavobacteria	Flavobacterium hydatis ATCC29551	AF335327	1	<1		82
Chlorobi		Pelodictyon luteolum DSM 273		1	<1		71
		Chlorobium limicola DSM 245	NZ_AAHJ01000037	1	<1		64
Firmicutes		Peptoniphilus asaccharolyticus ATCC14963	AY691273	1	<1		62
		Streptococcus thermophilus CNRZ1066	NC_006449	1	<1		99
		Paenibacillus illinoisensis z898		3	1		62-83
Proteobacteria	Alpha	Bartonella sp. NVH1	AF071193	1	<1		70
		Bartonella doshiae R18(T) NCTC 12862	AF014832	1	1		75
		Bartonella koehlerae ATCC700693T		1	<1		75
	Beta	Methylophilus methylotrophus ATCC53528D	AY837568	1	<1		79
	Delta/Epsilon	Bdellovibrio bacteriovorus ATCC15356	NC_005363	4	2		76-87
		Campylobacter lari RM2100	NZ_AAFK01000001	1	<1		64
Fungi		Candida albicans ATCC14053	AY837557	1	<1		67

Table 12 continued

Phylogenetic group		Nearest CPN-60 database neighbor	Genbank Accession number	# unique sequences	Control library (%)	Transgenic library (%)	Nucleotide sequence identity (%)
Chlamydiae		Chlamydia trachomatis A/HAR-13	CP000051	2	<1	<1	72-74
		Parachlamydia sp. UWE25	NC_005861	6	2	1	67-70
Bacteroidetes	Sphingobacteria	Cytophaga hutchinsonii ATCC33406D	AY922373	8	2	2	67-77
		Sphingobacterium multivorum ATCC35656	DQ212063	49	18	12	72-81
	Flavobacteria	Flavobacterium johnsoniae UW101 ATCC17061	AAPM01000007	7	2	1	70-95
		Flavobacterium ferrugineum ATCC13524	AF335328	152	48	54	71-85
		Chryseobacterium meningosepticum ATCC13253	AF440231	13	5	3	65-74
		Chryseobacterium indologenes ATCC29897	AY123684	9	1	3	74-90
Chlorobi		Chlorobium tepidum TLS	NC_002932	3	1	<1	72
		Chlorobium chlorochromatii CaD3	NC_007514	8	7	<1	70-73
Firmicutes		Aneurinibacillus aneurinolyticus ATCC12856	AY691204	3	<1	1	62-64
		Alicyclobacillus acidocaldarius CECT4328		2	<1	<1	63-65
Chloroflexi		Chloroflexus aurantiacus J-10-fl	NZ_AAAH02000009	3	<1	1	63-68
Proteobacteria	Delta/Epsilon	Desulfotalea psychrophila LSv54	NC_006138	2	<1	<1	68
	Gamma	Unknown sp. X-bacteria symbiont of Amoeba proteus	M86549	5	<1	2	76-87
		Legionella micdadei	X57520	3	<1	1	72-77
Fusobacteria		Fusobacterium prausnitzii ATCC27768	AY691258	7	3	1	72-74
Chlamydiae		Chlamydia pecorum FcStra	AF109789	1		<1	67

Table 12 continued

Phylogenetic group		Nearest cpn60 database neighbor	Genbank Accession number	# unique sequences	Control library (%)	Transgenic library (%)	Nucleotide sequence identity (%)
Bacteroidetes	Bacteroides	Bacteroides ureolyticus ATCC43606D	AY922366	1		<1	67
		Bacteroides uniformis ATCC8492	AF440237	1		<1	72
Chlorobi		Pelodictyon phaeoclathratiforme BU-1	NC_007512	2		1	70
Acidobacteria		Solibacter usitatus Ellin6076	NZ_AAIA01000003	1		<1	81
Firmicutes		Lactobacillus gasseri ATCC9857	AY123652	2		1	64
		Clostridium intestinale ATCC49213	AY691231	1		<1	65
		Eubacterium ruminantium ATCC17233	AY691254	1		<1	64
		Ruminococcus obeum ATCC29174	AY691300	1		<1	68
		Lactobacillus homohiochii ATCC15434	AF429685	1		<1	60
		Halothermothrix orenii H168	NZ_AAOZ01000008	2		1	65
Proteobacteria	Alpha	Rickettsia prowazekii Madrid E	NC_000963	1		<1	70
		Magnetospirillum gryphiswaldense	BX640510	1		<1	76
		Agrobacterium vitis NCPPB3268	AY837546	1		<1	69
		Bartonella henselae CAL-1	AF304020	1		<1	71
	Beta	Nitrospira multiformis ATCC25196	CP000103	1		<1	79
	Delta/Epsilon	Helicobacter hepaticus ATCC51449	NC_004917	1		<1	67
		Geobacter uraniumreducens Rf4	NZ_AAON01000017	1		<1	71
		Helicobacter pametensis ATCC51478	AY787945	1		<1	64
		Lawsonia intracellularis	AB218756	2		1	71-72
		Anaeromyxobacter dehalogenans 2CP-C	NZ_AAHD01000022	1		<1	70
	Gamma	Microbulbifer degradans 2-40	NZ_AABI03000012	4		2	81
		Providencia alcalifaciens ATCC9886	AY691291	1		<1	76
Deinococcus-Thermus		Deinococcus geothermalis DSM 11300	CP000359	1		<1	64
Planctomycetes		Blastopirellula marina DSM 3645	NZ_AANZ01000045	2		1	74-76
Total				337			62-90

Table 13. Relative abundance of clones recovered from CPN-60 libraries with respect to different taxa from soil samples obtained from the rhizospheres of transgenic and non-transgenic hybrid aspen. Taxonomic assignment of clone sequences (215/library) was based on nearest neighbours obtained using a FASTA search of DNA sequences catalogued in the CPN-60 database.

Phylogenetic group	Relative CPN-60 clone abundance %	
	Transgenic	Control
Fungi	**0**	**<1**
Deinococcus-Thermus	**<1**	**0**
Acidobacteria	**<1**	**0**
Chloroflexi	**1**	**0**
Planctomycetes	**1**	**0**
Fusobacteria	**1**	**3**
Chlamydiae	**2**	**4**
Firmicutes	**5**	**3**
Proteobacteria	**11**	**6**
Alphaproteobacteria	*2*	*2*
Betaproteobacteria	*<1*	*<1*
Deltaproteobacteria	*3*	*3*
Gammaproteobacteria	*5*	*1 **
Bacteriodetes/Chlorobi	**79**	**83**
Bacteroides	*1*	*<1*
Sphingobacteria	*14*	*20*
Flavobacteria	*62*	*57*
Chlorobi	*2*	*5*

* indicates a statistically significant difference ($p \leq 0.05$) in clone abundance, using pooled datasets. P-values were based on a χ^2 test with Yates Correction, or Fisher's Exact test.

Analysis of CPN-60 library sequences at the amino acid level revealed a similar taxonomic composition, with minor differences. Results of Blast-P searches for the most homologous amino acid sequences in the database are summarized in Table 14. Translation of the 337 unique nucleotide sequences resulted in 305 unique peptide sequences, and library sequences were 55-100% identical to reference sequences catalogued in the database. The relative frequencies of recovery of amino acid sequences associated with major phylogenetic groups from transgenic and control libraries are shown in Table 15. Seventy nine percent of the clones from each library were assigned to the Bacteroidetes/Chlorobi. Sequences most closely associated with the Proteobacteria were recovered with frequencies of 10% and 9% from transgenic and control soils respectively. The slight differences in taxonomic assignment of clones based on sequence comparisons at the nucleotide and amino acid levels reflect the uncertainty associated with assigning taxonomy to community members using sequence based methods, and highlight the importance of analyzing data in different ways. The only statistically significant difference between the transgenic (5%) and control (1%) libraries was observed in the recovery of clones assigned to the Gammaproteobacteria ($p \leq 0.05$, using Fisher's exact test). This difference was reflected at both the amino acid and nucleotide levels. Taxonomic composition of transgenic and control libraries based on analysis at the amino acid level, is shown graphically in Figure 6. Within the Bacteroidetes/Chlorobi, the majority of clones were assigned to the Flavobacteria, followed by the Sphingobacteria, and the Bacteroides. Only one sequence from each library was most closely associated with a member of the Chlorobi (Figure 7).

Table 14. Taxonomic assignment of 305 CPN60 peptide sequences, translated from 337 nucleotide sequences, showing occurrence of each in control and transgenic libraries. Classification was based on the nearest amino acid sequence neighbor in the CPN-60 database, obtained using a Blast-P search.

Phylogenetic group		Nearest CPN-60 database neighbor	Genbank peptide accession number	Number of unique sequences	Clones in control library (%)	Clones in transgenic library (%)	% peptide identity
Chlamydiae		Chlamydia muridarum MoPn ATCCVR-123	NP_296764	1	<1		72
		Chlamydophila caviae GPIC ATCC VR-813	NP_829507	1	<1		76
		Chlamydophila psittaci 6BC ATCCVR-125	AAV37130	1	<1		73
Bacteroidetes	Bacteroides	Dysgonomonas gadei z872		1	1		77
Chlorobi		Prosthecochloris aestuarii DSM 271	EAN22173	2	1		73-76
Proteobacteria	Alpha	Magnetospirillum magnetotacticum MS-1	ZP00055267	1	1		85
		Rhodospirillum rubrum	YP_425678	2	1		59-80
	Beta	Methylophilus methylotrophus ATCC53528D	AAV84352	1	<1		86
	Delta/Epsilon	Bdellovibrio bacteriovorus ATCC15356	NP_967123	4	2		81-96
		Arcobacter butzleri ATCC49616	AAZ94810	4	2		69-70
		Geobacillus stearothermophilus NUB36212		1	<1		57
		Desulfotomaculum ruminis ATCC23193	AAT96599	1	<1		63
Fungi		Candida albicans ATCC14053	AAV84341	1	<1		69
Chlamydiae		Simkania negevensis strain Z	AA063770	2	<1	<1	77
		Parachlamydia sp. UWE25	YP_007029	5	1	1	65-78
Bacteroidetes	Bacteroides	Rikenella microfusus ATCC29728	AAT96642	10	7	1	71-79
	Sphingobacteria	Rhodothermus marinus ITI 376	AAD37976	3	<1	1	76-77

Table 14 continued

Phylogenetic group		Nearest CPN-60 database neighbor	Genbank peptide accession number	Number of unique sequences	Clones in control library (%)	Clones in transgenic library (%)	% peptide identity
		Cytophaga hutchinsonii ATCC33406D	AAX24028	5	2	<1	79-81
		Sphingobacterium multivorum ATCC35656		38	18	14	74-90
	Flavobacteria	Flavobacterium hydatis ATCC29551	AAK32145	3	1	<1	63-93
		Flavobacterium ferrugineum ATCC13524	AAK32146	148	52	57	81-94
		Robiginitalea biformata HTCC2501	ZP_01120491	4	<1	1	67-78
		Flavobacterium johnsoniae UW101 ATCC 17601	EAS58989	3	1	1	98-100
Chlorobi		Chlorobium chlorochromatii CaD3	YP_379609	2	<1	<1	73-74
Firmicutes		Alicyclobacillus acidocaldarius CECT4328		3	1	<1	56-58
		Paenibacillus illinoisensis z898		3	1	1	57-89
		Ruminococcus obeum ATCC29174	AAT96650	3	<1	1	55-62
		Streptococcus thermophilus LMD-9	ZP_00389258	1	<1	0	99
Chloroflexi		Chloroflexus aurantiacus J-10-fl	ZP_00766918	6	<1	2	55-66
Proteobacteria	Delta/Epsilon	Anaeromyxobacter dehalogenans 2CP-C	ZP_00400853	5	<1	2	70-79
		Campylobacter coli NCTC11353		2	<1	<1	69
	Gamma	Unknown sp. X-bacteria symbiont of Amoeba proteus	AAC09381	3	1	<1	72-96
		Coxiella burnetii RSA 493	NP_820699	5	<1	2	79-82
Fusobacteria	Fusobacteria	Fusobacterium prausnitzii ATCC27768	AAT96608	7	4	1	71-83
Bacteroidetes	Bacteroides	Porphyromonas macacae ATCC33141	AAT96626	4		2	76-77
	Flavobacteria	Chryseobacterium indologenes ATCC29897	AAM88511	1		<1	97

Table 14 continued

Phylogenetic group		Nearest CPN-60 database neighbor	Genbank peptide accession number	Number of unique sequences	Clones in control library (%)	Clones in transgenic library (%)	% peptide identity
Acidobacteria		Solibacter usitatus Ellin6076	ZP_00520182	1		<1	88
Firmicutes		Bacillus psychrosaccharolyticus CECT 4074		2		1	55-56
		Aneurinibacillus aneurinolyticus ATCC12856		1		<1	64
		Bacillus thermoamylovorans		1		1	60
		Halothermothrix orenii H168	ZP_01188960	1		<1	61
Proteobacteria	Alpha	Rhodospirillum rubrum	ZP_00270903	1		<1	81
		Methylobacterium extorquens ATCC14718D	AAX24018	1		<1	72
Proteobacteria	Beta	Burkholderia thailandensis E264	YP_442004	1		<1	84
Proteobacteria	Delta/Epsilon	Wolinella succinogenes DSM 1740	NP_906559	1		<1	70
		Helicobacter cinaedi CCUG19503 S168		1		<1	57
Proteobacteria	Gamma	Microbulbifer degradans 2-40	ZP_00315051	4		2	85-86
		Francisella tularensis subsp. tularensis	AAT77113	1		<1	79
Planctomycetes		Blastopirellula marina DSM 3645	ZP_01094295	2		1	77-80
Total				305	215	215	55-100

Table 15. Relative abundance of clones with respect to different taxa from soil samples obtained from the rhizospheres of transgenic PPO over-expressing and unmodified hybrid aspen. Taxonomic assignment of clone sequences (215/library) was based on nearest neighbours obtained using a Blast-P search of amino acid sequences catalogued in the CPN-60 database.

Phylogenetic group	Relative CPN60 clone abundance %	
	Transgenic	Control
Fungi	**0**	**<1**
Deinococcus-Thermus	**0**	**0**
Acidobacteria	**<1**	**0**
Chloroflexi	**2**	**<1**
Planctomycetes	**1**	**0**
Fusobacteria	**1**	**4**
Chlamydiae	**1**	**3**
Firmicutes	**5**	**3**
Proteobacteria	**10**	**9**
Alphaproteobacteria	*1*	*2*
Betaproteobacteria	*<1*	*<1*
Deltaproteobacteria	*3*	*6*
Gammaproteobacteria	*5*	*1 **
Bacteriodetes/Chlorobi	**79**	**79**
Bacteroides	*3*	*4*
Sphingobacteria	*15*	*20*
Flavobacteria	*60*	*55*
Chlorobi	*<1*	*<1*

* indicates a statistically significant difference ($p \leq 0.05$) in clone abundance, using pooled datasets. P-values were based on a χ^2 test with Yates Correction, or Fisher's Exact test.

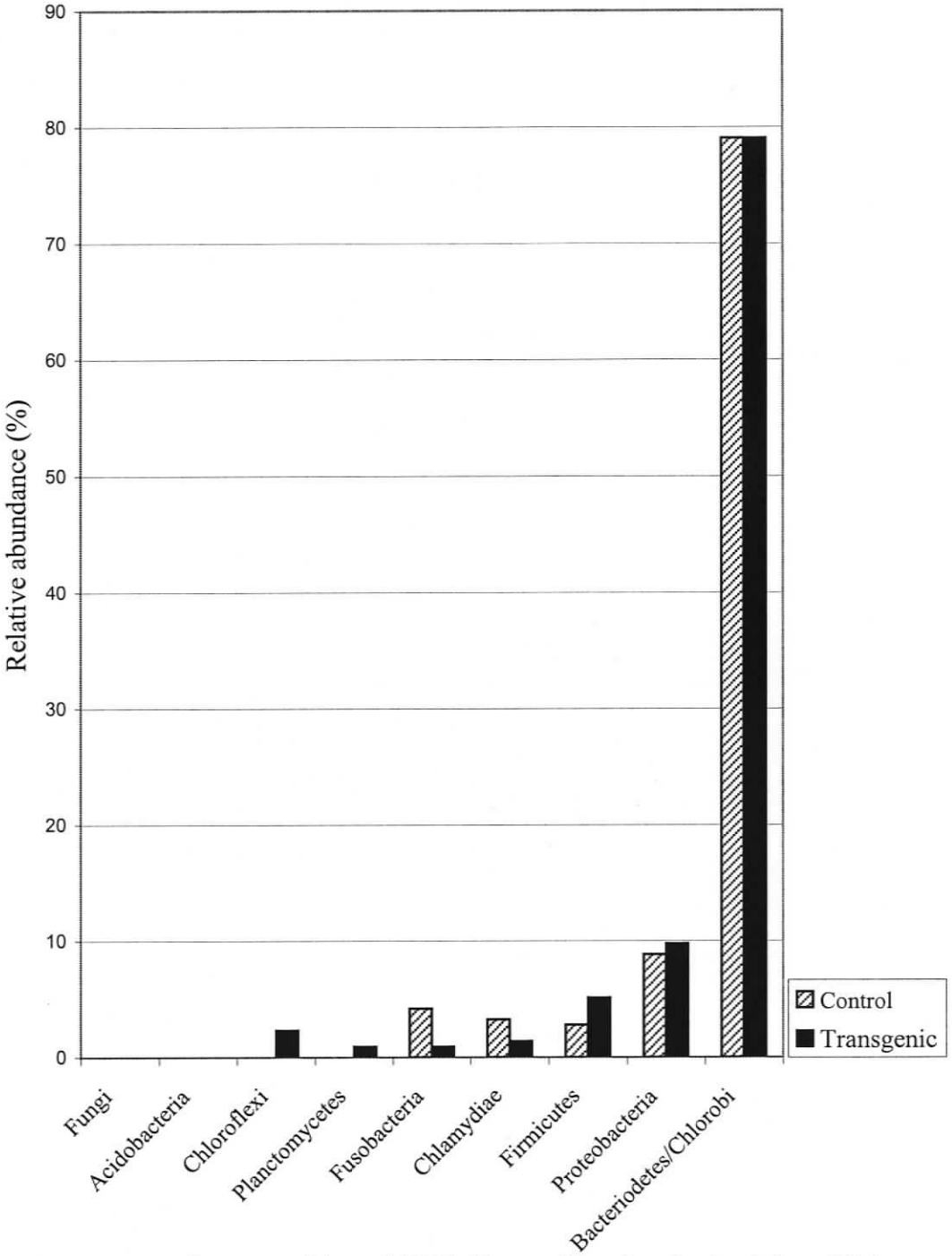


Figure 6. Taxonomic composition of CPN-60 gene libraries obtained from PPO overexpressing and non-transgenic hybrid aspen rhizospheres. Taxonomic assignment was based on the most homologous sequences obtained using a Blast-P search of Amino Acid sequences catalogued in the CPN-60 database. Replicate libraries were pooled, for a total of 215 clones/library. No significant differences in clone abundance were detected using $\chi 2$ analysis ($p<0.05$)

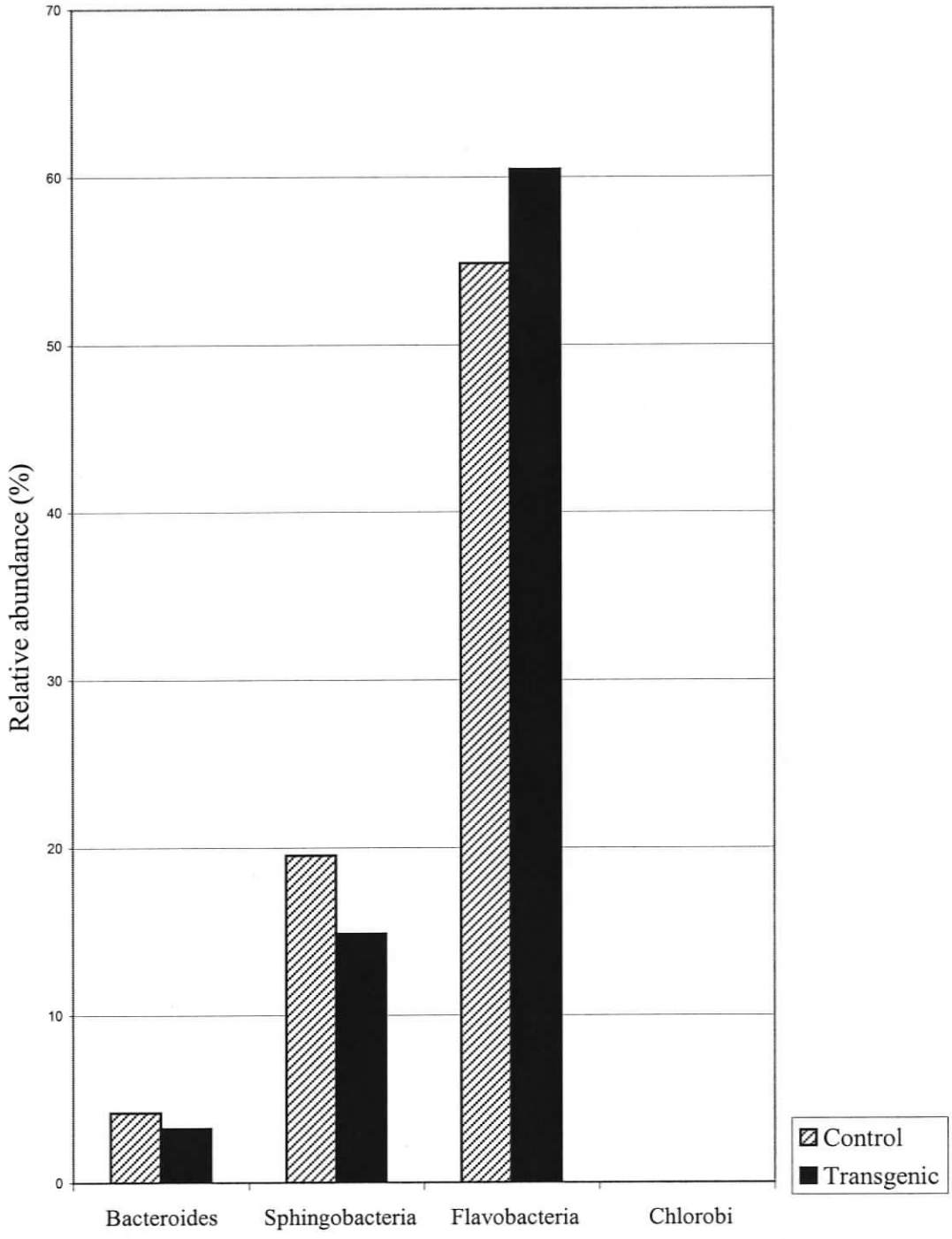


Figure 7. Taxonomic composition of the Bacteroidetes/Chlorobi group in CPN-60 libraries derived from the rhizospheres of PPO-overexpressing and non-transgenic hybrid aspen. Taxonomic assignment was based on nearest Blast-P amino acid sequence neighbors in the CPN-60 database. No significant differences in clone abundance were detected between pooled datasets ($\chi 2$ analysis, $p < 0.05$).

Library sequences at least 80% similar at the nucleotide level were grouped into OTUs, and a phylogenetic tree was constructed in order to illustrate the relationships of each OTU to their nearest neighbours in the CPN-60 database, and their frequency of occurrence in each library (Figure 8). In general, the initial estimates of clone taxonomy described above were supported by this analysis. In cases where % DNA identity was high, the relationships were clearly depicted by the branching pattern of the tree, and as expected, relationships were less clear when library sequences were distantly related to the CPN-60 reference sequences. The most frequently recovered sequences were most closely associated with the Bacteroidetes/Chlorobi group, and this subtree is shown separately (Figure 9).

3.4.2 Significance of difference between coverage curves for CPN-60 libraries

Significant differences between CPN-60 sequences in the pooled transgenic and control libraries were detected ($p < 0.025$), using the ∫-Libshuff algorithm, when each library was treated as the homologous library, suggesting that each harboured unique lineages. The greatest difference between the resulting coverage curves was observed at distances <0.2 (Figure 10). If a species unit is defined as sequences with genetic distances <0.2, then these results indicated that most differences being detected by this analysis were at the subspecies level. Significant differences were also detected between all of the replicate libraries ($p \leq 0.0043$), which were derived from individual trees. This result was expected, since each replicate library contained a different number of CPN-60 sequences (Table 3).

