BIOPROSPECTING FOR THERMOSTABLE GLYCOSIDE HYDROLASES:

A METATRANSCRIPTOMIC APPROACH

A THESIS SUBMITTED TO THE GRADUATE DIVISION OF THE
UNIVERSITY OF HAWAIʻI AT MĀNOA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

IN

MICROBIOLOGY

DECEMBER 2012

By

Jeffrey R. Lyford

Thesis Committee:

Maqsudul Alam, Chairperson
Paul Patek
Stephen Freeland

Keywords: metatranscriptomics, glycoside hydrolase, bioprospecting, lignocellulose, bioethanol

Acknowledgments:

I would like to thank Dr. Maqsudul Alam and Dr. Matthew Stott for their shared effort to make this project a reality. I thank Dr. Alam for initially allowing me to study in his laboratory and providing me the opportunity to study New Zealand. I owe Dr. Stott for any development I may have made as a scientist and a writer. Without his patient guidance, I would not be close to the point I am today. To this day, no words have sufficed to illustrate my development as a scientist because of Dr. Stott's efforts. I would also like to thank my committee members, Dr. Paul Patek and Dr. Stephen Freeland for their time and their suggestions in preparation of this thesis.

I would also like to thank my lab members at GNS Science in New Zealand: Dr. Xochitl Morgan, Chien-Yu Kevin Lee, Karen Houghton, and especially Jean Power. Without their daily support through the many hard times, my sanity would have departed long ago.

I would also like to thank my family for their support and encouragement through this arduous process.

Abstract:

The conversion of cellulosic material from linear form to monomer is the rate-limiting step in current biofuels production technologies from lignocellulosic material. The renewed focus on clean, sustainable, carbon-neutral fuels has resulted in increased interest in novel cellulolytic enzymes and microbial strains, with the aim to increase the efficiency of the above conversion. However, the availability of suitable cellulolytic enzymes has been restricted by the limited number of cellulolytic microbial strains in which these enzymes can be procured. One approach to increase efficiency has been to bioprospect novel thermophilic microbial strains, the logic being that an increase in production temperature will result in increased rates of lignocellulosic hydrolysis. Unfortunately, microbiologists have had limited success in cultivating cellulolytic thermophiles with only a small number of strains isolated.

The goal of this project was to circumvent the need to cultivate cellulolytic thermophiles by applying a metatranscriptomic approach to discover new thermostable cellulases commonly referred to as glycoside hydrolases (GH - enzymes that hydrolyze the glycosidic linkages between sugar molecules). This was achieved by employing a novel *in situ* enrichment technique to enrich for thermostable GHs directly from a geothermal environment. These GHs were being actively expressed by the resident microbial population to hydrolyse lignocellulose. This approach not only eliminated the need for cultivation, but also selected for the actively expressed GHs unregulated in response to the lignocellulosic material feedstock. This strategy removed the emphasis on identifying potentially relevant and substrate-active GHs from genomic data via genomic analysis of a cultivated microorganism or from environmental metagenomic surveys.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

LIST OF ABBREVIATIONS

A. D. L.	Above Detection Limit	N. A.	Not Applicable
AFEX	Ammonia Fiber Explosion	NAD	Nicotinamide adenine dinucleotide
ATP	Adenosine triphosphate	nt	nucleotide
BBL	Barrel of Oil	RBS	Ribosomal Binding Sites
BLAST	Basic Local Alignment Search Tool	RNA	ribonucleic acid
bp	base pairs	SDS	sodium dodecyl sulfate
Btu	British thermal units	SSF	Simultaneous Saccharification and Fermentation
ºC	Degrees Celsius	rRNA	ribosomal RNA
CBM	carbohydrate-binding domains	µ	micro *unit* (10^{-6}) (e.g. microliters - µl, micrograms - µg, etc.)
DEPC	Diethylpyrocarbonate		
DNA	deoxyribonucleic acid		
EC	Enzyme Commission		
EROI	Energy Returned on Investment		
g	gram		
Ga	billion years ago		
gDNA	genomic DNA		
GH	glycoside hydrolases		
HMM	Hidden Markov Models		
l	litre		
LB	Luria-Bertani		
mRNA	messenger RNA		

Chapter 1

Introduction

1.1 Fossil Fuels

Fossil fuels dominate energy consumption in the United States. Liquid fuels and other petroleum (37.25 quadrillion British thermal units (Btu) per year), natural gas (24.71 quadrillion Btu per year), and coal (20.76 quadrillion Btu per year) were responsible for the three highest figures for energy consumption in 2010 [36]. Energy generated by renewable technologies paled in comparison, with a total of only 15.17 quadrillion Btu produced in 2010 [36].

Energy utilization in the United States can be summarized according to the sectors in which the energy was used: residential end-use consumed 22.05 quadrillion Btu, the industrial sector used 30.26 quadrillion Btu, the commercial sector accounted for 18.22 quadrillion Btu, and the transportation sector was responsible for the consumption of 27.63 quadrillion Btu in 2010. For the transportation industry, of the total 27.63 quadrillion Btu, 26.88 quadrillion Btu per year of consumption was derived from liquid fuel and other petroleum products [36]. In 2010, the United States produced 5.47 million barrels of crude oil per day, while it imported 9.17 million barrels per day to reach the demand for production. Alone, the transportation industry was responsible for using 13.82 million barrels per day of liquid fuel (~72 % of the total liquid fuel consumption in 2010) [36]. This dependence on foreign oil will only be exacerbated by the finite nature of fossil fuels and has coiled the term “Peak Oil” which is being used to describe the current state of world oil production.

1.2 Peak Oil

The term “Peak Oil” is used to describe the distinct point at which the production from an oil well is at its highest [89]. After this point, the field enters a phase of irreversible decline. This model was used to generate a prediction of all of the world’s oil fields as a whole (Figure 1.1) [60].

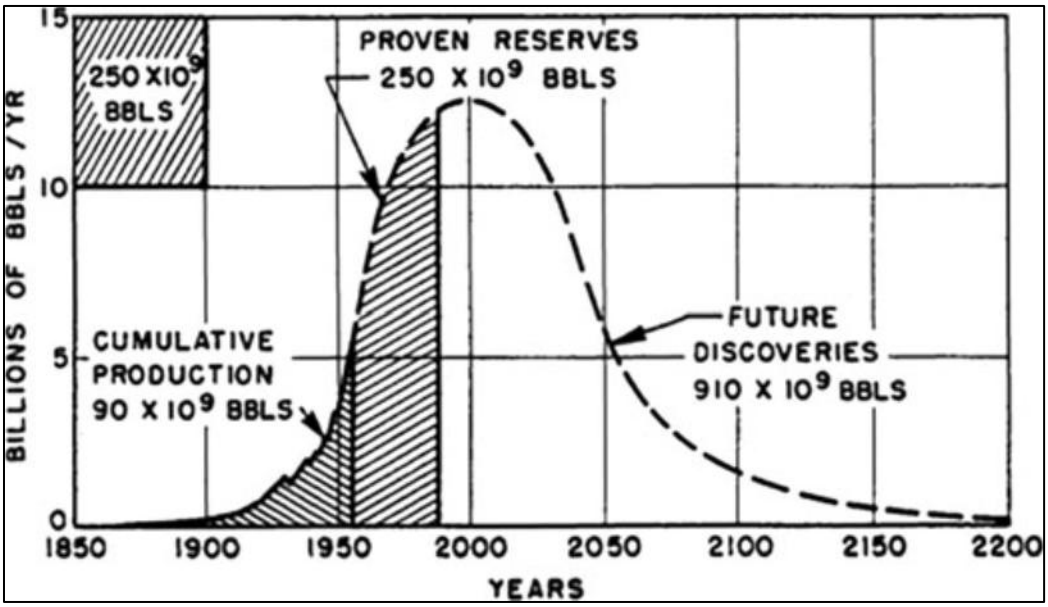


Figure 1.1. Hubbert Bell Curve Circa 1956. The bell curve predicts the peak production of all oil fields around the year 2000 [60]. After this point, the production of these fields will irreversibly decline for the duration of their exploitation. 1 barrel of oil (BBL) is defined as exactly 42 US liquid gallons.

Although an exact date for “Peak Oil” is not known, the prediction of “Peak Oil” has been only slightly refined since 1956. In 2011, a new figure was presented by Murphy and Hall [89] dividing global oil production into three eras (Figure 1.2). They conclude that the world is entering into an era where economic growth cannot be matched by an increase in oil production, thus leading to a climate of constrained oil supply.

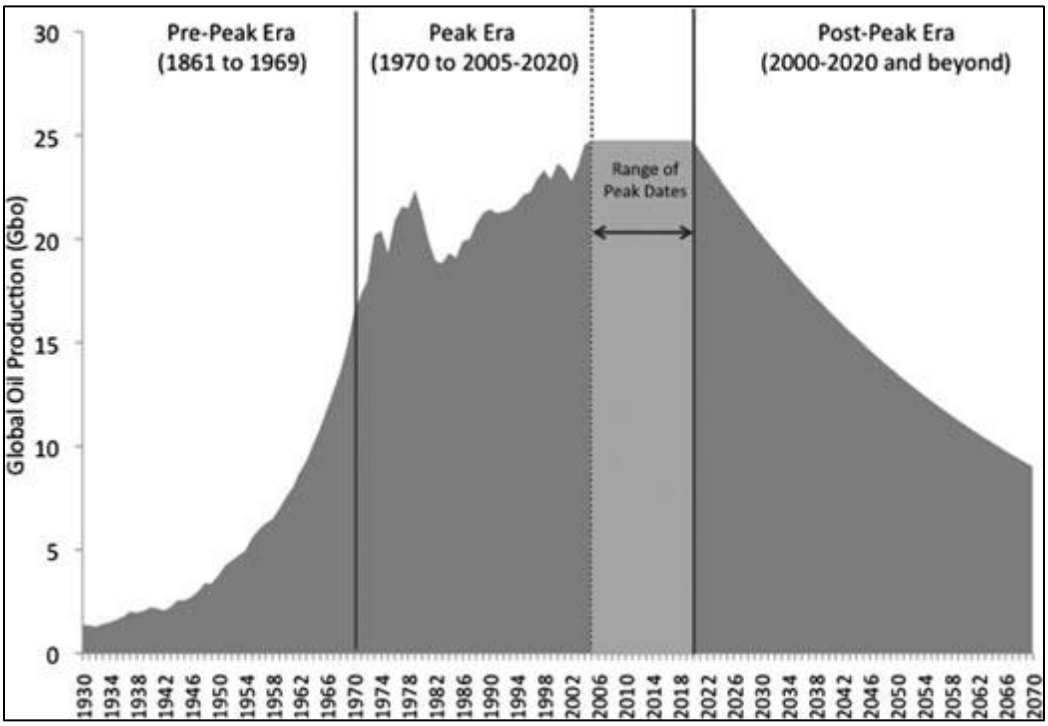


Figure 1.2. Murphy and Hall's Three Eras of Global Oil Production. The prediction of the end of the "Peak Era" coming between the years 2005-2020. After this time, decline in production occurs.

With the arrival and eminent departure from the "Peak Oil" era, the prospects for oil-based economic growth are bleak [89]. With a decrease in the energy returned on investment (EROI) of potential oil fields (e.g. Saudi Crude, deep sea, ultra-deep sea, oil sands, etc.) over time, the investment of the energy needed to harvest such reserves will become uneconomical [89]. Therefore, the investigation of alternative energy sources is of the utmost importance to account for economic growth and help countries gain energy independence. This energy independence would also alleviate the geo-political tensions generated by a constrained oil supply.

1.3 Alternative Energies

Alternative energies will serve to alleviate our dependence on foreign oil supplies. Although some alternative energy sources are promising, such as wind, hydro, solar, and nuclear power; these energy sources are geared toward electricity production for consumption. The transportation industry and the associated energy requirements in the United States cannot be met without imported oil. Total marketed (residential, commercial, industrial, transportation, and electric power) renewable energy shows projected growth of 2.6 % annually (Table 1.1). Contained within this field, bioethanol production is poised to dominate the switch to renewable energy in the transportation industry. The projected growth of ethanol used in E85 mixtures (ethanol/gasoline mixture) is projected to be 27.0 % annually from 2010 to 2035. Although various sources have been investigated for bioethanol production, cellulose is by far the most promising, boasting a projected annual growth of 56.6 % annually (Table 1.1) [36].

Table 1.1. Projected Annual Growth for Biomass Associated Energy by Source and Sector. Adapted from Energy Information Administration's Annual Energy Outlook 2012 [36].

Source and Sector	**Annual growth 2010-2035 (percent)**	**Sources of Ethanol**	**Annual growth 2010-2035 (percent)**
Ethanol used in E85 (Transportation)	27.0 %	Ethanol from cellulose	56.6 %
Biomass (Industrial)	1.2 %	Ethanol from corn and other starches	1.0 %
Biomass (Electric power)	4.4 %		
Total marketed renewable energy	2.6 %		

1.4 Bioethanol

Many starting materials have been successfully used to create bioethanol. The current major biomass sources for ethanol production are corn, coconut, sunflower, safflower, sugarcane, wheat, rice (including the corresponding stover), bagasse, grass, municipal waste, and coniferous and non-coniferous woody species [19,50]. Although all materials have the capability of being used as a starting biomass, the economics of conversion are unique to each. Biofuels produced from crops associated with food production (termed first generation biofuels) are usually composed of sugars and starch which can easily be extracted from the plant. However, the use of first generation crops such as corn, wheat, rice, and sugarcane compete for land already associated with food production [79]. Many argue the waste of potential in using this land for biofuel production, instead of much needed food crops.

Second generation biofuels use woody plants (coniferous/non-coniferous woody species, and some grasses) as a starting material. Although more complex and diverse in nature, the growth of these plants provides large amounts of lignocellulosic biomass and can also be harvested on land which is unfit for agriculture [79]. Due to the complex structure of the starting material, the sugar monomers are harder to extract, therefore providing a major area of research to try and improve efficiency and economics of biofuel generation from the feedstock. Once this is achieved, liquid fuel can be produced from crops and land dedicated to energy, without a threat to food production. The key lies in increasing the rate of conversion from lignocellulose to fermentable sugars, which improves efficiency and therefore leads to more economical production of biofuel.

1.5 Lignocellulose

Lignocellulose is a heterogeneous biomass which contains cellulose, hemicellulose, and lignin [139]. Cellulose is comprised of chains of glucose monomers linked by β-1,4 linkages (Figure 1.3). These chains form microfibrils by hydrogen bonding, resulting in a crystalline superstructure which is very hard to degrade (Figure 1.4). Lignin and hemicellulose are found in the spaces between microfibrils in primary and secondary cell walls, as well as middle lamellae [37].

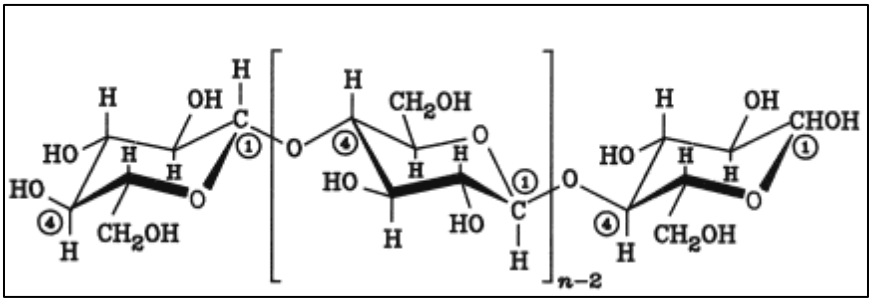


Figure 1.3. Glucose Monomers of a Cellulose Chain. Glucose molecules are arranged in β-1,4 linkages to create a cellulose chain. These chains are packed together through hydrogen bonding to form microfibrils (Figure 4).

Hemicellulose is more varied in structure and composition as opposed to cellulose. Types of hemicellulose include xylan, mannan, galactan and arabinan polymers [16]. Xylan, composed of the pentose (C5) sugar xylose, is the most abundant hemicellulose in nature. A representation of hemicelluloses containing different sugar monomers can be seen in Figure 1.5.

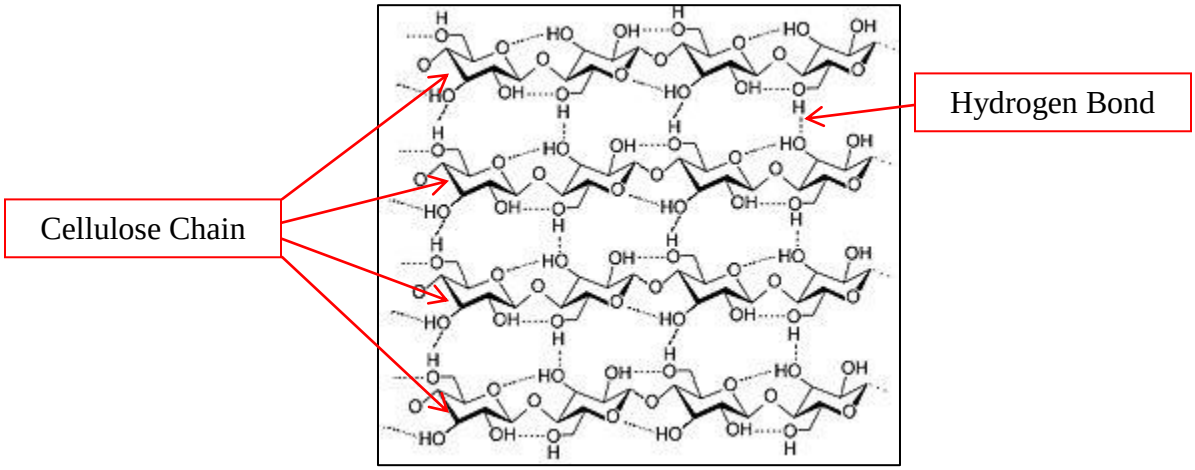


Figure 1.4. Crystalline Superstructure of Cellulose. Hydrogen bonding allows for packing of chains of cellulose creating tightly packed microfibrils. These microfibrils are surrounded by lignin and hemicellulose to create the superstructure of lignocellulose.

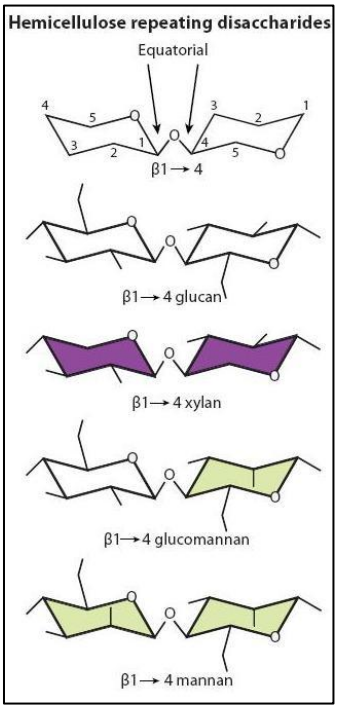


Figure 1.5. Hemicelluloses. Different sugar monomers can comprise hemicelluloses found in nature [116]. The varied structure does not allow tight packing by hydrogen bonding (Figure 1.4). Hemicellulose can also be branched, contributing to a non-uniform secondary structure.

Lignin is a heterogeneous molecule consisting of phenylpropand units such as p-coumaryl, coniferyl, gyaiacyl, syringyl, and sunapyl alcohol [74]. The composition of lignin is different between hardwoods and softwoods, although the specific structure of lignin is unknown [37]. This structure remains elusive due to its complex nature and lack of conservation between different species of plants. An example of a lignin molecule (and its component parts) can be seen in Figure 1.6 which illustrates the heterogeneity of this molecule and the complex nature of its composition. Older and woodier plants contain higher levels of lignin in their cell walls, making the cell walls hydrophobic and providing protection against pathogens [110]. This protection is also conferred in part by the reactive nature of lignin's component parts. Inhibitory compounds from the hydrolysate of lignocellulose have been shown to have an adverse effect on the ability of microbes to carry out fermentation [98].

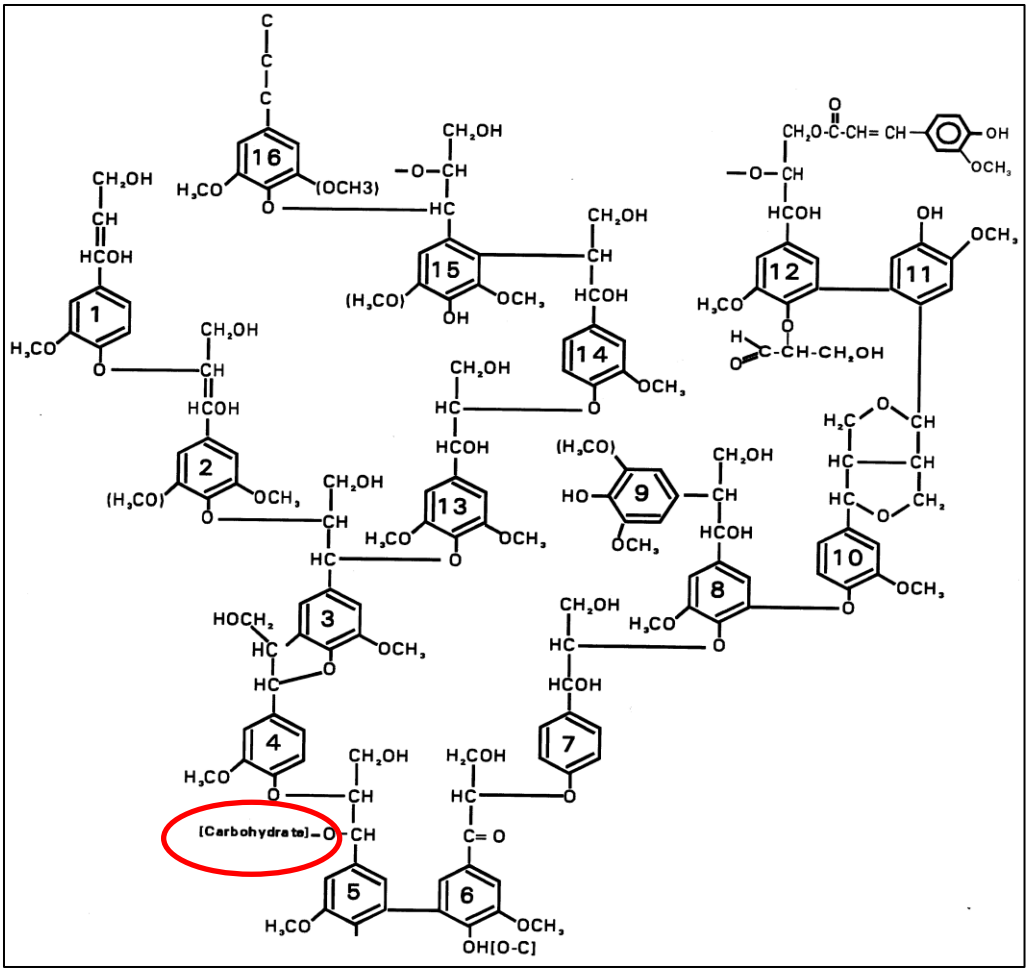


Figure 1.6. Lignin. Structure of the component parts of lignin, illustrating the complexity and heterogeneity of this superstructure [74]. The carbohydrate moiety (denoted in the red circle) illustrates a position at which the lignin and hemicellulose would form a covalent bond.

The composition of lignocellulosic starting material is unique in respect to its content of cellulose, hemicellulose, and lignin. Van Dyk *et al.* [139] compiled a list of current research undertaken, concerning the respective compositions of different lignocellulosic starting materials (Table 1.2) to illustrate these differences. Although each starting material is comprised of the three main components (cellulose, hemicellulose, and lignin), the heterogeneity of these starting materials makes the design of a universal

procedure for the bioconversion a challenge. Therefore, the bioconversion process must be completely understood for any progress to be made in this alternative energy.

Table 1.2. Review of Composition of Starting Materials. Approximate composition (as a percentage) of various biomass materials or agricultural waste products that can potentially be used for saccharification for biofuel production. This figure was adapted from [139] and references therein.

Biomass	**Cellulose (%)**	**Hemicellulose (%)**	**Lignin (%)**
Bermuda grass	25-47.8	13.3-35.7	6.4-19.4
Reed	39.5	29.8	24
Rapeseed stover	27.6	20.2	18.3
Willow	49.3	14.1	20
Reed canary straw	42.6	29.7	7.6
Corn cobs	35-45	35-42	4.5-15
Rice hulls	24-36.1	12-19.7	11-19.4
Melon shells	35	19	30
Corn stover (*Zea mays*)	39-41.7	19.1-20.5	15.1-18
Wheat straw	30-44	23-50	7.7-17
Rice straw (*Oryza sativa* sp.)	32.1-41	15-24	9.9-18
Bagasse	33.4-43	25-30	12.2-24
Newsprint (*Populus tristis*)	40-64.4	21.7-40	18-30
Douglas fir	50	17.8	28.3
Hay	26	26.8	4.7
Alfalfa	21.8	12.4	9.7
Verge grass	30.7	15.6	14.1

Table 1.2. (Continued) Review of Composition of Starting Materials.

Biomass	Cellulose (%)	Hemicellulose (%)	Lignin (%)
Coffee pulp	24	8.9	19.4
Sawdust	45	15.1	25.3
Coconut fibre	17.7	2.2	34
Switchgrass	31-45	22-31.4	12-18
Monterey pine (*Pinus radiata*)	41.7	20.5	25.9
Hybrid poplar	40	22	24
Nut shells	25–30	25–30	30–40
Leaves	15–20	80–85	0
Grasses (average)	25–40	25–50	10–30
Cotton seed hairs	80–95	5–20	0
S32 rye grass (early leaf)	21.3	15.8	2.7
S32 rye grass (seed setting)	26.7	25.7	7.3

1.6 Bioconversion of Lignocellulose

The conversion of a lignocellulosic biomass for ethanol production is an energy intensive process. Separated into three main energy intensive steps (pretreatment, hydrolysis and fermentation), Figure 1.7 illustrates the process of converting a lignocellulosic starting material into ethanol. Generally, the three steps in the bioconversion process are [74]:

1. The lignocellulosic starting material is pretreated to remove cross-linking lignin (delignification). This improves access for subsequent hydrolysis by increasing the surface area of the substrate.
2. The remaining cellulosic substrate is depolymerized. This can be accomplished either by chemical, physical, or enzymatic treatment steps [85]. The goal of this step is to allow the component parts of the cellulose and hemicellulose (polymerized six and five carbon sugars respectively) to be hydrolysed to monosaccharides for subsequent fermentation.
3. The fermentation process involves the conversion of the free monomeric sugars to energy, with ethanol and carbon dioxide as by-products. This process is routinely conducted using *Saccharomyces cerevisiae* (yeast), but several thermophilic anaerobic bacteria such as *Clostridium thermocellum* and *C. thermosaccharolyticum* also ferment sugars, forming ethanol [34].

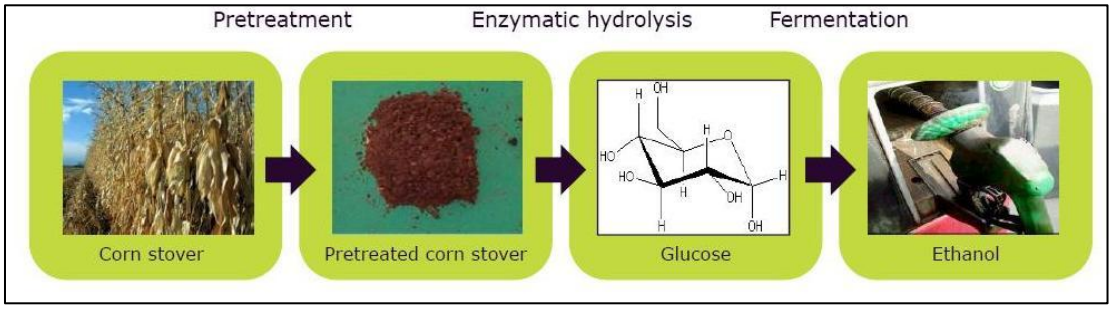


Figure 1.7. Bioconversion of Lignocellulose for Bioethanol Production. Adapted from www.novozyme.com. Using corn stover as an example, the major steps of the bioconversion process are highlighted.

The pretreatment steps involved in the bioconversion primarily support the subsequent hydrolysis [4]. Harmsen *et al.* [56] details the chemical and physical pretreatments. Presently, chemical and thermochemical pretreatment steps are among the

most effective. Among these industry standards are Ammonia Fiber Explosion – AFEX (conducted at 60 °C – 100 °C) and Steam Explosion/SO_2-Steam Explosion [4]. Pretreatment steps break down the large complex superstructure present in the starting material, while also removing the toxic lignin component. This allows for an increase in surface area of the substrates and therefore increases the effectiveness of the enzymatic bioconversion process. These pretreatment steps aside, the greatest potential for the enhancement of ethanol production from biomass lies in enzymatic hydrolysis of cellulose using cellulases [69].

Currently, methods are being employed which make use of chemical, physical, or enzymatic means to hydrolyse the heterogeneous starting material into its component parts. The enzymatic hydrolysis of lignocellulose is of special interest because it can be combined with physical conditions and chemical additives, which together increase the efficiency of the entire pretreatment process. The application of these enzymes, primarily glycoside hydrolases (GHs), for the hydrolysis of lignocellulose as a starting material account for 40 – 49 % of the total ethanol production cost [82].

1.7 Utilization of Cellulose by Microorganisms

Life has evolved the means to utilize the energy stored in cellulose. As discussed above, cellulose and hemicellulose are comprised of six and five carbon sugar monomers respectively. Although the superstructure of cellulose or hemicellulose cannot be used by organisms for energy production, the monomers of these structures are capable of being inserted into the metabolic pathways used for energy generation. There is also a challenge of transporting such a large and complex structure across the membrane into the cell.

Therefore, organisms attempting to use lignocellulose as an energy source will initially require the molecular machinery needed to convert the superstructure into its component parts. Even larger organisms, such as ruminants (who consume lignocellulosic material in their diet), rely on resident microorganisms living within them to deconstruct the cellulose and hemicellulose into easily utilized sugar monomers.

These metabolic pathways were being utilized to convert lignocellulose into its component parts, even before there was a complete understanding of their mechanism. Initially, whole organisms capable of this degradation were studied.

1.8 Lignocellulosic Bioconversion Using Traditional Microbiology

The application of microbial strains to aid the bioconversion of lignocellulose has been investigated through the many routes discussed in this section. Initially, whole organisms that were able to degrade cellulose were investigated heavily due to naturally occurring metabolic pathways discussed in the previous section. These organisms were selected by restricting the available energy sources in growth media to complex forms of simple sugars (cellulose, hemicellulose, and lignocellulose). With the advent of molecular tools, new approaches have been used to separate each step of the pathways involved for further optimization. These advances have allowed the investigation of organisms not capable of growth on cellulose alone, and whole microbial communities which contain constituents who are unable to be grown in the laboratory. However, as already mentioned, the original study of the breakdown of cellulose (and lignocellulose), using microbiology, began with whole organisms capable of growth on cellulose.

One of the first microorganisms to be used in lignocellulosic bioconversion was the fungus *Tricoderma reesi*. Originally isolated from a decaying canvas in World War II [111], *T. reesi* and its cellulose-active enzymes have been well studied. Today, it is the primary source of commercial cellulases [115].

Mesophilic cellulose-degrading microorganisms are widely studied as they are easily cultivated in the laboratory. For detailed reviews of the use of mesophilic microorganisms capable of growth on cellulose, please refer to references 82 and 145. Although mesophiles have been well studied, there is a heightened interest in thermophilic microorganisms (those growing optimally at 60 °C and above) due to the benefit of insensitivity to biologically challenging reaction conditions brought on by the various pretreatment steps involved in bioethanol production (Section 1.6)[20]. Specifically, elevated temperatures allow for an increase specific activity of enzymes, reduction of microbial contamination, increase in mass transfer rate due to lower fluid viscosity, and greater flexibility in the bioprocess [5]. Therefore, thermophilic cellulose-degrading microorganisms make an ideal target for increased lignocellulosic hydrolysis.

The order *Clostridiales* includes examples of some of the most well studied cellulose-degrading thermophilic microorganisms. The best studied organism of this order is *Clostridium thermocellum.* With an optimal growth temperature of 60 °C, it was originally isolated by Freier in 1988 [41]. This organism is of special importance and will be discussed more in-depth in later sections. Another member of this order, *Caldicellulosiruptor saccharolyticus*, has an optimal growth temperature of 70 °C and is capable of utilizing xylan, xylose, pectin, and arabinose as well as hemicellulose as a sole carbon source [108]. A related member of this order, *Caldicellulosiruptor kristjanssonii*,

is not as well studied but has been described to have an optimal temperature of 78 °C and is closely related to *Caldicellulosiruptor saccharolyticus* [21]. *Anaerocellum thermophilum* utilizes a broad spectrum of substrates such as glucose, galactose, and arabinose [125].

Thermotoga maritima, a member of the order *Thermotogales*, was found to have an optimal growth temperature of 80 °C and can hydolyze xylan, but not crystalline cellulose [30]. Other examples from this order have also been well studied because of their capability of growth on cellulosic substrates. These include *Thermotoga elfii* [112], *Thermotoga naphthophila* [129], and *Thermotoga petrophila* [129]; all of which were isolated in oil fields.

The order *Dictyoglomales* constitutes other examples of thermophilic and cellulolytic microorganisms central to the study of lignocellulose. Originally isolated in 1988 [124], it showed growth on xylan and cellulose. The production of xylanases in *Dictyoglomus thermophilum* Rt46B.1 [147] has been well studied and multiple members of this order have been confirmed to be cellulolytic as well as anaerobic [124,148].

Along with members from other orders, these microorganisms were initially studied for their capability to grow on cellulosic substrates. Table 1.3 contains a non-exhaustive list of thermophilic microorganisms described in the literature which are capable of growth on cellulose [20].

Table 1.3. Non-Exhaustive List of Thermophilic Microorganisms Capable of Growth on Cellulosic or Hemicellulosic Substrates. Table adapted from Blumer-Schuette [20]. Additions to this list were made with microorganisms relevant to the study of geothermal hot springs. The non-exhaustive nature of this table illustrates the amount of studies generated in this field.

Organism	Location	GC Content (%)	Topt (°C)	Source	Substrates[a]	Reference
Thermophilic *Archaea*						
Pyrococcus abyssi	North Fiji basin	44	96	Sea water	α, β	31
Pyrococcus furiosus	Vulcano Island, Italy	41	100	Marine sediment	α, β, T	40
Pyrococcus horikoshii	Okinawa Trough	42	98	Sea water	N. D.	51
Thermococcus kodakaraensis	Kodakara Island, Japan	52	85	Marine sediments	α, T	44
Sulfolobus solfataricus	Pisciarelli Solfatara, Italy	36	85	Solfataric hot spring	α, β, H, P	118

Table 1.3. (Continued) Non-Exhaustive List of Thermophilic Microorganisms Capable of Growth on Cellulosic or Hemicellulosic Substrates.

Organism	Location	GC Content (%)	Topt (°C)	Source	Substrates[a]	Reference
Thermophilic *Bacteria*						
Thermotoga maritima	Vulcano Island, Italy	46	80	Marine sediments	α, β, H, P	91
Thermotoga neapolitana	Lucrino, Italy	41	80	Marine sediments	α, β, H, P	140
Thermotoga lettingae	Netherlands	39	65	Sulfate-reducing bioreactor	α, β, H, P	11
Thermotoga naphthophila	Niigata, Japan	46	80	Kubiki oil reservoir	A	129
Thermotoga sp. RQ2	Ribeira Quente, the Azores	46	76-82	Marine sediments	N.D.	61
Thermotoga petrophila	Niigata, Japan	47	80	Kubiki oil reservoir	A	129

Table 1.3. (Continued) Non-Exhaustive List of Thermophilic Microorganisms Capable of Growth on Cellulosic or Hemicellulosic Substrates.

Organism	Location	GC Content (%)	Topt (°C)	Source	Substrates[a]	Reference
Thermotoga elfii	Africa	39	66	African oil field	A	112
*Caldicellulosiruptor saccharolyticus**	Taupo, New Zealand	35	70	Wood from hot spring	α, β, C, H, P	108
*Anaerocellum thermophilum***	Valley of Geysers, Russia	37	75	Plant residues from hot spring	α, β, C, H	125
Clostridium thermocellum	Louisiana, USA	39	60	Cotton bale	C	41
Dictyoglomus turgidum	Kamchatka, Russia	33.9	80	Hot spring	N. D.	124
Chthonomonas calidirosea Strain T49(T)	Tikitere, New Zealand	54.6	68	Hot spring	N. D.	75
Acidothermus cellulolyticus 11B	Yellowstone National Park	68.41	55	Acidic hot springs	N. D.	12

Table 1.3. (Continued) Non-Exhaustive List of Thermophilic Microorganisms Capable of Growth on Cellulosic or Hemicellulosic Substrates.

Organism	Location	GC Content (%)	Topt (°C)	Source	Substrates[a]	Reference
Spirochaeta thermophila DSM 6192	Raoul Island, Kermadec Archipelago, New Zealand	62	65	Brackish hot spring	C, T, β, α	2
Thermobacillus xylanilyticus	Northern France	57.5	55	Soil beneath a manure heap	N. D.	136
Fervidobacterium riparium	Kuril Islands, Russia	31	65	Hot spring	N. D.	104
Thermosipho affectus	Mid-Atlantic Ridge	27	70	Hydro-thermal vent	N. D.	105
Caldicellulosiruptor obsidiansis	Yellowstone National Park	35	78	Obsidian Pool (Hot Spring)	β, α, H, C, P	54
Geobacillus thermoleovorans T4	Kaohsiung, Taiwan	53.3	55–60	Sugar refinery wastewater	β, α, C	127

[a] Biomass substrates are abbreviated as follows: α: α-linked glucans; β: β-linked glucans; C: cellulose; T: chitin; H: hemicellulose; P: pectin.

* *Caldicellulosiruptor saccharolyticus* was originally referred to as *Caldocellum saccharolyticum.*

** *Anaerocellum thermophilum* is now classified as *Caldicellulosiruptor bescii*.

Along with representatives described above from the domain *Bacteria*, organisms from the domain *Archaea* are capable of growth in extreme temperatures (Table 1.3). The study of hyperthermophilic *Archaea* capable of degrading cellulose has been limited, due to an apparent inability to culture such cellulolytic strains in the laboratory. *Desulfurococcus fermentans* was the first isolated hyperthermophilic archaeon capable of growth on crystalline cellulose and capable of growing optimally at 81°C (with growth up to 89 °C) [101]. Recently, research has shifted from the study of individual organisms to groups of organisms who, as a whole, can degrade complex carbohydrates [94,138,1]. A microbial consortium enriched at 90 °C was found to degrade cellulose [52]. This enrichment was found to be a mixture of *Archaea* who together were able to degrade cellulose. However they could not be isolated as individuals. The research concluded that due to the inability of most hyperthermophilic microorganisms to be cultured, broader approaches must be applied as well as traditional culture based laboratory techniques for the discovery of hyperthermophilic cellulose-active enzymes [52].

1.9 Molecular Investigation of Cellulose-Active Enzymes

The molecular investigation of cellulose-active enzymes began with organisms capable of growth on cellulosic substrates as detailed in the previous section. These microorganisms which were known to be cellulolytic, were investigated at the genetic level to deduce the enzymes involved in the bioconversion of lignocellulose. Individual enzymes from these systems could be expressed in non-native organisms in an effort to make the bioconversion of lignocellulose more efficient.

It is postulated that as little as 1 % of microbial species are readily available for study by cultivation [52]. Organisms which contain cellulose-active enzymes are not always capable of growth on in the laboratory or on cellulose. With the development of high-throughput sequencing technologies, bioprospecting from the "uncultivatable" majority has become commonplace. Studies of whole communities of microorganisms capable of degrading lignocellulose are now not complete without sampling the uncultivable microbial component. Therefore, cellulose-active enzymes can be discovered in organisms or communities where none of the constituents are capable of isolation in pure culture. These enzymes then can be expressed in non-native organisms in what is called heterologous gene expression or purified in what is known as cell-free enzyme mixtures.

Today, the most common sources of heterologous expression of cellulose-active enzymes are fungi [35]. Many enzyme manufacturing companies (e.g. Novozyme) produce cellulose active enzymes from the fungus *T. reesi* in various mixtures for analytical and commercial purposes. *T. reesi*'s cellulose active enzymes have also been directly expressed in microorganisms capable of fermentation (*Saccharomyces cerevisiae*) in combination with another cellulose active enzyme gene from *Aspergillus aculeatus* [139] in an effort to increase the efficacy of the bioconversion of cellulose.

Although *T. reesi* was the first fungal species to be exploited to this end, those organisms which are thermophilic offer advantages (described in the previous section) by producing thermostable enzymes. Those fungal organisms which have been studied include *Sporotrichum thermophile* [17], *Thermoascus aurantiacus* [49], and *Thielavia terrestris* [48]. The non-culture based molecular techniques are powerful for all domains

of life, but have special importance when those representatives from the domain *Archaea* are studied due to the inability of most *Archaea* to be cultured on cellulosic substrates (Section 1.8).

Most hyperthermophilic *Archaea* are incapable of growth on crystalline cellulose. An investigation of their genomes surprisingly yields enzymes which are active on either cellulose or hemicellulose. Originally *Pyrococcus furiosus*, with an optimal temperature of 100 °C, was found to contain only one endo-1,4-glucanase (EglA) and was incapable of growth on cellulose or xylan [13]. Also with an optimal growth temperature of 100 °C, *Pyrococcus horikoshii* contains a β-glucosidase [6] but this enzyme has limited activity towards crystalline cellulose [70]. The extremely thermoacidophilic enzymes from *Sulfolobus* would work well in acid conditions, but no species of *Sulfolobus* have been reported to grow on crystalline cellulose [20]. However, an endo-xylanase [24] and bi-functional β-xylosidase/α-L-arabinosidase [88] have been found in the genome of *Sulfolobus solfataricus*.

The domain *Bacteria* offer interesting examples of cellulose-active enzyme discovery through molecular techniques. Individual cellulose-active enzymes from *Caldicellulosiruptor saccharolyticus* include CelB (GH44) [47] and CelA (GH03) [133]. *Anaerocellum thermophilum* was found to contain a cellulose-active enzyme CelA (GH03) [125]. As mentioned in the previous section, *Thermotoga maritima* [32] was targeted for study because of its growth on xylan. Once it was investigated at the genetic level, it was shown to change its expression of enzymes depending on exposure to different carbohydrates [30]. It is an interesting example of a cellulolytic thermophile because it does not grow on cellulose but its genome contains several simple β-1,4-

glucanases (Cel5A, Cel5B, Cel12A, Cel12B) that lack carbohydrate-binding domains (CBMs). As stated before, it grows well on xylan, this was found to be due to two encoded xylanases (XynA and XynB) with CBMs [30] in its genome.

Although much research has gone into the discovery of these cellulolytic enzymes, how they are utilized is still open for some debate in the bioethanol community. Both cell-free enzyme mixtures [137] and organisms called "consolidated bioprocessing biocatalysts" [46] have been promoted as options for the bioconversion of cellulose and lignocellulose. Although cell-free enzyme mixtures are usually capable of activity in conditions unsuitable for cellular growth, consolidated biocatalysts are enticing because they naturally include the molecular machinery for fermentation (ethanol production) and are discussed below.

Clostridium thermocellum, mentioned briefly in the previous section, is considered a candidate consolidated bioprocessing biocatalyst [76]. It is considered this because it is a thermophilic cellulolytic anaerobic microorganism, and possesses the molecular machinery to ferment sugars into ethanol. In addition, it degrades cellulose faster than contemporary cell-free enzyme mixtures [76]. Not only does this microorganism stand alone as a biocatalyst, but the enzymes it produces have been used in other fermenting microorganisms such as the yeast *S. cerevisiae*. In one instance, using a cellulosomal endoglucanase (EgE) from *Clostridium thermocellum* (in conjunction with BGL1 from *Saccharomycopsis fibuligera*) resulted in elevated ethanol production from *S. cerevisiae* [65]. In addition to this example, there have been studies of individual cellulose-active enzymes from the genome of *Clostridium thermocellum.* Vazana *et al.* [141] studied the cellulosomal and non-cellulosomal states of two glycoside hydrolases

(GH48 and GH9) from the genome of *C. thermocellum*, and found that free enzymes present in their natural state were more active on cellulosic substrates than those without substrate binding modules. They also noted the difference between the natural ratio of GHs produced by consolidated biocatalysts and those introduced as cell-free mixtures, a potential problem with the creation of cell-free enzyme mixtures reaching a bottleneck at certain steps in the bioconversion process.

