The Tannase Gene: Metaphylogenomics, Global Distribution and Presence in the Midgut Flora of the Forest Tent Caterpillar *Malacosoma disstria* Hübner

Michael Gasse

A Thesis

In

The Department

of

Biology

Presented in Partial Fulfillment of the Requirements

for the Degree of Master of Science (Biology) at

Concordia University

Montreal, Quebec, Canada

January 2014

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: Michael Gasse

Entitled: The Tannase Gene: Metaphylogenomics, Global Distribution and Presence in the Midgut Flora of the Forest Tent Caterpillar *Malacosoma disstria* Hübner.

and submitted in partial fulfillment of the requirements for the degree of

Master of Science (Biology)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final Examining Committee:

___Selvadurai Dayanandan__________Chair

___Ian Ferguson____________________Examiner

___David Walsh____________________Examiner

___Dan McLaughlin_________________External Examiner

___Emma Despland_________________Supervisor

Approved by ___Patrick Gulick_______________________________

Chair of Department or Graduate Program Director

___January 8th____2014 ___Joanne Locke______________________________

Dean of Faculty

ABSTRACT

The Tannase Gene: Metaphylogenomics, Global Distribution and Presence in the Midgut Flora of the Forest Tent Caterpillar *Malacosoma disstria* Hübner

Michael Gasse

Tannase enzymes hydrolyze tannins, a class of plant polyphenolics that defend against herbivory. Prior to the 1980's, most studies were focused on the tannase gene in fungi owing to the interest surrounding their industrial value. Since then a large number of bacterial tannase genes have also been discovered. I performed phylogenetic analysis on 110 fungal and bacterial tannase reference sequences in an effort to observe the relationships between fungal and bacterial tannase. The generated maximum likelihood tree shows eight strongly supported tannase clades, with a rift among fungal tannases, which either align with proteobacterial tannase or actinobacterial tannase. Metagenomes were used to assess the biogeographical distribution of the tannase clades, revealing that they may have environmental specificity.

An unpublished observation made by the Despland lab of Concordia University suggested that forest tent caterpillars (*Malacosoma disstria)* from western Canada are unable to survive on tannin-rich sugar maple foliage (*Acer saccharum*), whereas populations in eastern Canada develop on sugar maple. In Chapter 2 I examine this observation, and explore the possibility of tannase-expressing secondary symbionts in eastern *M. disstria.* For the first time, the microbiome of *M. disstria* has been described using 16S rRNA gene sequencing and is shown to contain several genera known for expressing the tannase gene.

Acknowledgements

Firstly, I would like to thank my family, lab mates and supervisor Dr. Emma Despland for their support and guidance. Special thanks to my colleague Dr. Jessica Ethier and co-supervisor Dr. David Walsh who both acted in a supervisory capacity during Dr. Emma Despland's maternity leave.

Secondly, I would like to thank Dr. Ian Ferguson for heroically stepping in as a committee member at the absolute last minute.

Lastly I would like to thank the Lagavulin distillery.

Table of Contents

Chapter 1: Metaphylogenomics and Global Distribution of the Tannase Gene

- 1.1 Introduction 1
- 1.2 Methods 6
    - 1.2.1 Reference Phylogeny 6
    - 1.2.2 Metagenomics 8
- 1.3 Results 9
    - 1.3.1 The T/FE Reference Phylogeny 9
    - 1.3.2 Functionally Characterized Tannase 13
    - 1.3.3 Udatha et al. (2011) Feruloyl Esterase Superfamilies 13
    - 1.3.4 Xenologs 17
    - 1.3.5 Global Biogeography of the T/FE Superfamily 17
    - 1.3.6 Phlorotannins 23
    - 1.3.7 Conserved Domains 23
    - 1.3.8 Conserved Cysteines in Tannase 26
- 1.4 Discussion 28

Chapter 2: Tannase Producing Symbionts in M. disstria

- 2.1 Introduction 39
- 2.2 Methods 49
    - 2.2.1 Insect Rearing 2011 49
    - 2.2.2 Insect Rearing 2012 50
    - 2.2.3 Caterpillar Dissection 51
    - 2.2.4 Culturing of Midgut Bacteria 51
    - 2.2.5 DNA Extraction and Bacterial 16S rRNA Amplification 52
    - 2.2.6 Sequencing 53
- 2.3 Results 54
    - 2.3.1 Insect Rearing 2011 54
    - 2.3.2 Abundance and Taxonomy of Culturable FTC Gut Bacteria 56
    - 2.3.3 16S rRNA Sequencing of Midgut Communities 57
- 2.4 Discussion 59

References 69

Chapter 1 - Metaphylogenomics and Global Distribution of the Tannase Gene

1.1 Introduction

Tannins are important plant polyphenols and are the fourth most plentiful biochemical component of plants after cellulose, hemicellulose and lignin (Swain, 1965). As a polyphenol, tannins consist of multiple phenol structural units. They are located in leaves, bark and wood and defend against herbivores as well as microbial and fungal pathogens (Lekha and Lonsane, 1997; Aguilar and Gutierrez-Sanchez, 2001). Tannins produce a bitter taste when ingested and in many cases have toxic effects, however the exact mechanism for toxicity depends on the herbivore. In vertebrate herbivores, the toxicity is based on the ability of tannins to bind to useful macromolecules like proteins, rendering them indigestible (Aguilar et al., 2007). For Lepidopteran herbivores, toxicity may instead be caused by the alkaline pH of the gut oxidizing tannins to form semiquinone radicals, quinines and other reactive oxygen species (Barbehenn and Constabel, 2011). In addition to tannins, another plant phenolic, ferulic acid, is also thought to be used in antiherbivore defense. It is found in cell walls and binds to useful polysaccharides rendering them unusable to herbivores (Bunzel et al., 2005).

Tannins are divided into four major groups; gallotannins, ellagitannins, complex tannins and condensed tannins. Gallotannins are typically gallic, digallic and chebulic acids bound to glucose through ester bonds, ellagitannins are made of ellagic acid bound to glucosides and complex tannins are ellagitannins bound to a flavonoid. Condensed tannins are constructed from solely flavonoids and are not easily hydrolysable (Ramirez-Coronel et al., 2004). Aguilar et al. (2007) described the basic structure of the four tannin groups (see Figure 1). The type of tannin used commercially is tannic acid. Tannic acid has as many as five digallates bound to glucose. It is used frequently as clarifying agents in alcoholic drinks and fruit juices (Lekha and Lonsane 1994; Belmares et al., 2004). Gallic acid, one of the breakdown products of tannase action, is a major compound in the synthesis of trimethroprim, an antibacterial drug (Sittig, 1988).

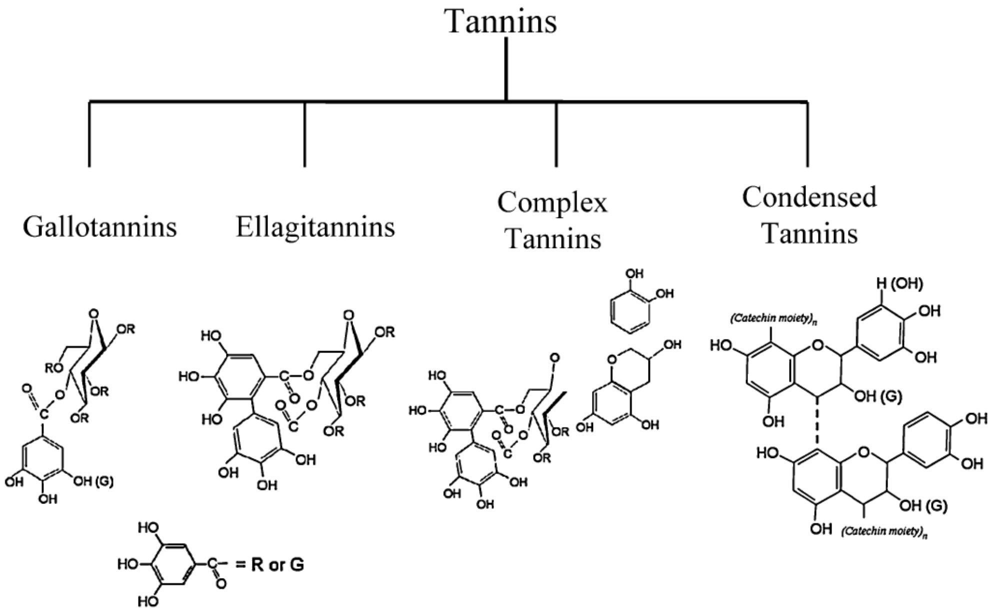


Figure 1 - The four tannin groups, as illustrated by Aguilar et al. (2007).

All but condensed tannins are hydrolysable by tannase, an enzyme belonging to the tannase/feruloyl esterase family (Aguilar, 2007). Tannase (E.C. 3.1.1.20) and feruloyl esterase (E.C. 3.1.1.73) belong to the same protein family (IPR011118 and Pfam PF07519), but have different functions. This enzyme family is expressed by certain bacteria and fungi, many of which are plant pathogens. The enzymes hydrolyze the ester bonds of hydrolysable tannins and feruloyl-polysaccharides, releasing the bound macromolecules that were previously indigestible. Tannase cleaves ester bonds between gallate and macromolecules (like carbohydrates) or complexes with multiple gallates (digallate). Ferulic acid, a hydroxycinnamic acid, is produced by plants and binds polysaccharides, creating the complex feruloyl polysaccharide. These complexes bind to lignin monomers that are active in cell wall development and help plants resist attack by microorganisms (Ishii, 1997). Once feruloyl esterases cleave the ester bond allowing the release of these polysaccharides, leftover ferulic acid may act as an antioxidant by reacting with free radicals like reactive oxygen species (Kampa et al., 2004).

Tannase/feruloyl esterase enzymes are expressed by a variety of aquatic as well as terrestrial microbes. Tannins inhibit microbial growth in aquatic environments through metal ion chelation (Hagerman, 2002) in addition to substrate deprivation. The tannase enzyme allows aquatic microorganisms to breakdown and absorb nutrients from dissolved organic matter more efficiently (Hagerman, 2002).

At the sequence level, tannase and feruloyl esterase are currently indistinguishable, causing confusion when annotating the genes without functional characterization. The majority of tannase sequences on record are often given both names but while both have similar function, they act on different plant polyphenols.

Traditionally, tannase and other esterases that act on similar substrates have been grouped under feruloyl esterases. Crepin et al. (2004) proposed 4 sub-classes for feruloyl esterases, types A through D, based on sequence data and the various substrates the enzymes use. Of these, tannase was designated as feruloyl esterase sub-class C for sharing sequence similarity with *Aspergillus niger* and *Talaromyces stipitatus* esterases that use methyl esters MFA (methyl ferulate), MSA (methyl sinapinate), MpCA (Methyl p-coumarate) and MCA (methyl caffeate) as substrates. Unfortunately, class C is characterized by only six fungal sequences, the majority of which are from *Aspergillus* species. As well, tannase is lumped into class C even though type C tannases were not tested on methyl gallate, digallate or glucogallin, three signature substrates of tannase activity. Without bacterial representatives and without functional characterization using gallate esters, it is unlikely that class C enzymes are truly representative of tannases. This idea is supported by the fact that the fungal and bacterial tannase sequences made available since 2004 do not fit neatly into the class structure defined by Crepin et al. (2004). In our reference phylogeny, the Crepin et al. (2004) class structure breaks apart, with members of each class ending up in very different tannase clades.

Following on the work by Crepin et al. (2004), Benoit et al. (2008) proposed 7 subfamilies rather than 4 sub-classes, but their analysis only included fungal sequences, creating a fungal bias for their particular classification method. Furthermore, their analysis only restructured the ABCD categories from Crepin at al. (2004) and did not account for new sequences. The system proposed by Crepin et al. (2004) was replaced recently by the phylogenetic analysis performed by Udatha et al. (2011).

Udatha et al. (2011) proposed 12 feruloyl esterase families including selected sequences described by Crepin et al. (2004) as feruloyl esterase types A through D. What they observed was a scattering of feruloyl esterases A through D across their 12 families. The naming system proposed by Crepin et al. (2004) had fallen apart with the use of new sequences calling for a new classification system. Udatha used machine learning of physicochemical properties and sequence derived descriptors. The 12 families proposed by Udatha et al. (2011) were based on the molecular weight of the protein, the number of amino acids in the sequence and the percentage of certain types of residues in the sequence (charged, aliphatic, aromatic, polar, non-polar). The downside is that these descriptors provide little information about the function of the protein or its evolutionary history. Proteins can share similar descriptors yet still be completely different, as the order of amino acids is very important for folding and function. As an example, it is not enough to know how many polar residues are in a sequence, what matters is precisely where those polar residues are. Creation of a family structure for a protein based purely on the length, weight and percent composition of certain amino acids is not sufficient to develop families that may share meaningful evolutionary information or function.

Banerjee et al. (2012) examined more physicochemical parameters such as isoelectric point and aliphatic index as well as studying the half-life of the protein from multiple species. In addition to identifying 5 highly conserved motifs, they constructed a neighbor-joining tree of 108 tannase sequences. They observed two major clades, one bacterial and the other mostly fungal with a few related bacterial sequences.

The objectives of this research were to provide an up-to-date view of tannase diversity through metaphylogenomic analysis of unique fungal and bacterial sequences that have been annotated as tannase. The inclusion of more recently discovered sequences was expected to reveal more accurate phylogenetic relationships than previous efforts. In addition, by BLASTing chosen sequences against environmental metagenomes it was expected to associate certain sequences with specific geographical locations and environmental conditions. Tannase clades in the reference phylogeny may potentially be defined by the collection of substrates they face in specific environments.

1.2 Methods

1.2.1 Reference phylogeny

Of all the protein-encoding sequences annotated as tannase/feruloyl esterase at NCBI, identical sequences and incomplete copies from the same species were removed leaving 110 sequences selected as the most complete and unique. The reference alignment is composed of 35 fungal, 74 bacterial and a single oomycete sequence. An effort was made to use a more equal representation of bacterial and fungal reference sequences, as previous studies either had a fungal bias or multiple sequences that may be considered repetitive at the generic level. Following the selection process, sequences were trimmed as many of the fungal sequences had extraneous non-tannase domains that would interfere with the alignment. These extra domains serve to cluster fungal sequences for reasons unrelated to the tannase gene. To this end, InterProScan was used to map out the domains of each sequence in order to remove all non-tannase domains. This has not been done in the past with this gene and may have interfered with alignments in previous work. It is currently unknown whether or not these extra domains are related to tannase in some way. The additional domains close to the tannase domain of fungi can be seen in Appendix A.

Trimmed sequences were aligned using MUSCLE in Mega 5 (Tamura et al., 2013) with the UPGMB clustering method for a maximum of eight iterations. Following the alignment, a maximum likelihood tree was generated using the Whelan and Goldman + Freq. (WAG +F) model with gamma distributed rate variation in invariant sites. WAG +F means the WAG matrix was used and +F refers to the fact that the stationary frequencies of the amino acids were estimated from the data set under analysis. The use of a maximum likelihood tree also provides more accurate topology than previous studies that used distance-based neighbor-joining trees. This improved topology also allows for the identification of new evolutionary clades as well as possible lateral gene transfer events. Sequences functionally characterized as tannase were included in a realignment of the reference phylogeny to assess possible functional identities for each group.

1.2.2 Metagenomics

A reference sequence representative of each of the 8 clades identified by the phylogentic analysis was queried against the Integrated Microbial Genomes with Microbiome Samples (img/m) database maintained by the US Department of Energy-supported Joint Genome Institute (JGI, URL:img.jgi.doe.gov). This was performed one clade at a time in an effort to discover novel tannase genes and attribute clades to particular microbiome types. Each of the multiple sequences of a clade was individually searched against the entire img/m database of 1233 environmental and 869 host-associated datasets (IMG version 4.0). The search was performed on December 7th, 2011 with an expect value cutoff of E^{-50}. The number of sequences matched and their corresponding microbiome type are listed in Table 1.

Table 1 - Microbiome hits obtained from img/m using reference clades as query.

Microbiome Type	# of Sequences
Air Filter	2
Chicken	1
Freshwater	69
Human	43
Insect	33
Marine	36
Panda	2
Plant	2
Plant Compost	25
Porifera	1
Rhizosphere	75
Soil	122
Thermal Spring	1
Wastewater	1

Given the large number of matches at low expect values, it was decided that it would be best to deal with only the most significant hits to form a clear idea of environmental tannase identity. To this end, the lower expect value cutoff of E^{-50} was used. Hits obtained from a searched clade were grouped together. Occasionally some microbiome sequences would appear as a significant result in more than one clade. To determine clade membership for microbiome hits, a list of the sequences in question were added to the alignment of the reference phylogeny. Following a reconstruction of the reference phylogeny, the microbiome sequences were located within one of the clades.

1.3 Results

1.3.1 The Tannase/Feruloyl Esterase Reference Phylogeny

The maximum likelihood tree representing the reference phylogeny shows 8 strongly supported tannase/feruloyl esterase clades (Figure 2). Expanded versions of the maximum likelihood tree with bootstrap value are included in Appendix B and C. Clades 1-4 are comprised of sequences mostly from proteobacteria (with a few exceptions) and are closely related to the fungi of clade 5. Clade 6 is composed of a small group of proteobacteria and a single actinobacteria. Clades 7 and 8 are made up mostly of fungi but also contain a significant number of actinobacteria.

Six genera have tannase sequences falling within a multitude of clades (Table 2). Clade 1 contains alpha, beta and gamma-proteobacteria with a significant number of marine genera such as *Shewanella* and *Marinomonas*. The clade also includes a single

Fusobacteria species, *Leptotrichia buccalis,* a member of the human oral flora (Ivanova et al., 2009). The tannase of *L. buccalis* is most closely related to that of several gamma proteobacteria of the bovine rumen; *Mannheimia succiniciproducens, Actinobacillus*

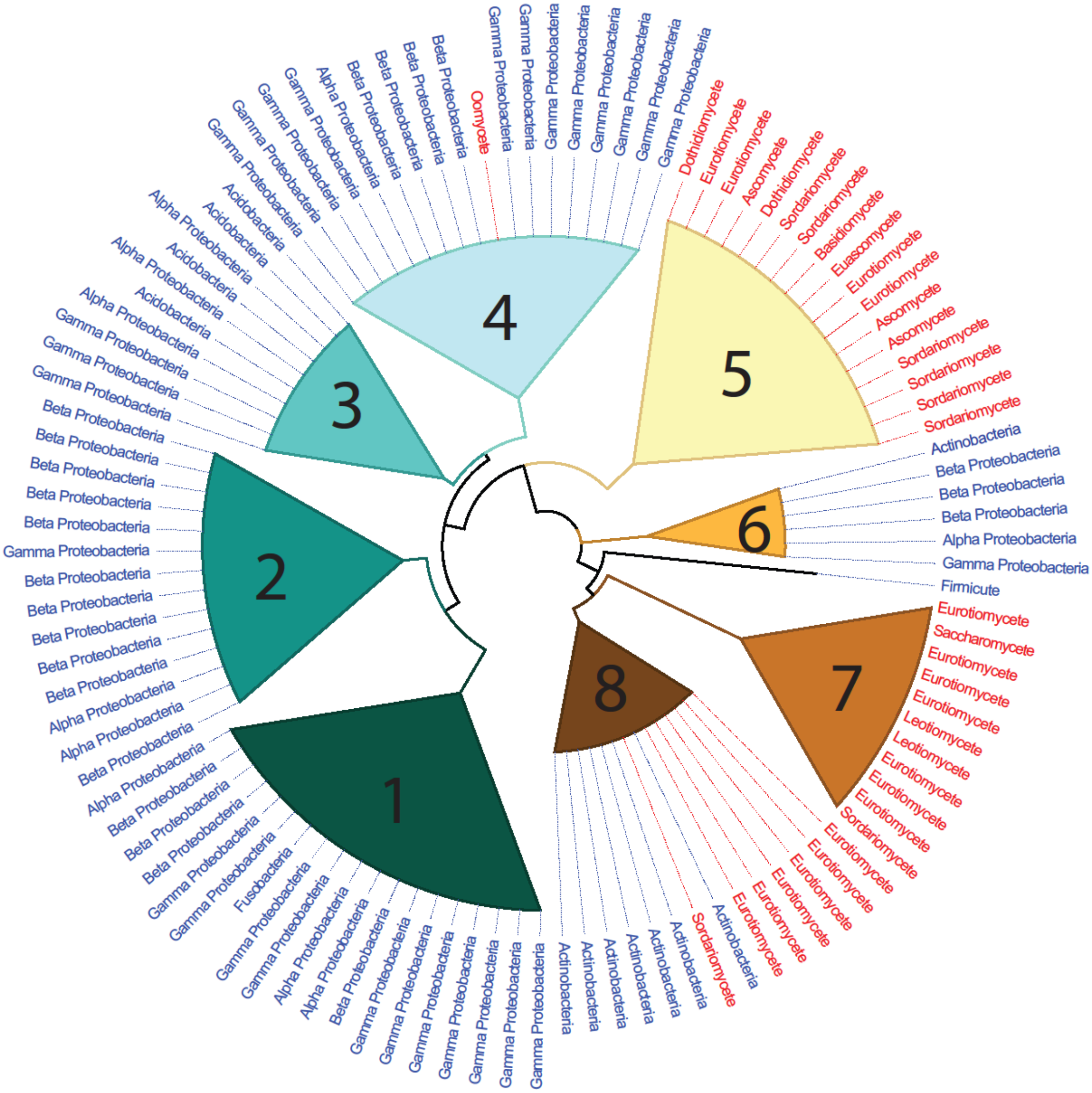


Figure 2 - A comprehensive reference phylogeny based on sequence analysis of 110 bacterial and fungal Tannase/Feruloyl Esterase (T/FE) proteins, showing 8 strongly supported clades.

succinogenes and *Mannheimia haemolytica. Mannheimia haemolytica* is known to cause bovine respiratory disease. The relationship between *L. buccalis, Mannheimia* and *Actinobacillus* tannases was explored using a reconstructed reference phylogeny including sequences originally not present in the reference alignment (Figure 3).

Table 2 - Genera having multiple clade membership in the reference phylogeny.

Organism	Class	Tannase Gene ID	Clade
Agrobacterium vitis S4	Alpha Proteobacteria	222148418	1
Agrobacterium radiobacter K84	Alpha Proteobacteria	222102103	6
Variovorax paradoxus S110	Beta Proteobacteria	239813100	6
Variovorax paradoxus ESP	Beta Proteobacteria	319796028	2
Frankia sp. EUN1f	Actinobacteria	288920905	8
Frankia sp. Eul1c	Actinobacteria	312197890	6
Penicillium chrysogenum	Eurotiomycete	255955963	7
Penicillium marneffei	Euascomycete	212538837	5
Burkholderia sp. CCGE1003	Beta Proteobacteria	307726323	2
Burkholderia sp. CCGE1001	Beta Proteobacteria	323529458	2
Burkholderia phymatum STM815	Beta Proteobacteria	186474898	2
Burkholderia phytofirmans PsJN	Beta Proteobacteria	187920718	2
Burkholderia sp. CCGE1002	Beta Proteobacteria	295699013	2
Burkholderia graminis C4D1M	Beta Proteobacteria	170693793	2
Burkholderia xenovorans LB400	Beta Proteobacteria	91781293	2
Burkholderia multivorans ATCC 17616	Beta Proteobacteria	161523057	4
Burkholderia multivorans CGD2M	Beta Proteobacteria	221198972	4
Burkholderia sp. 383	Beta Proteobacteria	78060680	4
Burkholderia sp. H160	Beta Proteobacteria	209516883	4
Burkholderia cenocepacia AU 1054	Beta Proteobacteria	107022853	6
Burkholderia cenocepacia MC03	Beta Proteobacteria	170734638	6
Aspergillus terreus NIH2624	Eurotiomycete	115388175	5
Aspergillus oryzae RIB40	Eurotiomycete	317147761	5
Aspergillus clavatus NRRL 1	Eurotiomycete	121699582	5
Aspergillus nidulans FGSC A4	Eurotiomycete	67904432	7
Aspergillus fumigatus Af293	Eurotiomycete	70982011	7
Aspergillus niger CBS 513.88	Eurotiomycete	145240885	7
Aspergillus oryzae RIB40	Eurotiomycete	169772335	7
Aspergillus flavus NRRL3357	Eurotiomycete	238490238	7
Aspergillus niger CBS 513.88	Eurotiomycete	317037506	8

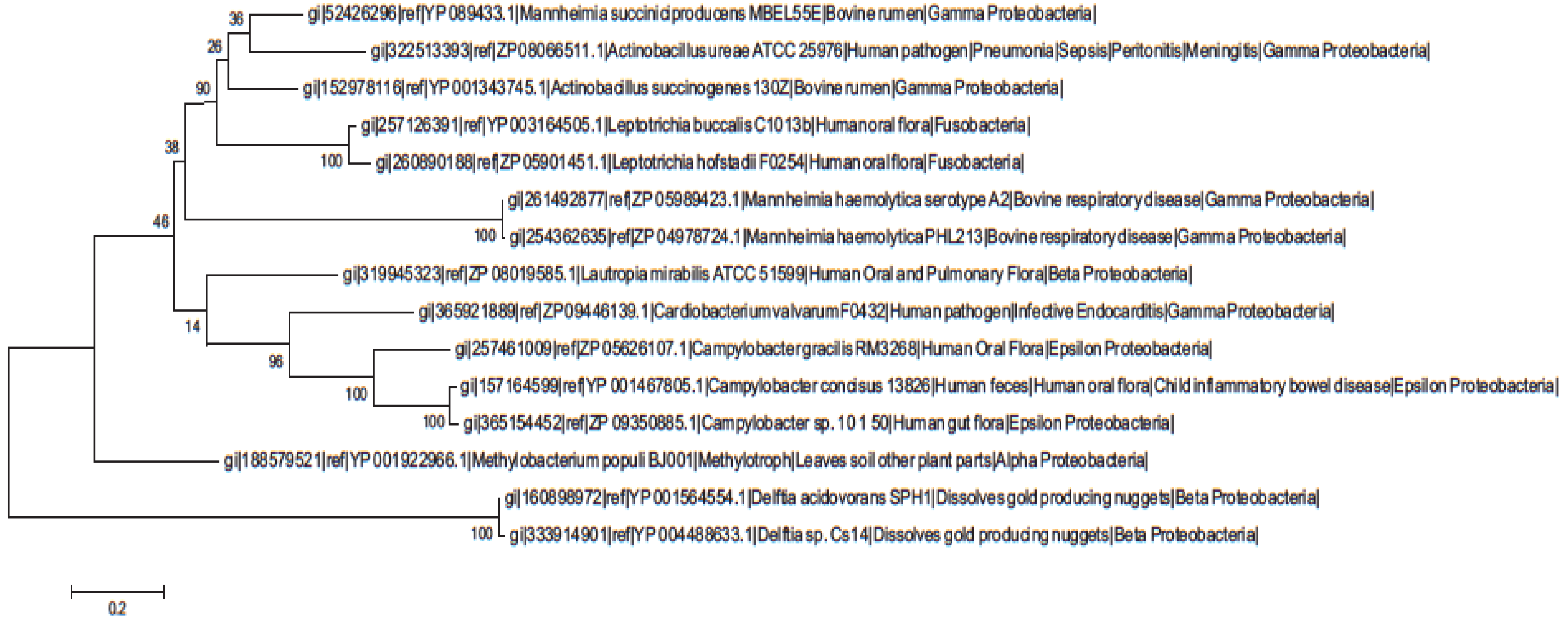


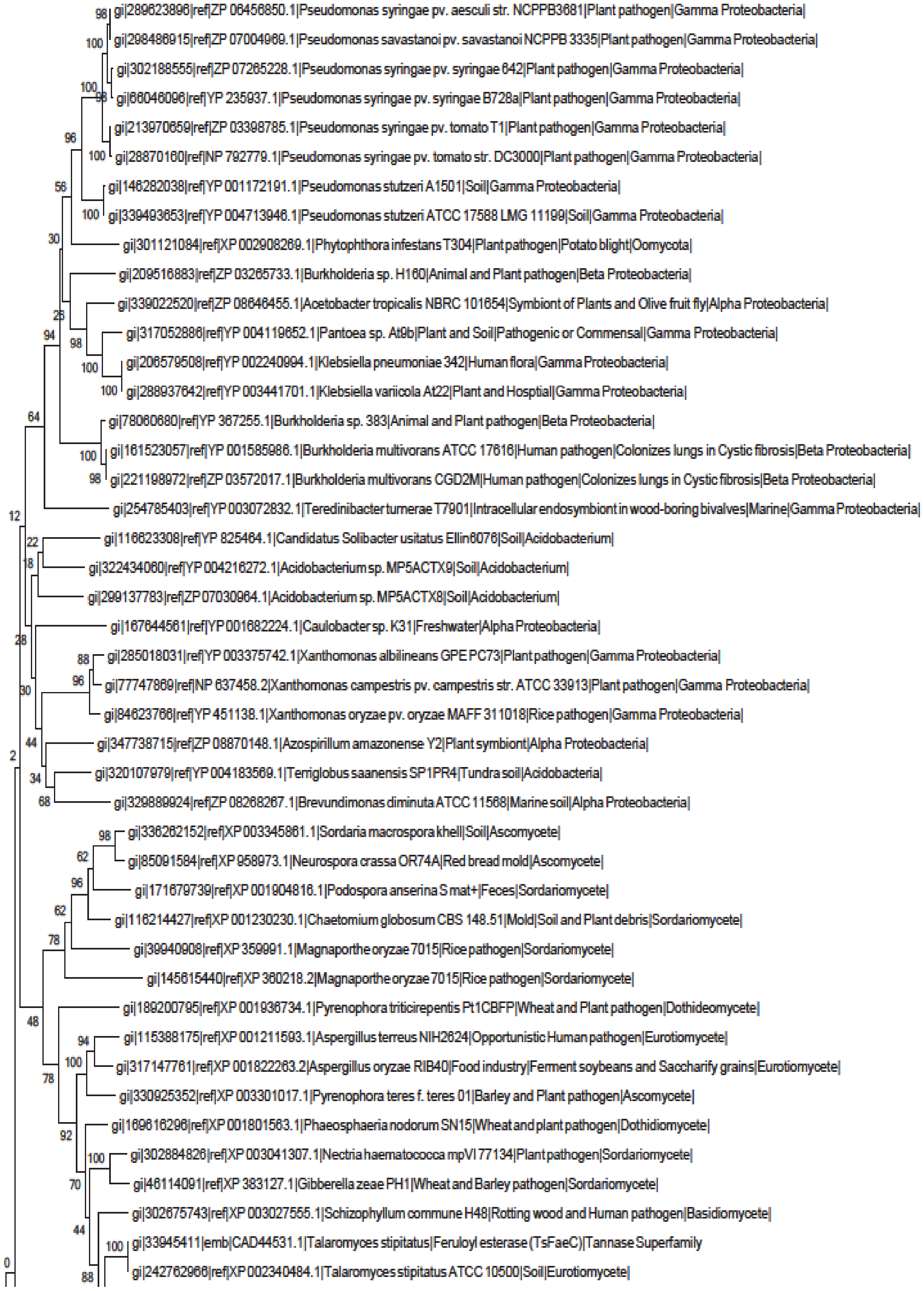
Figure 3 - ML tree of species with tannase most closely related to tannase of *L. buccalis.* The tree shows several different mammalian pathogens with highly related tannase sequences.

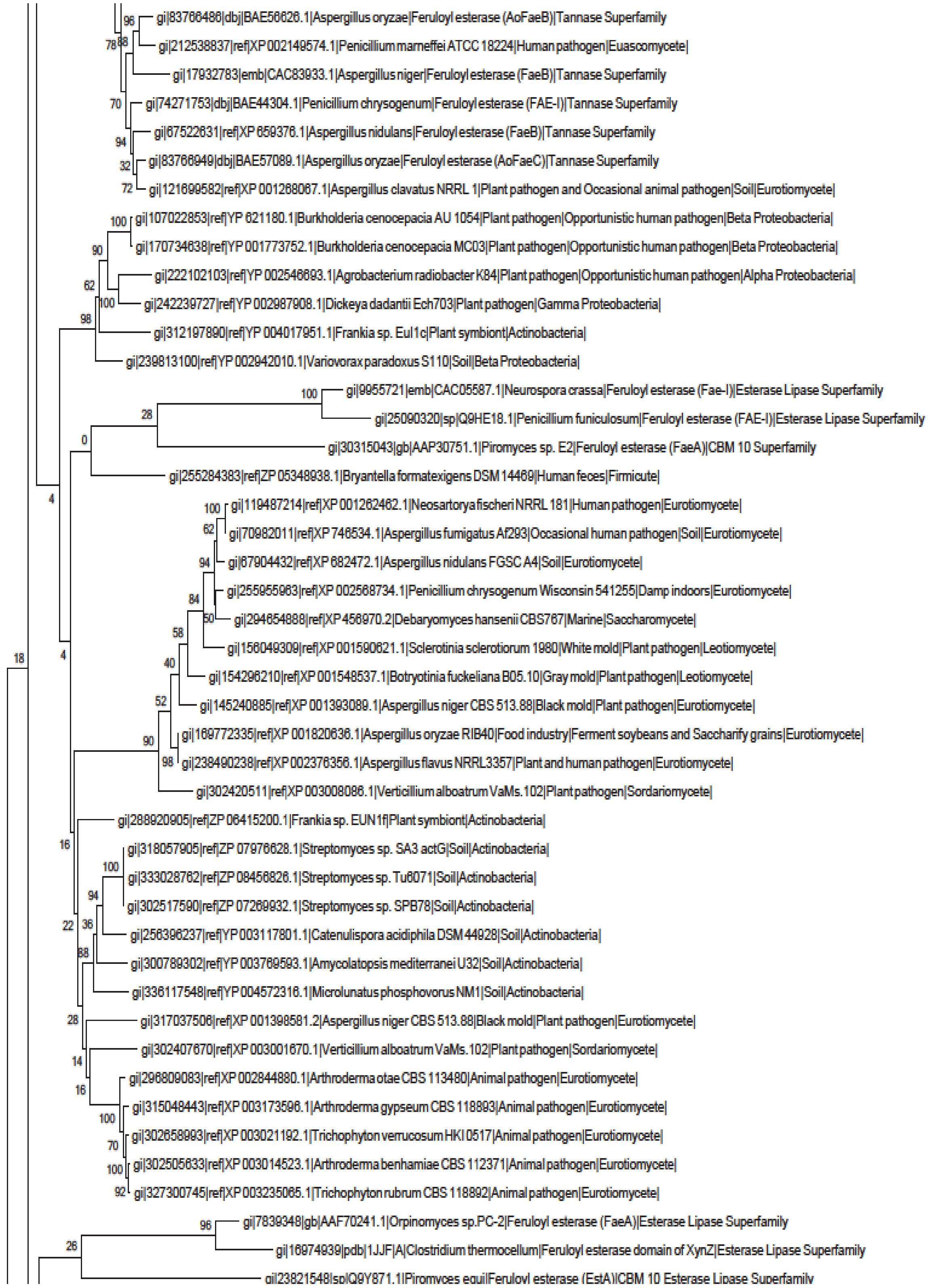
1.3.2 Functionally Characterized Tannase

Several sequences that have been biochemically validated as functional tannases were included in a reconstruction of the reference phylogeny in an attempt to attribute functions to some of the eight clades. *Streptococcus gallolyticus, Lactobacillus plantarum* and *Staphylococcus lugdunensis* all align with clade 4, while *Aspergillus awamori, Blastobotrys adeninivorans* and the two *Aspergillus oryzae* sequences align with clade 7.