Figure 8. Neighbor-joining tree representing the relationships of 142 groups of CPN-60 nucleotide sequences from the rhizosphere of transgenic PPO over-expressing and unmodified hybrid aspen seedlings to the most closely related sequences obtained from FASTA searches of the CPN-60 database. Clones were grouped into OTUs sharing 80% sequence similarity. The numbers at the nodes of the tree indicate bootstrap values for each node with 1,000 bootstrap resamplings (Values below 50% are not shown). The scale bar represents 0.1 substitutions per site. The numbers in brackets indicate the recovery of each OTU from (control:transgenic) libraries.

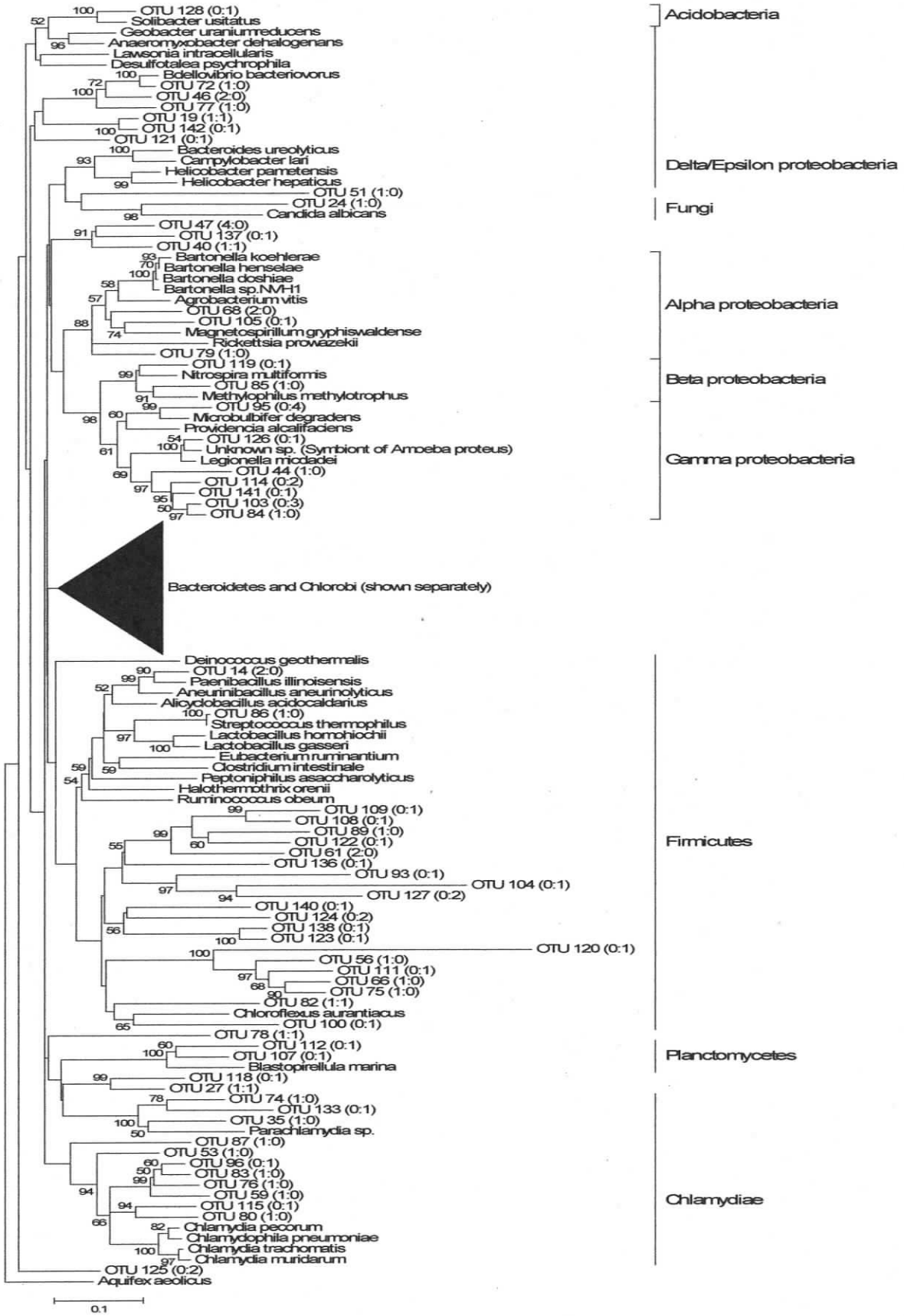

OTU 128 (0:1)
Solibacter usitatus
Geobacter uraniumreducens
Anaeromyxobacter dehalogenans
Lawsonia intracellularis
Desulfotalea psychrophila
Bdellovibrio bacteriovorus
OTU 72 (1:0)
OTU 46 (2:0)
OTU 77 (1:0)
OTU 19 (1:1)
OTU 142 (0:1)
OTU 121 (0:1)
Bacteroides ureolyticus
Campylobacter lari
Helicobacter pametensis
Helicobacter hepaticus
OTU 51 (1:0)
OTU 24 (1:0)
Candida albicans
OTU 47 (4:0)
OTU 137 (0:1)
OTU 40 (1:1)
Bartonella koehlerae
Bartonella henselae
Bartonella doshiae
Bartonella sp.NVH1
Agrobacterium vitis
OTU 68 (2:0)
OTU 105 (0:1)
Magnetospirillum gryphiswaldense
Rickettsia prowazekii
OTU 79 (1:0)
OTU 119 (0:1)
Nitrospira multiformis
OTU 85 (1:0)
Methylophilus methylotrophus
OTU 95 (0:4)
Microbulbifer degradens
Providencia alcalifaciens
OTU 126 (0:1)
Unknown sp. (Symbiont of Amoeba proteus)
Legionella micdadei
OTU 44 (1:0)
OTU 114 (0:2)
OTU 141 (0:1)
OTU 103 (0:3)
OTU 84 (1:0)
Bacteroidetes and Chlorobi (shown separately)
Deinococcus geothermalis
OTU 14 (2:0)
Paenibacillus illinoisensis
Aneurinibacillus aneurinolyticus
Alicyclobacillus acidocaldarius
OTU 86 (1:0)
Streptococcus thermophilus
Lactobacillus homohiochii
Lactobacillus gasseri
Eubacterium ruminantium
Clostridium intestinale
Peptoniphilus asaccharolyticus
Halothermothrix orenii
Ruminococcus obeum
OTU 109 (0:1)
OTU 108 (0:1)
OTU 89 (1:0)
OTU 122 (0:1)
OTU 61 (2:0)
OTU 136 (0:1)
OTU 93 (0:1)
OTU 104 (0:1)
OTU 127 (0:2)
OTU 140 (0:1)
OTU 124 (0:2)
OTU 138 (0:1)
OTU 123 (0:1)
OTU 120 (0:1)
OTU 56 (1:0)
OTU 111 (0:1)
OTU 66 (1:0)
OTU 75 (1:0)
OTU 82 (1:1)
Chloroflexus aurantiacus
OTU 100 (0:1)
OTU 78 (1:1)
OTU 112 (0:1)
OTU 107 (0:1)
Blastopirellula marina
OTU 118 (0:1)
OTU 27 (1:1)
OTU 74 (1:0)
OTU 133 (0:1)
OTU 35 (1:0)
Parachlamydia sp.
OTU 87 (1:0)
OTU 53 (1:0)
OTU 96 (0:1)
OTU 83 (1:0)
OTU 76 (1:0)
OTU 59 (1:0)
OTU 115 (0:1)
OTU 80 (1:0)
Chlamydia pecorum
Chlamydophila pneumoniae
Chlamydia trachomatis
Chlamydia muridarum
OTU 125 (0:2)
Aquifex aeolicus
0.1
Acidobacteria
Delta/Epsilon proteobacteria
Fungi
Alpha proteobacteria
Beta proteobacteria
Gamma proteobacteria
Firmicutes
Planctomycetes
Chlamydiae

Figure 9. Neighbor joining tree representing the relationships of CPN-60 nucleotide sequences most closely related to Bacteroidetes and Chlorobi reference sequences. Clones sharing 80% sequence similarity were grouped in to OTUs. The numbers at the nodes of the tree indicate bootstrap values for each node with 1,000 bootstrap resamplings (Values below 50% are not shown). The scale bar represents 0.1 substitutions per site. Numbers in brackets indicate the recovery of each OTU from (control : transgenic) libraries.


OTU 29 (1:0)
OTU 31 (7:2)
OTU 2 (10:16)
OTU 26 (3:1)
OTU 11 (1:2)
OTU 139 (0:3)
OTU 106 (0:1)
OTU 98 (0:4)
OTU 70 (1:0)
OTU 97 (0:2)
OTU 8 (28:24)
OTU 99 (0:1)
OTU 41 (1:2)
Flavobacterium ferrugineum
OTU 9 (3:9)
OTU4 (5:0)
OTU 129 (0:6)
OTU 64 (1:0)
OTU 32 (1:0)
OTU 37 (3:2)
OTU 52 (1:2)
OTU 25 (1:0)
OTU 1 (1:1)
OTU 22 (1:0)
OTU 23 (5:0)
OTU 73 (2:8)
OTU 12 (4:1)
OTU 50 (4:0)
OTU 62 (1:0)
OTU 71 (1:2)
OTU 36 (2:2)
OTU 5 (4:5)
OTU 54 (2:3)
OTU 48 (7:6)
OTU 135 (0:1)
OTU 43 (1:8)
OTU 101 (0:1)
OTU 30 (3:4)
OTU 102 (0:1)
OTU 116 (0:1)
OTU 60 (1:0)
OTU 34 (1:2)
OTU 57 (1:2)
OTU 49 (1:0)
OTU 67 (3:0)
OTU 63 (2:1)
Chryseobacterium indologenes
OTU 130 (0:1)
Chryseobacterium meningosepticum
OTU 38 (1:0)
Flavobacterium hyadatis
Flavobacterium johnsoniae
OTU 7 (3:2)
OTU 33 (1:0)
OTU 21 (2:2)
OTU 21 (1:0)
OTU 6 (1:0)
OTU 132 (0:1)
OTU 94 (0:3)
OTU 91 (1:0)
OTU 3(2:0)
OTU 69 (2:0)
Bacteroides uniformis
Prevotella heparinolytica
OTU 58 (1:0)
OTU 42 (1:1)
OTU 10 (2:0)
OTU 15 (3:1)
OTU 45 (2:0)
OTU 117 (0:1)
OTU 110 (0:2)
Cytophaga hutchinsonii
OTU 28 (1:1)
Fusobacterium prausnitzii
OTU 20 (4:1)
OTU 16 (1:0)
OTU 92 (1:0)
OTU 131 (0:1)
OTU 134 (0:1)
Sphingobacterium multivorum
OTU 55 (1:6)
OTU 18 (3:0)
OTU 39 (1:5)
OTU 113 (0:1)
OTU 13 (22:9)
OTU 17 (2:1)
OTU 90 (1:1)
OTU 81 (0:2)
OTU 65 (2:0)
Chlorobium chlorochromatii
Chlorobium tepidum
Pelodictyon phaeoclathratiforme
Chlorobium limicola
Pelodictyon luteolum
Bacteroidetes
Chlorobi
0.1

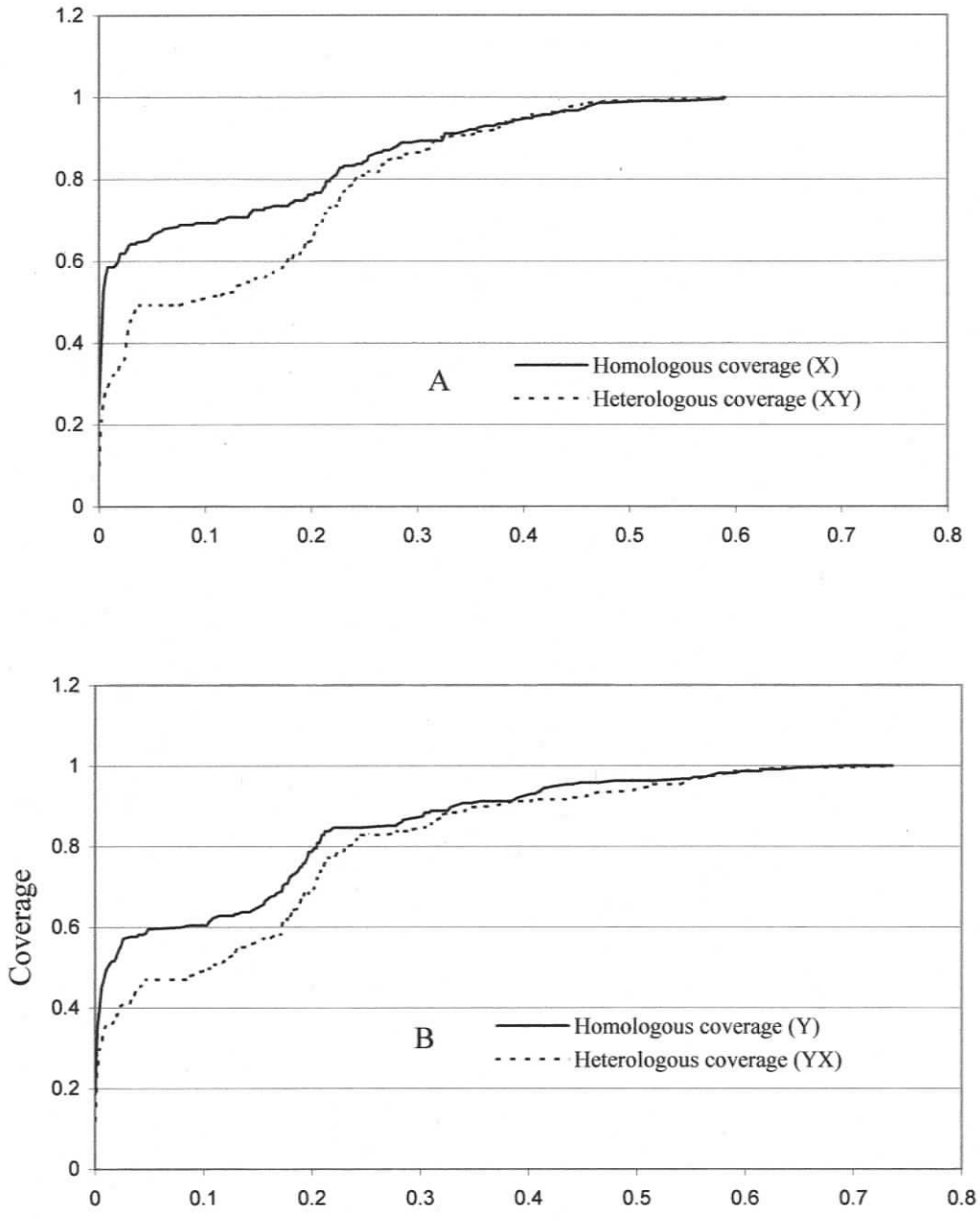


Figure 10. Comparison of CPN-60 gene libraries from transgenic PPO over-expressing and unmodified hybrid aspen rhizospheres based on coverage curves. (A) Transgenic (X) compared to control (Y) CPN-60 libraries (dCxy 327.2, p = 0.0000 *). (B) Control (Y) compared to transgenic (X) (dCyx=140, p-value=0.0013 *). Replicate libraries were pooled, for a total of 215 clones/library. The critical p-value for comparing 2 libraries was 0.025.

3.4.3 Diversity Indices for CPN-60 libraries

In order to describe and compare the relative diversity of the bacterial rhizosphere communities, CPN-60 sequences sharing 80% DNA sequence similarity were grouped into OTUs, and several diversity indices were calculated using pooled datasets.

In general, the species richness indices supported the results obtained from the 16S sequence libraries (Table 16). No significant differences in species richness were revealed between transgenic and control libraries, as each value for the diversity estimate from the transgenic library fell within the 95% confidence interval for the same estimate for the control library. The shannon index and ACE richness indices suggested a slightly higher richness associated with the transgenic library, supporting the results obtained from similar analyses of the 16S libraries. Interestingly, in contrast to analysis of the 16S libraries, the Chao richness index suggested that a slightly higher richness was associated with the control library. Similarly, 88 OTUs were recovered from control soils after sequencing 215 clones, compared to 84 recovered from transgenic soils. The collector's curves obtained for the control and transgenic libraries were highly similar however, suggesting that these differences in richness were insignificant (Figure 11). These discrepancies were likely due to biases inherent to a small sample size, and illustrate the value of estimating diversity using multiple measures.

The 3 indices used to describe levels of evenness within each library indicated that dominance in the control library was slightly higher than that observed in the transgenic library. Clones most closely related to *Flavobacterium ferrugineum* were assigned to OTU 8 and were recovered most frequently from both libraries. This OTU was recovered 28 times from the control library, and 24 times from the transgenic library. The bootstrap confidence estimate linking the sequences in OTU 8 with the nearest reference sequence was <50% however, suggesting a high level of uncertainty regarding their specific relationship (Figure 9).

Table 16. Diversity indices for bacterial CPN-60 OTUs associated with the rhizospheres of PPO-over-expressing and control hybrid aspen

Diversity parameter	Diversity Index	Rhizosphere soil type	
		Transgenic	Control
	Number of OTUs	84	88
Species richness (95% CI indicated in brackets)	Shannon (H')	3.93 (3.78-4.09)	3.89 (3.72-4.07)
	Chao Richness	162 (121-250)	183 (134-286)
	Ace Richness	182 (137-260)	180 (137-266)
Evenness or Dominance	Shannon Evenness (E)	0.887	0.869
	Simpson's (D)	0.029	0.037
	Berger Parker (d)	0.112	0.130

Replicate libraries were pooled, for a total of 215 clones/library. Operational taxonomic units (OTUs) were defined as partial CPN-60 sequences sharing 80% similarity. Where appropriate, 95% confidence intervals are indicated in brackets.

3.4.5 Collector's curves for CPN-60 libraries

The collector's curves for the CPN-60 libraries have not yet reached a plateau, indicating that unique sequences would continue to be recovered if more clones were sequenced. Although after sequencing 215 clones, 4 more OTUs were recovered from the control compared to the transgenic library, the curves cross after sampling 207 clones, suggesting that the difference in relative diversity was minor (Figure 11).

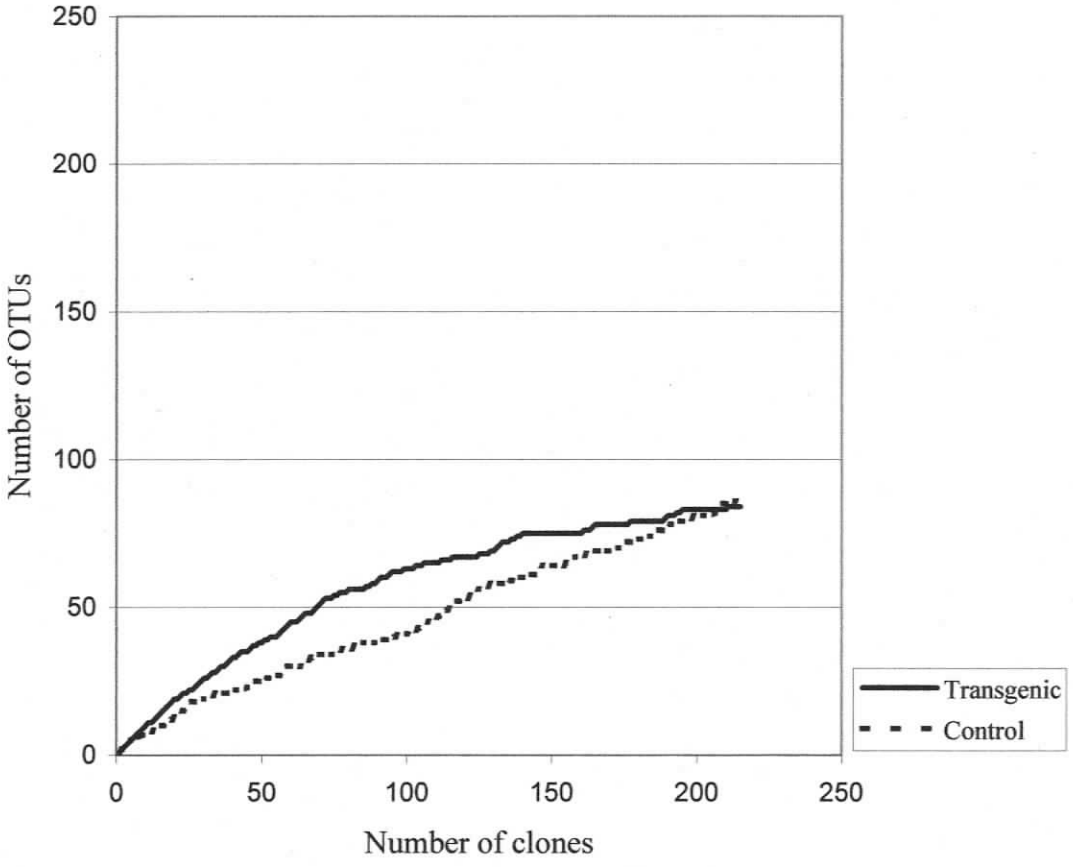


Figure 11. Collector's curves for CPN-60 gene libraries derived from the rhizospheres of transgenic PPO over-expressing and unmodified hybrid aspen. Clones were grouped into OTUs at a level of sequence similarity of >80%. Replicate libraries were pooled, for a total of 215 clones/library.

3.5 Enumeration of viable culturable fungi

Fungal plate counts were low in comparison to bacterial density, and significantly more CFUs were associated with one specific transgenic tree (Table 17). Since the difference in fungal density was not line-specific, it is likely that the observed variance reflected the underlying heterogeneity of the population and a small sample size.

Table 17. Enumeration of Viable Culturable Fungi (CFU/dry gram soil) after 48 hours of growth at room temperature and results of ANOVA (F=19.2, p=0.001), followed by a Student-Newman-Keuls test for differences between the mean CFU/dry gram of soil from each rhizosphere sample

Tree	Mean CFU/Dry Gram Soil (± SE)	Significantly different from
Transgenic Tree 19-9 (a)	1.41×10^3 (3.31×10^2)	b
Transgenic Tree 19-4 (b)	4.34×10^3 (4.45×10^2)	a,c,d
Control tree 717-10 (c)	2.12×10^3 (5.34×10^2)	b
Control tree 717-11 (d)	6.78×10^2 (4.44×10^1)	b

Differences in mean CFU/dry gram soil were considered significant if $p<0.05$. Means ± SE are the averages of 3 replicate 10^{-2} dilutions consisting of 3 plates each. Each series of dilution plates were generated from one soil sample derived from an individual tree.

3.6 Comparison of fungal rhizosphere communities based on 18S gene libraries

3.6.1 Taxonomic Composition of fungal 18S libraries

Fungal sequences were grouped into OTUs sharing 99% sequence similarity, and each OTU was assigned to a major fungal group by performing a phylogenetic analysis of the 58 different OTUs and their closest relatives in the genbank database. Blast-n search results for each OTU are summarized in Table 18, as well as the incidence of occurrence in pooled libraries derived from transgenic and non-transgenic rhizospheres. Phylogenetic analysis indicated that DNA from a diverse array of fungi was amplified from the rhizospheres. Basidiomycete, Ascomycete, Zygomycete, and Chytridiomycete sequences were recovered from both transgenic and control libraries (Figure 12, Figure

13). The majority of clones derived from soil obtained from both transgenic (68%) and control (61%) libraries grouped most closely with the Ascomycota (Table19).

Within the Ascomycota, clones were assigned to 11 groups, the most abundant being the Sordariomycetes, for both the transgenic and control libraries (Figure 14). This proportion was reflected in all replicate and pooled datasets (Table 19). Within the Ascomycetes, significantly more Saccharomycete sequences (13%) were recovered from the pooled control library compared to the pooled transgenic library (4%). In contrast, significantly more clones assigned to the Orbiliomycetes were recovered from the pooled transgenic library. For these two groups, no significant differences were observed in comparisons of replicate libraries from the same tree type (Table 19).

For the pooled datasets, significantly more of the clones from the control library belonged to Chytridiomycota (15%), compared to 7% of the clones from the transgenic library, but the difference between the proportions recovered from each of the control replicates (23% and 8%) was also significant for this group (Table 19). Within the Chytridiomycota, 2 groups were represented: the Spizellomycetes and the Chytridiomycetes. Significantly more Spizellomycetes were recovered from the pooled control libraries, but there were significant differences in recovery of these clones from the control replicates. Significantly more Chytridiomycete sequences were recovered from transgenic soils, and this was supported by comparisons between replicate libraries.