Although consolidated biocatalysts are enticing because of the prospect of creating a continuous bioreactor, cell-free enzyme mixtures are more versatile when the heterogeneity of different lignocellulosic starting materials is considered. Therefore, having a large group of cellulose active enzymes from which to choose and tailor the mixture's activity, based on the starting material, is currently the industry standard for companies such as Novozyme, Genecor, Dyadic International, Amano Enzymes Inc., and others. These cellulose active enzymes, termed glycoside hydrolases (GHs), contain a wide variety of actions on a wide variety of substrates.

1.10 Glycoside Hydrolases

Glycoside Hydrolases (GHs) are a class of enzymes responsible for the cleavage of the glycosidic bonds present in polysaccharides. Currently 130 families of glycoside hydrolases have been described and are curated on a web-site: www.cazy.org [25]. The enzymes are grouped by their domains and their respective activity on carbohydrates. Cellulose is converted to simple glucose monomers by three large groups of these enzymes, all of which are required to be present for the complete hydrolysis of cellulose.

Complete cellulose degradation involves three types of enzymes, endo-β-glucanases (e.g. EC 3.2.1.4), exo-β-glucanases (e.g. EC 3.2.1.91 and EC 3.2.1.176), and β-glucosidases (e.g. EC 3.2.1.21) [143] (EC corresponds to Enzyme Commission classification [144]). These enzymes act synergistically to produce monomeric glucose as an end product from cellulose. A representation of this bioconversion can be seen in Figure 1.8, adapted from an article by Lynd *et al*. [82]. The bioconversion of lignocellulose on the other hand is more complicated because of the complexity of the starting material, compared to cellulose.

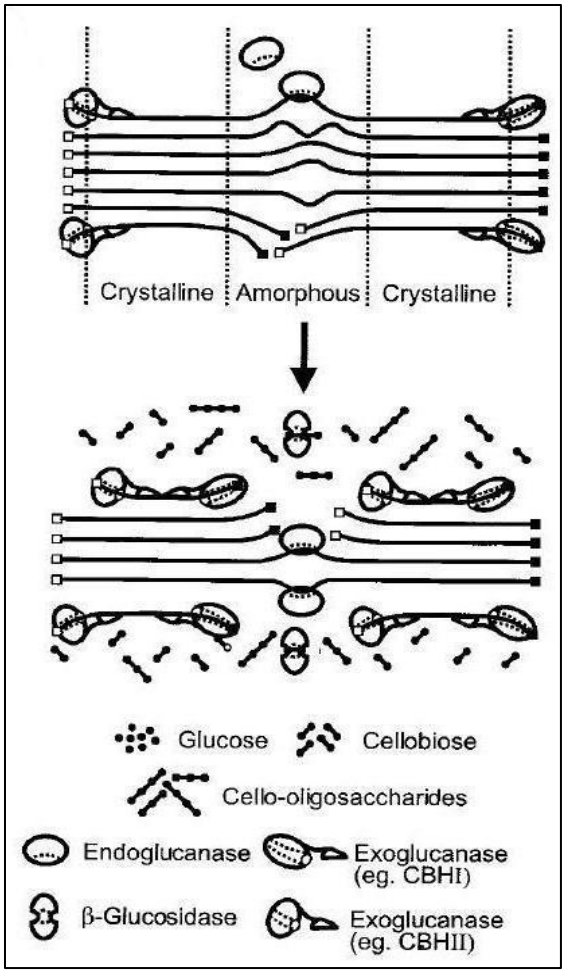


Figure 1.8. Cellulose Degradation by a Mixture of Enzymes. Adapted from Lynd *et al*. [82]. The action of the three groups of enzymes are synergistic. The product of endoglucanase activity is the substrate for exoglucanases, and the product of exoglucanase activity is the substrate for β-glucosidases. Therefore a mixture of all three groups is needed for complete degradation of cellulose.

Lignocellulose as a starting material is more complex than crystalline cellulose and therefore requires a broader spectrum enzyme cocktail for complete hydrolysis [83]. Most commercial mixtures rely on crude combinations of GHs with different activities to achieve the highest conversion [139]. Microorganisms capable of growth on lignocellulose usually have a large amount of diverse glycoside hydrolases in their genomes. *Clostridium phytofermentans* contains a repertoire of 108 putative glycoside hydrolases involved in the degradation of cellulose and hemicellulose [135]. Although

GHs are grouped within the same family, their substrate specificities and end-products may be different [8]. This makes the identification of families important as well as the identification of domains contained within the GH enzyme (which dictate specificity and action). It has recently been shown that certain families of GHs are absolutely required for the bioconversion of cellulose. GH family 9 and 48 have been identified to be necessary to the point that their removal from the system abolishes the bioconversion efficiency to the great degree [135,95]. While companies attempt the engineering of already known enzymes for increased activity; there is also interest in the discovery of new GHs from many different environments. This effort is constantly focused on trying to optimise the bioethanol production process by enhancing the enzymatic hydrolysis of the pretreated biomass.

1.11 Glycoside Hydrolases in Industry

The projected demand for industrial enzymes worldwide is increasing at an annual rate of 7.6 % and is expected to be $6 billion by 2012 [27]. Cellulases constitute approximately 20 % of the world enzyme markets alone [18]. These are predominantly used in textile detergents, paper, food/feed development and biorefinery industries as well as some other industries [28]. With economic development and increased public interest in alternative fuels, these numbers will only increase. Companies in major countries (various European countries, USA, Japan, China, and India) are leading the development of enzymes for use in the bioconversion of lignocellulose. Lignocellulosic active enzyme cocktails from the major providers (Table 1.4) originate from several main sources.

Table 1.4. Commercially Available GH Mixtures, Providers and Function. Adapted from [28] and [119]. Most of these enzyme mixtures have an optimal temperature of 50 – 60 °C.

Company Name	Microorganism	Brand Names of Enzyme Mixtures	Recommended Temperature	Application of the Formulation
Novozymes	*T. reesei*	Cellic CTec2	45 - 50 °C,	Lignocellulosic substrate hydrolysis
	T. longibrachiatum/ A. niger	Celluclast	40 - 50 °C	
	T. reesei	Cellic HTec2	45 - 50 °C	
	T. longibrachiatum	Viscozyme L	60 °C	
	A. niger	Novozymes 188	50 °C	
	T. longibrachiatum/ T. reesei	Bio-feed beta L	50 °C	
	T. longibrachiatum/ T. reesei	Energex L	50 °C	
	T. longibrachiatum /T. reesei	Ultraflo L	50 °C	

Table 1.4. (Continued) Commercially Available GH Mixtures, Providers and Function.

Company Name	Microorganism	Brand Names of Enzyme Mixtures	Recommended Temperature	Application of the Formulation
Genecor	*T. longibrachiatum/ T. reesei*	Spezyme CP	50 - 70 °C	Commercially available biomass enzymes developed specifically for second generation biofuels
	T. reesei	Accelerase®1500	50 °C	
	T. reesei	Multifect CL	50 °C	
	T. longibrachiatum/ T. reesei	GC 440	50 °C	
	T. longibrachiatum/ T. reesei	GC 880	50 °C	
	T. longibrachiatum/ T. reesei	GC 220	50 °C	
Dyadic International	*T. longibrachiatum*	AlternaFuel® 100P	50 °C	Conversion of lignocellulosic biomass to glucose for fermentation into ethanol

Table 1.4. (Continued) Commercially Available GH Mixtures, Providers and Function.

Company Name	Microorganism	Brand Names of Enzyme Mixtures	Recommended Temperature	Application of the Formulation
Amano Enzymes Inc.	*A. niger*	Cellulase DS	50 - 60 °C	Saccharification of lignocellulosics into fermentable sugars
	A. niger	Cellulase AP30K	60 °C	
	T. viride	Cellulase TAP106	50	
AB Enzymes GmbH	*T .longibrachiatum/ T. reesei*	ROHAMENT® CL	50 °C	A cellulase for biomass saccharification
Maps (India) Limited	*Bacillus sp.*	Palkolase HT	50 - 70 °C	Starch 31 liquefaction and saccharification
	Bacillus sp.	Palkolase LT	50 °C	
	Bacillus sp.	Palkodex	65 °C	
Speciality Enzymes & Biotechnologies Co.	*Bacillus sp.*	SEBfuel	50 °C	Liquefaction and saccharification

Table 1.4. (Continued) Commercially Available GH Mixtures, Providers and Function.

Company Name	Microorganism	Brand Names of Enzyme Mixtures	Recommended Temperature	Application of the Formulation
Iogen	*T. reesei*	Ultra-Low Microbial (ULM)	N. A.	Biomass saccharification
Biocatalysts Limited	*Trichoderma sp.*	Cellulase 13P	N. A.	Degradation of cellulose completely
	Trichoderma sp.	Cellulase 13L	N. A.	
Rhodia-Danisco (Vinay, France)	*T. longibrachiatum/ T. reesei*	Cellulase 2000L	50 °C	N. A.
50L Lyven (Colombelles, France)	*T. longibrachiatum/ T. reesei*	Cellulyve	50 °C	N. A.
Solvay Enzymes (Elkhart, IN)	*T. longibrachiatum/ T. reesei*	Cellulase TRL	50 °C	N. A.
Alko-EDC (New York, NY)	*T. longibrachiatum/ T. reesei*	Econase CE	50 °C	N. A.
Quest Intl. (Sarasota, FL)	*T. longibrachiatum/ T. reesei*	Biocellulase TRI	50 °C	N. A.

N.A.; Not applicable

Most commercially available cell-free enzyme mixtures available on the market (Table 1.4) have an optimal temperature of 50 - 60 °C. As discussed previously, thermostable enzymes are desirable. The lack of thermostable enzymes on the market today encourages research into environments which produce these enzymes. Although some mesophilic GHs have been engineered in an effort to increase thermostability [62], the discovery of new thermostable enzymes present in thermal environments is an enticing prospect. The term bioprospecting is used to describe this effort.

1.12 Bioprospecting for Novel Glycoside Hydrolases to Improve Bioconversion Efficiency

Bioprospecting has a broad set of definitions, but in general can be defined as the exploration of biodiversity for new resources with social and commercial value. However, at its core, bioprospecting ventures aim to discover novel chemicals, genes, metabolic pathways, structures, materials and behaviours [15]. Bioprospecting can target single strains of organisms, organism consortia, genes, enzymes or secondary metabolites.

Glycoside hydrolases (GHs) active at high temperatures are termed thermostable GHs. These offer many advantages in the bioconversion of lignocellulose. However, there is a paucity of thermostable bacterial isolates and glycoside hydrolases. This lack has prompted an increased focus on bioprospecting for thermophilic cellulolytic enzymes using a variety of strategies. One of the most prevalent strategies for selection of microorganisms capable of producing GHs is enriching for cellulolytic microorganisms using lignocellulosic substrates. When used to bioprospect for lignocellulose hydrolysis, lignocellulosic substrates are supplied to the environment in order to encourage the

growth of microorganisms that are capable of utilizing them as an energy source. This not only increases the amount of microorganisms capable of this bioconversion, but also causes the increased production of enzymes which are active on the substrate. There have been many examples of this when studying the bioconversion of lignocellulose [38,52,55,57,76,102,109]. Although this concept is not new, given the advent of new sequencing technologies, there have been more enrichment studies geared towards the entire genetic material present in the enrichment, rather than the individual organisms.

1.13 Bioprospecting Using "Meta-omics"

The terms metagenomics and metatranscriptomics refer to the study of whole community genetic material. This information is encoded in either deoxyribonucleic acid (DNA) or ribonucleic acid (RNA). The study of these sequences of information has been made possible by the advent of new sequencing technologies. As opposed to a genome, comprised of the information contained in only one organism, a metagenome is defined as the genetic material (DNA) recovered directly from a community of microorganisms. This is especially powerful when studying environments where cultivation of the constituents is a problem. A metatranscriptome is defined as the transcripts (messenger RNA or mRNA) recovered directly from an environmental sample, and is referred to as a metatranscriptomic study. The advantage of the analysis of the total transcriptome of an environmental sample is that that the expression of mRNA reflects the community's response to the current ecosystem conditions, and therefore identifies actively expressed enzymes as opposed to only the expression potential in the metagenome.

When bioprospecting from the genetic material of a community, genes are targeted which are predicted to have a specific functionality. Bioprospecting for GHs using a metagenomic study has been previously performed on soil [66], in the rumens of a Yak [33] and calf [39], the gut of a higher termite [92], and also the garden of a leaf-cutter ant [9]. All of these places are environments of natural lignocellulosic bioconversion resident microorganisms were found to produce GHs. These ventures were successful in identifying genes within the samples that potentially were active on a cellulosic substrate. In all cases, cloning and characterization was done on at least one gene identified in the samples to confirm action.

Although the use of a metagenomic study bypasses the need for pure culture of the community constituents, it still lacks some specificity as to the enzymes which are responsible for the degradation. The central dogma of biology states that genetic information is encoded in DNA, transcribed into mRNA, and then translated into active proteins. A study of total DNA from a community gives information on the capability of the community for function, but lacks specificity as to which proteins are actively being expressed. Metatranscriptomics not only circumvents the problems associated with cultivation, but also selects for actively expressed mRNA transcribed in response to the conditions to which the host organisms are exposed. This circumvents the need to identify the GHs active against the particular cellulosic substrate, which is not possible to do in metagenomic surveys.

1.14 Metatranscriptomics

A metatranscriptomic study is one of the total mRNA of a microbial community. A metatranscriptomic study presents many challenges involved with reliability, repeatability, redundancy and cost performance. Bacterial mRNA has a very short half-life and degrades rapidly. One group measured the average half-life of *E. coli* mRNA transcripts to be 6.8 minutes [117] and another measured an average half-life of 7.5 minutes [87]. Another group found a half-life range from 0.5 minutes to 50 minutes with a typical average half-life of about 3 minutes [130] and is congruent with the estimate of an average half-life of 2 minutes at 37°C [99]. The half-life of total mRNA in this instance was 6.8 min under the conditions tested. Therefore, special precautions must be taken to ensure transcript integrity. Even with these challenges, there has been much interest in this type of strategy.

The study of transcripts has been widely used and is well established; however, the use of metatranscriptomics to view the functionality of a community was relatively unexploited before the advent of next generation sequencing technologies. With these technologies there has been an explosion of interest in many unique environments in which the cultivation of community members has been a problem. Recently, metatranscriptomic techniques have been employed to bioprospect for GHs. The environments which have been studied are the gut of a termite [132], the rumen of Muskoxen [106], and a soil microbial community with an emphasis on plant matter degradation [10]. All of these environments have active degradation of lignocellulose; however they are all mesophilic in temperature and not necessarily conducive to the elevated temperatures required for greater conversion efficiency.

At the time of writing this review, there have been no metatranscriptomic studies of extreme environments which are known to produce enzymes capable of the degradation of lignocellulose (such as hot springs or soils). In addition to selecting an environment which will produce thermostable enzymes, pairing a metatranscriptomic study with an enrichment with lignocellulosic substrates provides many advantages. These advantages include an increased amount of microorganisms capable of lignocellulosic bioconversion and the increased production of enzymes involved in the bioconversion. Not only can this change be compared to the natural environment with a metatranscriptomic study, but the interplay between different microorganisms' enzymes can be seen by the reconstruction of metabolic pathways present in the metatranscriptome. This is in addition to the generation of sequences of transcripts which are not only actively being expressed, but code for enzymes which are actively involved in the bioconversion of lignocellulose. This will lead to the generation of a group of enzymes which could be expressed as a cell-free mixture or as heterologously expressed genes in a consolidated biocatalyst. Therefore the comparison of an enrichment with the natural environment using a metatranscriptomic technique will reveal GHs which are involved in the breakdown or hydrolysis of a lignocellulosic substrate.

1.15 Project Objectives

Due to our nation's dependence on foreign countries for fossil fuels, and the finite nature of these resources, alternative energies are being heavily investigated. Out of the many possibilities, bioethanol production is poised to alleviate the liquid fossil fuel dependence generated by the transportation industry. As well as being sustainable, bioethanol from lignocellulosic biomass does not compete with land fit for agriculture, a clear advantage over first generation biofuel biomass. A promising means of the bioconversion of lignocellulose is accomplished using whole organisms or individual enzymes involved in the bioconversion which are already present in nature. There are also advantages to the bioconversion process if the reaction occurs at an elevated temperature, making thermophilic organisms and thermostable enzymes of great interest.

Currently, there is a paucity of enzymes and organisms which meet these needs. Therefore, bioprospecting for novel thermostable GHs has been a focus for research to improve the bioconversion process. The selection of an environment which will produce thermostable GHs paired with the enrichment of microbes involved in the lignocellulosic breakdown is an enticing situation from which to bioprospect. Thus this project aims to bioprospect for novel thermophilic GHs from geothermal microbial communities actively degrading lignocellulose. Specifically, this project will be accomplished by these specific aims:

- **Specific Aim 1** – To develop methodology to enrich and isolate the mRNA from geothermal samples.
- **Specific Aim 2** – To bioprospect for mRNA encoding for thermostable GHs either by targeting i) geothermal ecosystems with known cellulolytic activity or ii)

enriching for cellulolytic and thermophilic bacteria in geothermal systems with unknown cellulolytic activity.

- **Specific Aim 3** – To generate a metatranscriptomic library of the bioprospecting targets and to annotate and identify candidate GHs.
- **Specific Aim 4** – To clone a target GH and confirm predicted function.

Chapter 2

Universal Materials and Methods

2.1 RNase-Free Environment

Solutions and materials were purchased or made RNase-free for all experiments. Glassware was made RNase-free by baking at 300 °C for four hours. Solutions were made RNase-free by treatment with 0.01 % [v/v] Diethylpyrocarbonate - DEPC (Sigma Aldrich) for 30 minutes at room temperature. All solutions were then sterilised at 121 °C at one atmosphere pressure for 30 minutes to inactivate the DEPC. For those solutions which could not be autoclaved (those including skim milk powder or phenol), heat was applied for 30 minutes on a hot-plate until the solution was brought to a boil for a minimum of 10 minutes to ensure inactivation of the DEPC. Plasticware was purchased from their respective manufacturers as RNase-free and not reused. RNase-free equipment and samples were handled in an aseptic manner to avoid contamination. Where possible, samples were processed in a UV-sterilised PCR hood.

2.2 Positive Controls for Nucleic Acid Extraction Procedure

Positive controls for the nucleic acid extraction were conducted using a pure cell culture. This ensured a large amount of nucleic acid which was easily visible using gel electrophoresis and could be reviewed for integrity. Log phase cell cultures were used for all extractions. This ensured the highest yield of RNA while decreasing the amount of degraded RNA in the sample as a result of cells that had died and lysed. Cell cultures were prepared as follows:

Luria-Bertani (LB) medium (BD Difco) was prepared by adding 25 g of powder per l of distilled H_2O, according to the manufacturer's instructions. This mixture was then sterilised at 121 °C at one atmosphere pressure for 30 minutes. 500 ml of sterile LB medium was inoculated with 1 ml of a previously frozen culture of *Escherichia coli*. The culture was then allowed to grow at 37 °C at 180 rpm in an orbital shaker (Ratek Instruments Pty Ltd). Cell growth was followed by measuring optical density. Samples were taken every hour and read at 600 nm on a Lambda 35 UV/Vis Spectrometer (Perkin Elmer) and the results were used to generate a standard growth curve (Appendix A). This was used to identify log phase growth, during which cells were harvested. Dilutions were made where appropriate. Once a time point was determined for optimal growth (and therefore containing the highest amount of undegraded nucleic acid), 500 ml cultures of *E. coli* were grown and harvested via centrifugation in an Avanti J-E Centrifuge (Beckman Coulter) into 10 ml volumes. The resulting pellets were then resuspended in 2 ml LB medium and placed in 2 ml Cryoware Cryogenic Vials (Nalgene) and stored at -80 °C until needed for the positive control experiments.

Extraction of total nucleic acid was performed on pure *E. coli* cells to verify the validity of the extraction procedure. One 2 ml Cryoware Cryogenic Vial (Nalgene) filled with log phase *E. coli* made previously was used and the contents were spun in a Centrifuge 5415R (Eppendorf) at maximum speed for 10 minutes at 4 °C. The resulting supernatant was discarded and the pellet was resuspended in 1 ml RNase-free distilled water and added to Lysing Matrix B Tubes (Qiagen). These were then subject to the general extraction procedure (Section 2.3).

2.3 General Nucleic Acid Extraction Procedure

Soil samples were collected from hot spring soil saturated with water. Soil was collected aseptically and placed into 50 ml RNase-free Falcon Tubes (Ambion). The full Falcon Tubes were immediately frozen by immersion in liquid nitrogen to ensure stasis of the sample for the duration of the trip back to the laboratory to avoid mRNA degradation (Section 1.14).

At the laboratory, ~0.5 g samples were aseptically transferred into Lysis Matix B tubes (Qiagen) while still frozen. 700 µl TPM Buffer (50 mM Tris-HCL[pH 5.0], 1.7 % [w/v] polyvinylpyrrolidone, 20 mM $MgCl_2$) was then added to the tube. This was immediately placed into the FastPrep FP120 cell disrupter (Thermo Electron Corporation) at 6.0 ms^{-1} for 40 seconds. Soil particles and cell debris were pelleted at 20,000 x g for 30 seconds at 4 ºC in a Centrifuge 5415R (Eppendorf). The supernatant was transferred to a RNase-free 2.1 ml microcentrifuge tube (Ambion). To the same cell disruptor tube, 700 µl of PBL buffer (5 mM Tris-HCl [pH 5.0], 5 mM Na_2EDTA, 0.1 % [w/v] sodium dodecyl sulphate (SDS), and 6 % [v/v] water-saturated phenol) was added. The cell disruption and centrifugation steps were repeated as described above and the two supernatants were pooled. Since the total volume was approximately 2 ml, the pooled supernatants were gently mixed and then split into two 2.1 ml RNase-free microcentrifuge tubes. The supernatants were first extracted with water-saturated phenol (pH 4.5), second with phenol-chloroform-isoamyl alcohol (25:24:1 [v/v/v], pH 4.5), and third with chloroform-isoamyl alcohol (24:1 [v/v], pH 5.5), each time using an equal volume of the phenol mixture as the aqueous phase transferred from the previous extraction step. The resulting aqueous phase was then mixed with 0.7 volumes of ice-cold

isopropanol and 0.1 volume of 3 M sodium acetate (pH 5.7). The mixture was inverted five times to mix and then subjected to incubation at -20 ºC for at least 30 minutes. After incubation, the tubes were spun at maximum speed (~20,000 x g) for 30 minutes. For samples where yield was an issue, incubation at -20ºC was conducted overnight and followed by centrifugation for 60 minutes at maximum speed (~20,000 x g) at 4 ºC. The pellet was washed with 70 % [v/v] ice cold ethanol, air dried, and resuspended in TE buffer (10 mM Tris-HCl, 1 mM EDTA [pH 8.0]). A schematic representation of the general extraction procedure is presented in Figure 2.1.

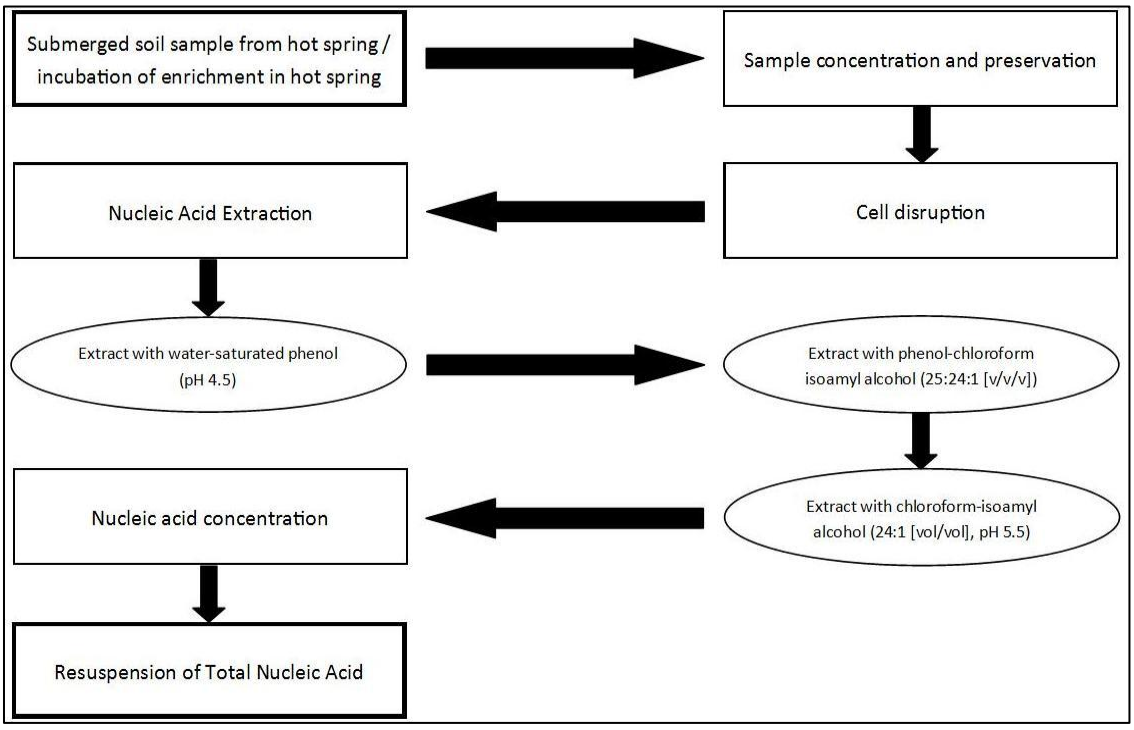


Figure 2.1. Schematic of Total Nucleic Acid Extraction.

2.4 mRNA Enrichment from Total Nucleic Acid

A total nucleic acid extraction will contain both community DNA and RNA. There are also contaminants such as humic substance and soluble iron which will co-extract from certain soil types and can inhibit downstream application of the DNA or RNA [120]. Figure 2.2 shows a schematic representation of the steps employed while enriching for mRNA transcripts and reverse transcribing them into cDNA for sequencing.

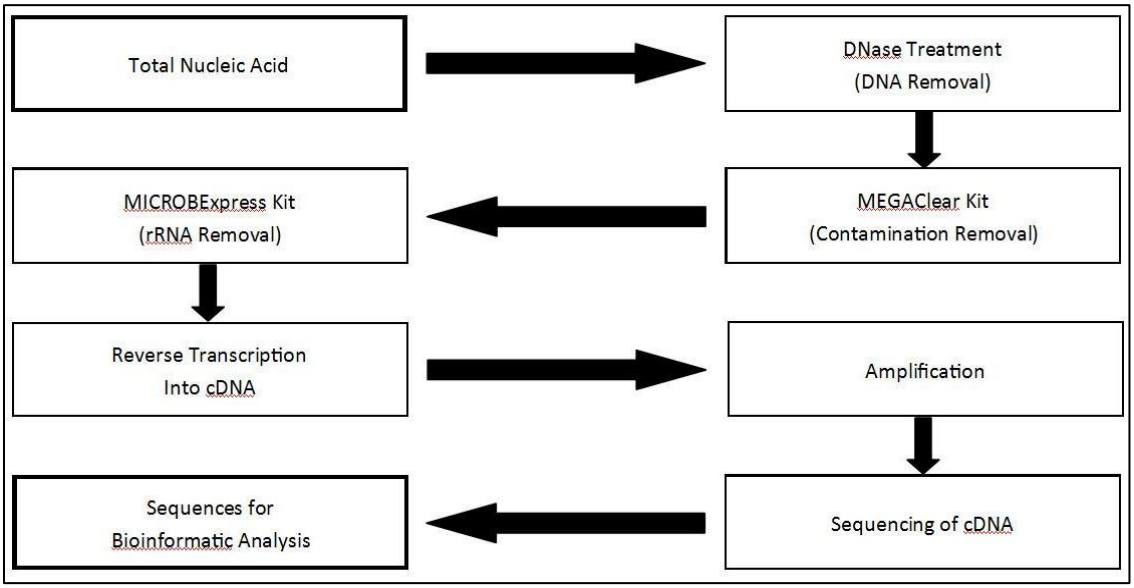


Figure 2.2. Schematic Representation of Enrichment of mRNA, Reverse Transcription, and Sequencing.

DNA was removed from the total nucleic acid extraction (Section 2.3) using TURBO DNase (Ambion) by incubation at 37 ºC for 30 minutes as per the manufacturer's instructions. The DNase in the sample was inactivated by extraction with phenol-chloroform-isoamyl alcohol (25:24:1 [v/v/v], pH 4.5) followed by chloroform-isoamyl alcohol (24:1 [v/v], pH 5.5). The resulting nucleic acids were then precipitated as described above using isopropanol and sodium acetate. After washing and air-drying the

resulting pellet, it was resuspended in 200 µl of DEPC-Treated Water (Ambion). The DNA-free sample was then treated to remove small sequences and contamination with the MEGAClear Kit (Ambion) as per the manufacturer's instructions. This step not only removed small degraded sequences (<200 base pairs), but also served to remove various Polymerase Chain Reaction (PCR) inhibitors which co-extracted with the total nucleic acid. The eluent was then reprecipitated as described above and then resuspended in no more than 15 µl of DEPC-Treated Water (Ambion). This was then treated with the MICROBExpress Kit (Ambion) to remove the rRNA from the total RNA by subtractive hybridization. The final product generated by the MICROBExpress Kit was fully enriched mRNA.

2.5 cDNA Library Generation and Sequencing

The resulting enriched mRNA was reverse transcribed using Superscript III (Promega) using Random Hexamer Primers (Invitrogen). The reaction was conducted in a Palm-Cycler Gradient Thermal Cycler (Corbett Research) using the manufacturer's instructions. The resulting reverse transcription reaction product (1 µl) was amplified with GenomePHI V2 (GE Healthcare) according to the manufacturer's instructions. The sample was sequenced using the 454 Corporation's GS-FLX instrument at The Center for Advanced Studies of Genomics, Proteomics, and Bioinformatics (ASGPB) at the University of Hawai‘i at Mānoa. Briefly, the resulting cDNA library was prepared for sequencing by the GS FLX DNA Library Preparation Method Manual, GS FLX emPCR Method Manual, and sequencing was done according to the GS FLX Sequencing Method Manual on the 454 GS FLX System (454 Life Sciences, Bradford, CT).

2.6 Gel Electrophoresis and Qubit Quantification

All extraction results detailed in this thesis were subject to visualization on the GeneGenius Gel Imaging System (Syngene) using either a 0.8 % or 1.0 % (w/v) agarose gel using SeaKem LE Agarose (Lonza) and 0.1 – 0.5 µl of RedSafe Nucleic Acid Staining Solution (iNtRON Biotechnology). Electrophoresis was conducted on an Owl EasyCast B1A Mini Gel Electrophoresis System (Thermo Scientific) with a PowerPac 300 (BioRad) for 30 - 45 minutes at 80 - 100 volts. All gel visualizations were run with 0.1 µl – 0.5 µl of a 1kB DNA ladder (Zymo Research) for scale. Figure 2.3 illustrates the typical constituents of a total nucleic acid extraction (Section 2.3) as well as the various lengths of the bands present on the 1 kB ladder included in all other gel visualizations.

In addition to gel electrophoresis, some samples were quantified using the Qubit Fluorometer (Invitrogen) as per the manufacturer's instructions. This system has both settings and calibration systems for measurement of DNA and RNA concentration of a sample.

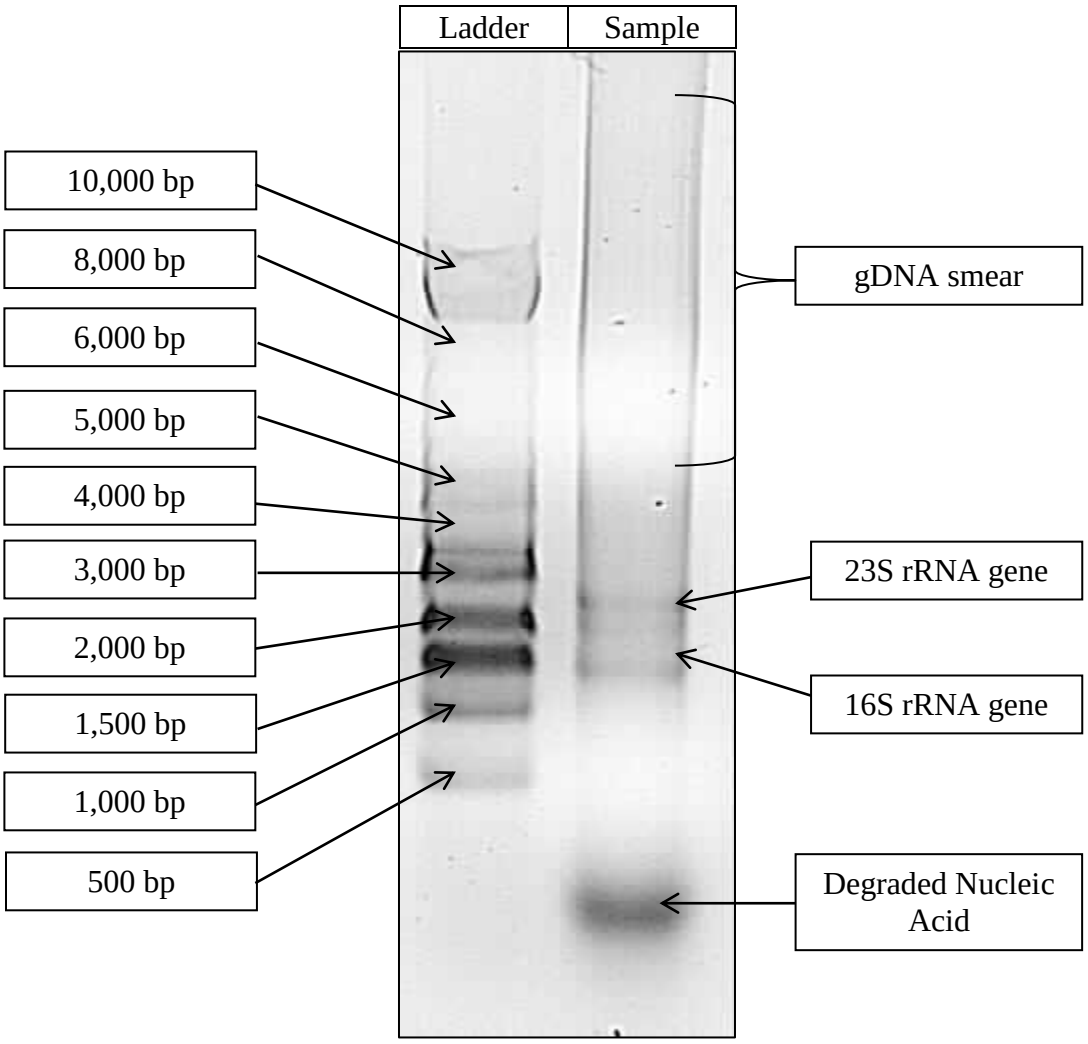


Figure 2.3. Representative Total Nucleic Acid Extraction Visualization. All aspects of the total nucleic acid component of an extraction can be seen including the genomic DNA smear (gDNA smear), 23S ribosomal RNA gene band (23S rRNA gene band), 16S rRNA gene band, and messenger RNA smear (mRNA smear). The values on the left of the figure represent the different length of the 1 kB DNA ladder in base pairs (bp).

2.7 Bioinformatic Analysis

Bioinformatic analysis of sequence information has expanded greatly in previous years. With great advances in computing power, large amounts of data can be mined for pertinent information in a relatively short amount of time. All sequences in this experiment were subject to a tailored pipeline of assembly and annotation originally developed for work with whole bacterial genomes. All programs contained within the pipeline were run on a local server at ASGPB at the University of Hawai‘i at Mānoa.

The output files generated by the 454 Pyrosequencing system were input into Newbler v2.6 [86] under variable parameters and assembled into sequences of various length. These nucleotide sequences, in FASTA format, were of inputted into Genemark.hmm for Prokaryotes v2.8 [80]. Genemark is a program which calls all potential proteins in all six reading frames of the input nucleotide sequence. In addition to using start codons as transcriptional starting sites, the program uses known Ribosomal Binding Sites (RBS) and Hidden Markov Models (HMM) to predict potential proteins which do not have a start codon. These predicted proteins were then interrogated using BLASTp v2.2.26 [3] against a standalone database downloaded on the 28th of July, 2011 from the BLAST ftp website (ftp://ftp.ncbi.nlm.nih.gov/blast/db/). This program assigned function to each transcript. Those which had even a slight hit to any of the various names of the cellulose-active enzymes were then separated for further study.

The amino acid sequences corresponding to those proteins with a predicted activity on cellulose were then interrogated using InterProScan v4.8 [107]. This program was used to verify the BLAST prediction and to also to assign domains, active sites, signal peptide, and transmembrane regions to the transcript. Those which were verified

by InterProScan were then subjected to BLASTx [3] and associated Conserved Domain Database [84] which was used to identify the most current prediction of the gene (the initial search was conducted using a standalone database downloaded July 2011), confirm functional domains, to assign closest taxonomy to the protein, and to trim the sequence to only the protein of interest. Substrate specificity of the candidate sequences identified using the CAZy glycoside hydrolase database [25]. A schematic of the generation of predicted function by BLASTp can be seen in Figure 2.4.

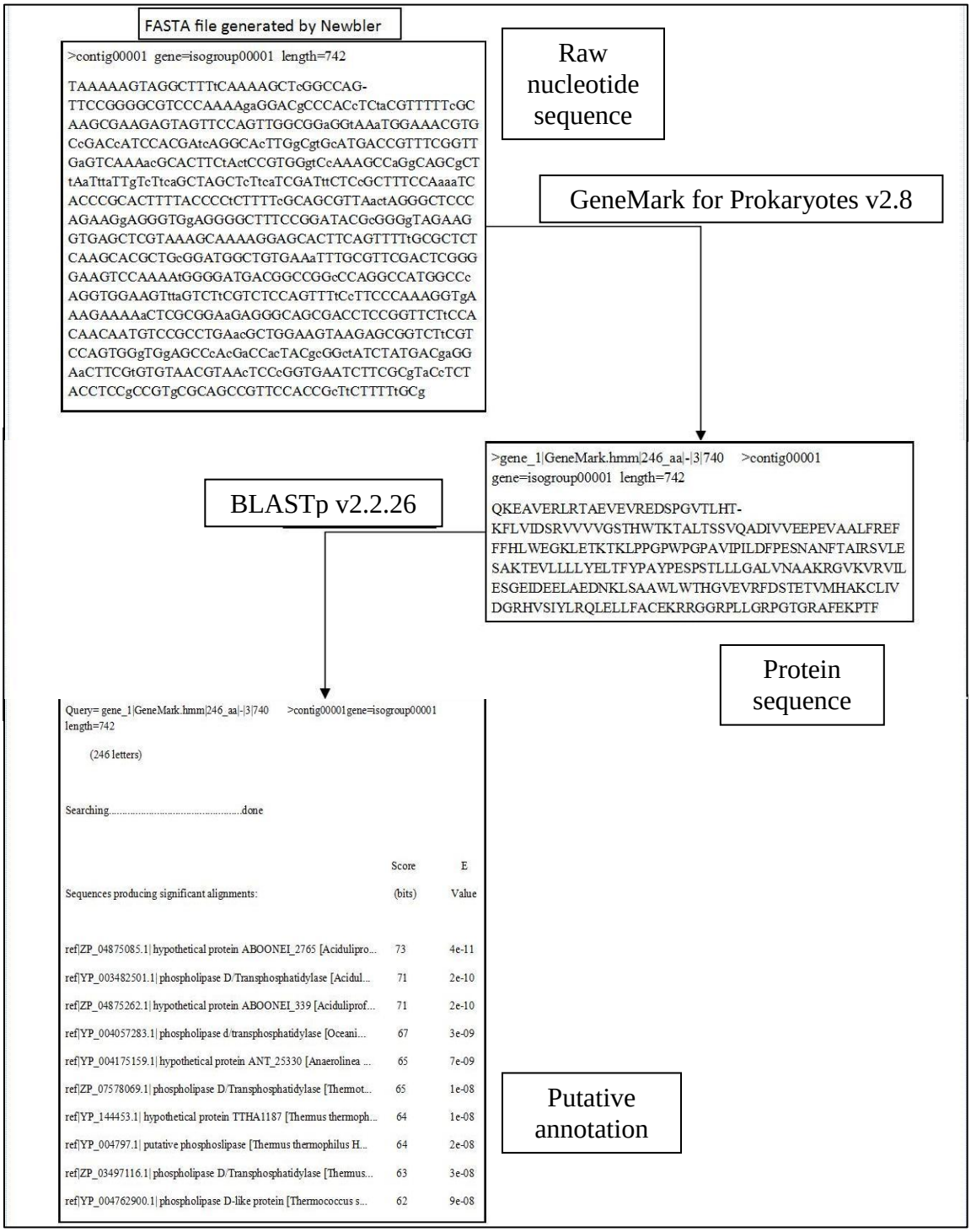


Figure 2.4. Schematic Representation of Informational Outputs Generated by Bioinformatic Programs.

Chapter 3

Results

3.1 Overview of Results Section

The research detailed in this results section has been split into two subsections; "Waikite" and "Ngatamariki". The purpose of this setup being that a significant proportion of this Master's thesis was devoted to troubleshooting the mRNA extraction protocol from a geothermal spring. The initial sample site chosen for this study was a spring at Waikite Valley in the Taupo Volcanic Zone and after considerable unresolvable issues with the RNA extraction process (detailed in Sections 3.2.2.1 to 3.2.2.4), this site was abandoned and a second site at Ngatamariki, also in the Taupo Volcanic Zone, was chosen to continue the research. As the applied methodologies and materials for each of the sample sites were significantly different, two site-specific Materials and Methods sections and Results sections are included in this chapter.

3.2 Waikite

The Waikite research site is located in the Taupo Volcanic Zone of the North Island of New Zealand (Figure 3.1). The Tē Manaroa Spring, the largest of the springs at the site, was described as an environment with a temperature of 95–99 °C and a pH range of 7.8–9.3 [121].