1.3.3 Udatha et al. (2011) Feruloyl Esterase Superfamilies

The distinction between available tannases and feruloyl esterases is not clear and the enzyme family is poorly characterized (Udatha et al., 2011). Though many sequences in the reference phylogeny are labelled as one or the other, some classified as feruloyl esterase show higher sequence similarity to tannase-labelled sequences than other feruloyl esterases. As these sequence alignments are based on highly conserved domains that are potentially essential to the function of the proteins, it is likely that many of these sequences have been named incorrectly. To test the most recent classification system for tannase, a reconstruction of the reference phylogeny was made to include representatives of the feruloyl esterase superfamilies, as described by Udatha et al. (2011) (Figure 4). Members of the Lipase and Esterase/Lipase Superfamily were scattered throughout the tree suggesting no close relationship between the Superfamily and the clades in the reference phylogeny, however all of the Tannase Superfamily representatives align best with fungal group 5. Several sequences have been functionally characterized as tannase in *Klebsiella* and *Pseudomonas* (Belur and Mugeraya, 2011), of


98
gi|289623896|ref|ZP 06456850.1|Pseudomonas syringae pv. aesculi str. NCPPB3681|Plant pathogen|Gamma Proteobacteria|
100
gi|298486915|ref|ZP 07004969.1|Pseudomonas savastanoi pv. savastanoi NCPPB 3335|Plant pathogen|Gamma Proteobacteria|
gi|302188555|ref|ZP 07265228.1|Pseudomonas syringae pv. syringae 642|Plant pathogen|Gamma Proteobacteria|
100
98
gi|66046096|ref|YP 235937.1|Pseudomonas syringae pv. syringae B728a|Plant pathogen|Gamma Proteobacteria|
96
gi|213970659|ref|ZP 03398785.1|Pseudomonas syringae pv. tomato T1|Plant pathogen|Gamma Proteobacteria|
100
gi|28870160|ref|NP 792779.1|Pseudomonas syringae pv. tomato str. DC3000|Plant pathogen|Gamma Proteobacteria|
56
gi|146282038|ref|YP 001172191.1|Pseudomonas stutzeri A1501|Soil|Gamma Proteobacteria|
100
gi|339493653|ref|YP 004713946.1|Pseudomonas stutzeri ATCC 17588 LMG 11199|Soil|Gamma Proteobacteria|
30
gi|301121084|ref|XP 002908269.1|Phytophthora infestans T304|Plant pathogen|Potato blight|Oomycota|
gi|209516883|ref|ZP 03265733.1|Burkholderia sp. H160|Animal and Plant pathogen|Beta Proteobacteria|
gi|339022520|ref|ZP 08646455.1|Acetobacter tropicalis NBRC 101654|Symbiont of Plants and Olive fruit fly|Alpha Proteobacteria|
26
94
98
gi|317052886|ref|YP 004119652.1|Pantoea sp. At9b|Plant and Soil|Pathogenic or Commensal|Gamma Proteobacteria|
100
gi|206579508|ref|YP 002240994.1|Klebsiella pneumoniae 342|Human flora|Gamma Proteobacteria|
100
gi|288937642|ref|YP 003441701.1|Klebsiella variicola At22|Plant and Hosptial|Gamma Proteobacteria|
64
gi|78060680|ref|YP 367255.1|Burkholderia sp. 383|Animal and Plant pathogen|Beta Proteobacteria|
100
gi|161523057|ref|YP 001585986.1|Burkholderia multivorans ATCC 17616|Human pathogen|Colonizes lungs in Cystic fibrosis|Beta Proteobacteria|
98
gi|221198972|ref|ZP 03572017.1|Burkholderia multivorans CGD2M|Human pathogen|Colonizes lungs in Cystic fibrosis|Beta Proteobacteria|
gi|254785403|ref|YP 003072832.1|Teredinibacter turnerae T7901|Intracellular endosymbiont in wood-boring bivalves|Marine|Gamma Proteobacteria|
12
22
gi|116623308|ref|YP 825464.1|Candidatus Solibacter usitatus Ellin6076|Soil|Acidobacterium|
18
gi|322434060|ref|YP 004216272.1|Acidobacterium sp. MP5ACTX9|Soil|Acidobacterium|
gi|299137783|ref|ZP 07030964.1|Acidobacterium sp. MP5ACTX8|Soil|Acidobacterium|
gi|167644561|ref|YP 001682224.1|Caulobacter sp. K31|Freshwater|Alpha Proteobacteria|
28
88
gi|285018031|ref|YP 003375742.1|Xanthomonas albilineans GPE PC73|Plant pathogen|Gamma Proteobacteria|
96
gi|77747869|ref|NP 637458.2|Xanthomonas campestris pv. campestris str. ATCC 33913|Plant pathogen|Gamma Proteobacteria|
30
gi|84623766|ref|YP 451138.1|Xanthomonas oryzae pv. oryzae MAFF 311018|Rice pathogen|Gamma Proteobacteria|
44
gi|347738715|ref|ZP 08870148.1|Azospirillum amazonense Y2|Plant symbiont|Alpha Proteobacteria|
2
34
gi|320107979|ref|YP 004183569.1|Terriglobus saanensis SP1PR4|Tundra soil|Acidobacteria|
68
gi|329889924|ref|ZP 08268267.1|Brevundimonas diminuta ATCC 11568|Marine soil|Alpha Proteobacteria|
98
gi|336262152|ref|XP 003345861.1|Sordaria macrospora khell|Soil|Ascomycete|
62
gi|85091584|ref|XP 958973.1|Neurospora crassa OR74A|Red bread mold|Ascomycete|
96
gi|171679739|ref|XP 001904816.1|Podospora anserina S mat+|Feces|Sordariomycete|
62
gi|116214427|ref|XP 001230230.1|Chaetomium globosum CBS 148.51|Mold|Soil and Plant debris|Sordariomycete|
78
gi|39940908|ref|XP 359991.1|Magnaporthe oryzae 7015|Rice pathogen|Sordariomycete|
gi|145615440|ref|XP 360218.2|Magnaporthe oryzae 7015|Rice pathogen|Sordariomycete|
gi|189200795|ref|XP 001936734.1|Pyrenophora triticirepentis Pt1CBFP|Wheat and Plant pathogen|Dothideomycete|
48
94
gi|115388175|ref|XP 001211593.1|Aspergillus terreus NIH2624|Opportunistic Human pathogen|Eurotiomycete|
100
gi|317147761|ref|XP 001822263.2|Aspergillus oryzae RIB40|Food industry|Ferment soybeans and Saccharify grains|Eurotiomycete|
78
gi|330925352|ref|XP 003301017.1|Pyrenophora teres f. teres 01|Barley and Plant pathogen|Ascomycete|
gi|169616296|ref|XP 001801563.1|Phaeosphaeria nodorum SN15|Wheat and plant pathogen|Dothidiomycete|
92
100
gi|302884826|ref|XP 003041307.1|Nectria haematococca mpVI 77134|Plant pathogen|Sordariomycete|
70
gi|46114091|ref|XP 383127.1|Gibberella zeae PH1|Wheat and Barley pathogen|Sordariomycete|
gi|302675743|ref|XP 003027555.1|Schizophyllum commune H48|Rotting wood and Human pathogen|Basidiomycete|
44
100
gi|33945411|emb|CAD44531.1|Talaromyces stipitatus|Feruloyl esterase (TsFaeC)|Tannase Superfamily
0
88
gi|242762966|ref|XP 002340484.1|Talaromyces stipitatus ATCC 10500|Soil|Eurotiomycete|


96
gi|83766486|dbj|BAE56626.1|Aspergillus oryzae|Feruloyl esterase (AoFaeB)|Tannase Superfamily
78
88
gi|212538837|ref|XP 002149574.1|Penicillium marneffei ATCC 18224|Human pathogen|Euascomycete|
gi|17932783|emb|CAC83933.1|Aspergillus niger|Feruloyl esterase (FaeB)|Tannase Superfamily
70
gi|74271753|dbj|BAE44304.1|Penicillium chrysogenum|Feruloyl esterase (FAE-I)|Tannase Superfamily
gi|67522631|ref|XP 659376.1|Aspergillus nidulans|Feruloyl esterase (FaeB)|Tannase Superfamily
94
32
gi|83766949|dbj|BAE57089.1|Aspergillus oryzae|Feruloyl esterase (AoFaeC)|Tannase Superfamily
72
gi|121699582|ref|XP 001268067.1|Aspergillus clavatus NRRL 1|Plant pathogen and Occasional animal pathogen|Soil|Eurotiomycete|
100
gi|107022853|ref|YP 621180.1|Burkholderia cenocepacia AU 1054|Plant pathogen|Opportunistic human pathogen|Beta Proteobacteria|
90
gi|170734638|ref|YP 001773752.1|Burkholderia cenocepacia MC03|Plant pathogen|Opportunistic human pathogen|Beta Proteobacteria|
gi|222102103|ref|YP 002546693.1|Agrobacterium radiobacter K84|Plant pathogen|Opportunistic human pathogen|Alpha Proteobacteria|
62
100
gi|242239727|ref|YP 002987908.1|Dickeya dadantii Ech703|Plant pathogen|Gamma Proteobacteria|
98
gi|312197890|ref|YP 004017951.1|Frankia sp. EuI1c|Plant symbiont|Actinobacteria|
gi|239813100|ref|YP 002942010.1|Variovorax paradoxus S110|Soil|Beta Proteobacteria|
100
gi|9955721|emb|CAC05587.1|Neurospora crassa|Feruloyl esterase (Fae-I)|Esterase Lipase Superfamily
28
gi|25090320|sp|Q9HE18.1|Penicillium funiculosum|Feruloyl esterase (FAE-I)|Esterase Lipase Superfamily
0
gi|30315043|gb|AAP30751.1|Piromyces sp. E2|Feruloyl esterase (FaeA)|CBM 10 Superfamily
gi|255284383|ref|ZP 05348938.1|Bryantella formatexigens DSM 14469|Human feces|Firmicute|
4
100
gi|119487214|ref|XP 001262462.1|Neosartorya fischeri NRRL 181|Human pathogen|Eurotiomycete|
62
gi|70982011|ref|XP 746534.1|Aspergillus fumigatus Af293|Occasional human pathogen|Soil|Eurotiomycete|
94
gi|67904432|ref|XP 682472.1|Aspergillus nidulans FGSC A4|Soil|Eurotiomycete|
gi|255955963|ref|XP 002568734.1|Penicillium chrysogenum Wisconsin 541255|Damp indoors|Eurotiomycete|
84
50
gi|294654888|ref|XP 456970.2|Debaryomyces hansenii CBS767|Marine|Saccharomycete|
58
gi|156049309|ref|XP 001590621.1|Sclerotinia sclerotiorum 1980|White mold|Plant pathogen|Leotiomycete|
18
4
40
gi|154296210|ref|XP 001548537.1|Botryotinia fuckeliana B05.10|Gray mold|Plant pathogen|Leotiomycete|
52
gi|145240885|ref|XP 001393089.1|Aspergillus niger CBS 513.88|Black mold|Plant pathogen|Eurotiomycete|
gi|169772335|ref|XP 001820636.1|Aspergillus oryzae RIB40|Food industry|Ferment soybeans and Saccharify grains|Eurotiomycete|
90
98
gi|238490238|ref|XP 002376356.1|Aspergillus flavus NRRL3357|Plant and human pathogen|Eurotiomycete|
gi|302420511|ref|XP 003008086.1|Verticillium alboatrum VaMs.102|Plant pathogen|Sordariomycete|
gi|288920905|ref|ZP 06415200.1|Frankia sp. EUN1f|Plant symbiont|Actinobacteria|
16
gi|318057905|ref|ZP 07976628.1|Streptomyces sp. SA3 actG|Soil|Actinobacteria|
100
gi|333028762|ref|ZP 08456826.1|Streptomyces sp. Tu6071|Soil|Actinobacteria|
94
gi|302517590|ref|ZP 07269932.1|Streptomyces sp. SPB78|Soil|Actinobacteria|
22
36
gi|256396237|ref|YP 003117801.1|Catenulispora acidiphila DSM 44928|Soil|Actinobacteria|
88
gi|300789302|ref|YP 003769593.1|Amycolatopsis mediterranei U32|Soil|Actinobacteria|
gi|336117548|ref|YP 004572316.1|Microlunatus phosphovorus NM1|Soil|Actinobacteria|
28
gi|317037506|ref|XP 001398581.2|Aspergillus niger CBS 513.88|Black mold|Plant pathogen|Eurotiomycete|
gi|302407670|ref|XP 003001670.1|Verticillium alboatrum VaMs.102|Plant pathogen|Sordariomycete|
14
gi|296809083|ref|XP 002844880.1|Arthroderma otae CBS 113480|Animal pathogen|Eurotiomycete|
16
gi|315048443|ref|XP 003173596.1|Arthroderma gypseum CBS 118893|Animal pathogen|Eurotiomycete|
100
gi|302658993|ref|XP 003021192.1|Trichophyton verrucosum HKI 0517|Animal pathogen|Eurotiomycete|
70
100
gi|302505633|ref|XP 003014523.1|Arthroderma benhamiae CBS 112371|Animal pathogen|Eurotiomycete|
92
gi|327300745|ref|XP 003235065.1|Trichophyton rubrum CBS 118892|Animal pathogen|Eurotiomycete|
96
gi|7839348|gb|AAF70241.1|Orpinomyces sp.PC-2|Feruloyl esterase (FaeA)|Esterase Lipase Superfamily
26
gi|16974939|pdb|1JJF|A|Clostridium thermocellum|Feruloyl esterase domain of XynZ|Esterase Lipase Superfamily
gi|23821548|sp|Q9Y871.1|Piromyces equi|Feruloyl esterase (EstA)|CBM 10 Esterase Lipase Superfamily

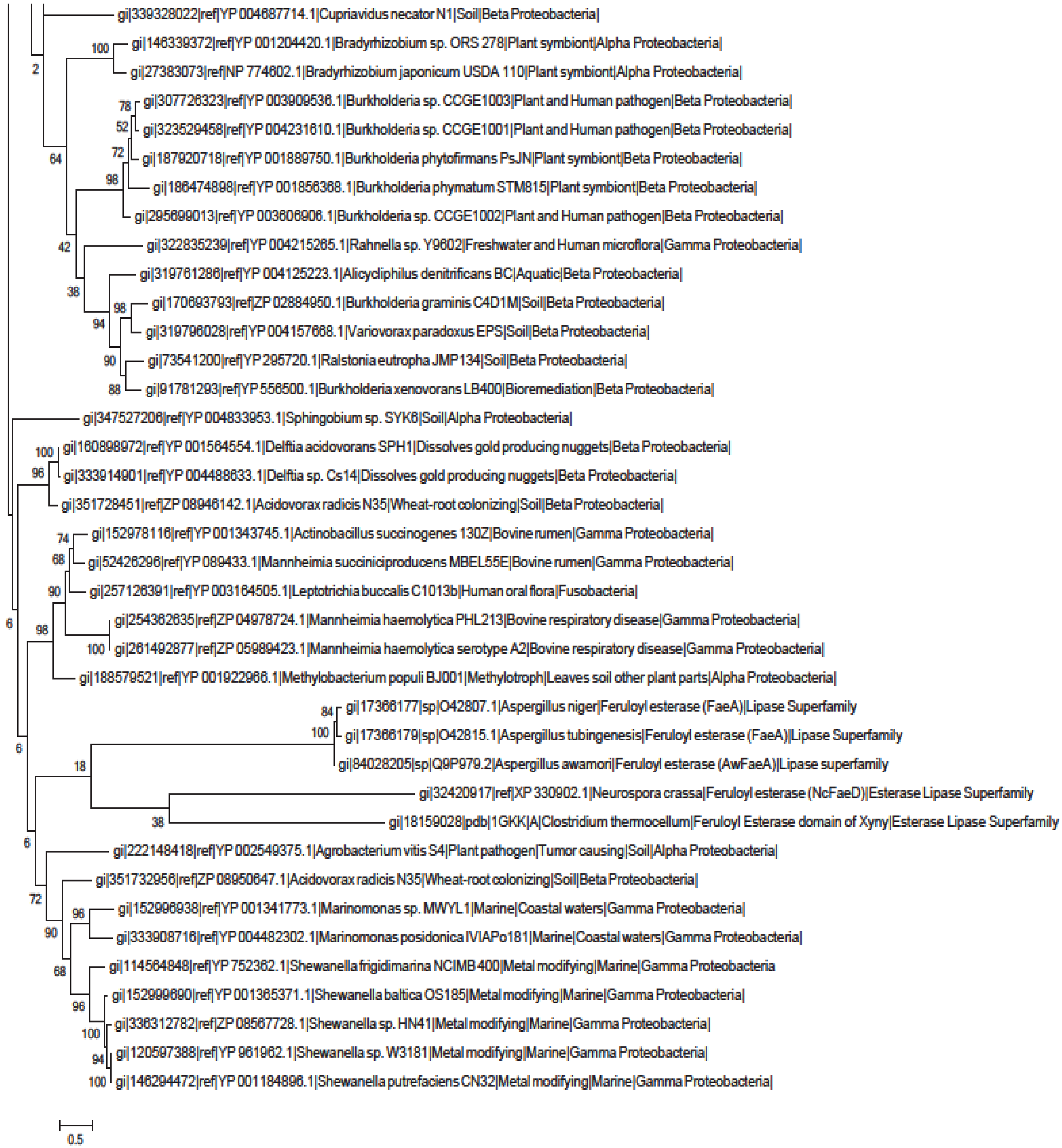


Figure 4 - Reference phylogeny with the addition of representatives from Udatha et al. (2011) FE superfamilies. The reconstruction of the reference phylogeny displays the fungal bias for the tannase superfamily in Udatha et al. (2011). The fungal sequences that make up the tannase superfamily all align with other fungi rather than functionally characterized bacterial tannases.

which 2 and 8 species respectively are included in the original reference phylogeny. Despite an Udatha et al. (2011) superfamily existing for tannase, Figure 4 shows *Klebsiella* and *Pseudomonas* to be relatively distant from the Udatha et al. (2011) Tannase Superfamily.

1.3.4 Xenologs

Xenologs are homologous sequences resulting from horizontal gene transfer. An example of this may be an organism with two different copies of a gene, each with different ancestry. Potential tannase xenologs are shown in the reference phylogeny within the rift between fungal tannases. *Aspergillus oryzae RIB40* and *Verticillium alboatrum VaMs. 102* are two fungal species for which there are different tannase sequences available in multiple reference genomes for their respective species. Some genomes contain two instances of the tannase gene. Table 3 shows several instances of multiple tannase gene copies within the same species. These gene copies are not identical however, as some align with proteobacterial clades 3 and 5, while the others align with the more actinobacterial clades 7 and 8.

1.3.5 Global biogeography of the T/FE superfamily

In an effort to gain understanding of how tannase and its identified clades are distributed biogeographically, each tannase clade was searched against the environmental microbiomes at the Integrated Microbiome Samples (img/m) database maintained by the US Department of Energy-supported Joint Genome Institute (JGI,

Table 3 - Tannase Xenologs in *A. oryzae* and *V. alboatrum*. Multiple copies of tannase from the same species that align to different bacterial tannase clades suggesting a different origin for each copy in fungi.

Species	Tannase Protein ID's	Reference Genome	Clade(s)
Aspergillus oryzae RIB40 (1)	XP_001822863.1	NW_001884671.1	8
Aspergillus oryzae RIB40 (2)	XP_001819091.1 and XP_001818628.2	NW_001884663.1	5 (both)
Aspergillus oryzae RIB40 (3)	XP_001821143.2 and XP_001820636.1	NW_001884666.1	7, 8
Aspergillus oryzae RIB40 (4)	XP_001827335.2 and XP_001827546.2	NW_001884682.1	5,7
Aspergillus oryzae RIB40 (5)	XP_001821840.1	NW_001884668.1	8
Aspergillus oryzae RIB40 (6)	XP_001822263.2	NW_001884670.1	5
Aspergillus oryzae RIB40 (7)	XP_001826830.2 and XP_001826685.1	NW_001884681.1	3, 7
Verticillium alboatrum VaMs. 102 (1)	XP_003008086.1	NW_003315037.1	7
Verticillium alboatrum VaMs. 102 (2)	XP_003007189.1 and XP_003007040.1	NW_003315036.1	5, 8
Verticillium alboatrum VaMs. 102 (3)	XP_003004044.1	NW_003315033.1	5
Verticillium alboatrum VaMs. 102 (4)	XP_003001670.1	NW_003315028.1	8
Verticillium alboatrum VaMs. 102 (5)	XP_003000281.1	NW_003315024.1	5

URL:img.jgi.doe.gov). The resulting analysis shows that inclusion in a particular tannase clade seems to be driven by environmental factors (Figure 5). Although there is some overlap regarding microbiome types, for the most part, each clade in the reference phylogeny is seen to dominate a certain microbiome. The majority of tannase sequences from the human microbiome samples fall within clade 1, with the rest belonging to clades 2, 3 and 4. With the human oral bacteria *L. buccalis* as a member of clade 1 it is interesting to see that it aligns with the large number of human oral microbiome samples that make up the majority of human microbiome samples.

The insect microbiome samples align with several tannase clades, however, the majority of these belong to species in clades 2 and 4, albeit for different reasons. Clade 2 is

Figure 5 - Donut plots showing the proportion of different microbiome sequences that align with each tannase clade. Marine aphotic zones are 1500 meters below sea level. The insect donut plot is further broken down into donut plots for the Coleoptera and Hymenoptera Order, for which there were 17 and 14 microbiomes respectively. The

Coleoptera microbiomes consisted of 4 different species of xylophages and the Hymenopteran microbiomes are based on 5 different species (honeybees and 4 leaf-cutter ants). Further details regarding microbiome matches can be found in Appendix D.

associated with 59% of the sequences associated with the order Coleoptera, while clade 4 is associated with 57% of the sequences associated with the order Hymenoptera.

Both rhizosphere and soil microbiomes are dominated by clade 3 and 2. Clade 3 is also the only clade to contain acidobacteria, a relatively new bacterial phylum that was previously resistant to cultivation. Acidobacteria are widespread in soil and rhizosphere environments, most notably in arctic and boreal zones (Rawat et al., 2012) but it appears that acidobacterial tannase or something extremely similar is also strongly represented in marine environments based on these alignments. In all, 59% and 20% of the marine aphotic and photic zones respectively align with tannase clade 3.

The marine aphotic microbiome is mostly shared by fungal clades 5, 7 and 8. This is misleading however, as the only marine aphotic zone microbiome samples available that align closely to the reference phylogeny come from the studies on the BP oil spill in the Gulf of Mexico. It is well known that post-spill, there has been selection for a specific group of fungi with the ability to degrade hydrocarbons (Bik et al., 2012). In addition to the presence of hydrocarbon-degrading enzymes in the metatranscriptome of post-spill fungi in the Gulf of Mexico, it appears that there is also tannase being actively produced by the same fungi. Overall, microbiome samples aligning to clades 5-8 are not very rare (Figure 6).

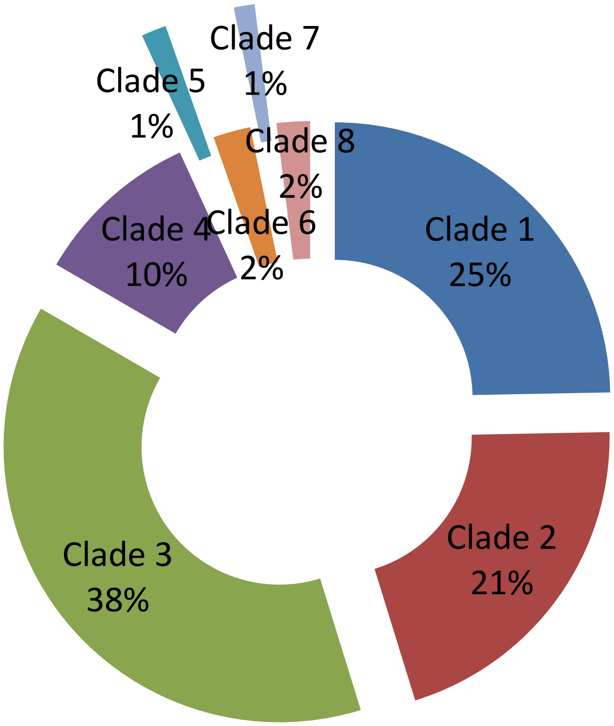


Figure 6 - Distribution of microbiome samples across clades.

Interestingly, the tannase within the microbiomes of honey bees in Tucson, Arizona does not fit completely in any of the tannase clades; it is most similar to fungal tannase clades 7 and 8. Regardless, the proximity of the honey bee micobiome tannase to the fungal tannase group including *Aspergillus fumigatus*, a fungus responsible for stonebrood disease in honey bees, is very interesting (Figure 7).

Clade 6 is very small, aligning with a few rhizosphere and Coleopteran sequences but it does not seem to have a clear environmental identity. Clade 6 does only align with 2% of currently available environmental metagenomes.

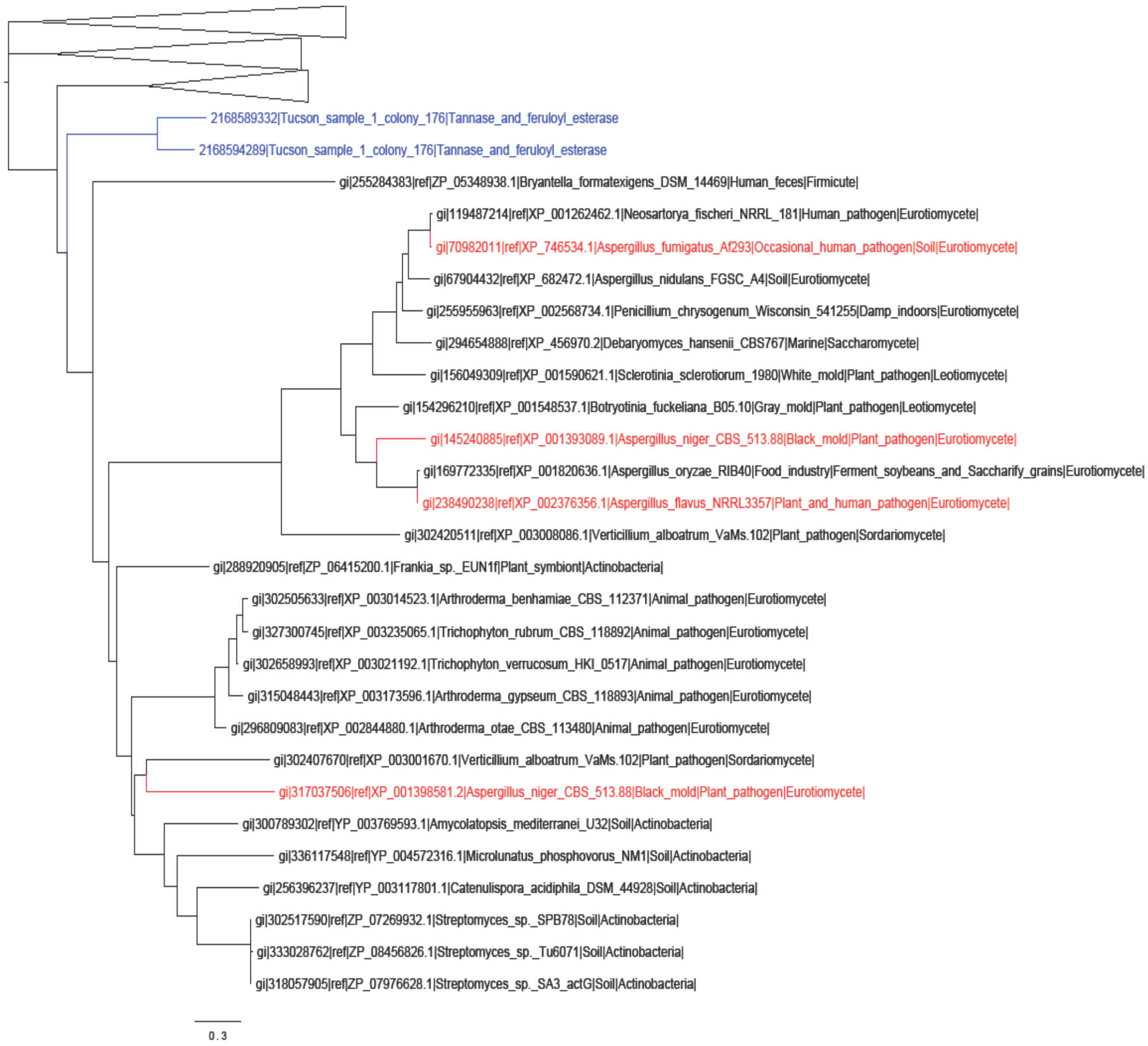


Figure 7 - The reference phylogeny including placement of the Tucson honey bee microbiomes (blue). The tree shows their proximity to *Aspergillus* tannase (red) of which *A. fumigatis* is responsible for stonebrood disease in honey bees.

1.3.6 Phlorotannins

Horvath (1981) defined tannins as any phenolic of a high molecular weight that contains enough hydroxyl groups (or other suitable groups like carboxyls) to form strong ester bonds with macromolecules like proteins and precipitate them. Phlorotannins (Figure 8) share similar structure with tannins (Figure 1). Both are polyphenols that present numerous hydroxyl groups necessary to form ester bonds, however phlorotannins are exclusively marine and produced by brown algae. The behaviour of the compound is similar to that of tannins, by discouraging herbivory and participating in the chelation of metal ions. Figure 5 shows the presence of tannase in marine environments, where both phlorotannins and terrestrial tannins are known to exist.

1.3.7 Conserved domains

From the reference alignment it was possible to observe several conserved domains that are either new, or extensions of domains discussed in previous work. A list of these domains can be seen in Table 4. The sequences under study commonly start with a Cysteine, though this is only conserved in 71% of sequences across all groups. A Valine present in 64% of sequences is paired with additional Valine in groups 7 and 8. Groups 7 and 8 also have a very early Histidine as part of an HGDVV region not seen in the other groups. Starting at site 11-23 (in the reference alignment, this applies to all site locations), all groups share variations of **P...W...R...G...GG...G** at 84% consensus. Most noticeable are the 3-6 copies of Glycine of which two are present in 90% of sequences, as well as the Proline which is present in 99% of all sequences.

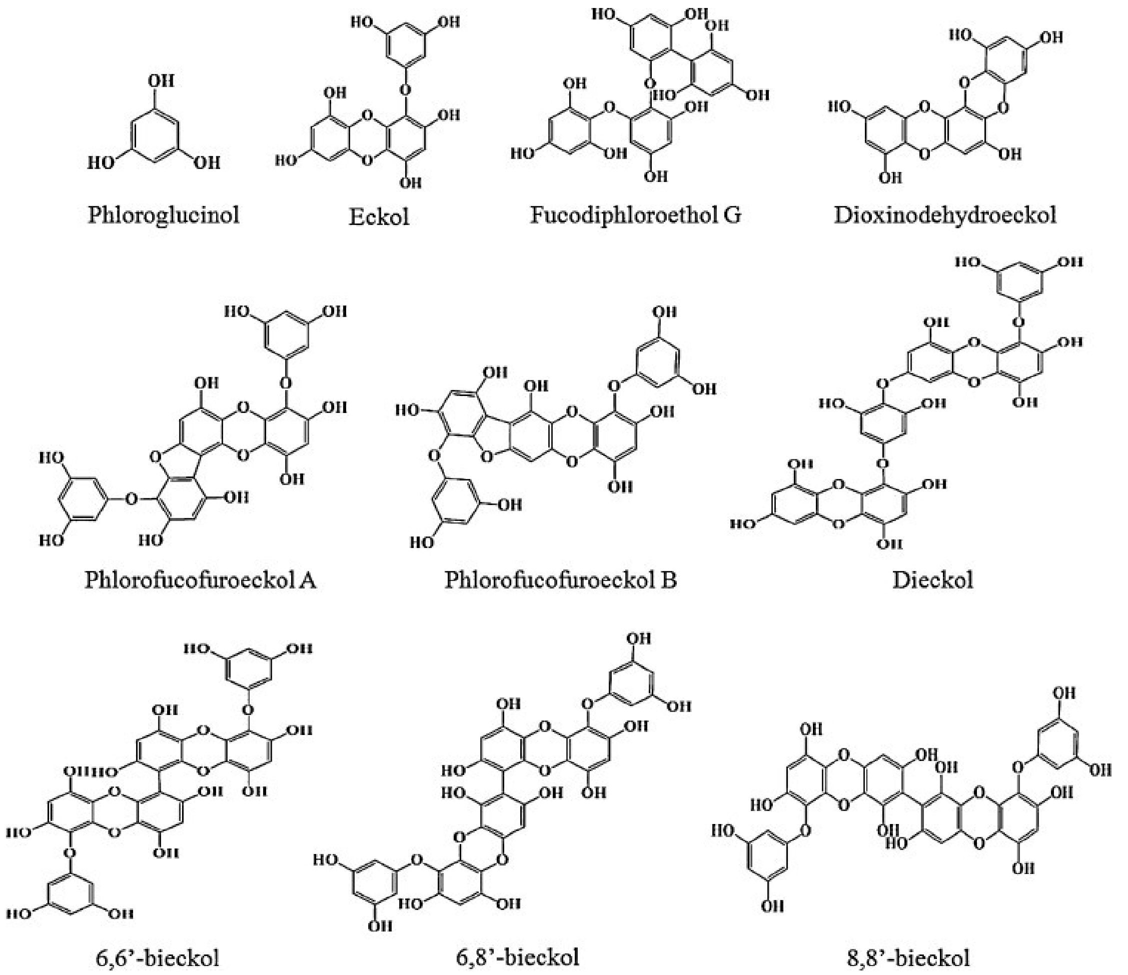


Figure 8 - The structures of phlorotannins, as illustrated by Wijesinghe and Jeon (2011).

Following this is a region at sites 69-74 where there is a conserved **G...D...G** in 80% of sequences. Only the second Glutamine is conserved in 98% of the sequences. Groups 7 and 8 are missing this Histidine, but have one much earlier in the sequence as mentioned earlier.

Table 4 - Tannase conserved domains obtained from the reference alignment. Domains were established using a cut-off of 80% presence among sequences.

Conserved domain	% Consensus
P...W...R...G...GG...G	84%
G...D...G	80%
K...Y...Y...GCS...GGR...P...DG...P	80%
CD...DG...DG...C...C	88%
D...L...GGK...G...D	80%
F...R...PG...HC...G	80%
WV...E...G...P	80%
C...P...G	80%

In most sequences there exists a highly conserved domain, present between amino acids 73-100. This is the first noticeable motif, **K...Y...Y...GCS...GGR...P...DG...P** contained in 80% of the sequences, of which **Y...Y...G...S...GGR...DG...P** are conserved in 94% of the sequences. This motif is an extension of the G...S...G motif discussed by Carter and Wells (1988) and Brady et al. (1990).

The next motif, known as **CD...DG...DG...C...C**, can begin from 198 to 218 and is present in 88% of the sequences. The only exception to this motif is group 6, which instead carries a **CD...GG...LG...C...C**. An alternate motif, **CD...G...DG...C...C** in 92% of the sequences, could serve as a compromise between the two motifs that differ in the amino acid preceding the first glycine.

Another motif, **D...L...GGK...G...D**, from 302 to 369, is conserved in 80% of the sequences. Only **G...K...G...D** residues are present at 93% consensus.