Some clone sequences were most homologous to fungal sequences of unknown identity that were derived from environmental samples (Table 19). These were recovered in greater frequency from the rhizospheres of transgenic seedlings, but variation between the two transgenic replicate libraries was also significant. Significant differences in abundance of other taxonomic groups were also detected between replicates, in cases where significant differences between pooled datasets were not observed (Table 19).

Table 18. Fungal sequences from the genbank database with the highest homology to each OTU (BLAST-n) and distribution of the 58 OTUs in 18S libraries obtained from rhizospheres of PPO over-expressing transgenic and control hybrid aspen

OTU	BLAST match	Accession Number	% Similarity	% Clones*	
				Control	Transgenic
1	*Rhizophlyctis rosea*	AY635829	98	14	
4	*Tomentella* sp. AFTOL-ID 1016	DQ092920	99	4	
51	Uncultured Sordariomycete	AJ635513	99	3	
45	*Piriformospora indica*	AY293147	97	2	
54	*Paecilomyces variotii*	AB023948	99	1	
46	*Nectria cinnabarina*	AB237663	97	<1	
47	*Tomentella* sp. AFTOL-ID 1016	DQ092920	97	<1	
48	Ophiostomatales sp.	AY315410	96	<1	
49	*Cancellidium* sp. HKUCC 10089	DQ144049	100	<1	
50	*Pleosporales* sp. IRB20-1	AB195631	98	<1	
52	*Holwaya mucida*	DQ257355	98	<1	
53	*Lachnum nudipes*	AY120859	99	<1	
55	*Nais inornata*	AF050482	99	<1	
56	*Graphium tectonae*	U43907	100	<1	
57	*Uncultured Chytridiomycete*	AJ635519	97	<1	
58	*Eladia saccula*	AB031391	99	<1	
29	*Togninia novae-zealandiae*	AY179958	98	1	<1
11	*Cadophora finlandica*	L76625	100	2	1
3	*Penicillium lagena*	AB032070	98	2	2
5	*Mortierella chlamydospora*	AF157143	99	5	2
8	*Piriformospora indica*	AY293147	98	8	4
2	*Penicillium namyslowskii*	AB028190	98	6	5
12	*Debaryomyces yamadae*	AB054266	98	13	4
7	*Chaetomium globosum*	DQ234257	99	23	23
39	*Hebeloma velutipes*	AY752972	99	2	2
36	*Oidiodendron tenuissimum*	AB015787	99	<1	<1
16	*Eladia saccula*	AB031391	100	5	8
9	*Hebeloma velutipes*	AY752972	98	1	7
14	*Monacrosporium sichuanense*	AY902788	99	1	6
18	Fungal clone UUF-54	AF095677	97	1	6
10	*Acremonium* sp. Dh3	AY392130	99	2	4
6	*Chytridiales* sp. JEL155	AF164315	99	1	3
23	*Spizellomyces acuminatus*	M59759	95	<1	2

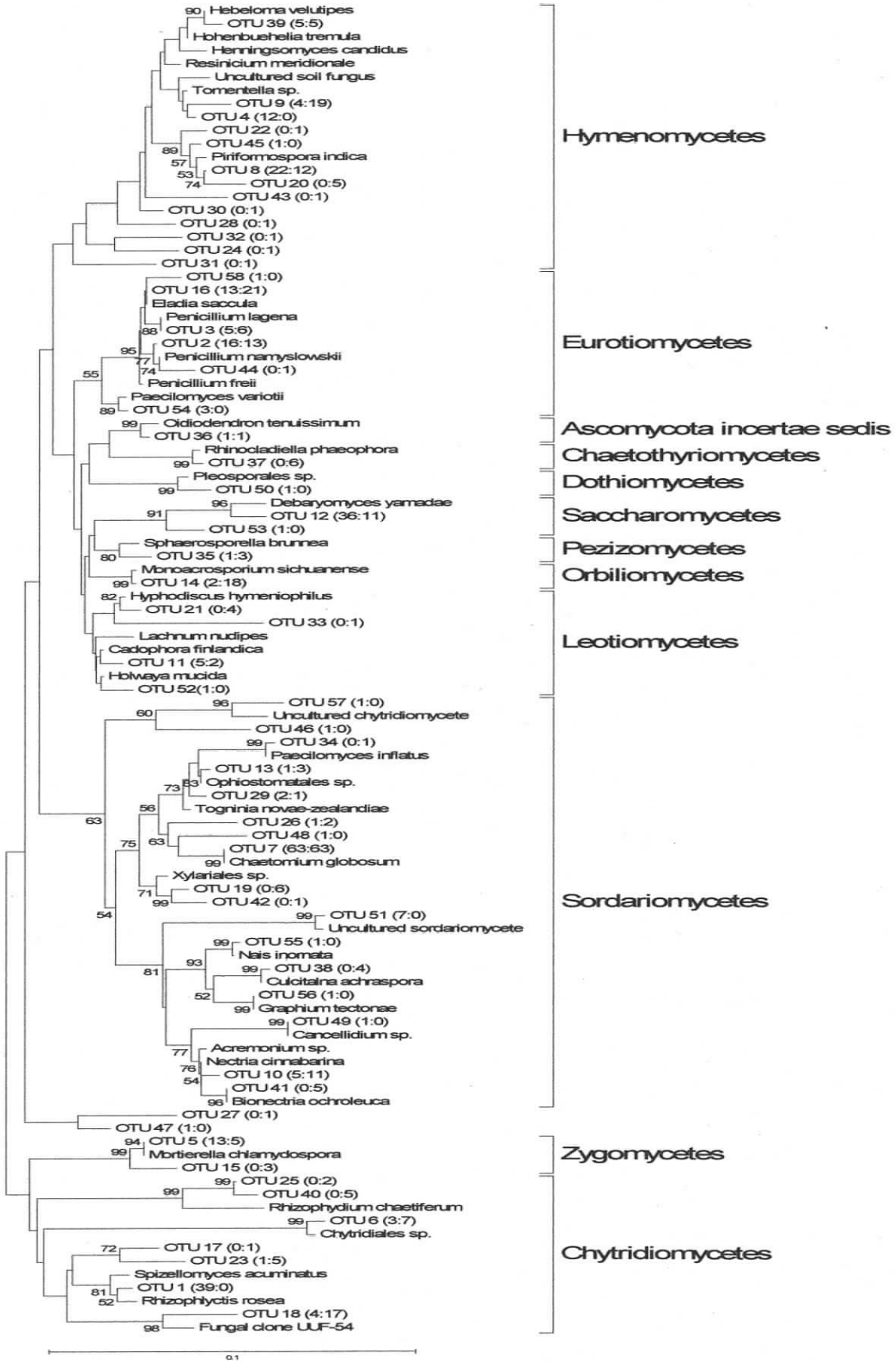
* Percentage of clones sharing the same OTU from the rhizosphere of transgenic or non-transgenic trees. Expected values for each BLAST match were 0.0.

Table 18 *continued*

OTU	BLAST match	Accession Number	% Similarity	% Clones*	
				Control	Transgenic
26	*Chaetomium globosum*	DQ234257	99	<1	1
13	*Ophiostomatales* sp.	AY315410	99	<1	1
35	*Sphaerosporella brunnea*	U53388	98	<1	1
17	*Monacrosporium sichuanense*	AY902788	94		3
37	*Rhinocladiella phaeophora*	AJ232950	99		2
38	*Culcitalna achraspora*	AY706332	99		2
32	*Resinicium meridionale*	AY293142	95		2
19	*Xylariales* sp. UB32-1	AB195633	98		2
31	*Penicillium freii*	AY640998	96		1
22	*Piriformospora indica*	AY293147	97		1
24	uncultured Sordariomycete	AJ635543	94		1
20	*Piriformospora indica*	AY293147	97		1
21	*Hyphodiscus hymeniophilus*	DQ227258	99		1
15	*Mortierella hyaline*	AY157493	99		<1
25	*Rhizophydium chaetiferum*	AF164263	93		<1
26	*Chaetomium globosum*	DQ234257	99		<1
27	uncultured soil fungus	AY382473	96		<1
28	*Henningsomyces candidus*	AF334916	97		<1
30	*Hohenbuehelia tremula*	DQ440645	100		<1
40	*Rhizophydium chaetiferum*	AF164263	99		<1
41	*Bionectria ochroleuca*	AJ783924	99		<1
42	*Xylariales* sp. UB32-1	AB195633	98		<1
43	*Hohenbuehelia tremula*	DQ440645	98		<1
44	*Penicillium namyslowskii*	AB028190	99		<1

* Percentage of clones sharing the same OTU from the rhizosphere of transgenic or non-transgenic trees. Expected values for each BLAST match were 0.0.

Figure 12. Neighbor-joining tree representing the phylogenetic relationships of 58 fungal OTUs from the rhizospheres of transgenic PPO over-expressing and control hybrid aspen to the most closely related sequences obtained from BLAST-n searches, showing the occurrence in (control : transgenic) libraries. The numbers at the nodes of the tree indicate bootstrap values for each node with 1,000 bootstrap resamplings (Values below 50% are not shown). The bar represents 0.1 substitutions per site.


Hebeloma velutipes
OTU 39 (5:5)
Hohenbuehelia tremula
Henningsomyces candidus
Resinicium meridionale
Uncultured soil fungus
Tomentella sp.
OTU 9 (4:19)
OTU 4 (12:0)
OTU 22 (0:1)
OTU 45 (1:0)
Piriformospora indica
OTU 8 (22:12)
OTU 20 (0:5)
OTU 43 (0:1)
OTU 30 (0:1)
OTU 28 (0:1)
OTU 32 (0:1)
OTU 24 (0:1)
OTU 31 (0:1)
Hymenomycetes
OTU 58 (1:0)
OTU 16 (13:21)
Eladia saccula
Penicillium lagena
OTU 3 (5:6)
OTU 2 (16:13)
Penicillium namyslowskii
OTU 44 (0:1)
Penicillium freii
Paecilomyces variotii
OTU 54 (3:0)
Eurotiomycetes
Cidiodendron tenuissimum
OTU 36 (1:1)
Ascomycota incertae sedis
Rhinocladiella phaeophora
OTU 37 (0:6)
Chaetothyriomycetes
Pleosporales sp.
OTU 50 (1:0)
Dothiomycetes
Debaryomyces yamadae
OTU 12 (36:11)
OTU 53 (1:0)
Saccharomycetes
Sphaerosporella brunnea
OTU 35 (1:3)
Pezizomycetes
Monoacrosporium sichuanense
OTU 14 (2:18)
Orbiliomycetes
Hyphodiscus hymeniophilus
OTU 21 (0:4)
OTU 33 (0:1)
Lachnum nudipes
Cadophora finlandica
OTU 11 (5:2)
Holwaya mucida
OTU 52(1:0)
Leotiomycetes
OTU 57 (1:0)
Uncultured chytridiomycete
OTU 46 (1:0)
OTU 34 (0:1)
Paecilomyces inflatus
OTU 13 (1:3)
Ophiostomatales sp.
OTU 29 (2:1)
Togninia novae-zealandiae
OTU 26 (1:2)
OTU 48 (1:0)
OTU 7 (63:63)
Chaetomium globosum
Xylariales sp.
OTU 19 (0:6)
OTU 42 (0:1)
OTU 51 (7:0)
Uncultured sordariomycete
OTU 55 (1:0)
Nais inornata
OTU 38 (0:4)
Culcitalna achraspora
OTU 56 (1:0)
Graphium tectonae
OTU 49 (1:0)
Cancellidium sp.
Acremonium sp.
Nectria cinnabarina
OTU 10 (5:11)
OTU 41 (0:5)
Bionectria ochroleuca
Sordariomycetes
OTU 27 (0:1)
OTU 47 (1:0)
OTU 5 (13:5)
Mortierella chlamydospora
OTU 15 (0:3)
Zygomycetes
OTU 25 (0:2)
OTU 40 (0:5)
Rhizophydium chaetiferum
OTU 6 (3:7)
Chytridiales sp.
OTU 17 (0:1)
OTU 23 (1:5)
Spizellomyces acuminatus
OTU 1 (39:0)
Rhizophlyctis rosea
OTU 18 (4:17)
Fungal clone UUF-54
Chytridiomycetes
0.1

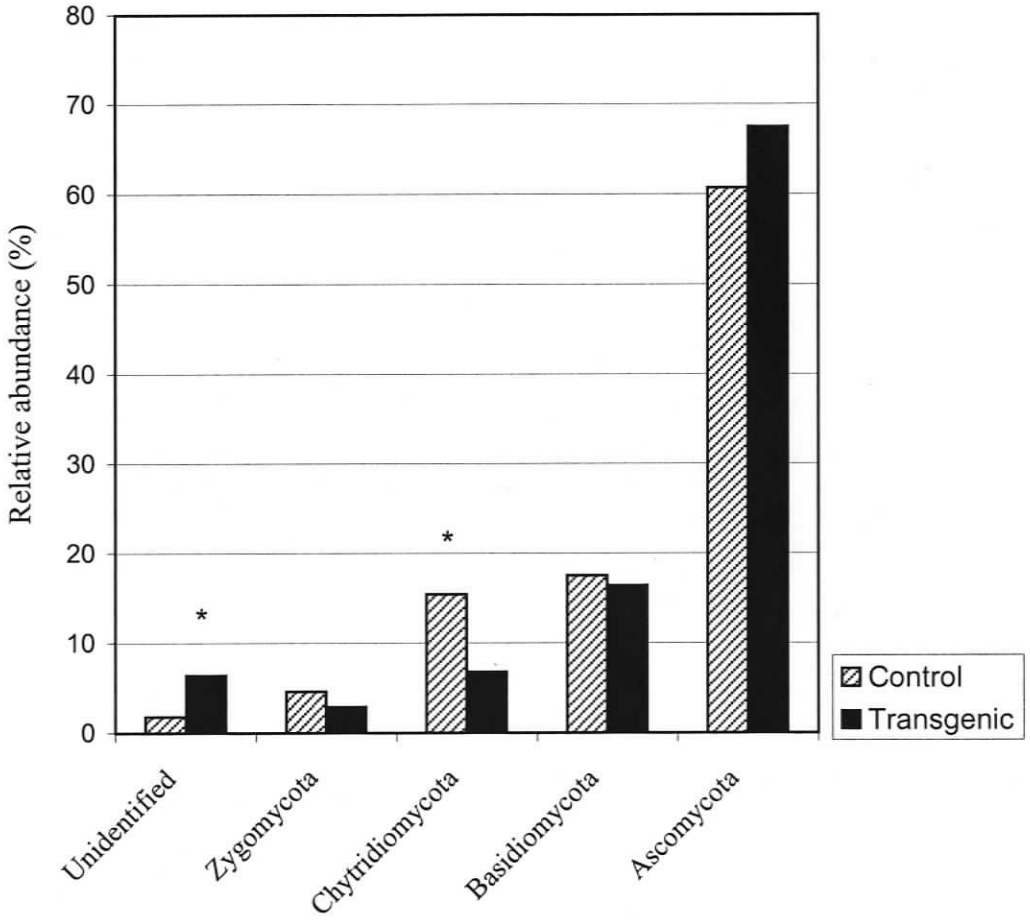


Figure 13. Relative abundance of 18S clones assigned to fungal taxa from transgenic PPO over-expressing and non-transgenic aspen rhizospheres. Replicates were pooled for a total of 280 clones/library. "*" indicates a significant difference in clone abundance between transgenic and control libraries ($\chi 2$ analysis, $p \leq 0.05$).

Table 19. Relative abundance of 18S clones with respect to different fungal taxa from duplicate soil samples obtained from the rhizospheres of PPO over-expressing and control hybrid aspen. Taxonomic assignment was based on the most homologous sequences in the Genbank database obtained with a BLAST-n search (280 18S clones/pooled library).

Phylogenetic group	Relative fungal clone abundance (%)					
	rRNA gene library from transgenic seedlings			rRNA gene library from control seedlings		
	Rep A	Rep B	Pooled	Rep A	Rep B	Pooled
Unidentified fungi	13 *	0	**6 ****	2	1	**2**
Zygomycota	4	1	**3**	6	4	**5**
Chytridiomycota	8	6	**7**	23 *	8	**15 ****
Spizellomycetes	*4*	*0*	***2***	*23 **	*6*	***14 *****
Chytridiomycetes	*4*	*6*	***5 *****	*0*	*2*	***1***
Basidiomycota	17	16	**16**	12	23	**18**
Hymenomycetes	*17*	*16*	***16***	*12*	*23 **	***18***
Ascomycota	58	77 *	**68**	57	64	**61**
Chaetothyriomycetes	*0*	*5 **	***3***	*0*	*0*	***0***
Incertae sedis	*0*	*1*	***<1***	*0*	*1*	***<1***
Dothiomycetes	*0*	*0*	***0***	*0*	*1*	***<1***
Pezizomycetes	*1*	*1*	***1***	*0*	*1*	***<1***
Orbiliomycetes	*10*	*4*	***7 *****	*1*	*0*	***<1***
Leotiomycetes	*2*	*3*	***3***	*4*	*1*	***3***
Mitosporic Ascomycota	*8*	*10*	***9***	*7*	*4*	***5***
Eurotiomycetes	*2*	*13 **	***8***	*4*	*13*	***9***
Saccharomycetes	*5*	*3*	***4***	*14*	*12*	***13 *****
Sordariomycetes	*30*	*38*	***34***	*28*	*32*	***30***

** indicates a statistically significant difference ($p \leq 0.05$) in clone abundance, using pooled datasets. P-values were based on a χ^2 Test with Yates Correction, or Fisher's Exact test. Statistically significant differences for comparisons of replicates within each library type are flagged with *

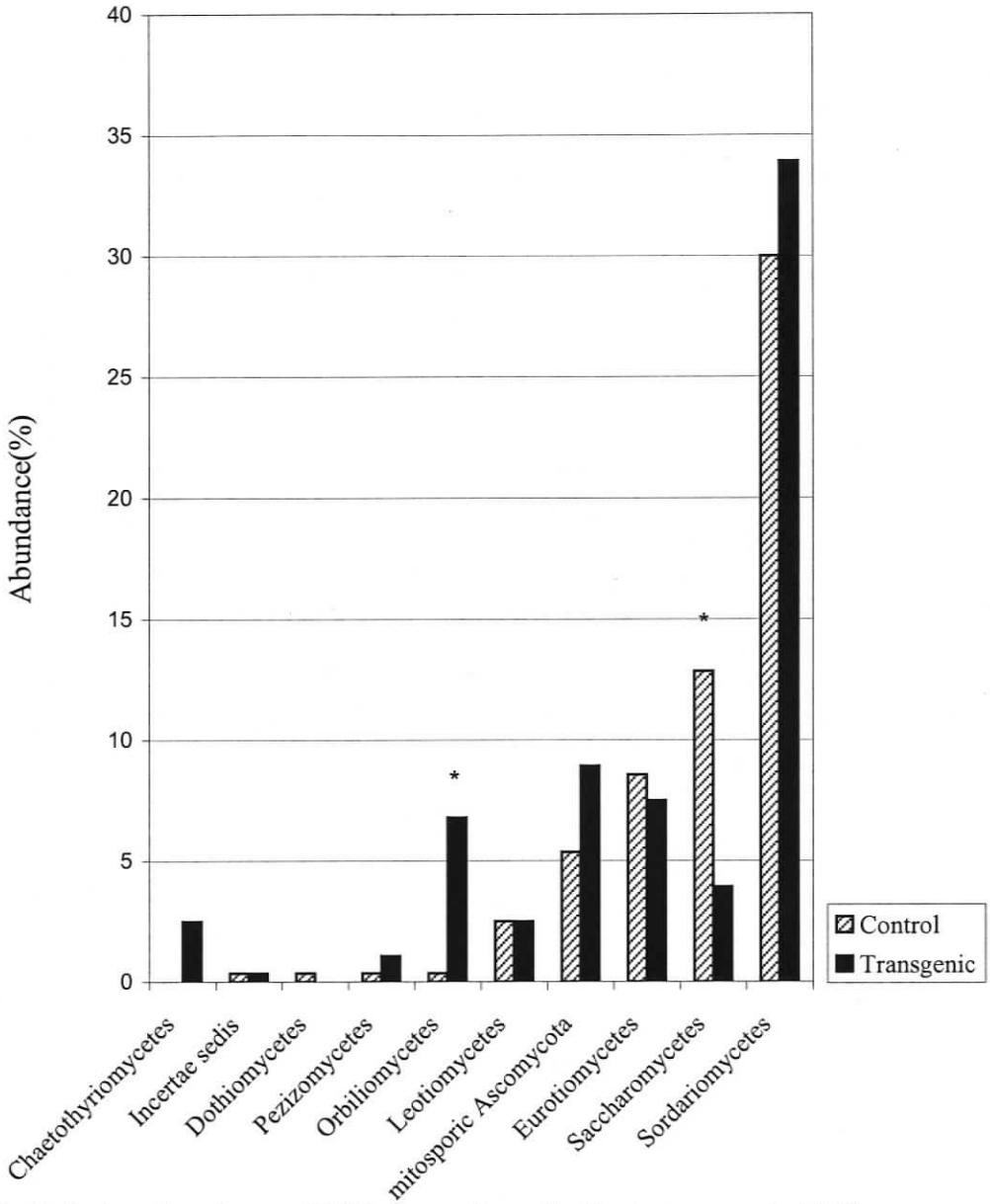


Figure 14. Relative abundance of 18S clones from duplicate transgenic PPO over-expressing and non-transgenic Aspen rhizosphere soil, assigned to groups within the Ascomycota. "*" indicates a significant difference in clone abundance between transgenic and control libraries ($\chi 2$ analysis, $p \leq 0.05$). Replicate libraries were pooled (280 clones/ pooled library).

3.6.2 Diversity Indices for fungal 18S libraries

In order to describe and compare the relative diversity of the fungal rhizosphere communities, sequences sharing 99% sequence similarity were grouped into OTUs, and several diversity indices were calculated.

To describe richness, defined as the number of different OTUs in the sample, 3 richness indices were calculated using pooled datasets. A slightly higher richness was recovered from the transgenic rhizosphere samples (44 OTUs), compared to the control samples (35 OTUs). The shannon index suggested that a higher richness was associated with the transgenic library, and the 95% confidence intervals for each library did not overlap. The Chao and ACE richness estimators also suggested that the transgenic library harboured a greater richness, but the confidence intervals for these estimates overlapped with those of the control library (Table 20).

The shannon evenness index and simpson's index of dominance suggested that a higher dominance of some fungal species were associated with the rhizospheres of control trees, compared to transgenic. Clones most closely related to *Rhizophlyctis rosea*, a Chytridiomycete, were assigned to OTU 1, accounted for 14% of the pooled control library, and were not recovered from the pooled transgenic library. Likewise, clones assigned to OTU 4, most closely related to the Basidiomycete, *Tomentella* sp. were not recovered from the transgenic library, but accounted for 4% of the control library. In contrast, OTU groups recovered from the pooled transgenic but not the pooled control library, represented 3% or less of the total library (Table 18).

The berger parker index, which expressed the proportional abundance of the most dominant species, was the same for both transgenic and control libraries (0.225). The most abundant OTU recovered from both libraries was most closely related to *Chaetomium globosum* (99% Bootstrap support), a Sordariomycete (Figure 12). This OTU accounted for 23% of the transgenic and control libraries (Table 18).

Table 20. Diversity indices for fungal OTUs associated with PPO over-expressing and non-transgenic hybrid aspen rhizospheres.

Diversity parameter	Diversity Index/ Richness Estimator	Rhizosphere soil type	
		Transgenic	Control
Species richness (95% CI indicated in brackets)	Number of OTUs	44	35
	Shannon (H')	3.06 (2.92-3.20)	2.65 (2.51-2.79)
	Chao Richness	101 (61.0-235)	75.0 (46.3-176)
	Ace Richness	67.2 (52.8-105)	56.0 (42.2-96.2)
Evenness or Dominance	Shannon Evenness (E)	0.808	0.745
	Simpson's (D)	0.078	0.107
	Berger Parker (d)	0.225	0.225

Fungal OTUs were defined as sequences having at least 99% sequence similarity. Replicate libraries were pooled, for a total of 280 sequences/library.

3.6.3 Significance of difference between coverage curves for fungal 18S libraries

Analysis using the ∫-Libshuff algorithm detected significant differences between the fungal sequences in all replicate and pooled 18S libraries (Figure 15). The 18S libraries were highly redundant however, and this likely increased the sensitivity to differences in community structure. Five hundred twenty sequences were grouped, based on 99% sequence similarity, into only 58 OTUs (Table 18). Therefore, although the level of sequence overlap between the libraries was high, and the coverage curves were highly similar, the results likely reflected differences in sequence abundance.