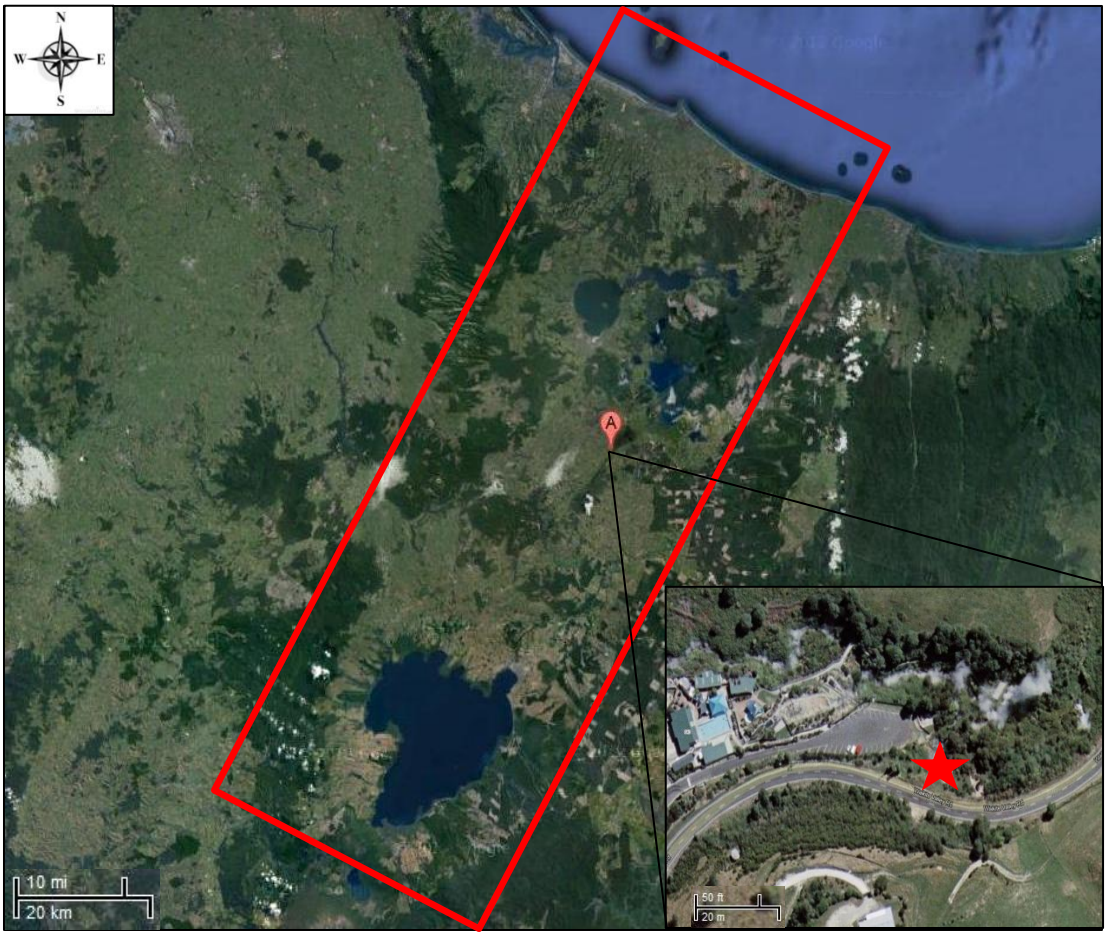


Figure 3.1. Location of Waikite in the Taupo Volcanic Zone of the North Island of New Zealand. The area donated in red is known as the Taupo Volcanic Zone. Located in the middle of this area, Waikite is denoted with the "A" marker. Inset is a zoomed image of the Waikite site. The hot spring itself is signified by the red star on the inset image.

Tē Manaroa Spring originates from a small orifice on an embankment and forms a small pool bounded by a concrete bund (Figure 3.2). The water from the spring is used as a hot water source for tourist bathing pools. Geochemical data generated from the spring water can be seen in Table 3.1.

Table 3.1. Physiochemical Measurements of the Waikite Spring Water. Dissolved amounts of elements as well as chemical species specific for Tē Manaroa Spring at Waikite. Other than the slightly basic pH and elevated temperature, this site holds little interesting physiochemical properties.

Species or Physical Measurement	**Units**	**Concentration**
Temperature	°C	96
Bicarbonate (total)	mg/L	315.61
pH		8.18
Ammonia (as total NH_3)	mg/L	<0.01
Aluminium (Al)	mg/L	<0.15
Arsenic (As)	mg/L	0.34
Boron (B)	mg/L	1.38
Bromide (Br)	mg/L	0.39
Calcium (Ca)	mg/L	6.84
Chloride (Cl)	mg/L	140.45
Conductivity	µS/cm	1013
Copper (Cu)	mg/L	<0.003
Fluoride (F)	mg/L	2.65
Iron (Fe)	mg/L	<0.08
Lithium (Li)	mg/L	2.22
Magnesium (Mg)	mg/L	0.173
Manganese (Mn)	mg/L	0.418
Nitrate (NO_3)	mg/L	<0.03
Phosphate (PO_4)	mg/L	<0.05
Potassium (K)	mg/L	8.39
Silica (as SiO_2)	mg/L	161.9
Sodium (Na)	mg/L	211.8
Strontium (Sr)	mg/L	0.0487
Sulfate (SO_4)	mg/L	33.14
Sulfide (total as H_2S)	mg/L	0.003
Carbon Dioxide (CO_2)	mmole/100mole H_2O	31.315
Hydrogen sulphide (H_2S)	mmole/100mole H_2O	0.576
Argon (Ar)	mmole/100mole H_2O	0.122
Helium (He)	mmole/100mole H_2O	0
Hydrogen (H_2)	mmole/100mole H_2O	0.152
Methane (CH_4)	mmole/100mole H_2O	0.045
Nitrogen (N_2)	mmole/100mole H_2O	3.783
Oxygen (O_2)	mmole/100mole H_2O	0.906

Figure 3.2. Location of Sampling Site at Tē Manaroa Spring, Waikite, New Zealand. The water's edge, denoted by the white silica deposits, was the site of sample collection. The samples contained decaying organic matter (green) and clay-like water-saturated soil (red).

The embankment wall is permanently moist due to the steam given off from the pool, facilitating the growth of moss and other organic matter growing at the water's edge. The large amount of decaying organic and lignocellulosic matter is available to the resident microorganisms in the embankment soil for use as an energy source (Figure 3.2). An X-Ray Diffraction (XRD) and X-Ray Fluorescence (XRF) analysis of the embankment materials were conducted by CRL Energy Ltd. (Wellington, New Zealand), showing that the mineral component of embankment was primarily silica and clay materials (SiO_2 63.9 % - Cristobalite, Tridymite, quartz; Al_2O_3 19.5 % - kaolinite, illite)

with elevated iron species (Fe_2O_3 3.33 % - possible iron ferrihydrate or goethite). The temperature of the soil on the embankment (where the sample was taken) was 60-65 °C.

A previous environmental diversity survey of the sample site showed that bacteria from the phyla *Chloroflexi,* candidate division OP10, and *Acidobacteria* dominated the microbial community [121]. The authors of this study were able to isolate representatives of the *Chloroflexi* (P373 & P352) and candidate division OP10 (P488). These isolates all demonstrated growth on the polysaccharides xanthan gum as well as xylan and sodium alginate. Isolates P488 and P352 were also able to grow on gellan gum, another complex polysaccharide. These isolates illustrate constituents of the resident microbial community which are capable of the bioconversion of complex carbohydrates. The genomes of phylogenetically identical isolates of P488 and P352 (T49 and T81 respectively) from another NZ geothermal system (Tikitere, Hell's Gate) have been sequenced (M. Stott, personal communications). Thus, the "Waikite" site was considered an ideal location to conduct an environmental transcriptomic study to identify and bioprospect for actively expressed GHs due to the known cellulolytic activity of dominant community members.

3.2.1 Materials and Methods Specific for Waikite

3.2.1.1 Sample Collection at Waikite

The soil used for this experiment was collected aseptically from a position just above the water line as discussed in the previous section (Section 3.2). The soil was removed from the embankment just above the waterline (denoted by the white silica deposits in Figure 3.2). The soil removed from the bank (Figure 3.3) was aseptically transferred to a 50 ml RNase-Free Falcon tube (Section 2.3) and frozen in liquid nitrogen for transportation to the laboratory.

Figure 3.3. Aseptic Removal of Waikite Soil for Collection. Soil was removed from just above the waterline (Figure 3.2). Samples were then collected into 50 ml RNase-Free Falcon Tubes from portions of the soil sample not in contact with the removal device.

After collection, and subsequent transportation to the laboratory, total nucleic acid extraction was attempted on the soil. After exhaustive method development with positive controls, it was determined that although there were microbes in the environment (DNA amplified from this site by Stott *et al.* in 2008 [121]), the general extraction procedure developed for this thesis was not sufficient for the conditions found at Waikite. Therefore, current literature was consulted for references to nucleic acid extractions from recalcitrant soils. In the literature, many references to the ability of clay-like soil to absorb nucleic acid were found [26,68,77,93,97,103,113,142,146]. These studies found that structures present in the soil not only provide protection from environmental conditions (DNases), but also are capable of binding extracellular DNA which resident cells can use for nutrients of incorporation into their genetic material [114]. Subsequently, these colloids inhibit complete enumeration of microorganisms present in soil using nucleic acid based techniques [42,43]. Therefore, blocking these colloids with substances chemically similar to nucleic acid (or non-native nucleic acid) was an option to facilitate the study the nucleic acids of these resident microorganisms.

In the literature, all attempts to counteract this absorption were targeted at blocking the absorption of nucleic acid on the soil colloids [42]. This was attempted by pre-applying skim milk or RNA (when native DNA was the target of PCR amplification) [128] and gDNA [42] to the samples prior to extraction. Therefore, blocking agents were prepared and incorporated into the extraction procedure in an attempt to allow for the extraction of nucleic acids.

3.2.1.2 Preparation of Blocking Agents for Waikite

Skim Milk Powder (Difco) was mixed to the appropriate concentration in RNase-free glassware according to the manufacturer's instructions (100g per l of H_20). The milk solution was made RNase-free by treating the solution with 0.1% [v/v] DEPC. The DEPC was inactivated by heating the milk to boiling for 10 - 15 minutes. The milk was then cooled to room temperature and subjected to a nucleic acid extraction (Section 2.3). Confirmation of the lack of nucleic acids present in the skim milk powder was also performed through a control amplification of just the milk powder with no soil (data not shown).

To the original 0.5 g of soil used for the nucleic acid extraction, 700 µl of skim milk was added to the bead beating tube at the same time as the soil. This tube was then centrifuged in a Centrifuge 5415R (Eppendorf) at maximum speed (20,000 x *g*) for 5 minutes. The supernatant was removed and discarded. After the addition of 700 µl of TPM Buffer, the general nucleic acid extraction (Section 2.3) was conducted.

As an alternative to the skim milk supplement, the addition of non-native gDNA to the sample was used in an attempt to block the soil colloids from absorbing native nucleic acids. To extract and purify gDNA for this procedure, total nucleic acid was extracted from the frozen log phase *E. coli* (Appendix A) using the procedure in Section 2.3. The total nucleic acid pool was then subject to treatment with RNase A (Promega) at 37 °C for 1 hour to remove all RNA from the sample. The RNase was inactivated by extraction with water-saturated phenol (pH 4.5) and nucleic acid precipitation was performed as before (Section 2.3). The resulting gDNA was then added to the 0.5 g of soil in the Lysing Matrix B tubes (Section 2.3) and the extraction was attempted.

3.2.2 Waikite Results

Due to the ubiquitous nature of RNases in the environment and the relative instability of bacterial RNA (Section 1.14), use of this procedure involved extensive method development. The steps involved in the development of this procedure are detailed in the following sections.

3.2.2.1 Extraction from Pure Culture *E. coli*

Initially, it was important to establish a working general extraction procedure for nucleic acids to ensure extraction on intact, full-length transcripts for subsequent sequencing and analysis. This was accomplished by first starting with extractions of *Escherichia coli* grown in pure culture. Initially, week-old cells were grown (Section 2.2) and used for in the general nucleic acid procedure (Section 2.3). Figure 3.4 illustrates the visualization of total nucleic acids yielded from a week old culture of *E. coli* when subjected to the above extraction procedure.

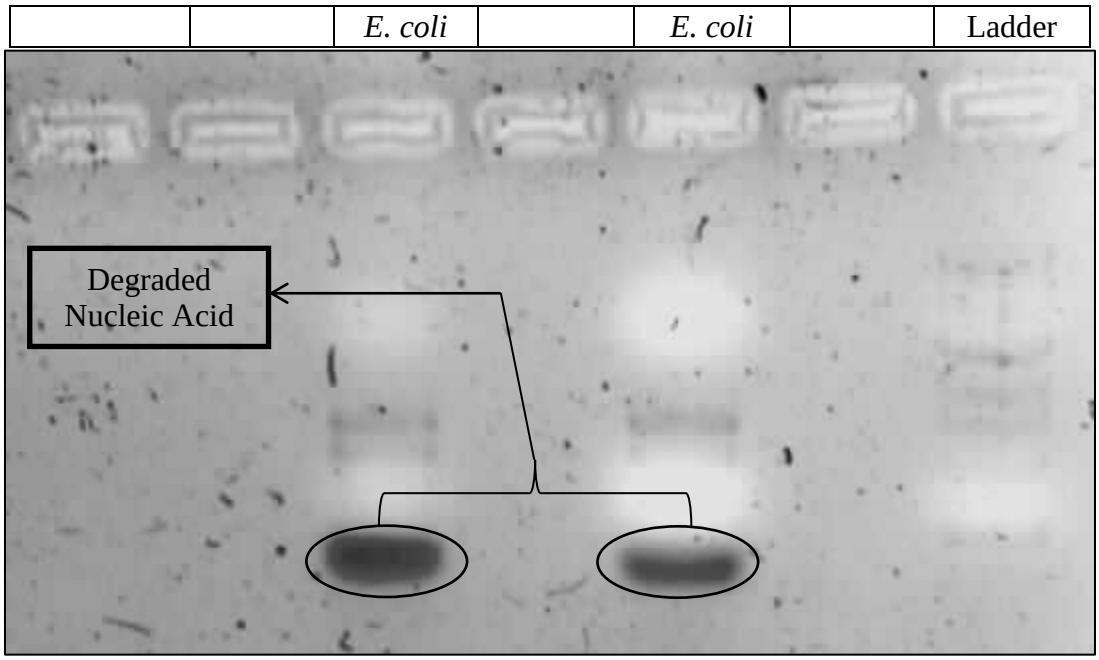


Figure 3.4. Total Nucleic Acid Extraction of Week Old *E. coli* Cells Grown in Pure Culture.

The extensive nucleic acid degradation observed in Figure 3.4 highlighted the need to ensure the integrity of the RNA transcripts present in the sample. Although mRNA transcripts vary in length and appear as a smear on most gel runs, degradation of an mRNA sample can be judged by the bands associated with the 16S and 23S ribosomal RNA (rRNA) gene. Since these molecules are present in an environmental sample in high numbers (relative to mRNA transcripts), they should appear as dark discrete bands in electrophoresis profiles. In a sample where RNA degradation has taken place (Figure 3.4), there would be no discrete rRNA bands and the highest concentration of sequences in the sample will be of short length (at the bottom of the lane). Much of the nucleic acid degradation could have been due to the age of the *E.coli* cells. Therefore determination of the time at which the culture was in "log" phase was important to the experiments involving mRNA (Appendix A). Identification of a "log" phase culture allows for a

standard culture in which the transcripts are of optimal integrity. Therefore, from this point on in the thesis, all cultured were grown to this time point, collected, and then frozen for future use as in Section 2.2.

After the generation of a log phase *E. coli* culture, nucleic acid was extracted (Section 2.3) to assess the amount of degradation of transcripts (Figure 3.5).

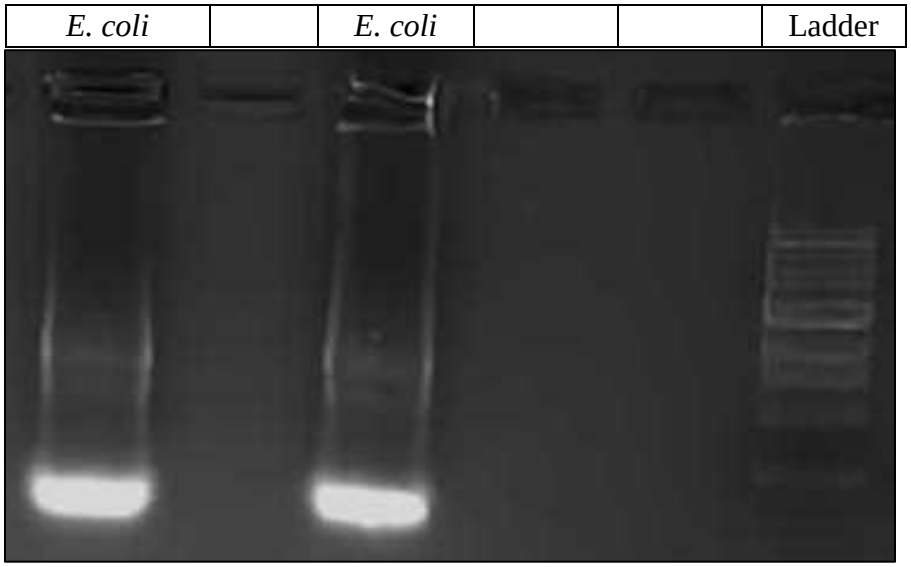


Figure 3.5. Extraction of Log Phase *E. coli*. A large amount of degraded nucleic acid was observed as the bright spots at the very bottom of the two *E. coli* lanes. There is no difference between the two lanes as far as extraction procedure and volumes loaded into the gel.

Significant degradation of nucleic acid was again evident in the extraction (Figure 3.5) suggesting that degradation was still occurring. The effect of incubation temperature on the extraction process was investigated.

The half-life of bacterial mRNA transcripts is notoriously short (Section 1.14). Therefore it was postulated that lowering the temperature would slow degradation and help preserve the integrity of mRNA transcripts present in the sample while the extraction was taking place. The reasoning being that at lower temperature, the enzymes

involved in degradation of RNA would operate at a much slower rate. With the aid of a chilled centrifuge and a PCR ice-block, the general extraction procedure was conducted in parallel at room temperature and in "chilled" conditions (4 °C). The resulting gel of the total nucleic acid extraction can be seen in Figure 3.6.

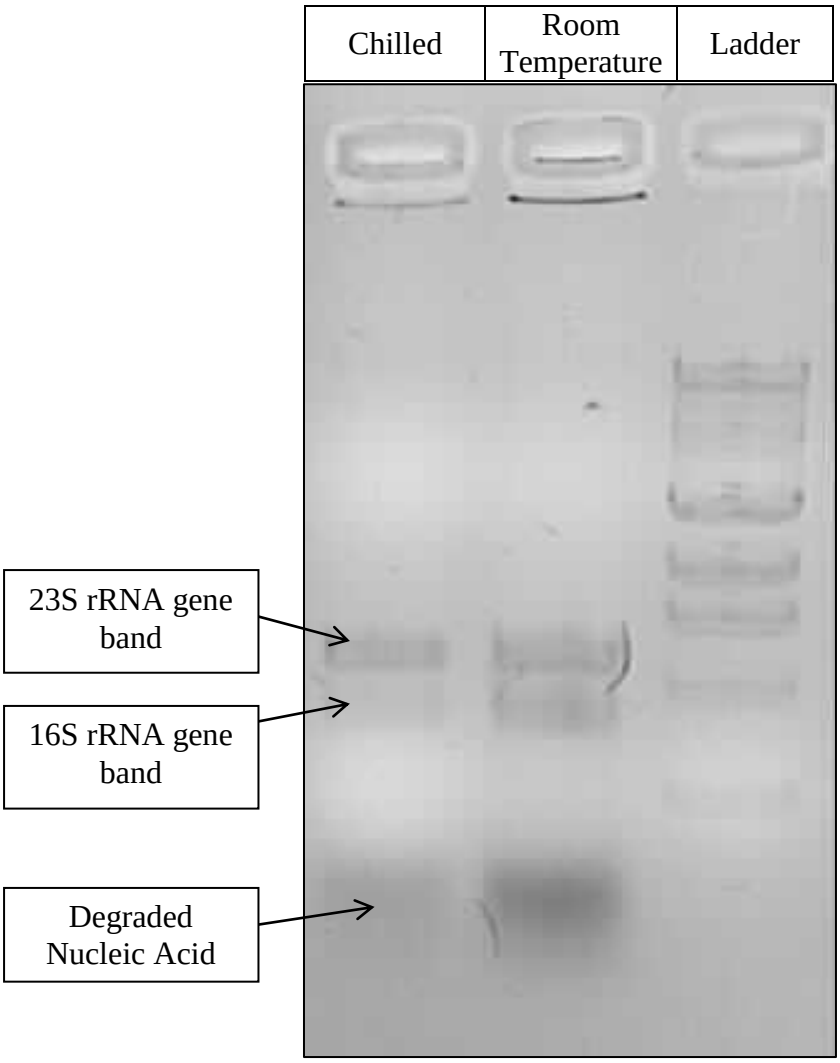


Figure 3.6. "Chilled" and "Room Temperature" Extractions. More degraded nucleic acid can be seen in the "Room Temperature" extraction. Both the 16S and 23S rRNA gene bands in both the middle and left lanes are intact.

Although both lanes present in Figure 3.6 show distinct rRNA gene bands (indicating intact mRNA transcripts present in the sample), the middle lane contains a more robust signal at the bottom of the lane. This indicates that there is more degradation occurring during an extraction conducted at room temperature than the extraction conducted at 4 °C. Therefore an extraction conducted in a "chilled" environment (4 °C) allows for more intact transcripts to be extracted from the sample. From this point on in the experiment, all extractions were conducted in a chilled environment using a PCR ice

block and all centrifuge steps conducted in a centrifuge chilled to 4 °C. At this point, with a relatively good extraction from pure culture, an extraction was conducted with the geothermal soil sample spiked with this pure culture *E. coli* to assess the nucleic acid yield and to see whether nucleic acid from *E. coli* could be detected in the soil samples.

3.2.2.2 Extraction of "Spiked" Soil

An extraction of soil spiked with log phase *E. coli* culture was used to assess the extraction efficiency. *E. coli* (Section 2.2) was added to 0.5 g of the Waikite soil (collected in Section 3.2.1.1) and then extracted at 4 °C as discussed in the preceding section. In parallel, nucleic acid was extracted from pure culture *E. coli* as a positive control to assess the extraction efficiency (Figure 3.7).

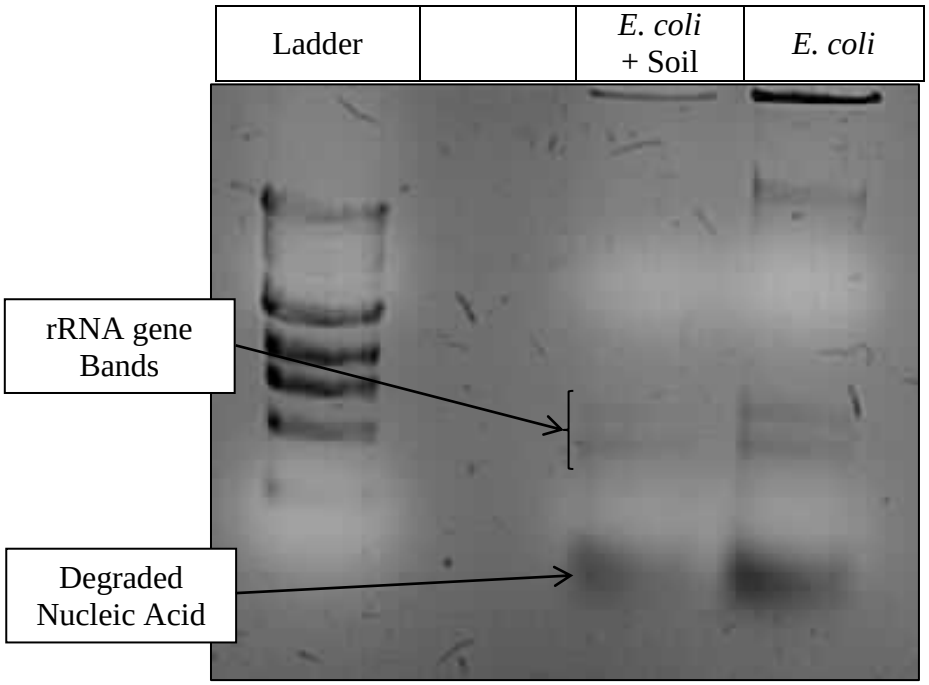


Figure 3.7. Extraction from "Spiked" Soil and *E. coli* Pure Culture. Although the soil does not absorb all of the nucleic acid from the sample, the amount of nucleic acid is decreased in the lane with the soil.

The decrease of signal intensity from the spiked soil sample, when compared to the positive control, clearly indicates that the geothermal soil was inhibiting the extraction efficiency. In order to test whether this reduction in extraction efficiency was due to a methodology issue or was specific to the Waikite soil, the same experiment was repeated, but with a different soil.

Soil collected from outside the Wilson Building (GNS Science, Wairakei, New Zealand) was used in this experiment as a control soil. This soil was "loamy" and very aerated. 0.5 g of this soil (wet weight) was spiked with a frozen *E. coli* as before and then subjected to the general extraction procedure (Section 2.3) as before (Figure 3.8).

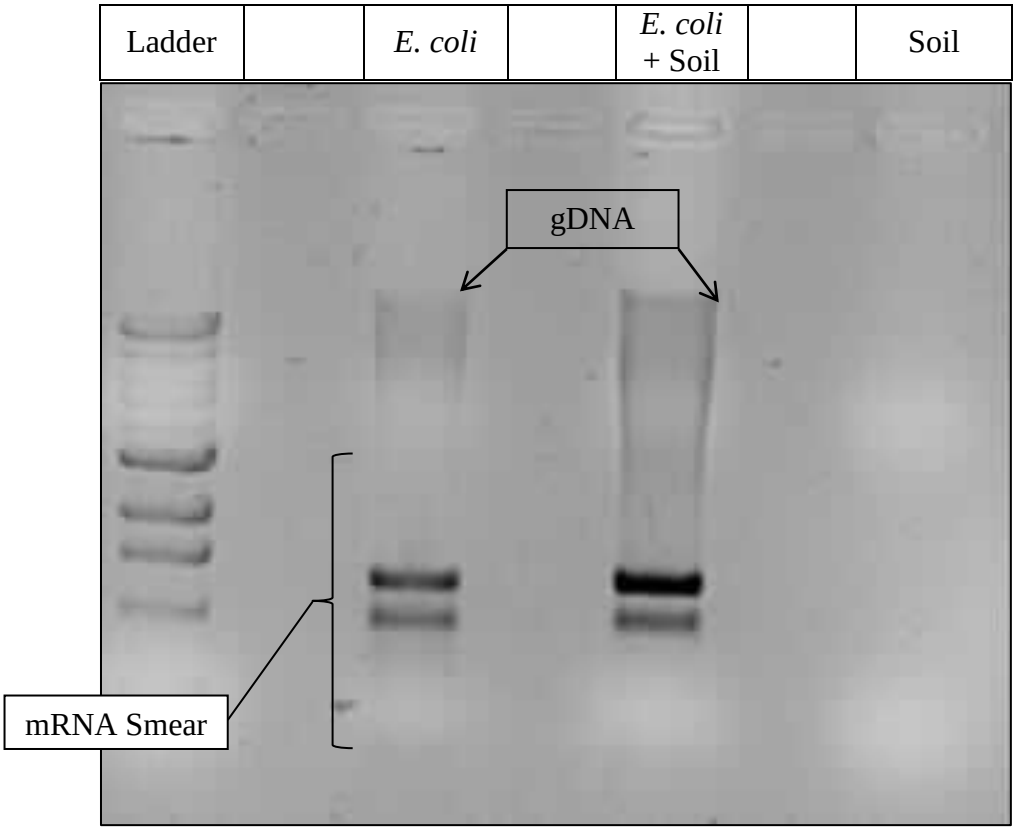


Figure 3.8. Wairakei Loamy Soil Spiked with *E. coli*. The distinct rRNA gene bands illustrate that all other transcripts maintained integrity during the extraction procedure.

Both lanes containing visible amounts of nucleic acid had highly intact banding and even contained visible smears of gDNA and mRNA. As judged by the signal intensity, it did not seem as though the Wairakei soil was inhibiting the extraction. Although there was no observable nucleic acid extracted from the soil-only sample. This was possibly due to a scaling issue.

The results of this experiment (Figure 3.8) suggest that the soil at Waikite was inhibiting the nucleic acid extraction efficiency. The high clay content of the Waikite soil [121], was again identified as a possible candidate for this inhibition due to the literature supporting clay-like soil inhibition of nucleic acid extraction (as discussed previously). Therefore quantification of total nucleic acid (Section 2.6) was done on a recreation of the experiment detailed in Figure 3.7. Table 3.2 illustrates the difference in total nucleic acid extraction efficiency from *E. coli* before and after the addition of Waikite soil.

Table 3.2. Quantification of Total Nucleic Acid Extraction. There was more nucleic acid extracted from the *E. coli* than with the addition of soil to the sample. This is indicative of clay colloids present in the soil.

	DNA	**RNA**	**Total**
E. coli	12.99 µg	24.69 µg	37.68 µg
Soil + *E. coli*	10.94 µg	22.76 µg	33.7 µg

There was a 3.989 µg (10%) loss of extracted nucleic acids from the sample containing *E. coli* and soil compared to that of *E. coli* extraction alone. A comparison of the extraction efficiency between "loamy" Wairakei soil and the Waikite soils strongly suggest that there is a significant loss of extractable nucleic material. These observations suggest that

i) the Waikite soil or a component of the soil was inhibiting the extraction or ii) the soil was adsorbing the extracted nucleic acids.

While it was known that the Waikite soil contained viable thermophilic microorganisms [121], the inability to extract usable quantities of may have been due to low *in situ* cell numbers. In order to assess the concentration of cells required generate a usable nucleic acid extraction, known concentrations of *E. coli* cells where inoculated into the Waikite soil. Cell numbers were determined (data not shown) and were serially-diluted into aliquots of 0.5 g Waikite soil and extracted (Table 3.3). Quantification was conducted as before (Section 2.6).

Table 3.3. Total RNA Extracted from Differing Amount of *E. coli* Cells. All final values were measured in a volume of 50 µl TE Buffer and are the result of 0.5 g of soil being inoculated with a known amount of *E. coli* cells.

Cell Amount	**10^{10} Cells**	**10^{9} Cells**	**10^{8} Cells**	**10^{7} Cells**	**10^{6} Cells**
Average (µg of RNA/ml)	199	59.60	18.05	7.26	1.50
Standard Deviation	1.00	0.42	0.05	0.065	0.1
Total µg RNA / Extraction	99 µg	2.98 µg	.5675 µg	.3627 µg	.075 µg

Although the data obtained for "10^{10} Cells" is at the upper limit of the detection range for the Qubit Quantification System (200 µg/ml) and an overestimation, this experiment serves to indicate how much nucleic acid is capable of being absorbed by 0.5 g of Waikite soil. Conversely, the cell numbers required to generate enough extractable nucleic acids could be the reason for the detection threshold. The data suggests that a nucleic acid extraction from between 10^{8} and 10^{9} cells per 0.5 g of soil will saturate the

soil adsorption sites. These data infers that at ~100x scale-up in the quantity of soil required for extraction to generate a usable quantity of nucleic acids.

The obvious difference between the two soils used in this experiment was the clay content of the Waikite soil (as previously noted [121]). Clearly, the adsorption capacity of Waikite soils was significant enough to reduce the nucleic acid extraction efficiency to below reasonable detection. Previous observations state that clay-like soils can significantly affect nucleic acid extraction efficiencies (Section 3.2.1.1). The inhibition was proposed to occur when the lysis of the cells exposed the nucleic acid to the soil environment [45]. With the presence of non-native nucleic acid, or another molecule of similar chemical nature to that of nucleic acids (possessing a net negative charge), it was proposed that the binding sites would preferentially bind the blocking agent, leaving the native nucleic acid free to be extracted. Subsequent experiments investigated whether the use of blocking agents could preferentially bind the adsorption sites on the soil colloids.

3.2.2.3 Blocking with Skim Milk

The ability of soil colloids to absorb nucleic acid has been well studied (Section 3.2.1.1). Skim milk powder was tested as a possible blocking since it had been used as a blocking agent to extract nucleic acids from clay-like soils [128]. The efficacy of skim milk as a blocking agent was tested by adding it to the Waikite soil and *E. coli* mixture. The skim milk was prepared and added to the extraction procedure (Section 3.2.1.2). The soil, milk, and *E. coli* mixture was then incubated for 1 - 24 hours at 4 °C to determine the optimal treatment time (Figure 3.9).

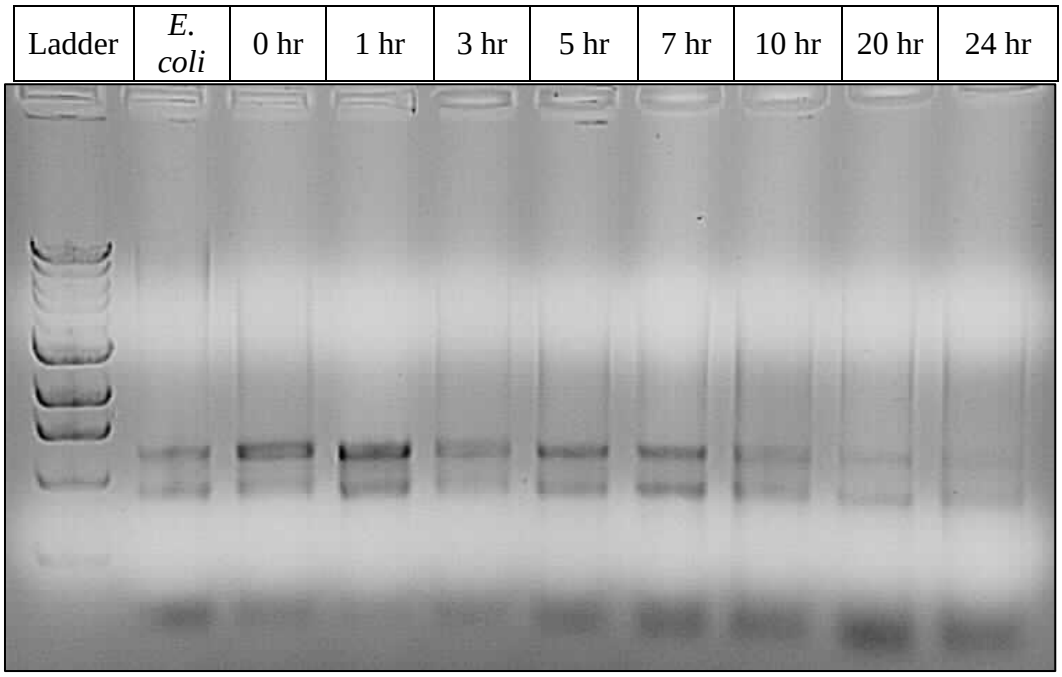


Figure 3.9. Incubation of Waikite Soil, *E. coli,* and Skim Milk. Different time points (in hours) of incubation of extraction at 4 °C using 0.5 g Waikite soil, *E. coli*, and 700 µl skim milk. The intact band corresponding to the rRNA genes along with small signal at the bottom of the lanes (degraded nucleic acid) signify that the extractions were successful, and the transcripts intact.

Nucleic acid extraction using skim milk as a blocking additive was successful in all incubations (Figure 3.9). Yet this could have also been a result of saturation of the binding sites due to the cell concentration used for inoculation (Table 3.3). The intensity of the degraded nucleic acid band increases as the incubation time increases. This indicates that extended incubations result in an elevated loss of intact nucleic acid. Although all solutions and additions to the soil were RNase-free, the soil itself contains RNases, therefore, it was likely that the longer nucleic acids were left in the soil matrix (exposed to the potential RNase), more degradation occurred. The quantification of the nucleic acid extracted from the soil (Table 3.4) confirms this observation. The general trend of these data (Figure 3.10) shows that incubation of the soil after one hour did not increase the amount of nucleic acid extracted from the sample. Likewise, the amount of

degradation of the nucleic acids in the sample increased with the longer incubations. These data lead to the optimal blocking time of 1 hour for the duration of the experiments involving skim milk.

Table 3.4. Quantification of Total Nucleic Acid Extracted from Skim Milk Incubation Experiment. Each extraction was done on 0.5 g Waikite soil + frozen *E. coli* + 700 µl of skim milk mixture. All values are of micrograms (µg) / ml of TE buffer and indicate total nucleic acid extracted from the each sample. "N. D." indicated that the quantification of nucleic acid at this time point was not done.

Time (Hours)	Conc. 1 (µg/ml)	Conc. 2 (µg/ml)	Conc. 3 (µg/ml)	Average (µg/ml)	Standard Dev.
0	200	323	317	280	69.34
1	Above Detection Limit	377	389	383	8.48
3	N. D.	N. D.	N. D.	N. D.	N. D.
5	Above Detection Limit	397	388	392.5	6.363
7	Above Detection Limit	374	385	379.5	7.7781
10	390	360	400	383.33	20.81
20	276	377	378	343.66	58.6
24	240	291	293	274.33	30.03

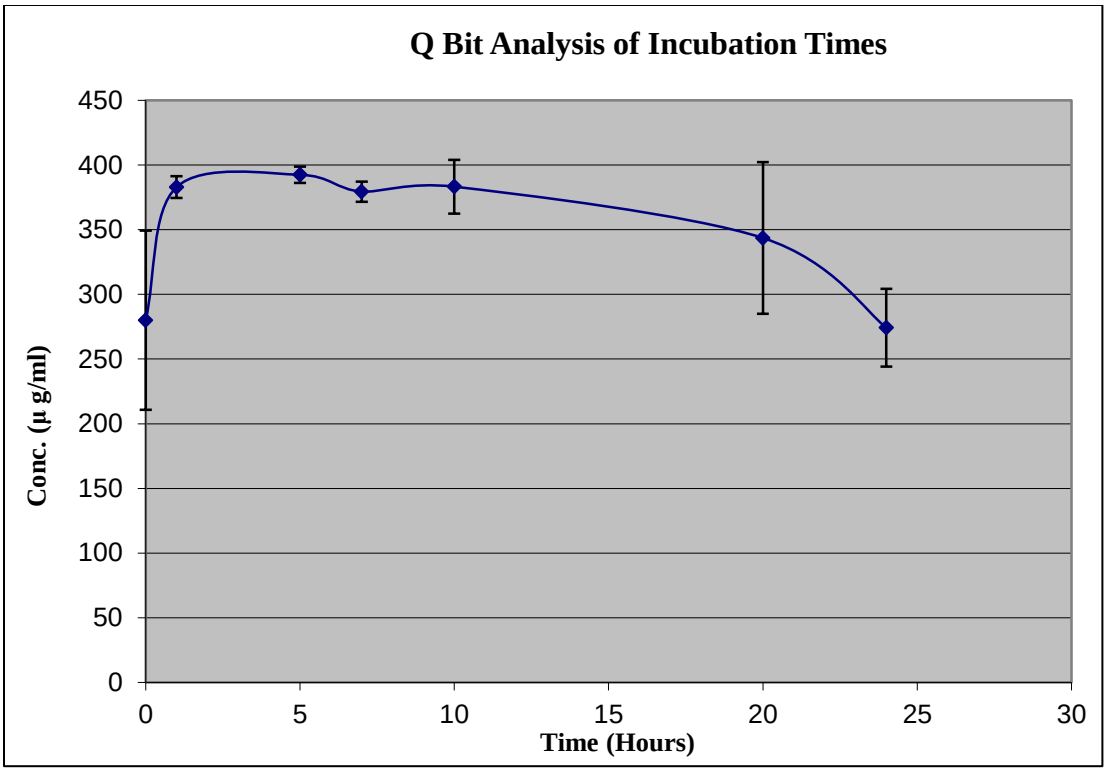


Figure 3.10. Concentration of Total Nucleic Acid. Concentration is in micrograms (µg)/ ml and are total nucleic acids extracted from each sample.

The optimal concentration skim milk powder required to block the adsorption sites was also investigated. Different volumes of skim milk powder (100 g / l H_2O) were added to 0.5 g of soil with frozen *E. coli* (Section 3.2.1.2). Total nucleic acids were quantified (Table 3.5) and checked for integrity via gel electrophoresis (Figure 3.11).

Table 3.5. Concentration of Nucleic Acid Extracted from Skim Milk-Blocked Soil. Amount is of total nucleic acid extracted from soil + milk + *E. coli*. All values are of total nucleic acid (DNA and RNA). Amount of skim milk is of a constant concentration.

Amount	Concentration (µg)
No soil	4.5 µg
No milk	4.2 µg
2 ml milk	4.3 µg
5 ml milk	4.9 µg
10 ml milk	3.9 µg

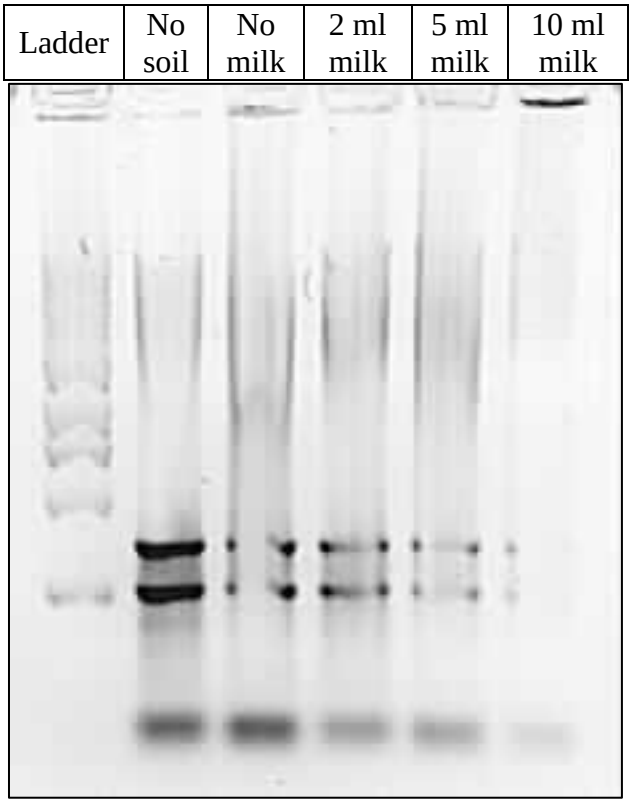


Figure 3.11. Nucleic Acid Integrity of Soil Extractions with Skim Milk. Total nucleic acid extraction of milk + soil + *E. coli* with varying amounts of milk added to the starting material. A smear of nucleic acid and intact rRNA gene bands can be easily seen in all lanes.

This extraction (Table 3.5) shows that 2 ml of skim milk is sufficient to block the clay in the soil. Any addition skim milk will result in the decreased yield of nucleic acid from the sample. These data indicated that the optimal conditions for nucleic acid extraction from the Waikite soil requires the addition of 2 ml of skim milk incubated for one hour.

After many extractions of the Waikite soil, the best extraction yielded only DNA. Figure 3.12 shows a typical extraction result. No rRNA gene bands were visible which suggests that the total RNA concentration was very low. With the failure of the skim milk-based blocking technique, an alternative approach to improve nucleic and RNA extraction efficiency was sought.

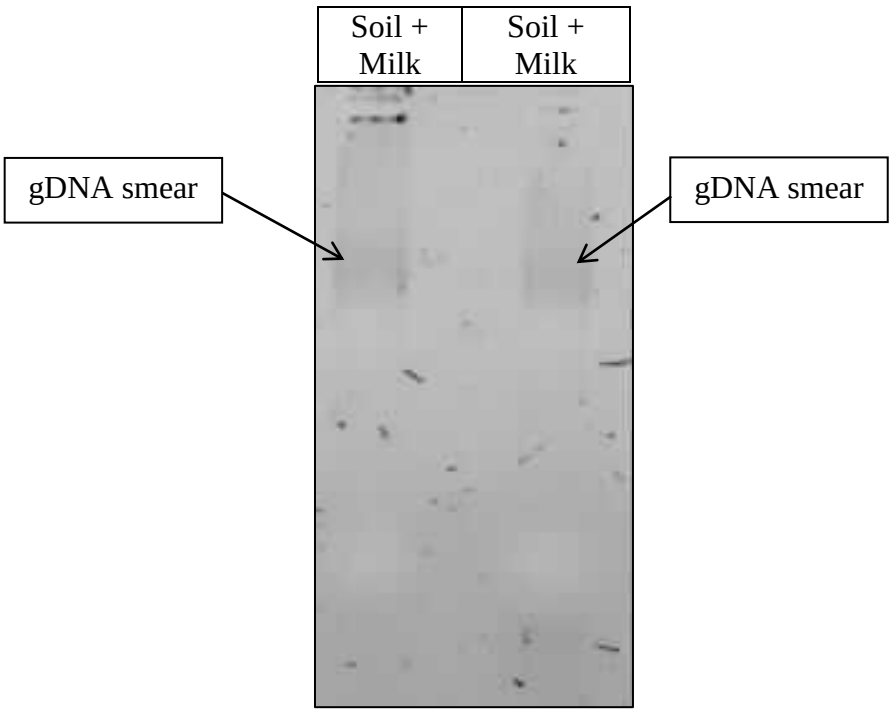


Figure 3.12. Extraction from Waikite Soil with the Addition of Skim Milk. Only a band containing gDNA could be extracted from the sample. No RNA was present during Qubit Quantification, verifying the result in this gel.