The region from 399 to 473 is **F...R...PG...HC...G**, conserved at 80%, but the only **G...H** remains conserved at 97%. The Histidine, located anywhere from 410 to 472 and recognized by the FRPGHC pattern is actually one of the only two amino acids in the alignment that are present in 100% of the sequences. This suggests that Histidine may be of some importance in tannase function.

The next conserved domain contains the other amino acid present in 100% of the sequences, Tryptophan. This Tryptophan is identified by proximity to the pattern **WV...E...G...P**, conserved in 80% of the sequences, but only **W...P** is retained in 95% of all sequences, while Tryptophan remains in all sequences. This Tryptophan occurs anywhere between 430 and 536 in the chosen sequences and its presence may also be of great importance for the proper functioning of the tannase enzyme.

All of the sequences end with some variant of **C...P...G** in 80% of the sequences, however only Proline is present in 97% of all sequences. This Proline is often bordered by Tyrosines, with the exception of group 5, where the Proline is bordered by other amino acids. Members of group 7 have Tryptophan instead, in the exact same positions where the others have Tyrosine.

1.3.8 Conserved Cysteines in Tannase

A study in 2007 by Battestin and Macedo, discovered reduced tannase activity when tannase is introduced to cysteine and 2-mercaptoethanol, suggesting the active site of tannase involves sulphur-containing amino acids. This would explain the large number of highly conserved Cysteines in the sequences of the reference phylogeny. Of the most conserved amino acid residues, the two that appear most often are glycine and cysteine. There are 18 highly conserved Glycines and 8 highly conserved Cysteines out of 57 amino acid residues that are conserved in over 80% of sequences.

A previous study refers to a catalytic triad of Ser133-His247-Asp194 in *Apergillus niger* most likely responsible for the feruloyl esterase function of the enzyme (Hermoso et al., 2004). In the reference alignment, the catalytic triad is Ser178-His610-Asp261. The change in distance between the location of the catalytic triad in *A. niger* and this alignment are easily explained by the introduction of gaps during the alignment and the many differences in the sequence between species. Variation in the location of this triad between sequences is discussed in Wong (2006). A comparison with most individual sequences yields similar positions, such as Ser127-His392-Asp193 for *A. niger CBS 513.88*. In an effort to determine if these sequences would form structures similar to those discovered from the crystal structure of *A. niger* feruloyl esterase, a sequence from the alignment was submitted to the RaptorX server at the University of Chicago, which generates the most likely protein structure using homology modelling. The sequence chosen was that of *Botryotinia fuckeliana B05.10* a plant pathogen with a putative tannase domain, and the best structure chosen by RaptorX reveals that the Histidine, Aspartic Acid and Serine residues (Ser108-His347-Asp308) conserved in 99% or more of the chosen sequences form a triad similar to the one discussed in Hermoso et al. (2004). The structure is displayed using RasWin and is shown in Figure 9.

Macedo and Battestin (2007) discuss the importance of cysteine residues for tannase function. Figure 10 shows two heavily conserved cysteines; cysteine 107 and 348 which are members of the **K...Y...Y...GCS...GGR...P...DG...P** and **F...R...PG...HC...G** conserved domains respectively. It should be noted that the cysteines in those conserved domains are only present in 80% of sequences. It can be seen in Figure 9 that these cysteines (colored yellow) appear directly at the mouth of the catalytic triad.

1.4 Discussion

Cleaning non-tannase domains from the reference sequences made the alignment of the sequences and subsequent phylogenetic trees much more accurate. The improved alignment allows for better visualization of conserved domains and better understanding of the evolutionary history of tannase. Future work should control for domains within reference sequences that are unrelated to the sequence under study. This unique restructuring of the tannase/feruloyl esterase phylogeny with a focus on global distribution shows that tannase clades may be defined by their environmental specificity. If so, it is possible that the differences in tannase enzymes reflect selection for enzymes best suited for a specific collection of substrates encountered in their environment. Future functional characterization using a variety of tannin compounds may well show that the tannase clades in the reference phylogeny are each be more suited to work with certain substrates over others. There is the possibility that tannases are merely feruloyl esterases with sufficient cysteines and therefore disulfide bridges to alleviate stress associated with tannin substrates. If this is the case, perhaps the most useful factor in organizing the tannase and feruloyl esterase family may be the environmental specificities suggested by the clades in this reference phylogeny. Despite the need for more microbiome sampling to obtain a more complete understanding of environmental specificity, the relationships shown using the reference phylogeny may serve as a guide for future study.

Fungi appear to have obtained their tannase genes from two main microbial sources, proteobacteria and actinobacteria, that have resulted in two different fungal tannase clades. This is supported by fungal tannase sequences of the same class, genus and species falling within separate bacterial groups.

The extensive transfer of tannase between distinct groups of bacteria, as expected, is also apparent in the tree. Of particular interest is the occurrence of fungal xenologs, multiple different tannase genes in the same species, implying different origins of the gene in the same fungal organism. These gene copies are not identical however, as some align with proteobacterial clades 3 and 5, while the others align with the more actinobacterial clades 7 and 8. This is highly suggestive of not only a bacterial origin for the tannase genes present in fungi, but several different bacterial origins of the gene within the same fungus. There may be even more examples of this in fungi not included in the

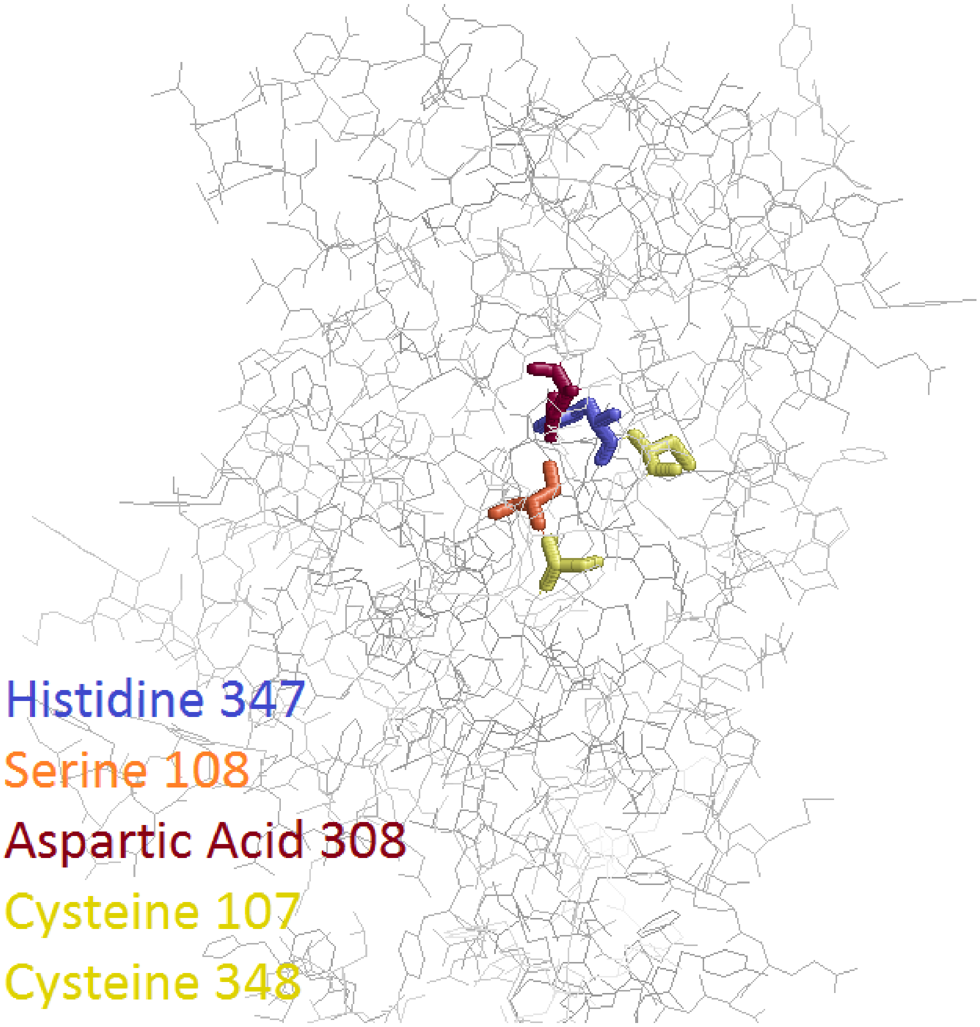


Figure 9 - Tertiary structure of tannase from *Botryotinia fuckeliana*. The sequence was generated by Raptor X, a program hosted at the University of Chicago for constructing tertiary structures from protein sequences using homology modelling. The structure shows the feruloyl esterase triad in blue, orange and bergundy, with potentially important conserved cysteines shown in yellow.

alignment, such as sequences removed to improve phylogenetic analysis, those in the NCBI nr database or very recent additions to the refseq database of NCBI. Alternatively, it may be possible that gene duplication and subsequent evolution of the gene within fungi account for the differences among fungal tannase copies and these copies were then passed on to bacteria, however, all of proteobacteria receiving one type of fungal tannase while all of actinobacteria receives the other seems unlikely. If the origin of the gene were truly fungal and then shared with bacteria, to see such a clear division between proteobacteria and actinobacterial tannase clades would not be expected. The split between the fungi and their association with different bacterial groups is suggestive of lateral gene transfer, especially considering the appearance of some similar species and genera in both fungal tannase clades, such as various *Aspergillus* species and *Verticillium alboatrum*.

The results from the reference phylogeny are different from the findings of Banerjee et al. (2012), who showed a single fungal group separate from bacteria. From this updated phylogentic analysis, the impact of lateral gene transfer in the evolution of T/FE can now be assessed. In fact, we find considerable evidence of lateral gene exchange throughout the tree.

One of many possible lateral gene transfer events suggested by the maximum likelihood tree is shown by the high tannase similarity between the *Leptotrichia* of human oral flora and bovine rumen flora.. It seems possible that repeated inoculation of the human oral cavity with the genus *Mannheimia* or *Actinobacillus succinogenes* may have resulted in the horizontal transfer of the tannase gene to our resident Fusobacteria. This potential horizontal transfer raises the question of what other genes may have been shared in addition to *L. buccalis* tannase (Lebu_1642). Evidence of the partial transfer of neighboring genes could support the idea that horizontal transfer indeed occurred. After performing BLAST searches using neighboring Lebu genes from the genome of *L. buccalis* against the non-redundant database, it becomes clear that there are other similarities in genes between these bacteria. The sequence of Lebu_1646, known as Antibiotic biosynthesis monooxygenase is 37% similar to *A. succinogenes 130Z* (Query coverage: 97%, E value: 9e-47) and 38% to *M. succiniciproducens MBEL55E* (Query coverage: 95%, E value: 9e-46). Lebu_1645, an aldo/keto reductase, is found to be 51% similar to a Tas protein in *M. succiniciproducens MBEL55E* (Query coverage: 99%, E value: 2e-109). Lastly, Lebu_1628, an acetyltransferase, is 42% similar to *A. succinogenes 130Z* (Query coverage: 89%, E value: 4e-12) and 41% similar to *M. succiniciproducens MBEL55E* (Query coverage: 88%, E value: 3e-12). Figure 3 allows a higher resolution look at all the species that share this similar version of tannase shared by the genera *Leptotrichia, Mannheimia* and *Actinobacillus*. The same relationship with rumen bacteria is shown, but this time *Actinobacillus ureae* is included beside *M. succiniciproducens MBEL55E*. *Actinobacillus ureae* is a human pathogen responsible for pneumonia, sepsis, peritonitis and meningitis, and its tannase similarity to *Mannheimia* strains responsible for bovine respiratory disease is very interesting. While it is not surprising that pathogens responsible for respiratory diseases in different mammals share sequence similarity, it is surprising to see tannase sequence similarity. Tannase works with plant defences and is not expected to assist the activity of a pathogen in mammals. Most likely, the close tannase sequence relationship is an indicator of the pathogens being closely related or of horizontal gene transfer. Other new sequences added to Figure 3 include *Lautropia mirabilis*, which is a normal member of the human oral and pulmonary flora. *Cardiobacterium valvarum*, a human pathogen responsible for infective endocarditis is also shown to have similar tannase. Also included are sequences for three *Campylobacter* species, which are epsilon proteobacteria present as normal human oral, fecal and gut flora, but also may be responsible for inflammatory bowel disease in children (Lastovica, 2009).

Much like the divide between fungal tannases, clade 4 is also suggestive of the establishment of tannase in eukaryotes having arisen from lateral gene transfer with bacteria. In clade 4, there are almost identical tannase sequences between *Klebsiella variicola At22* and *Pantoea sp. At9b*. The former is a fungal symbiont dispersed by leaf cutter ants (*Atta colombica)* in fungal gardens in Panama, the latter is a bacterial symbiont living within the leaf cutter ants of Panama. Given their proximity, opportunities for lateral gene transfer should be very frequent. Also in clade 4 is the tannase of the oomycete *Phytophthora infestans*, a water mold that causes potato blight. This eukaryote actually has tannase that is more similar to that of a large group of *Pseudomonas* than to the tannase in other eukaryotic life.

In terms of function, little is known about what, if anything differentiates a tannase from a feruloyl esterase at the level of the amino acid sequence. When functionally characterized tannases were added to the reference phylogeny, 4 of the 7 functionally characterized tannase sequences cluster with fungal clade 7. This is suggestive of the industrial bias in the functional characterization of this gene. As mentioned, the fungal tannase producers are historically the most studied tannase producers and most used in industry. The remaining three sequences are located within clade 4 containing the eukaryotic *Phytophthora infestans.* It is unclear whether or not sequences in clades other than 4 or 7 also have tannase function given the very low number of functionally characterized tannase sequences available. It seems there is a high diversity of tannases that require further biochemical characterization.

The Tannase Superfamily representatives characterized by Udatha et al. (2011) align with fungal clade 5. This is not surprising as all the representatives of this superfamily studied by Udatha et al. (2011) are fungal themselves. Despite some fungal tannase having higher similarity to bacterial tannase than other fungi, the extra non-tannase domains retained in the representatives from Udatha et al. (2011) are most likely clustering their fungal sequences together. Figure 4 shows *Klebsiella* and *Pseudomonas* to be quite distant from the Udatha et al. (2011) Tannase Superfamily. This is more evidence that the Udatha et al. (2011) superfamilies are biased towards fungal groups and microbial tannase does not fit well in this system of 12 Feruloyl Esterase Families.

Given the confusion in characterisation and nomenclature of tannase and feruloyl esterase, the interchangeable use of terms might make it appear as if the same sequence may create a protein that carries out both the function of a tannase and a feruloyl esterase. It would be interesting to determine if any of the sequences in the reference phylogeny can be functionally characterized as both, however there have been no efforts to test the same enzyme with substrates associated with both tannase and feruloyl esterase reactions. The catalytic triad typical of feruloyl esterases is present in each of the selected sequences, including those identified as tannase, suggesting all of the sequences may have feruloyl esterase function.

Increased functional characterization of available reference sequences, especially those belonging to bacteria, would allow the development of more accurate nomenclature related to substrates as well as permit investigation of which of the conserved domains are responsible for functional differences and where they exist in the evolutionarily distinct tannase clades observed. The superfamilies defined by Udatha et al. (2011) do not directly explain the functions that distinguish their families, however the sequences that make up their tannase superfamily all contain the same two cysteine residues that seem to appear at the mouth of the feruloyl esterase triad, while the other superfamilies do not. Despite the apparent fungal bias of the Udatha et al. (2011) study, it provides strong support for the idea that cysteine residues may be essential for a feruloyl esterase to function as a tannase. Further, it is possible that these residues and their ability to form stabilizing disulfide bridges may be the difference between feruloyl esterases that can serve as tannase and those that cannot. This could be tested by functionally characterizing all of the sequences available and comparing them to the sequences of the established clades in our reference phylogeny.

Clade 1 aligns with the majority of human microbiome samples. This is most likely due to the human oral Fusobacteria in clade 1 and the large amount of human oral microbiome samples available. If a similar number of samples were taken from elsewhere in the human microbiome it is likely that clades 2 and 4 especially would be better represented in terms of human-associated tannase. As an example, clade 2 contains a *Rahnella* species found in the human intestine and clade 4 contains an opportunistic *Burkholderia* species that colonizes the human lungs of cystic fibrosis pateints (Chaparro et al., 2001). It is possible that the human microbiome is not restricted to tannase clade 1, but that there may be areas of the human microbiome where specific tannase clades dominate. For example, clade 1 perhaps makes up the human oral microbiome, clade 2 dominates human intestine and clade 4 can be most commonly associated with the human lungs. It is currently unclear where clade 3 fits in the human microbiome.

The fact that clade 5 is not well represented in other microbiome samples may be due to uneven sampling, as terrestrial fungal tannase should be aligning with many soil and rhizosphere samples. As shown in Figure 6, the distribution of microbiome hits across clades is 25%, 21%, 38% and 10% respectively for clades 1-4. Only 6% of hits fall in clades 5-8, suggesting fungal tannase sequences are not well represented in current microbiome samples.

Clade 6 is very small, aligning with a few rhizosphere and Coleopteran sequences but it does not seem to have a clear environmental identity. Perhaps there is a specific niche for this tannase clade that is currently undersampled.

Interestingly, the tannase within the microbiomes of honey bees in Tucson, Arizona does not fit completely in any of the tannase clades; it is most similar to fungal tannase clades 7 and 8. The inability to place this tannase in any of the 8 clades may be due to independent evolution of the bacteria within honey bees. Lateral gene transfer between fungi and resident bacteria of honey bees surviving stonebrood disease may account for this similarity.

Given the structure similarity between tannins and phlorotannins and the potential for ester bonds as shown by multiple hydroxyl groups, there seems to be much potential for tannase action on phlorotannins by tannase-producing marine bacteria. Currently this avenue is completely unexplored. It is unknown whether the tannase enzyme can cleave the ester bonds of phlorotannins. It would be interesting to determine whether tannase enzymes produced in marine environments act only on tannins contained in dissolved organic matter or on phlorotannins produced by brown algae as well. This would also affect the understanding of the presence of acidobacteria (a family of tannase/feruloyl esterase producers) in chitons. Fukunaga et al. (2008) discovered novel acidobacteria in the chiton *Acanthopleura japonica* in 2003. The grazing activity of chitons includes brown algae in the photic zone, a preferred food of molluscan herbivores (Yates, 1989). If tannase produced by the acidobacteria does act on phlorotannins it would facilitate the digestion of brown algae by chitons.

The exploration of known microbiomes shows that the tannase enzyme has spread to a variety of animals that can also benefit from the degradation of tannins via microbial or fungal association. While the phylogenetic tree demonstrates the relationships between microbial and fungal tannase, much work is left to be done identifying the adaptive radiation of this enzyme. It is clear that there are most likely many other undiscovered examples of herbivores worldwide, terrestrial and perhaps marine, that use tannase to efficiently detoxify tannins. For example, the Japanese wood mouse uses tannase-producing *Lactobacillus* bacteria in their acclimation to acorn tannins (Sasaki et al., 2005). Only two of all the known insect orders are strongly represented in current environmental microbiomes, this suggests that perhaps the microbiomes of insects are undersampled or the metadata unaccessible. Considering the herbivorous nature of many insects and that several symbiotic events have been identified in insects, it may be likely that there are undiscovered tannase-producing symbionts in many insects.

Chapter 2 - Tannase Producing Symbionts in *Malacosoma disstria*

2.1 Introduction

Malacosoma disstria Hübner (Lepidoptera: Lasiocampidae), more commonly known as the forest tent caterpillar (FTC), can be found in most hardwood forests throughout North America. In the FTC life cycle, females mate once and lay all their eggs in a single egg mass. The eggs are laid in late summer around twigs of the host plant and must overwinter before emerging in early spring. From spring to late June, larvae go through five or six developmental stages known as instars in which they eat all the food necessary for their development; adults of this species are non-feeding. The sixth instar is usually only observed under lab conditions when the caterpillars are reared on poor-quality diet. In late June, individuals pupate and emerge as adult moths in July (Fitzgerald, 1995).

Approximately every decade, populations of FTC experience large outbreaks, during which the larvae cause severe defoliation of host trees. Outbreaks can last up to four years, and during this time defoliation by FTC has been linked to tree death (Fitzgerald, 1995). Even when outbreaks do not cause tree death, defoliation has a negative effect on photosynthesis, immune responses and other normal functions of the host tree. In eastern Canada, outbreaks are more or less synchronized, occurring every decade (Sippell, 1962), while outbreaks in western Canada are more asynchronous (Hildahl and Reeks, 1960). Cooke and Lorenzetti (2006) analyzed outbreaks in Quebec between the years of 1938 and 2002, and determined what they call an outbreak range that covers 95 019 834 acres, meaning outbreaks during that time covered at most 25% of Quebec's surface area. Due to the distribution of host-tree species in Quebec, FTC outbreaks in northern Quebec occur on aspen while outbreaks in southern Quebec occur on maple, the former observed to have an outbreak cycle of 9 years while the latter has a 13-year outbreak cycle between the years of 1938 and 2002 (Cooke and Lorenzetti, 2006). During an outbreak beginning in 2002 and ending in 2005 led to 650 000 acres of maple trees were defoliated in New York and 230 000 acres defoliated in Vermont. The outbreak continued in Vermont, defoliating an additional 343 000 acres in 2006 (Vermont Department of Forests, Parks & Recreation).

In eastern Canada, the main host trees for the forest tent caterpillar are trembling aspen (*Populus tremuloides)* and sugar maple (*Acer saccharum)* (Trudeau et al., 2010). They are also known to feed on several other secondary hosts, but their preferred host largely depends on where populations are geographically situated (Stehr and Cook, 1968). Parry and Goyer (2004) studied geographically isolated FTC fed on a variety of host trees and evaluated their success based on pupal mass, development time and survival. They found evidence that FTC are regionally adapted to specific host trees as opposed to being a generalist species. An unpublished observation made by the Despland lab at Concordia University is that western Canadian FTC experience high mortality on *Acer saccharum* (sugar maple), while eastern Canada FTC survive reasonably well on the same diet. This observation is consistent with the findings of Parry and Goyer regarding regional adaptation, but had to be confirmed through experimentation.

Sugar maple foliage contains large quantities of the phenolics known as tannins and is typically less nutritious than aspen. Biochemical analysis of the two types of foliage show that aspen contains twice as many soluble sugars as maple, while maple has the higher concentration of tannins (Lorenzetti et al., 1998). Sugar maple leaves also contain low protein (Barbehenn et al., 2009b). A diet of aspen foliage results in faster larval development and larger pupal weight over maple foliage (Lorenzetti, 1993). Quebec female FTC reared on sugar maple reach pupal masses between 210 and 400 mg, considered moderate for females (Fortin and Maufette, 2002). It is suspected that this is due to the low sugar and high tannin content of the foliage, which causes oxidative damage when the tannins are oxidized in the midgut. Mortality of western FTC on sugar maple may be due to the inability to deal with this oxidative damage (Nicol et al., 1997 and Barbehenn et al., 2005). Sugar maple leaves contain high levels of gallotannins and ellagitannins and it has been suggested that this phenolic combination is prone to creating oxidative stress (Barbehenn et al., 2005).

In vertebrate herbivores, tannins binding to macromolecules is the main reason for toxicity. In caterpillars, the high pH environment of the midgut oxidizes tannins, creating oxidative stress in midgut tissues (Barbehenn and Constabel, 2011). For caterpillars of species known to encounter tannins, there exist both phenolic-tolerant and phenolic-sensitive species, the former of which tend to have higher concentrations of antioxidants in their midgut (Barbehenn et al., 2008). *Lymantria dispar* (gypsy moth) is a species considered to be phenolic-tolerant, however sugar maple foliage still leads to poorer development as it contains less protein than red oak (Barbehenn et al., 2009b). Antioxidants, like ascorbate, are believed to be beneficial in dealing with mature tannin-rich foliage (Barbehenn et al., 2005). Glutathione is also helpful to defend against midgut oxidation however it is costly to produce, requiring short-supply amino acids (Barbehenn et al., 2013). In general, the use of antioxidants to defend against tannins is associated with a high metabolic cost, but this cost may be met by compensatory feeding (Barbehenn et al., 2009a). FTC caterpillars are considered phenolic-sensitive (Barbehenn et al., 2003) and sugar maple foliage creates high levels of peroxides in the midgut fluids of *O. leucostigma* and FTC (Barbehenn et al., 2005). Unlike *O. leucostigma* which has naturally higher levels of ascorbate that reduce the effect of peroxides, FTC's have naturally lower levels of ascorbate that lead to high, unmanaged levels of semiquinone radicals and protein carbonyls, which are commonly used as biomarkers of oxidative stress. FTC lack sufficient quantities of antioxidants necessary to abate foliar tannins but eastern individuals are still able to survive better on tannin-rich sugar maple than their western counterparts. This implies that a fundamental difference exists between western and eastern FTC. Potentially, this could be a physiological difference in the management of antioxidants and prooxidants, or the difference may be explained by the activity of the microbiome.

A potential benefit of a bacterial symbiont in FTC is the possibility of tannase enzyme production. Humans and other mammals are able to produce proline-rich proteins (PRP's) in their saliva that serve to bind tannins however, in animals without saliva or PRP's, tannase produced by symbionts may serve as an efficient means to counter the harmful effects of tannins (Shimada, 2006). As discussed in Chapter 1, tannase may be beneficial to any herbivore feeding on tannin-rich foliage, however it is only known to be expressed by bacteria and fungi, making symbiotic relationships essential for its use by herbivores. Bacterial symbionts occur in at least ten percent of insect species, usually by complementing the diet (Wernegreen, 2002). Typically these microbes exist either as primary endosymbionts and have a long co-evolutionary history with the host and are usually mandatory for host survival, or as secondary symbionts, which have formed a more recent association with their host (Moya et al., 2008).

Typically, tannins are either oxidised, egested or hydrolyzed (in some cases by tannase) in varying proportions depending on the species of caterpillar (Barbehenn and Constabel, 2011). In FTC, 20% of tannins are egested while the rest are chemically transformed, mostly in the midgut. It is suggested that majority of chemically transformed tannins in FTC are oxidized rather than hydrolized (Barbehenn and Martin, 1994) and hydrolysis occurs more often in other moths such as the autumnal moth (Salminen and Lempa, 2002). The products of hydrolysis in FTC may also have different fates. While ellagic acids are absorbed and excreted by Malphigian tubules, other products like gallic acid are oxidised (Barbehenn et al., 2009a). This means that even if tannase is present, some hydrolysis products like gallic acid may still lead to oxidative damage. The smaller size of these products also enables them to enter midgut tissues through the peritrophic membrane, causing further damage. Despite this, enzymatic modifications by tannase have been shown to increase antioxidant activity and would be beneficial to caterpillars suffering oxidative stress due to the oxidation of tannins. Lu and Chen (2007) showed that tannase-catalyzed hydrolysis of plant secondary metabolites known as catechin gallates led to a significant increase of antioxidant acitivity against hydrolysates such as DPPH radicals, superoxide anions and hydrogen peroxide. The tannase of *Aspergillus niger* has been shown to synthesize the antioxidant propyl gallate using tannic acid as a substrate (Sharma & Gupta, 2003).

Many instances of tannase-producing symbionts present in animals have been observed, such as the Japanese Wood Mouse *(Apodemus speciosus)* which uses the tannase of *Lactobacilli* and *Streptococcus gallolyticus* to break down the tannin-rich acorns of the deciduous oak species *Quercus crispula* (Shimada, 2006). Several other mammals, such as koalas, goats, sheep and ringtail possums, also host taxonomically novel bacterial species that express tannase (Osawa et al., 2000).

Chapter 1 discusses tannase-producing bacteria associated with the midguts of leaf-cutter ants and xylophagous beetles. In fact, symbiotic bacteria frequently play major roles in insect life. Bacteria producing essential amino acids for insect hosts, such as *Baumannia cicadellinicola* and *Sulcia muelleri*, a gamma proteobacteria and bacteroidetes respectively that live within the gut of *Homalodisca coagulata,* (Hemiptera) have been identified (Moran et al., 2005). Many aphids use microbes for dietary supplement as well, associating with a variety of *Buchnera* species for amino acid production, and *Hamiltonella defensa* for defense against parasitic wasps (Degnan et al., 2009). Stink bugs, *Riportus clavatus* are also known to populate their posterior midguts with new *Burkholderia* from the soil every generation (Kikuchi et al., 2005). Some *Burkholderia* strains are even capable of conferring resistance to Fenitrothion, a popular pesticide (Zhang et al., 2006). Termites especially host many microbes involved in various activities such as acetogenesis, uric acid recycling, sulfate reduction, nitrogen fixing, lactic acid production, methanogenesis and most importantly, cellulose degradation (Liu et al., 2011). Considering the many ways insects may benefit from bacteria, the gut profile of FTC may include bacteria with benefits other than or in addition to tannase activity.

Forest Tent Caterpillars belong to the order Lepidoptera, and several examples of microbial symbiosis within the order have been reported. Aside from adult moths that feed (some, like FTC do not), generally the presence of symbionts in the larval or caterpillar stage of Lepidopteran species has been controversial in the past. Food is passed through the digestive system of the caterpillar very rapidly and because of this it was believed that symbionts can only play a minor role in caterpillar digestion. This has been determined to not be the case, however. The velvetbean caterpillar *Anticarsia* *gemmatalis* has been shown to avoid the harmful effects of plant hosts rich in protease inhibitors with the help of bacterial lypolitic and proteolytic activity (Visotto et al., 2009).

Perhaps the most studied caterpillar microbiome is that of *Lymantria dispar* (gypsy moth). *Lymantria dispar* caterpillars are known to carry *Enterococcus* species in the midgut that confer resistance to penicillin-type beta-lactams by expressing genes for efflux pumps and lactamase (Allen et al., 2009). *Serratia* species also inhabit the *L. dispar* caterpillar gut to provide amino acids, but exist at subclinical levels until altered community structure causes it to be pathogenic. For *L. dispar*, although different microbiomes have been observed in caterpillars on different diets, *Enterococcus faecalis* appears in all individuals, and might be involved in reducing gut pH to prevent the activation of Bt toxin, which is activated at high pH (Broderick et al., 2004). In the caterpillar stage of five moth species, *L. dispar, V. cardui, M. sexta, P. rapae and H. virescens,* it has been shown that susceptibility to Bt toxin is dramatically reduced by eliminating *Enterobacter sp. NAB3* from the gut using antibiotics. It is also shown that reintroduction of the *Enterobacter* typically reintroduces toxicity. These examples stress the importance of Lepidopteran interaction with microbes.

If FTC rely on symbionts to degrade plant toxins, this knowledge could be quite valuable considering the impact FTC population outbreaks have on forests. In addition to pest species like FTC, maple trees and associated industries such as maple syrup production have been stressed in recent years due to climate change (Murphy et al., 2012). Avoiding further stress to important timber and non-timber forest products would help quell economic stress. Improved management of outbreaks on maple in eastern Canada could be enabled by an investigation of midgut bacteria in FTC. As discussed by Janson et al. (2008), bacterial symbionts have been known to promote plant defense tolerance in some insects. If this is the case in FTC, identification of the microbes responsible could not only improve understanding of the species, but also lead to pest control measures that do not involve either current low-specificity pesticides or viruses such as nuclear polyhedrosis that target all moths and butterflies. An attempt to induce the western FTC phenotype in eastern FTC could prove useful in quelling eastern outbreaks on sugar maple. The use of appropriate antibiotics could eliminate sugar maple as a food choice for FTC.

The objective of this study was to explore the midgut microbiome of FTC for the first time, in attempt to determine the presence of symbiotic tannase-producing bacteria. The presence of these bacteria may help explain host-specialization of different regional populations. Broderick et al. (2009) identified several bacterial species in the Lepidopteran larvae midguts of *Vanessa cardui, Manduca sexta, Pieris rapae, Pectinophora gossypiella* and *L. dispar.* These included species of *Lactococcus, Kelbsiella, Enterobacter, Pantoea, Enterococcus* and *Pseudomonas.* It was expected to identify similar bacteria in FTC, given the relatedness, shared physiology and lifestyle. Given the presence of *Bacillus thuringiensis* in *Bombyx mori* and *Ephestia kuehniella* (Broderick et al., 2009) it is likely that members of the *Bacillus* genus may be found in FTC as well other Lepidopteran microbiomes.

Cultivation-independent molecular studies using the 16S rRNA gene as a bacterial marker have shown that diversity in Lepidopterans is between 7 and 15 phylotypes, which is very small compared to the approximation of 50 phylotypes in termites (Ohkuma and Kudo, 1996) and 500 in humans (Wilson and Blitchington, 1996). The guts of Lepidopteran species are also chemically extreme, possessing a pH of 8-10 with some as high as 12.4, and containing terpenes and phenolics from foliage (Broderick et al., 2004). It seems reasonable given the relatedness of *L. dispar* and forest tent caterpillar, that if *L. dispar* can support a microbial community and derive benefits from it that may influence performance, then so could FTC.

As a first step in determining whether the inferior performance of western FTC on sugar maple is due to bacterial intervention, it was first necessary first survey the microbiome of FTC. Both culture and culture independent methods were used to determine the bacterial flora of FTC. Using culture-independent methods for studying microbiomes eliminates the need to predict optimal conditions for the growth of bacteria whose identity is not yet known. Additionally, rare bacteria or bacteria that are difficult to culture are more easily identified (Su et al., 2012). A complete picture of the microbiome can best be obtained by directly amplifying and sequencing all 16S rRNA genes in a genomic extract. As ribosomal RNA is a part of all self-replicating systems and evolves extremely slowly, knowledge of its nucleotide sequence allows us some means of identifying species and evaluating their relatedness (Woese and Fox, 1977). Once members of the microbiome are identified, it is then possible to evaluate the likelihood of tannase production and its ultimate use by FTC's.

2.2 Methods

2.2.1 Insect rearing 2011

In 2011, egg masses were obtained from Southern Ontario (44°33.5 N, 76°24.1 W), St. Esprit, Quebec (44.9000° N, 73.6667° W) and British Colombia Highway 16 West (53.83° N 123.14° W). The Ontario forest was predominantly sugar maple (*Acer saccharum*), the BC forest was trembling aspen (*Populus tremuloides*) outside of the distribution range of sugar maple, and the Quebec forest included both tree species. The numbers of egg masses obtained from the Ontario, BC and Quebec locations were 30, 22 and 22 respectively (8 extra Ontario egg masses were used in anticipation of their poorer performance on aspen). Considering the interest in microbe activity, egg masses were only briefly rinsed in household bleach (to minimize viral infections, a variation of Grisdale, 1985). They were subsequently held in an incubator with a 16h:8h light:dark photoperiod at 22 ^{o}C.

The egg masses from each region were hatched in the incubators. Half of the larvae from each region where reared on sugar maple foliage while the other half were reared on trembling aspen foliage. The foliage was untreated to avoid removing any environmental microbes. Caterpillars were reared in small Tupperware® containers (4 inches by 8 inches) until the fourth instar, at which stage their larger size required much larger Tupperware® (8 inches by 12 inches). At larger larval sizes, small living space may limit growth and facilitate the spread of the nuclear polyhedrosis virus.

It was only on the aspen treatment that any Quebec and BC individuals made it to the pupal stage. The pupae were weighed and sexed, and placed into individual creamer cups sealed with paper lids. Each group on the aspen treatment was given the opportunity to mate in tents with the choice of 2 trembling aspen branches, 2 sugar maple branches and 2 Spruce branches resting in 500mL Erlenmeyer flasks of water. However, oviposition was not observed because the tents were repeatedly disturbed by a groundhog.