3.6.4 Collector's curves for fungal 18S libraries

In comparison to the collector's curves obtained for the bacterial libraries, the collector's curves for the fungal libraries are approaching a plateau, indicating that the sequence diversity of the actual fungal population is more thoroughly covered by the clone libraries. Since a high level of redundancy was present in the library, sequencing more clones would not return substantial information. The collector's curves for the pooled and transgenic libraries indicated that a greater diversity of OTUs were recovered from the transgenic library after sampling 280 sequences (Figure 16).

3.6.5 Similarity coefficients for fungal 18S libraries

Similarity coefficients ranged from 0.391 to 0.500 and were all higher than those obtained for the bacterial libraries. No substantial differences between libraries were suggested by the similarity coefficients (Table 21).

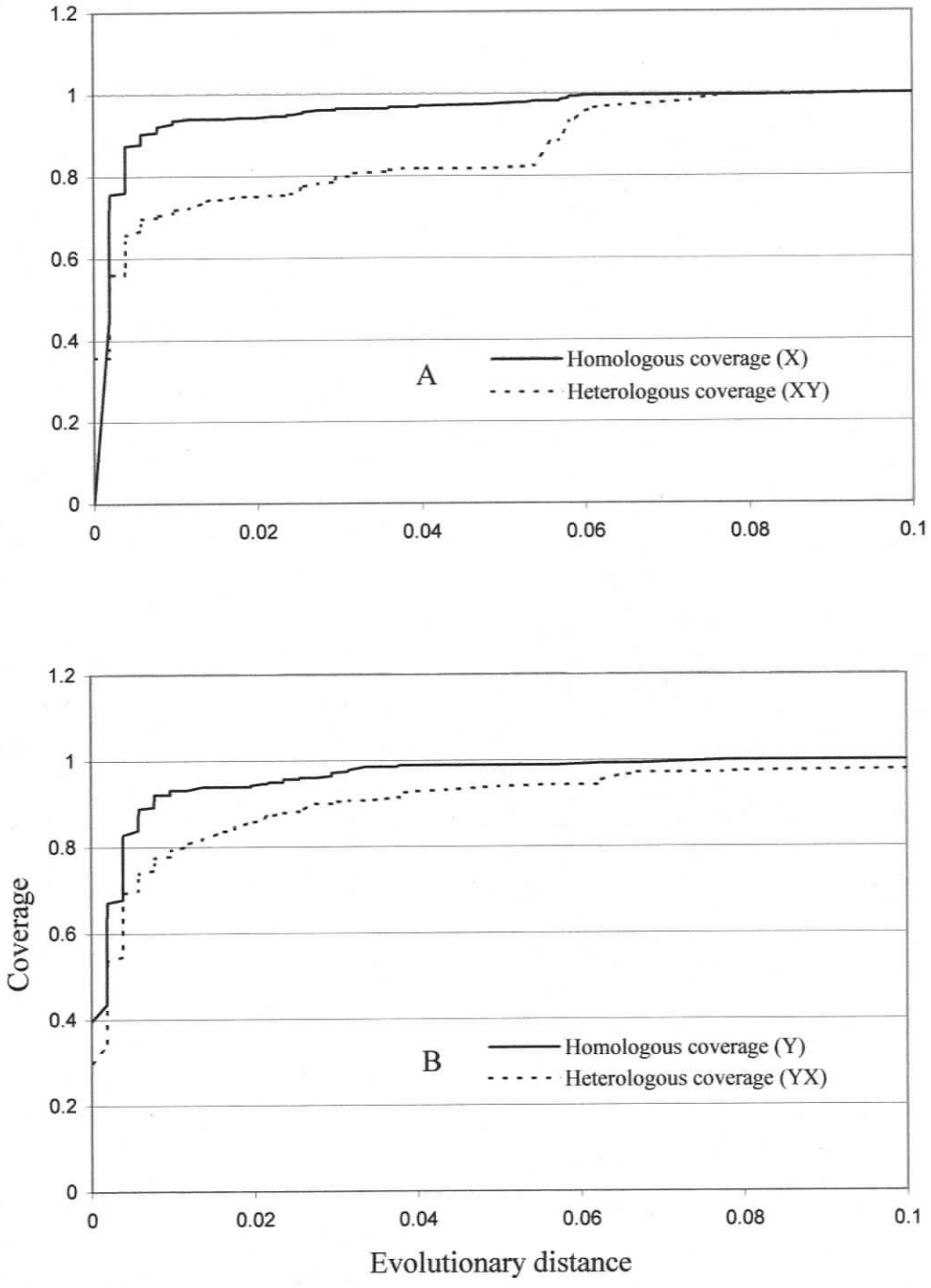


Figure 15. Comparison of fungal 18S gene libraries from transgenic PPO over-expressing and unmodified hybrid aspen rhizospheres based on coverage curves. (A) Transgenic (X) compared to control (Y) libraries (dCxy = 144.7, $p<0.0001$ *). (B) Control (Y) compared to transgenic (X) libraries (dCyx = 41.6, $p<0.0001$ *). Replicate libraries were pooled, for a total of 280 clones/library. The critical p-value for comparing 2 libraries was 0.025.

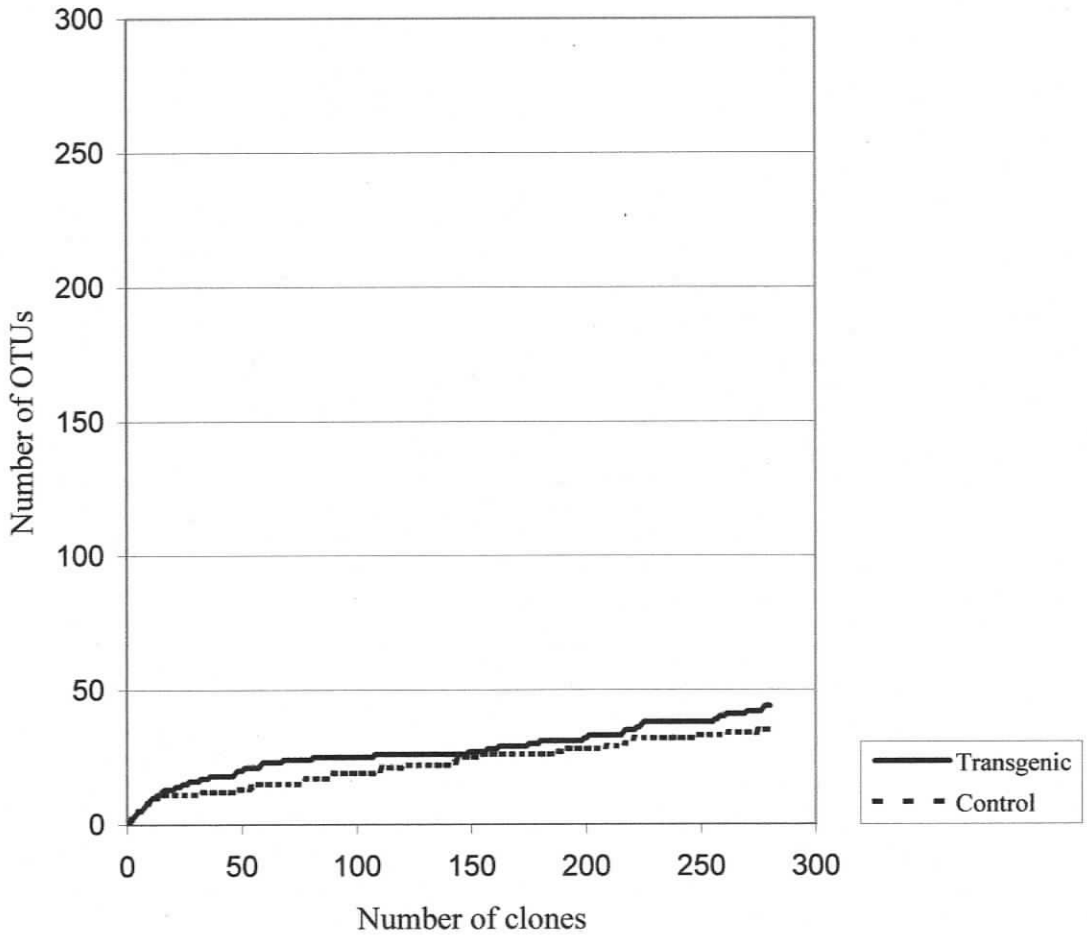


Figure 16. Collector's curves for fungal 18S gene libraries derived from the rhizospheres of transgenic PPO over-expressing and unmodified hybrid aspen. Replicate libraries were pooled, for a total of 280 sequences/library. Clones were grouped in to OTUs at a level of >99% sequence similarity.

Table 21. Pairwise comparisons of fungal 18S rRNA gene libraries obtained from duplicate rhizosphere soil samples associated with PPO over-expressing and non-transgenic hybrid aspen.

		Similarity coefficient		
Rhizosphere sample		Transgenic	Control	
		Tree 19-9	Tree 717-10	Tree 717 – 11
Transgenic	Tree 19-4	0.500	0.415	0.436
	Tree 19-9		0.468	0.571
Control	Tree 717-10			0.391

Clones having 99% sequence similarity were grouped into operational taxonomic units (OTUs). The similarity coefficients reflect proportions of shared OTUs. Similarity coefficients (S) were calculated by the method of McCaig *et al.*, 1999: S= 2C/(A+B) where C= # shared OTUs, A=# OTUs in Library A, B=#OTUs in Library B.

Discussion

The goal of this study was to determine the impact of transgenic polyphenol oxidase (PPO) over-expressing Aspen trees on rhizosphere community diversity and structure. By comparing the profiles of the fungal and bacterial communities associated with the roots of hybrid Aspen seedlings, changes in the rhizosphere can be detected. Previous studies have relied on assessing the diversity of the rhizosphere by analyzing the diversity of micro-organisms in culture. However, it is estimated that a large percentage of the rhizosphere community cannot be cultured by existing methods, and therefore is not detectable. The study detailed herein employed a highly discriminatory, culture independent approach that relied on the PCR amplification of partial gene sequences from members of the community directly from the sample, and thereby circumvented some of the problems of conventional means of measuring microbial diversity. Furthermore, the sequence datasets that were generated are additive and can be used to further develop tools to assess microbial diversity, as well as to monitor the effects of transgenic plants. Numerous attempts have been made to monitor the impact of transgenic crop species on soil microbial communities, but the present study is one of the few aimed at screening for the effects of transgenic trees on the soil microbial community. While caution needs to be taken when extrapolating results from a small number of samples, the dataset nevertheless provides insight into the natural variability in the composition of the bacterial and fungal communities associated with the rhizosphere of transgenic and non-transgenic hybrid aspen. The results presented in this study indicated that while there were some detectable differences in bacterial and fungal communities associated with PPO over-expressing and non-transgenic plants, these differences generally did not exceed the variation observed between individual trees of the same type.

4.1 Polyphenol oxidase activity in root tissue

The first step in the evaluation of the effects of PPO expression on soil microbes by transgenic Aspen was to establish that the trees were in fact overproducing PPO

(Bruinsma *et al.*, 2003; Cartwright and Lilley, 2004). Although the cauliflower mosaic virus (CaMV) promoter used in the construction of the transgenic plants generally drives constitutive transgene expression in most tissue types of many plant species (Benfey and Chua, 1990), confirmation of elevated PPO activity in the root tissue was of particular interest, since this is the portion of the plant most likely to influence the rhizosphere community. The data demonstrated that PPO activity in the transgenic plants was elevated 10 fold in the root tissue, relative to the non-transgenic plants. It had previously been shown that PPO activity in the leaves of these transgenic plants was elevated 50 fold, compared to unmodified plants (Wang and Constabel, 2004). Although the composition of the root exudates were not measured, elevated PPO activity levels in the root tissue confirmed the potential for a differential release of PPO or reactive o-quinones into the soil matrix. Polyphenol oxidase is localized within plastids (Mayer and Harel, 1979), and thus may be released from the roots into the rhizosphere following the discharge of damaged or dead root cells into the matrix.

The high variation in PPO activity observed in the root tissue of the trees used in this study was not unexpected. Significant differences in expression patterns of ß-glucoronidase (GUS) driven by a CaMV promoter have also been observed in tissue derived from primary roots, lateral roots, the root cap, and the elongation zone in transgenic *Allocasuarina verticillata* (Obertello, 2005). While levels of PPO activity were clearly elevated in transgenic plants relative to the untransformed controls, variation was also observed between individual plants. This observation was important because it demonstrated that although the hybrid aspen were clonal, there were subtle differences between individual plants, regardless of genetic modification.

4.2 Viable bacteria and fungi amenable to culture *in vitro*

Enumeration of the viable portions of the rhizosphere communities that were amenable to culture *in vitro* did not reveal clear differences between the bacterial or fungal cell densities harboured by transgenic and control trees. Variation between the heterotrophic plate counts from individual trees likely reflected the underlying

heterogeneity of the populations combined with the biases inherent to a small sample size. The lack of a clear change in total bacterial and fungal numbers may indicate that any difference in the components of the root exudates of transgenic and control plants were insufficient to significantly influence the total bacterial and fungal numbers. Changes in the populations of uncultured organisms however, would have been undetected by this method. Although in some cases enumeration of total bacterial and fungal populations has detected changes that have been correlated with transgenic plants, these methods are insensitive to alterations in specific components of the community. No differences were observed between the densities of the total bacterial population associated with the rhizospheres of transgenic T4-lysozyme producing potato plants (Heuer *et al.*, 2002; Lottman *et al.*, 1999) or opine-producing wheat (Oger *et al.*, 1997) and Black Nightshade (Mansouri *et al.*, 2002), when compared to their non-transgenic counterparts. In the latter examples however, the presence of opines allowed the enrichment of opine-utilizing bacteria, which remained undetected by aerobic heterotrophic plate counts. In contrast, leaves from cotton modified to express the *Bacillus thuringiensis* var. *kurstaki* (B.t.k.) endotoxin resulted in transient higher counts of bacteria and fungi cultured from soil for two of the three transgenic lines tested, but purified Bt toxin did not exert a detectable influence on soil microbial communities, suggesting that altered plant characteristics were responsible for the changes (Donegan *et al.*, 1995).

4.3 Comparison of the bacterial component of the rhizosphere communities

4.3.1 Bacterial community composition

The composition of rhizosphere communities has been shown to vary with factors including tree species (Priha *et al.*, 2001), plant health (Filion *et al.*, 2004), and environmental conditions (Graff and Conrad, 2005). In some cases, transgenic trees have also been shown to alter community composition (Oger *et al.* 1997, Tesfaye *et al.*, 2003), but the evaluation of minor shifts in composition is hindered by a lack of understanding about plant-microbe interactions in the rhizosphere. In the present study, the sequences amplified from extracted DNA represented a broad range of bacterial phylogenetic

groups, confirming that both the 16S and CPN-60 primers amplified DNA from a wide array of organisms. In all of the 16S libraries derived from both transgenic and unmodified trees, sequences assigned to the Proteobacteria were most abundant, and within this group, sequences affiliated with the Alphaproteobacteria accounted for 33-38% of each library. A similar proportion of Alphaproteobacterial clones (26%) was recovered from the rhizospheres of greenhouse grown hybrid aspen in an experiment aimed at determining impacts of flooding on community composition (Graff and Conrad, 2005). In the flooding experiment however, clones affiliated with the Bacillales were identified as being most prominent, accounting for 42% of the 50 clones sequenced from unflooded soil, but accounted for less than 2% of the libraries in the present study. Bacillales have also been found to dominate the rhizospheres of chrysanthemum (Duineveld *et al.*, 2001), and barley (Normander and Prosser, 2000), whereas Alphaproteobacterial clones were recovered in highest abundance from the rhizospheres of grass (McCaig *et al.*, 1999), and lodgepole pine (Chow *et al.*, 2002).

The phylogenetic representation in the CPN-60 libraries was significantly different than that observed in the 16S rRNA libraries, and this illustrated the bias potentially incurred by using a single gene locus to profile microbial diversity. In contrast to the 16S libraries, which were dominated by Proteobacterial sequences, the majority of clones in the CPN-60 libraries were most closely associated with the Bacteroidetes/Chlorobi group. Although some of the differences between the two libraries likely resulted from the preferential amplification of particular bacterial taxa by each primer set (Wintzingerode *et al.*, 1997), a higher level of uncertainty surrounded the taxonomic assignment of CPN-60 sequences. Each taxonomic assignment was an estimate, based on the similarity of the unknown sequences to reference sequences of known identity, and the number of known 16S rRNA sequences currently outnumbers the number of catalogued CPN-60 sequences.

Although the CPN-60 and 16S libraries depicted two very different views of the bacterial community and illustrated that each view of microbial diversity was incomplete, these methods allowed the comparison of the rhizosphere communities associated with the transgenic and unmodified trees. Although some major bacterial groups were differentially recovered from transgenic and control soils, these taxa represented minor components of the 16S and CPN-60 libraries. Based on 16S sequence analysis, a

significant decrease in Flavobacteria was observed in the transgenic library. Within the rhizosphere, Flavobacteria are likely involved in the breakdown of proteins and carbohydrates (Shewan and McMeekin, 1983), and also appear to actively decompose pesticides and insecticides (Singh and Walker, 2006). Flavobacteria have been recovered from diverse habitats, including freshwater and sediments (Shewan and McMeekin, 1983), and are generally present in low abundance in 16S libraries derived from soil (Janssen, 2006). Although this group represented a relatively minor component of the libraries, and the functional consequences of a change in number may be minimal, a cautionary approach might warrant further investigation. Similarly, the comparison of the major bacterial taxa present in CPN-60 libraries revealed that only a minor component of the community differed between transgenic and control plants. Significantly more sequences assigned to the Gammaproteobacteria were recovered from transgenic compared to control rhizosphere soils. The Gammaproteobacteria are the largest proteobacterial group, with 194 described genera that encompass a broad range of physiological diversity (Janssen, 2006), and the functional consequences of this change are therefore unclear. The overall lack of variation in the frequency of recovery of major bacterial groups derived from transgenic and control soils likely indicates that the impact of the transgenic plants on community composition was so low as to be undetectable by the approaches used.

4.3.2 Bacterial species diversity and community structure

One of the benefits of adopting a cloning and sequencing approach was that the dataset could be analyzed in numerous ways, allowing the measurement of several community parameters at various levels of resolution. Bacterial biodiversity in soil communities has been traditionally assessed in terms of the number of species in the community, and levels of evenness (Trevors, 1998), and in some cases, this has proven to be sensitive to environmental disturbance (e.g. Ibekwe *et al.*, 1998; Ramirez-Saad *et al.*, 2000). Thus, in addition to screening for changes in community composition, changes in bacterial species diversity were monitored.

Diversity indices, each differing in sensitivity to different abundance classes, and thus presenting a different view of the community, were calculated from sequence abundance data, in order to describe and compare various aspects of bacterial species diversity in the libraries. While there is sentiment in the microbial ecology literature that these estimates do not address the true total richness of soil microbial communities, they do allow comparison among treatments (Hughes *et al.*, 2001; Hill *et al.*, 2003). The Shannon Index is highly influenced by changes in rare as opposed to common species, and while its value may be inaccurate when coverage is low, it nonetheless allows the comparison of relative diversity. The corresponding shannon evenness index is also sensitive to changes in rare species, although low coverage results in overestimation of evenness. In contrast, the berger's parker index and simpson's index are sensitive to changes in dominant groups (Hill *et al.*, 2003). The application of multiple diversity measures thus allow the monitoring for changes in both rare and abundant OTU classes, which may be useful in monitoring for the unpredicted effects of transgenic plants on the rhizosphere community. The Chao richness estimator and the abundance-based coverage estimator (ACE) both estimate OTU richness without assuming a specific OTU abundance model, and are thus promising indices for studies of microbial communities. The high diversity present in microbial communities generally prevents exhaustive sampling, and it is unclear if any of the distribution models that are applied to macrobial communities accurately describe microbial communities (Hughes *et al.*, 2001). Chao (1984) noted that the Chao richness estimator is particularly useful for datasets skewed towards low abundance classes, which is likely to be the case for microbial communities.

Overall, the diversity indices did not reveal significant differences in terms of species richness or evenness between the libraries derived from transgenic and control plant rhizospheres. This suggests that any antibiotic effect resulting from the release of PPO or o-quinones from the roots of transgenic plants was undetectable using these methods. In chemically contaminated soils, a decrease in microbial diversity can result from a reduction of species richness due to pollutant toxicity (Trevors, 1998). A similar change would be expected if antimicrobial products with broad spectrum activity were introduced into the soil and were able to maintain activity. An increase in the dominance of a few species able to survive in the altered conditions would be expected to occur in

concert with a decrease in richness. An increase in certain populations can also occur when a suitable substrate becomes available for their enrichment, and this could also result in decreased diversity (Trevors, 1998). Thus, the present results also suggest that any change in the composition of root exudates was not sufficient to cause a detectable change in species richness. In some cases, minor changes in soil bacterial diversity have been detected in association with transgenic plants expressing antibacterial compounds (Heuer *et al.*, 2002; Rasche *et al.*, 2006), but there is no clear evidence linking the changes with the antibiotic properties of the transgenic products. It should be noted that the case specific nature of potential impacts of transgenic plants and the wide range of experimental methods used to detect changes hinders comparison among studies.

As with many other investigations aimed at comparing the richness of microbial species in soil and sediments (E.g. McCaig *et al.*1999, Filion *et al.*, 2004), numerous species were recovered with low frequency in the present study, and there were relatively few species common to both transgenic and control libraries. The lack of overlap between libraries was likely a result of low coverage of the underlying population. Low coverage also likely resulted in overestimation of species evenness by the shannon evenness indices (Hill *et al.*, 2003). The extremely high level of diversity within the rhizosphere communities was evidenced by the collector's curves, which are nearly linear. Higher coverage of the underlying populations could be attained by increasing the sample sizes.

The transgenic plants had no detectable effect on the recovery of the most dominant bacterial OTUs, as reflected by the calculated berger parker indices, but these groups accounted for minor proportions of the 16S ($\leq$ 5%) and CPN-60 ($\leq$ 13%) libraries. The OTU recovered with the highest frequency from the 16S rRNA libraries was most closely related to the genus *Burkholderia.* Members of the Betaproteobacteria, *Burkholderia* species are widespread in the environment, and are commonly recovered from plant rhizospheres (Richardson *et al.*, 2002; Falcão Salles, 2004). The OTU most frequently recovered from the CPN-60 dataset was distantly related to *Flavobacterium ferrugineum.* Although a lack of known CPN-60 reference sequences prevents strong conclusions regarding the specific identity of this group, the nearest neighbour in the CPN-60 sequence database possesses novel physiological capabilities. *Flavobacterium*

ferrugineum is a soil bacterium capable of growing on nitrogen free medium, without detectable nitrogenase activity, and may be able to use atmospheric nitrogen for growth, or may possess an unusual nitrogen fixing system (Xie and Yokota, 2006). The continual discovery of bacteria with unique physiologies from environmental samples illustrates that despite advances in the understanding of soil microbial ecology, microbiologists are only just beginning to grasp the true breadth of metabolic diversity present. This in itself may be a good reason to monitor and limit the impact of human activities on the microbial diversity in soil systems.

Although this is the first study to apply the Cramer-von Mises test statistic and Monte Carlo comparison test procedure to screen for differences in rhizosphere communities associated with transgenic and unmodified plants, similar methods have been used to compare a wide range of 16S libraries (E.g. Filion *et al.*, 2004; Wang *et al.*, 2005; Stach *et al.*, 2003). It has been demonstrated that these methods can successfully detect differences between soil communities harbouring different lineages (Schloss *et al.*, 2004), while correctly failing to find differences between libraries of the same composition (Singleton *et al.*, 2001). Significant differences between pooled 16S libraries derived from the transgenic and control plants were detected by ∫-libshuff analysis, but one control replicate library was identified as being the most dissimilar and likely responsible for the overall difference. The greatest difference in community structure was detected between the two control replicate libraries. This observation was significant in that it suggested that changes in community structure induced by the transgenic nature of the trees were small or even insignificant relative to the natural variability between the communities associated with individual plants of the same type. Analysis of CPN-60 libraries using ∫-libshuff detected significant differences between all replicate and pooled libraries, but most of the differences were at the subspecies level, between sequences sharing at least 80% similarity. After sampling 215 clones/pooled library, the protein-coding CPN-60 sequences may have been too divergent to allow meaningful comparison using the ∫-libshuff algorithm. Chaperonin-60 sequences have previously been used to differentiate between closely related bacterial species and strains indistinguishable by comparison of 16S sequences (Jian *et al.*, 2001; Goh *et al.*, 1996; Brousseau *et al.*, 2001; Kwok *et al.*, 2002) and this has been informative, especially in the field of medical

microbiology. In the context of impact assessment, the consequences of differences at the subspecies level are unclear. It is possible that changes in bacterial strains would have no functional consequences, (Schmalenberger and Tebbe, 2002), unless strains exhibiting phytopathogenic or antagonistic properties were selected for. A more meaningful evaluation of differences at the subspecies level would require that the changes be monitored in a narrower range of bacterial taxa.