An alternative blocking strategy was the addition of gDNA to the soil. Frostegård [42] found an increase in RNA extracted from clay-like soil when gDNA was added to the soil in an attempt to block the soil colloids. This was found to be a result of the competition of non-native nucleic acid for binding sites on the soil colloids. Also, RNA was used to compete for soil colloid binding sites when DNA was the target being extracted from clay-like soil [128]. This strategy was replicated in the Waikite soils with the addition of gDNA.

3.2.2.4 Blocking with Genomic DNA

Genomic DNA (gDNA) (Section 3.2.1.2) was prepared from *E. coli.* Briefly, the total nucleic acid extraction (Figure 3.13) was treated with RNase A and the resulting genomic DNA (Figure 3.14) was precipitated and stored at -80 °C and used as "processed" genomic DNA.

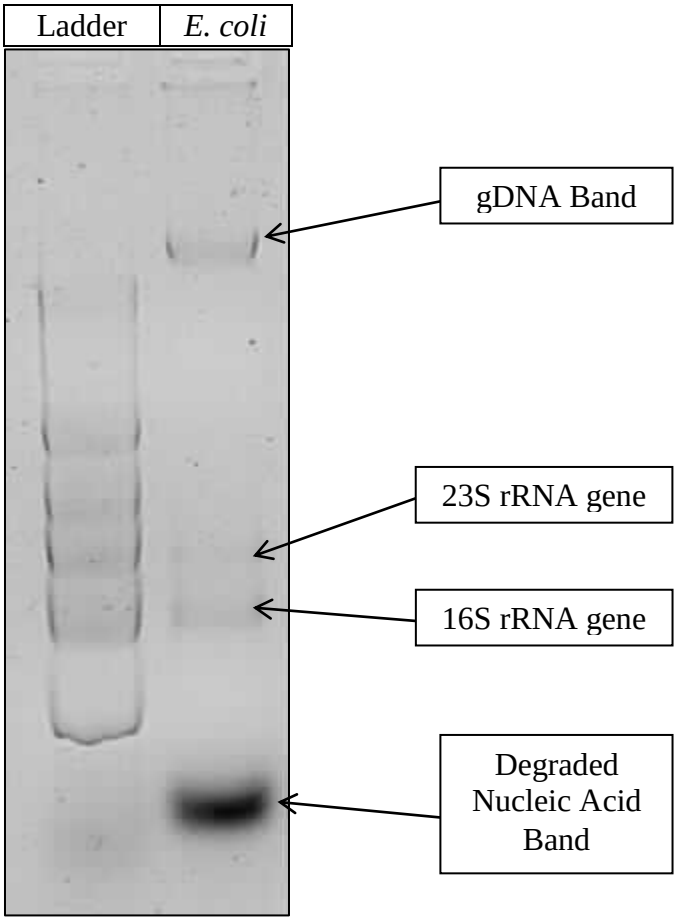


Figure 3.13. Total Nucleic Acid Extraction from Frozen Pure Culture *E. coli*. A band of genomic DNA (gDNA) can be seen at the top of the lane, the rRNA gene bands can be seen in the middle of the lane, and the degraded nucleic acid portion can be seen at the bottom of the lane. This represents a total nucleic acid extraction.

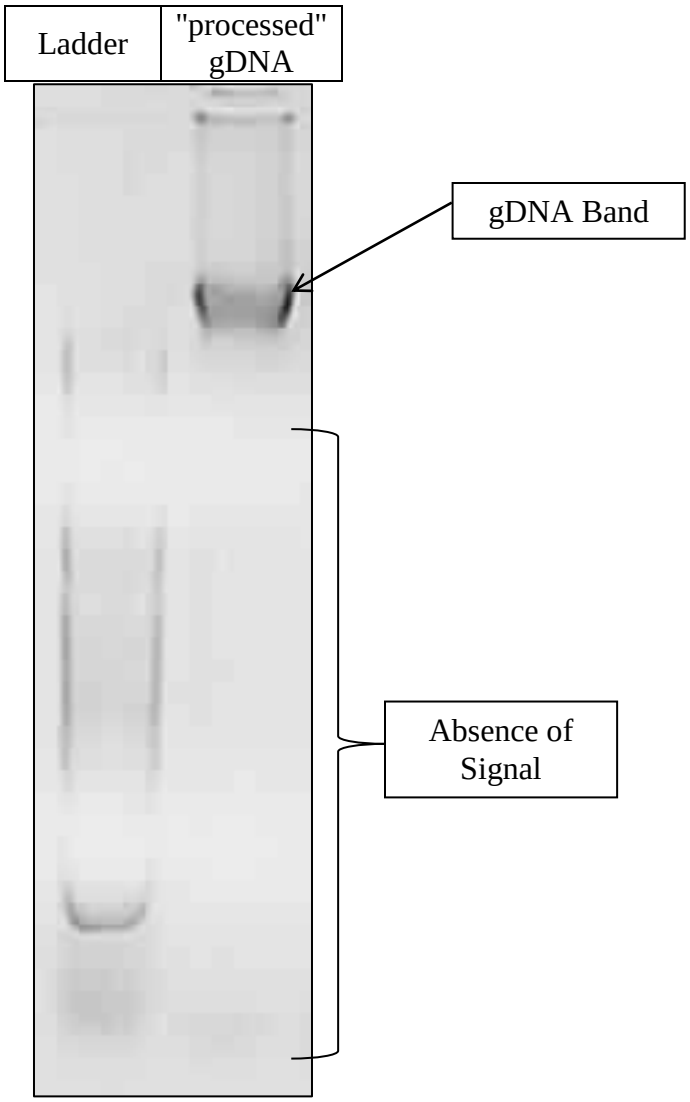


Figure 3.14. Processed gDNA from *E. coli*. The removal of the rRNA gene bands and degraded section (Figure 3.13) represent the removal of RNA from the sample by RNase A treatment.

Initially, the concentration of processed gDNA needed to sufficiently block the soil colloids was determined. Previously obtained data (Table 3.2) indicated that almost 4 ug of nucleic acid were being absorbed by 0.5 g of soil. Therefore, 4, 8, and 16 µg of processed gDNA were added to 0.5 g of soil along with 2 ml of skim milk in an attempt to block the soil colloids. The resulting amounts of RNA and DNA is presented in Table 3.6 and a visualization of the run can be seen in Figure 3.15.

Table 3.6. Quantification of Nucleic Acid Extracted from Soil with the Addition of Milk and Processed gDNA. A measurement of "Below Detection Limit" indicates that the amount was below the threshold of detection for the Qubit system.

Conditions	RNA	DNA
Milk	Below Detection Limit	Below Detection Limit
Soil	Below Detection Limit	0.017 µg
Soil + Milk	Below Detection Limit	Below Detection Limit
Soil + Milk + 4µg of gDNA	0.106 µg	0.329 µg
Soil + Milk + 8µg of gDNA	0.1735 µg	4.00 µg
Soil + Milk + 16µg of gDNA	0.233 µg	9.5 µg

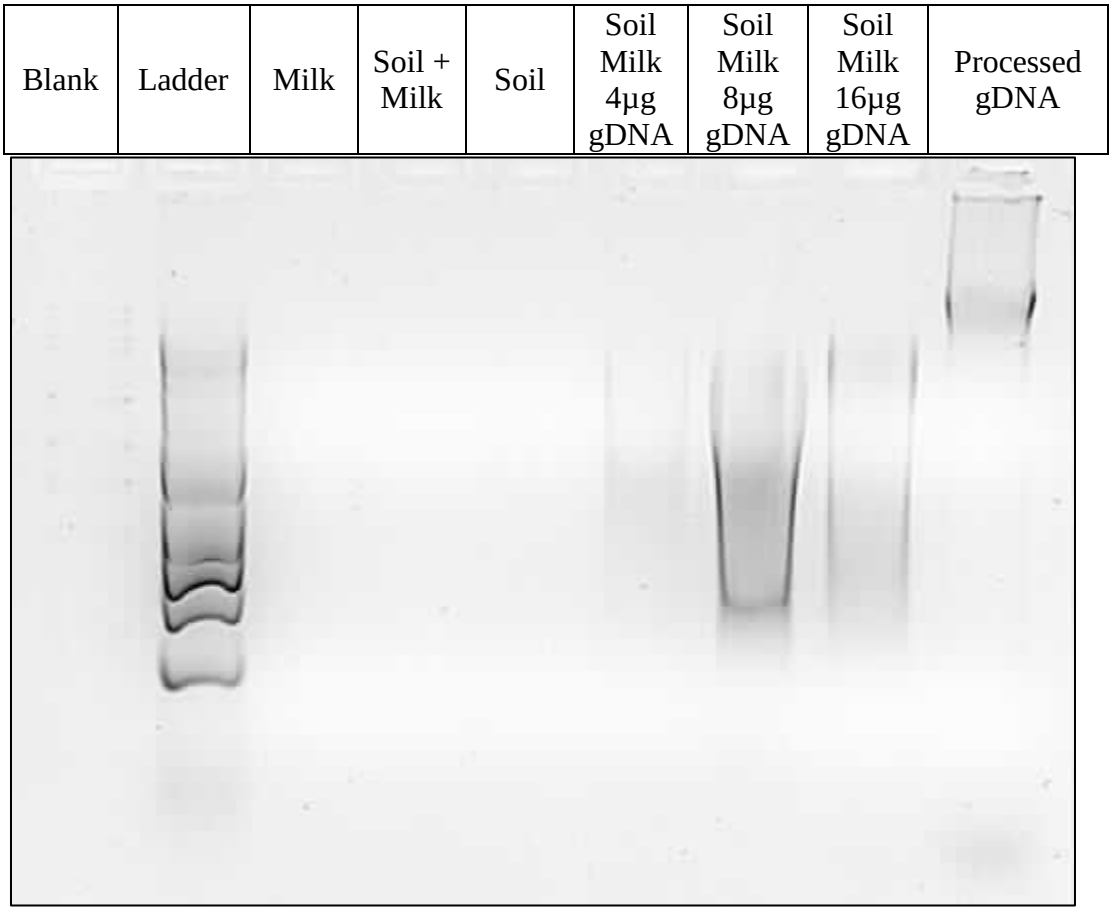


Figure 3.15. Visualization of Processed gDNA Blocking Experiment. Although there are amounts of nucleic acid visible in this gel, quantification with Qubit determined them to be primarily DNA. The generation of this smear (as opposed to the band seen in Figure 3.14) was due to the bead beating step shearing the transcripts. Absence of the rRNA gene bands confirms the lack of RNA in the sample.

This experiment shows the absorption of a standard amount (~4 µg) of non-native nucleic acid which was added to the soil and skim milk mixture. It was also postulated that any measure of RNA present in the sample (0.23 µg) was not above the noise level inherent in the Qubit Quantification System. With the options present in the current literature exhausted, very little RNA was being extracted from the soil obtained from Waikite. Even if this was not the case, in order to sequence the resulting cDNA, a certain threshold amount needed to be met (500 ng – 1 µg cDNA was needed for library creation – S. Hou, personal communications). This amount was the minimum amount which could be used for a 454 GSL FX Pyrosequencing run. Therefore, either a large-scale up of the extraction needed to be done (which was prohibitively expensive), or a replacement site needed to be chosen to continue the experiment.

At this point, there was a decision between the student and his advisors to move to another site more suitable for bioprospecting using a metatranscriptomic study.

3.2.3 Waikite Summary of Results

RNA was unable to be extracted successfully from the Waikite soils. The concentrations of RNA (including the rRNA fraction) did not exceed 0.23 µg which was significantly less than required for the generation and sequencing of a cDNA library of the environmental mRNA. The extraction protocol was extensively investigated and modified in an attempt to improve extraction efficiency. This included attempts to block nucleic acid binding to the dominant clay fraction in the Waikite soil using skim milk and gDNA. To the author's knowledge, extraction of RNA from clay soils and in particular

geothermal clay soils for the purpose of a metatranscriptomic study has not been attempted previously.

The author believes that either the resident microorganisms did not have an adequate cell density in the amount of soil used for the extraction (0.5 g wet weight) to compensate for the nucleic acids absorbed by the soil colloids. It was observed that when a certain threshold of cells is met, and the binding sites on the soil colloids are occupied, nucleic acid can readily be extracted from the soil (Table 3.3) as intact transcripts (Figure 3.6). It was also determined that the 0.5 g of soil was capable of binding 3.989 µg of total nucleic acids (Table 3.2). It was thought that the amount of nucleic acid absorbed was above the amount of nucleic acid present in the resident microorganisms. Although there were viable microorganisms present in the sample [121], the amount of DNA needed for detection of microorganism's 16S rRNA gene sequence pales in comparison to the amount of RNA needed for a metatranscriptomic study.

In the future, if this study is attempted again at this site, new techniques could be employed to alleviate the problem of nucleic acid absorption. Attempting to wash the whole cells off of the soil colloids before lysis could alleviate the problem of nucleic acid absorption, but would lead to bias based on how well the cells were bound to the soil or selectively washing one sort of microbial species from the soil and not another, therefore biasing any downstream analysis. This would also call for the fixing of the cells before any washing was attempted. Also, if the relative amount of extraction buffer was increased and the amount of soil kept the same, it could potentially free some of the nucleic acid from the soil colloids and allow for the extraction. This has the potential to make the entire procedure prohibitively expensive but would increase the amount of

buffer in which the freed nucleic acids were dissolved. In the same vein, if the procedure detailed in this write-up was scaled up to make the 0.23 µg total RNA an amount which then could be used for a metatrascriptomic study (once the mRNA was enriched), the experiment could then be accomplished. At this point, however, it remains prohibitively expensive and would be difficult to physically obtain that large quantity of soil without causing environmental damage to the delicate spring system. It would also be difficult to obtain a uniform and representative sample using such large volumes without crossing several ecosystems, and therefore bias the downstream analyses.

In consultation with the author's supervisors, a decision was made to investigate alternative sites to conduct a metatranscriptomic study to identify novel thermostable GHs. Continued method development of the Waikite site may have eventually resulted in a suitable RNA extraction. However, due to time constraints, continued work was considered risky and a secondary backup site was identified at Ngatamariki. While the site in question did not have the advantage of supporting a known cellulolytic microbial population in addition to the several isolated and sequenced strains like Waikite, the springs was relatively low in clay content and contained a visible biofilm in the vicinity of a cellulosic biomass. For this reason, work at Waikite was suspended and the RNA extraction focus moved to Ngatamariki.

3.3 Ngatamariki

Radiata Pool is located at the sample site Ngatamariki in the Taupo Volcanic Zone, New Zealand (Figure 3.16 and Figure 3.17). The pool presents perfect conditions for an experiment targeted at llignocellulosic active enzymes due to a large pine tree

(*Pinus radiata*) lying submerged in the middle of a 70 °C pool. Adjacent to the submerged pine tree was a significant biofilm (Figure 3.18) which was suspected of harbouring cellulolytic thermophiles. The spring also had a lack of dense, clay-like soil which was anticipated would alleviate the issues experienced at Waikite.

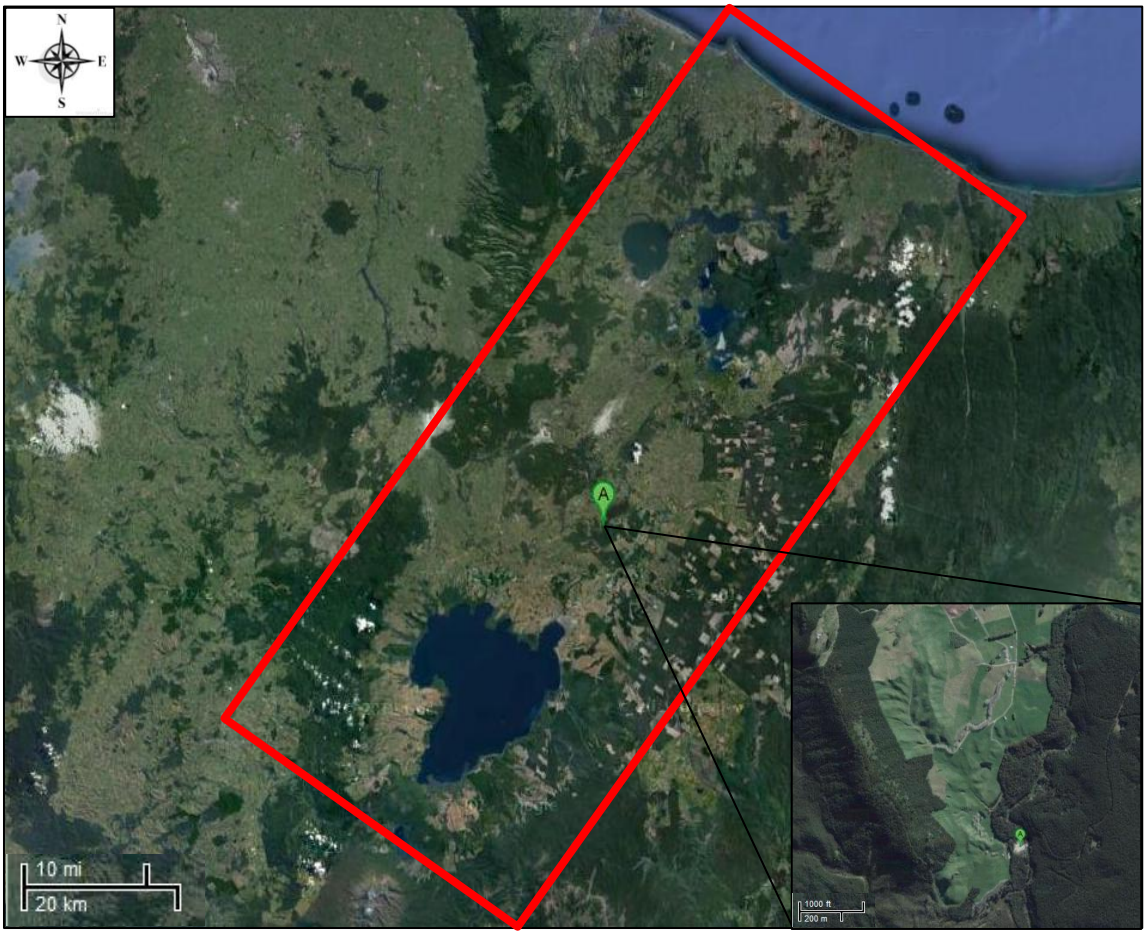


Figure 3.16. Location of Ngatamariki on the North Island of New Zealand. Ngatamariki is denoted by the "A" on the map. The area donated in the red box is the Taupo Volcanic Zone. The map is of the North Island of New Zealand. Inset is a zoomed image of the sampling site. Located amidst a *Pinus radiata* plantation, the site itself is surrounded by many trees of this genus and species (the dark green area on the inset map).

Figure 3.17. Relative Position of Radiata Pool at Ngatamariki. Location of Radiata Pool in relation to the other features of the Ngatamariki sample site. The Pine trees present in the middle of the site (denoted "Pine Trees") have all been killed by an eruption at the site in April 2005, and formed the "Larger Pool".

Figure 3.18. Radiata Pool. *Pinus radiata* can be seen to have fallen across the middle of the geothermal surface expression. There was also a larger tree lying perpendicular to these smaller trees.

While Radiata Pool appeared to be a reasonable candidate to investigate cellulolytic thermophiles and thermostable GHs, there was no empirical evidence at the time of this investigation that confirmed cellulolytic activity. While the biomass adjacent to submerged pine log was an obvious candidate (Figure 3.19), it should be noted that this geothermal feature also contains other constituents that can support microbial life.

Figure 3.19. Large *Pinus radiata* Lying Submerged in Radiata Pool. The grey biomass surrounds the submerged sections of the large pine tree. The border between the grey biomass and the green photosynthetic organisms denotes the change in temperature of the geothermal pool and the surrounding mesophilic environment.

Table 3.7 details physiochemical measurements taken of the water circulating in the pool. Many of the microbial constituents present in the spring were capable of acting as an energy source including aqueous phase H_2, CH_4, H_2S, and ferrous iron (M. Stott, personal communications). Also, simple organic compounds such as formate, acetate and butyrate are known to be present in the spring water (data not presented).Therefore in order to ensure the selection of cellulolytic thermophiles as a source of thermostable GH's, an enrichment was made which ensured the presence of microorganisms capable of cellulolysis.

Table 3.7. Physiochemical Measurements in Radiata Pool Spring Water.

Species or Physical Measurement	Units	Concentration
temperature	°C	71
Bicarbonate (total)	mg/L	569
pH		7.67
Aluminium (Al)	mg/L	0.22
Arsenic (As)	mg/L	0.17
Boron (B)	mg/L	9.5
Calcium (Ca)	mg/L	0.99
Caesium (Cs)	mg/L	0.45
Chloride (Cl)	mg/L	654
Conductivity	µS/cm	1013
Iron (Fe)	mg/L	<0.08
Lithium (Li)	mg/L	3.3
Magnesium (Mg)	mg/L	0.11
Manganese (Mn)	mg/L	0.009
Potassium (K)	mg/L	24
Silica (as SiO_2)	mg/L	251
Sodium (Na)	mg/L	649
Sulfate (SO_4)	mg/L	29
Sulfide (total as H_2S)	mg/L	1.6
Argon (Ar)	mmole/100mole H_2O	4.2
Helium (He)	mmole/100mole H_2O	0
Hydrogen (H_2)	mmole/100mole H_2O	2.2
Methane (CH_4)	mmole/100mole H_2O	2.9
Nitrogen (N_2)	mmole/100mole H_2O	90.1
Oxygen (O_2)	mmole/100mole H_2O	0.6

3.3.1 Materials and Methods Specific for Ngatamariki

3.3.1.1 Enrichment of Cellulolytic Microorganisms

Unlike the Waikite sampling site, cellulolytic microorganisms had not been confirmed at Ngatamariki. Therefore, substrates of differing compositions of cellulose were used to encourage the growth of microorganisms which were able to utilize the substrates for growth. Three different substrates were chosen as the starting material, each composed of different amount of cellulose. Whatman Cellulose Filter Paper (GE) is 100 % crystalline cellulose. Bleached kraft filter paper (Scion) is varying amounts of cellulose and hemicellulose but is relatively devoid of lignin. Unbleached kraft (Scion) is a cardboard-like substance which contains various amounts of cellulose, hemicellulose, and lignin. A close approximation of the composition of the unbleached kraft is: Lignin 3.3 %, Arabinan 0.6 %, Galactan 0.4 %, Glucan 80.8 %, Xylan 5.4 %, and Mannan 7.5 % (I. Suckling, personal communication).

Glass tubes were filled with shredded or intact substrate and capped. These tubes were then sterilized at 121 °C and one atmosphere of pressure for 30 minutes. After transportation to the sample site, 18G syringes were used to inject water from the geothermal pool into the tubes (Figure 3.20 and Figure 3.21). The tubes were bound together and then inserted adjacent to the fallen pine tree (*Pinus radiata*) to enrich for microorganisms already using the lignocellulose as an energy source (Figure 3.22). A wire cable was attached to the bound tubes to aid in retrieval.

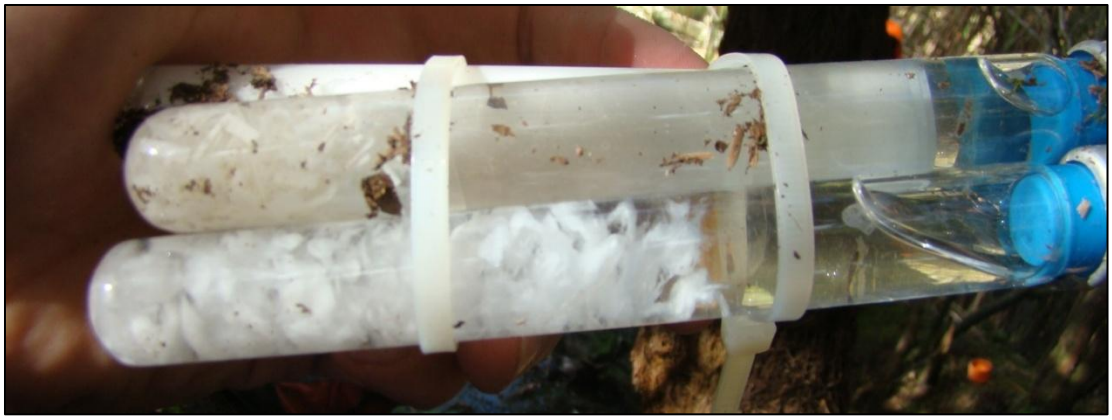


Figure 3.20. Enrichment Setup of Shredded Whatman Filter Paper and Shredded Bleached Kraft Filter Paper. The clear fluid and intact structure of the substrates illustrates a starting point in the enrichment setup. Both substrates were shredded before insertion into the tube to increase the surface area available to the microorganisms.

Figure 3.21. Shredded Unbleached Kraft and Intact Whatman Filter Paper. Substrate structure can be seen as well as clear water at the top of the tubes. The unbleached kraft was shredded to increase the surface area available to the resident microorganisms.

18G syringe needles were inserted into the tops of the tubes to permit the colonization by microorganisms and the expulsion of gas. This setup was left in the geothermal pool for two months before the progress of the enrichments was examined. However, inspection after the two months incubation showed a large amount of gas production which was not being expelled from the enrichments through the needles, restricting flow from the surrounding environment.

Figure 3.22. Relative Position of Enrichment Tubes in Radiata Pool at Two Months of Incubation. 18G Syringe needles were added to the tubes to permit the free movement of microorgansisms into the tubes. They were situated in an upright position in an attempt to keep the needles unblocked and allow for gas expulsion.

Figure 3.23. Relative Position of Enrichment Tubes before Removal from Environment. All four tubes can be seen in the upright position before the contents were removed and concentrated for transport back to the lab for extraction and enrichment.

Therefore, in an effort to increase the speed of colonization, it was decided that the caps should be removed and the tubes incubated for as long as possible. Tubes were incubated for a further two months (Figure 3.23) prior to processing and extraction.

On removal of the tubes from the spring, they were examined for the presence of "fluffy" cellular biomass and degradation of the substrate structure (indicating breakdown). Figure 3.24 shows the Whatman filter paper and the bleached kraft tubes.

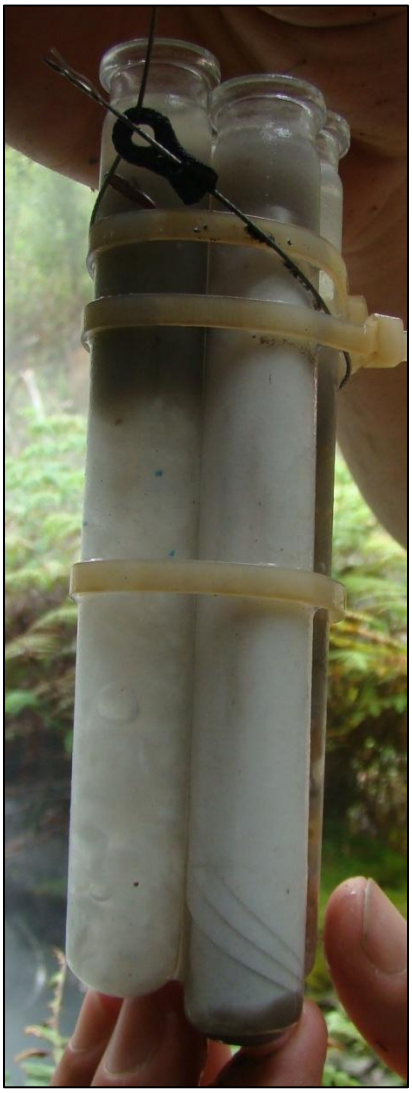


Figure 3.24. Tubes of Shredded Bleached Kraft Filter Paper and Intact Whatman Filter Paper after Four Months of Incubation. The grey biomass present in these tubes was thought to be microorganisms using energy sources other than the substrate present in the enrichment.

The tubes presented in Figure 3.24 have a large grey “fluffy” biomass, but this was only present at the very top of the tube. Figure 3.25 illustrates the shredded unbleached kraft and the intact Whatman filter paper. As opposed to the other tubes, the shredded unbleached kraft tube shows a biomass present throughout the entire tube with a slightly different color than the other tubes. Also, the presence of bubbles at the bottom of the tube illustrates the presence of metabolic activity since no gas was present at the initial time of insertion into the environment (Figure 3.21).

Figure 3.25. Shredded Unbleached Kraft and Intact Whatman Filter Paper. The presence of color change and air bubbles at the bottom of the tube show that metabolic activity present in the unbleached kraft tube must involve the substrate. Also, the change in structure of the unbleached kraft substrate shows degradation of the structure (Figure 3.21) and use of the substrate for an energy source.

3.3.1.2 Sample Concentration at Ngatamariki

The only change to the general extraction procedure (Section 2.3) was the use of filters to concentrate the biomass samples and the enrichment samples with the exception of the Whatman filter paper (see below). Sterile 0.22 µm filters were used to concentrate the water saturated soil to the point where nucleic acid of sufficient amount could be extracted. This procedure was used to not only concentrate the biomass sample, but was also used for the concentration of the enrichments.

Briefly, 0.22 µm cellulose nitrate (CN) membrane filters (Sartorius) were aseptically placed in a sterilized Swinnex 47 mm Filter Holder (Millipore) and were used to concentrate the samples of the biomass. Water-saturated biomass was passed into the filter with a SS+20ES 20 ml syringe (Terumo) to generate pressure. A total of 15 – 30 ml of water-saturated biomass was passed through each filter until no more could be passed through. The filters were removed aseptically and transferred to a 50 ml RNase-free Falcon Tubes (Ambion). A total of three to six filters were collected into the same Falcon Tube. The full Falcon Tubes were immediately snap frozen by immersion in liquid nitrogen to ensure stasis of the sample for the duration of the trip back to the laboratory. The filtrate was then extracted according to the protocol outlined in Section 2.3.

3.3.2 Ngatamariki Results

3.3.2.1 Enrichment Results

The enrichments were removed after a combined total of four months *in situ* incubation (Section 3.3.1.1). All enrichments were subject to the concentration and extraction procedure (Section 3.3.1.2) except the enrichment containing the intact Whatman filter paper. It was determined that there had not been enough utilization of the intact Whatman filter paper as a substrate to justify allocating resources to the preservation of the sample. It was observed that the surface area available to the resident microorgansisms was insufficient to encourage colonization by those which were able to use it as an energy source. All other enrichments and a sample taken from the grey biomass present in the environment (termed the "biomass sample") were used for the metatranscriptomic study.

3.3.2.2 Extraction of Total Nucleic Acids from Ngatamariki Samples

The total nucleic acid component was extracted (Section 3.2) from the biomass sample (Figure 3.26). The gDNA smear, intact 16S and 23S rRNA gene bands, a high quality mRNA smear, and relatively small amounts of degraded nucleic acid can be seen in the extraction. This was also done on the enrichment samples concentrated in Section 3.3.1.2. Unbleached kraft (Figure 3.27) and shredded Whatman filter paper (Figure 3.28) yielded nucleic acid. The bleached kraft enrichment did not yield detectable nucleic acid (either DNA or RNA).

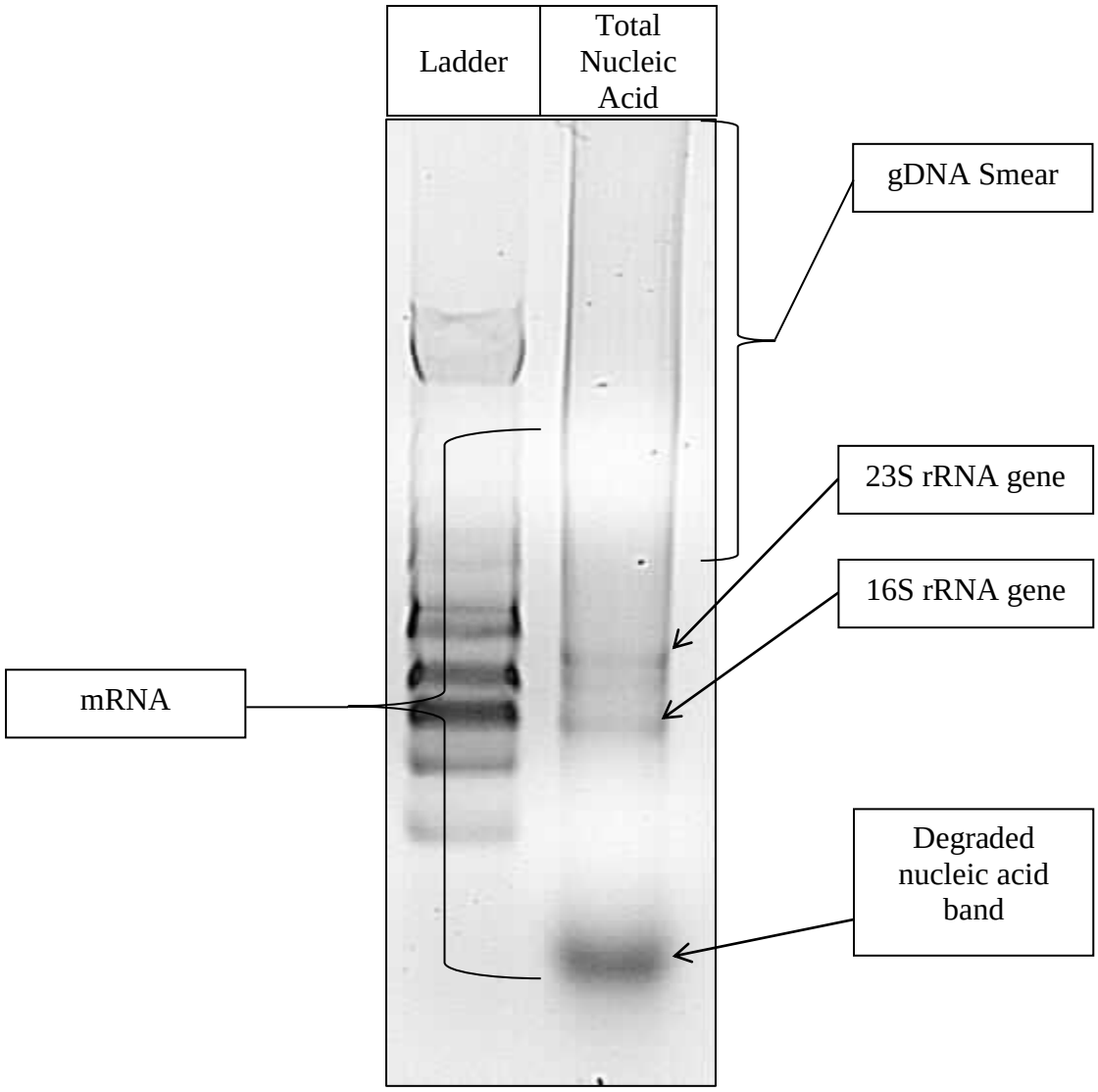


Figure 3.26. Total Nucleic Acid Extraction from the Biomass Sample Collected at Ngatamariki. A DNA component is visible at the top of the lane. Distinct rRNA gene bands are present along with a very minimal "degraded" nucleic acid component at the very bottom of the lane.

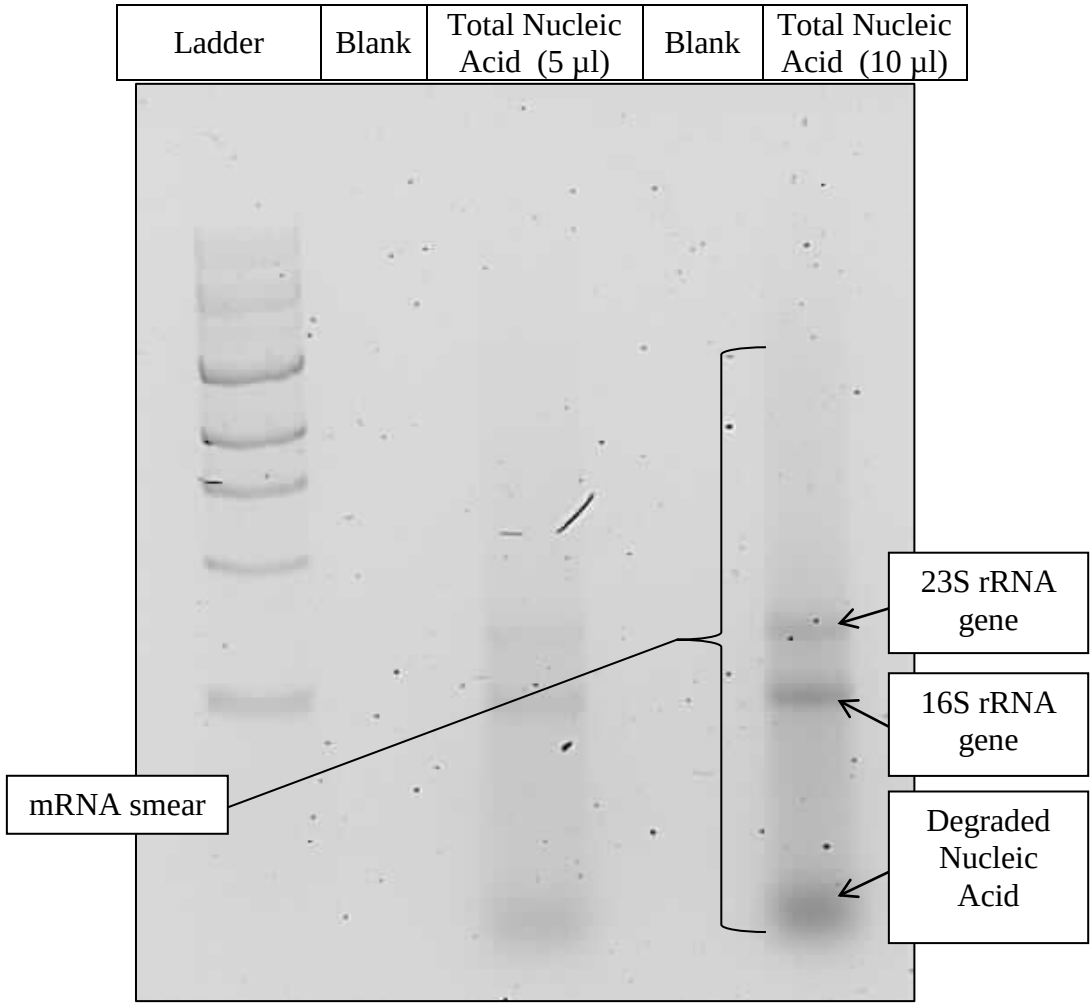


Figure 3.27. Total Nucleic Acid Extraction of the Unbleached Kraft Enrichment. DNA, rRNA gene bands, degraded nucleic acid section, and mRNA smear can clearly be observed in this visualization. The "Degraded Nucleic Acid" was of a similar intensity of the two rRNA gene bands. The mRNA smear is of a large variety of transcript size.

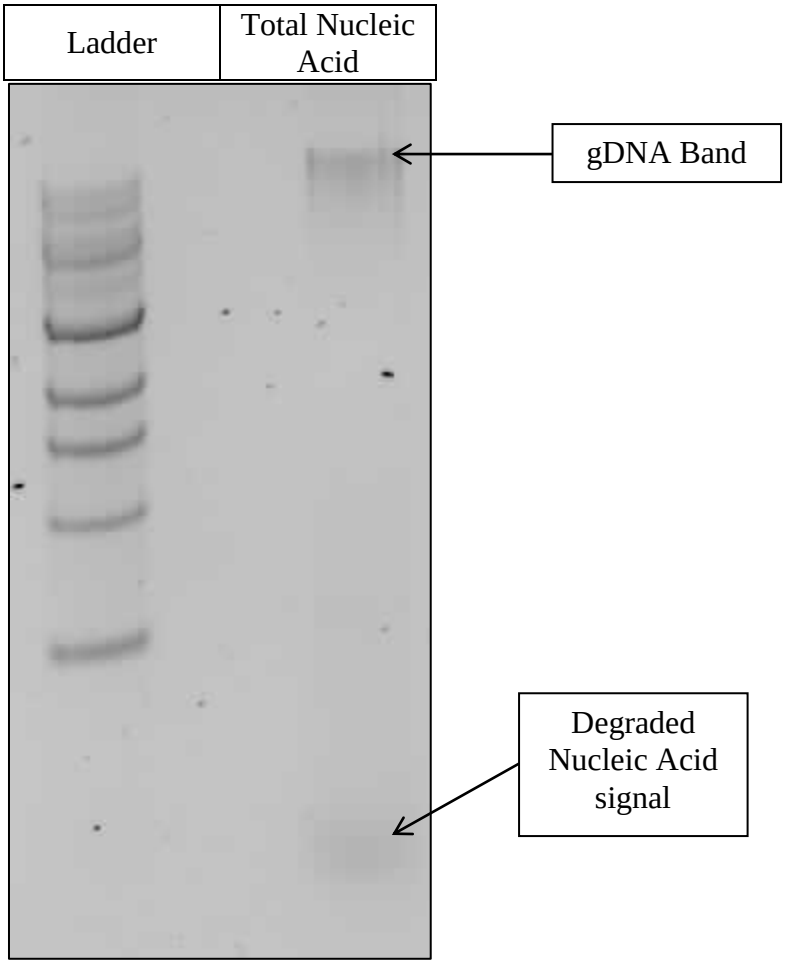


Figure 3.28. Total Nucleic Acid Extraction of the Shredded Whatman Filter Paper Enrichment. A strong gDNA band can be seen as well as a faint degraded signal. rRNA gene bands and mRNA smear are not readily visible in this gel. Due to relative concentrations, it was proposed that the RNA component would still be present, and mRNA enrichment was continued on this sample.

Observation of the shredded Whatman filter paper enrichment extraction (Figure 3.28) looked as though there was only DNA present in the extraction. The sample was subjected to mRNA enrichment because it was postulated the lack of signal was due to the relative concentration of DNA and RNA in the sample.

3.3.2.3 Messenger RNA Enrichment Method Development

Messenger RNA (mRNA) enrichment was conducted as previously described (Section 2.4). Initially, the total mRNA was extracted from *E. coli* (Figure 3.29) to qualify the enrichment procedure before attempting with a sample from the Ngatamariki site (Figure 3.26). The mRNA enrichment proceeds from left to right on Figure 3.29;

i. DNA was degraded via treatment with DNase, which results in a large component of degraded nucleic acid, can be seen at the bottom of the "DNase Treated" lane.

ii. Contamination and short RNA sequences < 200 base pairs (bp) removal can be clearly seen in the "MEGAClear" lane.

iii. The 16S and 23S rRNA gene sequences were removed by subtractive hybridization (by the MICROBExpress Kit) leaving only a smear of mRNA is seen in the final lane ("MICROBExpress" lane).

The resultant sample should include only mRNA transcripts which are free from contamination, DNA, and rRNA sequences. Quantification of the resultant RNA present in the sample after each step can be seen in Table 3.8.