2.2.2 Insect Rearing 2012

Egg masses were collected from Southern Ontario (Owen Sound) and British Colombia (Tweedsmuir South Provincial Park), but none were available from St-Esprit, Quebec. As well, egg masses were not bleached prior to hatching. In addition, midguts were sampled from the 3rd, 4th and 5th instars, and an effort was made to sample individuals affected by the nuclear polyhedrosis virus. Due to the difficulties experienced sequencing extracted DNA from colony-independant methods attempted in 2011, cultures were grown in 2012.

2.2.3 Caterpillar Dissection

In 2011, at the middle of the 4th instar, 40 midguts were taken from Quebec, Ontario and BC individuals on the trembling aspen treatment for a total of 120 aspen-fed midguts. In addition, 40 midguts were taken from Ontario individuals on the maple diet (the other populations did not show high enough survival on maple to permit sampling of midguts). Midguts were taken haphazardly from the different families. Insects were starved for a period of 24 hours prior to dissection. Dissection trays were sterilized between each insect. Each insect was anesthetized via freezing at -20 degrees Celsius for 20 minutes prior to dissection. Forceps were used to grip both ends of the insect. Careful pulling separated the head and midgut from the remainder of the caterpillar. Heads were severed from the midguts using sterile scissors and midguts were frozen at -80 degrees Celsius for microbial analysis. Midguts were not washed or emptied. In 2012, when only Ontario and BC individuals were available, 20 midguts were collected from 3rd, 4th and 5th instar Ontario individuals on both diets for a total of 120 midguts. Likewise, 20 midguts were collected from 3rd, 4th and 5th instar BC individuals on the Aspen diet and the 3rd and 4th instar BC individuals on the Maple diet, for a total of 100 midguts. No BC individuals reached the 5th instar on the Maple diet.

2.2.4 Culturing of midgut bacteria

In 2012, following dissection, 3 midguts from each region and treatment combination were homogenized in 200μL of Milli-Q ultrapure water and streaked on a variety of media. Four types of media were used for the isolation of midgut bacteria: Tryptic Soy

Agar (TSA), pH 7; TSA, pH 10; Maple Mash (MM), pH 7; MM pH, 10. TSA was used as it is a generally nutritious media, MM was created from puréed and sterilized maple foliage in the case that gut bacteria grew exclusively on sugar maple plant material. Maple foliage was surface-sterilized with ethanol prior to homogenization. Ethanol was air-dried before blending the leaves with some distilled water. The mixture was allowed to sit for two days before combination with agar. The alkaline versions of the media were created in light of the fact that *M. disstria* midguts have a naturally high pH. Pure cultures were obtained from all growth conditions. In total, 40 strains were isolated and stored in glycerol stocks.

2.2.5 DNA Extraction and bacterial 16S rRNA gene amplification by polymerase chain reaction (PCR)

Amplification of the 16S rRNA genes from the pure cultures was accomplished using colony PCR, the PCR products were sequenced using the Sanger sequencing service at the McGill Innovation Center. Sequences were identified by sequence similarity searches against the SILVA 16S rRNA database.

For culture-independent methods, different DNA extraction methods were attempted to obtain microbial DNA from the midguts. First, following sonication of the midgut tissue, DNA was extracted using a method similar to that outlined by Broderick (2004), with the following modifications. After adding 10% CTAB-5M NaCl, MPC (Master Pure Complete precipitating reagent) was used to clean the DNA and precipitate a protein pellet. DNA was then precipitated with isopropanol and resuspended in TE buffer.

DNA concentration was assessed by a nanospectrophotometer and samples were stored at -20 ^{o}C until PCR amplification. A gel was run with 1 uL of DNA against a 1kb ladder to ensure genomic DNA was obtained. Genomic DNA was obtained, however, following PCR, it seemed as if some PCR inhibitors remained in the extract, making amplification of the 16S rRNA gene problematic. A subsequent attempt was made to obtain DNA using a PowerLyzer Powersoil DNA Isolation bead beating kit from MO-BIO, but met with similar results. Extractions performed using a 25:24:1 solution of phenol:chloroform:isoamyl alcohol and a 24:1 solution of chloroform:isoamyl alcohol were successful. The high lipid content of the caterpillar gut appeared to be the problem and will inhibit PCR if not removed. This was resolved by removing it with chloroform. 16S rRNA was amplified using primers for the V5 region in combination with PHIRE enzyme. PCR products were obtained without issue.

2.2.6 Sequencing

Sequencing was accomplished using the Ion Personal Genome Machine Sequencer. Unfortunately, as the use of this machine is quite costly and shared with others, there was only space for two midgut samples; a 4th instar Ontario individual on the Aspen treatment from 2011 and a 5th instar Ontario individual on the Maple treatment from 2012.

2.3 Results

2.3.1 Insect Rearing 2011

Problems arose during this research season with the poor performance of individuals on the maple treatment. The deaths of all Quebec and BC individuals prior to pupation meant it was impossible to observe any post-pupation differences such as mate choice, oviposition or pupal weight. BC individuals experienced the highest mortality on the maple diet; all having died before the fifth instar. Quebec individuals were more successful in the fifth instar on the maple diet (over 50% survived to this stage), but the individuals that remained after midgut harvesting all died before pupation. Ontario individuals had a much stronger showing in the fifth instar on the maple diet (Figure 10).

Populations obtained from BC, Ontario and Quebec all performed similarly on trembling aspen foliage (Figure 11). Pupal weights were an average of 305.66 mg ± 97.63 mg, 318.62 mg ± 90.02 mg and 389.67 mg ± 119.89 mg for Quebec, BC and Ontario egg masses, respectively (N = 165, 172 and 160). Figure 10, on the other hand, displays differential success on sugar maple foliage based solely on where the egg masses were collected. This confirms that there is something fundamentally different in Ontario populations prior to hatching. We attempted to extract DNA from the guts and amplify bacterial 16S rRNA gene sequences in order to assess the composition of bacterial communities within FTC guts.

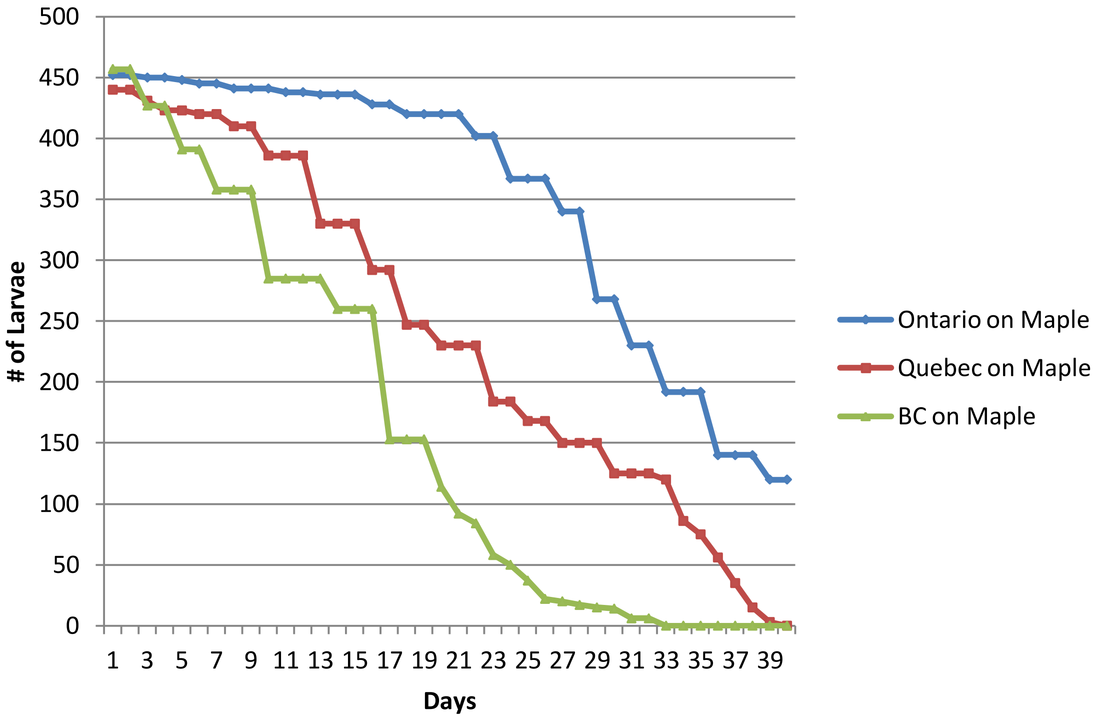


Figure 10 - Larval survival on the 2011 maple treatment starting at the hatching date and ending when the last individual pupated on day 40.

Initially, culture-independent methods performed in 2011 did not yield good-quality PCR product. This was possibly due to presence of natural PCR inhibitors, such as lipids, as well as the fact that the bacterial content of the guts is very low compared to the presence of chloroplast DNA. Despite extracting midguts from caterpillars in un-fed states, much plant material seems to have remained (this point was confirmed once culture-independent methods succeeded as many chloroplast sequences appeared in the results).

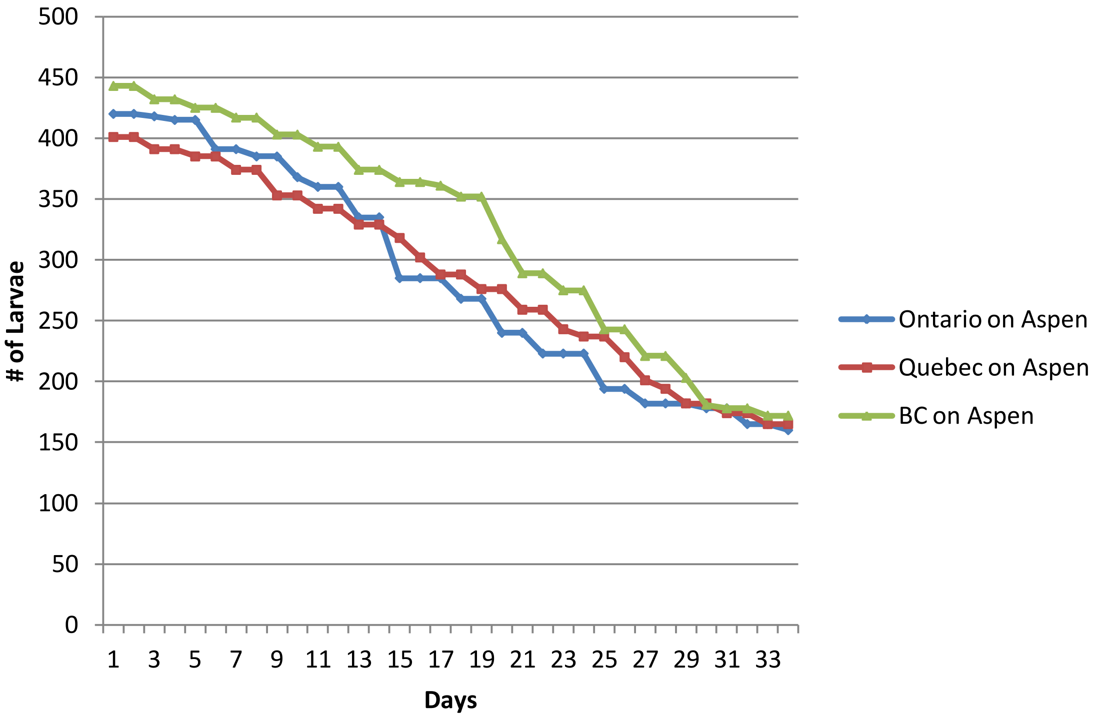


Figure 11 - Larval survival on the 2011 aspen treatment starting at the hatching date and ending when the last individual pupated on day 34.

2.3.2 Abundance and taxonomy of culturable FTC gut bacteria

Overall, a low number of colonies appeared (approximately 25-100 colony forming units (CFUs) for three guts combined) on all plates despite the large amount of gut material plated. Mostly, two morphotypes appeared on the plates independent of the media type. The remaining colonies that rarely appeared were not identified. 16S rRNA gene sequencing of colonies grown in 2012 identified the two most abundant strains, *Pseudomonas putida* and a *Pantoea* species. These two were previously identified in *Lymantria dispar* and *Pieris rapae* respectively (Broderick et al., 2009). *Pantoea agglomerans* was also identified as a member of the *L. dispar* midgut microbiome in Broderick et al. (2004). Of particular interest is the facultative anaerobe *Pantoea agglomerans*, found in our FTC guts. Not only is a facultative anaerobe ideal for residence in a midgut, but the genus *Pantoea* is also known for tannase production, e.g. *Pantoea sp. At-9b* (NCBI WP_013512920.1). At this point in the experiment it was questioned whether or not there were other bacteria present in the gut that are difficult to obtain cultures for, perhaps due to obligate anaerobic metabolism. In order to properly characterize the midgut microbiome of *M. disstria,* either anaerobic cultures needed to be made or the culture-independent molecular methods must be made to work, making use of the frozen midguts collected in both 2011 and 2012.

2.3.3 Culture-indepenent 16S rRNA sequencing of midgut commuities

With the revelation of using CIA and phenol-CIA to eliminate PCR inhibiting lipids, the opportunity arose to sequence two midgut samples. One midgut was a 4th instar Ontario individual on the aspen diet from 2011, the other was a 5th instar Ontario individual on the maple diet from 2012. In all, 15 species were identified in both the 4th and 5th instar midgut, however *Campylobacter sp.* and *Spiroplasma sp.* are missing in the 4th instar midgut while *Leucobacter sp.* and *Staphylococcus sp.* were not found in the 5th instar midgut. A list of identified bacteria are shown in Table 5, while Table 6 shows the percentage of sequences from each genus in both samples. In Table 6, chloroplasts have been removed as they originally accounted for 42.9% of sequences in each sample. This high chloroplast count was most likely due to leftover plant matter in the midguts despite the starvation period prior to dissection.

In some cases, sequences matched highly (query coverage = 100%, identity ≥ 99%) to those of many bacteria isolated from different host species. A list of known hosts to bacteria similar to those found in FTC can be seen in Table 8. *Lymantria dispar* and *P. rapae* do not appear under *Pantoea agglomerans*, despite evidence of those species hosting the bacteria, beacause Table 7 is based purely on sequence information available through NCBI.

Table 5 - FTC gut flora identified through culture-independent methods

4th Instar Ontario Larva on Aspen (2011)	5th Instar Ontario Larva on Maple (2012)	Lifestyle
Bacillus sp.	*Bacillus sp.*	Facultative Anaerobe
Staphylococcus sp.	---	Facultative Anaerobe
---	*Spiroplasma sp.*	Anaerobe
Lactobacillus sp.	*Lactobacillus sp.*	Facultative Anaerobe
Pantoeae agglomerans	*Pantoeae agglomerans*	Facultative Anaerobe
---	*Campylobacter sp.*	Microaerophilic
Bradyrhizobium sp.	*Bradyrhizobium sp.*	Facultative Anaerobe
Xanthomonas sp.	*Xanthomonas sp.*	Aerobe
Burkholderia sp.	*Burkholderia sp.*	Aerobe
Acidobacteria sp.	*Acidobacteria sp.*	Unknown
Sphingobacteria sp.	*Sphingobacteria sp.*	Aerobe
Pseudomonas putida	*Pseudomonas putida*	Aerobe
Leucobacter sp.	---	Aerobe
Comamonas sp.	*Comamonas sp.*	Obligate Aerobe
Alcaligenes sp.	*Alcaligenes sp.*	Aerobic, Potentially Anaerobic
Methylobacterium sp.	*Methylobacterium sp.*	Obligate Aerobe
Propionibacterium sp.	*Propionibacterium sp.*	Anaerobic

Table 6 - Proportions of FTC gut flora sequences in each sample.

Genera	% of Sequences in 4th Instar Ontario Larva on Aspen (2011)	% of Sequences in 5th Instar Ontario Larva on Maple (2012)
Acidobacteria sp.	8.8	12.6
Alcaligenes sp.	3.4	2.4
Bacillus sp.	25.3	24.4
Bradyrhizobium sp.	1.9	6.3
Burkholderia sp.	3.8	2.4
Campylobacter sp.	---	1.6
Comamonas sp.	9.6	14.2
Pantoea agglomerans	1.5	1.6
Lactobacillus sp.	29.1	14.2
Leucobacter sp.	1.1	---
Methylobacterium sp.	0.8	1.6
Propionibacterium sp.	4.2	7.9
Pseudomonas putida sp.	6.1	2.4
Sphingobacteria sp.	1.9	4.7
Spiroplasma sp.	---	3.1
Staphylococcus sp.	1.5	---
Xanthomonas sp.	0.8	0.8

2.4 Discussion

Species belonging to the following genera known to produce tannase have been identified in the guts of two FTC midguts from the same location in Ontario from two separate years: *Methylobacterium, Bradyrhizobium, Burkholderia, Xanthomonas, Acidobacterium, Pantoea,* and *Pseudomonas* are all genera with tannase-producing species as discussed in Chapter 1. While it is not clear whether the strains observed do produce tannase, there exists potential for one or more of these bacteria to be responsible for tannase-production within the FTC gut, enhancing its survival on sugar maple foliage.

In addition to tannase, residents of the FTC midgut flora may influence many aspects of the larval stage. Any bacterial influence over the level of antioxidants in the midgut

Table 7 - A list of other known hosts of bacteria similar to those identified in our FTC, obtained via NCBI (query coverage = 100%, identity ≥ 99%)

Genera	Hosts containing similar species
Spiroplasma sp.	*Acyrthosiphon pisum* (pea aphid, homoptera)
	Agathemera claraziana (walking stick)
	Anisosticta novemdecipunctata (coccinellidae)
	Antonina crawii (homoptera)
	Araneus diadematus (european garden spider)
	Chrysolina varians (leaf beetle)
	Curculio alvobittatus (weevil)
	Curculio elephas (weevil)
	Curculio glandium (weevil)
	Curculio sikkimensis (coleoptera)
	Danaus chrysippus (butterfly)
	Drosophila ananassae (fly)
	Drosophila atripex (fly)
	Fannia manicata (little housefly)
	Harmonia axyridis (Coleoptera, coccinellidae)
	Harpalus pennsylvanicus (beetle)
	Hepialus gonggaensis (ghost moth larvae)
	Laodelphax striatellus (planthopper, homoptera)
	Meimuna mongolica (hemiptera, cicadidae)
	Meta mengei (spider)
	Meta segmentata (spider)
	Mosquito midgut
	Ostrinia zaguliaevi (moth)
	Puto albicans (hemiptera)
	Tetragnatha montana (spider)
	Tipula oleracea (cranefly)
	Unidentified spider
Bacillus sp.	*Sarcophagidae* (flesh fly)
	Monochamus galloprovincialis (coleoptera)
	Vermicompost
	Hysteroneura setariae (aphid)
	Rhynchophorus ferrugineus (red palm weevil)

	Cataglyphis adult ants
Staphylococcus sp.	*Toxoptera aurantii* (aphid)
Lactobacillus sp.	Cecal contents Mus musculus
	Pasture gramineae
	Rat cecum
	mouse cecum
	Piglet caecum
	Chicken cecum
	Echidna feces
	mesophilic anaerobic digester fed with brown water and food waste
Pantoea agglomerans	Termite gut
	Curculio koreanus (weevil)
	Termite gut
	Diamondback moth (*Plutella xylostella*)
	Plant root of *Atractyloades macrocephala*
	Maize rhizosphere soil
	Sphenophorus levis (curculionidae, sugarcane weevil)
	Rhizosphere soil
	Plutella xylostella (adult moth)
	Rhynchophorus ferrugineus
	Drosophila sp.
Campylobacter sp.	*Bemisia tabaci* clone Q (Silverleaf whitefly, homoptera)
Comamonas sp.	*Drosophila sp.*
	Fig Wasp
	Diabrotica virgifera (adult western corn rootworm)

would certainly ease the stress of oxidative damage from the alkaline break-down of tannin compounds.

The presence of *Xanthomonas* in FTC is perhaps the most interesting. These *Xanthomonas* sequences match most closely to *X. oryzae, X. campestris* and *X. axonopodix*, major producers of the tannase enzyme. Despite being aerobic bacteria, this does not exclude it from being a potential gut symbiont. *Xanthomonas* species have previously been shown in the midgut microbiome of the red palm weevil *Rhynchophorus ferrugineus* (Khiyami &

Alyamani, 2008). *Xanthomonas axonopodis* is also known to inhabit the midgut of the Lepidopteran *Phyllocnistis citrella* Stainton, relying on the larvae for distribution (Chagas et al., 2001). As plant pathogens that can neutralize plant defences, the activity of *Xanthomonas* could be helpful to herbivorous insects. The match for the *Burkholderia* sequences is quite confusing as it matches several different 16S sequences with 100% query coverage and identity. Further investigation is required to determine the exact species identity, however, *B. multivorans* and *B. cenocepacia* are both tannase-producers. The bacteria *Lactobacillus sp.* was somewhat expected in FTC given the presence of another lactic acid producing bacteria, *Lactococcus lactis,* in the butterfly *Vanessa cardui* (Broderick et al., 2009). Tannase-producing *Lactobacillus* were also suggested as the reason Japanese wood mice were found to tolerate acorn tannins (Shimada, 2006).

In addition to tannase, the identified bacteria may have other benefits. The *Pantoea* 16S sequence identified in FTC most closely matches that of *P. agglomerans strain A9* isolated from the termite gut. Termite guts are known to harbor multiple specis of bacteria dedicated to degrading plant material, especially wood. The Xylanase used by termites and produced in *P. agglomerans* (Genbank: ELP24757.1) may also be of some use to FTC given its ability to break down the hemicellulose in plant cell walls. A *Pantoea* sp. has also been identified in the Lepidopteran *Pieris rapae* (Broderick et al., 2009). *Pseudomonas putida* may be of use to FTC given its ability to degrade organic solvents (Marques and Ramos, 1993). Its use in bioremediation could have potential benefits for FTC, neutralizing toxins and pesticides. *Pseudomonas putida* is also known to inhabit the midgut of *Lymantria dispar* (Broderick et al., 2009). *Alcaligenes sp.* sequences were abundant in the results. Their dominance in the FTC gut makes sense given their alkaline tolerance. Of particular interest is their ability to produce non-standard amino acids (Brock Biology of Microorganisms, 13th ed.), potentially supplementing the caterpillar diet.

For other identified bacteria, it is unclear whether or not they may provide some benefit to FTC, however, they have been identified in other insect species. The *Staphylococcus* species is a close relative of the *Staphylococcus sp. Cb6* within the aphid *Toxoptera aurantii* and *S. sciuri strain JS-3* obtained from the soil. This was only present in the 2011 sample and perhaps just a temporary infection from the phyllosphere, however, several *Staphylococcus* species are known to inhabit the gut of *L. dispar* (Broderick et al., 2004). *Bradyrhizobium* is known to form symbiotic relationships with many plants. In FTC, the *Bradyrhizobium* species most resembles *B. sp. Ppaniculata_4* and *B. japonicum* associated with a plants and rhizospheres. It is unclear whether or not the *Bradyrhizobium* is a regular member of the gut flora or an inocculation from the phyllosphere. As a facultative anaerobe, it could survive in the midgut. Acidobacteria are a relatively new phylum extremely abundant in soils, yet they rarely show up in cultures. The *Acidobacteria sp.* identified here shares 16S identity with uncultured soil *Acidobacteria,* most notably including a sample from a trembling aspen rhizosphere, one of the two host trees in this study. It is difficult to imagine what type of interaction may exist between this bacteria and FTC, especially given the mysterious nature of this new phylum. The *Sphingobacteria sp.* is closely related to *Sphingobacteria* among Bacteroidetes isolated from forest soils, as well as an uncultured bacterium from poplar plantation soils. This bacteria, as well as *Leucobacter* appear to be common soil residents and only present in the 2012 and 2011 gut samples respectively. Pinto-Tomas et al. (2011) identified *Methylobacterium sp., Bradyrhizobium sp.* and *Propionibacterium sp.* in many tropical caterpillars (Saturniidae, a large Lepidopteran family), all genera that are present in FTC. The *Campylobacter* species closely matches an uncultered secondary endosymbiont from the silverleaf whitefly. It appears to only be present in the 2012 gut, however, competition with the *Bacillus sp.* could account for its reduced visibility in the 2011 sample. Lastly, the *Comamonas sp.* matches the 16S of bacteria associated with several insects. Most notable of these are the Uncultured *Comamonas* and *Comamonas oodontitis* from the corn rootworm and termite respectively.

The gut profile described here for FTC shares some similarities with the gut profile of the wood-feeding huhu beetle larvae of *Prionoplus reticularis* described by Reid et al. (2011). Reid et al. (2011) described a community composed of *Acidobacteria, Xanthomonas, Acetobacteria, Burkholderia* and *Enterobacter*, all of which are present in the FTC samples except for *Acidobacteria* (*Pantoea agglomerans* was formerly referred to as *Enterobacter agglomerans*). The similarities between the two insects, both having to degrade plant matter in similar environments, could account for the similar microbial load in each gut.

The *Bacillus sp.*, *Pantoea agglomerans* and *Pseudomonas putida* bacteria present in the midgut were expected given their appearance in past studies on Lepidopteran gut microbiomes (Broderick et al., 2004 & 2009). Furthermore, *Pantoea agglomerans* and *Pseudomonas putida* were identified using both culture and culture-independent methods. The closest match for our *Bacillus* sequences were *Bacillus cereus* and *Bacillus thuringiensis*. Given the strength of the matches it is difficult to say which species are actually present. If it is *Bacillus thuringiensis* this would imply that FTC may be carriers of the Cry toxin as is the case with many other Lepidopterans. Kouassi et al. (2001) discovered that FTC is 100-fold more susceptible to *Bacillus thuringiensis* while feeding on sugar maple rather than aspen. However, the presence of *Bacillus cereus* would also be interesting given that a *Campylobacter sp.* is also present and the two species are known to compete with each other. As far as tannase production, members of the genus *Bacillus* are known to possess the gene but not specifically the species identified.

The presence of male-killing *Spiroplasma* in FTC may contribute to speciation by distorting the sex ratios and influencing a change in sexual selection as seen with *Wolbachia* species (Kageyama et al., 2012). The *Spiroplasma sp.* identified matches an overwhelming number of *Spiroplasma* species acting as male-killing endosymbionts. *Spiroplasma*, present in a number of insect orders including Lepidoptera, is also shown in a variety of arachnids. This makes sense as the origin of male-killing bacteria are said to be derived from arthropod-associated bacterial clades (Hurst and Jiggins, 2000). Male-killing bacteria typically infect a minority of females and are rarely seen in males at any developmental stage (Hurst and Jiggins, 2000), this could easily explain why no *Spiroplasma* was observed in the 2011 gut sample. It is possible that the 2011 gut was either a male or uninfected female. If *Spiroplasma* are in fact a regular member of FTC gut flora or even just a regional infection, this could have a major impact on the understanding of the FTC life cycle. Where competition exists among males for mating opportunities, the activity of a male-killing parasite could reverse the sex ratios of this situation, effectively increasing the proportion of females that fail to mate (Jiggins et al., 2000). Male-killing parasites acting very early on sibling males would definitely affect the outcome of a hatching egg mass, increasing resource access for females. The potential for speciation due to the altered sex ratios is a legitimate concern, as it has been observed from the similar action of *Wolbachia.* As discussed in Hurst and Jiggins (2000), the action of such bacteria tends to limit its own vertical transmission, however, if this *Spiroplasma* species were manipulated to be more widespread, it may have potential in terms of long-term pest control. If it is confirmed to act as a male-killing parasite within FTC, it may serve as a desirable alternative to pesticides. However, as with any biological control, the side-effects of such action must be considered at length, especially given the possibility for accelerating speciation.

Another important observation to note from the colonies grown in 2012 was the presence of a possible *Streptomyces* species that failed to sequence following colony PCR. It bears a striking resemblance to *Streptomyces noursei,* and appeared in all midguts harvested from individuals affected by the nuclear polyhedrosis virus. It could be that they are unrelated and the virus creates conditions preferable for *Streptomyces* which arrives from the phyllosphere, or it may also be possible that the *Streptomyces* species creates an environment suitable for virus or is responsible for the virus itself. Jakubowska et al. (2013) demonstrated that baculovirus infections in the Lepidopteran *Spodoptera exigua* lead to increased gut microbial loads, and in turn, the higher gut microbial load may enhance the virulence, pathogenicity and dispersion of baculovirus.

The clear differential success of eastern FTC on the sugar maple diet suggests some fundamental difference between them and other members of the species in western Canada. Whether this difference is in the insect's own biology or based on the gut microbiome is unknown. The heavy exposure of eastern FTC to bacteria typical of the sugar maple phyllosphere may have led to the incorporation of some of the bacteria into FTC's microbiome. It is unknown whether the bacterial species identified in the Ontario midguts are also present in the midguts of geographically different populations, however, the bacteria that have been identified suggest further study is justified. Crucially, midguts of western FTC must be explored for proper comparison.

For future work building microbe profiles of FTC based on geography, the methods should be different than used here. While the sequencing work done may have identified gut flora vertically transmitted to egg masses, some bacteria may merely be members of the phyllosphere that happened to be present in the midguts during sequencing. Firstly, future work should have lab setups at the chosen geographic locations so as to feed the resident FTC their local foliage, i.e. foliage from the same location the egg masses were laid. This eliminates the potentially problematic interactions of foreign phyllosphere bacteria with resident symbionts leading to spurious results. Due to competition, phyllosphere bacteria may eliminate vertically transmitted midgut bacteria and vice versa. Comparisons of gut flora between populations of FTC can be made much easier if the gut microbiome and microbiome of the respective environment are both clearly established and distinct for geographically distant populations. Secondly, in order to assess whether bacteria are vertically transmitted from an individuals parents or inoculations from the environment, FTC egg masses must be split and reared on either sterile artificial diet or local foliage.

Despite not having the opportunity, it is anticipated that sequencing FTC midguts from other geographic locations will reveal significantly different microbiomes, especially where tannase production is considered. Not only would this improve our understanding of a very important species, but the increase of microbiome samples in databases, such as at JGI, allow for much more precise metaphylogenomic analysis of genes. Increased sampling as well as more documented cases of symbiotic applications encourage a better understanding of the function of a gene as well as its global distribution. With regards to tannase much of its activity, especially in the insect world, seems yet to be explored.

References

Aguilar, C. N. & Gutierrez-Sanchez, G. (2001). Review: sources, properties, applications and potential uses of tannin acyl hydrolase. *Food Science & Technology International*, *7*(5): 373–382.

Aguilar, C. N., Rodríguez, R., Gutiérrez-Sánchez, G., Augur, C., Favela-Torres, E., Prado-Barragan, L. & Contreras-Esquivel, J. (2007). Microbial tannases: advances and perspectives. *Applied Microbiology and Biotechnology*, *76*(1): 47–59.

Allen, H. K., Cloud-Hansen, K. A., Wolinski, J. M., Guan, C., Greene, S., Lu, S. & Handelsman, J. (2009). Resident microbiota of the gypsy moth midgut harbors antibiotic resistance determinants. *DNA and Cell Biology*, *28*(3): 109–117.

Banerjee, A., Jana, A., Pati, B., Mondal, K. & Mohapatra, P. (2012). Characterization of tannase protein sequences of bacteria and fungi: an *in silico* Study. *The Protein Journal*, *31*(4): 306–327.

Barbehenn, R., Cheek, S., Gasperut, A., Lister, E. & Maben, R. (2005). Phenolic compounds in red oak and sugar maple leaves have prooxidant activities in the midgut fluids of *Malacosoma disstria* and *Orgyia leucostigma* caterpillars. *Journal of Chemical Ecology*, *31*(5): 969–988.

Barbehenn, R. V. & Constabel, P.C. (2011). Tannins in plant-herbivore interactions. *Phytochemistry*, *72*(13): 1551–1565.

Barbehenn, R.V., Jaros, A., Lee, G., Mozola, C., Weir, Q. & Salminen, J.-P. (2009a). Hydrolyzable tannins as "quantitative defenses": limited impact against *Lymantria dispar* caterpillars on hybrid poplar. *Journal of Insect Physiology 55*: 297-304.

Barbehenn, R.V., Jaros, A., Lee, G., Mozola, C., Weir, Q. & Salminen, J.-P. (2009b). Tree resistance to *Lymantria dispar* caterpillars: importance and limitations of foliar tannin composition. *Oecologia 159(4)*: 777-788.

Barbehenn, R.V., Kochmanski, J., Menachem, B. & Poirier, L.M. (2013). Allocation of cysteine for glutathione production in caterpillars with different antioxidant defense strategies: a comparison of *Lymantria dispar* and *Malacosoma disstria*. *Archives of Insect Biochemistry and Physiology 84*: 90-103.

Barbehenn, R.V., Maben, R.E. & Knoester, J.J. (2008). Linking phenolic oxidation in the midgut lumen with oxidative stress in the midgut tissues of a tree feeding caterpillar *Malacosoma disstria* (Lepidoptera: Lasiocampidae). *Environmental Entomology 37*: 1113-1118.

Barbehenn, R.V. & Martin, M.M. (1994). Tannin sensitivity in larvae of *Malacosoma disstria* (Lepidoptera): Roles of the peritrophic envelope and midgut oxidation. *Journal of Chemical Ecology 20*: 1985-2001.

Barbehenn, R. V., Walker, A. C. & Uddin, F. (2003). Antioxidants in the midgut fluids of a tannin-tolerant and a tannin-sensitive caterpillar: effects of seasonal changes in tree leaves. *Journal of Chemical Ecology 29(5):* 1099-1116.

Belmares, R., Contreras-Esquivel, J. C., Rodrı́guez-Herrera, R., Coronel, A. R. & Aguilar, C. N. (2004). Microbial production of tannase: an enzyme with potential use in food industry. *Food Science and Technology*, *37*(8), 857–864.

Belur, P.D. & Mugeraya, G. (2011). Microbial production of tannase: State of the art. *Research Journal of Microbiology, 6:* 25-40.

Benoit, I., Danchin, E. G. J., Bleichrodt, R.-J. & de Vries, R. P. (2008). Biotechnological applications and potential of fungal feruloyl esterases based on prevalence, classification and biochemical diversity. *Biotechnology Letters*, *30*(3): 387–396.

Bik, H., Halanych, K., Sharma, J. & Thomas, W. (2012). Dramatic shifts in benthic microbial eukaryote communities following the deepwater horizon oil spill. *PLoS ONE 7*(6), e38550.

Brady, J.P., Richmond, R.C. & Oakeshott, J.G. (1990). Cloning of the esterase-5 locus from *Drosophila pseudoobscura* and comparison with its homologue in *Drosophila melanogaster*. *Molecular Biology and Evolution 7*: 525-546.

Broderick, N, Raffa, K., Goodman, R. & Handelsman, J. (2004). Census of the bacterial community of the gypsy moth larval midgut by using culturing and culture-independent methods. *Applied and Environmental Microbiology*, *70*(1): 293–300.

Broderick, N., Robinson, C., McMahon, M., Holt, J., Handelsman, J. & Raffa, K. (2009). Contributions of gut bacteria to *Bacillus thuringiensis*-induced mortality vary across a range of Lepidoptera. *BMC Biology 7*(1): 11.

Bunzel, M., Ralph, J. & Steinhart, H. (2005). Association of non-starch polysaccharides and ferulic acid in grain amaranth (*Amaranthus caudatus L.*) dietary fiber. *Molecular Nutrition & Food Research*, *49*(6): 551–559.

Carter, P. & Wells, J. A. (1988). Dissecting the catalytic triad of a serine protease. *Nature*, *332*(6164): 564–568.