4.4 Comparison of the fungal component of the rhizosphere communities

4.4.1 Composition of the fungal community

It is not yet possible to generate a clear picture of the dominant fungal taxa commonly associated with the rhizospheres of plants, because relatively few culture-independent studies have been initiated. It is also challenging to compare data from independent studies of diversity because a wide range of molecular markers and methodologies, each with inherent biases, have been used to investigate fungal communities. In the present study, sequences assigned to the four major fungal groups; the Zygomycota, Chytridiomycota, Basidiomycota and Ascomycota; were represented in the 18S libraries. Members of the Ascomycota were clearly dominant, and accounted for over 60% of the sequences recovered from both the transgenic PPO over-expressing and control rhizosphere soils. Ascomycetes include a wide range of root associated saprobic taxa (Jumponnen and Johnson, 2005), that can often assimilate cellulose (Kendrick, 1992), so the recovery of this group from rhizosphere soils was not unexpected. Analysis of the rhizosphere of wheat using multiple primer sets also revealed the presence of all four major fungal groups, but Smit *et al.* (1999) did not recover basidiomycete sequences using the 18S-specific (EF4/Fung5) primer set, and suggested that these primers preferentially amplified ascomycete sequences. Basidiomycete sequences were found to dominate grassland soils however, suggesting that the reported primer bias was less significant than originally thought, and the difference in representation more likely reflected a decreased fungal diversity in rhizosphere soil compared to soil that was not in contact with plant roots (Hunt *et al.*, 2004; Jumpponen and Johnson, 2005).

While all of the 18S rRNA sequences recovered in the present study were fungal in origin, there have been conflicting reports in the literature about the ability of the primer set EF4/fung5 to selectively amplify fungal DNA. It is challenging to design primers that capture a broad range of fungal diversity, but still specifically amplify fungal sequences in complex environmental samples. The 18S coding region is a commonly used target, but shares high homology with other eukaryotes, including plants and soil invertebrates, and co-amplification of DNA from these organisms often results. The successful recovery of fungal sequences in the present study were similar to the results of Smit *et al.* (1999), who used this primer set to profile fungal diversity in wheat rhizospheres, but contrasted those of Borneman and Hartin (2000), who amplified a significant number of non-fungal templates from cultured organisms and avocado grove soil using similar PCR conditions. While primer bias and specificity currently hinder attempts to exhaustively describe soil fungal diversity and complicate comparisons between studies, it was possible to use these methods to screen for differences between the rhizosphere communities associated with PPO over-expressing and control trees.

The data presented in this study showed that grouping sequences at the phylum level revealed only minor differences between the fungal taxa associated with the rhizospheres of transgenic and unmodified trees. Significantly fewer Chytridiomycete sequences were recovered from the pooled transgenic library, but one control replicate library was responsible for the observed difference. Within the chytrid fungi, members of the order Spizellomycetales are common soil inhabitants, while others are largely aquatic (Kendrick, 1992). Sequences most closely related to fungi of unknown identity also were recovered differentially from pooled transgenic and control libraries, but one transgenic replicate library was responsible for the observed difference. Most of the observed variation between replicate libraries likely reflected the discontinuous distribution of fungal populations, rather than the influence of genetic modification of the trees. Similarly, fungal taxa were shown to occur sporadically on the roots of non-transgenic and transgenic potato expressing the antimicrobial peptide, magainin, and were not associated with any particular plant line (O'Callaghan *et al.* 2004). The distribution of fungi in the soil matrix is less homogenous than bacteria, as fungal populations are strongly influenced by antagonistic chemical conflicts and competition for substrates

(Mikola and Setälä, 1998). Many fungi also spend a significant proportion of their lifecycle growing as mycelium, which is a vegetative thallus comprised of fine branching hyphal tubes (Kendrick, 1992). Studies of fungal diversity based on richness and abundance may also be complicated because both hyphae and spores may be present when DNA is extracted from environmental samples, and highly sporulating species may be overrepresented (Guidot *et al.*, 2003).

Within the Ascomycota, data from replicate libraries showed a difference in the recovery of groups from transgenic versus control soils, and supported the observation that transgenic modification may have resulted in minor qualitative changes in community composition. Qualitative differences in fungal community composition were also observed between transgenic potatoes with an altered starch composition and unmodified plants, but these differences were minor relative to differences observed between non-transgenic potato cultivars (Milling *et al.*, 2004). Significantly more Saccharomycete sequences, most closely related to *Debaryomyces yamadae* were recovered from the control soils in the present study. Members of the genus *Debaryomyces* have been isolated from a broad range of natural habitats, including air, pollen, and soil (Martorell *et al.*, 2005). Related genera, including *Saccharomyces* and *Cryptococcus*, are common rhizosphere inhabitants (Slávikovᡠand Vadkertiovᡠ2000), whose exudates positively influence colonization by mycorrhizal fungi (Sampedro *et al.*, 2004). Fast growing organisms, such as yeasts, have previously been found to be sensitive to environmental changes (van Bruggen and Semenov, 2000) and further investigation into changes in this population may be warranted. Significantly more nematode-trapping fungi, belonging to the Orbiliomycetes, most closely related to *Monacrosporium sichuanense* (re-classified as *Dactylellina sichuanensis*) (Li et al., 2006) were recovered from the transgenic soils, but there is no apparent reason that the transgenic plants would be expected to influence these fungi. Unexpected changes in their abundance may indirectly affect plant health however, since in conjunction with a range of other genera, this fungal group may influence the density of plant parasitic nematodes in the surrounding soil (Jansonn, 1982).

4.4.2 Fungal diversity and community structure

The 18S coding region is highly conserved, and fungal sequences were thus grouped into OTUs sharing 99% sequence similarity to monitor for changes in fungal species diversity. No important differences in terms of fungal species richness were detected between transgenic and control libraries, although the control libraries were characterized by a slightly higher level of dominance. The results presented here agree with a study using DGGE analysis that did not detect an influence of transgenic potatoes with an altered starch composition on fungal diversity (Milling *et al.*, 2004).

Most abundant OTUs were common to both the transgenic and control libraries, suggesting that the dominant populations within the fungal communities associated with each tree type were highly similar. Interestingly, the most dominant OTU was recovered in identical abundance from both the transgenic and control libraries. Sequences in this group were most closely related to *Chaetomium globosum*, a cellulolytic Sordariomycete that has been used as a biocontrol fungus to reduce a range of leaf and soil-borne diseases caused by fungal pathogens (Aggarwal *et al.*, 2004; Tomilova and Shternshis, 2006).

Although overall species richness was similar, and most abundant species were present both in transgenic and control libraries, some fungal species were recovered differentially. For example, clones most closely related to *Rhizophlyctis rosea* were only recovered from control soils, and accounted for a significant proportion of the library. *Rhizophlyctis rosea* is a chytrid fungus that is an important cellulose decomposer in soils (Willoughby, 1998). Similarly, OTU 4, most closely associated with the basidiomycete, *Tomentella* sp., accounted for 4% of the control library, but no sequences assigned to this group were derived from the transgenic soils. *Tomentella*-like fungi are ectomycorrhizal, and commonly colonize aspen roots (Visser *et al.*, 1998). The reasons for the differential recovery of these fungal species are unclear, but this dissimilarity may indicate that transgenic modification of the plants resulted in minor changes in fungal community composition. It is possible that changes in the abilities of mycorrhizal fungi to colonize the roots of the transgenic plants may significantly influence the health of the trees, and

the monitoring of changes in species composition over longer time periods may be valuable.

The Cramer-Von Mises test statistic and Monte-Carlo test procedure detected significant differences in community structure between all replicate and pooled libraries. Since differences in the relative abundance of fungal taxa were observed, this result was not unexpected. The 18S libraries were highly redundant, however, and fewer dissimilar sequences are required to distinguish libraries of low complexity using this method (Singleton *et al.*, 2001), and it is likely that ∫-libshuff was very sensitive to minor differences in the lineages harboured by each sample.

The collectors curves for both transgenic and control datasets suggested that the libraries thoroughly represented the underlying sequence diversity of the fungal population, and that sequencing more clones would not return substantial information. The 18S coding region is relatively conserved however, and does not allow distinction between closely related species. More species may have been present in the underlying fungal population than detected by the 18S libraries. The main strength of using an 18S rDNA approach was that it allowed the comparison of dominant populations belonging to a broad range of fungal taxa, but targeting more variable gene sequences, such as the internally transcribed spacer (ITS) regions (White *et al.*, 1990), would be suitable for targeting specific fungal groups (Hunt *et al.*, 2004) and would help to detect differences in the less dominant components of the community.

4.5 Evaluation of the CPN-60 gene target in terms of monitoring impact

This was the first study to screen for impacts of transgenic plants on rhizosphere diversity using CPN-60 methods. A secondary research objective was to evaluate the utility of these methods as tools to be used in impact assessment, as multiple approaches are likely to better detect changes in communities. The degenerate CPN-60 primers successfully amplified a broad range of bacterial taxa that differed from those represented in the 16S rRNA libraries, and thus improved the resolution derived from using a single amplification target. Diversity indices based on the partial CPN-60 sequences generally supported the data generated by the 16S rRNA sequence analysis. It was possible to

capture the 500 bp portion of the CPN-60 gene by sequencing in a single direction, which minimized the cost of sequencing and allowed the analysis of a larger sample size.

Although successful amplification of fungal DNA from reference strains was achieved using the CPN-60 primers, only one clone in the gene libraries was identified as a fungal sequence. The failure to detect fungal sequences in the libraries was not accounted for by a failure to isolate genomic DNA from fungi in the soil, since successful amplification from a broad range of taxa was achieved using the fungal 18S rRNA primers. Preferential annealing of the CPN-60 primers to bacterial DNA templates was likely a contributing factor to the under-representation of fungi in the libraries, as annealing bias is acknowledged as a problem associated with using the CPN-60 degenerate primers (Hill *et al.*, 2006). Recently, a specific mixture of PCR primers were found to improve the representation of bacteria with high G+C content in CPN-60 gene libraries (Hill et al., 2006), and it is possible that the same mixture would also allow amplification of fungal sequences. Less fungal DNA is generally recovered from soil samples than bacterial DNA (Harris, 1994) however, and optimal PCR conditions for the amplification of fungal and bacterial sequences from complex starting DNA template mixtures would likely differ. Therefore, for impact assessments, profiling of the bacterial and fungal components of the rhizosphere using the CPN-60 gene target may require the generation of two separate gene libraries.

Inclusion of CPN-60 gene sequences allowed sequence alignment and analysis at both the nucleotide and amino acid levels. Sequence variation within the protein-coding CPN-60 gene is generally higher than in 16S rRNA genes, which encode structural RNA (Brousseau *et al.*, 2001), and a greater diversity of CPN-60 sequences were recovered from the soils. At this level of sampling, this sequence diversity resulted in high levels of background noise when libraries were compared based on coverage. However, if more samples were processed, or if narrower phylogenetic groups were targeted, comparison of CPN-60 libraries using the ∫-libshuff algorithm may have more meaning. In comparison to 16S rRNA gene sequences, relatively few CPN-60 sequences are known, and thus a higher uncertainty was associated with assigning taxonomy to environmental clones. Many catalogued CPN-60 sequences were derived from studies of animal gut microflora and medically important microbes. The representation in the database is thus biased

towards these organisms. However, as the database expands in the future, it will be possible to estimate the identity of environmental CPN-60 sequences with increased precision. When the database is more complete, reassessment of the sequence libraries generated in this study will allow identification at a higher level of resolution than is presently possible. The number of studies aimed at profiling microbial communities using CPN-60 methods have increased substantially in the last 5 years (e.g. Hill *et al.*, 2002; Hill *et al.*, 2005a, Hill *et al.*, 2005b; Dumonceaux *et al.*, 2006), and the continual addition of reference sequences to the database will contribute to the value of these methods.

Since rRNA based methods dominate phylogenetic and microbial community studies, numerous tools have been developed to aid in sequence analysis, and CPN-60 methods would benefit from the development of similar tools. It is suspected that CPN-60 gene fragments amplified from environmental samples would result in fewer chimeric products than those commonly found in 16S libraries (Hill *et al.*, 2002), but tools have not been developed yet to allow rapid screening for these PCR artefacts in CPN-60 libraries. Chimeric sequences are essentially sequences derived from multiple DNA templates, whose inclusion in analysis may lead to overestimation of diversity and inaccurate taxonomic identification (Ashelford *et al.*, 2005). Similarly, tools available on the Ribosomal Database Project website that allow the rapid comparison of 16S libraries and taxonomic classification of clones have no parallel in the CPN-60 based methods. Practical and efficient approaches are required for screening for the impacts of transgenic plants on microbial communities, and 16S methods currently offer more options to researchers. However, considerable progress has been made in the establishment of the CPN-60 gene as a target for use in microbial community and phylogenetic studies in a relatively short time and further development of these methods shows considerable promise for application to monitoring for broad changes in soil microbial diversity.

4.6 General Discussion

Since root-associated organisms are potentially influenced by subtle differences in plant characteristics such as root morphology, metabolism, and root exudation (Cartwright and Lilley, 2004; Garbeva *et al.*, 2004), some variation in rhizosphere

communities associated with individual plants of the same line would be expected. Potential variation between individual plants was minimized by comparing a clonal species, and no obvious differences were in morphology. In general, the results suggested that elevated PPO activity had minimal impact on bacterial community structure and only minor differences in community composition were observed between transgenic and control libraries. The observed differences generally did not exceed the variation between individual trees of the same type. Variations between bacterial rhizosphere communities associated with individual plants of the same transgenic lines were also documented in corn varieties (Gyamfi *et al.*, 2002, Schmalenberger and Tebbe, 2003). Similarly, although some fungal OTUs were recovered solely from either transgenic or control soils, potentially warranting further investigation, differences between replicate libraries were also substantial, and likely reflected in part, the natural heterogeneity of the underlying population.

Impact assessment requires that changes in communities imposed by transgenic plants be evaluated in the context of natural fluctuations in community composition. Bacterial rhizosphere community composition is also influenced by factors including season, cultivar type (Dunfield and Germida, 2003), root zone, and different plant developmental stages (Duineveld *et al.*, 2001; Di Cello *et al.*, 1997). Although less well characterized, observed changes in fungal community structure have also been correlated with plant age (Gomes *et al.*, 2003) and root zone (Kurakov and Kostina, 2001). Information about natural fluctuations in rhizosphere communities have been identified as being valuable for determining the impacts of transgenic plants and in establishing a baseline for determining whether observed differences are significant (Kowalchuk *et al.*, 2003; Cartwright and Lilley, 2004).

Statistical significance was assigned to some of the differences observed between gene libraries. The link between statistically significant or numerical differences and biological significance remains unclear. In a soil system with such high diversity, a high level of redundancy in function would be expected within the bacterial community (Nannipieri et al., 2003). A change in a single phylogenetic group or bacterial species within the community may not necessarily result in a change in the overall functioning or

stability of the community, since other species or groups are likely to be able to fill the same ecological niche (Griffiths *et al.*, 2000).

As is the case with studies of this nature, some variation detected between libraries would be expected to be due to soil sampling, and efforts were thus made to standardize the sampling procedure so that any biases influenced each library to the same degree. Soil is an innately heterogeneous matrix, containing many microhabitats suitable for microbial growth (Trevors, 1998). As a result, bacteria may be highly aggregated in soil and extrapolation of results to the entire population or comparisons between samples would be influenced by the presence or absence of these dense clumps in each sample. In anticipation of the challenges associated with soil sampling, soil collected from each rhizosphere was thoroughly homogenized prior to DNA extraction. It is possible that a relatively small sample size and low coverage may also have contributed to some of the observed variation between libraries. However, the benefits of adopting a high resolution sequencing approach allowed for the screening of broad differences between transgenic and control libraries.

Although coverage was higher, variation between replicate fungal libraries generally exceeded that observed in the bacterial libraries, and this likely reflected a higher level of heterogeneity of the soil fungal populations. Although it presents methodological challenges for comparative studies, heterogeneity itself may be an important property of the system and worthy of note. While sequencing more clones from the existing 18S libraries would not return significant information, sampling a third replicate library for each tree type would provide additional insight into the significance of the variation observed between transgenic and control trees. Genomic DNA was extracted from smaller soil sample sizes so that minor populations were less likely to be masked by dominant ones in larger samples, allowing for a more complete inventory of microbial diversity. This sampling method may have also contributed in part to the observed variation between replicates, since Ranjard *et al.* (2003) found that when DNA was extracted from smaller soil samples (< 1 gram), variations in fungal community structure were detected between replicates, whereas soil sample sizes of 0.125-4 grams did not influence the assessment of bacterial diversity from a single homogenized soil sample.

The present study screened for general, unpredictable effects of PPO over-expressing plants on rhizosphere community structure. In conjunction with the monitoring of indicator groups or potentially vulnerable processes, Kowalchuk *et al.* (2003) identified this approach as being valuable in successfully assessing the impacts of transgenic plants. Broad changes in biodiversity are common measures of impact since unexpected changes may be detected (Cartwright and Lilley, 2004), and biodiversity is important in that it may contribute to the ability of the soil system to absorb changes and still persist (Botton *et al.*, 2006). While a clear link between diversity and ecosystem function has not been established, a minimum number of species is likely required for an ecosystem to function under stable conditions, and a greater species number is necessary for maintaining stability of processes under fluctuating conditions (Griffiths *et al.*, 2000; Loreau, 2001).

Indicator species or groups are organisms whose loss is likely to have a significant effect on soil ecosystem function (Kowalchuk *et al.*, 2003; Cartwright and Lilley, 2004; Lilley *et al.*, 2006). Potential indicators are also identified as being vulnerable to change, non-redundant in function, and accessible experimentally (Kowalchuk *et al.*, 2003). It has been proposed that a case-by-case approach be taken to determine the impacts of transgenic plants. The monitoring of diversity is a common first step that may also be useful in the identification of indicator species or narrower groups of organisms for subsequent monitoring (Kowalchuk *et al.*, 2003; Cartwright and Lilley, 2004). Despite some of the gaps in the understanding of plant-soil systems, several groups of microbes have been identified as meeting these criteria. Potential fungal indicators include the arbuscular mycorrhizal fungi, whose interaction with plants can be sensitive to change (Bending *et al.*, 2004), and wood degrading members of the Basidiomycota (Kowalchuk *et al.*, 2003). Symbiotic nitrogen-fixing and ammonium-oxidizing bacteria have also been identified as being potentially useful indicator groups (Chang *et al.*, 2001). In agricultural systems, soil suppressiveness is critical to sustainability, and key genera that are antagonists to plant pathogens have also been identified as indicator species. In cases where crops are impacted by a narrow range of pathogens, changes in specific pathogens themselves may also be indicative of functionally significant perturbations (Kowalchuk, 2003).

The rhizosphere is a dynamic environment, and root associated microorganisms continuously experience a range of changing parameters that influence microbial community structure. Different microbial groups are continuously selected for under natural conditions, largely in response to changes in root exudation induced by factors such as plant developmental stage(Di Cello *et al.*, 1997), nutritional status (Yang and Crowley, 2000), and season (Dunfield and Germida, 2003); and also as a consequence of various microbial interactions (Whipps, 2001). The results supported the observation that overall, shifts in microbial community structure in response to genetic modification of plants appear to be within the range of magnitude of these natural changes.

4.7 Conclusions

In general, the results of the present research supported the findings of other studies aimed at assessing the impacts of transgenic plants on the soil microbial community (reviewed in Cartwright and Lilley, 2004), in that relatively minor differences in the abundance of fungal and bacterial taxa were associated with transgenic trees compared to unmodified trees. Three observations suggest that the changes noted in this study were minor. First, the observed differences were in non-dominant groups of bacteria and fungi, as opposed to the most abundant taxa. Secondly, the differences did not exceed the variation observed between individual trees of the same type. Finally, most differences were observed in taxa that perform general, as opposed to specialized functions. Both PPO over-expressing and non-transgenic plant rhizospheres harboured high levels of bacterial and fungal diversity, and no clear differences were revealed in terms of richness or evenness. The fungal community was characterized by a higher level of heterogeneity than the bacterial communities, and it is suspected that this contributed to some of the observed differences between the fungal communities associated with the transgenic and unmodified trees.

4.8 Future direction

The most convincing assessments of the impacts of transgenic plants involve multiple experiments that investigate changes in community structure and function under both

laboratory and field conditions, using a range of methods. For this study, universal primers that amplified DNA from a wide range of fungal and bacterial phyla were used, allowing the detection of broad changes in community composition and diversity. While this approach has been identified as being valuable in detecting unexpected effects of transgenic plants on the microbial community (Kowalchuk et al., 2003), conclusions could be strengthened with data obtained from complementary approaches. For example, comparison of community fingerprints generated using DGGE would allow for the analysis of larger sample sizes. Targeting narrower phylogenetic groups would allow for more complete coverage of the underlying populations and thus more statistical power. Furthermore, the detection and monitoring of indicator species may provide more conclusive evidence as to the functional significance of any observed differences. The importance of the observed changes would also be clarified by comparing the effects due to genetic modification to changes induced by a range of other factors, such as soil type, vegetation stage, and pathogen exposure (Rasche *et al.*, 2006).

It was demonstrated that PPO activity was elevated in the roots of the PPO over-expressing trees, but future assessment would benefit from further characterization of the transgenic aspen, and the link of transgene expression to the diversity of the soil microbial community. In poplar leaves, PPO is present in a latent form, requiring partial proteolysis for full enzymatic activity (Constabel *et al.*, 2000), and it would therefore be interesting to determine whether active forms of PPO or reactive quinone products are present in the surrounding soil, and if they are, how long they persist in the environment. Since additional commitment would be required to thoroughly assess the impacts of these transgenic trees, their benefits need to be clearly established. Evidence so far suggests that the transgenic aspen have an increased resistance to insect herbivory (Wang and Constabel, 2004), and Polyphenol oxidases in general appear to play a role in disease resistance (Li and Steffens, 2002). However, the enzymatic properties and expression patterns of PPO isoforms vary within plant species (Wang and Constabel, 2003), and between plant species (Constabel and Ryan, 1998), suggesting that the enzyme has a broad range of physiological roles. Thus, the next step in risk assessment may be to test if the poplar leaf PPO that was over-expressed in the aspen influences plant pathogenic fungi and bacteria. Further characterization of the mechanisms linking polyphenol

oxidases to enhanced disease resistance may also help to identify components of the soil microbial community that may be impacted by over-expression. It is possible that release of transgenic proteins into the soil from fallen leaves may be an important route of contact in this case, since PPO activity in leaf tissue (Wang and Constabel, 2004) is significantly higher than in the root tissue.

The continued development of methods used to study soil microbial communities are needed to improve the assessment of the impacts of transgenic plants. The data generated in this study can be used to further evaluate the utility of the cloning/sequencing approach in monitoring strategies, and to advance the development of statistical methods used to compare gene libraries. Although this study was limited to a relatively small sample size, soil and tissue were collected from numerous trees, and stored samples could easily be used to expand the size of the gene libraries. In addition, tissue and soil samples were collected from hybrid Poplar over-expressing a novel cationic peptide with antimicrobial properties, which could be analyzed in the future. The sequence dataset will be available for future reference, and as we increase the present understanding about what is needed to assess the impact of transgenic trees and continue to expand upon knowledge regarding the link between microbial diversity and soil function, the data can be revisited to reveal additional information.

Literature cited

Aggarwal, R., A.K. Tewari, K.D. Srivastava, D.V. Singh. 2004. Role of antibiosis in the biological control of spot blotch (*Cochliobolus sativus*) of wheat by *Chaetomium globosum*. *Mycopathologia*. 157: 369-377.

Altschul, S.F., W. Gish, W. Miller, E.W. Myers, D.J. Lipman. 1990. Basic local alignment search tool. *Journal of Molecular Biology*. 215: 403-410.

Arenholtz, I., K. Harms, J. de Vries, W. Wackernagel. 2000. Increased killing of *Bacillus subtilis* on the hair roots of transgenic T4 lysozyme-producing potatoes. *Applied and Environmental Microbiology*. 66(5): 1862-1865.