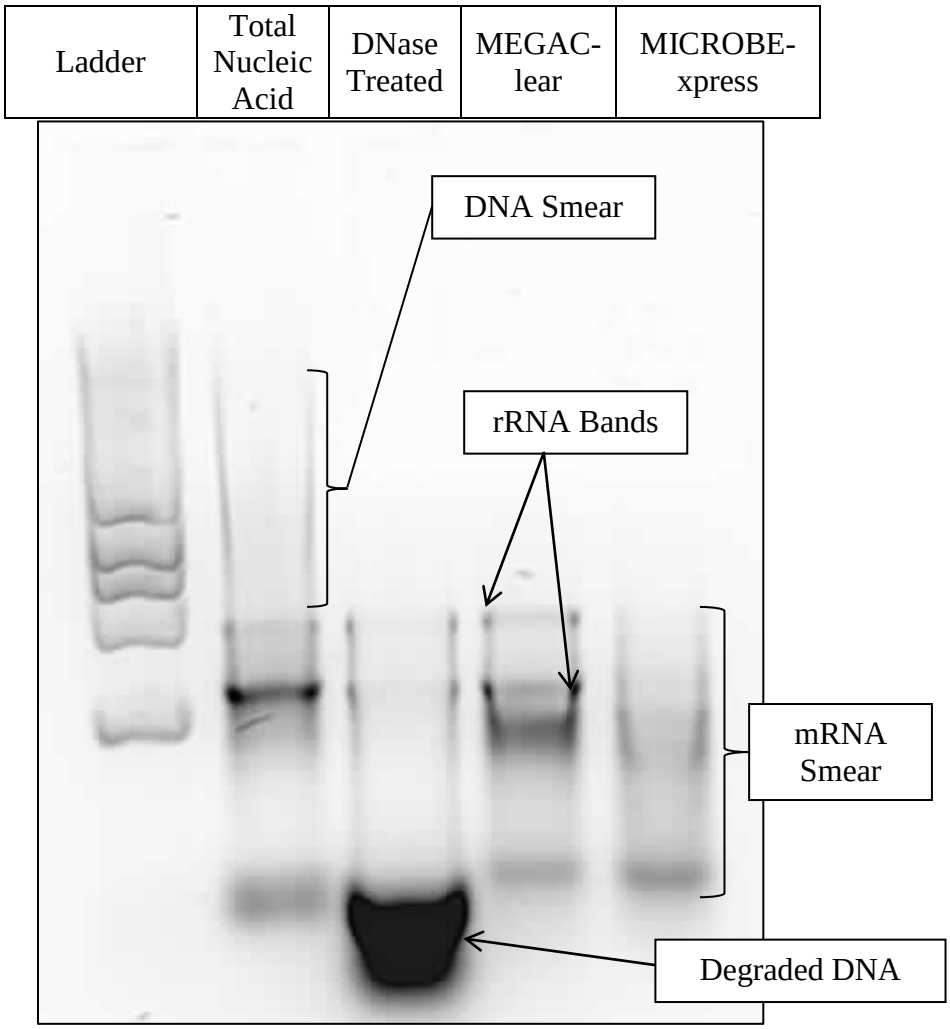


Figure 3.29. Process of Enrichment of mRNA Component of Nucleic Acid Extraction from *E. coli.* Each lane represents the sample after the treatment with the denoted kit. Removal of the DNA Smear results in Degraded DNA. Degraded DNA is removed and then rRNA bands are removed yielding a smear of mRNA transcripts of varying length.

Table 3.8. Quantification at Points Throughout Enrichment of *E. coli* mRNA. As expected, the amount of mRNA from an enrichment was well below the amount of total RNA extracted from the sample. The "A.D.L." (Above Detection Limit) measurement after the DNase Treatment was due to the large amount of degraded nucleic acid present before its removal by the MEGAclear Kit.

Treatment	RNA Amount
Original Extraction	47.15 µg
DNase Treated	A.D.L.
MEGAClear Treated	30.95 µg
MICROBExpress Treated	6 µg

As was anticipated, mRNA made up only a small proportion of total RNA extracted from the cells, accounting for roughly 12 % of total RNA.

3.3.2.3 Results of Enrichment of mRNA from Environmental and Enrichment Samples

Total nucleic acid from the biomass sample (Figure 3.26) was used in mRNA enrichment (Section 2.4) and yielded enriched mRNA transcripts (Figure 3.30). The shredded Whatman filter paper and unbleached kraft samples (Figures 3.27 and 3.28 respectively) were enriched for mRNA in parallel (Section 2.4) and were visualized on the same gel (Figure 3.31).

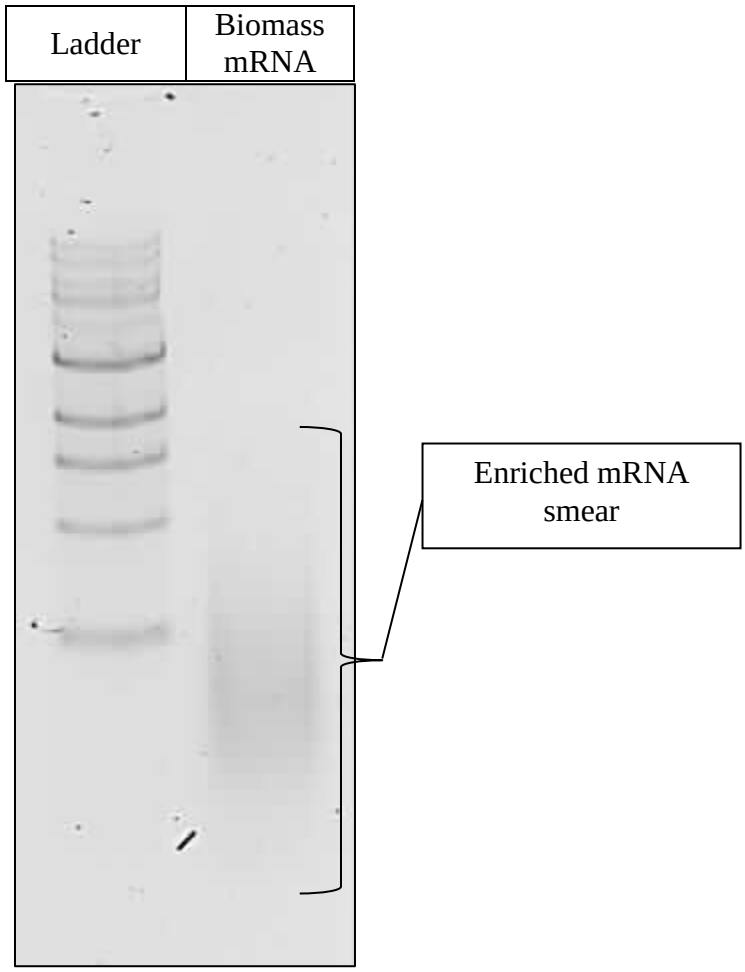


Figure 3.30. Fully Enriched mRNA of the Biomass Sample from Ngatamariki. The removal of DNA, rRNA genes, and the degraded nucleic acid components of the sample can be verified by this gel. The signal generated by the mRNA illustrates mRNA transcripts of varying sequence length.

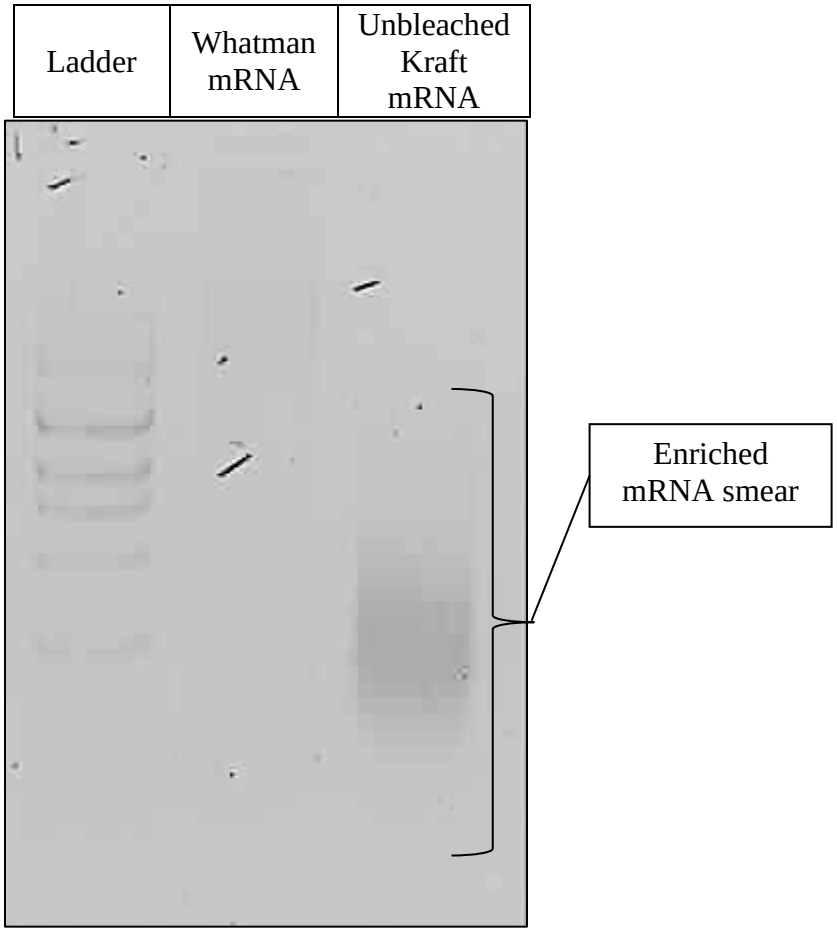


Figure 3.31.Fully Enriched mRNA from the Whatman and Unbleached Kraft Samples Collected at Ngatamariki. The removal of DNA, rRNA genes, and the degraded nucleic acid components of the sample can be verified by this gel. The relative size of the transcripts present in these samples can also be viewed. Most of the transcripts from the shredded Whatman filter paper sample can be seen to be of longer size, indicating they might be of DNA origin.

Although the enrichment of mRNA from the unbleached kraft yielded a smear of mRNA of the appropriate length, the smear present in the shredded Whatman filter paper sample was of a size which would usually correspond to DNA (see Figure 3.31 for comparison). Although this observation was made, the sample was still used for reverse transcription and amplification.

3.3.2.3 Reverse Transcription

DNase and MEGAClear treated total nucleic acid from *E. coli* ("MEGAClear" lane - Figure 3.29) was treated subject to reverse transcription (SuperScriptIII Kit) and RNase A treatment. Successful cDNA library generation was checked via Polymerase Chain Reaction (PCR) targeting the 16S rRNA gene sequence of the library generated from an *E. coli* culture. Universal bacterial 16S rRNA gene primers (9F and 1492R [72]) were used to target DNA. Positive reactions in both the 1 and 2 µl reverse transcription samples confirm successful cDNA creation from RNA sequences (Figure 3.32).

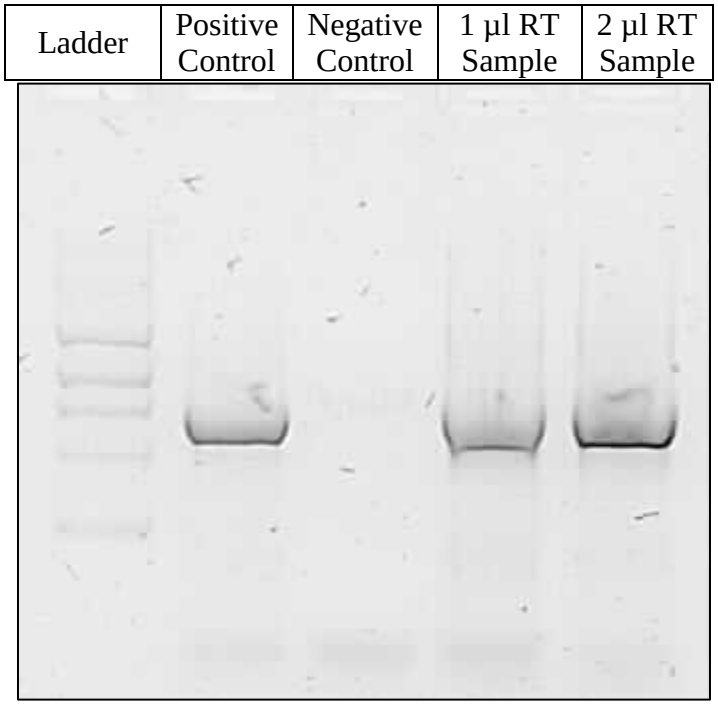


Figure 3.32. PCR Verification of Reverse Transcription Using Total RNA from *E. coli*. Both 1 µl and 2 µl yielded a positive reaction when run with standard 16S universal bacterial primers. The bands present in the figure illustrate the amplification of the bacterial 16S ribosomal gene sequences present in the sample.

The biomass, unbleached kraft, and Whatman mRNA enrichments (Figure 3.30 and Figure 3.31 respectively) were reverse transcribed as previously described (Section 2.5). The low concentrations of enriched mRNA from all Ngatamariki samples dictated that the cDNA libraries required amplification prior to pyrosequencing. Amplification was conducted using the GenomePHI Kit (GE Healthcare) using 1 µl of the RT product as template yielding a total of 1.67 µg of cDNA for the biomass sample, 1.15 µg of cDNA for the unbleached kraft sample, and 1.22 µg of cDNA for the shredded Whatman filter paper sample. The cDNA libraries were then pyrosequenced at the Advanced Studies of Genomics, Proteomics, and Bioinformatics (ASGPB) at The University of Hawai‘i at Mānoa.

3.3.2.4 Bioinformatic Analysis

3.3.2.4.1 Assembly

The biomass and unbleached kraft cDNA libraries were pyrosequenced at the ASGPB at the University of Hawai‘i at Mānoa (Section 2.5). Two quarter plate runs were used to sequence the biomass samples and one quarter run for the unbleached kraft sample (Table 3.9). The remaining quarter plate run was designated for the shredded Whatman filter paper sample. Upon library creation and sequencing of this sample, unusable data was generated (S. Hou, personal communications). No further analysis was conducted on the shredded Whatman filter paper sample.

The metrics produced by the Newbler assembly program (Section 2.7) for the biomass samples and the unbleached kraft enrichment sample can be seen in Table 3.10 and 3.11 respectively.

Table 3.9. Files Generated by Sequencing at ASGPB. Detailing the number of reads and number of bases in each file generated by ASGPB.

File Name (.sff)	Number of Bases	Number of Reads
Biomass 1	83,741,854	269,857
Biomass 2	87,744,083	257,948
Unbleached Kraft	105,788,327	290,734

Table 3.10. Biomass General Assembly Metrics. Details generated by the program Newbler when assembling the two output files associated with the biomass sample.

Section	Aspect	Value	Average
runMetrics			
	inputFileNumReads	527,805	
	inputFileNumBases	171,485,937	
	numberSearches	247,836	
	seedHitsFound	126,095,775	
	overlapsFound	19,841,753	15.74%
	overlapsReported	17,575,275	88.58%
	overlapsUsed	1,099,806	6.26%
readAlignmentResults			
	HDCY7HI01		
	numAlignedReads	190,943	70.76%
	numAlignedBases	58,628,192	70.17%
	inferredReadError	140,5713	2.40%
	HCLGKXD04		
	numAlignedReads	182,316	70.68%
	numAlignedBases	61223599	69.96%
	inferredReadError	1341637	2.19%
consensusResults			
	numAlignedReads	373,259	70.72%
	numAlignedBases	119,851,791	70.06%
	inferredReadError	2,747,350	2.29%
	numberOfIsogroups	5,658	
	avgContigCnt	2.3	
	largestContigCnt	430	

Table 3.10. (Continued) Biomass General Assembly Metrics.

Section	Aspect	Value	Average
consensusResults			
	numberWithOneContig	4,600	
	avgIsotigCnt	1.6	
	largestIsotigCnt	94	
	numberWithOneIsotig	4,635	
	numberOfIsotigs	9,035	
	avgContigCnt	3.4	
	largestContigCnt	18	
	numberOfBases	20,946,550	
	avgIsotigSize	2,318	
	N50IsotigSize	5,797	
	largestIsotigSize	16,051	
	numberOfContigs	5,393	
	numberOfBases	5,924,845	
	avgContigSize	1,098	
	N50ContigSize	1,184	
	largestContigSize	12,806	
	Q40PlusBases	5,436,964	91.77%
	Q39MinusBases	487,881	8.23%
	numberOfContigs	13,265	
	numberOfBases	7,659,797	

Table 3.11. Unbleached Kraft General Assembly. Details generated by the program Newbler when assembling the output file associated with the unbleached kraft sample.

Section	Aspect	Value	Average
runMetrics			
	inputFileNumReads	290,734	
	inputFileNumBases	105,788,327	
	numberSearches	220,510	
	seedHitsFound	43,521,816	
	overlapsFound	3,993,649	9.18%
	overlapsReported	3,643,722	91.24%
	overlapsUsed	406,870	11.17%
readAlignmentResults			
	HIPM5R203		
	numAlignedReads	167,576	57.64%
	numAlignedBases	58,047,057	55.01%
	inferredReadError	1308,866	2.25%

Table 3.11. (Continued) Unbleached Kraft General Assembly.

Section	Aspect	Value	Average
consensusResults			
	numAlignedReads	167,576	57.64%
	numAlignedBases	58,047,057	55.01%
	inferredReadError	1,308,866	2.25%
	numberOfIsogroups	4,693	
	avgContigCnt	2.8	
	largestContigCnt	219	
	numberWithOneContig	3579	
	avgIsotigCnt	1.9	
	largestIsotigCnt	97	
	numberWithOneIsotig	3,588	
	numberOfIsotigs	9,099	
	avgContigCnt	4	
	largestContigCnt	19	
	numberWithOneContig	3,579	
	numberOfBases	23,170,306	
	avgIsotigSize	2,546	
	N50IsotigSize	5,347	
	largestIsotigSize	18,448	
	numberOfContigs	5,677	
	numberOfBases	6,540,763	
	avgContigSize	1,152	
	N50ContigSize	1,255	
	largestContigSize	12,573	
	Q40PlusBases	6,082,462	92.99%
	Q39MinusBases	458,301	7.01%
	numberOfContigs	12,953	
	numberOfBases	8,274,828	

Since only one file was used in the assembly of the unbleached kraft sample, the average for the values of "numAlignedReads" and "numAlignedBases" are less than in the biomass dataset. If this were an assembly of a genome, it would be expected that the value be much closer to 100 % (especially with higher coverage) but this value is good for transcriptomic data. An alternative indicator to a quality sequencing run is the "Q40PlusBases" metric. Since both assemblies had Q40 values in the 90 % range, it was

assumed that the data generated was of high quality negating the need to examine the transcripts individually.

3.3.2.4.2 Annotation

Annotation was accomplished (Section 2.7) and a list of putative GH proteins generated from the BLASTp output file was linked to the original nucleotide sequence of the candidate amino acid sequences (schematic of this process can be found in Figure 3.33). InterProScan v4.8 [107], BLASTx [3], and associated Conserved Domain Database [84] was used to identify the most current prediction of the gene (the initial search was conducted using a standalone database downloaded July 2011), confirm functional domains, to assign closest taxonomy to the protein, and to trim the sequence to only the protein of interest. Substrate specificity of the candidate GHs identified using the CAZy GH database [25]. Not all of the putative GH sequences appeared to be complete transcripts when compared to the most closely related GH.

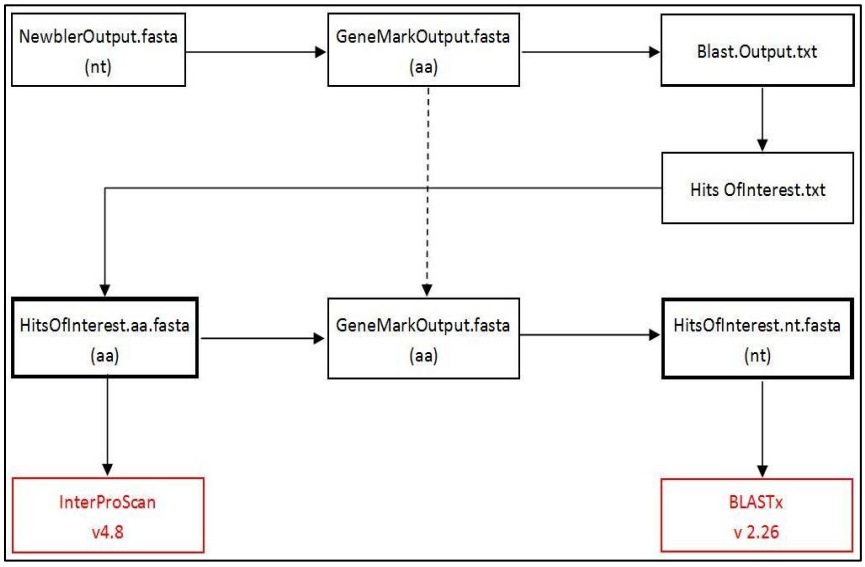
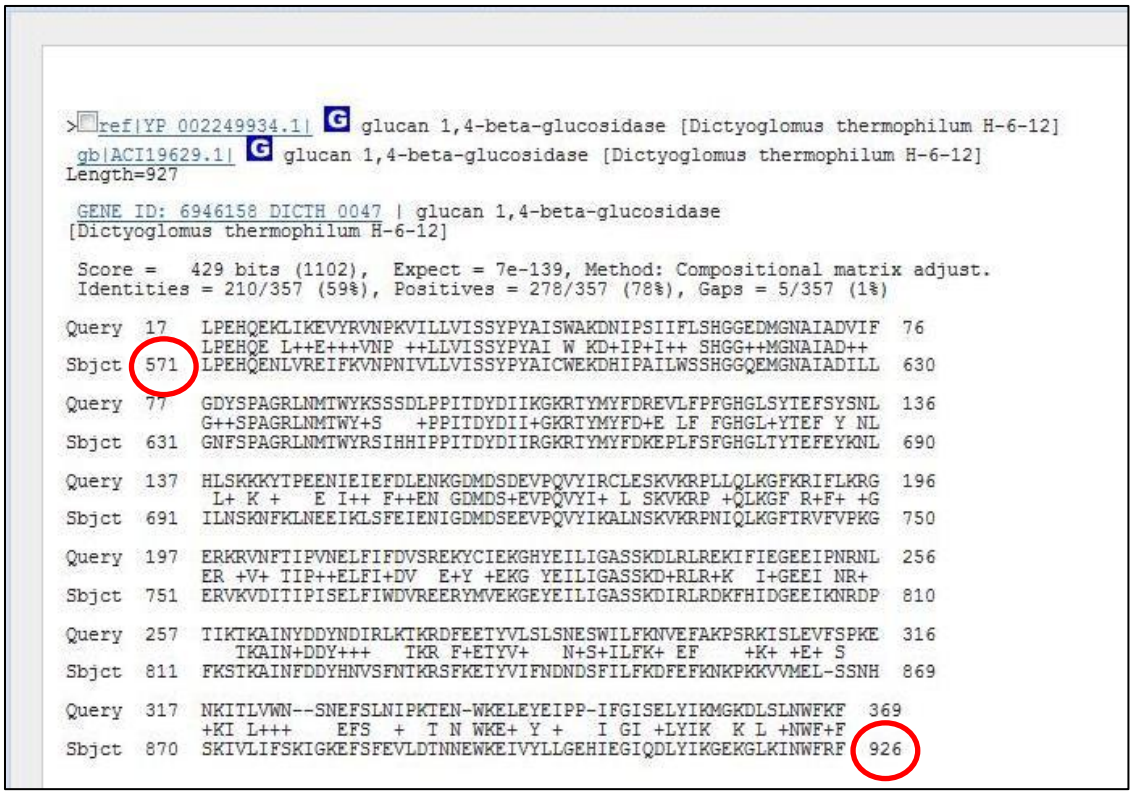


Figure 3.33. Schematic of Hits of Interest. This diagram shows the pipeline used for function assignment and the subsequent study of the sequences which had some function associated with GHs.

Tables 3.12 and 3.14 lists full length sequences of the biomass and the unbleached kraft samples respectively. However, it should be noted that while the truncated GHs for the biomass and unbleached kraft samples have been listed "non-full length transcripts" in tables 3.13 and 3.15 respectively, these predictions are based solely on a pairwise comparison with the closest relative. It is entirely conceivable that the truncated candidate GH genes may not be truncated and could retain GH activity.

In several cases, non-full length mRNA transcripts were annotated as sequential genes (e.g. gene_ 566 & gene_567, Table 3.14). These transcripts required manual assembly and were examined in the assembly program Newbler. Any regions of suspect (repeats or areas of low read depth – denoted by a lowercase letter in the assembled sequence (Appendix B & C)) were examined with the highest BLASTx hit as a reference. This allowed for some sequences, denoted by a "*" (Table 3.12 through Table 1.15), to be concatenated into a longer, though not always full length, transcripts (e.g. gene_29901 was originally annotated as gene_29901 and gene_29902 (Figure 3.34 through 3.37)). Initially, BLASTp [3] was used to annotate gene_29901 (Figure 3.34) and gene_29902 (Figure 3.35) individually. The top hit for each was the same enzyme and organism: ref|YP_002249934.1| glucan 1,4-beta-glucosidase [*Dictyoglomus thermophilum* H-6-12].



>ref|YP_002249934.1| glucan 1,4-beta-glucosidase [Dictyoglomus thermophilum H-6-12]
gb|ACI19629.1| glucan 1,4-beta-glucosidase [Dictyoglomus thermophilum H-6-12]
Length=927

GENE_ID: 6946158 DICTH_0047 | glucan 1,4-beta-glucosidase
[Dictyoglomus thermophilum H-6-12]

Score = 429 bits (1102), Expect = 7e-139, Method: Compositional matrix adjust.
Identities = 210/357 (59%), Positives = 278/357 (78%), Gaps = 5/357 (1%)

```
Query  17   LPEHQEKLIKEVYRVNPKVILLVISSYPYAISWAKDNIPSIIFLSHGGEDMGNAIADVIF  76
            LPEHQE L++E+++VNP ++LLVISSYPYAI W KD+IP+I++ SHGG++MGNAIAD++
Sbjct  571  LPEHQENLVREIFKVNPNIVLLVISSYPYAICWEKDHIPAILWSSHGGQEMGNAIADILL  630

Query  77   GDYSPAGRLNMTWYKSSSDLPPITDYDIIKGKRTYMYFDREVLFPFGHGLSYTEFSYSNL  136
            G++SPAGRLNMTWY+S   +PPITDYDII+GKRTYMYFD+E LF FGHGL+YTEF Y NL
Sbjct  631  GNFSPAGRLNMTWYRSIHHIPPITDYDIIRGKRTYMYFDKEPLFSFGHGLTYTEFEYKNL  690

Query  137  HLSKKKYTPEENIEIEFDLENKGDMDSDEVPQVYIRCLESKVKRPLLQLKGFKRIFLKRG  196
             L+ K +   E I++ F++EN GDMDS+EVPQVYI+ L SKVKRP +QLKGF R+F+ +G
Sbjct  691  ILNSKNFKLNEEIKLSFEIENIGDMDSEEVPQVYIKALNSKVKRPNIQLKGFTRVFVPKG  750

Query  197  ERKRVNFTIPVNELFIFDVSREKYCIEKGHYEILIGASSKDLRLREKIFIEGEEIPNRNL  256
            ER +V+ TIP++ELFI+DV  E+Y +EKG YEILIGASSKD+RLR+K  I+GEEI NR+
Sbjct  751  ERVKVDITIPISELFIWDVREERYMVEKGEYEILIGASSKDIRLRDKFHIDGEEIKNRDP  810

Query  257  TIKTKAINYDDYNDIRLKTKRDFEETYVLSLSNESWILFKNVEFAKPSRKISLEVFSPKE  316
               TKAIN+DDY+++   TKR F+ETYV+   N+S+ILFK+ EF    +K+ +E+ S
Sbjct  811  FKSTKAINFDDYHNVSFNTKRSFKETYVIFNDNDSFILFKDFEFKNKPKKVVMEL-SSNH  869

Query  317  NKITLVWN--SNEFSLNIPKTEN-WKELEYEIPP-IFGISELYIKMGKDLSLNWFKF  369
            +KI L+++    EFS  +  T N WKE+ Y +   I GI +LYIK  K L +NWF+F
Sbjct  870  SKIVLIFSKIGKEFSFEVLDTNNEWKEIVYLLGEHIEGIQDLYIKGEKGLKINWFRF  926
```

Figure 3.34. BLASTp Screen Capture of Top Hit Generated by Gene_29901. The homology corresponds to position 571 - 926 of the nucleotide sequence of the gene from *Dictyoglomus*. The red circles indicate the nucleotide position at which the subject gene (from *Dictyoglomus*) began (571) and ended (926).

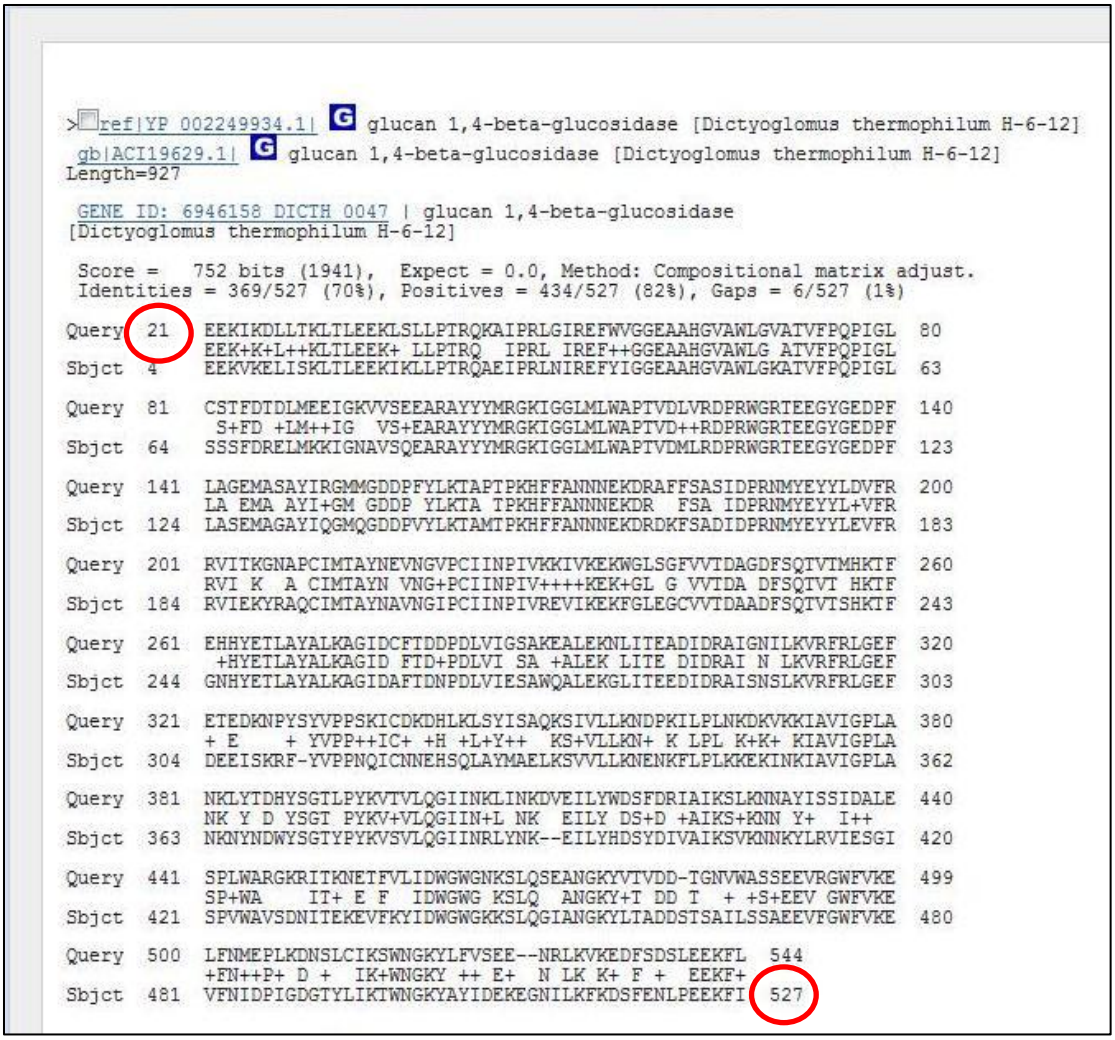


Figure 3.35. BLASTp Screen Capture of Top Hit Generated by Gene_29902. The homology corresponds to position 4 – 527 of the nucleotide sequence of the gene from *Dictyoglomus*. The red circles indicate the nucleotide position at which the subject gene (from *Dictyoglomus*) began (4) and ended (527).

The generation of the same hit from the same organism was reason enough to investigate the original nucleotide sequence generated by the assembly. In addition, the two genes covered different regions of the top hit from positions 527 - 571 in the nucleotide sequence of the gene from *Dictyoglomus*. As anticipated, both transcripts originated from the same contiguous nucleotide sequence (isotig03536). The BLASTx result of isotig03536 (Figure 3.36) confirmed the hypothesis.

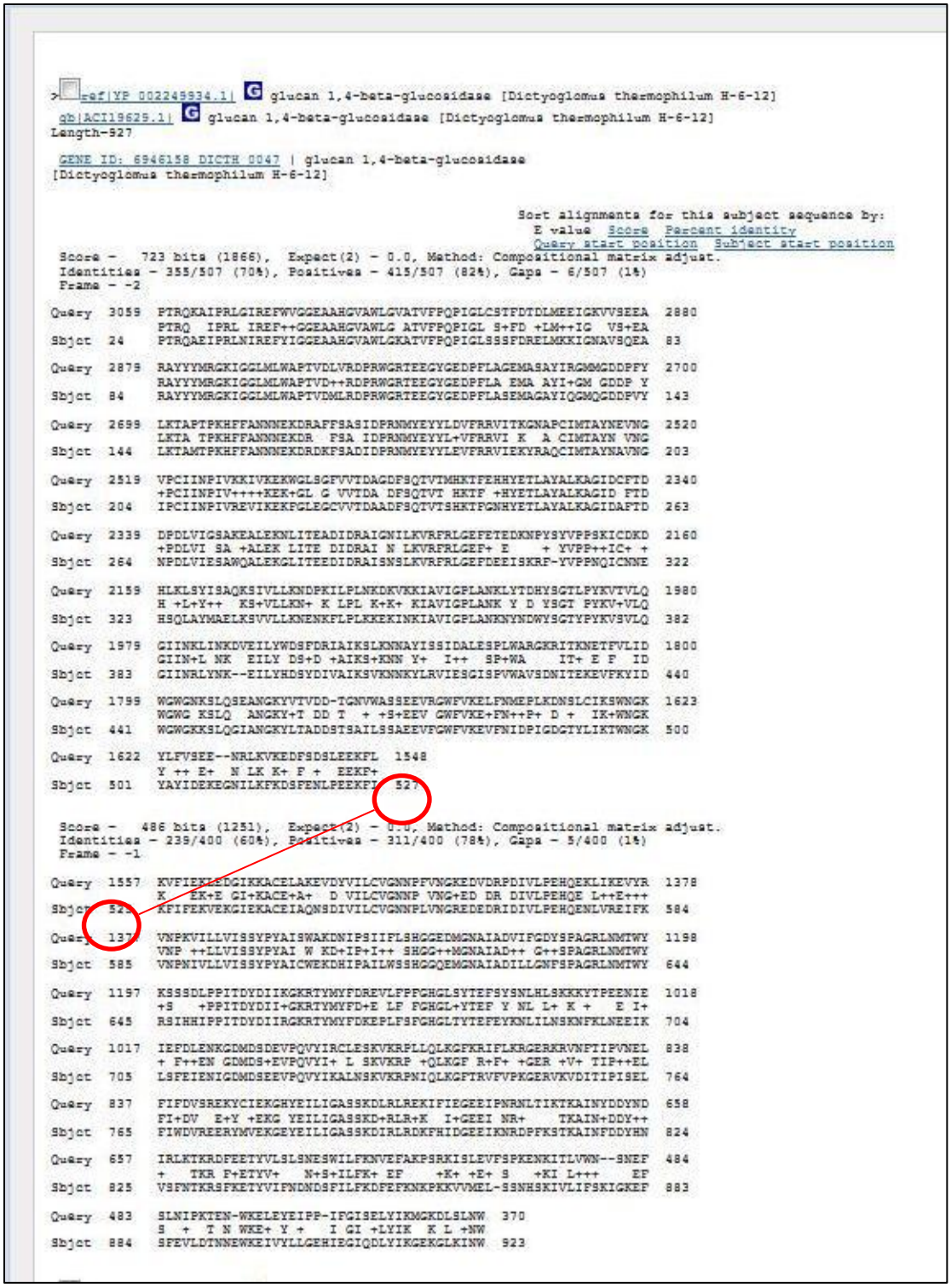


Figure 3.36. BLASTx Screen Capture of Top Hit Generated by Isotig03536. The entire transcript (isotig03536) codes for the same GH in *Dictyoglomus*. The problem in the assembly was anticipated to be around nucleotide 525 – 527. The red circles indicate the nucleotide position at which the subject gene (from *Dictyoglomus*) terminated in the first reading frame (527) and began in the second reading frame (525).

When the nucleotide sequence was investigated more thoroughly, a reading frame issue was determined to be the cause of the separation of related genes in the transcript (Figure 3.37). This was corrected manually as described above.

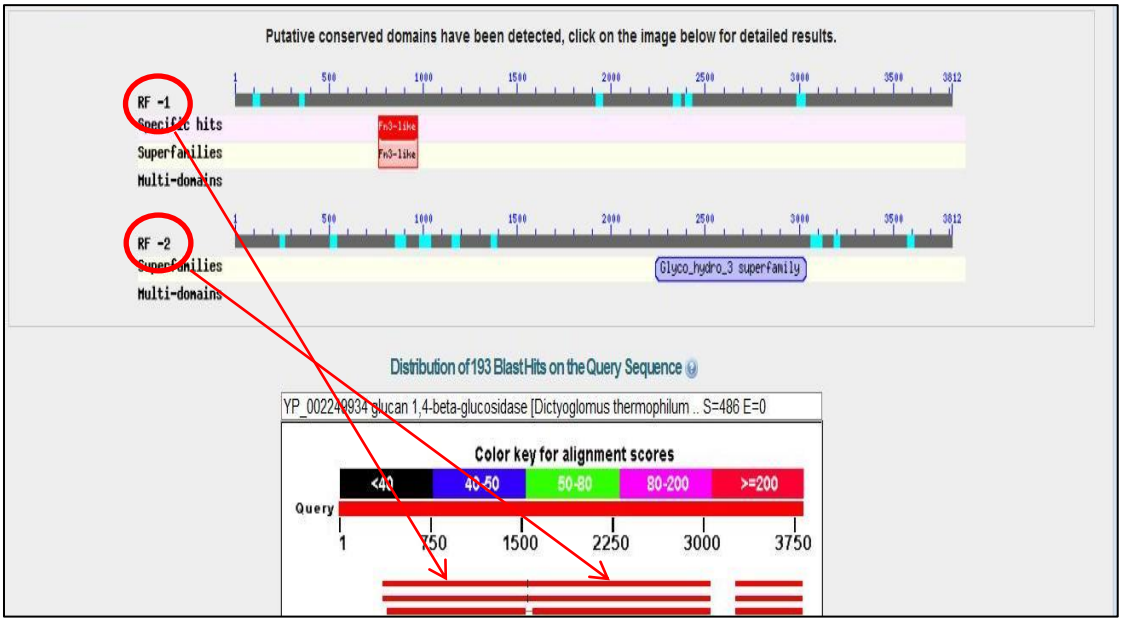


Figure 3.37. Screen Capture of Graphic Summary of BLASTx and Functional Domain Identification [84] of Isotig03536. Different reading frames (RF) can be observed as the reason for the problem in annotation of the transcripts. RF; reading frame number. The red circles indicate the reading frames for the corresponding sections of the gene from *Dictyoglomus*.

After manual assembly, BLASTx was used to verify the result (Figure 3.38 and 3.39).

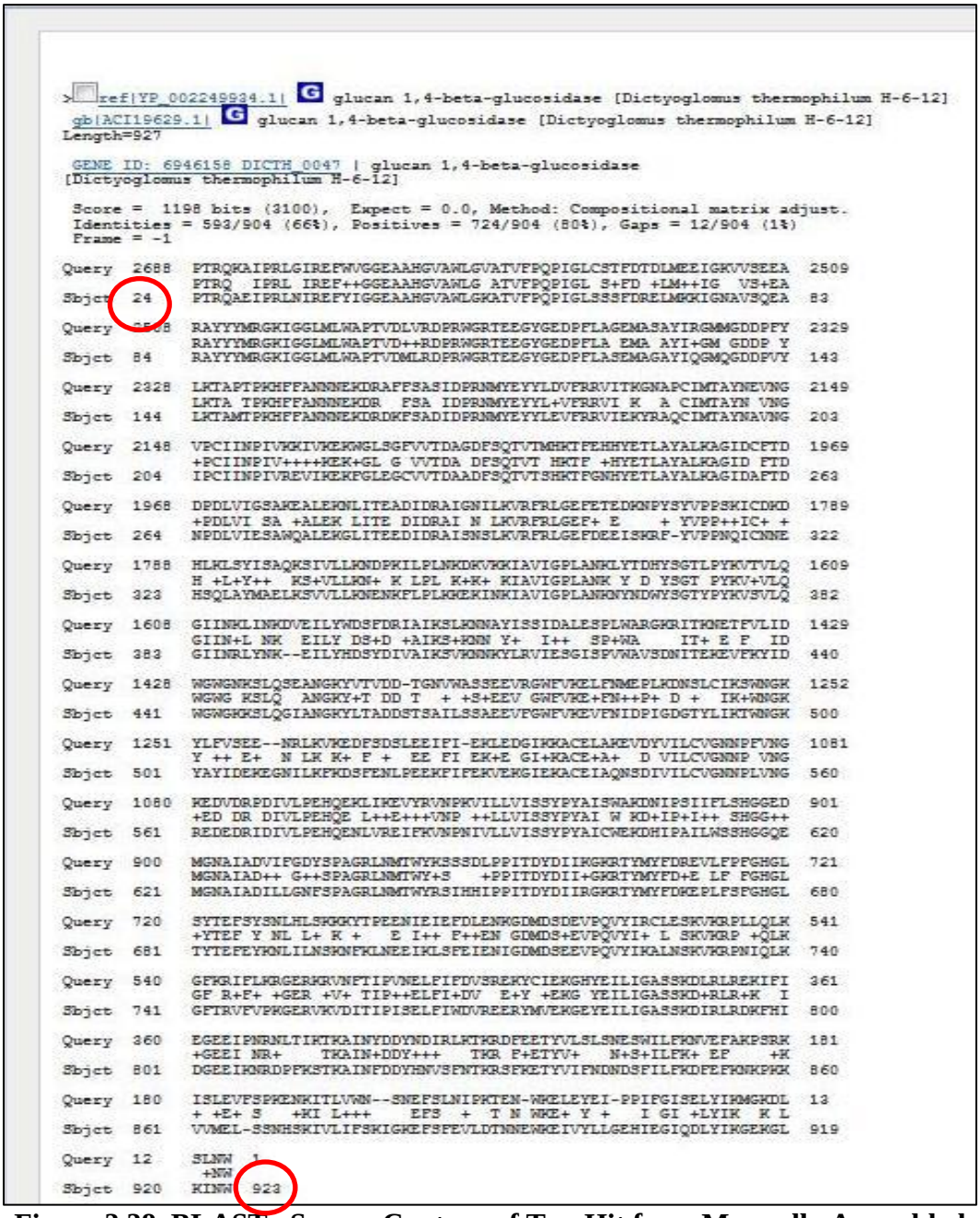


Figure 3.38. BLASTx Screen Capture of Top Hit from Manually Assembled Gene_29901 (Isotig03536). This shows a BLASTx hit covering most of the sequence of the top hit. The red circles indicate the nucleotide positions at which the subject gene (from *Dictyoglomus*) began (24) and ended (922). This indicates almost complete coverage of the 927 nucleotide (nt) long gene.