Caspi, R., Foerster, H., Fulcher, C.A., Kaipa, P., Krummenacker, M., Latendresse, M., Paley, S., Rhee, S.Y., Shearer, A.G., Tissier, C., Walk, T.C., Zhang, P. & Karp, P.D. (2012) The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Research 36*: 623-631

Chagas, M.C.M., Parra, J.R.P., Namekata, T., Hartung, J.S. & Yamamoto, P.T. (2001). *Phyllocnistis citrella* Stainton (Lepidoptera: Gracillariidae) and its relationship with the citrus canker bacterium *Xanthomonas axonopodis* pv citri in Brazil. *Sociedade Entomologica do Brasil 30(1).*

Chaparro, C., Maurer, J., Gutierrez, C., Krajden, M., Chan, C., Winton, T., Keshavjee, S., Scavuzzo, M., Tullis, E., Hutcheon, M. & Kesten, S. (2001). Infection with *Burkholderia cepacia* in Cystic Fibrosis. *American Journal of Respiratory and Critical Care Medicine 163(1)*: 43-48.

Cooke, B.J. & Lorenzetti, F. (2006). The dynamics of forest tent caterpillar outbreaks in Quebec, Canada. *Forest Ecology and Management 226(1-3):* 110-121.

Crepin, V. F., Faulds, C. B. & Connerton, I. F. (2004). Functional classification of the microbial feruloyl esterases. *Applied Microbiology and Biotechnology*, *63*(6): 647–652.

Degnan, P. H., Yu, Y., Sisneros, N., Wing, R. A., & Moran, N. A. (2009). *Hamiltonella defensa*, genome evolution of protective bacterial endosymbiont from pathogenic ancestors. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(22): 9063–9068.

Felsenstein, J. (1985). Confidence Limits on Phylogenies: An Approach Using the Bootstrap. *Evolution*, *39*(4): 783.

Fitzgerald, T. D. (1995). *The Tent Caterpillars*. Cornell University Press. Ithaca, NY.

Fortin, M. and Mauffette, Y. (2002) The suitability of leaves from different canopy layers for a generalist herbivore (Lepidoptera: Lasiocampidae) foraging on sugar maple. *Canadian Journal of Forest Research,* 32(3): 379-389.

Fukunaga, Y., Kurahashi, M., Yanagi, K., Yokota, A. & Harayama, S. (2008). *Acanthopleuribacter pedis* gen. nov., sp. nov., a marine bacterium isolated from a chiton, and description of Acanthopleuribacteraceae fam. nov., Acanthopleuribacterales ord. nov., Holophagaceae fam. nov., Holophagales ord. nov. and Holophagae classis nov. in the phylum "Acidobacteria". *International Journal of Systematic and Evolutionary Microbiology*, *58*(Pt 11): 2597–2601.

Grigoriev, I.V., Nordberg, H., Shabalov, I., Aerts, A., Cantor, M., Goodstein, D., Kuo, A., Minovitsky, S., Niktin, R., Ohm, R.A., Otillar, R., Poliakov, A., Ratnere, I., Riley, R., Smirnova, T., Rokhsar, D. & Dubchak, I. (2011). The Genome Portal of the Department of Energy Joint Genome Institute. *Nucleic Acids Research 42:* 1-7

Grisdale, D. G. "*Malacosoma disstria*" In P. Singh & R. F. Moore (Eds.) *Handbook of Insect Rearing* (1985) Volume 2: 369-379. Elsevier. New York, NY.

Hagerman, A.E. (2002). Tannin Handbook. *Miami University, Oxford OH 45056* www.users.muohio.edu/hagermae/

Hildahl, V. & Reeks, W. A. (1960). Outbreaks of the forest tent caterpillar, *M. disstria* Hbn., and their effects on stands of trembling aspen in Manitoba and Saskatchewan. *Canadian Entomology* 92: 199-209.

Hermoso, J., Sanz-Aparicio, J., Molina, R., Juge, N., González, R. & Faulds, C. (2004). The crystal structure of feruloyl esterase A from *Aspergillus niger* suggests evolutive functional convergence in feruloyl esterase family. *Journal of Molecular Biology*, *338*(3): 495–506.

Horvarth, P. J. (1981). The nutritional and ecological significance of *Acer*-tannins and related polyphenols. *M. S. Thesis, Cornell University, Ithaca, NY.*

Hurst, G. D. & Jiggins, F. M. (2000). Male-killing bacteria in insects: mechanisms, incidence, and implications. *Emerging Infectious Diseases*, *6*(4): 329–336.

Ishii, T. (1997). Structure and functions of feruloylated polysaccharides. *Plant Science*, *127*(2): 111–127.

Ivanova, N., Gronow, S., Lapidus, A., Copeland, A., Glavina Del Rio, T., Nolan, M. & Klenk, H.-P. (2009). Complete genome sequence of *Leptotrichia buccalis* type strain *(C-1013-b). Standards in Genomic Sciences*, *1*(2): 126–132.

Jakubowska, A.K., Vogel, H. & Herrero, S. (2013). Increase in gut microbiota after immune suppression in Baculovirus-infected larvae. *PLoS Pathogens 9(5)*: e1003379.

Janson, E. M., Stireman, J. O., Singer, M. S. & Abbot, P. (2008). Perspective: phytophagous insect-microbe mutualisms and adaptive evolutionary diversification. *Evolution* 62: 997–1012.

Jiggins, F. M., Hurst, G. D. D. & Majerus, M. E. N. (2000). Sex-ratio-distorting *Wolbachia* causes sex-role reversal in its butterfly host. *Proceedings of the Royal Society B: Biological Sciences 267(1438):* 69–73.

Jones, D. T., Taylor, W. R. & Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Computer Applications in the Biosciences : CABIOS*, *8*(3): 275–282.

Kageyama, D., Narita S. & Watanabe M. (2012). Insect sex determination manipulated by their endosymbionts: indices, mechanisms and implications. *Insects 3*: 161-199.

Kampa, M., Alexaki, V.-I., Notas, G., Nifli, A.-P., Nistikaki, A., Hatzoglou, A. & Castanas, E. (2004). Antiproliferative and apoptotic effects of selective phenolic acids on T47D human breast cancer cells: potential mechanisms of action. *Breast Cancer Research : BCR*, *6*(2): R63–74.

Kanehisa, M. & Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research 28:* 27-30.

Khiyami, M. & Alyamani, E. (2008). Aerobic and facultative anaerobic bacteria from gut of red palm weevil (*Rhynchophorus ferrugineus). African Journal of Biotechnology 7(10)*: 1432-1437.

Kikuchi, Y., Meng, X.-Y. & Fukatsu, T. (2005). Gut symbiotic bacteria of the genus *Burkholderia* in the broad-headed bugs *Riptortus clavatus* and *Leptocorisa chinensis* (Heteroptera: Alydidae). *Applied and Environmental Microbiology*, *71*(7): 4035–4043.

Kouassi, K. C., Lorenzetti, F., Guertin, C., Cabana, J. & Mauffette, Y. (2001). Variation in the susceptibility of the forest tent caterpillar (Lepidoptera: Lasiocampidae) to *Bacillus thuringiensis* variety *kurstaki HD-1*: effect of the host plant. *Journal of Economic Entomology*, *94*(5): 1135–1141.

Lastovica, A. J. (2009). Clinical relevance of *Campylobacter concisus* isolated from pediatric patients. *Journal of Clinical Microbiology*, *47*(7): 2360.

Lekha, P. K. & Lonsane, B. K. (1994). Comparative titres, location and properties of tannin acyl hydrolase produced by *Aspergillus niger PKL 104* in solid-state, liquid surface an submerged fermentations. *Process Biochemistry*, *29*(6): 497–503.

Lekha, P.K. & Lonsane, B.K. (1997). Production and application of tannin acyl hydrolase: state of the art. *Advances in Applied Microbiol*ogy 44: 215–260.

Letunic, I. & Bork, P. (2007). Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics 23(1):* 475-478.

Letunic, I. & Bork, P. (2011) Interactive Tree of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Research 39:* 1-4.

Liu, N., Yan, X., Zhang, M., Xie, L., Wang, Q., Huang, Y. & Zhou, Z. (2011). Microbiome of fungus-growing termites: a new reservoir for lignocellulase genes. *Applied and Environmental Microbiology*, *77*(1): 48–56.

Lorenzetti, F. (1993). Performances Relatives de la Livrée des Forêts *Malacosoma disstria* Hbn. sur L'érable à Sucre *Acer saccharum* Marsh. Sain et Dépéri et sur le Peuplier Faux-tremble *Populus tremuloïdes* Michx. en Relation avec la Chimie Foliaire. *MSc thesis, Université du Québec à Montréal.*

Lorenzetti, F., Mauffette, Y. & Bauce, É. (1998). Relationship between foliar chemistry and insect performance: the forest tent caterpillar. *Sugar Maple Ecology and Health: An International Symposium Hosted By The Northeastern Research Station, USDA Forest Service, 2-4 June, Warren, Pennsylvania, USA.*

Lu, M.-J. & Chen, C. (2007). Enzymatic modification of tannase increase the antioxidant activity of green tea. *Food Research International 41*: 130-137.

Macedo, G. A. & Battestin, V. (2007). Effects of temperature, pH and additives on the activity of tannase produced by *Paecilomyces variotii*. *Electronic Journal of Biotechnology*, *10*(2): 191–199.

Madigan, M. T., Stahl, D. & Martinko, J. M. (2011) *Brock Biology of Microorganisms*. Benjamin Cummings.

Marqués, S. & Ramos, J. L. (1993). Transcriptional control of the *Pseudomonas putida* TOL plasmid catabolic pathways. *Molecular microbiology*, *9*(5): 923–929.

Moran, N. A., Tran, P. & Gerardo, N. M. (2005). Symbiosis and insect diversification: an ancient symbiont of sap-feeding insects from the bacterial phylum Bacteroidetes. *Applied and Environmental Microbiology*, *71*(12): 8802–8810.

Moya, A., Peretó, J., Gil, R. & Latorre, A. (2008). Learning how to live together: genomic insights into prokaryote-animal symbioses. *Nature Reviews. Genetics*, *9*(3): 218–229.

Murphy, B. L., Chretien, A. R., Jayne, L., Ramsey, D. & Brown, L. J. (2012). Non-timber forest products, maple syrup and climate change. *Journal of Rural and Community Development,* 7(3): 42-64.

Nicol, R. W., Arnason, J. T., Helson, B. & Abou-Zaid, M. M. (1997). Effect of host and nonhost trees on the growth and development of the forest tent caterpillar *M. disstria* (Lepidoptera: Lasiocampidae). *Canadian Entomology, 129*: 991-999.

Ohkuma, M. & Kudo, T. (1996). Phylogenetic diversity of the intestinal bacterial community in the termite *Reticulitermes speratus*. *Applied and Environmental Microbiology*, *62*(2): 461–468.

Osawa, R., Kuroiso, K. & Goto, S. (2000). Isolation of tannin-degrading lactobacilli from humans and fermented foods. *Applied and Environmental Microbiology,* 66(7) 3093-3097.

Parry, D. & Goyer, R. A. (2004). Variation in the suitability of host tree species among geographically discrete populations of forest tent caterpillar. *Environmental Entomology,* 33: 1477-1487.

Pinto-Tomás, A. A., Sittenfeld, A., Uribe-Lorío, L., Chavarría, F., Mora, M., Janzen, D. H. & Simon, H. M. (2011). Comparison of midgut bacterial diversity in tropical caterpillars (Lepidoptera: Saturniidae) fed on different diets. *Environmental Entomology*, *40*(5): 1111–1122.

Pruitt, K., Brown, G., Tatusova, T. & Maglott, D. (2012). The NCBI handbook: chapter 18, the reference sequence (RefSeq) database. http://www.ncbi.nlm.nih.gov/books/NBK21091/

Quevillon E., Silventoinen V., Pillai S., Harte N., Mulder N., Apweiler R. & Lopez R. (2005). InterProScan: protein domains identifier. *Nucleic Acids Research 33 (Web Server issue): W116-W120*

Ramirez-Coronel, M. A., Marnet, N., Kolli, V. S. K., Roussos, S., Guyot, S. & Augur, C. (2004). Characterization and estimation of proanthocyanidins and other phenolics in coffee pulp (*Coffea arabica*) by thiolysis-high-performance liquid chromatography. *Journal of Agricultural and Food Chemistry*, *52*(5): 1344–1349.

Rawat, S. R., Männistö, M. K., Starovoytov, V., Goodwin, L., Nolan, M., Hauser, L. & Häggblom, M. M. (2012). Complete genome sequence of *Terriglobus saanensis* type strain SP1PR4(T), an Acidobacteria from tundra soil. *Standards in Genomic Sciences*, *7*(1): 59–69.

Reid, N. M., Addison, S. L., Macdonald, L. J. & Lloyd-Jones, G. (2011). Biodiversity of active and inactive bacteria in the gut flora of wood-feeding huhu beetle larvae (*Prionoplus reticularis*). *Applied and Environmental Microbiology*, *77*(19): 7000–7006.

Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, *4*(4): 406–425.

Salminen, J.-P. & Lempa, K. (2002). Effects of hydrolysable tannins on a herbivorous insect: fate of individual tannins in insect digestive tract. *Chemoecology 12*: 203-211.

Sasaki, E., Shimada, T., Osawa, R., Nishitani, Y., Spring, S. & Lang, E. (2005). Isolation of tannin-degrading bacteria isolated from feces of the Japanese large wood mouse, *Apodemus speciosus*, feeding on tannin-rich acorns. *Systematic and Applied Microbiology*, *28*(4): 358–365.

Sayle, R. & Milner-White, J. E. (1995) RasMol: Biomolecular graphics for all. *Trends in Biochemical Sciences (TIBS) 20(9)*: 374

Sharma, S. & Gupta, M.N. (2003) Synthesis of antioxidant propyl gallate using tannase from *Aspergillus niger van Teighem* in nonaqueous media. *Bioorganic & Medical Chemistry Letters 13(3)*: 395-397

Shimada, T. (2006). Salivary proteins as a defense against dietary tannins. *Journal of Chemical Ecology, 32*: 1149-1163.

Sippell, W. L. (1962). Outbreaks of the forest tent caterpillar, *M. disstria* Hbn., a periodic defoliator of broad-leafed trees in Ontario, *Canadian Entomology, 94:* 408-416.

Sittig, M. (1988). *Pharmaceutical Manufacturing Encyclopedia* (2nd Edition). Noyes Publications. Park Ridge, NJ.

Stehr, F. W. & Cook, E. F. (1968). *A revision of the genus Malacosoma* Hübner *in North America (Lepidoptera: Lasiocampidae): systematics, biology, immatures, and parasites*. Smithsonian Institution Press. Washington, DC.

Su, C., Lei, L., Duan, Y., Zhang, K.-Q. & Yang, J. (2012). Culture-independent methods for studying environmental microorganisms: methods, application, and perspective. *Applied Microbiology and Biotechnology*, *93*(3): 993–1003.

Swain, T. (1965). The tannins. *Plant Biochemistry* (1965): 552-580

Tamura, K., Stecher, G., Peterson, D., Filipski, A. and Kumar, S. (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution 30*: 2725-2729.

The Mendeley Support Team. (2011). Getting Started with Mendeley. *Mendeley Desktop*. London: Mendeley Ltd. Retrieved from http://www.mendeley.com

Trudeau, M., Mauffette, Y., Rochefort, S., Han, E. & Bauce, E. (2010). Impact of host tree on forest tent caterpillar performance and offspring overwintering mortality. *Environmental Entomology*, *39*(2): 498–504.

Udatha, D. B. R. K. G., Kouskoumvekaki, I., Olsson, L. & Panagiotou, G. (2011). The interplay of descriptor-based computational analysis with pharmacophore modeling builds the basis for a novel classification scheme for feruloyl esterases. *Biotechnology Advances*, *29*(1): 94–110.

Visotto, L. E., Oliviera, M. G. A., Guedes, R. N. C., Ribon, A. O. B., Good-God, P. I. V. (2009). Contribution of gut bacteria to digestion and development of the velvetbean caterpillar *Anticarsia gemmatalis*. *Journal of Insect Physiology*, 55: 185-191.

Wernegreen, J. J. (2002). Genome evolution in bacterial endosymbionts of insects. *Nature Reviews. Genetics*, *3*(11): 850–861.

Wijesinghe, W. A. J. P. & Jeon, Y.-J. (2011). Biological activities and potential cosmeceutical applications of bioactive components from brown seaweeds: a review. *Phytochemistry Reviews*, *10*(3): 431–443.

Wilson, K. H. & Blitchington, R. B. (1996). Human colonic biota studied by ribosomal DNA sequence analysis. *Applied and Environmental Microbiology*, *62*(7): 2273–2278.

Woese, C. R. & Fox, G. E. (1977) Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proceedings of the National Academy of Sciences USA* 74: 5088-5090.

Wong, D. W. S. (2006). Feruloyl esterase: a key enzyme in biomass degradation. *Applied Biochemistry and Biotechnology*, *133*(2): 87–112.

Yates, K.R. (1989). The feeding ecology of the gumboot chiton, *Cryptochiton stelleri* (Middendorff, 1846). *Ph.D. Thesis, Oregon State University, Corvallis, OR.*

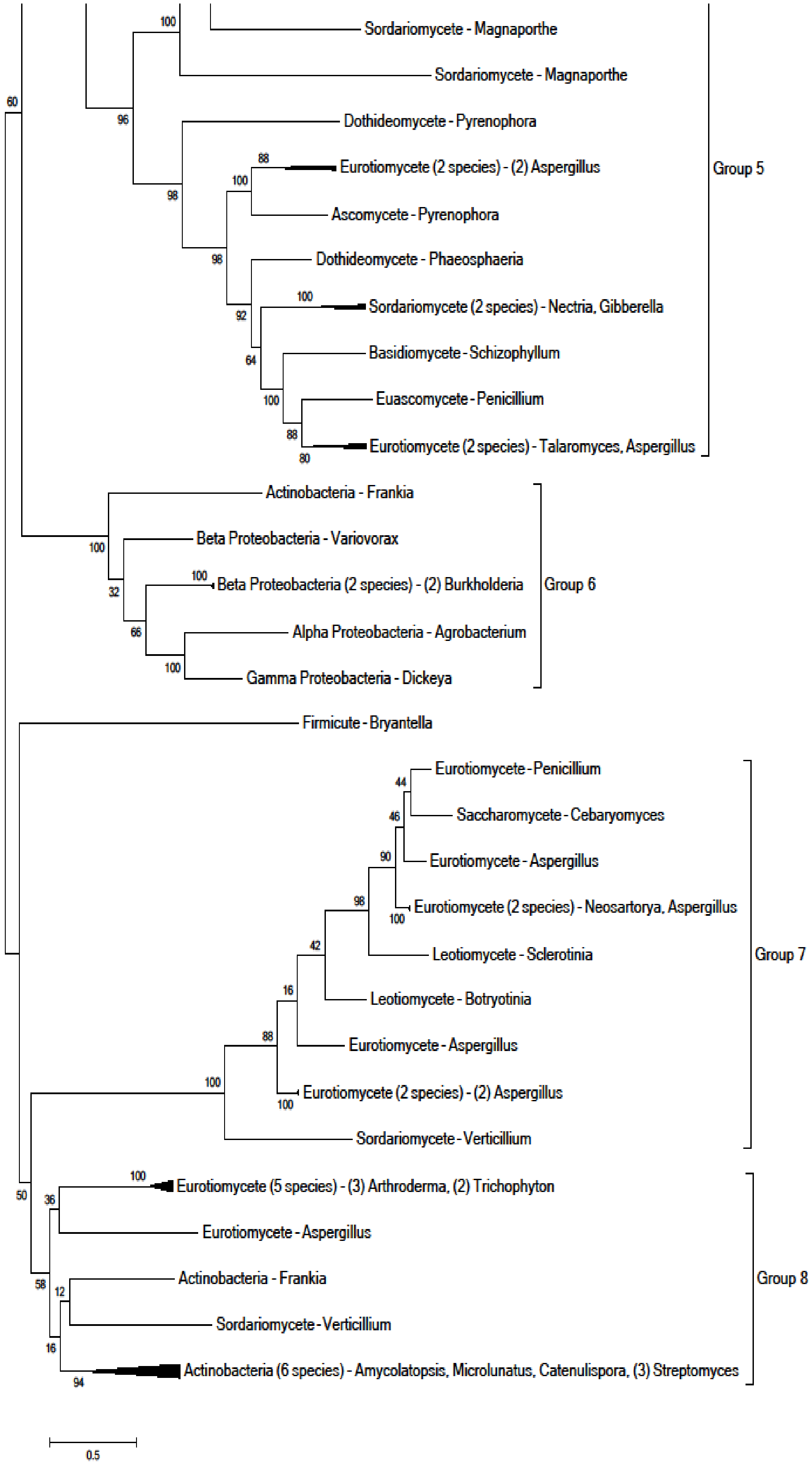
Zhang, Z., Hong, Q., Xu, J., Zhang, X. & Li, S. (2006). Isolation of fenitrothion-degrading strain *Burkholderia sp. FDS-1* and cloning of mpd gene. *Biodegradation*, *17*(3): 275–283.

Appendix A - Additional non-tannase domains removed from tannase sequences

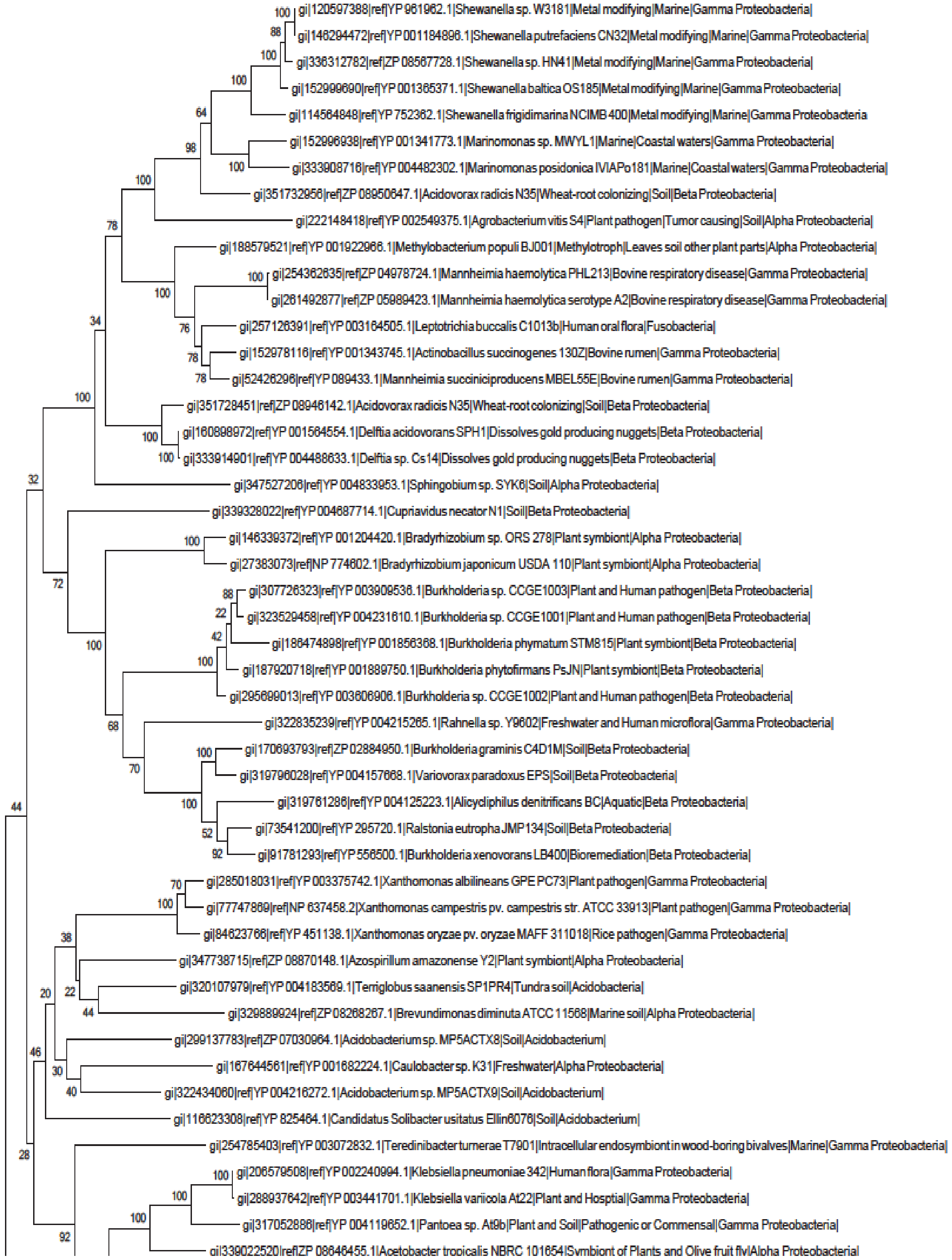
Domains Downstream of Tannase/feruloyl esterase (IPR011118)	Gene ID's
Outer membrane protein beta-barrel (IPR011250)	46114091, 302884826
Prokaryotic lipoprotein (noIPR)	73541200, 77747869, 91781293, 107022853, 114564848, 120597388, 121699582, 146294472, 152996938, 152999690, 160898972, 170734638, 242239727, 302420511, 319796028, 333914901, 336312782
Tetratricopeptide-like helical (IPR011990)	115388175

Domains Upstream of Tannase/feruloyl esterase (IPR011118)	Gene ID's
CTP synthase (IPR004468)	116214427
CTP synthase N-terminal (IPR017456)	116214427
Cytochrome P450 (IPR001128)	317037506
Glutamyl-tRNA(Gln) amidotransferase, B subunit (IPR004413)	336262152
Aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferase, subunit B/E (IPR017959)	336262152
Glutamine amidotransferase type 1	116214427

Appendix B - Reference phylogeny with bootstrap values and some genera grouped.


Sordariomycete - Magnaporthe
Sordariomycete - Magnaporthe
Dothideomycete - Pyrenophora
Eurotiomycete (2 species) - (2) Aspergillus
Ascomycete - Pyrenophora
Dothideomycete - Phaeosphaeria
Sordariomycete (2 species) - Nectria, Gibberella
Basidiomycete - Schizophyllum
Euascomycete - Penicillium
Eurotiomycete (2 species) - Talaromyces, Aspergillus
Group 5
Actinobacteria - Frankia
Beta Proteobacteria - Variovorax
Beta Proteobacteria (2 species) - (2) Burkholderia
Alpha Proteobacteria - Agrobacterium
Gamma Proteobacteria - Dickeya
Group 6
Firmicute - Bryantella
Eurotiomycete - Penicillium
Saccharomycete - Cebaryomyces
Eurotiomycete - Aspergillus
Eurotiomycete (2 species) - Neosartorya, Aspergillus
Leotiomycete - Sclerotinia
Leotiomycete - Botryotinia
Eurotiomycete - Aspergillus
Eurotiomycete (2 species) - (2) Aspergillus
Sordariomycete - Verticillium
Group 7
Eurotiomycete (5 species) - (3) Arthroderma, (2) Trichophyton
Eurotiomycete - Aspergillus
Actinobacteria - Frankia
Sordariomycete - Verticillium
Actinobacteria (6 species) - Amycolatopsis, Microlunatus, Catenulispora, (3) Streptomyces
Group 8
100
60
96
98
88
100
98
92
100
64
100
88
80
100
32
100
66
100
44
46
90
98
100
42
16
88
100
100
100
50
36
58
12
16
94
0.5

Appendix C - Reference phylogeny with bootstrap values with all species shown.




94
90
gi|161523057|ref|YP 001585986.1|Burkholderia multivorans ATCC 17616|Human pathogen|Colonizes lungs in Cystic fibrosis|Beta Proteobacteria|
100
gi|221198972|ref|ZP 03572017.1|Burkholderia multivorans CGD2M|Human pathogen|Colonizes lungs in Cystic fibrosis|Beta Proteobacteria|
100
32
gi|78060680|ref|YP 367255.1|Burkholderia sp. 383|Animal and Plant pathogen|Beta Proteobacteria|
gi|209516883|ref|ZP 03265733.1|Burkholderia sp. H160|Animal and Plant pathogen|Beta Proteobacteria|
gi|301121084|ref|XP 002908269.1|Phytophthora infestans T304|Plant pathogen|Potato blight|Oomycota|
44
100
gi|146282038|ref|YP 001172191.1|Pseudomonas stutzeri A1501|Soil|Gamma Proteobacteria|
72
gi|339493653|ref|YP 004713946.1|Pseudomonas stutzeri ATCC 17588 LMG 11199|Soil|Gamma Proteobacteria|
100
gi|213970659|ref|ZP 03398785.1|Pseudomonas syringae pv. tomato T1|Plant pathogen|Gamma Proteobacteria|
100
gi|28870160|ref|NP 792779.1|Pseudomonas syringae pv. tomato str. DC3000|Plant pathogen|Gamma Proteobacteria|
98
gi|289623896|ref|ZP 06456850.1|Pseudomonas syringae pv. aesculi str. NCPPB3681|Plant pathogen|Gamma Proteobacteria|
100
gi|298486915|ref|ZP 07004969.1|Pseudomonas savastanoi pv. savastanoi NCPPB 3335|Plant pathogen|Gamma Proteobacteria|
100
gi|302188555|ref|ZP 07265228.1|Pseudomonas syringae pv. syringae 642|Plant pathogen|Gamma Proteobacteria|
100
gi|66046096|ref|YP 235937.1|Pseudomonas syringae pv. syringae B728a|Plant pathogen|Gamma Proteobacteria|
100
gi|336262152|ref|XP 003345861.1|Sordaria macrospora khell|Soil|Ascomycete|
90
gi|85091584|ref|XP 958973.1|Neurospora crassa OR74A|Red bread mold|Ascomycete|
100
gi|171679739|ref|XP 001904816.1|Podospora anserina S mat+|Feces|Sordariomycete|
60
70
gi|116214427|ref|XP 001230230.1|Chaetomium globosum CBS 148.51|Mold|Soil and Plant debris|Sordariomycete|
100
gi|39940908|ref|XP 359991.1|Magnaporthe oryzae 7015|Rice pathogen|Sordariomycete|
gi|145615440|ref|XP 360218.2|Magnaporthe oryzae 7015|Rice pathogen|Sordariomycete|
gi|189200795|ref|XP 001936734.1|Pyrenophora triticirepentis Pt1CBFP|Wheat and Plant pathogen|Dothideomycete|
96
88
gi|115388175|ref|XP 001211593.1|Aspergillus terreus NIH2624|Opportunistic Human pathogen|Eurotiomycete|
100
gi|317147761|ref|XP 001822263.2|Aspergillus oryzae RIB40|Food industry|Ferment soybeans and Saccharify grains|Eurotiomycete|
98
gi|330925352|ref|XP 003301017.1|Pyrenophora teres f. teres 01|Barley and Plant pathogen|Ascomycete|
gi|169616296|ref|XP 001801563.1|Phaeosphaeria nodorum SN15|Wheat and plant pathogen|Dothidiomycete|
98
100
gi|302884826|ref|XP 003041307.1|Nectria haematococca mpVI 77134|Plant pathogen|Sordariomycete|
92
gi|46114091|ref|XP 383127.1|Gibberella zeae PH1|Wheat and Barley pathogen|Sordariomycete|
64
gi|302675743|ref|XP 003027555.1|Schizophyllum commune H48|Rotting wood and Human pathogen|Basidiomycete|
gi|212538837|ref|XP 002149574.1|Penicillium marneffei ATCC 18224|Human pathogen|Euascomycete|
100
88
gi|121699582|ref|XP 001268067.1|Aspergillus clavatus NRRL 1|Plant pathogen and Occasional animal pathogen|Soil|Eurotiomycete|
80
gi|242762966|ref|XP 002340484.1|Talaromyces stipitatus ATCC 10500|Soil|Eurotiomycete|
gi|312197890|ref|YP 004017951.1|Frankia sp. EuI1c|Plant symbiont|Actinobacteria|
gi|239813100|ref|YP 002942010.1|Variovorax paradoxus S110|Soil|Beta Proteobacteria|
100
100
gi|107022853|ref|YP 621180.1|Burkholderia cenocepacia AU 1054|Plant pathogen|Opportunistic human pathogen|Beta Proteobacteria|
32
gi|170734638|ref|YP 001773752.1|Burkholderia cenocepacia MC03|Plant pathogen|Opportunistic human pathogen|Beta Proteobacteria|
66
gi|222102103|ref|YP 002546693.1|Agrobacterium radiobacter K84|Plant pathogen|Opportunistic human pathogen|Alpha Proteobacteria|
100
gi|242239727|ref|YP 002987908.1|Dickeya dadantii Ech703|Plant pathogen|Gamma Proteobacteria|
gi|255284383|ref|ZP 05348938.1|Bryantella formatexigens DSM 14469|Human feces|Firmicute|
44
gi|255955963|ref|XP 002568734.1|Penicillium chrysogenum Wisconsin 541255|Damp indoors|Eurotiomycete|
46
gi|294654888|ref|XP 456970.2|Debaryomyces hansenii CBS767|Marine|Saccharomycete|
90
gi|67904432|ref|XP 682472.1|Aspergillus nidulans FGSC A4|Soil|Eurotiomycete|
gi|119487214|ref|XP 001262462.1|Neosartorya fischeri NRRL 181|Human pathogen|Eurotiomycete|
98
100
gi|70982011|ref|XP 746534.1|Aspergillus fumigatus Af293|Occasional human pathogen|Soil|Eurotiomycete|
42
gi|156049309|ref|XP 001590621.1|Sclerotinia sclerotiorum 1980|White mold|Plant pathogen|Leotiomycete|
16
gi|154296210|ref|XP 001548537.1|Botryotinia fuckeliana B05.10|Gray mold|Plant pathogen|Leotiomycete|
88
gi|145240885|ref|XP 001393089.1|Aspergillus niger CBS 513.88|Black mold|Plant pathogen|Eurotiomycete|
gi|169772335|ref|XP 001820636.1|Aspergillus oryzae RIB40|Food industry|Ferment soybeans and Saccharify grains|Eurotiomycete|
100
100
gi|238490238|ref|XP 002376356.1|Aspergillus flavus NRRL3357|Plant and human pathogen|Eurotiomycete|
gi|302420511|ref|XP 003008086.1|Verticillium alboatrum VaMs.102|Plant pathogen|Sordariomycete|


gi|302505633|ref|XP 003014523.1|Arthroderma benhamiae CBS 112371|Animal pathogen|Eurotiomycete|
gi|327300745|ref|XP 003235065.1|Trichophyton rubrum CBS 118892|Animal pathogen|Eurotiomycete|
gi|302658993|ref|XP 003021192.1|Trichophyton verrucosum HKI 0517|Animal pathogen|Eurotiomycete|
gi|315048443|ref|XP 003173596.1|Arthroderma gypseum CBS 118893|Animal pathogen|Eurotiomycete|
gi|296809083|ref|XP 002844880.1|Arthroderma otae CBS 113480|Animal pathogen|Eurotiomycete|
gi|317037506|ref|XP 001398581.2|Aspergillus niger CBS 513.88|Black mold|Plant pathogen|Eurotiomycete|
gi|288920905|ref|ZP 06415200.1|Frankia sp. EUN1f|Plant symbiont|Actinobacteria|
gi|302407670|ref|XP 003001670.1|Verticillium alboatrum VaMs.102|Plant pathogen|Sordariomycete|
gi|300789302|ref|YP 003769593.1|Amycolatopsis mediterranei U32|Soil|Actinobacteria|
gi|336117548|ref|YP 004572316.1|Microlunatus phosphovorus NM1|Soil|Actinobacteria|
gi|256396237|ref|YP 003117801.1|Catenulispora acidiphila DSM 44928|Soil|Actinobacteria|
gi|318057905|ref|ZP 07976628.1|Streptomyces sp. SA3 actG|Soil|Actinobacteria|
gi|302517590|ref|ZP 07269932.1|Streptomyces sp. SPB78|Soil|Actinobacteria|
gi|333028762|ref|ZP 08456826.1|Streptomyces sp. Tu6071|Soil|Actinobacteria|
76
76
66
100
50
36
58
12
16
94
64
82
100
78
0.5