Ashelford, K.E., N. Chuzhanova, J.C. Fry, A.J. Jones, and A.J. Weightman. 2006. New screening software shows that most recent large 16s rRNA gene clone libraries contain chimeras. *Applied and Environmental Microbiology*. 72(9). *In press*

Baumgarte, S., C.C. Tebbe. 2005. Field studies on the environmental fate of the Cry1Ab Bt-toxin produced by transgenic maize (MON810) and its effect on bacterial communities in the maize rhizosphere. *Molecular Ecology*. 14: 2539-2551.

Bending, G.D., M.K. Turner, F. Rayns, M. Marx, M. Wood. 2004. Microbial and biochemical soil quality indicators and their potential for differentiating areas under contrasting agricultural management regimes. *Soil Biology and Biochemistry*. 36: 1785-1792.

Benfey, PN, and N.H. Chua. 1990. The cauliflower mosaic virus 35S promoter: combinatorial regulation of transcription in plants. *Science*. 250: 959-966.

Berthelet, M., L.G. Whyte, C.W. Greer. 1996. Rapid, direct extraction of DNA from soils for PCR analysis using polyvinylpolypyrrolidone spin columns. *FEMS Microbiology Letters*. 138: 17-22.

Blackwood, C.B., J.S. Buyer. 2004. Soil microbial communities associated with Bt and non-Bt corn in three soils. *Journal of Environmental Quality*. 33(3): 832-836.

Borneman, J., R.J. Hartin. 2000. PCR primers that amplify fungal rRNA genes from environmental samples. *Applied and Environmental Microbiology*. 66(10): 4356-4360.

Botton, S., M. van Heusden, J.R. Parsons, H. Smidt, N. van Straalen. 2006. Resilience of microbial systems towards disturbance. *Critical Reviews in Microbiology*. 32: 101-112.

Bradford, M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principles of protein-dye binding. *Analytical Chemistry*. 72: 248-254.

Bradshaw, H.D., R. Ceulemans, J. Davis, R. Stettler. 2000. Emerging model systems in plant biology: Poplar (*Populus*) as a model forest tree. *Journal of Plant Growth Regulators*. 19: 306-313.

Brousseau, R., J.E. Hill, G. Prefontaine, S. Goh, J. Harel, S.M. Hemmingsen. 2001. *Streptococcus suis* serotypes characterized by analysis of chaperonin 60 gene sequences. *Applied and Environmental Microbiology*. 67(10): 4828-4833.

Bruinsma, M., G.A. Kowalchuk, J.A.van Veen. 2003. Effects of genetically modified plants on microbial communities and processes in soil. *Biology and Fertility of Soils*. 37: 329-337.

Butler, J.L., M.A. Williams, P.J. Bottomley, D.D. Myrold. 2003. Microbial community dynamics associated with rhizosphere carbon flow. *Applied and Environmental Microbiology*. 69(11): 6793-6800.

Cartwright, C.D., A.K.Lilley. 2004. Mechanisms for investigating changes in soil ecology due to GMO releases (Defra report EPG 1/5/214). *Department for Environment, Food, and Rural Affairs*

Chang, Y., A.K. Hussain, J.R. Stephen, M.D. Mullen, D.C. White, A. Peacock. 2001. Impact of herbicides on the abundance and structure of indigenous B-subgroup ammonia oxidizer communities in soil microcosms. *Environmental Toxicology and Chemistry*. 20(11): 2462-2468.

Chao, A. 1984. Non-parametric estimation of the number of classes in a population. *Scandinavian Journal of Statistics*. 11: 265-270.

Chao, A., S.M. Lee. 1992. Estimating the number of classes via sample coverage. *Journal of the American Statistics Association*. 87: 210-217.

Chow, M.L., C.C. Radomski, J.M. McDermott, J. Davies, P.E. Axelrood. 2002. Molecular characterization of bacterial diversity in Lodgepole Pine (*Pinus contorta*) rhizosphere soils from British Columbia forest soils differing in disturbance and geographic source. *FEMS Microbiology and Ecology*. 42: 347-357.

Cole J.R., B. Chai, R.J. Farris, Q. Wang, S.A. Kulam, D.M. McGarrell, G.M. Garrity, J.M. Tiedje. 2005. The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis. *Nucleic Acids Research*. 33 (Database Issue): D294-D296.

Constabel, C.P., C.A. Ryan. 1998. A survey of wound- and methyl jasmonate-induced leaf polyphenyl oxidase in crop plants. *Phytochemistry*. 47: 507-511.

Constabel, C.P., L. Yip, J.J. Patton, M.E. Christopher. 2000. Polyphenyl oxidase from hybrid Poplar. Cloning and expression in response to wounding and herbivory. *Plant Physiology*. 124: 285-295.

De Boer, W., L.B. Folman, R.C. Summerbell, L. Boddy. 2005. Living in a fungal world: impact of fungi on soil bacterial niche development. *FEMS Microbiology Reviews*. 29:795-811.

Di Cello, F., A. Bevivino, L. Chiarini, R. Fani, D. Paffetti, S. Tavacchioni, C. Dalmastri. 1997. Biodiversity of a *Burkholderia cepacia* population isolated from the maize rhizosphere at different plant growth stages. *Applied and Environmental Microbiology*. 63: 4485-4493.

Donegan, K., C.J. Palm, V.J. Fieland, L.A. Porteous, L.M. Ganio, D.L. Schaller, L.Q. Bucao, R.J. Seidler. 1995. Changes in levels, species and DNA fingerprints of soil microorganisms associated with cotton expressing the *Bacillus thuringiensis* var. *kurstaki* endotoxin. *Applied Soil Ecology*. 2: 111-124.

Duineveld, B.M., G.A. Kowalchuk, A. Keijzer, J.D. VanElsas, and J.A. Van Veen. 2001. Analysis of bacterial communities in the rhizosphere of chrysanthemum via denaturing gradient gel electrophoresis of PCR-amplified 16s rRNA as well as DNA fragments coding for 16s rRNA. *Applied and Environmental Microbiology*. 67: 172-178.

Dumonceaux, T.J., J.E. Hill, C.P. Pelletier, M.G. Paice, A.G. Van Kessel, S.M. Hemmingsen. 2006. Molecular characterization of microbial communities in Canadian pulp and paper activated sludge and quantification of a novel *Thiothrix eikelboomii*-like bulking filament. *Canadian Journal of Microbiology*. 52(5): 494-500.

Dunfield, K.E., J.J. Germida. 2001. Diversity of bacterial communities in the rhizosphere and root interior of field-grown genetically modified *Brassica napus*. *FEMS Microbiology Ecology*. 38: 1-9.

Dunfield, K.E., J.J. Germida. 2003. Seasonal changes in the rhizosphere microbial communities associated with field-grown genetically modified canola (*Brassica napus*). *Applied and Environmental Microbiology*. 69(12): 7310-7318.

Dunfield, K.E., J.J. Germida, 2004. Impact of genetically modified crops on soil and plant-associated microbial communities. *Journal of Environmental Quality*. 33: 806-815.

Ewing, B., L. Hillier, M.C. Wendl., P. Green. 1998. Base-calling of automated sequencer traces using Phred. I Accuracy assessment. *Genome Research*. 8: 175-185.

Falcão Salles, J., J.A. van Veen, J.D. van Elsas. 2004. Multivariate analyses of *Burkholderia* species in soil: effect of crop and land use history. *Applied and Environmental Microbiology*. 70(7): 4012-4020.

F.A.O., 2004. Preliminary review of biotechnology in forestry, including genetic modification. Forest Genetic Resources Working Paper FGR/59E. Forest Resources Development Service, Forest Resources Division. Rome, Italy.

Felsenstein, J. 1989. PHYLIP - Phylogeny Inference Package (Version 3.2). *Cladistics*. 5: 164-166.

Fenning, T.M., J. Gershenzon. 2002. Where will the wood come from? Plantation forests and the role of biotechnology. *Trends in Biotechnology*. 20(7): 291-296.

Filion, M., R.C. Hamelin, L. Bernier, M. St-Arnaud. 2004. Molecular profiling of rhizosphere microbial communities associated with healthy and diseased Black Spruce (*Picea mariana*) seedlings grown in a nursery. *Applied and Environmental Microbiology*. 70(6): 3541-3551.

Garbeva, P., J.A. van Veen, J.D. van Elsas. 2004. Microbial diversity in soil: selection of microbial populations by plant and soil type and implications for disease suppressiveness. *Annual Review of Plant Phytopathology*. 42: 243-270.

Goh, S.H., S. Potter, J.O. Wood, S.M. Hemmingsen, R.P. Reynolds, A.W. Chow. 1996. HSP60 gene sequences as universal targets for microbial species identification: studies with coagulase-negative staphylococci. *Journal of Clinical Microbiol*ogy. 34: 818-823.

Good, I.J. 1953. The population frequency of species and the estimation of population parameters. *Biometrika*. 40: 237-264.

Gomes, N.C., O. Fagbola, R. Costa, N.G. Rumjanek, A. Buchner, L. Mendona-Hagler, K. Smalla. 2003. Dynamics of fungal communities in bulk and maize rhizosphere soil in the tropics. *Applied and Environmental Microbiology*. 69(7): 3758-3766.

Graff, A., R. Conrad. 2005. Impact of flooding on soil bacterial communities associated with poplar (*Populus* sp.) trees. *FEMS Microbiology and Ecology*. 53: 401-415.

Grayston, S.J., D. Vaughan, D. Jones. 1996. Rhizosphere carbon flow in trees, in comparison with annual plants: the importance of root exudation and its impact on microbial activity and nutrient availability. *Applied Soil Ecology*. 5: 29-56.

Griffiths, B.S., K. Ritz, R.D. Bardgett, R. Cook, S. Christensen, F. Ekelund, S.J. Sorensen, E. Baath, J. Bloem, P.C. de Ruiter, J. Dolfing, B. Nicolardot. 2000. Ecosystem response of pasture soil communities to fumigation-induced microbial diversity reductions: an examination of the biodiversity-ecosystem function relationship. *Oikos*. 90: 279-294.

Guidot, A., J. Debaud, A. Effosse, R. Marmeisse. 2003. Below-ground distribution and persistence of an ectomycorrhizal fungus. *New Phytologist*. 161: 539-547.

Gyamfi, S., U. Pfeifer, M. Stierschneider, A. Sessitsch. 2000. Effects of transgenic glufosinate-tolerant oilseed rape (*Brassica napus*) and the associated herbicide application on eubacterial and *Pseudomonas* communities in the rhizosphere. *FEMS Microbiology Ecology*. 41: 181-190.

Harris, D. 1994. Analyses of DNA extracted from microbial communities. In *Beyond the Biomass*. Eds. K. Ritz, J. Dighton, K.E. Giller. John Wiley and Sons, Chichester, UK; 111-118.

Heuer, H., K. Smalla. 1999. Bacterial phyllosphere communities of *Solanum tuberosum* L. and T4-lysozyme-producing transgenic variants. *FEMS Microbiology Ecology*. 28: 357-371.

Heuer, H., R.M. Kroppenstedt, J. Lottmann, G. Berg, K. Smalla. 2002. Effects of T4 lysozyme release from transgenic potato roots on bacterial rhizosphere communities are negligible relative to natural factors. *Applied and Environmental Microbiology*. 68(3): 1325-1335.

Hill, J.E., R.P. Seipp, M. Betts, L. Hawkins, A.G. Van Kessel, W.L. Crosby, S.M. Hemmingsen. 2002. Extensive profiling of a complex microbial community by high-throughput sequencing. *Applied and Environmental Microbiology*. 68(6): 3055-3066.

Hill, J.E., S.L. Penny, K. G. Crowell, S. Goh, S.M. Hemmingsen. 2004. *cpnDB*: A chaperonin sequence database. *Genome Research*. 14: 1669-1675.

Hill, J.E., S.H. Goh, D.M. Money, M. Doyle, A. Li, W.L. Crosby, M.Links, A. Leung, D. Chan, S.M. Hemmingsen. 2005a. Characterization of vaginal microflora of healthy, nonpregnant women by chaperonin-60 sequence-based methods. *American Journal of Obstetrics and Gynecology*. 193: 682-692.

Hill, J.E., S.M. Hemmingsen, B.G. Goldade, T.J. Dumonceaux, J. Klassen, R.T. Zijlstra, S.H. Goh, A.G. Van Kessel. 2005b. Comparison of ileum microflora of pigs fed corn-, wheat- or barley-based diets by chaperonin-60 sequencing and quantitative PCR. *Applied and Environmental Microbiology*. 71(2): 867-875.

Hill, J.E., J.R. Town, S.M. Hemmingsen. 2006. Improved template representation in cpn60 polymerase chain reaction (PCR) product libraries generated from complex templates by application of a specific mixture of PCR primers. *Environmental Microbiology*. 8(4): 741-746.

Hill, T.C., K.A. Walsh, J.A. Harris, B.F. Moffett. 2003. Using ecological diversity measures with bacterial communities. *FEMS Microbiology Ecology*. 43: 1-11.

Horton, T.R., T.D. Bruns. 2001. The molecular revolution in ectomycorrhizal ecology: peeking in to the black box. *Molecular Ecology*. 10: 1855-1871.

Hughes, J.B., J. J. Hellmann, T.H. Ricketts, B.J. Bohannan. 2001. Counting the uncountable: Statistical approaches to estimating microbial diversity. *Applied and Environmental Microbiology*. 67(10): 4399-4406.

Hunt, J., L. Boddy, P.F. Randerson, H.J. Rogers. 2004. An evaluation of 18s rDNA approaches for the study of fungal diversity in grassland soils. *Microbial Ecology*. 47: 385-395.

Ibekwe, A.M., S.K. Papiernik, J. Gan, S.R. Yates, C. Yang, D.E. Crowley. 2001. Impact of fumigants on soil microbial communities. *Applied and Environmental Microbiology.* 67(7): 3245-3257.

James, C. 2005. Preview: Global status of commercialized transgenic crops: 2005. *ISAAA Briefs* No. 34. ISAAA: Ithica, NY

Janssen, P.H. 2006. Identifiying the dominant soil bacterial taxa in libraries of 16s rRNA and 16s rRNA genes. *Applied and Environmental Microbiology.* 72(3): 1719-1728.

Jansonn, H. 1982. Predacity by nematophagous fungi and its relation to the attraction of nematodes. *Microbial Ecology.* 8: 233-240.

Jian, W., L. Zhu, X. Dong. 2001. New approach to phylogenetic analysis of the genus *Bifidobacterium* based on partial HSP60 gene sequences. *International Journal of Systematic and Evolutionary Microbiology.* 51: 1633-1638.

Johnson, G.C. 1995. A comparison of the efficacy of six media for isolating wood-decaying hymenomycetes. *Canadian Journal of Microbiology.* 41(1): 104-107.

Jumponnen, A., L.C. Johnson. 2005. Can rDNA analyses of diverse fungal communities in soil and roots detect effects of environmental manipulations – a case study from tallgrass prairie. *Mycologia.* 97(6): 1177-1194.

Kendrick. B. 1992. *The Fifth Kingdom.* Focus Information Group, Inc., Newburyport, MA, USA; 1-406.

Kent, A.D., E.W. Triplett. 2002. Microbial communities and their interactions in soil and rhizosphere ecosystems. *Annual Review of Microbiology.* 56(1): 211-236.

Kowalchuk, G.A., M. Bruinsma, J.A. van Veen. 2003. Assessing responses of soil microorganisms to GM plants. *Trends in Ecology and Evolution.* 18(8): 403-410.

Krebs, C. 1999. Species diversity measures. In *Ecological Methodology 2nd Ed.* Addison Wesley Educational Publishers Inc., New York, NY, USA; 411-451.

Kumar, S., K. Tamura, M. Nei. 2004. Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Briefings in Bioinformatics.* 5: 150-163.

Kurakov, A.V., N.V. Kostina. 2001. Spatial peculiarities in the colonization of the plant rhizoplane by microscopic fungi. *Microbiology.* 70(2): 165-174.

Kwok, A.Y., J.T. Wilson, M. Coulthart, L. Ng, L. Mutharia, A. Chow. 2002. Phylogenetic study and identification of human pathogenic *Vibrio* species based on partial *hsp60* gene sequences. *Canadian Journal of Microbiology.* 48: 903-910.

Li, Y., R. Jeewon, K.D. Hyde, M. Mo, K. Zhang. 2006. Two new species of nematode-trapping fungi: relationships inferred from morphology, rDNA, and protein gene sequence analysis. *Mycological Research*. 110(7): 798-800.

Li, L., J.C. Steffens. 2002. Overexpression of polyphenyl oxidase in transgenic tomato plants results in enhanced bacterial disease resistance. *Planta*. 215(2): 239-247.

Lilley, A.K., M.J. Bailey, C. Cartwright, S.L. Turner, P.R. Hirsch. 2006. Life on earth: the impact of GM plants on soil ecology? *Trends in Biotechnology*. 24(1): 9-14.

Loreau, M. 2001. Microbial diversity, producer-decomposer interactions and ecosystem processes: a theoretical model. *Proceedings of the Royal Society of London – Biological Sciences*. 268: 303-309.

Lottmann, J, H. Heuer, K. Smalla, G. Berg. 1999. Influence of transgenic T4-lysozyme producing potato plants on potentially beneficial plant-associated bacteria. *FEMS Microbiology Ecology*. 29: 365-377.

Lottmann, J., H. Heuer, J. de Vries, A. Mahn, K. During, W. Wackernagel, K. Smalla, G. Berg. 2000. Establishment of introduced antagonistic bacteria in the rhizosphere of transgenic potatoes and their effect on the bacterial community. *FEMS Microbiology Ecology*. 33: 41-49.

Lui, B., Q. Zeng, F. Yan, H. Xu, C. Xu. 2005. Effects of transgenic plants on soil microorganisms. *Plant and Soil*. 271: 1-13.

Luis, S., R.M. Manshardt, M.M. Fitch, J.L. Slightom, J.C. Sanford, and D. Gonsalves. 1997. Pathogen-derived resistance provides papaya with effective protection against papaya ringspot virus. *Molecular Breeding*. 3: 161-168.

Lukow, T., P.F. Dunfield, W. Liesack. 2000. Use of the T-RFLP technique to assess spatial and temporal changes in the bacterial community structure within an agricultural soil planted with transgenic and non-transgenic potato plants. *FEMS Microbiology Ecology*. 32: 241-247.

Maguire, M., Coates, A., Henderson, B. 2002. Chaperonin 60 unfolds its secrets of cellular communication. *Cell Stress and Chaperones*. 7(4): 317-329.

Magurran, A. 1988. *Ecological diversity and its measurement*. Princeton University Press, Princeton, NJ, USA; 1-179.

Mansouri, H., A. Petit, P. Oger, Y. Dessaux. 2002. Engineered rhizosphere: the trophic bias generated by opine-producing plants is independent of the opine type, the soil origin and the plant species. *Applied and Environmental Microbiology*. 68(5): 2562-2566.

Mathew, A.G., H.A. Parpia. 1971. Food browning as a polyphenyl reaction. *Advances in Food Research*. 19: 75-145.

Martin, A.P. 2002. Phylogenetic approaches for describing and comparing the diversity of microbial communities. *Applied and Environmental Microbiology*. 68(8): 3673-3682.

Martin, J. M. Mayhew, T. Langer, F. Harl. 1993. The reaction cycle of GroEL and GroES in chaperonin-assisted protein folding. *Nature (London)*. 366: 228-233.

Martorell, P., M.T. Fernández-Espinar, A. Querol. 2005. Sequence-based identification of species belonging to the genus Debaryomyces. *FEMS Yeast Research*. 5: 1157-1165.

Mayer, A.M., E. Harel. 1979. Polyphenyl oxidases in plants. *Phytochemistry*. 18: 193-215.

McCaig, A.E., A. Glover, J.I. Prosser. 1999. Molecular analysis of bacterial community structure and diversity in unimproved and improved upland grass pastures. *Applied and Environmental Microbiology*. 65(4): 1721-1730.

Mikola, J., H. Setälä. 1998. Productivity and trophic-level biomasses in a microbial-based soil food web. *Oikos*. 82: 158-168.

Milling, A., K. Smalla, F.X. Maidl, M. Schloter, J. C. Munch. 2004. Effects of transgenic potatoes with an altered starch composition on the diversity of soil and rhizosphere bacteria and fungi. *Plant and Soil*. 266: 23-39.

Mohammadi, M., H. Kazemi. 2002. Changes in peroxidase and polyphenyl oxidase activities in susceptible and resistant wheat heads inoculated with *Fusarium graminearum* and induced resistance. *Plant Science*. 162: 491-498.

Morgan, J.A., G.D. Bending, P.J. White. Biological costs and benefits to plant-microbe interactions in the rhizosphere. *Journal of Experimental Botany*. 56(417): 1729-1739.

Nannipieri, P., J. Ascher, M.T. Ceccherini, L. Landi, G. Pietramellara, G. Renella. 2003. Microbial diversity and soil functions. *European Journal of Soil Science*. 54: 655-670.

Nevill, R.J., P.M. Hall, J. Beale. 1995. Forest health research needs in British Columbia. *Forestry Chronicle*. 71(4): 489-496.

Normander, B., J.I. Prosser. 2000. Bacterial origin and community composition in the barley phytosphere as a function of habitat and presowing conditions. *Applied and Environmental Microbiology*. 66(10): 4372-4377.

Nübel, U., B. Engelen, A. Felske, J. Snaidr, A. Wieshuber, R. I. Amann, W. Ludwig, and H. Backhaus. 1996. Sequence heterogeneities of genes encoding 16S rRNAs in *Paenibacillus polymyxa* detected by temperature gradient gel electrophoresis. *Journal of Bacteriology*. 178(19): 5636-5643.

Obertello, M., C. Santi, M. Sy, L. Laplaze, F. Auguy, D. Bogusz, C. Franche. 2005. Comparison of four constitutive promoters for the expression of transgenes in the tropical nitrogen-fixing tree *Allocasuarina verticillata*. *Plant Cell Reports*. 24: 540-548.

O'Callaghan, M., E.M. Gerard, N.W. Waipara, S.D. Young, T.R. Glare, P.J. Barrell, A.J. Conner. 2004. Microbial communities of *Solanum tuberosum* and magainin-producing transgenic lines. *Plant and Soil*. 266: 47-56.

Oger, P., A. Petit, Y Dessaux. 1997. Genetically engineered plants producing opines alter their biological environment. *Nature Biotechnology*. 15: 369-372.

Ohtsuka, E., S. Matsuki, M. Ikehara, Y. Takahashi, K. Matsubara. 1985. An alternative approach to deoxyoligonucleotides as hybridization probes by insertion of deoxyinosine at ambiguous codon positions. *Journal of Biological Chemistry*. 260: 2605-2608.

Page , R.D., E.C. Holmes. 1999. Measuring genetic change. In *Molecular evolution: A Phylogenetic Approach*. Blackwell Science, UK; 135-172.

Pearson, W., D. Lipman. 1988. Improved tools for biological sequence comparison. *Proceedings of the National Academy of Sciences (USA)*. 85: 2444-2448.

Peter, M.G. 1989. Chemical modifications of biopolymers by quinones and quinone methides. *Angewandte Chemie International Edition in English*. 28(5): 555-570.

Priha, O., S.J. Grayston, R. Hiukka, T. Pennanen, A. Smolander. 2001. Microbial community structure and characteristics of the organic matter in soils under *Pinus sylvestris*, *Picea abies* and *Betula pendula* at two forest sites. *Biology and Fertility of Soils*. 22: 17-24.

Raj, S.N., B.R. Sarosh, H.S. Shetty. 2006. Induction and accumulation of polyphenyl oxidase activities as implicated in development of resistance against pearl millet downy mildew disease. *Functional Plant Biology*. 33: 563-571.

Ramirez-Saad, H.C., A. Sessitsch, W.M. de Vos, A.D. Akkermans. 2000. Bacterial community changes and enrichment of Burkholderia-like bacteria induced by chlorinated benzoates in a peat-forest soil-microcosm. *Systematic and Applied Microbiology*. 23: 591-98.

Ranjard, L. F. Poly, S. Nazaret. 2000. Monitoring complex bacterial communities by culture-independent molecular techniques: application to soil environment. *Research in Microbiology*. 151(3): 167-177.

Ranjard, L., D.P. Lejon, C. Mougel, L. Schehrer, D. Merdinoglu, R. Chaussod. 2003. Sampling strategy in molecular microbial ecology: influence of soil sample size on DNA fingerprinting analysis of fungal and bacterial communities. *Environmental Microbiology*. 5(11): 1111-1120.