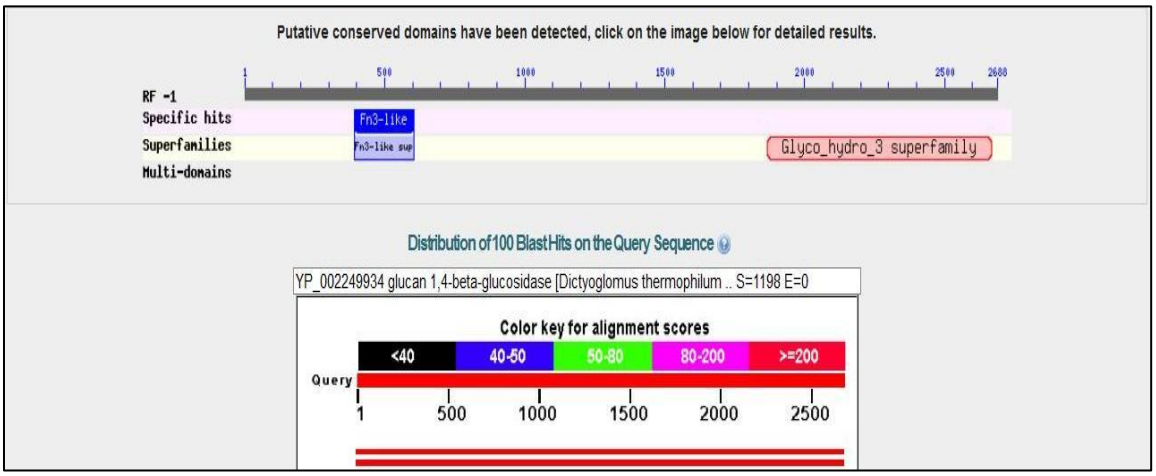


Figure 3.39. Screen Capture of Graphic Summary of BLASTx and Functional Domain Identification [84] of Manually Assembled Gene_29901 (Isotig03536). Showing only one reading frame, the transcript codes for the same protein as the original two transcripts (Figure 3.34 and Figure 3.35) in one complete transcript.

Although all sequentially annotated genes were examined, some could not be joined. These were annotated as sequential genes (denoted by a "**" in Table 3.12 through Table 3.15), and showed enough difference to justify their annotation as separate transcripts. This evidence was either a separation in coding regions of >100 bp or the annotation of a different GH family or top BLAST hit. Annotation of a single transcript as sequential genes was probably due to fragmentation during the extraction process or low read depth generated by the library creation (Section 2.5). Concatenation and BLAST analysis of these candidate genes confirmed that these originated from single transcripts with either sequence deletions or inserts forcing the annotation of truncated transcripts. After the manual assembly of transcripts a final list of GHs was achieved (Table 3.12 through Table 3.15).

Table 3.12. Full Length GHs from Biomass Sample. The length of the protein is in amino acid (aa) residues. Also included is the E-Values generated by BLASTx and the closest related organism from the top hit. The function was predicted from the information generated by BLASTp and confirmed by InterProScan and BLASTx. CBM refers to a Carbohydrate-Binding Module [25]. The numbers in the brackets (e.g. 3.2.1.25) indicate the Enzyme Commission number (EC number) [144] corresponding to the top BLAST hit.

ID	Length (aa)	GH Family	Known or Predicted Function	BLAST E-value	Organism
gene_337	438	GH77	4-alpha-glucotransferase (3.2.1.25)	5.00E-165	*Hydrogenobacter thermophilus* TK-6
gene_34538	345	GH5	Endoglucanase (3.2.1.4)	2.00E-152	Uncultured candidate division OP1 bacterium
gene_36258	844	GH38	Alpha-mannosidase (3.2.1.24)	1.00E-166	*Rubrobacter xylanophilus* DSM 9941

Table 3.13. Non-Full Length GHs from the Biomass Sample. The length of the protein is in amino acid (aa)residues. Also included is the E-Values generated by BLASTx and the closest related organism from the top hit. The function was predicted from the information generated by BLASTp and confirmed by InterProScan and BLASTx. CBM refers to a Carbohydrate-Binding Module [25]. The numbers in the brackets (e.g. 3.2.1.1) indicate the Enzyme Commission number (EC number) [144] corresponding to the top BLAST hit.

ID	Length (aa)	GH Family	Known or Predicted Function	BLAST E-value	Organism
gene_3317	216	GH57	Alpha-amylase (3.2.1.1)	5.00E-100	*Hydrogenobacter thermophilus* TK-6
gene_36543	130	GH4	Alpha-galactosidase (3.2.1.20)	5.00E-45	Uncultured archaeon
gene_37359	128	GH38	Alpha-mannosidase (3.2.1.24)	9.00E-08	*Archaeoglobus veneficus* SNP6
gene_37914	232	GH77	4-alpha-glucotransferase (3.2.1.25)	1.00E-89	*Thermus brockianus*
gene_38546	291	GH15	Glucan 1,4-alpha-glucosidase (EC 3.2.1.3)	3.00E-121	*Sulfurihydrogenibium azorense*
gene_40561	180	GH13	Alpha-amylase (3.2.1.1)	2.00E-56	*Deinococcus geothermalis* DSM 11300
gene_40746	184	GH57	Amylopullulanase (3.2.1.41)	1.00E-12	*Desulfurispirillum indicum* S5
gene_41054	99	GH13	Pullulanase (3.2.1.41)	3.00E-12	Uncultured candidate division OP1 bacterium
gene_41570	548	GH57	Alpha-amylase (3.2.1.1)	4.00E-88	*Marinitoga piezophila* KA3

Table 3.13. (Continued) Non-Full Length GHs from the Biomass Sample.

ID	Length (aa)	GH Family	Known or Predicted Function	BLAST E-value	Organism
gene_42832	173	GH13	Alpha-amylase (3.2.1.1)	5.00E-10	*Herpetosiphon aurantiacus* DSM 785
gene_43551	162	GH13:: CBM20	Neopullulanase (3.2.1.135)	8.00E-13	*Caloramator australicus* RC3
gene_43916	153	GH16	Glucan endo-1,3-beta-D-glucosidase (3.2.1.39)	2.00E-104	*Fervidobacterium pennivorans* DSM 9078
gene_44066**	191	GH57	Alpha-amylase (3.2.1.1)	1.00E-14	*Granulicella tundricola* MP5ACTX9
gene_44068**	279	GH78	Alpha-L-rhamnosidase (3.2.1.40)	1.00E-37	*Opitutaceae bacterium* TAV5
gene_44261	226	GH1	Beta-glucosidase (3.2.1.21)	2.00E-159	*Fervidobacterium* sp. YNP
gene_44279	234	GH2	Family 2 sugar binding domain	3.00E-53	*Pedobacter heparinus* DSM 2366
gene_45023	221	GH4	6-phospho-beta-glucosidase (3.2.1.86)	4.00E-46	*Phyllobacterium* sp. YR531
gene_45198	229	GH31	Alpha-galactosidase (3.2.1.22)	1.00E-65	*Thermobaculum terrenum* ATCC BAA-798
gene_45258	234	GH4	Alpha-glucosidase (3.2.1.20)	6.00E-146	*Thermotoga lettingae* TMO
gene_45382	230	GH65	Kojibiose phosphorylase (2.4.1.230)	7.00E-16	*Chloroflexus aurantiacus* J-10-fl

Table 3.13. (Continued) Non-Full Length GHs from the Biomass Sample.

ID	Length (aa)	GH Family	Known or Predicted Function	BLAST E-value	Organism
gene_45450	219	GH57	Amylopullulanase (3.2.1.41)	5.00E-45	*Coprothermobacter proteolyticus* DSM 5265
gene_45820	144	GH31	Alpha-glucosidase (3.2.1.20)	7.00E-13	*Candidatus* "Koribacter versatilis" Ellin345
gene_45835	211	GH2	Putative beta-glucosidase (3.2.1.21)	7.00E-66	*Microcoleus vaginatus* FGP-2
gene_45889	159	GH4	Alpha-galactosidase (3.2.1.22)	4.00E-63	*Thermoanaerobacter tengcongensis* MB4
gene_46159	41	GH38	Alpha-mannosidase (3.2.1.24)	2.00E-08	*Salinibacterium* sp. PAMC 21357
gene_46259	132	GH67	Alpha-glucouronidase (3.2.1.139)	1.00E-05	*Pedobacter saltans* DSM 12145
gene_46489	130	GH2	Putative beta-glucosidase (3.2.1.21)	8.00E-30	*Oscillatoria* sp. PCC 6506
gene_46625	154	GH13	Isoamylase (3.2.1.68)	1.00E-26	*Fervidobacterium nodosum* Rt17-B1
gene_46781	172	GH38	Alpha-mannosidase (3.2.1.24)	7.00E-53	*Dictyoglomus turgidum* DSM 6724
gene_46988	61	GH48	Endo-beta-1,4-glucanase (3.2.1.4)	8.00E-10	*Bacteroides cellulosolvens* ATCC 35603

Table 3.13. (Continued) Non-Full Length GHs from the Biomass Sample.

ID	Length (aa)	GH Family	Known or Predicted Function	BLAST E-value	Organism
gene_47005**	88	GH1	Beta-glucosidase (3.2.1.21)	1.00E-16	*Anaerolinea thermophilia* UNI-1
gene_47009**	178	GH31	Alpha-glucosidase (3.2.1.20)	3.00E-87	*Dictyoglomus turgidum* DSM 6724
gene_47799	115	GH1	Beta-glucosidase (3.2.1.21)	6.00E-59	*Caldilinea aerophila* DSM 14535 = NBRC 104270
gene_47977	146	GH28	Endo-polygalacturonase (3.2.1.15)	8.00E-54	*Caldicellulosiruptor kristjanssonii* 177R1B
gene_48127	66	GH57	Alpha-amylase (3.2.1.1)	5.00E-10	*Spirochaeta thermophila* DSM 6192
gene_48326	141	GH57	Pullulanase (3.2.1.41)	7.00E-86	*Dictyoglomus turgidum* DSM 6724
gene_48477	136	GH31	Alpha-xylosidase Yicl (3.2.1.177)	7.00E-74	*Dictyoglomus thermophilum* H-6-12
gene_48822	113	GH65	Kojibiose phosphorylase (2.4.1.230)	3.00E-36	*Chloroflexus aggregans* DSM 9485

** - Transcripts which were annotated as sequential genes yet still retained individual identity due to supporting evidence of difference according to spatial separation, GH families, predicted function, or related organism.

Table 3.14. Full Length GHs from Unbleached Kraft Sample. The length of the protein is in amino acid (aa) residues. Also included is the E-Values generated by BLASTx and the closest related organism from the top hit. The function was predicted from the information generated by BLASTp and confirmed by InterProScan and BLASTx. CBM refers to a Carbohydrate-Binding Module [25]. The numbers in the brackets (e.g. 3.2.1.176) indicate the Enzyme Commission number (EC number) [144] corresponding to the top BLAST hit.

ID	Length (aa)	GH Family	Known or Predicted Function	BLAST E-value	Organism
gene_15250*	809	GH48 :: CBM3 :: CBM3	Cellulose 1,4-beta-cellobiosidase (3.2.1.176)	0.00E+00	*Cauldicellulosiruptor lactoaceticus* 6A
gene_28236	336	GH31	Endoglucanase (3.2.1.4)	0.00E+00	*Dictyoglomus thermophilum* H-6-12
gene_28379	472	GH05 :: CBMX2	Cellulase (3.2.1.4)	0.00E+00	*Dictyoglomus turgidum* DSM 6724
gene_28797	567	GH03	Beta-D-glucosidase (3.2.1.21) / Beta-N-acetylhexosaminidase (3.2.1.52)	0.00E+00	*Dictyoglomus thermophilum* H-6-12
gene_28824	811	GH02	Beta-mannosidase (3.2.1.25)	0.00E+00	*Dictyoglomus thermophilum* H-6-12
gene_28956**	517	GH57	Alpha-amylase (3.2.1.1)	0.00E+00	*Dictyoglomus thermophilum* H-6-12
gene_28957**	685	GH57	Alpha-amylase (3.2.1.1)	0.00E+00	*Dictyoglomus thermophilum* H-6-12
gene_29370*	700	Putative Glycosidase	Glycosidase PH1107-related	4.00E-176	*Dictyoglomus thermophilum* H-6-12

Table 3.14. (Continued) Full Length GHs from Unbleached Kraft Sample.

ID	Length (aa)	GH Family	Known or Predicted Function	BLAST E-value	Organism
gene_29387	570	GH20	Beta-hexosaminidase (3.2.1.52)	0.00E+00	*Carboxydibrachium pacificum* DSM 12653
gene_29901*	899	GH03	Glucan 1,4-beta-glucosidase (3.2.1.74)	0.00E+00	*Dictyoglomus thermophilum* H-6-12
gene_30428**	288	GH12	Endoglucanase (3.2.1.4)	7.00E-124	*Dictyoglomus turgidum* DSM 6724
gene_30429**	758	GH31	Alpha-glucosidase (3.2.1.20)	0.00E+00	*Dictyoglomus turgidum* DSM 6724
gene_30476	418	GH01	Beta-glucosidase (3.2.1.21)	0.00E+00	*Dictyoglomus turgidum* DSM 6724
gene_30590	452	GH04	Alpha -glucosidase (3.2.1.20)	0.00E+00	*Dictyoglomus turgidum* DSM 6724
gene_31557**	360	GH10	Beta-1,4-xylanase (3.2.1.8)	0.00E+00	Uncultured bacterium
gene_31560**	751	GH31	Alpha-glucosidase (3.2.1.20)	0.00E+00	*Thermoanaerobacter mathranii* subsp. *Mathranii* str. A3
gene_32645	400	GH01	Beta-galactosidase (3.2.1.23)	8.00E-180	*Dictyoglomus thermophilum* H-6-12

Table 3.14. (Continued) Full Length GHs from Unbleached Kraft Sample.

ID	Length (aa)	GH Family	Known or Predicted Function	BLAST E-value	Organism
gene_33199	321	GH04	Alpha-glucosidase (3.2.1.20)	2.00E-169	*Thermotoga thermarum* DSM 5069
gene_34397	466	GH04	Alpha-glucosidase (3.2.1.20)	0.00E+00	*Thermotoga thermarum* DSM 5069
gene_34635	416	GH04	6-phospho-beta-glucosidase (3.2.1.86)	0.00E+00	*Dictyoglomus thermophilum* H-6-12
gene_34676	287	GH05	Cellulase (3.2.1.4)	6.00E-144	*Dictyoglomus thermophilum* H-6-12
gene_34865	896	GH38	Alpha-mannosidase (3.2.1.24)	5.00E-134	*Rubrobacter xylanophilus* DSM 9941
gene_36826	567	GH20	N-acetyl-beta-hexosaminidase (3.2.1.52)	0.00E+00	*Dictyoglomus turgidum* DSM 6724
gene_36936	305	GH32	Beta-frucosidase (3.2.1.153)	5.00E-110	*Marinithermus hydrothermalis* DSM 14884
gene_37408	326	GH10	Beta-1,4-xylanase (3.2.1.8)	0.00E+00	Uncultured bacterium
gene_37470	525	GH05	Cellulase (3.2.1.4)	0.00E+00	*Dictyoglomus thermophilum* H-6-12

Table 3.14. (Continued) Full Length GHs from Unbleached Kraft Sample.

ID	Length (aa)	GH Family	Known or Predicted Function	BLAST E-value	Organism
gene_37548	566	GH16	Laminarinase (3.2.1.6)	0.00E+00	*Dictyoglomus thermophilum* H-6-12
gene_566**	422	GH04	Alpha-galactosidase (3.2.1.10)	7.00E-162	*Thermoanaerobacter pseudethanolicus* ATCC 33223
gene_567**	326	Putative Glycosidase	Putative glycosylase	8.00E-161	*Mahella australiensis* 50-1 BON
gene_new	686	GH13	Alpha-amylase (3.2.1.1)	0.00E+00	*Dictyoglomus thermophilum* H-6-12

* - Transcripts which were originally annotated as separate genes and after manual assembly were able to be annotated as single transcripts

** - Transcripts which were annotated as sequential genes yet still retained individual identity due to supporting evidence of difference according to spatial separation, GH families, predicted function, or related organism.

Table 3.15. Non-Full Length Unbleached Kraft GHs. The length of the protein is in amino acid (aa) residues. Also included is the E-Values generated by BLASTx and the closest related organism from the top hit. The function was predicted from the information generated by BLASTp and confirmed by InterProScan and BLASTx. CBM refers to a Carbohydrate-Binding Module [25]. The numbers in the brackets (e.g. 3.2.1.41) indicate the Enzyme Commission number (EC number) [144] corresponding to the top BLAST hit.

ID	Length (aa)	GH Family	Known or Predicted Function	BLAST E-value	Organism
gene_13844	196	GH57	Amylopullanase (3.2.1.41)	1.00E-36	*Coprothermobacter proteolyticus* DSM 5265
gene_14665	383	GH57	Alpha-amylase (3.2.1.1)	3.00E-53	*Thermus aquaticus* Y51MC23
gene_15474	846	GH10 :: GH44 :: CBM3	CelE [Endoglucanase (3.2.1.4)]	0.00E+00	*Caldicellulosiruptor* sp. Tok7B.1
gene_15773	285	GH10	Endo-beta-1,4-xylanase (3.2.1.8)	2.00E-130	*Spirochaeta thermophila* DSM 6578
gene_23927	598	GH03	Beta-glucosidase B (3.2.1.21)	0.00E+00	*Dictyoglomus turgidum* DSM 6724
gene_27008	93	GH01	Beta-glucosidase A (3.2.1.21)	1.00E-27	*Granulicella mallensis* MP5ACTX8
gene_27761	517	GH57	Amylopullanase (3.2.1.41)	2.00E-97	*Kosmotoga olearia* TBF 19.5.1
gene_28603	305	GH44	Beta-mannanase (3.2.1.78) / Endoglucanase A (3.2.1.4)	4.00E-143	*Dictyoglomus thermophilum* H-6-12
gene_29847**	235	GH10	Endo-1,4-beta-xylanase (3.2.1.32)	4.00E-111	*Dictyoglomus turgidum* DSM 6724

Table 3.15. (Continued) Non-Full Length Unbleached Kraft GHs.

ID	Length (aa)	GH Family	Known or Predicted Function	BLAST E-value	Organism
gene_29850**	570	GH31	Alpha-glucosidase (3.2.1.20)	0.00E+00	*Dictyoglomus turgidum* DSM 6724
gene_30141*	1031	GH10 :: CBM9 :: CBM9	Endo-beta-1,4-xylanase (3.2.1.8)	0.00E+00	*Caldicellulosiruptor* sp. Rt69B.1
gene_31680	224	GH57	Alpha-amylase (3.2.1.1)	3.00E-63	*Candidatus* "Desulforudis audaxviator" MP104C
gene_31697	198	GH57	Amylopullanase (3.2.1.41)	6.00E-92	*Dictyoglomus turgidum* DSM 6724
gene_32767	99	GH13	Alpha-amylase (3.2.1.1)	2.00E-15	*Kosmotoga olearia* TBF 19.5.1
gene_33001	78	GH38	Alpha-mannosidase (3.2.1.24)	2.00E-25	*Dictyoglomus turgidum* DSM 6724
gene_33327	317	GH02	Beta-galactosidase (3.2.1.23) / Beta-glucuronidase (3.2.1.31)	9.00E-74	*Microvirga* sp. WSM3557
gene_34145	93	GH04	Alpha-glucosidase (3.2.1.20)	2.00E-29	*Spirochaeta thermophila* DSM 6578
gene_34308	118	GH38	Alpha-mannosidase (3.2.1.24)	2.00E-16	*Thermococcus barophilus* MP
gene_35777**	120	GH18	Chitinase A1 (3.2.1.14)	2.00E-44	*Dictyoglomus thermophilum* H-6-12
gene_35778**	438	GH36	Alpha-galactosidase AgaN (3.2.1.22)	0.00E+00	*Dictyoglomus turgidum* DSM 6724

Table 3.15. (Continued) Non-Full Length Unbleached Kraft GHs.

ID	Length (aa)	GH Family	Known or Predicted Function	BLAST E-value	Organism
gene_36357*	443	GH57	Alpha-amylase (3.2.1.1)	0.00E+00	*Dictyoglomus turgidum* DSM 6724
gene_36443	286	GH03	Xylan 1,4-beta-xylosidase (3.2.1.37)	4.00E-175	*Caldicellulosiruptor kristjanssonii* 177R1B
gene_36604	87	GH13	Isoamylase (3.2.1.68)	4.00E-18	*Fervidobacterium pennivorans* DSM 9078
gene_36731	229	GH02	Beta-galactosidase (3.2.1.23)	2.00E-51	*Thermobaculum terrenum* ATCC BAA-798
gene_36926	245	GH04	Alpha-glucosidase (3.2.1.20)	5.00E+153	*Dictyoglomus thermophilum* H-6-12
gene_37140	243	GH31	Alpha-xylosidase (3.2.1.177)	2.00E-106	*Thermotoga neapolitana* DSM 4359
gene_37349	315	GH43	Alpha-L-arabinofuranosidase (3.2.1.55) / Arabinase (3.2.1.99)	2.00E-160	*Caldicellulosiruptor hydrothermalis* 108
gene_37391	199	GH38	Alpha-mannosidase (3.2.1.24)	6.00E-60	*Dictyoglomus turgidum* DSM 6724
gene_37856	296	GH15	Glucoamylase (3.2.1.3)	7.00E-88	*Candidatus* Kuenenia Stuttgartiensis"
gene_38193	276	GH04	6-phospho-beta-glucosidase BglT (3.2.1.86)	8.00E-140	*Dictyoglomus thermophilum* H-6-12
gene_38194*	272	GH57	Alpha-amylase (3.2.1.1)	8.00E-115	*Hydrogenobacter thermophilus* TK-6
gene_38497	141	GH01	Beta-galactosidase (3.2.1.23)	1.00E-65	*Dictyoglomus turgidum* DSM 6724

Table 3.15. (Continued) Non-Full Length Unbleached Kraft GHs.

ID	Length (aa)	GH Family	Known or Predicted Function	BLAST E-value	Organism
gene_38682	167	GH38	Alpha-mannosidase (3.2.1.24)	1.00E-27	*Rubrobacter xylanophilus* DSM 9941
gene_39002	194	GH05	Beta-D-mannosidase (3.2.1.25)	2.00E-122	*Caldicellulosiruptor saccharolyticus* DSM 8903
gene_39774	154	GH03	Beta-glucosidase B (3.2.1.21)	1.00E-55	*Dictyoglomus turgidum* DSM 6724
gene_40132	182	GH01	Beta-glucosidase (3.2.1.21)	4.00E-89	*Dictyoglomus turgidum* DSM 6724
gene_40145	178	GH31	Alpha-glucosidase (3.2.1.20)	4.00E-64	*Solitalea canadensis* DSM 4303
gene_40284	162	GH26	Beta-1,4-mannanase (3.2.1.78)	4.00E-110	*Caldicellulosiruptor saccharolyticus*
gene_40458	89	GH05	Endoglucanase H (3.2.1.4)	5.00E-55	*Dictyoglomus thermophilum* H-6-12
gene_563*	141	GH01	Beta-glucosidase A2 (3.2.1.21)	3.00E-63	Uncultured organism

* - Transcripts which were originally annotated as separate genes and after manual assembly were able to be annotated as single transcripts

** - Transcripts which were annotated as sequential genes yet still retained individual identity due to supporting evidence of difference according to spatial separation, GH families, predicted function, or related organism.

3.3.2.4.2 Comparison of Glycoside Hydrolases

In an effort to make a comparison of glycoside hydrolases (GHs) found in both samples, the total amount of nucleotide base pairs were compared with the total number of nucleotide base pairs in the sample assembly (Table 3.16).

Table 3.16. Comparison of Nucleotide Bases Dedicated to GH Transcripts. The percentage total number of base pairs used to generate all the contigs of the respective assemblies (Table 3.10 and 3.11).

GHs from Unbleached Kraft (nt)	26,945	GHs from Biomass (nt)	8,373
Total Bases (nt)	23,170,306	Total Bases (nt)	20,946,550
Percent Total Bases Dedicated to GHs	0.12%	Percent Total Bases Dedicated to GHs	0.04%

Although there were more nucleotide bases in the unbleached kraft sample, the number of nucleotides comprising the GH transcripts was also increased. Table 3.17 is a comparison of GH families present in each sample by the amount of base pairs they represent in the total nucleotides contained in the GH transcripts.

Table 3.17. Relative Proportion of Nucleotides of Glycoside Hydrolase Families. The relative proportion was calculated using the total nucleotide base pairs of all transcripts which were found to be of GHs (Table 3.16).

GH Family	Biomass	Base Pairs	Biomass (%)	Unbleached Kraft	Base Pairs	Unbleached Kraft (%)
GH01	3	429	5.1	6	1,375	5.1
GH02	3	575	6.9	3	1,357	5
GH03	0	0	0	5	2,504	9.3
GH04	4	744	8.9	8	2,691	10
GH05	1	345	4.1	5	1,567	5.8
GH10	0	0	0	6	3,083	11.4
GH12	0	0	0	1	288	1.1
GH13	5	768	9.2	3	872	3.2
GH15	1	291	3.5	1	296	1.1
GH16	1	153	1.8	1	566	2.1
GH18	0	0	0	1	120	0.4
GH20	0	0	0	2	1,137	4.2
GH26	0	0	0	1	162	0.6
GH28	1	146	1.7	0	0	0
GH31	4	687	8.2	6	2,836	10.5
GH32	0	0	0	1	305	1.1
GH36	0	0	0	1	438	1.6
GH38	4	1,185	14.2	5	1,458	5.4
GH43	0	0	0	1	315	1.2
GH44	0	0	0	1	305	1.1
GH48	1	61	0.7	1	809	3
GH57	7	1,565	18.7	9	3,435	12.7
GH65	2	343	4.1	0	0	0
GH67	1	132	1.6	0	0	0
GH77	2	670	8	0	0	0
GH78	1	279	3.3	0	0	0
Other	0	0	0	2	1,026	3.8
Sum	**41**	**8,373**	**100**	**70**	**26,945**	**100**

A different view was generated of this data for ease of viewing (Figures 3.40 and 3.41). Both charts are out of the total number of nucleotide base pairs generated by the GH hits in each sample (Table 3.17).

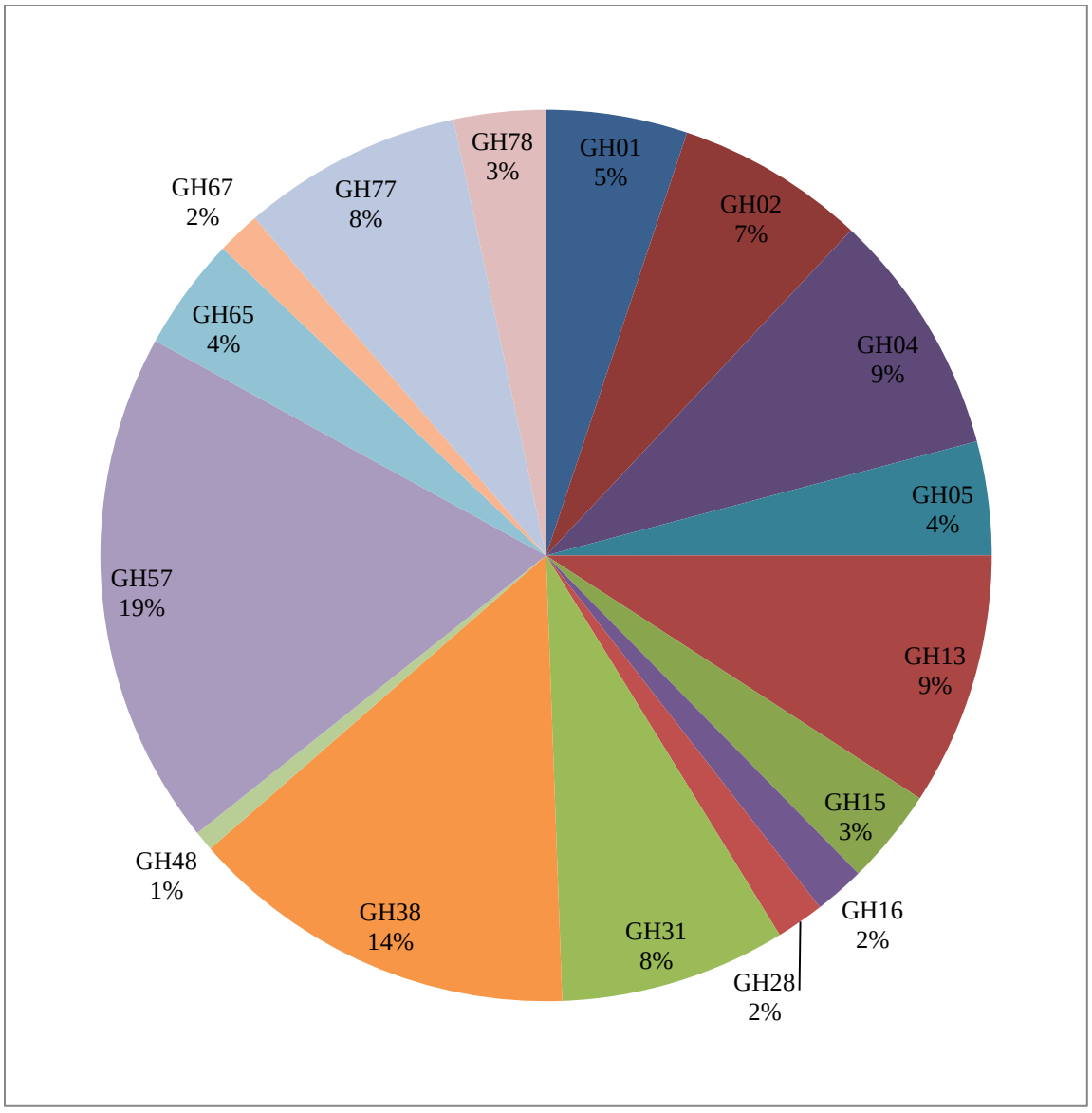


Figure 3.40. Relative proportion of Nucleotides of Glycoside Hydrolase Families from the Biomass Sample. The relative proportions were generated from the total number of nucleotides in the biomass sample which were associated with a GH hit (non-full length and full length transcripts).

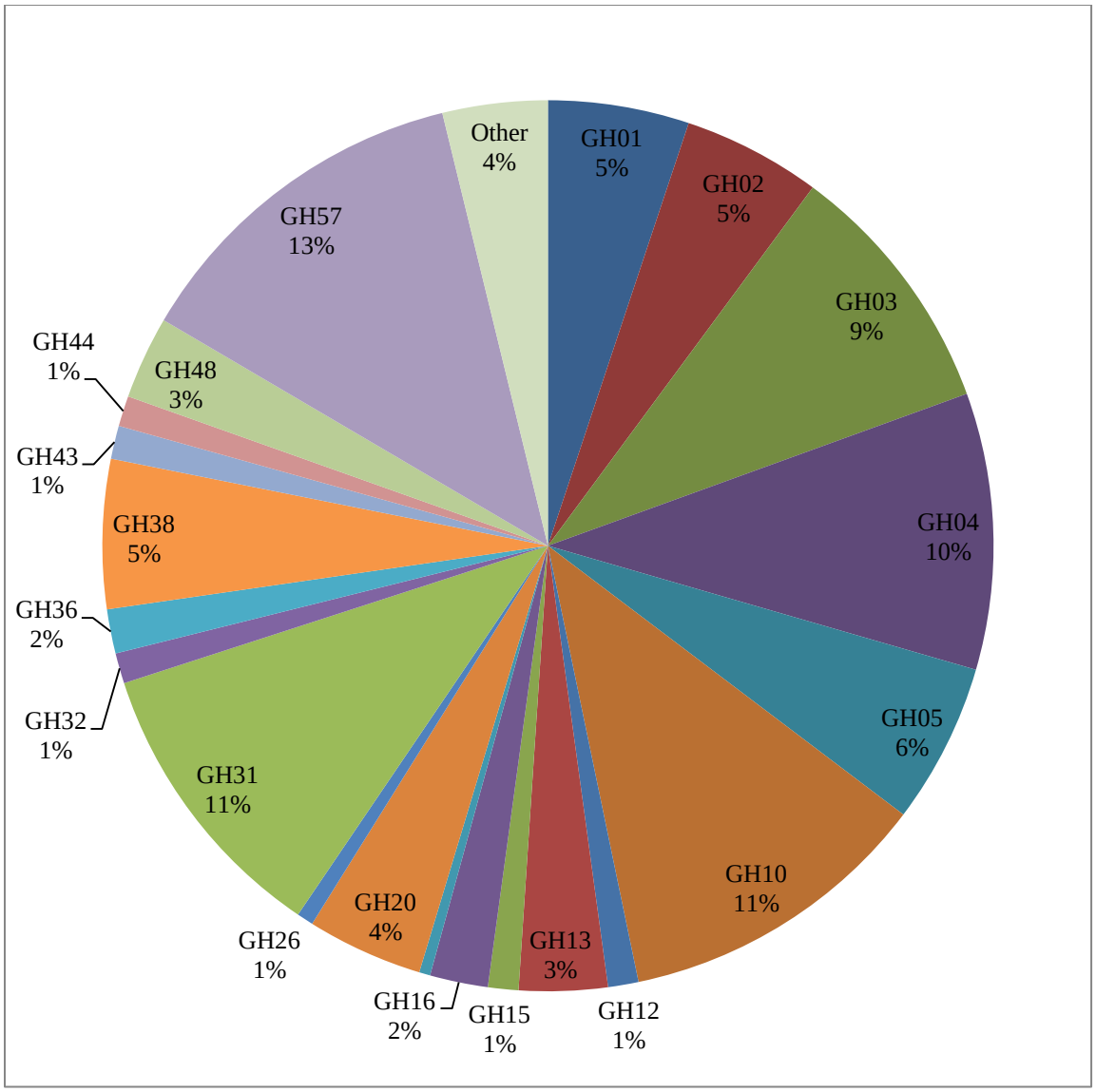


Figure 3.41. Relative Proportion of Nucleotides of Glycoside Hydrolase Families from the Unbleached Kraft Sample. The relative proportions were generated from the total number of nucleotides in the biomass sample which were associated with a GH hit (non-full length and full length transcripts).

A comparison of the relative abundances of GH families in the biomass (Tables 3.12 and 1.14) and unbleached kraft samples (Tables 1.13 and 1.15) is presented in Table 3.18. A juxtaposition of these two data sets in one graph (Figure 3.42) illustrates the difference in GH families of transcripts generated by both samples as a proportion of total nucleotides coding for GH transcripts.

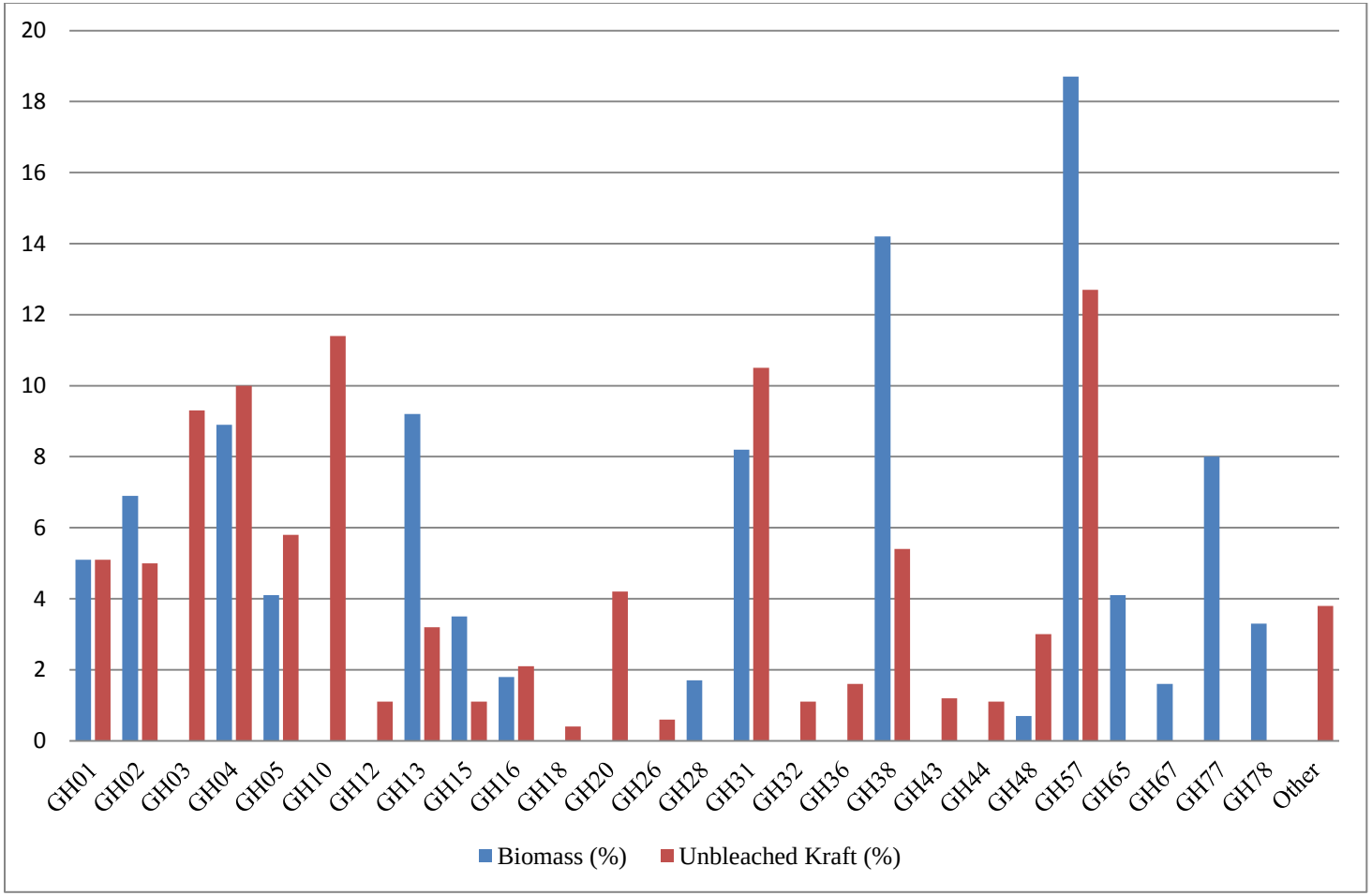


Figure 3.42. Relative Proportion of Nucleotides of GH Families Generated from Both Samples. The proportion of total nucleotides coding for GH transcripts (Table 3.17) for each sample. A lack of bar indicates the lack of representation of GH family in sample.

Table 3.18. Comparison of GH Family Distribution. Comparison made between the two samples taken during the course of this experiment.

GH Family	Biomass	Biomass (%)	Unbleached Kraft	Unbleached Kraft (%)
GH01	3	7.3	6	8.6
GH02	3	7.3	3	4.3
GH03	0	0.0	5	7.1
GH04	4	9.8	8	11.4
GH05	1	2.4	5	7.1
GH10	0	0.0	6	8.6
GH12	0	0.0	1	1.4
GH13	5	12.2	3	4.3
GH15	1	2.4	1	1.4
GH16	1	2.4	1	1.4
GH18	0	0.0	1	1.4
GH20	0	0.0	2	2.9
GH26	0	0.0	1	1.4
GH28	1	2.4	0	0.0
GH31	4	9.8	6	8.6
GH32	0	0.0	1	1.4
GH36	0	0.0	1	1.4
GH38	4	9.8	5	7.1
GH43	0	0.0	1	1.4
GH44	0	0.0	1	1.4
GH48	1	2.4	1	1.4
GH57	7	17.1	9	12.9
GH65	2	4.9	0	0.0
GH67	1	2.4	0	0.0
GH77	2	4.9	0	0.0
GH78	1	2.4	0	0.0
Other	0	0.0	2	2.9
Sum	**41**	**100**	**70**	**100**

A graphical representation was made of GH families (Table 3.18) present in the biomass sample (Figure 3.43) and unbleached kraft sample (Figure 3.44). The relative proportion was generated from total transcripts and not dependent on length of transcript (Figure 3.45).

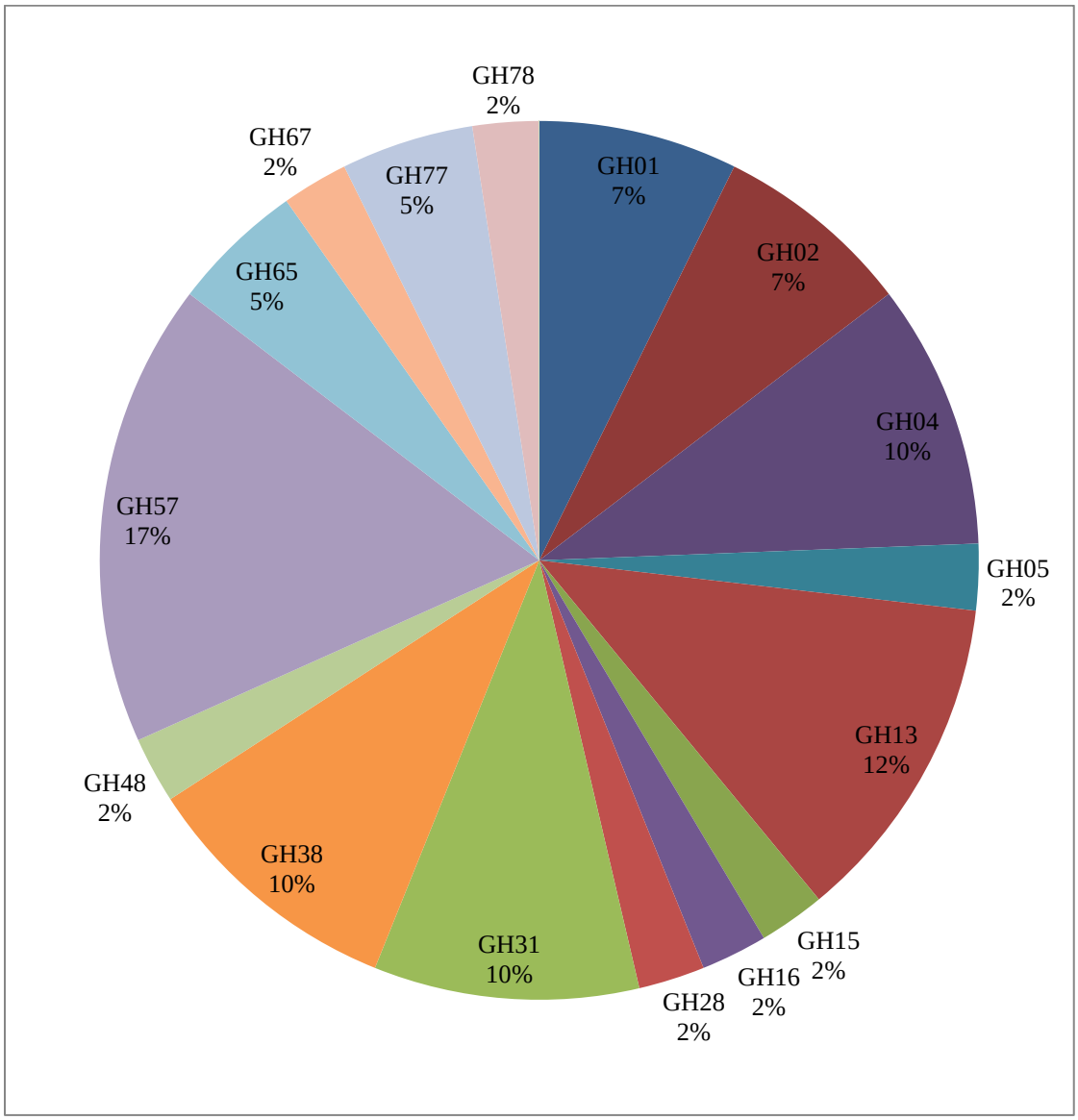


Figure 3.43. Relative Proportion of Glycoside Hydrolase Families from the Biomass Sample. All GH families which were not represented in this sample were removed for ease of viewing.