Appendix D - Microbiome list organized by tannase clade membership

Clade	Gene ID	Locus tag	Gene Product Name	Genome Name	Ecosystem	Specific Ecosystem
1	2003404028		hypothetical protein	Air microbial communities Singapore indoor air filters 1	Air	Filter
1	2003486041		Bll4200 protein	Air microbial communities Singapore indoor air filters 2	Air	Filter
1	BMHBC_06067040	BMHBC_06067040	Tannase and feruloyl esterase.	Biofuel metagenome 1 (454/Illumina combined assembly)	Freshwater	Biofuel
1	LBLACPB1_01141330	LBLACPB1_01141330	Tannase and feruloyl esterase.	Fresh water microbial communities from LaBonte Lake, Laramie, Wyoming, sample from algal/cyanobacterial bloom material peak-bloom 1 (algal/cyano bloom peak-bloom 1)	Freshwater	Algal/Cyanobacterial Bloom
1	LBLACPB1_01644250	LBLACPB1_01644250	Lysophospholipase	Fresh water microbial communities from LaBonte Lake, Laramie, Wyoming, sample from algal/cyanobacterial bloom material peak-bloom 1 (algal/cyano bloom peak-bloom 1)	Freshwater	Algal/Cyanobacterial Bloom
1	M3P_100039883	M3P_100039883	Tannase and feruloyl esterase	Lotic microbial communities from Mississippi River at two locations in the state of Minnesota, sample from River Site 1, Mississippi Headwaters	Freshwater	Mississippi River
1	M3P_100135281	M3P_100135281	Tannase and feruloyl esterase	Lotic microbial communities from Mississippi River at two locations in the state of Minnesota, sample from River Site 1, Mississippi Headwaters	Freshwater	Mississippi River
1	M3P_100156701	M3P_100156701	Tannase and feruloyl esterase	Lotic microbial communities from Mississippi River at two locations in the state of Minnesota, sample from River Site 1, Mississippi Headwaters	Freshwater	Mississippi River
1	M3P_100307212	M3P_100307212	Tannase and feruloyl esterase	Lotic microbial communities from Mississippi River at two locations in the state of Minnesota, sample from River Site 1, Mississippi Headwaters	Freshwater	Mississippi River
1	M3P_100397281	M3P_100397281	Tannase and feruloyl esterase	Lotic microbial communities from Mississippi River at two locations in the state of Minnesota, sample from River Site 1, Mississippi Headwaters	Freshwater	Mississippi River
1	M3P_10086221	M3P_10086221	Tannase and feruloyl esterase	Lotic microbial communities from Mississippi River at two locations in the state of Minnesota, sample from River Site 7, Mississippi Headwaters (Site 7)	Freshwater	Mississippi River

1	2006736305		Bll4200 protein	Methylotrophic community from Lake Washington sediment combined (v2)	Freshwater	Lake Washington Methylotrophic Community
1	2006240924	LWME_40924	Bll4200 protein	Methylotrophic community from Lake Washington sediment Methane enrichment	Freshwater	Lake Washington Methylotrophic Community
1	2006344790	LWMOL_44790	Bll4200 protein	Methylotrophic community from Lake Washington sediment Methanol enrichment	Freshwater	Lake Washington Methylotrophic Community
1	2007496403		Bll4200 protein	Oak Ridge Pristine Groundwater FRC FW301	Freshwater	Uranium-Contaminated Groundwater
1	2007497911		Bll4200 protein	Oak Ridge Pristine Groundwater FRC FW301	Freshwater	Uranium-Contaminated Groundwater
1	2007522986		Bll4200 protein	Oak Ridge Pristine Groundwater FRC FW301	Freshwater	Uranium-Contaminated Groundwater
1	2007556081		Bll4200 protein	Oak Ridge Pristine Groundwater FRC FW301	Freshwater	Uranium-Contaminated Groundwater
1	2007582462		hypothetical protein	Oak Ridge Pristine Groundwater FRC FW301	Freshwater	Uranium-Contaminated Groundwater
1	comb1_0768.00004250	comb1_0768.00004250	Tannase and feruloyl esterase.	Sediment microbial communities from Lake Washington, Seattle, for Methane and Nitrogen Cycles, original sample replicate 1	Freshwater	Lake Washington Sediment

1	WSSedA1BaDRAFT_10036321	WSSedA1BaDRAFT_10036321	Tannase and feruloyl esterase	Wetland microbial communities from Twitchell Island in the Sacramento Delta, sample from surface sediment Feb2011 Site A1 Bulk (Wetland Surface Sediment Feb2011 Site A1 Bulk, Assem Ctgs IBYY 2011 Sep Assem)	Freshwater	Wetland Sediment
1	WSSedA1BaDRAFT_10135741	WSSedA1BaDRAFT_10135741	Tannase and feruloyl esterase	Wetland microbial communities from Twitchell Island in the Sacramento Delta, sample from surface sediment Feb2011 Site A1 Bulk (Wetland Surface Sediment Feb2011 Site A1 Bulk, Assem Ctgs IBYY 2011 Sep Assem)	Freshwater	Wetland Sediment
1	WSSedA2CDRAFT_0387441	WSSedA2CDRAFT_0387441	Tannase and feruloyl esterase	Wetland microbial communities from Twitchell Island in the Sacramento Delta, sample from surface sediment Feb2011 Site A2 Cattail (Wetland Surface Sediment Feb2011 Site A2 Cattail Sept 2011 assem)	Freshwater	Wetland Sediment
1	WSSedB1CaDRAFT_100029766	WSSedB1CaDRAFT_100029766	Tannase and feruloyl esterase	Wetland microbial communities from Twitchell Island in the Sacramento Delta, sample from surface sediment Feb2011 Site B1 Cattail (Wetland Surface Sediment Feb2011 Site B1 Cattail, Assem Ctgs Sep 2011 assem)	Freshwater	Wetland Sediment
1	WSSedL1CaDRAFT_1000030113	WSSedL1CaDRAFT_1000030113	Tannase and feruloyl esterase	Wetland microbial communities from Twitchell Island in the Sacramento Delta, sample from surface sediment Feb2011 Site L1 Cattail (Wetland Surface Sediment Feb2011 Site L1 Cattail, Assem Ctgs Sep 2011 assem)	Freshwater	Wetland Sediment
1	WSSedL1CaDRAFT_100043734	WSSedL1CaDRAFT_100043734	Tannase and feruloyl esterase	Wetland microbial communities from Twitchell Island in the Sacramento Delta, sample from surface sediment Feb2011 Site L1 Cattail (Wetland Surface Sediment Feb2011 Site L1 Cattail, Assem Ctgs Sep 2011 assem)	Freshwater	Wetland Sediment
1	C1384448__gene_49952	C1384448__gene_49952	flagellar biosynthetic protein fliQ	Human Buccal mucosa microbiome from visit number 1 of subject 159490532	Human	Buccal
1	C2693091__gene_114598	C2693091__gene_114598	hypothetical protein	Human Buccal mucosa microbiome from visit number 2 of subject 159814214	Human	Buccal
1	SRS011098_Baylor_scaffold_332	SRS011098_Baylor_scaffold_33283__gene_4351	hypothetical protein	Human Supragingival plaque microbiome from visit number 1 of subject 158458797	Human	Supragingival

	83__gene_43511	1				
1	C4700393__gene_272383	C4700393__gene_272383	tannase/feruloyl esterase family protein	Human Supragingival plaque microbiome from visit number 1 of subject 158479027	Human	Supragingival
1	SRS011126_Baylor_scaffold_67919__gene_82888	SRS011126_Baylor_scaffold_67919__gene_82888	flagellar biosynthetic protein fliQ	Human Supragingival plaque microbiome from visit number 1 of subject 158479027	Human	Supragingival
1	SRS017304_Baylor_scaffold_30051__gene_45757	SRS017304_Baylor_scaffold_30051__gene_45757	hypothetical protein	Human Supragingival plaque microbiome from visit number 1 of subject 159207311	Human	Supragingival
1	SRS017304_Baylor_scaffold_4118__gene_6196	SRS017304_Baylor_scaffold_4118__gene_6196	tannase and feruloyl esterase	Human Supragingival plaque microbiome from visit number 1 of subject 159207311	Human	Supragingival
1	SRS017227_Baylor_scaffold_110087__gene_142374	SRS017227_Baylor_scaffold_110087__gene_142374	flagellar biosynthetic protein fliQ	Human Supragingival plaque microbiome from visit number 1 of subject 159814214	Human	Supragingival
1	SRS017227_Baylor_scaffold_125882__gene_165836	SRS017227_Baylor_scaffold_125882__gene_165836	tannase and feruloyl esterase	Human Supragingival plaque microbiome from visit number 1 of subject 159814214	Human	Supragingival
1	SRS017227_Baylor_scaffold_158511__gene_221024	SRS017227_Baylor_scaffold_158511__gene_221024	flagellar biosynthetic protein fliQ	Human Supragingival plaque microbiome from visit number 1 of subject 159814214	Human	Supragingival

1	C3129047__gene_176309	C3129047__gene_176309	flagellar biosynthetic protein fliQ	Human Supragingival plaque microbiome from visit number 1 of subject 160319967	Human	Supragingival
1	C3497502__gene_234651	C3497502__gene_234651	flagellar biosynthetic protein fliQ	Human Supragingival plaque microbiome from visit number 1 of subject 160400887	Human	Supragingival
1	C2534940__gene_179059	C2534940__gene_179059	tannase/feruloyl esterase family protein	Human Supragingival plaque microbiome from visit number 1 of subject 763840445	Human	Supragingival
1	C2598810__gene_191434	C2598810__gene_191434	hypothetical protein	Human Supragingival plaque microbiome from visit number 1 of subject 764143897	Human	Supragingival
1	SRS015574_WUGC_scaffold_78217__gene_100413	SRS015574_WUGC_scaffold_78217__gene_100413	flagellar biosynthetic protein fliQ	Human Supragingival plaque microbiome from visit number 1 of subject 764285508	Human	Supragingival
1	SRS015899_WUGC_scaffold_14024__gene_19317	SRS015899_WUGC_scaffold_14024__gene_19317	tannase and feruloyl esterase	Human Supragingival plaque microbiome from visit number 1 of subject 764305738	Human	Supragingival
1	SRS019591_WUGC_scaffold_34334__gene_42111	SRS019591_WUGC_scaffold_34334__gene_42111	hypothetical protein	Human Supragingival plaque microbiome from visit number 1 of subject 765620695	Human	Supragingival
1	SRS022536_LANL_scaffold_90610__gene_202726	SRS022536_LANL_scaffold_90610__gene_202726	tannase and feruloyl esterase	Human Supragingival plaque microbiome from visit number 1 of subject 809635352	Human	Supragingival
1	SRS024289_LANL_scaffold_6284	SRS024289_LANL_scaffold_62846__gene_88593	hypothetical protein	Human Supragingival plaque microbiome from visit number 2 of subject 158944319	Human	Supragingival

	6__gene_88593					
1	SRS024289_LANL_scaffold_66258__gene_95126	SRS024289_LANL_scaffold_66258__gene_95126	tannase/feruloyl esterase family protein	Human Supragingival plaque microbiome from visit number 2 of subject 158944319	Human	Supragingival
1	SRS024289_LANL_scaffold_66459__gene_95522	SRS024289_LANL_scaffold_66459__gene_95522	flagellar biosynthetic protein fliQ	Human Supragingival plaque microbiome from visit number 2 of subject 158944319	Human	Supragingival
1	SRS049318_LANL_scaffold_76559__gene_98628	SRS049318_LANL_scaffold_76559__gene_98628	flagellar biosynthetic protein fliQ	Human Supragingival plaque microbiome from visit number 2 of subject 246515023	Human	Supragingival
1	SRS015803_WUGC_scaffold_48882__gene_68701	SRS015803_WUGC_scaffold_48882__gene_68701	flagellar biosynthetic protein fliQ	Human Supragingival plaque microbiome from visit number 2 of subject 763435843	Human	Supragingival
1	SRS015063_WUGC_scaffold_14129__gene_13712	SRS015063_WUGC_scaffold_14129__gene_13712	hypothetical protein	Human Supragingival plaque microbiome from visit number 2 of subject 763577454	Human	Supragingival
1	C3637963__gene_214362	C3637963__gene_214362	flagellar biosynthetic protein fliQ	Human Supragingival plaque microbiome from visit number 2 of subject 763840445	Human	Supragingival
1	C2094478__gene_113694	C2094478__gene_113694	tannase/feruloyl esterase family protein	Human Supragingival plaque microbiome from visit number 2 of subject 764224817	Human	Supragingival
1	C2270814__gene_131318	C2270814__gene_131318	flagellar biosynthetic protein fliQ	Human Supragingival plaque microbiome from visit number 2 of subject 764892411	Human	Supragingival
1	SRS019906_WU	SRS019906_WUG	tannase and	Human Supragingival plaque microbiome from visit number 2 of	Human	Supragingival

	GC_scaffold_26968__gene_31103	C_scaffold_26968__gene_31103	feruloyl esterase	subject 764892411		
1	SRS016740_Baylor_scaffold_37079__gene_41239	SRS016740_Baylor_scaffold_37079__gene_41239	hypothetical protein	Human Tongue dorsum microbiome from visit number 1 of subject 160421117	Human	Tongue
1	SRS042643_WUGC_scaffold_25099__gene_34690	SRS042643_WUGC_scaffold_25099__gene_34690	tannase and feruloyl esterase	Human Tongue dorsum microbiome from visit number 1 of subject 404239096	Human	Tongue
1	SRS042643_WUGC_scaffold_25100__gene_34693	SRS042643_WUGC_scaffold_25100__gene_34693	tannase and feruloyl esterase	Human Tongue dorsum microbiome from visit number 1 of subject 404239096	Human	Tongue
1	C2057225__gene_138510	C2057225__gene_138510	tannase and feruloyl esterase	Human Tongue dorsum microbiome from visit number 1 of subject 764508039	Human	Tongue
1	SRS016037_WUGC_scaffold_13676__gene_20176	SRS016037_WUGC_scaffold_13676__gene_20176	tannase/feruloyl esterase family protein	Human Tongue dorsum microbiome from visit number 1 of subject 764508039	Human	Tongue
1	SRS019219_WUGC_scaffold_67899__gene_104201	SRS019219_WUGC_scaffold_67899__gene_104201	hypothetical protein	Human Tongue dorsum microbiome from visit number 1 of subject 765034022	Human	Tongue
1	SRS022621_Baylor_scaffold_204	SRS022621_Baylor_scaffold_2042__gene_2251	tannase and feruloyl esterase	Human Tongue dorsum microbiome from visit number 2 of subject 158499257	Human	Tongue

	2__gene_2251					
1	SRS024441_LANL_scaffold_32495__gene_39205	SRS024441_LANL_scaffold_32495__gene_39205	flagellar biosynthetic protein fliQ	Human Tongue dorsum microbiome from visit number 2 of subject 159571453	Human	Tongue
1	SPBFG_1220040	SPBFG_1220040	Tannase and feruloyl esterase.	Dendroctonus frontalis Fungal community	Insect	Southern Pine Beetle
1	ACODT_3112880	ACODT_3112880	Tannase and feruloyl esterase.	Fungus garden microbial communities from Atta colombica in Panama, sample from dump top (Dump top)	Insect	Atta colombica
1	AglaG_05916270	AglaG_05916270	Tannase and feruloyl esterase.	Larvae (6 May 2010 assembly)	Insect	Anaplophora glabripennis
1	2228960184	PaMGMa_11	Tannase and feruloyl esterase.	Panchlora_midgut_metagenome (Midgut August 2011 assem)	Insect	Panchlora
1	2225226354		KEGG ID:K09252	Wild Panda gut microbiome from Saanxi China, sample from individual w5 (GB9)	Panda	Saanxi China
1	F10SI0310_0102.00005760	F10SI0310_0102.00005760	Tannase and feruloyl esterase.	Line P sample F_10_SI03_10 (F_10_S103_10)	Marine	10m
1	ASA123_01198670	ASA123_01198670	Tannase and feruloyl esterase.	Marine sediment archaeal communities from Santa Barbara Basin, CA, that are methane-oxidizing, sample 3-6 cm (ANME Sed A12 3-6 cm)	Marine	Sediment 3-6cm
1	M940CN_10053262	M940CN_10053262	Tannase and feruloyl esterase	Microbial Communities from Little Sippewissett Salt Marsh, Woods Hole, MA that are anoxygenic and photosynthetic, Marine photosynthetic community that grows at 940nm (Marine_940nm_cellulose)	Marine - Salt Marsh	1ft - Little Sippewissett
1	M940M_10002302	M940M_10002302	Tannase and feruloyl esterase	Microbial Communities from Little Sippewissett Salt Marsh, Woods Hole, MA that are anoxygenic and photosynthetic, Marine photosynthetic community that grows at 940nm with malate (Marine_940nm_malate)	Marine - Salt Marsh	1ft - Little Sippewissett
1	2201344357	MRSJC2b_58228	Tannase and feruloyl esterase.	Mesophilic rice straw/compost enrichment metagenome: eDNA_1 (Mesophilic 454/Illumina Combined June 2011 assem)	Plant Compost	Rice Straw/Compost Enrichment

1	2201348890	MRSJC2b_61036	Tannase and feruloyl esterase.	Mesophilic rice straw/compost enrichment metagenome: eDNA_1 (Mesophilic 454/Illumina Combined June 2011 assem)	Plant Compost	Rice Straw/Compost Enrichment
1	2201380084	MRSJC2b_708681	Dipeptidyl aminopeptidases/acylaminoacyl-peptidases	Mesophilic rice straw/compost enrichment metagenome: eDNA_1 (Mesophilic 454/Illumina Combined June 2011 assem)	Plant Compost	Rice Straw/Compost Enrichment
1	2201460927	MRSJC2b_98533	Tannase and feruloyl esterase.	Mesophilic rice straw/compost enrichment metagenome: eDNA_1 (Mesophilic 454/Illumina Combined June 2011 assem)	Plant Compost	Rice Straw/Compost Enrichment
1	2201586457	MRSJC2b_150020	Tannase and feruloyl esterase.	Mesophilic rice straw/compost enrichment metagenome: eDNA_1 (Mesophilic 454/Illumina Combined June 2011 assem)	Plant Compost	Rice Straw/Compost Enrichment
1	2049432526	PBDCA2_9065440	Tannase and feruloyl esterase.	Poplar biomass bioreactor microbial communities from Brookhaven National Lab, NY, sample from total biomass decay community (13 April 2010 assembly with 454 paired-end)	Plant Compost	Poplar Biomass
1	2200358422	TRSJC2b_40342	Lysophospholipase	Thermophilic rice straw/compost enrichment metagenome: eDNA_2 (Thermophilic 454/Illumina Combined June 2011 assem)	Plant Compost	Rice Straw
1	2200464770	TRSJC2b_8199	Lysophospholipase	Thermophilic rice straw/compost enrichment metagenome: eDNA_2 (Thermophilic 454/Illumina Combined June 2011 assem)	Plant Compost	Rice Straw
1	2200499591	TRSJC2b_8010	Poly(3-hydroxybutyrate) depolymerase	Thermophilic rice straw/compost enrichment metagenome: eDNA_2 (Thermophilic 454/Illumina Combined June 2011 assem)	Plant Compost	Rice Straw
1	ARcpr5oldR_0025942	ARcpr5oldR_0025942	Tannase and feruloyl esterase	Arabidopsis rhizosphere microbial communities from University of North Carolina, sample from Arabidopsis cpr5 old rhizosphere (Arabidopsis cpr5 old rhizosphere, Nov 2011 assem)	Rhizosphere	Arabidopsis
1	2214012114	ARWTCa_233991	Tannase and feruloyl esterase.	Arabidopsis rhizosphere microbial communities from University of North Carolina, sample Wild type Col-0 (Arabidopsis rhizosphere microbiome-wild type Col-0 454/Illumina 2011 July Assem)	Rhizosphere	Arabidopsis
1	MBSR1b_0039.0	MBSR1b_0039.0000	Tannase and	Miscanthus rhizosphere microbial communities from Kellogg Biological	Rhizosphere	Miscanthus

	0007770	7770	feruloyl esterase.	Station, MSU, sample Bulk Soil Replicate 1 : eDNA_1 (Bulk soil 1 April 2011 assembly)		
1	MBSR1b_0602.00000210	MBSR1b_0602.00000210	Tannase and feruloyl esterase.	Miscanthus rhizosphere microbial communities from Kellogg Biological Station, MSU, sample Bulk Soil Replicate 1 : eDNA_1 (Bulk soil 1 April 2011 assembly)	Rhizosphere	Miscanthus
1	SwBSRL2_0190.00006970	SwBSRL2_0190.00006970	Tannase and feruloyl esterase.	Switchgrass rhizosphere microbial community from Michigan, US, sample from Rose Lake bulk soil RL2 (Bulk soil RL2 April 2011 assembly)	Rhizosphere	Switchgrass
1	SwRhRL2b_0058.00002780	SwRhRL2b_0058.00002780	Tannase and feruloyl esterase.	Switchgrass rhizosphere microbial community from Michigan, US, sample from Rose Lake rhizosphere RL2 (Rhizosphere RL2 April 2011 assembly)	Rhizosphere	Switchgrass
1	ICChiseqgaiiDRAFT_06090831	ICChiseqgaiiDRAFT_06090831	Tannase and feruloyl esterase	Soil microbial communities from Great Prairies, sample from Iowa, Continuous Corn soil (Iowa, Continuous Corn soil, Jan 2012 Assem MSU hiseq+gaii)	Soil	Corn Soil
1	ICChiseqgaiiDRAFT_06657431	ICChiseqgaiiDRAFT_06657431	Tannase and feruloyl esterase	Soil microbial communities from Great Prairies, sample from Iowa, Continuous Corn soil (Iowa, Continuous Corn soil, Jan 2012 Assem MSU hiseq+gaii)	Soil	Corn Soil
1	ICChiseqgaiiDRAFT_07672791	ICChiseqgaiiDRAFT_07672791	Tannase and feruloyl esterase	Soil microbial communities from Great Prairies, sample from Iowa, Continuous Corn soil (Iowa, Continuous Corn soil, Jan 2012 Assem MSU hiseq+gaii)	Soil	Corn Soil
1	ICChiseqgaiiDRAFT_07933941	ICChiseqgaiiDRAFT_07933941	Tannase and feruloyl esterase	Soil microbial communities from Great Prairies, sample from Iowa, Continuous Corn soil (Iowa, Continuous Corn soil, Jan 2012 Assem MSU hiseq+gaii)	Soil	Corn Soil
1	ICChiseqgaiiDRAFT_20812321	ICChiseqgaiiDRAFT_20812321	Tannase and feruloyl esterase	Soil microbial communities from Great Prairies, sample from Iowa, Continuous Corn soil (Iowa, Continuous Corn soil, Jan 2012 Assem MSU hiseq+gaii)	Soil	Corn Soil
1	ICChiseqgaiiDRAFT_22089901	ICChiseqgaiiDRAFT_22089901	Tannase and feruloyl esterase	Soil microbial communities from Great Prairies, sample from Iowa, Continuous Corn soil (Iowa, Continuous Corn soil, Jan 2012 Assem MSU hiseq+gaii)	Soil	Corn Soil
1	2001278801		hypothetical protein	Soil microbial communities from Minnesota Farm	Soil	Minnesota Farm
1	A2_c1_00130480	A2_c1_00130480	Tannase and feruloyl esterase.	Soil microbial communities from permafrost in Bonanza Creek, Alaska, sample from Active Layer A2 (A2_CLC_pe)	Soil	Permafrost
1	A2_c1_0	A2_c1_00872640	Tannase and	Soil microbial communities from	Soil	Permafrost

	0872640		feruloyl esterase.	permafrost in Bonanza Creek, Alaska, sample from Active Layer A2 (A2_CLC_pe)		
1	A5_c1_01752900	A5_c1_01752900	Tannase and feruloyl esterase.	Soil microbial communities from permafrost in Bonanza Creek, Alaska, sample from Active Layer A5 (A5_CLC_pe)	Soil	Permafrost
1	PRSSG_00618020	PRSSG_00618020	Tannase and feruloyl esterase.	Soil microbial communities from Puerto Rico rain forest, that decompose switchgrass, sample from feedstock-adapted consortia SG only (SG only)	Soil	Rainforest
1	PRSSG2_01175290	PRSSG2_01175290	Lysophospholipase	Soil microbial communities from Puerto Rico rain forest, that decompose switchgrass, sample from feedstock-adapted consortia SG only (SG only, May 2011 assembly)	Soil	Rainforest
1	PRSSG2_01906080	PRSSG2_01906080	Tannase and feruloyl esterase.	Soil microbial communities from Puerto Rico rain forest, that decompose switchgrass, sample from feedstock-adapted consortia SG only (SG only, May 2011 assembly)	Soil	Rainforest
1	FACEMDA_6564290	FACEMDA_6564290	Tannase and feruloyl esterase.	Soil microbial communities from sample at FACE Site 1 Maryland Estuary CO2- (Maryland Estuary ambient)	Soil	Estuary 10cm
1	FACENCEE_1086640	FACENCEE_1086640	Tannase and feruloyl esterase.	Soil microbial communities from sample at FACE Site 3 Nevada Test Site Creosote CO2+	Soil	5cm
2	BMHBC_21746100	BMHBC_21746100	Tannase and feruloyl esterase.	Biofuel metagenome 1 (454/Illumina combined assembly)	Freshwater	Bioreactor
2	BMHB3a_1498443	BMHB3a_1498443	Tannase and feruloyl esterase.	Biofuel metagenome 3 (Biofuel metagenome 3 July 2011 assem)	Freshwater	Bioreactor
2	LBLACPB2_04120390	LBLACPB2_04120390	Tannase and feruloyl esterase.	Fresh water microbial communities from LaBonte Lake, Laramie, Wyoming, sample from algal/cyanobacterial bloom material peak-bloom 2 (algal/cyano bloom peak-bloom 2)	Freshwater	LaBonte Lake Algal/Cyano Bloom
2	ADL20m3uS_00305130	ADL20m3uS_00305130	Tannase and feruloyl esterase.	Freshwater microbial communities from Antarctic Deep Lake, sample 24m 3.0um (24 m 3.0 um Sept 2010 combined)	Freshwater	24m Antarctic Deep Lake

2	M3P_100645941	M3P_100645941	Tannase and feruloyl esterase	Lotic microbial communities from Mississippi River at two locations in the state of Minnesota, sample from River Site 1, Mississippi Headwaters	Freshwater	Mississippi River
2	2006686738		Bll7962 protein	Methylotrophic community from Lake Washington sediment combined (v2)	Freshwater	Lake Washington Sediment
2	2006740772		Tannase	Methylotrophic community from Lake Washington sediment combined (v2)	Freshwater	Lake Washington Sediment
2	2006234707	LWME_34707	Bll7962 protein	Methylotrophic community from Lake Washington sediment Methane enrichment	Freshwater	Lake Washington Sediment
2	2007455067		Bll7962 protein	Oak Ridge Pristine Groundwater FRC FW301	Freshwater	Uranium-Contaminated Groundwater
2	LWSO_09203020	LWSO_09203020	Tannase and feruloyl esterase.	Sediment microbial communities from Lake Washington, Seattle, for Methane and Nitrogen Cycles, original sample replicate 1 (Original sample replicate 1)	Freshwater	Lake Washington Sediment
2	LWSO2_04753170	LWSO2_04753170	Tannase and feruloyl esterase.	Sediment microbial communities from Lake Washington, Seattle, for Methane and Nitrogen Cycles, original sample replicate 2 (Original sample replicate 2 12C fraction)	Freshwater	Lake Washington Sediment
2	LWAeNNiSIP_54070	LWAeNNiSIP_54070	Tannase and feruloyl esterase.	Sediment microbial communities from Lake Washington, Seattle, for Methane and Nitrogen Cycles, sample Microc enrich af exp to meth lab w 13Ccarbon-no added nitrate (Aerobic without added nitrate, 13C SIP)	Freshwater	Lake Washington Sediment
2	LWAeNiSIP_281130	LWAeNiSIP_281130	Tannase and feruloyl esterase.	Sediment microbial communities from Lake Washington, Seattle, for Methane and Nitrogen Cycles, sample SIP 13C methane aerobic+nitrate (Aerobic with added nitrate, 13C SIP)	Freshwater	Lake Washington Sediment
2	LWAnNN_03353080	LWAnNN_03353080	Tannase and feruloyl esterase.	Sediment microbial communities from Lake Washington, Seattle, for Methane and Nitrogen Cycles, sample SIP 13Cmethane anaerobic no nitrate (Anaerobic no nitrate SIP Nov 2010 with PE)	Freshwater	Lake Washington Sediment
2	WSSedA1BaDRAFT_1012	WSSedA1BaDRAFT_101272	Tannase and feruloyl	Wetland microbial communities from Twitchell Island in the Sacramento Delta, sample from surface sediment	Freshwater	Groundwater

	7281	81	esterase	Feb2011 Site A1 Bulk (Wetland Surface Sediment Feb2011 Site A1 Bulk, Assem Ctgs IBYY 2011 Sep Assem)		
2	WSSedA1BaDRAFT_10567831	WSSedA1BaDRAFT_10567831	Tannase and feruloyl esterase	Wetland microbial communities from Twitchell Island in the Sacramento Delta, sample from surface sediment Feb2011 Site A1 Bulk (Wetland Surface Sediment Feb2011 Site A1 Bulk, Assem Ctgs IBYY 2011 Sep Assem)	Freshwater	Groundwater
2	WSSedB1B2DRAFT_00001510	WSSedB1B2DRAFT_00001510	Tannase and feruloyl esterase	Wetland microbial communities from Twitchell Island in the Sacramento Delta, sample from surface sediment Feb2011 Site B1 Bulk (Wetland Surface Sediment Feb2011 Site B1 Bulk Feb 2012)	Freshwater	Groundwater
2	WSSedL1CaDRAFT_1000078813	WSSedL1CaDRAFT_1000078813	Tannase and feruloyl esterase	Wetland microbial communities from Twitchell Island in the Sacramento Delta, sample from surface sediment Feb2011 Site L1 Cattail (Wetland Surface Sediment Feb2011 Site L1 Cattail, Assem Ctgs Sep 2011 assem)	Freshwater	Groundwater
2	SRS049959_WUGC_scaffold_27216__gene_51704	SRS049959_WUGC_scaffold_27216__gene_51704	hypothetical protein	Human Stool microbiome from visit number 1 of subject 765701615	Human	Stool
2	C1905634__gene_149935	C1905634__gene_149935	hypothetical protein	Human Stool microbiome from visit number 2 of subject 159490532	Human	Stool
2	CLOFG_1884830	CLOFG_1884830	Tannase and feruloyl esterase.	Cyphomyrmex longiscapus fungus garden	Insect	Cyphomyrmex longiscapus
2	SPBFG_1828390	SPBFG_1828390	Tannase and feruloyl esterase.	Dendroctonus frontalis Fungal community	Insect	Southern Pine Beetle
2	DPOFG_164020	DPOFG_164020	Tannase and feruloyl esterase.	Dendroctonus ponderosae fungus gallery (Hybrid pine) (MPB hybrid gallery)	Insect	Mountain Pine Beetle

2	DPOFG_680160	DPOFG_680160	Tannase and feruloyl esterase.	Dendroctonus ponderosae fungus gallery (Hybrid pine) (MPB hybrid gallery)	Insect	Mountain Pine Beetle
2	ACODT_547830	ACODT_547830	Tannase and feruloyl esterase.	Fungus garden microbial communities from Atta colombica in Panama, sample from dump top (Dump top)	Insect	Atta colombica
2	DPOLB_386180	DPOLB_386180	Tannase and feruloyl esterase.	Mountain Pine Beetle microbial communities from McBride, British Columbia, Canada, sample from Lodgepole pine (Lodgepole pine)	Insect	Mountain Pine Beetle (BC)
2	XylAfBA_294450	XylAfBA_294450	Tannase and feruloyl esterase.	Xyleborus affinis microbiome from Bern, Switzerland, sample of adult community (Ambrosia beetle adult)	Insect	Xyleborus affinis
2	XAGC_00436170	XAGC_00436170	Tannase and feruloyl esterase.	Xyleborus affinis microbiome from Bern, Switzerland, sample of gallery community (Gallery community)	Insect	Xyleborus affinis
2	XAGC_01372890	XAGC_01372890	Tannase and feruloyl esterase.	Xyleborus affinis microbiome from Bern, Switzerland, sample of gallery community (Gallery community)	Insect	Xyleborus affinis
2	XAGC_01610110	XAGC_01610110	Tannase and feruloyl esterase.	Xyleborus affinis microbiome from Bern, Switzerland, sample of gallery community (Gallery community)	Insect	Xyleborus affinis
2	XAGC_03328080	XAGC_03328080	Tannase and feruloyl esterase.	Xyleborus affinis microbiome from Bern, Switzerland, sample of gallery community (Gallery community)	Insect	Xyleborus affinis
2	XAGC_03454830	XAGC_03454830	Tannase and feruloyl esterase.	Xyleborus affinis microbiome from Bern, Switzerland, sample of gallery community (Gallery community)	Insect	Xyleborus affinis
2	RicEn_483310	RicEn_483310	Tannase and feruloyl	Endophytic microbiome from Rice	Plant	Rice Endophytic Microbiom

			esterase.			e
2	RicEn_596880	RicEn_596880	Tannase and feruloyl esterase.	Endophytic microbiome from Rice	Plant	Rice Endophytic Microbiome
2	2062004879	sg4i_00311270	Tannase and feruloyl esterase.	Thermophilic enrichment culture SG0.5JP960 (454-Illumina assembly) - version 2 (454-Illumina assembly v2)	Plant Compost	Switchgrass
2	2053620660	sgz454il_00143330	Tannase and feruloyl esterase.	Thermophilic enrichment culture SG0.5Z960 (454-Illumina assembly) (454-Illumina assembly)	Plant Compost	Switchgrass
2	ARcpr5yngRDRAFT_0086241	ARcpr5yngRDRAFT_0086241	Tannase and feruloyl esterase	Arabidopsis rhizosphere microbial communities from University of North Carolina, sample from Arabidopsis cpr5 young rhizosphere (Arabidopsis cpr5 young rhizosphere, Nov 2011 assem)	Rhizosphere	Arabidopsis
2	2104943239	ARMCC_00948310	Tannase and feruloyl esterase.	Arabidopsis rhizosphere microbial communities from University of North Carolina, sample Mutant cpr5 (cpr5 454/Illumina combined assembly)	Rhizosphere	Arabidopsis
2	2105407170	ARMCC_05587620	Tannase and feruloyl esterase.	Arabidopsis rhizosphere microbial communities from University of North Carolina, sample Mutant cpr5 (cpr5 454/Illumina combined assembly)	Rhizosphere	Arabidopsis
2	MBSR1b_0092.00002550	MBSR1b_0092.00002550	Tannase and feruloyl esterase.	Miscanthus rhizosphere microbial communities from Kellogg Biological Station, MSU, sample Bulk Soil Replicate 1 : eDNA_1 (Bulk soil 1 April 2011 assembly)	Rhizosphere	Miscanthus
2	MBSR1b_0863.00004920	MBSR1b_0863.00004920	Tannase and feruloyl esterase.	Miscanthus rhizosphere microbial communities from Kellogg Biological Station, MSU, sample Bulk Soil Replicate 1 : eDNA_1 (Bulk soil 1 April 2011 assembly)	Rhizosphere	Miscanthus
2	SRBS_1613270	SRBS_1613270	Tannase and feruloyl esterase.	Switchgrass rhizosphere microbial community from Michigan, US, sample from East Lansing bulk soil	Rhizosphere	Switchgrass