Rasche, F., V. Hodl, C. Poll, E. Kandeler, M.H. Gerzabek, J.D. van Elsas, A. Sessitsch. 2006. Rhizosphere bacteria affected by transgenic potatoes with antibacterial activities compared with the effects of soil, wild-type potatoes, vegetation stage and pathogen exposure. *FEMS Microbiology Ecology*. 56: 219-235.

Richardson, J., D.E. Stead, J.G. Elphinstone, R.H. Coutts. 2002. Diversity of *Burkholderia* isolates from woodland rhizosphere environments. *Journal of Applied Microbiology*. 93: 616-630.

Roselló-Mora, R., R. Amann. 2001. The species concept for prokaryotes. *FEMS Microbiology Reviews*. 25: 39-67.

Saibil, H., and Ranson, N. 2002. The chaperonin folding machine. *Trends in Biochemical Science*. 27: 627-632.

Sampedro, I., E. Aranda, J.M. Seervino, S. Fracchia, I. Garcia-Romera, J. A. Ocampo, A. Godeas. 2004. Improvement by soil yeasts of arbuscular mycorrhizal symbiosis of soybean (*Glycine max*) colonized by *Glomus mosseae*. *Mycorrhiza*. 14: 229-234.

Saxena, D., and G. Stotzky. 2000. Insecticidal toxin from *Bacillus thuringiensis* is released from roots of transgenic *Bt* corn in vitro and in situ. *FEMS Microbiology Ecology*. 33(1): 35-39.

Schloss, P.D., B.R. Larget, J. Handelsman. 2004. Integration of microbial ecology and statistics: a test to compare gene libraries. *Applied and Environmental Microbiology*. 70(9): 5485-5492.

Schloss, P.D., J. Handelsman. 2005. Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Applied and Environmental Microbiology*. 71(3): 1501-1506.

Schloss, P.D. and J. Handelsman. 2006. Introducing TreeClimber, a test to compare microbial community structures. *Applied and Environmental Microbiology*. 72(4): 2379-2384.

Schmalenberger, A., C.C. Tebbe. 2002. Bacterial community composition in the rhizosphere of a transgenic, herbicide resistant maize (*Zea mays*) and comparison to its non-transgenic cultivar *Bosphore*. *FEMS Microbiology Ecology*. 40: 29-37.

Schröder, P., A. Hartmann. 2003. New developments in rhizosphere research. *Journal of Soils and Sediments*. 3(4): 227.

Sherman, T.D., K.C. Vaughn, S.O. Duke. 1991. A limited survey of the phylogenetic distribution of polyphenyl oxidase. *Phytochemistry*. 30(8): 2499-2506.

Shewan, J.M., T.A. McMeekin. 1983. Taxonomy (and ecology) of *Flavobacterium* and related genera. *Annual Reviews of Microbiology*. 37: 233-252.

Siciliano , S.D., J.J. Germida. 1999. Taxonomic diversity of bacteria associated with the roots of field grown transgenic *Brassica napus* cv. Quest, compared to the non-transgenic *B. napus* cv. Excel and *B. rapa* cv. Parkland. *FEMS Microbiology Ecology*. 29: 263-272.

Singh, B.K., and A. Walker. 2006. Microbial degradation of organophosphorus compounds. *FEMS Microbiology Reviews*. 30: 428-471.

Singleton, D.R., M.A. Furlong, S.L. Rathbun, W.B. Whitman. 2001. Quantitative comparisons of 16s rRNA gene sequence libraries from environmental samples. *Applied and Environmental Microbiology*. 67(9): 4374-4376.

Sláviková, E., R. Vadkertiová. 2000. The occurrence of yeasts in the forest soils. *Journal of Basic Microbiology*. 40(3): 207-212.

Smit, E., P. Leeflang, B. Glandorf, J.D. van Elsas, K. Wernars. 1999. Analysis of fungal diversity in the wheat rhizosphere by sequencing of cloned PCR-amplified genes encoding 18S rRNA and temperature gradient gel electrophoresis. *Applied and Environmental Microbiology*. 65(6): 2614-2621.

Sokal, R.R., F.J. Rolf. 1995. *Biometry: the principles and practice of statistics in biological research. 3rd Ed.* W.H. Freeman, New York, NY, USA; 1-887.

Stach, J.E., L.A. Maldonado, D.G. Masson, A.C. Ward, M. Goodfellow, A.T. Bull. 2003. Statistical approaches for estimating Actinobacterial diversity in marine sediments. *Applied and Environmental Microbiology*. 69(10): 6189-6200.

Tang, W, R.J. Newton. 2003. Genetic transformation of conifers and its application in forest biotechnology. *Plant Cell Reports*. 22: 1-15.

Tesfaye, M., N.S. Dufault, M.R. Dornbusch, D.L. Allan, C.P. Vance, D.A. Samac. 2003. Influence of enhanced malate dehydrogenase expression by alfalfa on diversity of rhizobacteria and soil nutrient availability. *Soil Biology and Biochemistry*. 35(8): 1103-1113.

Tomilova, O.G., M.V. Shternshis. 2006. The effect of a preparation from *Chaetomium* fungi on the growth of phytopathogenic fungi. *Applied Biochemistry and Microbiology*. 42(1): 67-71.

Trevors, J.T. 1998. Bacterial biodiversity in soil with an emphasis on chemically-contaminated soils. *Water, Air, and Soil Pollution*. 101: 45-67.

Van Bruggen, A.H., A.M. Semenov. In search of biological indicators for soil health and disease suppression. *Applied Soil Ecology*. 15: 13-24.

Van Frankenhuyzen, K., T. Beardmore. 2004. Current status and environmental impact of transgenic forest trees. *Canadian Journal of Forestry*. 34: 1163-1180.

Vaughn, K.C., S.O. Duke. 1984. Function of polyphenyl oxidase in higher plants. *Physiologia Plantarum*. 60: 106-112.

Viale, A.M., A.K. Arakaki, F.C. Soncini, R.G. Ferreyra. 1994. Evolutionary relationships among eubacterial groups as inferred from GroEl (Chaperonin) sequence comparisons. *International Journal of Systematic Bacteriology*. 44: 527-533.

Vierheilig, H., M. Alt, J. Lange, M. Gut-Rella, A. Wiemken, T. Boller. 1995. Colonization of transgenic tobacco constitutively expressing pathogenesis-related proteins by the vesicular-arbuscular mycorrhizal fungus *Glomus mosseae*. *Applied and Environmental Microbiology*. 61(8): 3031-3034.

De Vries, J., K. Harms, I. Broer, G. Kriete, A. Mahn, K. During, W. Wackernagel. 1999. The bacteriolytic activity in transgenic potatoes expressing a chimeric T4 lysozyme gene and the effect of T4 lysozyme on soil- and phytopathogenic bacteria. *Systematic and Applied Microbiology*. 22: 280-286.

Visser, S., D. Maynard, R.M. Danielson. 1998. Response of ecto- and arbuscular mycorrhizal fungi to clear-cutting and the application of chipped aspen wood in a mixed wood site in Alberta, Canada. *Applied Soil Ecology*. 7: 257-269.

Wang, J., and C.P. Constabel. 2003. Biochemical characterization of two differentially expressed polyphenol oxidases from hybrid poplar. *Phytochemistry*. 64: 115-121.

Wang, J., and C.P. Constabel. 2004. Polyphenyl oxidase overexpression in transgenic *Populus* enhances resistance to herbivory by forest tent caterpillar (*Malacosoma disstria*). *Planta*. 220: 87-96.

Wang, M., S. Ahrne, B. Jeppsson, G. Molin. 2005. Comparison of bacterial diversity along the human intestinal tract by direct cloning and sequencing of 16S rRNA genes. *FEMS Microbiology Ecology*. 54: 219-231.

Whipps, J.M. 2001. Microbial interactions and biocontrol in the rhizosphere. *Journal of Experimental Botany*. 52: 487-511.

White, T.J., T. Bruns, S. Lee, J.W. Taylor. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In *PCR Protocols: A Guide to Methods and Applications*. Eds. Innis, M.A., D.H. Gelfand, J.J. Sninsky, T.J White. Academic Press Inc., New York, NY, USA; 315-322.

Willoughby, L.G. 1998. A quantitative ecological study on the monocentric soil chytrid, *Rhizophlyctis rosea,* in Provence. *Mycological Research*. 102(11): 1338-1342.

Wintzingerode, F.V., U.B. Gobel, E. Stackebrandt. 1997. Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiology Reviews*. 21(3): 213-229.

Wititsuwannakul, D., N. Chareonthiphakorn, M. Pace, R. Wititsuwannkul. 2002. Polyphenyl oxidases from latex of *Hevea brasiliensis*: purification and characterization. *Phytochemistry*. 61: 115-121.

Wolfenbarger, L.L., P.R. Phifer. 2000. The ecological risks and benefits of genetically engineered plants. *Science*. 290: 2088-2093.

Xie, C., A. Yokota. 2006. Reclassification of (*Flavobacterium*) *ferrugineum* as *Terrimonas ferruginea* gen. nov., comb. nov., and description of *Terrimonas lutea* sp. nov., isolated from soil. *International Journal of Systematic and Evolutionary Microbiology*. 56: 1117-1121.

Yang, C., D.E.Crowley. 2000. Rhizosphere microbial community structure in relation to root location and plant iron nutritional status. *Applied and Environmental Microbiology*. 66(1): 345-351.

Zuberer, D. 1994. Recovery and enumeration of viable bacteria. In *Methods of soil analysis Part 2 Microbiological and biochemical properties*. Eds. Weaver, R.W, S. Angle, P. Bottomley, D. Bedicek, S. Smith, A. Tabatabai, and A. Wollum. Soil Science Society of America, Madison, WI, USA; 119-143.

Appendix 1

Table A1. Assignment of bacterial 16S sequences to hierarchical taxa. Operational taxonomic units (OTUs) were defined as sequences sharing 97% sequence similarity, and were classified using the Ribosomal Database Project Naïve Bayesian Classifier (Library Compare Tool). The numbers following each hierarchical rank are bootstrap confidence estimates (%).

OTU	Major Group		Subgroup		Order		Family		Genus	
1	Chloroflexi	76	Anaerolineae	76	Anaerolinaeles	76	Anaerolinaeceea	76	Anaerolinea	76
2	Proteobacteria	99	Alphaproteobacteria	98	Rhizobiales	78	Bradyrhizobiaceae	44	Balneimonas	35
3	Proteobacteria	100	Gammaproteobacteria	99	Chromatiales	59	Chromatiaceae	40	Thiorhodococcus	14
4	Actinobacteria	100	Actinobacteridae	100	Actinomycetales	100	Micrococcineae	100	Agreia	42
5	Verrucomicrobia	94	Verrucomicrobiae	94	Verrucomicrobiales	94	Verrucomicrobiaceae	54	Verrucomicrobium	52
6	Proteobacteria	84	Deltaproteobacteria	63	Myxococcales	51	Polyangiaceae	51	Chondromyces	51
7	Proteobacteria	99	Alphaproteobacteria	99	Rhizobiales	89	Hyphomicrobiaceae	49	Devosia	18
8	Proteobacteria	100	Gammaproteobacteria	100	Pseudomonadales	100	Pseudomonadaceae	100	Cellvibrio	100
9	Proteobacteria	37	Deltaproteobacteria	14	Myxococcales	7	Cystobacteraceae	7	Anaeromyxobacter	6
10	Actinobacteria	100	Actinobacteria	100	Rubrobacterales	100	Rubrobacteraceae	100	Conexibacter	67
11	Proteobacteria	100	Alphaproteobacteria	100	Rhizobiales	100	Hyphomicrobiaceae	100	Devosia	93
12	Proteobacteria	100	Alphaproteobacteria	100	Rhizobiales	100	Hyphomicrobiaceae	85	Devosia	85
13	Proteobacteria	87	Deltaproteobacteria	51	Myxococcales	34	Cystobacteraceae	14	Stigmatella	10
14	Proteobacteria	100	Gammaproteobacteria	100	Xanthomonadales	100	Xanthomonadaceae	100	Rhodanobacter	94
15	Firmicutes	35	Bacilli	29	Bacillales	29	Bacillaceae	23	Ureibacillus	20
16	Proteobacteria	100	Deltaproteobacteria	99	Myxococcales	99	Polyangiaceae	99	Chondromyces	99
17	Acidobacteria	100	Acidobacteria	100	Acidobacteriales	100	Acidobacteriaceae	100	Acidobacterium	100
18	Actinobacteria	79	Rubrobacteridae	74	Rubrobacterales	74	Rubrobacteraceae	74	Conexibacter	48
19	Firmicutes	52	Clostridia	52	Clostridiales	52	Clostridiaceae	52	Sulfuricurvum	52
20	Proteobacteria	99	Alphaproteobacteria	99	Rhodospirillales	87	Rhodospirillaceae	82	Phaeospirillum	39
21	Proteobacteria	100	Betaproteobacteria	96	Burkholderiales	39	Incertae sedis 5	17	Thiobacter	12
22	Actinobacteria	100	Actinobacteria	100	Actinomycetales	100	Frankineae	67	Sporichthya	42
23	Actinobacteria	100	Actinobacteria	100	Actinomycetales	100	Micrococcineae	100	Microbacterium	99
24	Verrucomicrobia	100	Verrucomicrobiae	100	Verrucomicrobiales	100	Opitutaceae	100	Opitutus	100
25	Proteobacteria	100	Gammaproteobacteria	99	Xanthomonadales	96	Xanthomonadaceae	96	Nevskia	62
26	Proteobacteria	100	Alphaproteobacteria	100	Rhizobiales	100	Bradyrhizobiaceae	100	Bradyrhizobium	100
27	Gemmatimonadetes	67	Gemmatimonadetes	67	Gemmatimonadales	67	Gemmatimonadaceae	67	Gemmatimonas	67
28	Acidobacteria	100	Acidobacteria	100	Acidobacteriales	100	Acidobacteriaceae	100	Acidobacterium	100

Table A1 continued.

OTU	Major Group		Subgroup		Order		Family		Genus	
29	Proteobacteria	96	Gammaproteobacteria	90	Xanthomonadales	57	Xanthomonadaceae	57	Lysobacter	13
30	Verrucomicrobia	99	Verrucomicrobiae	99	Verrucomicrobiales	99	Opitutaceae	99	Opitutus	99
31	Acidobacteria	100	Acidobacteria	100	Acidobacteriales	100	Acidobacteriaceae	100	Acidobacterium	100
32	Proteobacteria	45	Deltaproteobacteria	31	Myxococcales	22	Myxococcaceae	21	Corallococcus	21
33	Proteobacteria	83	Alphaproteobacteria	80	Rhodospirillales	56	Acetobacteraceae	54	Kozakia	11
34	Proteobacteria	100	Alphaproteobacteria	100	Rhizobiales	100	Hyphomicrobiaceae	100	Pedomicrobium	73
35	Proteobacteria	100	Betaproteobacteria	100	Burkholderiales	87	Comamonadaceae	71	Xenophilus	21
36	Actinobacteria	100	Actinobacteria	100	Actinomycetales	100	Micrococcaceae	100	Arthrobacter	98
37	Proteobacteria	96	Alphaproteobacteria	96	Rhodospirillales	62	Acetobacteraceae	41	Stella	29
38	Proteobacteria	100	Alphaproteobacteria	100	Rhizobiales	100	Beijerinckiaceae	85	Methylocapsa	48
39	Proteobacteria	100	Alphaproteobacteria	100	Rhizobiales	100	Hyphomicrobiaceae	56	Rhodoplanes	43
40	Proteobacteria	100	Betaproteobacteria	100	Burkholderiales	75	Incertae sedis 5	41	Thiobacter	35
41	Genera_incertae sedis TM7	100	TM7	100						
42	Proteobacteria	100	Alphaproteobacteria	100	Rhizobiales	100	Rhizobiaceae	98	Rhizobium	80
43	Proteobacteria	33	Gammaproteobacteria	13	Chromatiales	12	Chromatiaceae	12	Chromatium	11
44	Proteobacteria	100	Alphaproteobacteria	100	Rhizobiales	96	Hyphomicrobiaceae	43	Angulomicrobium	37
45	Proteobacteria	100	Betaproteobacteria	100	Burkholderiales	97	Comamonadaceae	73	Ramlibacter	60
46	Proteobacteria	100	Alphaproteobacteria	100	Rhizobiales	99	Hyphomicrobiaceae	44	Angulomicrobium	37
47	Proteobacteria	100	Alphaproteobacteria	100	Rhizobiales	100	Brucellaceae	86	Brucella	78
48	Proteobacteria	100	Betaproteobacteria	99	Rhodocyclales	65	Rhodocyclaceae	65	Rhodocyclus	57
49	Thermomicrobia	37	Thermomicrobia	37	Thermomicrobiales	37	Thermomicrobiaceae	37	Thermomicrobium	37
50	Proteobacteria	100	Alphaproteobacteria	100	Rhodospirillales	100	Rhodospirillaceae	74	Inquilinus	53
51	Proteobacteria	100	Alphaproteobacteria	100	Caulobacterales	97	Caulobacteraceae	97	Caulobacter	79
52	Proteobacteria	100	Alphaproteobacteria	100	Caulobacterales	97	Caulobacteraceae	97	Caulobacter	79
53	Proteobacteria	100	Betaproteobacteria	100	Burkholderiales	99	Comamonadaceae	96	Variovorax	52
54	Gemmatimonadetes	75	Gemmatimonadetes	75	Gemmatimonadales	75	Gemmatimonadaceae	75	Gemmatimonas	75
55	Proteobacteria	98	Alphaproteobacteria	95	Rhodospirillales	69	Acetobacteraceae	69	Kozakia	49
56	Actinobacteria	96	Actinobacteria	96	Actinomycetales	63	Nocardioidaceae	36	Actinopolymorpha	34
57	Gemmatimonadetes	82	Gemmatimonadetes	82	Gemmatimonadales	82	Gemmatimonadaceae	82	Gemmatimonas	82
58	Proteobacteria	100	Alphaproteobacteria	99	Caulobacterales	66	Caulobacteraceae	66	Brevundimonas	43
59	Actinobacteria	100	Actinobacteria	100	Actinomycetales	100	Nocardioidaceae	99	Nocardioides	90

Table A1 continued

OTU	Major Group		Subgroup		Order		Family		Genus	
60	Proteobacteria	100	Alphaproteobacteria	100	Rhizobiales	100	Phyllobacteriaceae	99	Mesorhizobium	99
61	Proteobacteria	100	Alphaproteobacteria	100	Rhizobiales	95	Hyphomicrobiaceae	81	Rhodoplanes	45
62	Proteobacteria	100	Alphaproteobacteria	100	Rhodospirillales	100	Acetobacteraceae	100	Acidiphilium	19
63	Genera incertae sedis TM7	100	TM7	100						
64	Proteobacteria	100	Betaproteobacteria	100	Burkholderiales	100	Incertae sedis 5	94	Leptothrix	82
65	Proteobacteria	98	Deltaproteobacteria	98	Myxococcales	98	Polyangiaceae	98	Chondromyces	98
66	Proteobacteria	98	Alphaproteobacteria	84	Rhodospirillales	51	Rhodospirillaceae	40	Rhodocista	29
67	Proteobacteria	100	Betaproteobacteria	100	Burkholderiales	92	Comamonadaceae	89	Comamonas	26
68	Proteobacteria	94	Betaproteobacteria	48	Burkholderiales	39	Comamonadaceae	26	Caenibacterium	16
69	Verrucomicrobia	92	Verrucomicrobiae	92	Verrucomicrobiales	92	Verrucomicrobiaceae	46	Verrucomicrobium	45
70	Proteobacteria	93	Deltaproteobacteria	90	Myxococcales	81	Nannocystaceae	43	Plesiocystis	26
71	Proteobacteria	97	Alphaproteobacteria	96	Rhodospirillales	52	Acetobacteraceae	52	Rubritepida	22
72	Proteobacteria	87	Deltaproteobacteria	79	Desulfuromonales	42	Desulfuromonaceae	36	Malonomonas	36
73	Proteobacteria	100	Gammaproteobacteria	90	Xanthomonadales	88	Xanthomonadaceae	88	Nevskia	51
74	Proteobacteria	96	Alphaproteobacteria	96	Rhodospirillales	65	Acetobacteraceae	56	Acidisphaera	17
75	Actinobacteria	100	Actinobacteria	100	Actinomycetales	100	Nocardioidaceae	98	Nocardioides	95
76	Proteobacteria	100	Alphaproteobacteria	100	Rhizobiales	96	Hyphomicrobiaceae	86	Devosia	74
77	Proteobacteria	100	Alphaproteobacteria	100	Caulobacterales	100	Caulobacteraceae	100	Caulobacter	100
78	Proteobacteria	100	Alphaproteobacteria	100	Rhizobiales	100	Rhizobiaceae	100	Rhizobium	100
79	Proteobacteria	92	Deltaproteobacteria	61	Myxococcales	35	Cystobacteraceae	33	Anaeromyxobacter	30
80	Proteobacteria	100	Alphaproteobacteria	100	Rhizobiales	100	Hyphomicrobiaceae	58	Devosia	50
81	Proteobacteria	100	Gammaproteobacteria	100	Oceanospirillales	98	Halomonadaceae	95	Chromohalobacter	46
82	Proteobacteria	100	Alphaproteobacteria	99	Rhizobiales	99	Hyphomicrobiaceae	73	Angulomicrobium	49
83	Proteobacteria	95	Alphaproteobacteria	90	Rhodospirillales	46	Acetobacteraceae	23	Acidisphaera	13
84	Actinobacteria	100	Actinobacteria	100	Actinomycetales	100	Intrasporangiaceae	75	Ornithinicoccus	69
85	Proteobacteria	100	Alphaproteobacteria	100	Rhizobiales	100	Hyphomicrobiaceae	100	Xanthobacter	86
86	Firmicutes	35	Clostridia	25	Clostridiales	18	Clostridiaceae	14	Thermohalobacter	14
87	Gemmatimonadetes	92	Gemmatimonadetes	92	Gemmatimonadales	92	Gemmatimonadaceae	92	Gemmatimonas	92
88	Proteobacteria	100	Deltaproteobacteria	100	Sorangineae	100	Polyangiaceae	100	Chondromyces	100
89	Verrucomicrobia	100	Verrucomicrobiae	100	Verrucomicrobiales	100	Opitutaceae	100	Opitutus	100
90	Proteobacteria	100	Betaproteobacteria	100	Burkholderiales	100	Burkholderiaceae	100	Burkholderia	100