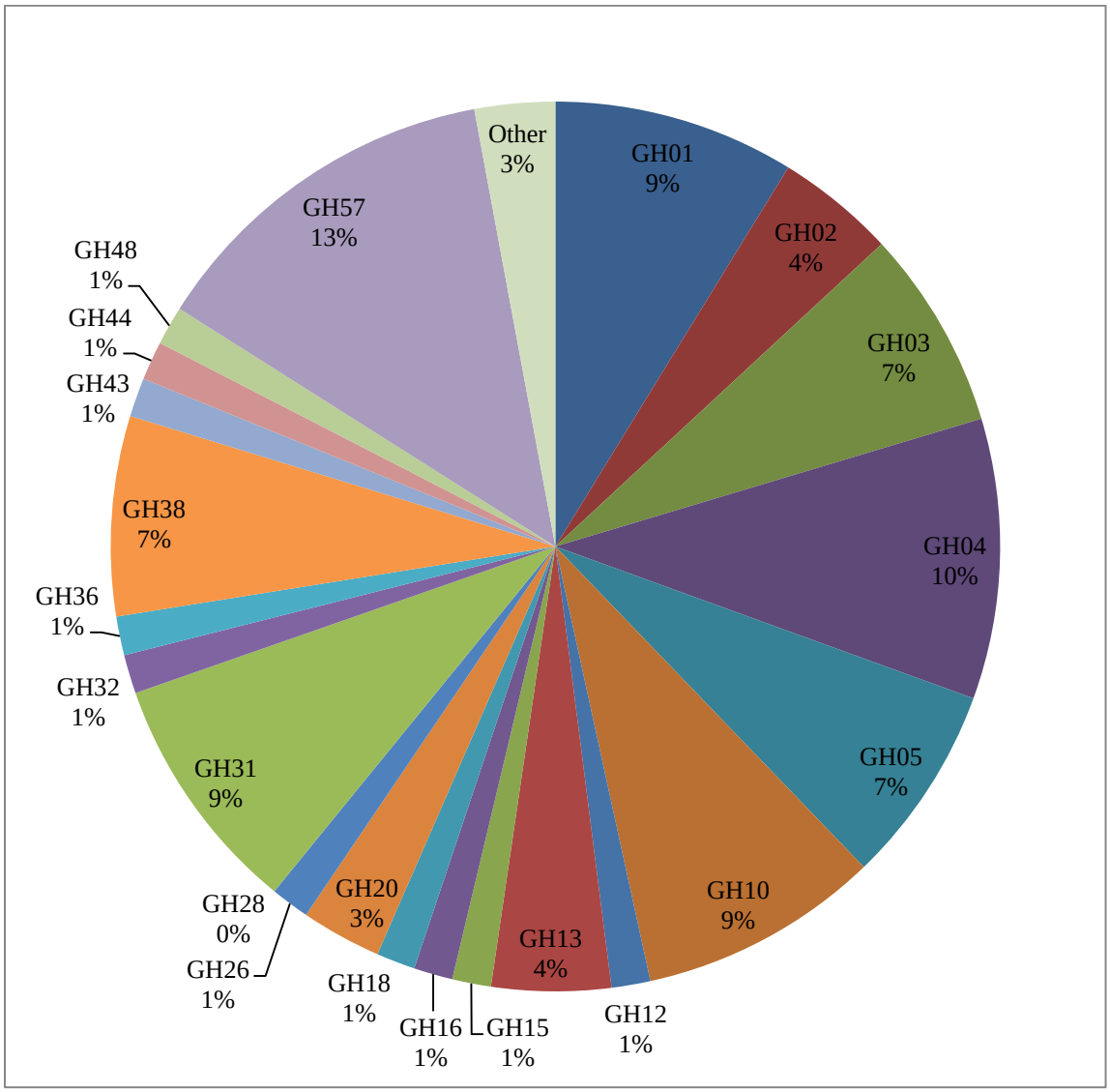


Figure 3.44. Relative Proportion of Glycoside Hydrolase Families from the Unbleached Kraft Sample. All GH families which were not represented in this sample were removed for ease of viewing.

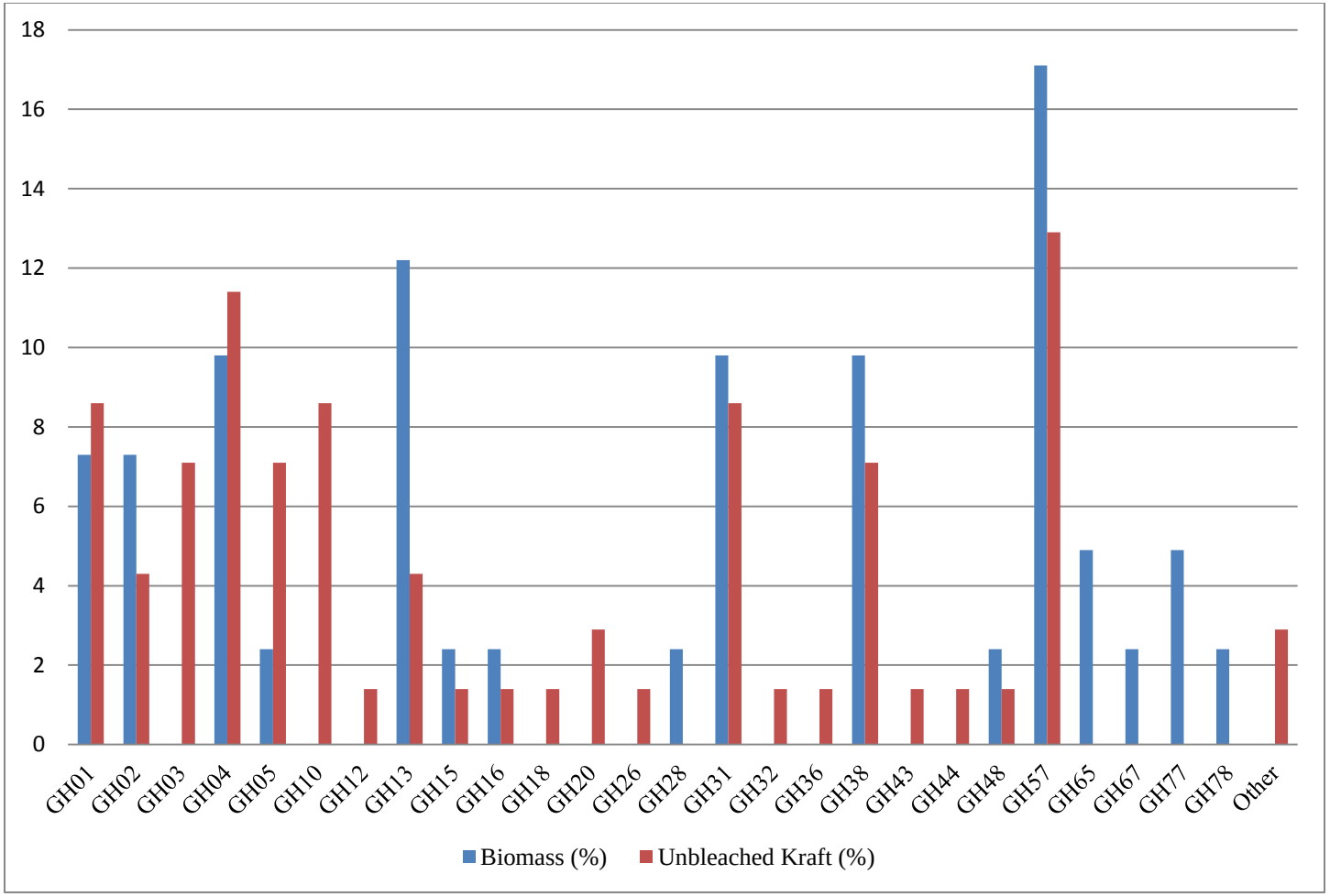


Figure 3.45. Relative Proportion of GH Families Generated from Both Samples. Percentage generated from comparison of total transcripts corresponding to GHs (Table 3.18) for each sample. A lack of bar indicates the lack of representation of the GH family in sample.

A local BLASTn [3] was run using the larger GH group (unbleached kraft) as the database. The BLAST generated a list of transcripts which were expressed in both samples (Table 3.19).

Table 3.19. BLAST Result of GH Transcript Overlap. Those GHs expressed in both the biomass and the unbleached kraft samples. The note "full" denotes a full length transcript and the note "non" denotes a non-full length transcript.

Biomass	Unbleached Kraft	GH Family
gene_3317(non)	gene_38194(non)	GH57
gene_36258(full)	gene_34865(full)	GH38
gene_41570(non)	gene_27761(non)	GH57
gene_45450(non)	gene_13844(non)	GH57
gene_45889(non)	gene_566(full)	GH4
gene_46781(non)	gene_33001(non)	GH38
gene_47009(non)	gene_30429(full)	GH31

The phylogeny of GH transcripts was examined (Table 3.20). The phylogeny was generated from the top BLAST hit resulting from BLASTx [3] challenge (Tables 3.12 through 3.15). It represents the phylum level of *Bacteria* and *Achaea* for both the biomass (Figure 3.46) and unbleached kraft (Figure 3.47) samples. A juxtaposition of these two data in one graph (Figure 3.48) illustrates the difference in phylogeny of GH transcripts generated by both samples.

Table 3.20. Comparison of Phylogeny of All GH Transcripts. Representation of the phylogeny generated from the top BLAST hit of GH transcripts from both samples (Figures 3.12 through 3.15).

Phylum	Biomass	Biomass (%)	Unbleached Kraft	Unbleached Kraft (%)
Acidobacteria	3	7.1	1	1.4
Actinobacteria	2	4.8	2	2.9
Aquificales	3	7.1	1	1.4
Bacteroidetes	3	7.1	1	1.4
Chloroflexi	6	14.6	1	1.4
Chrysiogenetes	1	2.4	0	0.0
Cyanobacteria	2	4.8	0	0.0
Deinococcus - Thermus	2	4.8	2	2.9
Dictyoglomi	4	9.5	35	50.0
Euryarchaeota	1	2.4	1	1.4
Firmicutes	4	9.5	13	18.6
Planctomycetes	0	0.0	1	1.4
Proteobacteria	0	0.0	1	1.4
Spirochetes	1	2.4	2	2.9
Thermotoga	5	11.9	6	8.6
Unknown	3	7.1	3	4.3
Verrucomicrobia	1	2.4	0	0.0
SUM	**41**	**100**	**70**	**100**

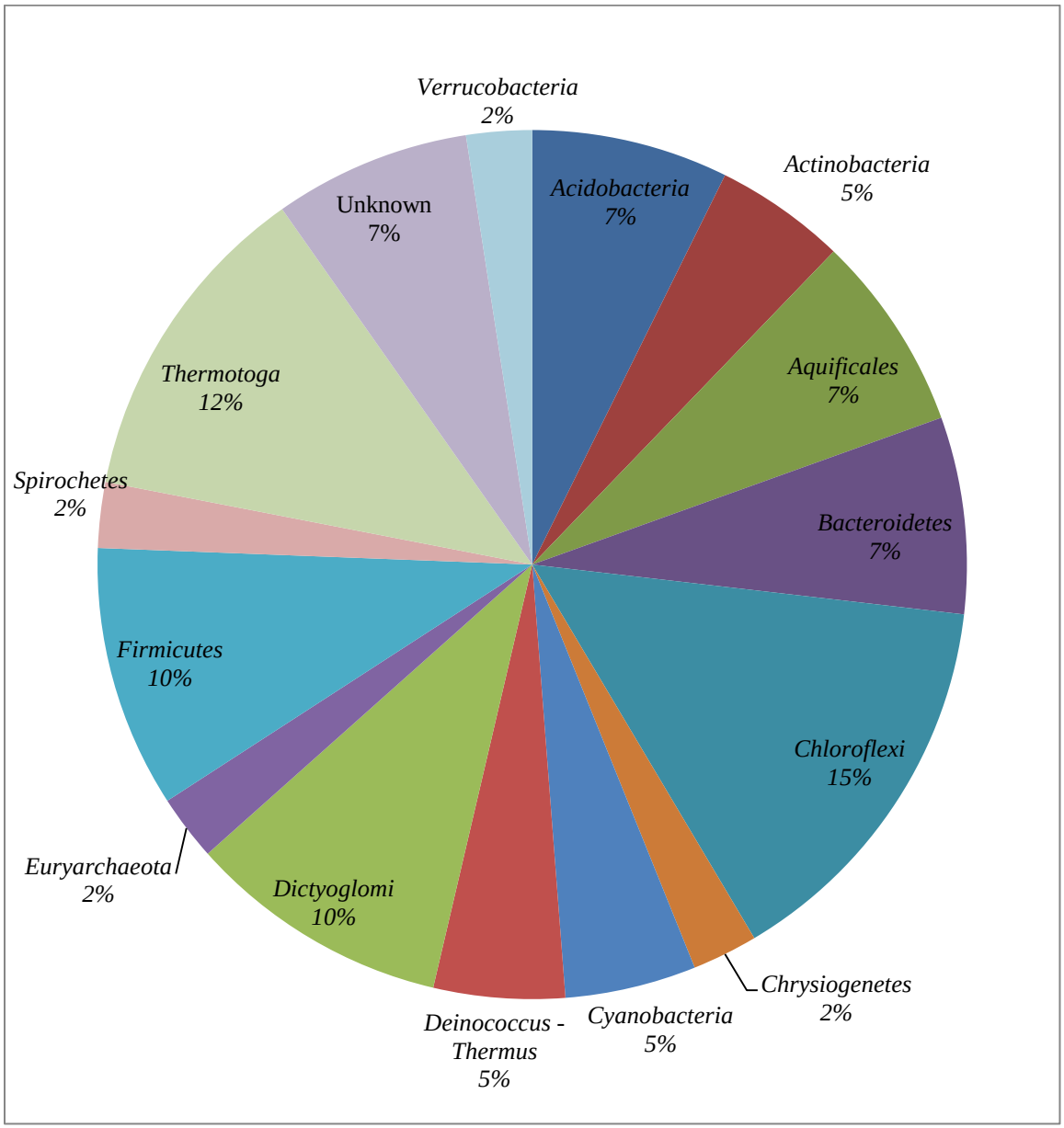


Figure 3.46. Biomass Sample Phylogeny of All GHs. Representative phyla of GH hits as generated from the top BLAST hit organized as a proportion of the total GH hits from the biomass sample.

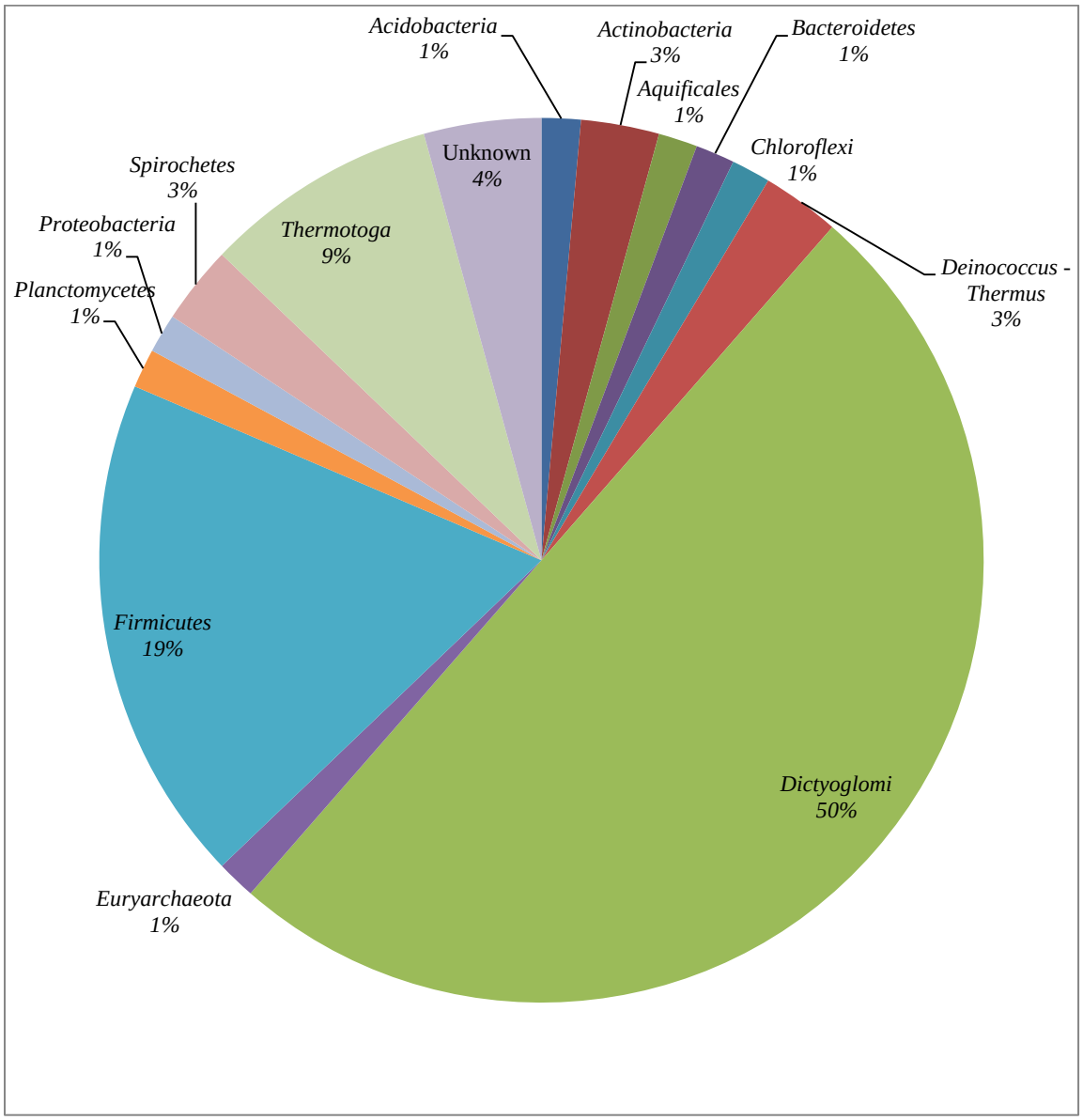


Figure 3.47. Unbleached Kraft Sample Phylogeny of All GHs. Representative phyla of GH hits as generated from the top BLAST hit organized in the per cent of total GH hits from the unbleached kraft sample.

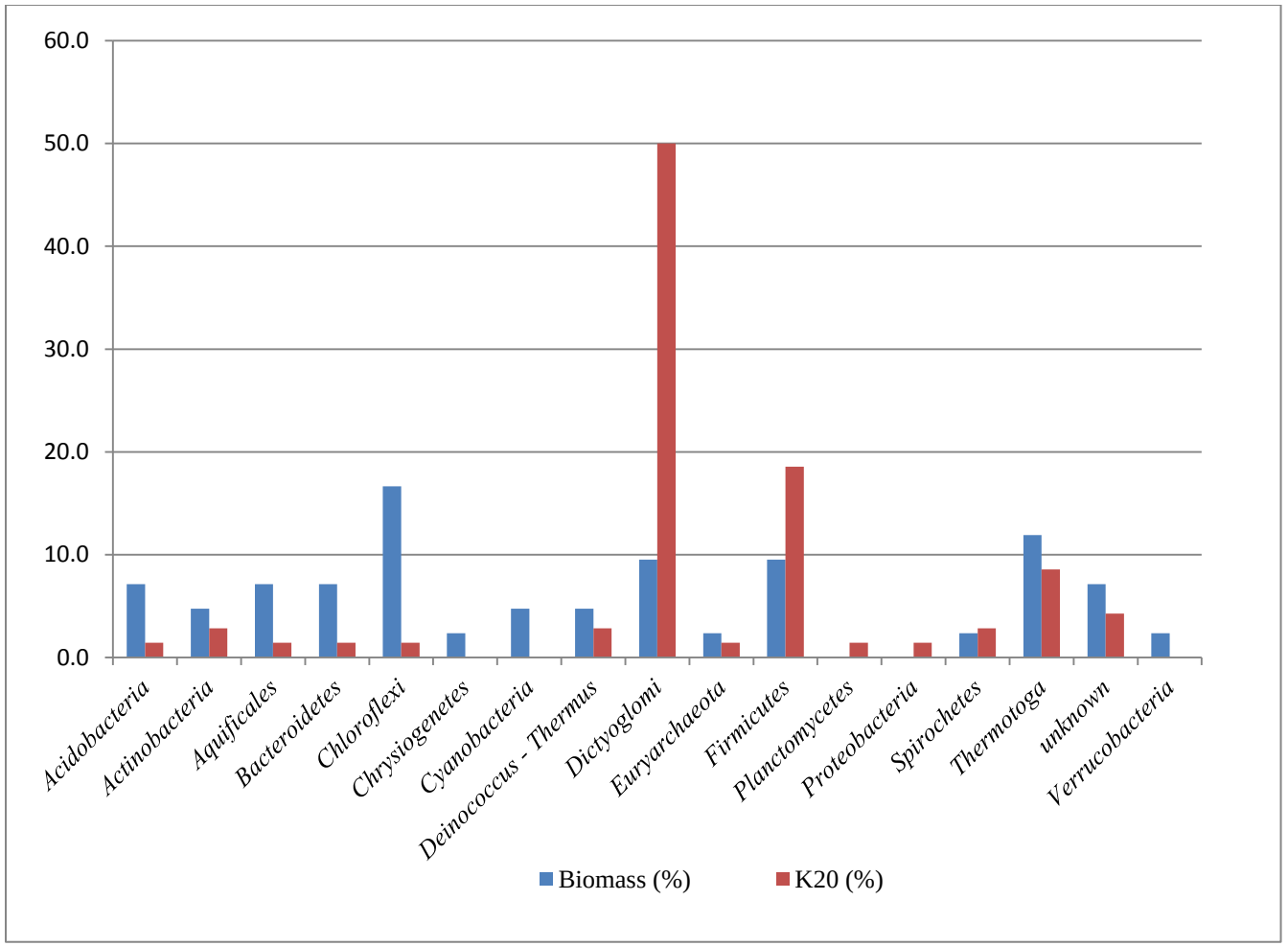


Figure 3.48. Relative Proportion of Phylum Generated from GH Hits from Both Samples. A comparison of the two samples and the relative proportion of GH phylogeny as given by the BLAST output.

The relative proportions are similar when the GH phylogeny is calculated as a percentage of total nucleotides in the sample (coding for transcripts associated with GH function) (Table 3.21).

Table 3.21. Relative Proportion of Nucleotides of Phylum Represented in Each Sample. Representation of the proportion of nucleotides associated with phylogeny generated from the top BLAST hit of GH transcripts from both samples (Figures 3.12 through 3.15).

Phylum	Biomass Nucleotides	Biomass (%)	Unbleached Kraft Nucleotides	Unbleached Kraft (%)
Acidobacteria	614	7.3	93	0.3
Actinobacteria	885	10.6	1063	3.9
Aquificales	945	11.3	272	1
Bacteroidetes	427	5.1	178	0.7
Chloroflexi	1076	12.9	229	0.8
Chrysiogenetes	184	2.2	0	0
Cyanobacteria	341	4.1	0	0
Deinococcus - Thermus	412	4.9	688	2.6
Dictyoglomi	627	7.5	14621	54.3
Euryarchaeota	0	0	188	0.4
Firmicutes	686	8.2	6132	22.8
Planctomycetes	0	0	296	1.1
Proteobacteria	0	0	317	1.2
Spirochetes	66	0.8	378	1.4
Thermotoga	1315	12.7	1733	6.4
Unknown	574	6.9	827	3.1
Verrucomicrobia	221	2.6	0	0
SUM	**8373**	**100**	**26,945**	**100**

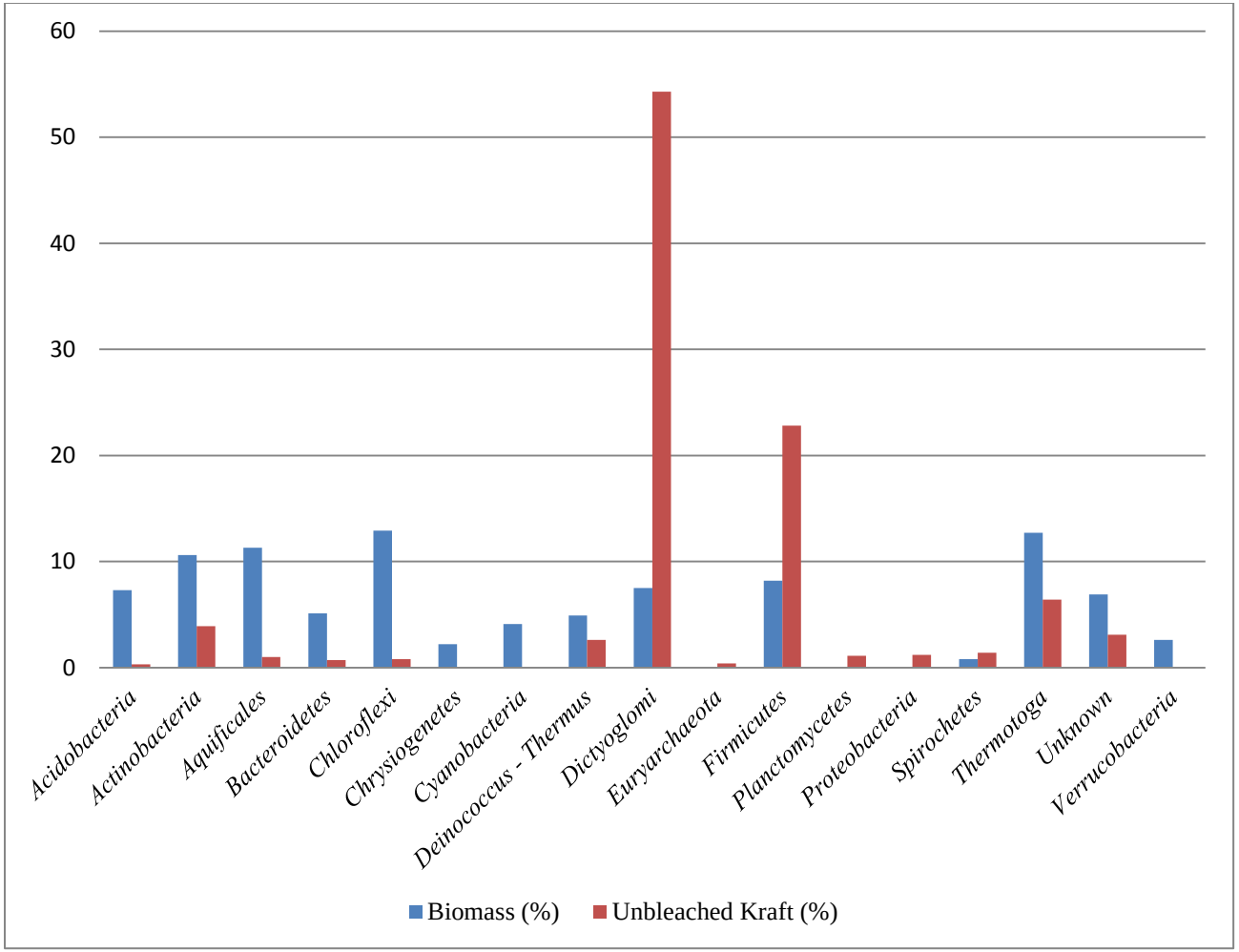


Figure 4.49. Relative Proportion of Nucleotides of Phylum Generated from GH Hits from Both Samples. A comparison of the two samples and the relative proportion of nucleotides associated with GH phylogeny as given by the BLAST output (Table 3.12 through 3.15).

Lastly, transcripts were grouped by EC number [144] of the top BLAST hit (Tables 3.12 through 3.15). This produced data similar to grouping by GH family, however it allowed the categorization of action (endoglucanase/exoglucanase) and linkage target specificity (α- or β- linked bonds) on a general level. Sugar monomers of polysaccharides are linked together by either α- or β- linked bonds (Section 1.5). Enzymes are specific for either one bond or the other, and can be separated by linkage specificity (Table 3.22). Exclusively detected transcripts from the biomass (Table 3.23) and unbleached kraft sample (Table 3.24) illustrate a common trend in linkage specificity.

Table 3.22. EC Number, GH Family, Endoglucanse/Exoglucanase Action, and Linkage Target Specificity of GH Transcripts. Endoglucanase/exoglucanase activity (Section 1.10) and linkage target specificity (Section 1.5) can be assessed by grouping transcripts by EC number [144]. Linkage target specificity corresponds to specify for either α- or β- bonds which connect sugar monomers of polysaccharides.

EC Number	GH Family	Biomass	Unbleached Kraft	Endoglucanase /Exoglucanase Activity*	α/β Activity
2.4.1.230	GH65	2	0	exo	α
3.2.1.1	GH57	6	8	endo	α
3.2.1.3	GH15	1	1	exo	α
3.2.1.4	GH05	2	7	endo	β
3.2.1.6	GH16	0	1	endo	β
3.2.1.8	GH10	0	4	endo	β
3.2.1.10	GH04	0	1	exo	α
3.2.1.14	GH18	0	1	endo	β
3.2.1.15	GH28	1	0	endo	α
3.2.1.20	GH04	4	9	exo	α
3.2.1.21	GH01	5	7	exo	β
3.2.1.22	GH31	2	1	exo	α
3.2.1.23	GH02	0	4	exo	β
3.2.1.24	GH38	4	5	exo	α
3.2.1.25	GH77	2	2	exo	β
3.2.1.32	GH10	0	1	endo	β
3.2.1.37	GH03	0	1	endo	β
3.2.1.39	GH16	1	0	endo	β
3.2.1.41	GH57	4	3	endo	α
3.2.1.40	GH78	1	0	exo	α

Table 3.22. (Continued) EC Number, GH Family, Endoglucanse/Exoglucanase Action, and Linkage Target Specificity of GH Transcripts.

EC Number	GH Family	Biomass	Unbleached Kraft	Endoglucanase /Exoglucanase Activity*	α/β Activity
3.2.1.52	GH20	0	2	exo	β
3.2.1.55	GH43	0	1	exo	α
3.2.1.68	GH04	1	1	endo	α
3.2.1.74	GH03	0	1	exo	β
3.2.1.78	GH26	0	2	endo	β
3.2.1.86	GH04	1	2	endo	β
3.2.1.135	GH13	1	0	endo	α
3.2.1.139	GH67	1	0	exo	α
3.2.1.153	GH32	0	1	exo	β
3.2.1.176	GH48	0	1	endo	β
3.2.1.177	GH31	1	1	exo	α
Unknown	Unknown	1	2	Unknown	Unknown
SUM	**N. A.**	**41**	**70**	**N. A.**	**N. A.**

* - Endoglucanase/Exoglucanase and Linkage Specificity determination was made by referencing (http://www.chem.qmul.ac.uk/iubmb/) for additional information corresponding to the EC number examined.
N. A. - Not applicable.

Table 3.23. Information on Transcripts Exclusively Detected in the Biomass Sample. Endoglucanase/exoglucanase activity (Section 1.10) and linkage target specificity (Section 1.5) can be assessed by grouping transcripts by EC number [144]. Linkage target specificity corresponds to specificity for either α- or β- bonds which connect sugar monomers of polysaccharides.

EC Number	GH Family	Biomass	Endoglucanase /Exoglucanase Activity*	α/β Activity
2.4.1.230	GH65	2	exo	α
3.2.1.15	GH28	1	endo	α
3.2.1.39	GH16	1	endo	β
3.2.1.40	GH78	1	exo	α
3.2.1.135	GH13	1	endo	α
3.2.1.139	GH67	1	exo	α

* - Endoglucanase/Exoglucanase and Linkage Specificity determination was made by referencing (http://www.chem.qmul.ac.uk/iubmb/) corresponding to the EC number examined.

Table 3.23. Information on Transcripts Exclusively Detected in the Unbleached Kraft Sample. Endoglucanase/exoglucanase activity (Section 1.10) and linkage target specificity (Section 1.5) can be assessed by grouping transcripts by EC number [144].

EC Number	GH Family	Unbleached Kraft	Endoglucanase /Exoglucanase Activity*	α/β Activity
3.2.1.6	GH16	1	endo	β
3.2.1.8	GH10	4	endo	β
3.2.1.10	GH04	1	exo	α
3.2.1.14	GH18	1	endo	β
3.2.1.23	GH02	4	exo	β
3.2.1.32	GH10	1	endo	β
3.2.1.37	GH03	1	endo	β
3.2.1.52	GH20	2	exo	β
3.2.1.55	GH43	1	exo	α
3.2.1.74	GH03	1	exo	β
3.2.1.78	GH26	2	endo	β
3.2.1.153	GH32	1	exo	β
3.2.1.176	GH48	1	endo	β

* - Endoglucanase/Exoglucanase and Linkage Specificity determination was made by referencing (http://www.chem.qmul.ac.uk/iubmb/) corresponding to the EC number examined.

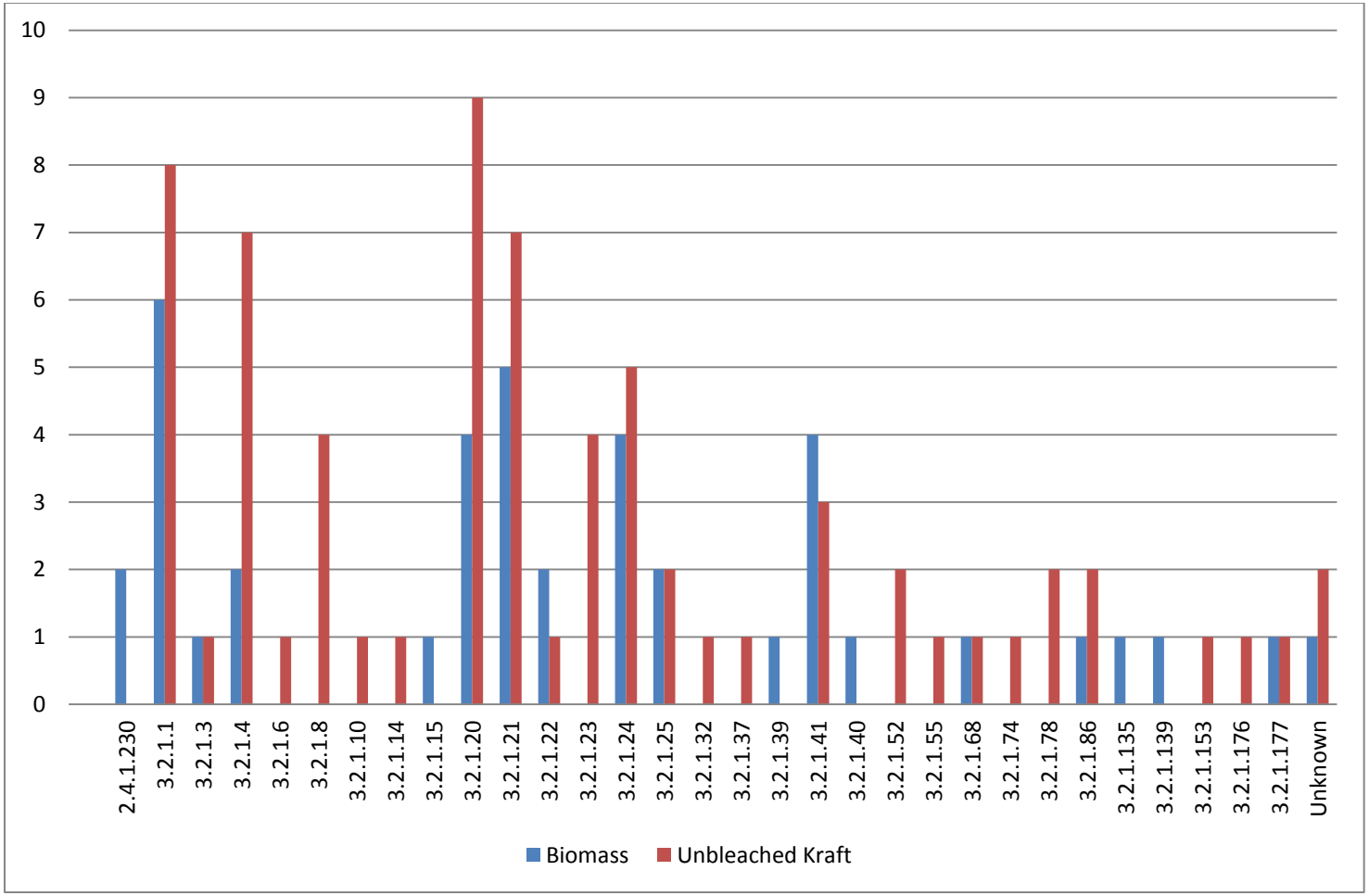


Figure 4.50. Amount EC Number Representation from Each Sample. Number of transcripts with predicted activity of a given EC Number [144].

3.3.3 Ngatamariki Summary of Results

Through the use of an enrichment, microorganisms were selected who possessed the ability to utilize lignocellulose as an energy source. Sequence data was generated and annotated with a bioinformatics pipeline which revealed the presence of transcripts coding for Glycoside Hydrolases (GHs). These transcripts were measured at the transcriptional level in the native environment (biomass) and the enrichment (unbleached kraft). Finally, the bioprospecting endeavour to identify GHs was complete at the sample site Ngatamariki.

Chapter 4

Discussion

4.1 Introduction

The specific aims of this Master's thesis (Section 1.15) allowed for the successful bioprospecting for thermostable transcripts associated with glycoside hydrolases (GHs) activity. Initially, methodology was developed to enrich and isolate mRNA transcripts from geothermal samples. Due to unresolved issues at the initial sample site, Waikite (Section 3.2), a second site was chosen to continue the project. Unlike Waikite, resident microorganisms at the second site (Ngatamariki - Section 3.3) were not known to be cellulolytic. Other energy sources than lignocellulose were present in the pool (Table 3.7), indicating that the presence of cellulolytic microorganisms was not guaranteed. Therefore, an enrichment was created to ensure the production of the target enzymes from thermophilic and cellulolytic microorganisms. Extraction and enrichment of mRNA transcripts, present in the environment and enrichment, were conducted. This generated a metatranscriptomic library of transcripts which annotated and identified candidate GHs. GHs detected in environment and enrichment were compared based on function (Table 3.22), GH family (Table 3.17), and phylogeny of the top BLAST hit (Table 3.20). The unforeseen issues with the method development and troubleshooting the Waikite samples reduced the time available to clone a target GH and confirm predicted function. However, this project was successful in generating sequences of active thermostable GHs which could potentially hold value for the bioethanol industry.

4.2 Method Development of Nucleic Acid Extraction from Geothermal Soils

Methodology was successfully developed to enrich and isolate mRNA transcripts through the use of *Escherichia coli* as a positive control (Section 3.2.2.1). The extraction of nucleic acid from geothermal soil was initially a problem (Section 3.2.2.2) due to the presence of clay-like soil colloids. These colloids were believed to be the inhibitory agent present in the soil at Waikite (Section 3.2.3). Several attempts were made to liberate nucleic acids from the soil, including blocking with skim milk (Section 3.2.2.3) and processed gDNA (Section 3.2.2.4). Due to temporal and monetary constraints associated with these unresolved issues, a second site was selected to continue the work of this project.

A site known to have low concentrations of the putative inhibitory agents (that restricted nucleic acid extraction at Waikite) was chosen instead of further method development at the original site. Ngatamariki, the second site chosen for this experiment, did not contain a large clay component in the soil. Therefore, nucleic acid extraction was not a problem and was achieved with relative ease (Section 3.3.2.2). However, Ngatamariki was not known to harbour cellulolytic microorganisms and an enrichment strategy was devised to select for cellulolytic microorganisms and therefore ensure detection of transcripts related to cellulolysis.

4.3 Enrichment for Cellulolytic Microorganisms at Ngatamariki

The spring at Ngatamariki was not known to contain cellulolytic microorganisms (Section 3.3). Introducing an abundant energy source would allow microorganisms capable of utilizing this energy to represent an increased proportion of the population

through natural selection (Section 1.12). At the end of the incubation period, there was observable degradation of one of the provided lignocellulosic substrates (Figure 3.25). Extraction and enrichment of mRNA transcripts was conducted on three out of four enrichments. Out of these three, one enrichment produced usable sequence data when the resulting cDNA was sequenced. It was unfortunate that the enrichment with Whatman filter paper and the bleached kraft paper were not able to generate sequence data as the additional data would have allowed for a more detailed comparison of GH transcript activities present in different environments.

4.4 Assembly and Annotation of Metatranscriptomic Library

The biomass and unbleached kraft samples generated usable sequences from the cDNA libraries (Section 3.3.2.4.1). These assemblies contained large amounts of sequences generated from mRNA transcripts of various functionalities. Machine and manual annotation generated transcripts related to the functionality of glycoside hydrolases (GHs) (Section 3.3.2.4.2). Sequences of interest were examined in more detail with current databases after initial screening.

The assembly generated by the biomass sample (20,946,550 base pairs) was smaller than the unbleached kraft sample (23,170,306 base pairs) despite the extra quarter plate sequencing run completed for the biomass cDNA (Table 3.9). This could be evidence that increased read depth (as generated by more sequencing data) leads to related sequences (corresponding to the same transcript) being assembled correctly instead of as slightly different separate sequences. This observation may also be a product

of more usable data being generated by the unbleached kraft sample due to extensive method development during the process of the experiment.

Metatranscriptomics detects only actively transcribed enzymes (Section 1.14), thus metatranscriptomic approach bypasses the need to assess which of the potential GHs are responsible for the cellulolytic activity on different substrates. For example, *Clostridium phytofermentans,* a known cellulolytic microorganism has been shown to contain many genes related to GH function (108 putative GHs) (Section 1.10). However, which of these GH's (or combinations thereof) acts on particular substrates can only be surmised via annotation or extensive investigations. Thus a metatranscriptomic approach eliminates a large amount of work and potentially useless information generated by a metagenomic study and shows the true power of a metatranscriptomic endeavour when bioprospecting for a specific activity. In this vein, the GH transcripts generated from both samples allow for comparison of GH family (Table 3.17), function (Table 3.22), and phylogeny of the top BLAST hit (Table 3.20).

4.4.1 Glycoside Hydrolases

There was an increased amount of GH transcripts detected in the enrichment (70) than the native environment (41). These transcripts were sorted into full length and non-full length transcripts according to coverage of the top BLAST hit (Table 3.12 through 3.15). There was an increase in full length GH transcripts in the enrichment (30) than the native environment (3). This was also accompanied by an increase percentage of base pairs attributed to GH transcripts detected in the enrichment (0.12 %) compared to the native environment (0.04 %), despite the increased amount of total nucleotides in the

enrichment assembly (Section 4.4). The unbleached kraft sample detected 29 more transcripts dedicated to GH production and 19,038 more nucleotide base pairs than in the natural environment.

The GH transcripts from both samples were related to both endoglucanase / exoglucanase action (Table 3.22), due to both types being needed for complete hydrolysis of complex polysaccharides (Section 1.10). No trends can be seen in the production of endoglucanases or exoglucanases in the two different environments. However, endoglucases have the ability to greatly influence the bioconversion of lignocellulose (Section 1.10). Therefore, full length endoglucases found in this experiment (gene_34538, gene_28236, gene_30428) provide potentially industrially applicable thermostable GHs (Section 1.11).

Polysaccharides are sugar monomers linked in either α- or β- glycosidic bonds (Section 1.5) depending on the type of polysaccharide. Specific enzymes hydrolyse different linkages. The majority of transcripts exclusively detected in the biomass sample were active on α-linked polysaccharides (Table 3.23) [25]. In contrast, the majority of transcripts exclusively detected in the unbleached kraft sample were active on β-linked polysaccharides (Table 3.24) [25]. Importantly, the unbleached kraft was composed chiefly of glucan (80.8 %), a β-linked polysaccharide (Section 3.3.1.1) indicating that the enrichment strategy successfully targeted GHs active against the substrate. The detection of α-linked polysaccharides in the environment was not attempted, although considering that the spring contained large quantities of degrading tree and plant matter, it was suspected that starch from plant cells would be readily available. In addition, glycogen, an α-linked polysaccharide (α-glucan), is an ubiquitous energy storage molecule among

bacteria [29]. Therefore, the metabolism of α-linked polysaccharides should be expected in a normal microbiological community. Despite the lack of α-linked enzyme activity needed for lignocellulosic bioethanol production, starch ethanol (not discussed in this thesis) is also an area of research pertaining to bioethanol production [131].

The categorization of function was conducted on the basis of top BLAST hit. Although an E-value of 0.0 was recorded for many GH transcripts, others were not as homologous. Enzyme Commission (EC) number [144] categorization illustrated a general trend of linkage specificity in the two samples. However, different EC numbers can belong to the same GH family. Therefore, GH family was also used for more general characterization of transcripts present in the respective samples.