2	SwBSRL2_0098.00003410	SwBSRL2_0098.00003410	Tannase and feruloyl esterase.	Switchgrass rhizosphere microbial community from Michigan, US, sample from Rose Lake bulk soil RL2 (Bulk soil RL2 April 2011 assembly)	Rhizosphere	Switchgrass
2	SwBSRL2_0442.00002250	SwBSRL2_0442.00002250	Carboxylesterase type B	Switchgrass rhizosphere microbial community from Michigan, US, sample from Rose Lake bulk soil RL2 (Bulk soil RL2 April 2011 assembly)	Rhizosphere	Switchgrass
2	SwRhRL2b_0142.00006410	SwRhRL2b_0142.00006410	Dipeptidyl aminopeptidases/acylaminoacyl-peptidases	Switchgrass rhizosphere microbial community from Michigan, US, sample from Rose Lake rhizosphere RL2 (Rhizosphere RL2 April 2011 assembly)	Rhizosphere	Switchgrass
2	SwRhRL2b_0479.00001630	SwRhRL2b_0479.00001630	Predicted esterase	Switchgrass rhizosphere microbial community from Michigan, US, sample from Rose Lake rhizosphere RL2 (Rhizosphere RL2 April 2011 assembly)	Rhizosphere	Switchgrass
2	SwRhRL2b_0746.00005440	SwRhRL2b_0746.00005440	Tannase and feruloyl esterase.	Switchgrass rhizosphere microbial community from Michigan, US, sample from Rose Lake rhizosphere RL2 (Rhizosphere RL2 April 2011 assembly)	Rhizosphere	Switchgrass
2	prs_07590700	prs_07590700	Tannase and feruloyl esterase.	Luquillo Experimental Forest Soil, Puerto Rico	Soil	Forest Soil
2	ICChiseqgaiiDRAFT_06771071	ICChiseqgaiiDRAFT_06771071	Tannase and feruloyl esterase	Soil microbial communities from Great Prairies, sample from Iowa, Continuous Corn soil (Iowa, Continuous Corn soil, Jan 2012 Assem MSU hiseq+gaii)	Soil	Corn Soil
2	ICChiseqgaiiDRAFT_18706221	ICChiseqgaiiDRAFT_18706221	Tannase and feruloyl esterase	Soil microbial communities from Great Prairies, sample from Iowa, Continuous Corn soil (Iowa, Continuous Corn soil, Jan 2012 Assem MSU hiseq+gaii)	Soil	Corn Soil
2	ICChiseqgaiiDRAFT_19394761	ICChiseqgaiiDRAFT_19394761	Tannase and feruloyl esterase	Soil microbial communities from Great Prairies, sample from Iowa, Continuous Corn soil (Iowa, Continuous Corn soil, Jan 2012 Assem MSU hiseq+gaii)	Soil	Corn Soil
2	ICChiseqgaiiDRAFT_20977741	ICChiseqgaiiDRAFT_20977741	Tannase and feruloyl esterase	Soil microbial communities from Great Prairies, sample from Iowa, Continuous Corn soil (Iowa, Continuous Corn soil, Jan 2012 Assem MSU hiseq+gaii)	Soil	Corn Soil

2	ICChiseqgaiiDRAFT_21222171	ICChiseqgaiiDRAFT_21222171	Tannase and feruloyl esterase	Soil microbial communities from Great Prairies, sample from Iowa, Continuous Corn soil (Iowa, Continuous Corn soil, Jan 2012 Assem MSU hiseq+gaii)	Soil	Corn Soil
2	ICChiseqgaiiDRAFT_23918661	ICChiseqgaiiDRAFT_23918661	Tannase and feruloyl esterase	Soil microbial communities from Great Prairies, sample from Iowa, Continuous Corn soil (Iowa, Continuous Corn soil, Jan 2012 Assem MSU hiseq+gaii)	Soil	Corn Soil
2	ICChiseqgaiiDRAFT_23947011	ICChiseqgaiiDRAFT_23947011	Tannase and feruloyl esterase	Soil microbial communities from Great Prairies, sample from Iowa, Continuous Corn soil (Iowa, Continuous Corn soil, Jan 2012 Assem MSU hiseq+gaii)	Soil	Corn Soil
2	ICChiseqgaiiDRAFT_24011761	ICChiseqgaiiDRAFT_24011761	Tannase and feruloyl esterase	Soil microbial communities from Great Prairies, sample from Iowa, Continuous Corn soil (Iowa, Continuous Corn soil, Jan 2012 Assem MSU hiseq+gaii)	Soil	Corn Soil
2	2001289588		hypothetical protein	Soil microbial communities from Minnesota Farm	Soil	Minnesota Farm
2	2001336858		hypothetical protein	Soil microbial communities from Minnesota Farm	Soil	Minnesota Farm
2	A2_c1_00749130	A2_c1_00749130	Tannase and feruloyl esterase.	Soil microbial communities from permafrost in Bonanza Creek, Alaska, sample from Active Layer A2 (A2_CLC_pe)	Soil	Permafrost
2	A5_c1_01104300	A5_c1_01104300	Tannase and feruloyl esterase.	Soil microbial communities from permafrost in Bonanza Creek, Alaska, sample from Active Layer A5 (A5_CLC_pe)	Soil	Permafrost
2	B3_all_c_01895100	B3_all_c_01895100	Tannase and feruloyl esterase.	Soil microbial communities from permafrost in Bonanza Creek, Alaska, sample from Bog Site B3 (B3_all_CLC)	Soil	Permafrost
2	P3_CLC_00766570	P3_CLC_00766570	Tannase and feruloyl esterase.	Soil microbial communities from permafrost in Bonanza Creek, Alaska, sample from Permafrost Layer P3 (P3_CLC)	Soil	Permafrost
2	FACEMDA_3710290	FACEMDA_3710290	Tannase and feruloyl esterase.	Soil microbial communities from sample at FACE Site 1 Maryland Estuary CO2- (Maryland Estuary ambient)	Soil	Estuary

2	FACEORE_1057030	FACEORE_1057030	Tannase and feruloyl esterase.	Soil microbial communities from sample at FACE Site 5 Oak Ridge CO2+ (Oak Ridge elevated CO2)	Soil	Oak Ridge
2	FACEORE_3916260	FACEORE_3916260	Tannase and feruloyl esterase.	Soil microbial communities from sample at FACE Site 5 Oak Ridge CO2+ (Oak Ridge elevated CO2)	Soil	Oak Ridge
2	FACEORE_4591450	FACEORE_4591450	Tannase and feruloyl esterase.	Soil microbial communities from sample at FACE Site 5 Oak Ridge CO2+ (Oak Ridge elevated CO2)	Soil	Oak Ridge
2	FWIRA_00595410	FWIRA_00595410	Tannase and feruloyl esterase.	Soil microbial communities from sample at FACE Site Metagenome WIR_Amb2 (WIR_Amb2)	Soil	WIR
2	FWIRA_00861820	FWIRA_00861820	Tannase and feruloyl esterase.	Soil microbial communities from sample at FACE Site Metagenome WIR_Amb2 (WIR_Amb2)	Soil	WIR
2	FWIRA_02558820	FWIRA_02558820	Tannase and feruloyl esterase.	Soil microbial communities from sample at FACE Site Metagenome WIR_Amb2 (WIR_Amb2)	Soil	WIR
2	FWIRA_07174880	FWIRA_07174880	Tannase and feruloyl esterase.	Soil microbial communities from sample at FACE Site Metagenome WIR_Amb2 (WIR_Amb2)	Soil	WIR
2	FWIREl_02524770	FWIREl_02524770	Tannase and feruloyl esterase.	Soil microbial communities from sample at FACE Site Metagenome WIR_Elev2 (WIR_Elev2)	Soil	WIR
2	FWIREl_06791330	FWIREl_06791330	Tannase and feruloyl esterase.	Soil microbial communities from sample at FACE Site Metagenome WIR_Elev2 (WIR_Elev2)	Soil	WIR
2	FWIRElOz_07163260	FWIRElOz_07163260	Tannase and feruloyl	Soil microbial communities from sample at FACE Site Metagenome WIR_ElevOz2 (WIR_ElevOz2)	Soil	WIR

			esterase.			
2	FWIRElOz_08330650	FWIRElOz_08330650	Tannase and feruloyl esterase.	Soil microbial communities from sample at FACE Site Metagenome WIR_ElevOz2 (WIR_ElevOz2)	Soil	WIR
2	FWIROz_01833710	FWIROz_01833710	Tannase and feruloyl esterase.	Soil microbial communities from sample at FACE Site Metagenome WIR_Oz2 (WIR_Oz2)	Soil	WIR
2	FWIROz_07795870	FWIROz_07795870	Tannase and feruloyl esterase.	Soil microbial communities from sample at FACE Site Metagenome WIR_Oz2 (WIR_Oz2)	Soil	WIR
2	FWIROz_08907730	FWIROz_08907730	Tannase and feruloyl esterase.	Soil microbial communities from sample at FACE Site Metagenome WIR_Oz2 (WIR_Oz2)	Soil	WIR
2	FNCDEF_08807730	FNCDEF_08807730	Tannase and feruloyl esterase.	Soil microbial communities from sample at FACE Site North Carolina NCD_ElevF (NCD_ElevF)	Soil	North Carolina
2	FNTS007_04129080	FNTS007_04129080	Tannase and feruloyl esterase.	Soil microbial communities from sample at FACE Site NTS_007 Nevada Test Site (NTS_007)	Soil	Nevada
2	FNTS067_09776970	FNTS067_09776970	Tannase and feruloyl esterase.	Soil microbial communities from sample at FACE Site NTS_067 Nevada Test Site (NTS_067)	Soil	Nevada
2	FNTS_07598000	FNTS_07598000	Tannase and feruloyl esterase.	Soil microbial communities from sample at FACE Site NTS_071 Nevada Test Site (NTS_071)	Soil	Nevada
2	LCrCPGB2aDRAFT_0061351	LCrCPGB2aDRAFT_0061351	Tannase and feruloyl esterase	Soil microbial communities sample from Light Crust, Colorado Plateau, Green Butte 2 (Light Crust Colorado Plateau Green Butte 2, Oct 2011 assem)	Soil	Colorado Plateau

2	LCrCPGB2aDRAFT_0115811	LCrCPGB2aDRAFT_0115811	Tannase and feruloyl esterase	Soil microbial communities sample from Light Crust, Colorado Plateau, Green Butte 2 (Light Crust Colorado Plateau Green Butte 2, Oct 2011 assem)	Soil	Colorado Plateau
2	2000096020		hypothetical protein	Sludge/US, Phrap Assembly	Wastewater	EBPR Reactors
3	WSSedL2TaDRAFT_10308721	WSSedL2TaDRAFT_10308721	Tannase and feruloyl esterase	CECUM_4-1 (Microbiome Characterization)	Aves	Chicken Cecum
3	TB_LI09_4DRAFT_100237413	TB_LI09_4DRAFT_100237413	Tannase and feruloyl esterase	Groundwater microbial communities from subsurface biofilms in sulfidic aquifier in Frasassi Gorge, Italy, sample from two redox zones- LI09_4 (Targeted Biofilm samples from two redox zones-LI09_4, Oct 2011 Assem)	Freshwater	Groundwater Subsurface Biofilms
3	2006454565	LWFA_54565	Conserved hypothetical protein	Methylotrophic community from Lake Washington sediment Formaldehyde enrichment	Freshwater	Sediment
3	2006244155	LWME_44155	Tannase	Methylotrophic community from Lake Washington sediment Methane enrichment	Freshwater	Sediment
3	2006251512	LWME_51512	Tannase	Methylotrophic community from Lake Washington sediment Methane enrichment	Freshwater	Sediment
3	2006295663	LWMOL_95663	Conserved hypothetical protein	Methylotrophic community from Lake Washington sediment Methanol enrichment	Freshwater	Sediment
3	2006368104	LWMAM_68104	Tannase	Methylotrophic community from Lake Washington sediment Methylamine enrichment	Freshwater	Sediment
3	2006422035	LWMAM_22035	Tannase	Methylotrophic community from Lake Washington sediment Methylamine enrichment	Freshwater	Sediment
3	2007446894		Tannase	Oak Ridge Pristine Groundwater FRC FW301	Freshwater	Uranium Contaminated Groundwater
3	2007490376		Tannase	Oak Ridge Pristine Groundwater FRC FW301	Freshwater	Uranium Contaminated

						Groundwater
3	2007505310		hypothetical protein	Oak Ridge Pristine Groundwater FRC FW301	Freshwater	Uranium Contaminated Groundwater
3	2007505743		hypothetical protein	Oak Ridge Pristine Groundwater FRC FW301	Freshwater	Uranium Contaminated Groundwater
3	2007555530		hypothetical protein	Oak Ridge Pristine Groundwater FRC FW301	Freshwater	Uranium Contaminated Groundwater
3	2007586879		hypothetical protein	Oak Ridge Pristine Groundwater FRC FW301	Freshwater	Uranium Contaminated Groundwater
3	2007591923		Tannase	Oak Ridge Pristine Groundwater FRC FW301	Freshwater	Uranium Contaminated Groundwater
3	LWAnN_03395130	LWAnN_03395130	Tannase and feruloyl esterase.	Sediment microbial communities from Lake Washington, Seattle, for Methane and Nitrogen Cycles, sample SIP 13C-methane anaerobic+nitrate (Anaerobic + nitrate SIP Nov 2010 with PE)	Freshwater	Lake Washington Sediment
3	WSSedA1BaDRAFT_10332591	WSSedA1BaDRAFT_10332591	Tannase and feruloyl esterase	Wetland microbial communities from Twitchell Island in the Sacramento Delta, sample from surface sediment Feb2011 Site A1 Bulk (Wetland Surface Sediment Feb2011 Site A1 Bulk, Assem Ctgs IBYY 2011 Sep Assem)	Freshwater	Groundwater
3	WSSedA1BaDRAFT_10599771	WSSedA1BaDRAFT_10599771	Tannase and feruloyl esterase	Wetland microbial communities from Twitchell Island in the Sacramento Delta, sample from surface sediment Feb2011 Site A1 Bulk (Wetland Surface Sediment Feb2011 Site A1 Bulk, Assem Ctgs IBYY 2011 Sep Assem)	Freshwater	Groundwater
3	WSSedB1CaDRAFT_100068592	WSSedB1CaDRAFT_100068592	Tannase and feruloyl esterase	Wetland microbial communities from Twitchell Island in the Sacramento Delta, sample from surface sediment Feb2011 Site B1 Cattail (Wetland	Freshwater	Groundwater

				Surface Sediment Feb2011 Site B1 Cattail, Assem Ctgs Sep 2011 assem)		
3	WSSedL1CaDRAFT_1000004448	WSSedL1CaDRAFT_1000004448	Tannase and feruloyl esterase	Wetland microbial communities from Twitchell Island in the Sacramento Delta, sample from surface sediment Feb2011 Site L1 Cattail (Wetland Surface Sediment Feb2011 Site L1 Cattail, Assem Ctgs Sep 2011 assem)	Freshwater	Groundwater
3	WSSedL1CaDRAFT_1000004449	WSSedL1CaDRAFT_1000004449	Tannase and feruloyl esterase	Wetland microbial communities from Twitchell Island in the Sacramento Delta, sample from surface sediment Feb2011 Site L1 Cattail (Wetland Surface Sediment Feb2011 Site L1 Cattail, Assem Ctgs Sep 2011 assem)	Freshwater	Groundwater
3	SRS015264_WUGC_scaffold_68364__gene_160067	SRS015264_WUGC_scaffold_68364__gene_160067	feruloyl esterase	Human Stool microbiome from visit number 1 of subject 763982056	Human	Stool
3	SRS078176_LANL_scaffold_22345__gene_51104	SRS078176_LANL_scaffold_22345__gene_51104	feruloyl esterase	Human Stool microbiome from visit number 3 of subject 159227541	Human	Stool
3	SPBFG_433620	SPBFG_433620	Tannase and feruloyl esterase.	Dendroctonus frontalis Fungal community	Insect	Southern Pine Beetle
3	SWWA_01059590	SWWA_01059590	Tannase and feruloyl esterase.	Sirex noctilio microbiome from Pennsylvania, sample of adult community (Adult June 2010 assembly)	Insect	Sirex noctilio
3	2155498707	CHXU_0076.00001830	Tannase and feruloyl esterase.	16-5 In Plume (16-5 In Plume)	Marine	1125m
3	KGI_S2_ANT04_2345mDRAFT_10014345	KGI_S2_ANT04_2345mDRAFT_10014345	Tannase and feruloyl esterase	King George Island site S2 sample ANT 04_23.45m (King George Island site S2 sample ANT 04_23.45m, Dec 2011 Assem)	Marine	23.45m

3	F10SI0310_0113.00002470	F10SI0310_0113.00002470	Tannase and feruloyl esterase.	Line P sample F_10_SI03_10 (F_10_S103_10)	Marine	10m
3	F10SI0310_0360.00004330	F10SI0310_0360.00004330	Tannase and feruloyl esterase.	Line P sample F_10_SI03_10 (F_10_S103_10)	Marine	10m
3	F10SI0310_0492.00008120	F10SI0310_0492.00008120	Tannase and feruloyl esterase.	Line P sample F_10_SI03_10 (F_10_S103_10)	Marine	10m
3	F10SI0310a_00556440	F10SI0310a_00556440	Tannase and feruloyl esterase.	Line P sample F_10_SI03_10 (sample_F_10_SI03_10 June 2011 assem)	Marine	10m
3	A09P04_1300_00512320	A09P04_1300_00512320	Tannase and feruloyl esterase.	Line P sample_A_09_P04_1300 (A_09_P04_1300 June 2011 assembly)	Marine	1300m
3	F10SI03135_0023.00008280	F10SI03135_0023.00008280	Tannase and feruloyl esterase.	Line P sample_F_10_SI03_135 (F_10_SI03_135)	Marine	135m
3	F10SI03135_0125.00015780	F10SI03135_0125.00015780	Tannase and feruloyl esterase.	Line P sample_F_10_SI03_135 (F_10_SI03_135)	Marine	135m
3	F10SI03150a_00061500	F10SI03150a_00061500	Tannase and feruloyl esterase.	Line P sample_F_10_SI03_150 (sample_F_10_SI03_150 June 2011 assem)	Marine	150m
3	J08P26500_0222.00000790	J08P26500_0222.00000790	Tannase and feruloyl esterase.	Line P sample_J_08_P26_500 (J_08_P26_500)	Marine	500m
3	M940CN_10158471	M940CN_10158471	Tannase and feruloyl	Microbial Communities from Little Sippewissett Salt Marsh, Woods Hole, MA that are anoxygenic and	Marine	Littoral region Lake Sippewisse

			esterase	photosynthetic, Marine photosynthetic community that grows at 940nm (Marine_940nm_cellulose)		tt
3	M940CN_10244571	M940CN_10244571	Tannase and feruloyl esterase	Microbial Communities from Little Sippewissett Salt Marsh, Woods Hole, MA that are anoxygenic and photosynthetic, Marine photosynthetic community that grows at 940nm (Marine_940nm_cellulose)	Marine	Littoral region Lake Sippewissett
3	2200943446	MRSJC2b_24655	Tannase and feruloyl esterase.	Mesophilic rice straw/compost enrichment metagenome: eDNA_1 (Mesophilic 454/Illumina Combined June 2011 assem)	Plant Compost	Rice-Straw
3	2200957225	MRSJC2b_257110	Tannase and feruloyl esterase.	Mesophilic rice straw/compost enrichment metagenome: eDNA_1 (Mesophilic 454/Illumina Combined June 2011 assem)	Plant Compost	Rice-Straw
3	2201040861	MRSJC2b_303910	Tannase and feruloyl esterase.	Mesophilic rice straw/compost enrichment metagenome: eDNA_1 (Mesophilic 454/Illumina Combined June 2011 assem)	Plant Compost	Rice-Straw
3	2201071425	MRSJC2b_323362	Tannase and feruloyl esterase.	Mesophilic rice straw/compost enrichment metagenome: eDNA_1 (Mesophilic 454/Illumina Combined June 2011 assem)	Plant Compost	Rice-Straw
3	2201423657	MRSJC2b_848511	Predicted peptidase	Mesophilic rice straw/compost enrichment metagenome: eDNA_1 (Mesophilic 454/Illumina Combined June 2011 assem)	Plant Compost	Rice-Straw
3	2201623187	MRSJC2b_1676781	Tannase and feruloyl esterase.	Mesophilic rice straw/compost enrichment metagenome: eDNA_1 (Mesophilic 454/Illumina Combined June 2011 assem)	Plant Compost	Rice-Straw
3	2201639334	MRSJC2b_17567	Tannase and feruloyl esterase.	Mesophilic rice straw/compost enrichment metagenome: eDNA_1 (Mesophilic 454/Illumina Combined June 2011 assem)	Plant Compost	Rice-Straw
3	2200318239	TRSJC2b_28123	Protease II	Thermophilic rice straw/compost enrichment metagenome: eDNA_2 (Thermophilic 454/Illumina Combined June 2011 assem)	Plant Compost	Rice-Straw

3	2200325443	TRSJC2b_30134	Tannase and feruloyl esterase.	Thermophilic rice straw/compost enrichment metagenome: eDNA_2 (Thermophilic 454/Illumina Combined June 2011 assem)	Plant Compost	Rice-Straw
3	2200675764	TRSJC2b_20017	Tannase and feruloyl esterase.	Thermophilic rice straw/compost enrichment metagenome: eDNA_2 (Thermophilic 454/Illumina Combined June 2011 assem)	Plant Compost	Rice-Straw
3	BBAY34_10454471	BBAY34_10454471	Tannase and feruloyl esterase	Rhopaloeides odorabile metagenome replicate 1 (BBAY34)	Porifera	Marine Sponge
3	ARcpr5oldR_0117121	ARcpr5oldR_0117121	Tannase and feruloyl esterase	Arabidopsis rhizosphere microbial communities from University of North Carolina, sample from Arabidopsis cpr5 old rhizosphere (Arabidopsis cpr5 old rhizosphere, Nov 2011 assem)	Rhizosphere	Arabidopsis
3	ARSoilOldRDRAFT_0232261	ARSoilOldRDRAFT_0232261	Tannase and feruloyl esterase	Arabidopsis rhizosphere microbial communities from University of North Carolina, sample from Arabidopsis soil old (Arabidopsis soil old, Nov 2011 assem)	Rhizosphere	Arabidopsis
3	MBSR1b_0008.00001780	MBSR1b_0008.00001780	Tannase and feruloyl esterase.	Miscanthus rhizosphere microbial communities from Kellogg Biological Station, MSU, sample Bulk Soil Replicate 1 : eDNA_1 (Bulk soil 1 April 2011 assembly)	Rhizosphere	Miscanthus
3	MBSR1b_0009.00008130	MBSR1b_0009.00008130	Tannase and feruloyl esterase.	Miscanthus rhizosphere microbial communities from Kellogg Biological Station, MSU, sample Bulk Soil Replicate 1 : eDNA_1 (Bulk soil 1 April 2011 assembly)	Rhizosphere	Miscanthus
3	MBSR1b_0049.00006700	MBSR1b_0049.00006700	Predicted peptidase	Miscanthus rhizosphere microbial communities from Kellogg Biological Station, MSU, sample Bulk Soil Replicate 1 : eDNA_1 (Bulk soil 1 April 2011 assembly)	Rhizosphere	Miscanthus
3	MBSR1b_0100.00004010	MBSR1b_0100.00004010	Tannase and feruloyl esterase.	Miscanthus rhizosphere microbial communities from Kellogg Biological Station, MSU, sample Bulk Soil Replicate 1 : eDNA_1 (Bulk soil 1 April 2011 assembly)	Rhizosphere	Miscanthus
3	MBSR1b_0112.00006620	MBSR1b_0112.00006620	Predicted peptidase	Miscanthus rhizosphere microbial communities from Kellogg Biological Station, MSU, sample Bulk Soil Replicate 1 : eDNA_1 (Bulk soil 1 April	Rhizosphere	Miscanthus

				2011 assembly)		
3	MBSR1b_0278.00005670	MBSR1b_0278.00005670	Tannase and feruloyl esterase.	Miscanthus rhizosphere microbial communities from Kellogg Biological Station, MSU, sample Bulk Soil Replicate 1 : eDNA_1 (Bulk soil 1 April 2011 assembly)	Rhizosphere	Miscanthus
3	MBSR1b_0297.00000760	MBSR1b_0297.00000760	Lysophospholipase	Miscanthus rhizosphere microbial communities from Kellogg Biological Station, MSU, sample Bulk Soil Replicate 1 : eDNA_1 (Bulk soil 1 April 2011 assembly)	Rhizosphere	Miscanthus
3	MBSR1b_0335.00000190	MBSR1b_0335.00000190	Tannase and feruloyl esterase.	Miscanthus rhizosphere microbial communities from Kellogg Biological Station, MSU, sample Bulk Soil Replicate 1 : eDNA_1 (Bulk soil 1 April 2011 assembly)	Rhizosphere	Miscanthus
3	MBSR1b_0342.00007690	MBSR1b_0342.00007690	Tannase and feruloyl esterase.	Miscanthus rhizosphere microbial communities from Kellogg Biological Station, MSU, sample Bulk Soil Replicate 1 : eDNA_1 (Bulk soil 1 April 2011 assembly)	Rhizosphere	Miscanthus
3	MBSR1b_0387.00006080	MBSR1b_0387.00006080	Tannase and feruloyl esterase.	Miscanthus rhizosphere microbial communities from Kellogg Biological Station, MSU, sample Bulk Soil Replicate 1 : eDNA_1 (Bulk soil 1 April 2011 assembly)	Rhizosphere	Miscanthus
3	MBSR1b_0471.00007240	MBSR1b_0471.00007240	Tannase and feruloyl esterase.	Miscanthus rhizosphere microbial communities from Kellogg Biological Station, MSU, sample Bulk Soil Replicate 1 : eDNA_1 (Bulk soil 1 April 2011 assembly)	Rhizosphere	Miscanthus
3	MBSR1b_0599.00004360	MBSR1b_0599.00004360	Tannase and feruloyl esterase.	Miscanthus rhizosphere microbial communities from Kellogg Biological Station, MSU, sample Bulk Soil Replicate 1 : eDNA_1 (Bulk soil 1 April 2011 assembly)	Rhizosphere	Miscanthus
3	MBSR1b_0622.00002300	MBSR1b_0622.00002300	Tannase and feruloyl esterase.	Miscanthus rhizosphere microbial communities from Kellogg Biological Station, MSU, sample Bulk Soil Replicate 1 : eDNA_1 (Bulk soil 1 April 2011 assembly)	Rhizosphere	Miscanthus
3	MBSR1b_0634.00007340	MBSR1b_0634.00007340	Tannase and feruloyl esterase.	Miscanthus rhizosphere microbial communities from Kellogg Biological Station, MSU, sample Bulk Soil Replicate 1 : eDNA_1 (Bulk soil 1 April 2011 assembly)	Rhizosphere	Miscanthus
3	MBSR1b	MBSR1b	Tannase	Miscanthus rhizosphere microbial	Rhizospher	Miscanthus

	_0671.00006880	_0671.00006880	and feruloyl esterase.	communities from Kellogg Biological Station, MSU, sample Bulk Soil Replicate 1 : eDNA_1 (Bulk soil 1 April 2011 assembly)	e	
3	MBSR1b_0702.00006180	MBSR1b_0702.00006180	Tannase and feruloyl esterase.	Miscanthus rhizosphere microbial communities from Kellogg Biological Station, MSU, sample Bulk Soil Replicate 1 : eDNA_1 (Bulk soil 1 April 2011 assembly)	Rhizosphere	Miscanthus
3	MBSR1b_0746.00000060	MBSR1b_0746.00000060	Tannase and feruloyl esterase.	Miscanthus rhizosphere microbial communities from Kellogg Biological Station, MSU, sample Bulk Soil Replicate 1 : eDNA_1 (Bulk soil 1 April 2011 assembly)	Rhizosphere	Miscanthus
3	MBSR1b_0753.00006860	MBSR1b_0753.00006860	Tannase and feruloyl esterase.	Miscanthus rhizosphere microbial communities from Kellogg Biological Station, MSU, sample Bulk Soil Replicate 1 : eDNA_1 (Bulk soil 1 April 2011 assembly)	Rhizosphere	Miscanthus
3	MBSR1b_0782.00006940	MBSR1b_0782.00006940	Tannase and feruloyl esterase.	Miscanthus rhizosphere microbial communities from Kellogg Biological Station, MSU, sample Bulk Soil Replicate 1 : eDNA_1 (Bulk soil 1 April 2011 assembly)	Rhizosphere	Miscanthus
3	MBSR1b_0879.00008420	MBSR1b_0879.00008420	Tannase and feruloyl esterase.	Miscanthus rhizosphere microbial communities from Kellogg Biological Station, MSU, sample Bulk Soil Replicate 1 : eDNA_1 (Bulk soil 1 April 2011 assembly)	Rhizosphere	Miscanthus
3	MBSR1b_0887.00006700	MBSR1b_0887.00006700	Tannase and feruloyl esterase.	Miscanthus rhizosphere microbial communities from Kellogg Biological Station, MSU, sample Bulk Soil Replicate 1 : eDNA_1 (Bulk soil 1 April 2011 assembly)	Rhizosphere	Miscanthus
3	MBSR1b_0894.00001650	MBSR1b_0894.00001650	Tannase and feruloyl esterase.	Miscanthus rhizosphere microbial communities from Kellogg Biological Station, MSU, sample Bulk Soil Replicate 1 : eDNA_1 (Bulk soil 1 April 2011 assembly)	Rhizosphere	Miscanthus
3	MBSR1b_0933.00000780	MBSR1b_0933.00000780	Tannase and feruloyl esterase.	Miscanthus rhizosphere microbial communities from Kellogg Biological Station, MSU, sample Bulk Soil Replicate 1 : eDNA_1 (Bulk soil 1 April 2011 assembly)	Rhizosphere	Miscanthus
3	MRS1b_0075.00000890	MRS1b_0075.00000890	Tannase and feruloyl esterase	Miscanthus rhizosphere microbial communities from Kellogg Biological Station, MSU, sample Rhizosphere Soil Replicate 1: eDNA_1 (Rhizosphere	Rhizosphere	Miscanthus

			.	replicate 1 April 2011 assembly)		
3	MRS1b_0338.00000610	MRS1b_0338.00000610	Tannase and feruloyl esterase.	Miscanthus rhizosphere microbial communities from Kellogg Biological Station, MSU, sample Rhizosphere Soil Replicate 1: eDNA_1 (Rhizosphere replicate 1 April 2011 assembly)	Rhizosphere	Miscanthus
3	MRS1b_0388.00002210	MRS1b_0388.00002210	Tannase and feruloyl esterase.	Miscanthus rhizosphere microbial communities from Kellogg Biological Station, MSU, sample Rhizosphere Soil Replicate 1: eDNA_1 (Rhizosphere replicate 1 April 2011 assembly)	Rhizosphere	Miscanthus
3	MRS1b_0394.00000780	MRS1b_0394.00000780	Tannase and feruloyl esterase.	Miscanthus rhizosphere microbial communities from Kellogg Biological Station, MSU, sample Rhizosphere Soil Replicate 1: eDNA_1 (Rhizosphere replicate 1 April 2011 assembly)	Rhizosphere	Miscanthus
3	MRS1b_0526.00002930	MRS1b_0526.00002930	Tannase and feruloyl esterase.	Miscanthus rhizosphere microbial communities from Kellogg Biological Station, MSU, sample Rhizosphere Soil Replicate 1: eDNA_1 (Rhizosphere replicate 1 April 2011 assembly)	Rhizosphere	Miscanthus
3	MRS1b_0688.00003340	MRS1b_0688.00003340	Lysophospholipase	Miscanthus rhizosphere microbial communities from Kellogg Biological Station, MSU, sample Rhizosphere Soil Replicate 1: eDNA_1 (Rhizosphere replicate 1 April 2011 assembly)	Rhizosphere	Miscanthus
3	MRS1b_0841.00000780	MRS1b_0841.00000780	Tannase and feruloyl esterase.	Miscanthus rhizosphere microbial communities from Kellogg Biological Station, MSU, sample Rhizosphere Soil Replicate 1: eDNA_1 (Rhizosphere replicate 1 April 2011 assembly)	Rhizosphere	Miscanthus
3	SwBSRL2_0090.00005360	SwBSRL2_0090.00005360	Tannase and feruloyl esterase.	Switchgrass rhizosphere microbial community from Michigan, US, sample from Rose Lake bulk soil RL2 (Bulk soil RL2 April 2011 assembly)	Rhizosphere	Switchgrass
3	SwBSRL2_0176.00000590	SwBSRL2_0176.00000590	Tannase and feruloyl esterase.	Switchgrass rhizosphere microbial community from Michigan, US, sample from Rose Lake bulk soil RL2 (Bulk soil RL2 April 2011 assembly)	Rhizosphere	Switchgrass
3	SwBSRL2_0218.00004140	SwBSRL2_0218.00004140	Tannase and feruloyl esterase.	Switchgrass rhizosphere microbial community from Michigan, US, sample from Rose Lake bulk soil RL2 (Bulk soil RL2 April 2011 assembly)	Rhizosphere	Switchgrass
3	SwBSRL	SwBSRL2	Tannase	Switchgrass rhizosphere microbial	Rhizospher	Switchgras