Table A1 continued

OTU	Major Group		Subgroup		Order		Family		Genus	
91	Proteobacteria	100	Alphaproteobacteria	100	Caulobacterales	100	Caulobacteraceae	100	Phenylobacterium	100
92	Actinobacteria	100	Actinobacteria	100	Rubrobacterales	100	Rubrobacteraceae	100	Conexibacter	96
93	Proteobacteria	97	Alphaproteobacteria	97	Rhodospirillales	56	Acetobacteraceae	38	Saccharibacter	14
94	Actinobacteria	100	Actinobacteria	100	Actinomycetales	100	Mycobacteriaceae	99	Mycobacterium	99
95	Proteobacteria	100	Alphaproteobacteria	100	Rhizobiales	99	Hyphomicrobiaceae	44	Rhodoplanes	36
96	Proteobacteria	53	Gammaproteobacteria	27	Chromatiales	13	Chromatiaceae	9	Thiococcus	6
97	Proteobacteria	100	Betaproteobacteria	100	Burkholderiales	100	Oxalobacteraceae	100	Telluria	76
98	Proteobacteria	38	Deltaproteobacteria	24	Myxococcales	12	Cystobacteraceae	9	Melittangium	4
99	Chloroflexi	71	Anaerolineae	69	Anaerolinaeles	69	Anaerolinaeceea	69	Anaerolinea	69
100	Proteobacteria	100	Alphaproteobacteria	100	Rhizobiales	100	Rhizobiaceae	58	Rhizobium	35
101	Proteobacteria	100	Alphaproteobacteria	100	Rhizobiales	99	Hyphomicrobiaceae	51	Rhodoplanes	26
102	Gemmatimonadetes	70	Gemmatimonadetes	70	Gemmatimonadales	70	Gemmatimonadaceae	70	Gemmatimonas	70
103	Proteobacteria	49	Deltaproteobacteria	25	Syntrophobacterales	12	Syntrophobacteraceae	11	Thermodesulforhabdus	8
104	Proteobacteria	100	Betaproteobacteria	100	Burkholderiales	100	Burkholderiaceae	100	Wautersia	100
105	Proteobacteria	97	Alphaproteobacteria	92	Rhodospirillales	76	Rhodospirillaceae	73	Phaeospirillum	43
106	Proteobacteria	99	Alphaproteobacteria	96	Rhodospirillales	56	Acetobacteraceae	50	Acidocella	22
107	Actinobacteria	100	Actinobacteria	100	Actinomycetales	100	Sporichthyaceae	30	Sporichthya	30
108	Proteobacteria	100	Alphaproteobacteria	100	Rhizobiales	100	Hyphomicrobiaceae	100	Hyphomicrobium	100
109	Proteobacteria	100	Gammaproteobacteria	100	Pseudomonadales	100	Pseudomonadaceae	100	Pseudomonas	100
110	Actinobacteria	100	Actinobacteria	100	Actinomycetales	100	Kineosporiaceae	16	Cryptosporangium	14
111	Cyanobacteria	82	Cyanobacteria	82	Subsection 1	48	Family 1.1	48	Microcystis	23
112	Firmicutes	19	Clostridia	17	Clostridiales	9	Acidaminococcaceae	3	Anaeromusa	1
113	Gemmatimonadetes	59	Gemmatimonadetes	59	Gemmatimonadales	59	Gemmatimonadaceae	59	Gemmatimonas	59
114	Chloroflexi	79	Anaerolineae	78	Anaerolinaeles	78	Anaerolinaeceea	78	Anaerolinea	78
115	Proteobacteria	100	Alphaproteobacteria	100	Rhizobiales	100	Rhizobiaceae	58	Ensifer	58
116	Actinobacteria	100	Actinobacteria	100	Actinomycetales	100	Nocardioidaceae	35	Nocardioides	27
117	Acidobacteria	100	Acidobacteria	100	Acidobacteriales	100	Acidobacteriaceae	100	Acidobacterium	100
118	Acidobacteria	45	Acidobacteria	45	Acidobacteriales	45	Acidobacteriaceae	45	Acidobacterium	32
119	Proteobacteria	100	Betaproteobacteria	100	Burkholderiales	89	Incertae sedis 5	56	Schlegelella	34
120	Bacteroidetes	100	Sphingobacteria	100	Sphingobacteriales	100	Sphingobacteriaceae	95	Pedobacter	57
121	Proteobacteria	100	Alphaproteobacteria	100	Rhodospirillales	91	Acetobacteraceae	85	Stella	77
122	Actinobacteria	94	Rubrobacteridae	88	Rubrobacterales	88	Rubrobacteraceae	88	Conexibacter	77

Table A1 continued

OTU	Major Group		Subgroup		Order		Family		Genus	
123	Firmicutes	100	Bacilli	65	Bacillales	64	Paenibacillaceae	54	Paenibacillus	38
124	Proteobacteria	100	Alphaproteobacteria	100	Rhizobiales	100	Methylobacteriaceae	96	Methylobacterium	96
125	Proteobacteria	100	Gammaproteobacteria	100	Xanthomonadales	100	Xanthomonadaceae	100	Hydrocarboniphaga	63
126	Proteobacteria	33	Alphaproteobacteria	14	Rickettsiales	7	Incertae sedis 4	7	Caedibacter	7
127	Cyanobacteria	28	Cyanobacteria	28	Subsection 1	15	Family 1.1	15	Microcystis	8
128	Proteobacteria	57	Deltaproteobacteria	35	Syntrophobacterales	23	Syntrophobacteraceae	22	Desulforhabdus	19
129	Proteobacteria	100	Alphaproteobacteria	100	Rhizobiales	98	Hyphomicrobiaceae	30	Angulomicrobium	15
130	Cyanobacteria	25	Cyanobacteria	25	Subsection 1	16	Family 1.1	16	Dactylococcopsis	16
131	Actinobacteria	94	Actinobacteridae	48	Actinomycetales	40	Nocardioidaceae	15	Actinopolymorpha	15
132	Proteobacteria	94	Deltaproteobacteria	90	Myxococcales	84	Nannocystaceae	44	Nannocystis	29
133	Proteobacteria	100	Betaproteobacteria	100	Burkholderiales	87	Incertae sedis 5	58	Thiomonas	31
134	Proteobacteria	100	Alphaproteobacteria	99	Rhizobiales	99	Hyphomicrobiaceae	74	Rhodoplanes	66
135	Proteobacteria	100	Alphaproteobacteria	100	Rhizobiales	99	Hyphomicrobiaceae	44	Ancylobacter	15
136	Proteobacteria	100	Betaproteobacteria	100	Burkholderiales	99	Incertae sedis 5	77	Leptothrix	29
137	Proteobacteria	99	Alphaproteobacteria	98	Rhodospirillales	58	Acetobacteraceae	57	Rubritepida	20
138	Proteobacteria	100	Alphaproteobacteria	100	Rhizobiales	100	Hyphomicrobiaceae	100	Devosia	93
139	Gemmatimonadetes	66	Gemmatimonadetes	66	Gemmatimonadales	66	Gemmatimonadaceae	66	Gemmatimonas	66
140	Proteobacteria	99	Gammaproteobacteria	95	Xanthomonadales	65	Xanthomonadaceae	65	Fulvimonas	13
141	Proteobacteria	93	Alphaproteobacteria	92	Rhodospirillales	69	Acetobacteraceae	69	Swaminathania	36
142	Actinobacteria	100	Actinobacteria	100	Actinomycetales	100	Nocardioidaceae	40	Nocardioides	24
143	Proteobacteria	100	Alphaproteobacteria	98	Rhodospirillales	96	Rhodospirillaceae	74	Inquilinus	40
144	Proteobacteria	100	Alphaproteobacteria	100	Caulobacterales	100	Caulobacteraceae	100	Asticcacaulis	80
145	Proteobacteria	100	Alphaproteobacteria	100	Rhizobiales	100	Rhizobiaceae	98	Rhizobium	93
146	Gemmatimonadetes	99	Gemmatimonadetes	99	Gemmatimonadales	99	Gemmatimonadaceae	99	Gemmatimonas	99
147	Proteobacteria	44	Deltaproteobacteria	21	Syntrophobacterales	14	Syntrophobacteraceae	14	Desulfovirga	13
148	Proteobacteria	100	Alphaproteobacteria	100	Caulobacterales	100	Caulobacteraceae	100	Brevundimonas	100
149	Genera_incertae sedis BRC1	42	BRC1	42						
150	Proteobacteria	99	Alphaproteobacteria	99	Rhodospirillales	86	Acetobacteraceae	76	Stella	68
151	Actinobacteria	99	Actinobacteria	99	Actinomycetales	82	Micromonosporaceae	62	Longispora	58
152	Proteobacteria	100	Betaproteobacteria	100	Nitrosomonadales	47	Gallionellaceae	31	Gallionella	31
153	Bacteroidetes	100	Flavobacteria	100	Flavobacteriales	100	Flavobacteriaceae	100	Flavobacterium	100

Table A1 continued

OTU	Major Group		Subgroup		Order		Family		Genus	
154	Proteobacteria	100	Betaproteobacteria	100	Burkholderiales	76	Incertae sedis 5	58	Thiobacter	52
155	Gemmatimonadetes	96	Gemmatimonadetes	96	Gemmatimonadales	96	Gemmatimonadaceae	96	Gemmatimonas	96
156	Proteobacteria	100	Alphaproteobacteria	100	Rhodospirillales	87	Acetobacteraceae	70	Stella	52
157	Proteobacteria	100	Alphaproteobacteria	100	Rhodospirillales	94	Rhodospirillaceae	89	Inquilinus	87
158	Proteobacteria	99	Alphaproteobacteria	96	Rhizobiales	40	Bradyrhizobiaceae	17	Balneimonas	15
159	Proteobacteria	100	Betaproteobacteria	100	Burkholderiales	100	Burkholderiaceae	100	Burkholderia	100
160	Bacteroidetes	100	Sphingobacteria	100	Sphingobacteriales	100	Sphingobacteriaceae	100	Pedobacter	89
161	Proteobacteria	100	Gammaproteobacteria	96	Legionellales	32	Legionellaceae	18	Legionella	18
162	Proteobacteria	100	Alphaproteobacteria	100	Rhizobiales	100	Phyllobacteriaceae	56	Pseudaminobacter	52
163	Gemmatimonadetes	91	Gemmatimonadetes	91	Gemmatimonadales	91	Gemmatimonadaceae	91	Gemmatimonas	91
164	Proteobacteria	100	Alphaproteobacteria	99	Rhodospirillales	81	Rhodospirillaceae	77	Phaeospirillum	40
165	Proteobacteria	100	Gammaproteobacteria	100	Xanthomonadales	100	Xanthomonadaceae	100	Nevskia	69
166	Actinobacteria	100	Actinobacteria	100	Actinomycetales	100	Micrococcaceae	100	Arthrobacter	98
167	Proteobacteria	99	Gammaproteobacteria	59	Chromatiales	35	Chromatiaceae	35	Chromatium	24
168	Proteobacteria	45	Deltaproteobacteria	25	Myxococcales	14	Cystobacteraceae	12	Anaeromyxobacter	9
169	Proteobacteria	100	Alphaproteobacteria	100	Rhizobiales	100	Rhizobiaceae	72	Rhizobium	72
170	Proteobacteria	54	Deltaproteobacteria	29	Syntrophobacterales	23	Syntrophaceae	19	Desulfobacca	19
171	Proteobacteria	100	Alphaproteobacteria	100	Rhizobiales	100	Rhizobiaceae	100	Agrobacterium	76
172	Proteobacteria	100	Betaproteobacteria	100	Burkholderiales	99	Oxalobacteraceae	99	Duganella	80
173	Proteobacteria	100	Alphaproteobacteria	100	Rhizobiales	100	Rhizobiaceae	66	Rhizobium	66
174	Chloroflexi	90	Anaerolineae	90	Anaerolinaeles	90	Anaerolinaeceea	90	Anaerolinea	90
175	Bacteroidetes	100	Flavobacteria	100	Flavobacteriales	100	Flavobacteriaceae	100	Flavobacterium	74
176	Spirochaetes	97	Spirochaetes	97	Spirochaetales	97	Spirochaetaceae	97	Spirochaeta	97
177	Proteobacteria	100	Alphaproteobacteria	100	Rhizobiales	100	Brucellaceae	65	Brucella	62
178	Genera_incertae_sedis_TM7	100	TM7	100						
179	Proteobacteria	100	Alphaproteobacteria	100	Rhizobiales	100	Bradyrhizobiaceae	71	Bosea	70
180	Proteobacteria	100	Gammaproteobacteria	100	Legionellales	57	Coxiellaceae	57	Aquicella	49
181	Proteobacteria	100	Alphaproteobacteria	100	Caulobacterales	91	Caulobacteraceae	91	Asticcacaulis	67
182	Actinobacteria	97	Actinobacteridae	52	Actinomycetales	52	Nocardioidaceae	26	Actinopolymorpha	15
183	Proteobacteria	100	Alphaproteobacteria	100	Rhizobiales	99	Hyphomicrobiaceae	58	Xanthobacter	22
184	Proteobacteria	96	Gammaproteobacteria	86	Chromatiales	34	Chromatiaceae	34	Isochromatium	13

Table A1 continued

OTU	Major Group		Subgroup		Order		Family		Genus	
185	Firmicutes	54	Clostridia	35	Clostridiales	23	Peptococcaceae	20	Thermoterrabacterium	18
186	Proteobacteria	38	Deltaproteobacteria	27	Syntrophobacterales	10	Syntrophaceae	6	Desulfobacca	6
187	Proteobacteria	100	Alphaproteobacteria	100	Rhodospirillales	53	Acetobacteraceae	32	Paracraurococcus	13
188	Gemmatimonadetes	44	Gemmatimonadetes	44	Gemmatimonadales	44	Gemmatimonadaceae	44	Gemmatimonas	44
189	Proteobacteria	100	Betaproteobacteria	100	Burkholderiales	93	Oxalobacteraceae	82	Duganella	28
190	Firmicutes	41	Clostridia	31	Halanaerobiales	16	Halobacteroidaceae	9	Natroniella	8
191	Proteobacteria	100	Gammaproteobacteria	100	Xanthomonadales	90	Xanthomonadaceae	90	Rhodanobacter	75
192	Actinobacteria	100	Actinobacteria	100	Actinomycetales	100	Nocardioidaceae	52	Friedmanniella	25
193	Proteobacteria	86	Alphaproteobacteria	78	Rhizobiales	38	Rhodobiaceae	26	Parvibaculum	26
194	Actinobacteria	100	Rubrobacteridae	100	Rubrobacterales	100	Rubrobacteraceae	100	Conexibacter	76
195	Proteobacteria	100	Gammaproteobacteria	100	Legionellales	100	Coxiellaceae	100	Aquicella	100
196	Proteobacteria	96	Deltaproteobacteria	87	Myxococcales	79	Polyangiaceae	64	Chondromyces	64
197	Firmicutes	100	Bacilli	100	Bacillales	100	Bacillaceae	89	Bacillus	65
198	Proteobacteria	100	Alphaproteobacteria	100	Rhodospirillales	56	Acetobacteraceae	33	Stella	17
199	Proteobacteria	92	Deltaproteobacteria	61	Myxococcales	44	Polyangiaceae	25	Chondromyces	25
200	Proteobacteria	100	Deltaproteobacteria	100	Bdellovibrionales	100	Bdellovibrionaceae	100	Bdellovibrio	100
201	Bacteroidetes	38	Sphingobacteria	27	Sphingobacteriales	27	Flexibacteraceae	23	Arcicella	11
202	Proteobacteria	100	Alphaproteobacteria	99	Rhizobiales	98	Hyphomicrobiaceae	62	Angulomicrobium	43
203	Proteobacteria	100	Alphaproteobacteria	100	Rhizobiales	100	Hyphomicrobiaceae	54	Devosia	51
204	Proteobacteria	100	Gammaproteobacteria	100	Pseudomonadales	100	Pseudomonadaceae	100	Pseudomonas	80
205	Actinobacteria	100	Actinobacteria	100	Actinomycetales	100	Micrococcaceae	100	Arthrobacter	100
206	Actinobacteria	74	Actinobacteria	74	Sphaerobacterales	74	Sphaerobacteraceae	74	Sphaerobacter	74
207	Gemmatimonadetes	69	Gemmatimonadetes	69	Gemmatimonadales	69	Gemmatimonadaceae	69	Gemmatimonas	69
208	Proteobacteria	100	Betaproteobacteria	100	Methylophilales	100	Methylophilaceae	100	Methylobacillus	70
209	Proteobacteria	100	Gammaproteobacteria	99	Xanthomonadales	96	Xanthomonadaceae	96	Luteimonas	76
210	Proteobacteria	98	Alphaproteobacteria	97	Rhodospirillales	60	Acetobacteraceae	49	Stella	31
211	Proteobacteria	100	Alphaproteobacteria	98	Rhizobiales	87	Hyphomicrobiaceae	57	Angulomicrobium	22
212	Proteobacteria	100	Betaproteobacteria	100	Burkholderiales	77	Incertae sedis 5	55	Thiobacter	55
213	Chlamydiae	100	Chlamydiae	100	Chlamydiales	100	Parachlamydiaceae	77	Parachlamydia	76
214	Bacteroidetes	100	Sphingobacteria	100	Sphingobacteriales	100	Sphingobacteriaceae	99	Pedobacter	57
215	Chloroflexi	29	Anaerolineae	29	Anaerolinaeles	29	Anaerolinaeceea	29	Anaerolinea	29
216	Chlamydiae	99	Chlamydiae	99	Chlamydiales	99	Simkaniaceae	41	Rhabdochlamydia	35

Table A1 continued

OTU	Major Group		Subgroup		Order		Family		Genus	
217	Proteobacteria	100	Alphaproteobacteria	100	Caulobacterales	100	Caulobacteraceae	100	Caulobacter	94
218	Gemmatimonadetes	80	Gemmatimonadetes	80	Gemmatimonadales	80	Gemmatimonadaceae	80	Gemmatimonas	80
219	Proteobacteria	99	Alphaproteobacteria	90	Rhodospirillales	77	Acetobacteraceae	62	Stella	24
220	Chloroflexi	94	Anaerolineae	94	Anaerolinaeles	94	Anaerolinaeceea	94	Anaerolinea	94
221	Proteobacteria	89	Deltaproteobacteria	87	Myxococcales	78	Haliangiaceae	63	Haliangium	63
222	Bacteroidetes	95	Flavobacteria	77	Flavobacteriales	77	Flavobacteriaceae	61	Weeksella	25
223	Proteobacteria	100	Alphaproteobacteria	99	Rhodospirillales	64	Acetobacteraceae	43	Saccharibacter	14
224	Proteobacteria	100	Gammaproteobacteria	100	Xanthomonadales	55	Xanthomonadaceae	55	Fulvimonas	29
225	Proteobacteria	100	Alphaproteobacteria	99	Rhodospirillales	67	Rhodospirillaceae	38	Inquilinus	17
226	Proteobacteria	77	Deltaproteobacteria	33	Myxococcales	19	Cystobacteraceae	19	Anaeromyxobacter	19
227	Cyanobacteria	56	Cyanobacteria	56	Subsection 1	37	Family 1.1	37	Gloeocapsa	34
228	Proteobacteria	100	Betaproteobacteria	100	Burkholderiales	55	Incertae sedis 5	39	Thiobacter	37
229	Proteobacteria	50	Deltaproteobacteria	36	Syntrophobacterales	19	Syntrophobacteraceae	18	Desulforhabdus	10
230	Proteobacteria	100	Gammaproteobacteria	100	Legionellales	100	Legionellaceae	100	Legionella	100
231	Proteobacteria	100	Alphaproteobacteria	100	Caulobacterales	100	Caulobacteraceae	100	Phenylobacterium	98
232	Actinobacteria	100	Actinobacteria	100	Actinomycetales	100	Mycobacteriaceae	100	Mycobacterium	100
233	Proteobacteria	100	Betaproteobacteria	100	Burkholderiales	100	Burkholderiaceae	100	Pandoraea	100
234	Proteobacteria	100	Alphaproteobacteria	99	Rhizobiales	99	Hyphomicrobiaceae	65	Rhodoplanes	59
235	Proteobacteria	100	Alphaproteobacteria	100	Rhizobiales	100	Rhizobiaceae	92	Sinorhizobium	40
236	Proteobacteria	100	Betaproteobacteria	100	Burkholderiales	100	Burkholderiaceae	100	Burkholderia	100
237	Proteobacteria	97	Gammaproteobacteria	71	Legionellales	27	Coxiellaceae	27	Rickettsiella	25
238	Proteobacteria	100	Betaproteobacteria	100	Nitrosomonadales	100	Nitrosomonadaceae	100	Nitrosospira	100
239	Proteobacteria	100	Betaproteobacteria	100	Burkholderiales	94	Incertae sedis 5	46	Thiobacter	41
240	Proteobacteria	99	Gammaproteobacteria	96	Xanthomonadales	87	Xanthomonadaceae	87	Silanimonas	32
241	Chloroflexi	77	Anaerolineae	77	Anaerolinaeles	77	Anaerolinaeceea	77	Anaerolinea	77
242	Proteobacteria	49	Deltaproteobacteria	29	Myxococcales	13	Cystobacteraceae	13	Anaeromyxobacter	8
243	Proteobacteria	99	Alphaproteobacteria	99	Rhodospirillales	75	Acetobacteraceae	47	Stella	34
244	Proteobacteria	100	Betaproteobacteria	100	Methylophilales	89	Methylophilaceae	89	Methylophilus	89
245	Actinobacteria	100	Actinobacteria	100	Actinomycetales	100	Nocardioidaceae	45	Nocardioides	42
246	Proteobacteria	81	Deltaproteobacteria	63	Myxococcales	23	Cystobacteraceae	23	Anaeromyxobacter	23
247	Acidobacteria	43	Acidobacteria	43	Acidobacteriales	43	Acidobacteriaceae	43	Acidobacterium	43
248	Proteobacteria	100	Alphaproteobacteria	100	Rhodospirillales	100	Acetobacteraceae	100	Acidisphaera	29

Table A1 continued

OTU	Major Group		Subgroup		Order		Family		Genus	
249	Proteobacteria	98	Alphaproteobacteria	97	Rhizobiales	46	Brucellaceae	30	Mycoplana	30
250	Proteobacteria	100	Alphaproteobacteria	100	Rhodospirillales	100	Acetobacteraceae	82	Stella	80
251	Proteobacteria	99	Gammaproteobacteria	87	Chromatiales	56	Ectothiorhodospiraceae	28	Alkalilimnicola	17
252	Chloroflexi	49	Anaerolineae	48	Anaerolinaeles	48	Anaerolinaeceea	48	Anaerolinea	48
253	Bacteroidetes	94	Sphingobacteria	48	Sphingobacteriales	48	Flexibacteraceae	36	Microscilla	19
254	Firmicutes	100	Bacilli	100	Bacillales	98	Bacillaceae	44	Saccharococcus	9
255	Proteobacteria	100	Gammaproteobacteria	99	Legionellales	86	Coxiellaceae	86	Aquicella	69
256	Chloroflexi	97	Anaerolineae	97	Anaerolinaeles	97	Anaerolinaeceea	97	Anaerolinea	97
257	Proteobacteria	100	Betaproteobacteria	100	Burkholderiales	100	Burkholderiaceae	100	Burkholderia	100
258	Proteobacteria	98	Alphaproteobacteria	98	Rhodospirillales	64	Acetobacteraceae	63	Kozakia	26
259	Proteobacteria	92	Deltaproteobacteria	84	Myxococcales	74	Haliangiaceae	66	Haliangium	66
260	Proteobacteria	90	Alphaproteobacteria	89	Rhodospirillales	57	Acetobacteraceae	57	Kozakia	16
261	Proteobacteria	100	Alphaproteobacteria	100	Rhizobiales	93	Hyphomicrobiaceae	60	Rhodoplanes	40
262	Bacteroidetes	100	Sphingobacteria	100	Sphingobacteriales	100	Sphingobacteriaceae	97	Pedobacter	93
263	Proteobacteria	100	Betaproteobacteria	97	Burkholderiales	87	Burkholderiaceae	49	Burkholderia	46
264	Proteobacteria	100	Alphaproteobacteria	100	Rhodospirillales	100	Acetobacteraceae	100	Acetobacteria	53

Appendix 2. Equations used to calculate diversity indices.

The equations used to calculate the confidence intervals, where appropriate, are described in the DOTUR manual, (www.plantpath.wisc.edu/fac/joh/**DOTUR**/documentation.html)

i) Shannon diversity index (H')

$$H' = - \sum_{i=1}^{Sobs} \frac{S_i}{N} \ln \frac{S_i}{N}$$

Where:
S_i/N = the proportion of individuals found in the i^{th} OTU

ii) Chao Richness Estimator (S_{Chao1})

$$S_{Chao1} = S_{obs} + \frac{n_1 (n_1-1)}{2 (n_2 + 1)}$$ When $n_1 > 0$ and $n_2 \geq 0$ and when $n_1 = 0$ and $n_2 = 0$

$$S_{Chao1} = S_{obs} + \frac{n_1^2}{2n_2}$$ When $n_1 = 0$ and $n_2 \geq 0$

Where:
S_{Chao1} = Richness Estimate
S_{obs} = Observed number of species
n_1 = Number of OTUs with only one sequence
n_2 = Number of OTUs with only two sequences

iii) Ace Richness Estimator (S_{ACE})

$$S_{ACE} = S_{abund} + \frac{S_{rare}}{C_{ACE}} + \frac{n_1}{C_{ACE}} (\gamma^2_{ACE})$$

$$\gamma^2_{ACE} = \max \left[\frac{S_{rare}}{C_{ACE}} \frac{\sum_{i=1}^{10} (i (i-1) n_i)}{N_{rare} (N_{rare}-1)} - 1, 0 \right]$$

$$C_{ACE} = 1 - \frac{n_1}{N_{rare}}$$

$$N_{rare} = \sum_{i=1}^{10} i n_i$$

Where n_i = The number of OTUs with i individuals
S_{rare} = The number of OTUs with 10 or fewer individuals
S_{abund} = The number of OTUs with more than 10 individuals

iv) Shannon Evenness Index (E)

$$E = \frac{H'}{H_{max}} = \frac{H'}{\ln S}$$

Where H' = the Shannon diversity index
H_{max}= the maximum diversity possible
S = the number of OTUs recovered

v) Simpson's (D)

$$D = \frac{\sum_{i=1}^{Sobs} S_i (S_i-1)}{N (N-1)}$$

Where Si = the number of individuals in the ith OTU
N = the total number of individuals

vi) Berger-Parker Index (d)

$$d = \frac{N_{max}}{N}$$

Where N_{max} = the number of individuals in the most abundant OTU
N = the number of individuals