	2_0251.00005050	_0251.00005050	and feruloyl esterase.	community from Michigan, US, sample from Rose Lake bulk soil RL2 (Bulk soil RL2 April 2011 assembly)	e	s
3	SwBSRL2_0307.00002290	SwBSRL2_0307.00002290	Tannase and feruloyl esterase.	Switchgrass rhizosphere microbial community from Michigan, US, sample from Rose Lake bulk soil RL2 (Bulk soil RL2 April 2011 assembly)	Rhizosphere	Switchgrass
3	SwBSRL2_0396.00002080	SwBSRL2_0396.00002080	Tannase and feruloyl esterase.	Switchgrass rhizosphere microbial community from Michigan, US, sample from Rose Lake bulk soil RL2 (Bulk soil RL2 April 2011 assembly)	Rhizosphere	Switchgrass
3	SwBSRL2_0604.00004110	SwBSRL2_0604.00004110	Tannase and feruloyl esterase.	Switchgrass rhizosphere microbial community from Michigan, US, sample from Rose Lake bulk soil RL2 (Bulk soil RL2 April 2011 assembly)	Rhizosphere	Switchgrass
3	SwBSRL2_0927.00000460	SwBSRL2_0927.00000460	Tannase and feruloyl esterase.	Switchgrass rhizosphere microbial community from Michigan, US, sample from Rose Lake bulk soil RL2 (Bulk soil RL2 April 2011 assembly)	Rhizosphere	Switchgrass
3	SwBSRL2_0981.00000880	SwBSRL2_0981.00000880	Tannase and feruloyl esterase.	Switchgrass rhizosphere microbial community from Michigan, US, sample from Rose Lake bulk soil RL2 (Bulk soil RL2 April 2011 assembly)	Rhizosphere	Switchgrass
3	SwRhRL2b_0040.00000760	SwRhRL2b_0040.00000760	Tannase and feruloyl esterase.	Switchgrass rhizosphere microbial community from Michigan, US, sample from Rose Lake rhizosphere RL2 (Rhizosphere RL2 April 2011 assembly)	Rhizosphere	Switchgrass
3	SwRhRL2b_0101.00002060	SwRhRL2b_0101.00002060	Tannase and feruloyl esterase.	Switchgrass rhizosphere microbial community from Michigan, US, sample from Rose Lake rhizosphere RL2 (Rhizosphere RL2 April 2011 assembly)	Rhizosphere	Switchgrass
3	SwRhRL2b_0324.00006170	SwRhRL2b_0324.00006170	Tannase and feruloyl esterase.	Switchgrass rhizosphere microbial community from Michigan, US, sample from Rose Lake rhizosphere RL2 (Rhizosphere RL2 April 2011 assembly)	Rhizosphere	Switchgrass
3	SwRhRL2b_0475.00000180	SwRhRL2b_0475.00000180	Tannase and feruloyl esterase	Switchgrass rhizosphere microbial community from Michigan, US, sample from Rose Lake rhizosphere RL2 (Rhizosphere RL2 April 2011 assembly)	Rhizosphere	Switchgrass

			.			
3	SwRhRL2b_0768.00009200	SwRhRL2b_0768.00009200	Tannase and feruloyl esterase.	Switchgrass rhizosphere microbial community from Michigan, US, sample from Rose Lake rhizosphere RL2 (Rhizosphere RL2 April 2011 assembly)	Rhizosphere	Switchgrass
3	SwRhRL3b_0187.00001650	SwRhRL3b_0187.00001650	Tannase and feruloyl esterase.	Switchgrass rhizosphere microbial community from Michigan, US, sample from Rose Lake RL3 (Rhizosphere RL3 April 2011 assembly)	Rhizosphere	Switchgrass
3	SwRhRL3b_0349.00001160	SwRhRL3b_0349.00001160	Tannase and feruloyl esterase.	Switchgrass rhizosphere microbial community from Michigan, US, sample from Rose Lake RL3 (Rhizosphere RL3 April 2011 assembly)	Rhizosphere	Switchgrass
3	prs_01848950	prs_01848950	Tannase and feruloyl esterase.	Luquillo Experimental Forest Soil, Puerto Rico	Soil	Forest Soil
3	prs_06112210	prs_06112210	Tannase and feruloyl esterase.	Luquillo Experimental Forest Soil, Puerto Rico	Soil	Forest Soil
3	ICChiseqgaiiDRAFT_03449211	ICChiseqgaiiDRAFT_03449211	Tannase and feruloyl esterase	Soil microbial communities from Great Prairies, sample from Iowa, Continuous Corn soil (Iowa, Continuous Corn soil, Jan 2012 Assem MSU hiseq+gaii)	Soil	Corn Soil
3	ICChiseqgaiiDRAFT_05089841	ICChiseqgaiiDRAFT_05089841	Tannase and feruloyl esterase	Soil microbial communities from Great Prairies, sample from Iowa, Continuous Corn soil (Iowa, Continuous Corn soil, Jan 2012 Assem MSU hiseq+gaii)	Soil	Corn Soil
3	ICChiseqgaiiDRAFT_06047071	ICChiseqgaiiDRAFT_06047071	Tannase and feruloyl esterase	Soil microbial communities from Great Prairies, sample from Iowa, Continuous Corn soil (Iowa, Continuous Corn soil, Jan 2012 Assem MSU hiseq+gaii)	Soil	Corn Soil
3	ICChiseqgaiiDRAFT_06059821	ICChiseqgaiiDRAFT_06059821	Tannase and feruloyl esterase	Soil microbial communities from Great Prairies, sample from Iowa, Continuous Corn soil (Iowa, Continuous Corn soil, Jan 2012 Assem MSU hiseq+gaii)	Soil	Corn Soil
3	ICChiseqgaiiDRAFT_07103801	ICChiseqgaiiDRAFT_07103801	Tannase and feruloyl esterase	Soil microbial communities from Great Prairies, sample from Iowa, Continuous Corn soil (Iowa, Continuous Corn soil, Jan 2012 Assem MSU hiseq+gaii)	Soil	Corn Soil
3	ICChiseq	ICChiseq	Tannase	Soil microbial communities from Great	Soil	Corn Soil

	gaiiDRAFT_07338781	gaiiDRAFT_07338781	and feruloyl esterase	Prairies, sample from Iowa, Continuous Corn soil (Iowa, Continuous Corn soil, Jan 2012 Assem MSU hiseq+gaii)		
3	ICChiseqgaiiDRAFT_07434911	ICChiseqgaiiDRAFT_07434911	Tannase and feruloyl esterase	Soil microbial communities from Great Prairies, sample from Iowa, Continuous Corn soil (Iowa, Continuous Corn soil, Jan 2012 Assem MSU hiseq+gaii)	Soil	Corn Soil
3	ICChiseqgaiiDRAFT_07861131	ICChiseqgaiiDRAFT_07861131	Tannase and feruloyl esterase	Soil microbial communities from Great Prairies, sample from Iowa, Continuous Corn soil (Iowa, Continuous Corn soil, Jan 2012 Assem MSU hiseq+gaii)	Soil	Corn Soil
3	ICChiseqgaiiDRAFT_07921801	ICChiseqgaiiDRAFT_07921801	Tannase and feruloyl esterase	Soil microbial communities from Great Prairies, sample from Iowa, Continuous Corn soil (Iowa, Continuous Corn soil, Jan 2012 Assem MSU hiseq+gaii)	Soil	Corn Soil
3	ICChiseqgaiiDRAFT_07922081	ICChiseqgaiiDRAFT_07922081	Tannase and feruloyl esterase	Soil microbial communities from Great Prairies, sample from Iowa, Continuous Corn soil (Iowa, Continuous Corn soil, Jan 2012 Assem MSU hiseq+gaii)	Soil	Corn Soil
3	ICChiseqgaiiDRAFT_07936911	ICChiseqgaiiDRAFT_07936911	Tannase and feruloyl esterase	Soil microbial communities from Great Prairies, sample from Iowa, Continuous Corn soil (Iowa, Continuous Corn soil, Jan 2012 Assem MSU hiseq+gaii)	Soil	Corn Soil
3	ICChiseqgaiiDRAFT_07961191	ICChiseqgaiiDRAFT_07961191	Tannase and feruloyl esterase	Soil microbial communities from Great Prairies, sample from Iowa, Continuous Corn soil (Iowa, Continuous Corn soil, Jan 2012 Assem MSU hiseq+gaii)	Soil	Corn Soil
3	ICChiseqgaiiDRAFT_07961201	ICChiseqgaiiDRAFT_07961201	Tannase and feruloyl esterase	Soil microbial communities from Great Prairies, sample from Iowa, Continuous Corn soil (Iowa, Continuous Corn soil, Jan 2012 Assem MSU hiseq+gaii)	Soil	Corn Soil
3	ICChiseqgaiiDRAFT_08137932	ICChiseqgaiiDRAFT_08137932	Tannase and feruloyl esterase	Soil microbial communities from Great Prairies, sample from Iowa, Continuous Corn soil (Iowa, Continuous Corn soil, Jan 2012 Assem MSU hiseq+gaii)	Soil	Corn Soil
3	ICChiseqgaiiDRAFT_08295223	ICChiseqgaiiDRAFT_08295223	Tannase and feruloyl esterase	Soil microbial communities from Great Prairies, sample from Iowa, Continuous Corn soil (Iowa, Continuous Corn soil, Jan 2012 Assem MSU hiseq+gaii)	Soil	Corn Soil
3	ICChiseqgaiiDRAFT_08307961	ICChiseqgaiiDRAFT_08307961	Tannase and feruloyl esterase	Soil microbial communities from Great Prairies, sample from Iowa, Continuous Corn soil (Iowa, Continuous Corn soil, Jan 2012 Assem MSU hiseq+gaii)	Soil	Corn Soil
3	ICChiseqgaiiDRAFT_08548472	ICChiseqgaiiDRAFT_08548472	Tannase and feruloyl esterase	Soil microbial communities from Great Prairies, sample from Iowa, Continuous Corn soil (Iowa, Continuous Corn soil, Jan 2012 Assem MSU hiseq+gaii)	Soil	Corn Soil

3	ICChiseqgaiiDRAFT_08652281	ICChiseqgaiiDRAFT_08652281	Tannase and feruloyl esterase	Soil microbial communities from Great Prairies, sample from Iowa, Continuous Corn soil (Iowa, Continuous Corn soil, Jan 2012 Assem MSU hiseq+gaii)	Soil	Corn Soil
3	ICChiseqgaiiDRAFT_08701301	ICChiseqgaiiDRAFT_08701301	Tannase and feruloyl esterase	Soil microbial communities from Great Prairies, sample from Iowa, Continuous Corn soil (Iowa, Continuous Corn soil, Jan 2012 Assem MSU hiseq+gaii)	Soil	Corn Soil
3	ICChiseqgaiiDRAFT_08840961	ICChiseqgaiiDRAFT_08840961	Tannase and feruloyl esterase	Soil microbial communities from Great Prairies, sample from Iowa, Continuous Corn soil (Iowa, Continuous Corn soil, Jan 2012 Assem MSU hiseq+gaii)	Soil	Corn Soil
3	ICChiseqgaiiDRAFT_08841541	ICChiseqgaiiDRAFT_08841541	Tannase and feruloyl esterase	Soil microbial communities from Great Prairies, sample from Iowa, Continuous Corn soil (Iowa, Continuous Corn soil, Jan 2012 Assem MSU hiseq+gaii)	Soil	Corn Soil
3	ICChiseqgaiiDRAFT_08898833	ICChiseqgaiiDRAFT_08898833	Tannase and feruloyl esterase	Soil microbial communities from Great Prairies, sample from Iowa, Continuous Corn soil (Iowa, Continuous Corn soil, Jan 2012 Assem MSU hiseq+gaii)	Soil	Corn Soil
3	ICChiseqgaiiDRAFT_16734081	ICChiseqgaiiDRAFT_16734081	Tannase and feruloyl esterase	Soil microbial communities from Great Prairies, sample from Iowa, Continuous Corn soil (Iowa, Continuous Corn soil, Jan 2012 Assem MSU hiseq+gaii)	Soil	Corn Soil
3	ICChiseqgaiiDRAFT_21329571	ICChiseqgaiiDRAFT_21329571	Tannase and feruloyl esterase	Soil microbial communities from Great Prairies, sample from Iowa, Continuous Corn soil (Iowa, Continuous Corn soil, Jan 2012 Assem MSU hiseq+gaii)	Soil	Corn Soil
3	ICChiseqgaiiDRAFT_22213772	ICChiseqgaiiDRAFT_22213772	Tannase and feruloyl esterase	Soil microbial communities from Great Prairies, sample from Iowa, Continuous Corn soil (Iowa, Continuous Corn soil, Jan 2012 Assem MSU hiseq+gaii)	Soil	Corn Soil
3	ICChiseqgaiiDRAFT_22340261	ICChiseqgaiiDRAFT_22340261	Tannase and feruloyl esterase	Soil microbial communities from Great Prairies, sample from Iowa, Continuous Corn soil (Iowa, Continuous Corn soil, Jan 2012 Assem MSU hiseq+gaii)	Soil	Corn Soil
3	ICChiseqgaiiDRAFT_22348411	ICChiseqgaiiDRAFT_22348411	Tannase and feruloyl esterase	Soil microbial communities from Great Prairies, sample from Iowa, Continuous Corn soil (Iowa, Continuous Corn soil, Jan 2012 Assem MSU hiseq+gaii)	Soil	Corn Soil
3	ICChiseqgaiiDRAFT_23296761	ICChiseqgaiiDRAFT_23296761	Tannase and feruloyl esterase	Soil microbial communities from Great Prairies, sample from Iowa, Continuous Corn soil (Iowa, Continuous Corn soil, Jan 2012 Assem MSU hiseq+gaii)	Soil	Corn Soil
3	ICChiseqgaiiDRAFT_23518582	ICChiseqgaiiDRAFT_23518582	Tannase and feruloyl esterase	Soil microbial communities from Great Prairies, sample from Iowa, Continuous Corn soil (Iowa, Continuous Corn soil, Jan 2012 Assem MSU hiseq+gaii)	Soil	Corn Soil

3	ICChiseqgaiiDRAFT_23577921	ICChiseqgaiiDRAFT_23577921	Tannase and feruloyl esterase	Soil microbial communities from Great Prairies, sample from Iowa, Continuous Corn soil (Iowa, Continuous Corn soil, Jan 2012 Assem MSU hiseq+gaii)	Soil	Corn Soil
3	ICChiseqgaiiDRAFT_23592391	ICChiseqgaiiDRAFT_23592391	Tannase and feruloyl esterase	Soil microbial communities from Great Prairies, sample from Iowa, Continuous Corn soil (Iowa, Continuous Corn soil, Jan 2012 Assem MSU hiseq+gaii)	Soil	Corn Soil
3	ICChiseqgaiiDRAFT_23748321	ICChiseqgaiiDRAFT_23748321	Tannase and feruloyl esterase	Soil microbial communities from Great Prairies, sample from Iowa, Continuous Corn soil (Iowa, Continuous Corn soil, Jan 2012 Assem MSU hiseq+gaii)	Soil	Corn Soil
3	ICChiseqgaiiDRAFT_23841332	ICChiseqgaiiDRAFT_23841332	Tannase and feruloyl esterase	Soil microbial communities from Great Prairies, sample from Iowa, Continuous Corn soil (Iowa, Continuous Corn soil, Jan 2012 Assem MSU hiseq+gaii)	Soil	Corn Soil
3	ICChiseqgaiiDRAFT_24333352	ICChiseqgaiiDRAFT_24333352	Tannase and feruloyl esterase	Soil microbial communities from Great Prairies, sample from Iowa, Continuous Corn soil (Iowa, Continuous Corn soil, Jan 2012 Assem MSU hiseq+gaii)	Soil	Corn Soil
3	2001272573		hypothetical protein	Soil microbial communities from Minnesota Farm	Soil	Farm Silage Surface Soil
3	2001318000		hypothetical protein	Soil microbial communities from Minnesota Farm	Soil	Farm Silage Surface Soil
3	2001321339		hypothetical protein	Soil microbial communities from Minnesota Farm	Soil	Farm Silage Surface Soil
3	2001358121		hypothetical protein	Soil microbial communities from Minnesota Farm	Soil	Farm Silage Surface Soil
3	2001374400		hypothetical protein	Soil microbial communities from Minnesota Farm	Soil	Farm Silage Surface Soil
3	A2_c1_01173610	A2_c1_01173610	Tannase and feruloyl esterase.	Soil microbial communities from permafrost in Bonanza Creek, Alaska, sample from Active Layer A2 (A2_CLC_pe)	Soil	Permafrost
3	A2_c1_01176950	A2_c1_01176950	Tannase and feruloyl esterase.	Soil microbial communities from permafrost in Bonanza Creek, Alaska, sample from Active Layer A2 (A2_CLC_pe)	Soil	Permafrost
3	A5_c1_00521920	A5_c1_00521920	Tannase and feruloyl	Soil microbial communities from permafrost in Bonanza Creek, Alaska, sample from Active Layer A5	Soil	Permafrost

			esterase.	(A5_CLC_pe)		
3	B3_all_c_00863630	B3_all_c_00863630	Tannase and feruloyl esterase.	Soil microbial communities from permafrost in Bonanza Creek, Alaska, sample from Bog Site B3 (B3_all_CLC)	Soil	Permafrost
3	P1_C_00430530	P1_C_00430530	Tannase and feruloyl esterase.	Soil microbial communities from permafrost in Bonanza Creek, Alaska, sample from Permafrost Layer P1 (P1_CLC_pe)	Soil	Permafrost
3	P3_CLC_00743880	P3_CLC_00743880	Tannase and feruloyl esterase.	Soil microbial communities from permafrost in Bonanza Creek, Alaska, sample from Permafrost Layer P3 (P3_CLC)	Soil	Permafrost
3	P3_CLC_00782920	P3_CLC_00782920	Tannase and feruloyl esterase.	Soil microbial communities from permafrost in Bonanza Creek, Alaska, sample from Permafrost Layer P3 (P3_CLC)	Soil	Permafrost
3	FACENCA_1911310	FACENCA_1911310	Tannase and feruloyl esterase.	Soil microbial communities from sample at FACE Site 2 North Carolina CO2-	Soil	North Carolina
3	FACENCA_2065720	FACENCA_2065720	Tannase and feruloyl esterase.	Soil microbial communities from sample at FACE Site 2 North Carolina CO2-	Soil	North Carolina
3	FACENCE_4420620	FACENCE_4420620	Tannase and feruloyl esterase.	Soil microbial communities from sample at FACE Site 2 North Carolina CO2+ (North Carolina Elevated CO2)	Soil	North Carolina
3	FWIRA_07201760	FWIRA_07201760	Tannase and feruloyl esterase.	Soil microbial communities from sample at FACE Site Metagenome WIR_Amb2 (WIR_Amb2)	Soil	WIR
3	FWIRA_10269840	FWIRA_10269840	Tannase and feruloyl esterase.	Soil microbial communities from sample at FACE Site Metagenome WIR_Amb2 (WIR_Amb2)	Soil	WIR

3	FNCDEF_01099530	FNCDEF_01099530	Tannase and feruloyl esterase.	Soil microbial communities from sample at FACE Site North Carolina NCD_ElevF (NCD_ElevF)	Soil	North Carolina
3	FNCDEF_04240990	FNCDEF_04240990	Tannase and feruloyl esterase.	Soil microbial communities from sample at FACE Site North Carolina NCD_ElevF (NCD_ElevF)	Soil	North Carolina
3	FNCDEF_05251080	FNCDEF_05251080	Tannase and feruloyl esterase.	Soil microbial communities from sample at FACE Site North Carolina NCD_ElevF (NCD_ElevF)	Soil	North Carolina
3	FNCDEF_10538270	FNCDEF_10538270	Tannase and feruloyl esterase.	Soil microbial communities from sample at FACE Site North Carolina NCD_ElevF (NCD_ElevF)	Soil	North Carolina
3	FNTS007_03728110	FNTS007_03728110	Tannase and feruloyl esterase.	Soil microbial communities from sample at FACE Site NTS_007 Nevada Test Site (NTS_007)	Soil	Nevada
3	GBSCECS77c_10008529	GBSCECS77c_10008529	Tannase and feruloyl esterase	Sediment microbial communities from Great Boiling Spring, Nevada, sample from cellulolytic enrichment CS 77C (GBS Cellulolytic enrichment CS 77C sediment, Combined June 2011 assem)	Thermal Spring	Great Boiling Spring Nevada
4	BMHBC_12406440	BMHBC_12406440	Tannase and feruloyl esterase.	Biofuel metagenome 1 (454/Illumina combined assembly)	Freshwater	Lake Berkely, CA
4	TB_LI09_4DRAFT_100796731	TB_LI09_4DRAFT_100796731	Tannase and feruloyl esterase	Groundwater microbial communities from subsurface biofilms in sulfidic aquifier in Frasassi Gorge, Italy, sample from two redox zones- LI09_4 (Targeted Biofilm samples from two redox zones-LI09_4, Oct 2011 Assem)	Freshwater	Subsurface Biofilm Groundwater
4	Incfw_10151891	Incfw_10151891	Tannase and feruloyl esterase	Lentic microbial communities from Lake Waban, Wellesley MA, that are anoxygenic and photosynthetic, photosynthetic consortia incandescent light (FW_incandescent_CN)	Freshwater	Photosynthetic consortia Lentic Lake
4	WSSedB1CaDRA	WSSedB1CaDRAFT	Tannase and	Wetland microbial communities from Twitchell Island in the Sacramento	Freshwater	Groundwater

	FT_100012315	_100012315	feruloyl esterase	Delta, sample from surface sediment Feb2011 Site B1 Cattail (Wetland Surface Sediment Feb2011 Site B1 Cattail, Assem Ctgs Sep 2011 assem)		
4	WSSedL1CaDRAFT_101503131	WSSedL1CaDRAFT_101503131	Tannase and feruloyl esterase	Wetland microbial communities from Twitchell Island in the Sacramento Delta, sample from surface sediment Feb2011 Site L1 Cattail (Wetland Surface Sediment Feb2011 Site L1 Cattail, Assem Ctgs Sep 2011 assem)	Freshwater	Groundwater
4	C2659911__gene_225957	C2659911__gene_225957	tannase /feruloyl esterase family protein	Human Stool microbiome from visit number 2 of subject 763536994	Human	Stool
4	SRS022725_LANL_scaffold_115001__gene_255744	SRS022725_LANL_scaffold_115001__gene_255744	hypothetical protein	Human Supragingival plaque microbiome from visit number 1 of subject 370425937	Human	Supragingival Plaque
4	SRS015440_WUGC_scaffold_36830__gene_44997	SRS015440_WUGC_scaffold_36830__gene_44997	tannase	Human Supragingival plaque microbiome from visit number 1 of subject 764083206 replicate 1	Human	Supragingival Plaque
4	ACEFG_346400	ACEFG_346400	Tannase and feruloyl esterase.	Atta cephalotes fungus garden (ACEF)	Insect	Atta cephalotes Panama
4	CLOFG_2910980	CLOFG_2910980	Tannase and feruloyl esterase.	Cyphomyrmex longiscapus fungus garden	Insect	Cyphomyrmex longiscapus
4	CLOFG_688070	CLOFG_688070	Tannase and feruloyl esterase.	Cyphomyrmex longiscapus fungus garden	Insect	Cyphomyrmex longiscapus
4	SPBFG_690170	SPBFG_690170	Tannase and feruloyl esterase	Dendroctonus frontalis Fungal community	Insect	Southern Pine Beetle

			.			
4	DPOLFG_869650	DPOLFG_869650	Tannase and feruloyl esterase.	Dendroctonus ponderosae beetle community (MPB hybrid beetle) (Lodgepole pine)	Insect	Mountain Pine Beetle
4	AECFG_907800	AECFG_907800	Tannase and feruloyl esterase.	Fungus garden combined (combined)	Insect	Acromyrmex echinatior Panama
4	ACOFGB_1482090	ACOFGB_1482090	Tannase and feruloyl esterase.	Fungus garden microbial communities from Atta colombica in Panama, sample from fungus garden bottom (Fungus garden bottom)	Insect	Atta colombica Panama
4	ACOFGT_37500	ACOFGT_37500	Tannase and feruloyl esterase.	Fungus garden microbial communities from Atta colombica in Panama, sample from fungus garden top	Insect	Atta colombica Panama
4	ACOFGT_719190	ACOFGT_719190	Tannase and feruloyl esterase.	Fungus garden microbial communities from Atta colombica in Panama, sample from fungus garden top	Insect	Atta colombica Panama
4	2228678910	PaMGMa_13	Tannase and feruloyl esterase.	Panchlora_midgut_metagenome (Midgut August 2011 assem)	Insect	Panchlora
4	SWWA_01563680	SWWA_01563680	Tannase and feruloyl esterase.	Sirex noctilio microbiome from Pennsylvania, sample of adult community (Adult June 2010 assembly)	Insect	Sirex noctilio
4	XAGC_02507900	XAGC_02507900	Tannase and feruloyl esterase.	Xyleborus affinis microbiome from Bern, Switzerland, sample of gallery community (Gallery community)	Insect	Xyleborus affinis
4	A_09_P04_10m_0591.00003840	A_09_P04_10m_0591.00003840	Tannase and feruloyl esterase.	Line P sample_A_09_P04_10 (A_09_P04_10)	Marine	10m
4	LP_A_0	LP_A_09	Tannase	Line P sample_A_09_P04_10 (Line P	Marine	10m

	9_P04_10DRAFT_10027884	_P04_10DRAFT_10027884	and feruloyl esterase	sample_A_09_P04_10, April 2012 Assem)		
4	DelMOWin2010_1000028992	DelMOWin2010_10002899 2	Tannase and feruloyl esterase	Marine microbial communities from Delaware Coast, sample from Delaware MO Winter December 2010 (Delaware MO Winter December 2010, Nov 2011 assem)	Marine	Neritic Zone
4	2225199738		KEGG ID:K09252	Wild Panda gut microbiome from Saanxi China, sample from individual w2 (GB1)	Panda	Saanxi China
4	GBAN_859250	GBAN_859250	Tannase and feruloyl esterase.	Compost Minireactor Metagenome (final assembly)	Plant Compost	Switchgrass
4	2201331161	MRSJC2b_560471	Tannase and feruloyl esterase.	Mesophilic rice straw/compost enrichment metagenome: eDNA_1 (Mesophilic 454/Illumina Combined June 2011 assem)	Plant Compost	Rice-straw
4	2201550829	MRSJC2b_134652	Tannase and feruloyl esterase.	Mesophilic rice straw/compost enrichment metagenome: eDNA_1 (Mesophilic 454/Illumina Combined June 2011 assem)	Plant Compost	Rice-straw
4	MBSR1b_0125.00007950	MBSR1b_0125.00007950	Tannase and feruloyl esterase.	Miscanthus rhizosphere microbial communities from Kellogg Biological Station, MSU, sample Bulk Soil Replicate 1 : eDNA_1 (Bulk soil 1 April 2011 assembly)	Rhizosphere	Miscanthus
4	MBSR1b_0469.00006640	MBSR1b_0469.00006640	Tannase and feruloyl esterase.	Miscanthus rhizosphere microbial communities from Kellogg Biological Station, MSU, sample Bulk Soil Replicate 1 : eDNA_1 (Bulk soil 1 April 2011 assembly)	Rhizosphere	Miscanthus
4	MRS1b_0839.00000510	MRS1b_0839.00000510	Uncharacterized protein conserved in bacteria	Miscanthus rhizosphere microbial communities from Kellogg Biological Station, MSU, sample Rhizosphere Soil Replicate 1: eDNA_1 (Rhizosphere replicate 1 April 2011 assembly)	Rhizosphere	Miscanthus
4	MRS1b_0951.00000450	MRS1b_0951.00000450	Tannase and feruloyl esterase.	Miscanthus rhizosphere microbial communities from Kellogg Biological Station, MSU, sample Rhizosphere Soil Replicate 1: eDNA_1 (Rhizosphere replicate 1 April 2011 assembly)	Rhizosphere	Miscanthus

4	SRBS_3983230	SRBS_3983230	Tannase and feruloyl esterase.	Switchgrass rhizosphere microbial community from Michigan, US, sample from East Lansing bulk soil	Rhizosphere	Switchgrass Rhizosphere
4	SwRhRL3b_0193.00001180	SwRhRL3b_0193.00001180	Tannase and feruloyl esterase.	Switchgrass rhizosphere microbial community from Michigan, US, sample from Rose Lake RL3 (Rhizosphere RL3 April 2011 assembly)	Rhizosphere	Switchgrass Rhizosphere
4	SwRhRL3b_0492.00002050	SwRhRL3b_0492.00002050	Tannase and feruloyl esterase.	Switchgrass rhizosphere microbial community from Michigan, US, sample from Rose Lake RL3 (Rhizosphere RL3 April 2011 assembly)	Rhizosphere	Switchgrass Rhizosphere
4	SwRhRL3b_0888.00001880	SwRhRL3b_0888.00001880	Tannase and feruloyl esterase.	Switchgrass rhizosphere microbial community from Michigan, US, sample from Rose Lake RL3 (Rhizosphere RL3 April 2011 assembly)	Rhizosphere	Switchgrass Rhizosphere
4	ICChiseqgaiiDRAFT_07971852	ICChiseqgaiiDRAFT_07971852	Tannase and feruloyl esterase	Soil microbial communities from Great Prairies, sample from Iowa, Continuous Corn soil (Iowa, Continuous Corn soil, Jan 2012 Assem MSU hiseq+gaii)	Soil	Corn Soil
4	ICChiseqgaiiDRAFT_17042281	ICChiseqgaiiDRAFT_17042281	Tannase and feruloyl esterase	Soil microbial communities from Great Prairies, sample from Iowa, Continuous Corn soil (Iowa, Continuous Corn soil, Jan 2012 Assem MSU hiseq+gaii)	Soil	Corn Soil
4	2001343078		hypothetical protein	Soil microbial communities from Minnesota Farm	Soil	Minnesota
4	FNTS067_08338380	FNTS067_08338380	Tannase and feruloyl esterase.	Soil microbial communities from sample at FACE Site NTS_067 Nevada Test Site (NTS_067)	Soil	FACE
4	LCrCPGB2aDRAFT_0008013	LCrCPGB2aDRAFT_0008013	Tannase and feruloyl esterase	Soil microbial communities sample from Light Crust, Colorado Plateau, Green Butte 2 (Light Crust Colorado Plateau Green Butte 2, Oct 2011 assem)	Soil	Colorado
5	2155433491	CHXU_0135.00001710	Tannase and feruloyl esterase.	16-5 In Plume (16-5 In Plume)	Marine	1125m Gulf of Mexico
5	2155452151	CHXU_0320.00000	Tannase and	16-5 In Plume (16-5 In Plume)	Marine	1125m Gulf of Mexico

		990	feruloyl esterase.			
5	2155488937	CHXU_0671.00000750	Tannase and feruloyl esterase.	16-5 In Plume (16-5 In Plume)	Marine	1125m Gulf of Mexico
5	2155504424	CHXU_0081.00000730	Tannase and feruloyl esterase.	16-5 In Plume (16-5 In Plume)	Marine	1125m Gulf of Mexico
5	2200468225	TRSJC2b_84082	Tannase and feruloyl esterase.	Thermophilic rice straw/compost enrichment metagenome: eDNA_2 (Thermophilic 454/Illumina Combined June 2011 assem)	Plant Compost	Rice-straw
5	WSSedA1BaDRAFT_10188181	WSSedA1BaDRAFT_10188181	Tannase and feruloyl esterase	Wetland microbial communities from Twitchell Island in the Sacramento Delta, sample from surface sediment Feb2011 Site A1 Bulk (Wetland Surface Sediment Feb2011 Site A1 Bulk, Assem Ctgs IBYY 2011 Sep Assem)	Freshwater	Groundwater
6	ARSoilOldRDRAFT_0058541	ARSoilOldRDRAFT_0058541	Tannase and feruloyl esterase	Arabidopsis rhizosphere microbial communities from University of North Carolina, sample from Arabidopsis soil old (Arabidopsis soil old, Nov 2011 assem)	Rhizosphere	Arabidopsis rhizosphere
6	2006681821		Conserved hypothetical protein	Methylotrophic community from Lake Washington sediment combined (v2)	Freshwater	Lake Washington Sediment
6	MBSR1b_0216.00003820	MBSR1b_0216.00003820	Tannase and feruloyl esterase.	Miscanthus rhizosphere microbial communities from Kellogg Biological Station, MSU, sample Bulk Soil Replicate 1 : eDNA_1 (Bulk soil 1 April 2011 assembly)	Rhizosphere	Miscanthus rhizosphere
6	ICChiseqgaiiDRAFT_07580282	ICChiseqgaiiDRAFT_07580282	Tannase and feruloyl esterase	Soil microbial communities from Great Prairies, sample from Iowa, Continuous Corn soil (Iowa, Continuous Corn soil, Jan 2012 Assem MSU hiseq+gaii)	Soil	Corn Soil
6	A2_c1_00777490	A2_c1_00777490	Tannase and feruloyl esterase.	Soil microbial communities from permafrost in Bonanza Creek, Alaska, sample from Active Layer A2 (A2_CLC_pe)	Soil	Permafrost

6	A5_c1_00274690	A5_c1_00274690	Tannase and feruloyl esterase.	Soil microbial communities from permafrost in Bonanza Creek, Alaska, sample from Active Layer A5 (A5_CLC_pe)	Soil	Permafrost
6	A5_c1_01593900	A5_c1_01593900	Tannase and feruloyl esterase.	Soil microbial communities from permafrost in Bonanza Creek, Alaska, sample from Active Layer A5 (A5_CLC_pe)	Soil	Permafrost
6	NTS_CREO_AMB_4237030	NTS_CREO_AMB_4237030	Tannase and feruloyl esterase.	Soil microbial communities from sample at FACE Site 3 Nevada Test Site Creosote CO2-	Soil	0-5cm Creosote
6	XylAfBA_346770	XylAfBA_346770	Tannase and feruloyl esterase.	Xyleborus affinis microbiome from Bern, Switzerland, sample of adult community (Ambrosia beetle adult)	Insect	Xyleborus affinis Fungus associated
7	2155462174	CHXU_0420.00000360	Tannase and feruloyl esterase.	16-5 In Plume (16-5 In Plume)	Marine	1125m Gulf of Mexico
7	2155513366	CHXU_0890.00001470	Tannase and feruloyl esterase.	16-5 In Plume (16-5 In Plume)	Marine	1125m Gulf of Mexico
7	2155670091	CHXW_0916.00000630	Tannase and feruloyl esterase.	52-4 In plume (52-4 In Plume)	Marine	1210m Gulf of Mexico
7	CHXX_0518.00000990	CHXX_0518.00000990	Tannase and feruloyl esterase.	52-1 Below Plume (52-1 Below Plume)	Marine	1300m Gulf of Mexico
7	FACEORE_893960	FACEORE_893960	Tannase and feruloyl esterase.	Soil microbial communities from sample at FACE Site 5 Oak Ridge CO2+ (Oak Ridge elevated CO2)	Soil	0-5cm Oak Ridge FACE site 5
8	2155457814	CHXU_0376.00001020	Tannase and feruloyl	16-5 In Plume (16-5 In Plume)	Marine	1125m Gulf of Mexico

			esterase.			
8	2155459430	CHXU_0393.00000460	Tannase and feruloyl esterase.	16-5 In Plume (16-5 In Plume)	Marine	1125m Gulf of Mexico
8	2155517574	CHXU_0093.00000590	Tannase and feruloyl esterase.	16-5 In Plume (16-5 In Plume)	Marine	1125m Gulf of Mexico
8	ACODB_12254260	ACODB_12254260	Tannase and feruloyl esterase.	Fungus garden microbial communities from Atta colombica in Panama, sample from dump bottom (Dump bottom)	Insect	Atta colombica Panama
8	ACODT_8700690	ACODT_8700690	Tannase and feruloyl esterase.	Fungus garden microbial communities from Atta colombica in Panama, sample from dump top (Dump top)	Insect	Atta colombica Panama
8	CHXX_0580.00001060	CHXX_0580.00001060	Tannase and feruloyl esterase.	52-1 Below Plume (52-1 Below Plume)	Marine	1300m Gulf of Mexico
8	NTS_CREO_AMB_3872240	NTS_CREO_AMB_3872240	Tannase and feruloyl esterase.	Soil microbial communities from sample at FACE Site 3 Nevada Test Site Creosote CO2-	Soil	0-5cm Creosote
8	NTS_CREO_AMB_5602690	NTS_CREO_AMB_5602690	Tannase and feruloyl esterase.	Soil microbial communities from sample at FACE Site 3 Nevada Test Site Creosote CO2-	Soil	0-5cm Creosote