4.4.2 Glycoside Hydrolase Families

More GH families were exclusively detected in the enrichment (11) than those exclusive to the native environment (5). As with the EC number comparison of GHs, those GH families exclusively detected in the biomass sample (GH28, 65, 67, 77, and 78) (Table 3.23) are of enzymes which have activity on α-linked polysaccharides [25]. With the exception of Family 36, all GH families expressed exclusively in the unbleached kraft enrichment (GH12, 18, 20, 26, 32, 43, and 44) (Table 3.24) were active on β-linked polysaccharides (the chief component of the substrate provided in the enrichment). It was postulated that those transcripts present in the biomass sample were not involved in lignocellulosic breakdown because of the lack of β-linked activity (Section 4.4.1). Rather, as discussed previously, resident microorganisms expressed enzymes involved in oligosaccharide scavenging and metabolism.

GH Family 48 has been shown to be crucial in the growth of *Clostridium thermocellum* crystalline cellulose [95]. Therefore, it was expected that GH Family 48 was detected in both samples (gene_46988 and gene_15250). Detection of GH Family 9 was expected in both samples due to previous reports of its requirement in the bioconversion process of cellulose [135] (Section 1.10). GH9 proteins are all β -1,4-glucanases [135] (Endoglucanases (EC 3.2.1.4))[25]. Multiple enzymes with an endoglucanase function were found in both samples. It is plausible that enzymes with a related function negated the need for GH9 expression.

Xylan was present in the unbleached kraft sample (5.4 %). Detection of GH10 transcripts was drastically increased in the enrichment sample mRNA pool (Table 3.18). This family encodes either an endo-1,4- β -xylanase (EC 3.2.1.8) or an endo-1,3- β -xylanase (EC 3.2.1.32). No GH transcript in the biomass sample had the predicted function of a xylanase (Figure 3.12 and 3.13). The production of xylanases was in response to changes in the environment.

As described above, the GH Family 36 transcript (gene_35778) was the only transcript expressed exclusively in the unbleached kraft sample which was α-acting. This lone transcript may be a product of a bacterial operon. It is denoted by a "**" (Figure 3.15). This indicates that there was another GH transcript (gene_35777) in the proximity of the transcript. Therefore it is plausible that the transcription of this sequence was merely an artifact of the expression of a set of genes with related functionalities.

Differences were observed in the relative amount of GH families present in the two samples when measurement was done with relative transcript amount (Figure 3.45) and relative nucleotide amount (Figure 3.42). A difference can be seen in relative

proportion of GH31 in the two samples between the two figures. Although the proportion of transcript number is higher in the biomass sample, those transcripts are all non-full length genes. In the enrichment, those transcripts corresponding to GH31 are half full length and half non-full length genes. GH31 is a complex family to study in this manner due to broad functionalities contained within this family. Gene_30429, corresponding to an alpha-glucosidase (3.2.1.20), and gene_28236, coding for Endoglucanase (3.2.1.4), are both are attributed to family GH31 (Table 3.14). Therefore, difference observed by the two figures could be attributed to the types of transcripts in which a difference was seen.

Comparison of GH family composition allows for general trends to be observed. The microbial population changes its expression profile depending on environmental changes. The microbial population also changes its composition. BLAST generates an organism in which the top hit originates. The phylogeny of population constituents can be studied in this way.

4.4.3 Glycoside Hydrolase Phylogeny

Phylum level phylogeny of the top BLAST hit (Table 3.12 through 3.15) allowed for comparison of diversity (Table 3.20). While the unbleached kraft sample has a higher number of total GHs, there is still an overrepresentation of the phylum *Dictyoglomi* (50 %) and *Firmicutes* (18.6 %). The chief genus representative of phylum *Firmicutes* (*Caldicellulosiruptor*) and *Dictyoglomi* (*Dictyoglomus*) are both from thermophilic and cellulolytic genera (Table 1.3). GH transcripts from *Dictyoglomi* in the unbleached kraft sample (35) are close in number to transcripts detected in the entire biomass data set (41). This appears to be evidence that the dominant microorganism detected producing GHs in

the enrichment sample was of the phylum *Dictyoglomi*. However, this may be an artifact of the examination of using transcripts associated with GH function. What can be extrapolated without question is that the phylum *Dictyoglomi* was the chief producer of GH transcripts in the enrichment. An examination of the total transcripts generated by the sequencing project could establish whether it was dominant producer of all transcripts or just those with GH activity.

A comparison of the phylogeny of GHs present in the two samples indicates that the phylogenetic diversity decreases. By selecting for those microorganisms which can utilize lignocellulose as an energy source, an environment was created where those microorganisms would be overrepresented when cellulose-active enzymes were examined. It is interesting to note the similarity between relative proportions generated by raw transcript number (Figure 3.47) and total nucleotide proportion (Figure 3.48) unlike GH31 in the previous section.

4.4 Future Work

Several improvements could be made to this experiment if a study similar to this is conducted in the future. As discussed previously (Section 3.2.3), if an experiment is attempted at the Waikite site in the future, a massive scale-up of the extraction procedure or washing of fixed whole-cells before extraction could possibly alleviate the problems associated with nucleic acid extraction.

Concerning the enrichment setup at Ngatamariki, the anaerobic tube caps and bungs should be removed for the duration of the incubation if tubes are used again for the housing of the enrichments. This would encourage the colonization by resident

microorganisms and lead to an increase in microorganism and transcript numbers. Also, an increase in incubation time would generate a larger microbial population than was observed in this experiment. The generation of sequence data from the two addition enrichments would make comparison of the samples more in-depth.

The phylogenetic composition of the community can be created in a more reproducible way than was done in this experiment. Before the depletion of DNA with DNase (Section 2.4), an amount of DNA necessary for deep sequencing of taxonomic DNA sequences (16S rRNA gene – 16S rDNA) should be separated and retained. A deep sequencing project of this nature allows for the most accurate estimation of population composition [96]. Examination could then be conducted on the relative population dynamics not only supplied by transcripts associated with a single function.

Verification of manual annotation of transcripts must be done with PCR to confirm correct sequence alignment. This can be accomplished easily with sequence data generated by this experiment and PCR primers designed to sit at the ends of the sequence. Since all sequences were concatenated in the middle of the larger sequence, verification must only be done on the sequence in the middle of the larger sequence and not at the ends (where the primers would be placed). This allows for verification of manual annotation efforts.

Sequences of interest (Tables 3.12 through 3.15) were related to GH functionality. However, the data generated by this experiment contains more novel GHs than were captured by this bioinformatic technique. Although this pipeline served to capture thermostable GHs from an environment which had not been studied in the past, the GHs found (and especially those of "full length transcripts") would have had to be of certain

relatedness to pre-existing GHs already available to the market. This was a product of using homology sequence matching at the basis for classification (BLAST). In the future, as protein prediction software becomes better at assigning function by characterization of domains and active site position (rather than sequence homology), this sequence data could be used to find completely novel GHs. In addition, not all top-hits generated from BLAST are of thermostable enzymes (e.g. gene_42832) and transcripts found in this study could be thermophilic versions of mesophilic enzymes, an enticing prospect for the bioethanol community (Section 1.11).

GH transcripts generated by this experiment must also be cloned, purified, and classified. Those sequences that were not "full length transcripts", will probably have decreased activity on a cellulosic substrate (if any activity). Regardless, they should still be investigated using these molecular tools. Those sequences which were "non-full length transcripts", which still contain active sites related to GH activity, could hold potentially new domain arrangements which have not been encountered before. These sequences, while not generating a "full length transcript" could hold the key to a truly new class of GH proteins with unknown activities and functionalities.

If a target were selected for initial cloning, gene_36936 should be strongly considered. With a predicted function of a β-frucosidase (E.C. 3.2.1.153), it possesses the highest E-value (5.00E-110) of all full length GH transcripts generated by the unbleached kraft enrichment (Table 3.14). This suggests that it is more divergent from the BLAST hit than any of the other full length transcripts detected in the enrichment, suggesting a possible augmentation in action, specificity, or optimal temperature. Also, the organism from which the top hit was generated was *Marinithermus hydrothermalis* DSM 14884.

This is the only transcript from this organism in the entire dataset generated from this experiment (as well as the only transcript relating to GH Family 32). In addition, it is one of the few "exo" acting GHs in the dataset with a primary substrate of fructose. This enzyme could potentially have many applications, including in the production of high-fructose corn syrup (HFCS), a staple product in American food production. If cloning was done, it is recommended that the functional analysis be done according to previous examples of the study of this enzyme class [78].

Another alternative would be to clone, purify, and characterize a transcript that while lacking in novel characteristics, would hold value for the bioethanol industry. Gene_28236 (Figure 1.14) codes for an Endoglucanase (EC 3.2.1.4) from *Dictyoglomus thermophilum* H-6-12 and represents an example of a transcript sequence of value to the bioethanol community (Section 1.10). The endoglucanase activity would potentially increase the production of cello-oligosaccharides (Figure 1.8) and would drastically increase the rate of the bioconversion process. Enzyme kinetics should be measured and compared to the known kinetics for related enzymes since no kinetics have been measured for the endoglucanase originating from *Dictyoglomus thermophilum* H-6-12.

Further examination is required for transcripts similar to gene_30428 and gene_30429 (Table 3.14) – those appearing with a "**". They are sequentially annotated genes yet they remain separate genes due to their different GH family and predicted function (Table 3.14). Evidence supports their annotation as separate genes, however the closest related organism was *Dictyoglomus turgidum* DSM 6724 for both transcripts (with relatively low E-values). Further examination is required (with the genome of *Dictyoglomus turgidum* DSM 6724 as a reference) to determine if they are part of a

multi-domain protein, operon, or from unrelated sections of the genome of *Dictyoglomus turgidum* DSM 6724.

This experiment created a multitude of information related to the metabolism of both communities examined. Commercial packages (Ingenuity Pathways Analysis, GeneGO, or Pathway Studio) could be used to generate these interspecies pathways [58]. Detailed interspecies metabolic pathways can be generated from the transcripts, especially in the biomass sample (in which cellulose is not the dominant energy source). These pathways would yield a plethora of metabolic data when paired with a deep sequencing endeavour generating reliable phylogenetic distribution of constituents of the population. In addition, mining other enzymes which are active on lignocellulose can be accomplished using the sequences generated in this experiment. For example, the enzyme lignase (or ligninase) is important to the lignocullosic bioconversion because it is involved in the breakdown of lignin (a component of lignocellulose – Section 1.5).

Chapter 5

Conclusion

The mRNA extraction from hot spring microbial communities and bioprospecting for thermostable glycoside hydrolases (GHs) using a metatranscriptomic approach was accomplished yielding 111 total transcripts with GH functionality. These transcripts are of interest to the bioethanol industry through their involvement in the bioconversion of lignocellulose. The nature of this endeavour lead to the detection of active transcripts from a population of microorganisms, circumventing the problems with cultivation experienced in working with thermophiles and cellulolytic microorganisms. Bypassing the need to identity enzymes which are active in the sample, metatranscriptomics allowed for direct detection of enzymes actively involved in the bioconversion process. If other environments are identified to harbour cellulolytic thermophiles, this experiment would be valuable to reproduce.

Appendices

Appendix A: Selection of "Log" Phase *E. coli* Culture

The use of cells growing in "log" phase is important for the study of mRNA. A culture which is in the "lag" phase of development will have too few cells and a relatively small amount of mRNA transcripts. A "stationary" and/or "death" phase culture will have a higher relative proportion of dead cells which have lysed and spilled their cellular contents into the medium. These cellular contents will contain large amounts of DNA and RNA which will be degraded by living cells and subsequently used as nutrients. Therefore a culture in the "log" phase of development will have a higher amount of mRNA transcripts (than the "lag" phase) while still maintaining transcript integrity. Three of the four phases of culture growth can be observed in the graph plotted in Figure A.1.

Table A.1 is data generated from *E. coli* grown according to the Section 2.2. A graph of these data can be viewed in Figure A.1 where a clear distinction can be made as to the different phases of growth of the culture.

Table A.1. Absorbance Generated from an *E. coli* Culture. Absorbance readings were taken at 600nm and illustrate a growth in cell density of the culture.

Time (Hours)	Absorbance (600nm)
0	0.00006
1	0.0005
2	0.0023
3	0.0118
4	0.0363
5	0.0917
6	0.2216
7	0.4162
8	0.7394
9	1.1075
10	1.2855
11	1.3951
12	1.4951

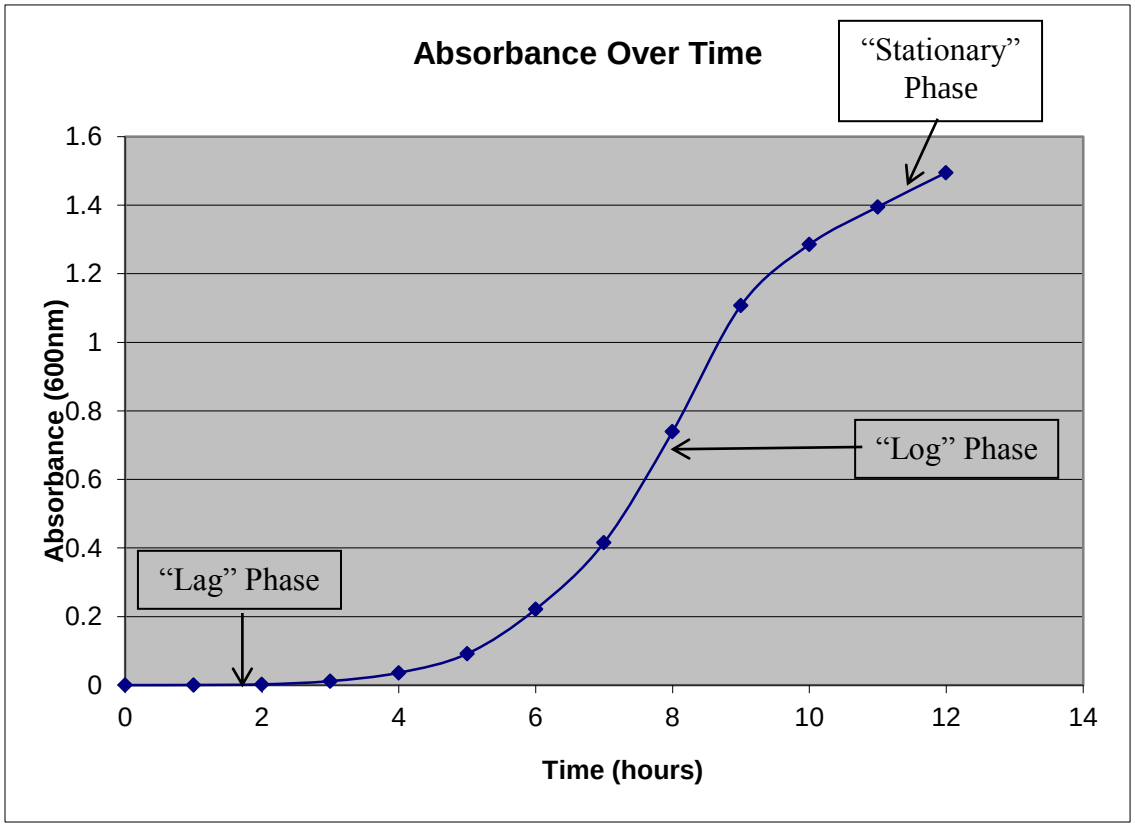


Figure A.1. Absorbance of a Culture of *E. coli* as Compared to Time. Absorbance was taken at 600nm and a standard growth curve can be observed from the data.

Appendix B: Nucleic Acid Sequences of the Biomass Sample GHs

>gene_3317 >contig03295

AACACcTCTTGGAGCTGTCCcCACGGCGTGGAAAGGTGGAGAAGCGATTGCG
GATGCTCTACGGGAGGacTtCCCGGCTGGCATCAAAAGTGGAGAAAgCCcctAA
GAGAAGGACTTGAAGCGGTAAGGTCAAGGGTAAAGGAAgTcgCTTACAACGA
GCTTGATAAGTATTTTcACagTGTGGAAGAAGCCCTTTGGGACTTtGTGgATACC
TTAATGgACCTGCCCATTGaCcAATACTTGGAAAGGCAcTtAAAGAGAGAtgTAA
GCAAGgAggaaaaGGTAAGAgTgctAAAaCTCCTGAAcGCTATAAAATACgCTCACC
tctCCTTTTCTTCcGACGGAtGGTTTTTtGCGGACATATCTGGCATAGAGAGTGTT
AATAACTTGCTTTTtGCAAAGAAGGCAATAGACCTTTTGGGAGAtGAAACTGC
TGAAAAGATGCTTCTGAGGGAGCTTGagCTTGCCCCAAGCAACGTGCAAAGCT
TTGGGAAtGGCTTAGGAGTTtGGAGGCAGTTAGTaTTGCCCAAGGTCTTTACCT
ACagaGAtGTGGCAAGgACTGCGgTAGTTATGtATCTTTCTGAGTTAAAGAaGGA
AAaGGACTACATAGGACATTGGgAgTTTGAAGTCTTAGAGATGAACGAGgtC

>gene_337 >contig01540

aTCTGGTATCTCTTTAAGTCTCCACTCCCAATTTCCCTCCTTGGTTCCGGGAGT
GTTCATCCTTGCCTCTTTTCCCAAGTTTAGGACGTCTTGCATAGGCACAATGC
ATGCTTTTGCCACTGACATGTAAGCAAGCCTTATAAGGGTTGGGCTTACCTCC
TCCTCACTTAgAGCTtTTCCCACATACCTAAAAAGTCTATCTCTtGCCTCTTTtGT
tAGCTCTtCCTTAAACCAaCCTCTTACGGGCagcaTTGTCGTGCGTGCTCGTATAT
ACGAAAAAGTCcCTCTCGTgaTTGTGGGGCATAAAGGGGTGGtCCTCTtcTtCAA
AGgCAAAGACTAATACCTTCATACCGGGAATGCCAACTCcTtCCTTACTTCTTC
CACaTCTGGGGTTATAAACCCTAAGTCTCCGCTATAAAGGGAAGATTAGGAaC
TCTTCTTTTAGcTTGGTAAAGAACtCtCTTGCGGGGGCTTTTACCCATCTtcctcTT
ATGGCTGtcTTTCTCCTGCTGGAATCTCGTAATAAGCAACAAATCCCcTAAAGT
GGTCAAGccttActaGgTCAAAAaGTTTTaGgTTGTGTCTTACccTCTTTATCCACCA
ATCAAAGTCTCTcTCTTTtAGTTTGTCCCAATCATAAACAGGATTTCCCCAAAG
CTGACCCGTAGGGCTAAAGTAATCaGGtGGAaCTCCTGCCACcACATTTAGCTT
GAATATCTCCCTATTTACCCACACATCACTACTATTAAATGCCGGATATATGG
GCATGTCCCCTACTATCCTTATACCCTTTtGGTTTGCGTAGgaCTTTAaaCTTTTC
CACTGcTtGAAAAAtAGGAATTGATAAAAGCACTCTCTCTTTATgaGCTCCTCCA
tACCTTCCACGTAATCCCTCTCCCACTCGTTCCAcGGCTTTTTAAATGTCTTTCG
TaGGGCTTCAAACCTGCAATAATCCTCAACCCAGTaGgAGTTTTCTTCTAAAAA
ATCCCAATActcTttaTTTtCCTCAAAGTTTTTGTATGCTTTTTCAAGGAGTTTTTCc
TTTATCTCCCAGGCTACTTGgTAATCTACCCTATCTCCTTCAAGTTTTATGCCT
TCTATGTCCTCTCTTTTtAGGAGTCCctCCTCCACCAAAAGCTCaGGgCTAATAA
AtAGGGGGTTTCCCGCAAATAGAGAAGTGCTGTAATAGGGAGAGTTTCCGTG
CTCCAAGTGAGTGGGgTTTAGTgaTAATACCTGCCAAAGGCTCTGCCTTCCTTC
TTTAAGAAAaTCCACAAACTGGTATGCCTTTGGTCC

>gene_34538 >isotig02658

GTGGAACTTCTCAAGCGGTTATCGGAGGCCCCTGGtGTTCCCGGGCGGGAAGA
AGCGATCCGCGCCATCGTCCGCGAGGCCCTAAGCGGGTACGTGGACGAAATG
GAagTCGATCTCCTGGGGAACCTCATCGCCCACAAAAAGGGtgCGGGCCCGAA
GGTtGTGGTCGCCGCCCACATGGACGAGATCGGATTTTTTGtGTCCCACATCGA

CGaGgAAAGCGGATACCTGCGGTTGGAACCGGTGGGCGGGTTCGATGCCCGC
ACCCTGATCGCTTCCCGGGTAGTGGTTCACACGGAAACCGGCCCCCTCCTCG
GCCTCATCGGAATTAAACCGGTGCACATCCTCACCGAGGAGGAGAAGAAAAa
GGAAATtcGCATTCAGGACCTTTTCGTGGACGTAGGGCTCCCCGGAAAAGAAG
TTAAAGAGCGGGTGAAGGTGGGAGACCCCGTGACCCTCTACCAAACCTTTGC
CGAGGTGGGGGAACTAGTGTCCGGGAAAGCTTtGGATGACCGGGTTGGCCTA
TACGTGGgCATTGAGGCCTTGCGGCGCCTAAAGAACCACAAGGCCGACATTT
TCTTCGTCGCCACCGTGCAGGAAGAGGTGGGgCTTCGGGGCGCCCGCACCAG
CGcTTtCCAAATCGCCCCGGAAATCGGCGTGGCCTTGGATGTgACGctGGCCTG
CGACATGCCCGGCGTGCaGCCCCACGAACAGGTCACCAAGCtcGGAAAAGGCG
CGGCCATAAAACTCAAGGATTCCGCCTCCATCTCCCAtCCtGGACTtGTCCGGT
TCCTCATGAAaCTCGCGGAGGAAAaGGGCATACCTtACCAGCTCGAGATCCTtC
CCCGCGGTGGCACCGATGCCGGGGCCATCCAGCTTgcccGGGAGGGtGCGGCA
GTGGTGACCATCTCCGTCCctACCCGCTATGTCCACACCGTCGTGGaGgCCGCC
CAcAAAAAGGaCATCGAGGCCGCTATCCAACTTTtGACCGCTTTCCtcGAGCGCT
GCCAcGAAGCCGACCTTCGCCTC

>gene_36258 >isotig01683
cTTttGGACtACctCGAGGTGAAGCCGGAAAaCGaGGCTGTGCTCCGCAAATACGT
CGCGgAAGGgAGGCTCCTCATCGGACcTtGGTATGTgCAGCCGGACGAATTTTT
GGtATCGGgTgAGTCCTTGATACGAAACCTaCTTTATGGCACGCGGATTAGTCG
AGAGTTCGGTAaTGTAATGTTGGAAGGTTACGTTCCcGATGCGTTCGGCcACAT
CGCTCAGCTTCCCcagaTCcTTCAAGGATTCGGTATTAACTCAGCCTACGTCATG
CGGGGGgCGgACtCGGCGTGCGCGAAAGCAAACGGCCCGGATTTTGTCTGGCG
TGCTCCTGACGgCTCGGCCGTaCTTTGTCACGTAATGGAGACGGGATATTGCA
ACGCCGACCAACTTTCGGCTGACCCcAAGAaGgTTCcTCCCcCcGTTGtAAGCCT
GATAGAAGCTGGGCTTCTTAAACCTCAAAGTAaTCCTCTCTtAGGGTTTTTGCA
AGAACTTTCtcGGCGAAGCCGTACTGAaGGGATCCTCCTTATGAACGGTTGCGA
TCACCGAGGCCCTCAAGAAGACCTgCCAGAGGTGGTAGACGAACTAAATGTG
CGTTTtCcTCATTTTCTGTTCATGATCGGTACCCTcAaCGACTACACTAGACTGC
TTGAGCAAACAAGGgACACACTTCCAGAGGTTCGCGGCGAATTTAGAACCTC
CACACGGCACCCCATTCTTGCCGGAGTCCTTTCTACTCGAGTCTATCTAAAGC
AGgCGAATCACCGTGTGGAAACGCTGCTTCTTAAGTATGCAGAGCCCTTATCC
GCGTTAGCACGAGTGCTTGGCGGAAAAGaGTTAAaGCATTTCCTtGATTTAGCG
TGgAAACgCTGCttgaaCCaTGCCcATGACTCGATTTGCGGtACGGGgaTCGATTGtG
TACatCGGGAAATGGagTGTCGcTTtCTCcGGGCAgAAGAaaTTGGTAAAAAGaTA
GTaCATGCaggTctttttGCTCTtGCTCaaaaaTatttGGTAGGGggaGACGTtGTTTCCGTTT
TCGTTTTCAATCCTTGCCCgTGGGAGAGATTAGACGAGGTTTCCGCGGACATC
CCTGCAGAGGCCGTTAAAGGGGcGgTCCTTGACCATGaGGGCAGAGAAATTCc
gTTCGTAGTCAAAGGCAAAAGCCTTGTATCTGAAAGAGTTCTTGGGGgCGTTA
TTCATCATGAAAAAaGCCACCATtACTTTCCcGgCAGCTCTCCCCCCcTTTGGCT
TTGCTACATTTCGTaTCATCCCTCGCCAACCGCAGAaaaAAgCCGAAAGCCTCTT
CaTagACGActACACATTAGAGAATGAGTTCTACAGAGTAGCAATCAGGGAaGA
CGGggCgaTtGATCTCTTGGACAAAAaaCAGGTaTTACCTATCGTGGgCTAAACTT
TttaGAGGACTCcGGGgAcGCAGgTGATGAGTATAaCTTTTCGCCTCCTGTTGAAC
AGGCCGTTATTACCTCAACTGGTCTCCGAGGAGAAATAGAGCGTGCTGAAGA

CGTCCCGTGGAAAGGGTCTGTGaGGGTTAAaCTACGACTGGTGGTCCCAAaGG
AGCTGGaCGCGCATCGAAAAtCTCGCAGCAAAGAAaCGGTCGAAGTGCCTATT
GTTTTTACAGTCTCCCTTAAGCAGGGAaTAaGACGAGTGGAAATTGAGGCTTG
GGTGGAAAACCGCGCCcGtGACCATCGGcTACGTGTAGCTTTTCCCaCCGGGAT
CGCCACGGAaTCCTCCATCGCTGATGAtGCCTTCTGGGTAATaGAgcggccaacgcgc
ccatccccctggtgaagacgggattgaggttccccctgttacacatccgcaaaaggctttgtagcggtcgagaacgagggaaa
GGGTTTTGCCATCCTtAACcGAGGTCTTCCCGAATACGAaGTCACTCCAGATGG
CACTATCTATCTTACACTTCTTCGGTGCGTGGGATGGCTTTCCCGAGAcGACCT
TaTTACGCGCCGAGGCCATgCcGgCCCTCCCTATGAGACTCCGGAAGCGCAGT
GCTTAGGCATGCATCGCTTTGAGTATGCGATCTATACTTaTTCAGGCACTTGGg
AGAACTCCGGAGTtAtACAAGCAGCGCAGGAGTACTCTGCCCCTCCTGTGGCA
aTAGCTTTGAGAGGCGTGCTCAAaCCATGCTACGACAaGATCTTtCACGTAAaG
CCGAATGCCcTTGTCTTAAGCGCTGTGAAGCCCCcAGAAGAaGGGgATGGCCTT
GTGGTGCGGGTCTACAaTCCGACGCGCAAGGAaGTTATGGCCGAGCTCGAAaC
CGCTTTTCCtATAGGCGAAGCTAAAGAAaTACGCcTTGATGAGTCTTCCCTTGC
GGCGATGGTTCCAGCTTCTTCGCACCTCCTGCGCTTTCCCGTGCGCTCTGGGG
AAaTAAAAaCGTTAaGAATCAGCATTAGAAAa

>gene_36543 >isotig03162
GGCTGCCcGGACaGTATCGCCGATCGGAAAGAGCCCGAAAAGCTTGTACATA
GTATAGGCAGCAGGTGACAGTTGGTCGTTCCAAGGGTTGAAAAGGAATTCCT
CGCTCGCCCAAAAGTCTTGAGCCTTTTCGaGAAACCATTTATCAATGAGTGGA
TAAGCATCTTCTCCGTTGTAGGTGAATGTTGTCAGCCAAATGCAATGATTAAG
ACCAGCTACTTGGAACACCACCTTGTTTACATTAAGCCCGAGAATTTGAATTA
AATCTTTAACTTTAGCGTACCCGTGGCACATCCCGACAATCCTTATTGAAGTT
TCTCGTCCAATTAGTGTCGAGCTTTCTAACACAGGATTCGCTGACTGAAGCAA
CCATGCTTTtGGACATAC

>gene_37359 >isotig03305
tCAaTtAAaTTCcATCCaCAAATTCCACAGATTTtAAGTAAGGCAGGTTATAAAtA
TTTCAGATTCTGGAGACCACATTCTTCTTTAAGTGaAAAAAaTATTCCTTATGA
GTTTATATGGGAGGGAATTGATGGGAGTAAAaTTATTtGTtCAAGGGGTTCTtAT
tCAGGTCTCTGTTATAGAGAAATTTTTACAGATAAGGAtAAaTTTTTAAAAAAC
TGGGAGAAAGCAAaGGAAAGATTTtAtCAAATTGAAGTTGAATATGCtGGAAAa
TTATCAAGGACAGGAATTTTATGGATAAGTCAGGGGATGGATGATTTAAGGC
CATTGAgAaGTGTTTATCCATTTGATGAGAAAaTTGATTTAATTGAATTTATGG
AAGAGtGGAATAAAAGAGaAAAAaTACCTTTAATTTTtGGgACACCCATTGATT
ATTTTAAAGATATAGAGAAAGAGAAGTTGCCA

>gene_37914 >isotig03430
TGTCCTGTGGGTCTCTTGGGCCAGGTGGAAAAGGTTcTGGGCCAAGGcgggggg
AAGGTGaGgcccGGAAaGGCGCCACCGCCAATTGCCCTCCGCGCGGCTTGGGTA
ATTCATGCGGGCTTCGCTCCCcAAACCgaGgaCGTCCTGCATGGGGATTACGGC
TAACTTtGCCCGGCTCCTCAATGCCACGGTGATcaGGGCCCACGGGATTTCctCC
TCTTTTCTCGCGGCGATCCCGTGCAATTCCAGGTATTCGCGGaCCCTTTTGCGc
ACcGGTTCctcGgCGgTGCGGAACCAGCCCAGcGCCGTgTCGTTGTCGTGGGTTC

CCGTgTATACCACCACGTTTCCATGGGCGGGAAAGTTTtCGGGAAGATGAGGG
TTATCGGGTTtCCCGTCGAaCGCGAACTGCAAAACGCGCATCCCAGGAAAGCC
GAATTTGTCCCGCAGTTCCTCCACGTctGGAGTGATGACCCCcAGGTCTTCCGC
AaGGATGGGAAGCTCACCGAGTtCCCGGCGCATTTGGgTGAAAAGTTtCTCTCC
CGGGGCGGGGACCCAgcGTCCCCGcACGGCCGTGGGTTCCTCCGCCGGgaTCTC
CcAATacGCAGCAAaGCCGCGGAAGTGGTCGaTGCGtAAAaGaTCGCaGAGGCGC
AGGgCCTGGCGCACCCGGgCGATCCACcactGAAaaCCCTCGGCTtC

>gene_38546 >isotig03654
CcATACACCtATTCCGGATGCTtCACCaTTaaGATGaTTAAaCATtCCAACaTAGGA
AtAAAAaTCTCTCATATTCATGAaTGAaTCAAAaTTTATAAACATATtCCcATTaC
TTAAAACTATCTCTTTTTCATTTTATCTATTACCTCTTTaGCATTTTAaTTtCTTT
TCTGTAAGAAaCTATTTtaTTTtGGCGTTCATGGGTTCAaGaGAAaGacAGActgCAT
CCGGCAATATCTATGGATATAAAAaGTTTATATTACAGAAGcTTACTTATAAC
AaGGGCTCATATAGATAATAGAGGAGCAATTATTGCTGCAAATGATACATCT
ATcTAtAAATTTAaCAAAGACCATTATAGCTAtCTTTGGCCAAGAGATGGAAGT
TTTGTTGCAATAGCTTTGGATAATGCAGGTtATTTTCATATTaCGGAAAAaTTCT
TtAAATTTtGTGCAGATAagATAACAGAAGATGGTTAtcTACTTCATAAaTACTCG
CCTGATGGTTCTGTTGGTTCTTCcTGGCATCCTTGGTCAGATGAAGAAGAGAA
TTTtCAATTACCTATACAGGAAGACGAAACTGCCTTAGTTATATATGCCCTTTA
TAATCATTATAAAGTCAGcAAAAATATAGAATTTATAGATTTTATGTATAACA
AATATGTAAGACCGGCTGCTGAATTTAtGATTTCTTAtAGAGATGAAAAAaCAA
AaTTACCACTTGAAaGTtATGATTTATGGgAAGAAAAAaGAGGTAtATTTACATA
TACTGCAAGCACCGTTtATGCCGGATTAAATACTGCTTCaTTTTTGGCAAaTAT
AACCGGTAATAAAGAAGAaTCAATTAGaTaTAAAAAAGCATCAGAaGAGGTAA
GAGAaGCAaTATTAAAATATCTTtATgA

>gene_40561 >isotig04529
GAGGTAATCGCGCACTGCCGGATTGTCGTGGTTCCATTCCGGAAGGTTCtCGC
AGCCTGCCCAGCACGaGTACCCGGGCTTTTCCTCCTCCCAAACAAACCAATCC
CAGTAGGGCGAATCTTTtCCccGTtCCTTtGCGGCgCGGAAGAAaGGgTGTCCGGC
GCCGCAGTGgTtGAAAaCCCcGTCGAGGATGAggCGcATACCCCGCCGgTGtAGC
TCCTCCACCAAAACTTtcaGGgCCTCgTtCCCGCCCAAaGCGGGATCAACCTCGA
AATAGTCCTCCGTgTCGTAACGATGATTGGTGgAAGCCCGGAAAATTGGGGTG
AGATAAAGaGCaGTTACCCCAAGGGCctCAAGGTAATCGATCTTctCCaGgATCC
CCCAAAGATCGCCCCcGAAAAAGCTGTGCGGcGTGGGAAGTTCTCCCCAGGG
CCTCGTTCCCggaggATCGTTATCAGCCCGGCCGTTGcGGAAGCGgTCCGGGAA
AATTtGGTAAAaCACGGCATCCCGAACCCaGgatgg

>gene_40746 >isotig04566
CTCGCGgTATAGaGGAAGAACGGACTTtACGAGCTCCATCTGcACTTgcGCCAG
GATCTCGAGgCGCTTGGTCGCCGGCCAGGAAAGtACGGATTTTaGgcGCTCATC
GTCcTCAAGCGCGGTTTGcGCGGTCCAAGAAAGATAAAACCAGGCCCACaaatttt
CGTCGGAGGAAGGGGAAACCGAaGGAAAGCCGCGCCGCGCAAGGGGCATCG
GGATTAGTCCGgAACGTATCTCCTTAAGTtCTTCCCCAGTTCCTTCcGGCGTAT
ACGGgTCTTTGtATTTTtGCGCGCAAAGGAGGCGAAGCTGCTCGGTGAGGgAGG

GGGTGAAGTTCACCGTGCaGGgAAAGCCGGTGAGGCGAAGATTTTGGGCCAT
GTCGTAATAATCcTTGGCCGCCCGCAGGCGAACCCACGGGAGCTCGGCcTTCT
CgccccccGGAAGGAGGTACCACGGCTGATGCATATGCCAGAGGAAGGCCAC

>gene_41054 >isotig04636
AGCCCGATACCCcAGTGGAGCCCcAACCTCTtCGTtATCGCCGCCCcTGGCGCG
CGCTCCGTGGCCGTAGTCGGCGATTTCTCCGCTTGGCAACCTATCCCCCtATCC
GATCCCGATGGGGACGGAATCTGGATGGCCTCCATCCCCCTGCCCCCcGGCCG
GTACGAATACGCATTCATCGTGGATGGAAAaTGGGTGGGACAGGATCCCcTGG
CTGAGGAATATCTGAGGACATTCGGGGAATACGTGTCCGTGCGCTATGTAGG
AGGGAACCGTGAGGGCTCTTCT

>gene_41570 >isotig04785
aaGCTTTTtGGGAaGaGGGCGGTGGgtGTTTGGCCGCCGGAGCAGGCCGTGtCCG
AAAAaGCCGTGGCCcTTTTGGGGCAAGCCGGCTTCcTCTGGACCGTGGCCGAT
GAGTGGACCctGAGCGCCGCCCTGGGCCGCGGGCCAAgCCCGGCGGAGCTGGC
CCAGCCGTGGCGCCTGGGAAACATCGCCATCTTCTTCCGCGACCACACAATTT
CCGACCAaATCtCCTTCTCCTACGGcAacAAGCCcACTTCCGAAGCcGTGgCGGA
TTTCATGAAtcGGgTTCGCCGATATtGGGAAGCCCTTCCTAATCCCGAAGAGTA
CGtCCTTGTGGTGGCCCTGGAcGGGGAGAACTGGATGTTCATGGCTGGGTACC
CCGATAACGGCCGGGAGTTtcTGCGGgctcTcTATCgTGCCCTCCTTCAGGCGGAT
TGGGTGAAGACCGTGACCCcTGGGgAATTCCTGGCCCAGCATGGGGtCAAGGC
CTCCCTTTCCCGCCTTCCCACCGGctCGTGGgCtGGcGACCtCTCCACTtGGCAGG
GCGAGCCtGAaGAGGACGAGGCgtGgAAGGcTTtGGCCCAGGCCCGCGAGGCCG
TGTtCTCCCGCGACCCcAAtCcGCGGgCCCTtCAAGCCCtCTACGCcGCGgAGGGct
CGgACTGGTTctGGTGGTACGGCGACGACCAGGACTCCGGGActGAtGATCTCTT
CGACTGGCTgTtcAAGGCCCatcTGgTGGGcGCCTATCGCGCTGCGGGCTATGCG
GACGAaGAAATCCCcGCTTCCCTTTtCCTGCGCCTGAttGTGCCcTtCCGCGCAAa
CTTGGGGgAAGtcAAaCCcAATTGGATGGAAAAATCACgTCctCcGGGGgAGTGG
GCCGAGGCCGCGGTCTACCcGGGCGCcGGGGAGATCAAAAGCTtcGCcGTaGgT
TACGGTGAAaCCTACCTCTATGTgCGGGcGGATCTtGcGAGGCCagCgCGGgAGC
TGGTCGGCTCGGATCTCCGCCTCGtCcTCTATGCCAGCGGAAAGCctagcGaCAAa
GTgAaCATCcTCACCCGgCACGCTGGGGTaCAGCTTGGGTtCGCGTTTGCCTCCG
CGGTGGAGCTGgAGTTCGCgAAGATCCGGGCCGATGGAACGGGGTATGTGTT
CCGCTACGGCGCGGATGGGAAGGGCGGATGGCGCTTAACCTCGCCGGTTCGA
ACTCTCCTTTCTCGCAAAGCCGCCGTGGATGAAGTGGTGGAATTCATGGTGCC
cTTtGAGGAGCTCGGGGTGgAACCGGGGCAGAGCCTCACCGTGGCCGtGGTTTT
GGAGCGGCCGGGCGAGGTTTTCGGGAAAGCTCCGGAGCGGCCCcTCTTTGCC
CGcATCCCCACCCTcATCCgCGGGACGGAGATCTGGGCCATGgAGGATCCCAA
GGGCGACGACTTTGGGCCcGGAACTTACGTGTACCCcTTGAACAAGGTGTTTA
CGGAGGCgCCcGGGCTTTTCGATCTCGTTCGCTACGCGGTCTATGATGCTGGC
AACGCCTGGCAACTGGCGTTCGACTTCCcGGCCCTTCCcAACCcGTGGAACGG
CCCG

>gene_42832 >isotig05247
TATCGGTCGCCCTCGCTCGTCGGCGCAAATCCAACTGTCCTGACCGGCGTCAA
ACTTCGCCCGCTGCTcTTtGctCAGTTCGCTTTCCGGTCTCGGTCCGTGCGGGAT
CAagTCAAGCAAAACCCTCATGCCCAACTTGTGGGCTTTTGCGACGAGGGCTT
TCAAGTCCTCAGGGCTACCGACATCGGGGCGAATTGCGAAGTAGTCAAGGGG
AGCGTAGACGCCGGGGGTGTATGGCGGCAGCCAAATCGTCTTGACACCAAGT
CGcTTCAGGTGGGgCAGCAGACGGAGGATGTTGCGGTAGGTGTGTTGGATGG
CGGGTTCGTAAAAaGCGCCGTCAAAAaCATCCTCCGGCGTGTCGGGCAGCGgT
TTCAAGCCGAGCGCATCGTAAGCGTCCCAAATTCGCTCCAaCAAaGTTCGCTC
GCCGCCTTCCGCCAcGATGAAaaaTtGCGTtCCCAGTTCGGCAACCCAGCC

>gene_43551 >isotig05576
GaTTGGATCACTTCTGtCaCAGTGgctGGTGAGTTCAACGGgTGGAatCCCACcGct
ACTCCCATgaCTTACGACCCGgCcACCGGGCTTTGGTCGGTGACCGTTCGCCTT
GATCCGGGAACGGTGCGCTACAAGTACgtGATAAACGGCAtGGGTGGgAAaCA
TGTCCACCGgTCACCCGgtAggcGGCGGGCCgtGGATCCTGcGGATGGTTACGTG
GACGACGGGtaCGGTGGCCAAACGCtgTGCGGATTGTGCGGAGCGTGTGCAAA
GCTGACCTCGTTCCCGTGACCTTCCGCTACaTCCCGCGGgaTGGGCAGAgcATT
ACGCGGATTAGCGTTCCTGGGGCaTGGCCGGCGTGGACAGGATACGGctGGTG
GaTCCAgCATGCTCCGAGCTTTCTCtCAGCCCGACGGCACTTGGAGCGTGACCC
TCATGCTCCCGCCCGGCACTTTCATCGAGTACAAGTACTACATcgAGgACAaCT
GGGGGACAGGTTGGGTCCCGTGGATGCCGAATTTTGCGGGTTCACCT

>gene_43916 >isotig05739
CTCTTCACCACTTTCCATTTGGTAGACCCGCACATAATCCACATACATTCTTG
CaGGGAATGGTGTCtCGTTTGTCGGATAACCGGGCCAATATCCACCAACTGCA
AGGTTCAATATAATGAAAAaGTCGTTATCGAACACCCAAGCTTTtCCACGTGA
CTCTATCGCTGACTTTGTAACGATGTGGTAAACTgTGTtGTCAACATACCACAC
AaTCTTTTCCTCGTCCcACATtATTCCAAAAaCGTGGAAATCACTTGTAAAaTCT
GGTGTTGGTGGTTCCAATCTTATCCTTCCTTGAATACCGTTTGAACCAAAATA
ACCAGGCCCATGCAAAGTACCATACGCCGTCCATTTATCGTGTCCcAAGAATT
CAACAATATCGATTTCTCCACAAGTTGGCCAACCAACATACCTGAAGTTGCTA
CCAAGCATCCAGAACGCTGGCcAAAGACCTTT

>gene_44066 >isotig05811
GGGGTTTATATGCTCAAATACCCTGTCTTTGGATTCCAGGGTTTCCAAGAAGT
TTCTaCTTTCTTCCCTCAAAaCaCctGTTCCCTTTATTtCGTCTAAAaTGtCcAGCAA
CCTTTTGCAATACTCGTAGTGTATTGAAAAACGCCTTTCCGCATAGTCCCGAG
CACTCCAGGTTGTTATCAAAAATTGCCAATCGGAGGATTGCAACAGTAAAAG
TTCTCTCCCcAACTGCTTTAAAACCCTTtCGTAAAGCTCATTATTACGGTATGC
CTCAACCAGAGCGTACTTAAaGAaCCTATCTTCAATCTCATAAaTCTTTtCCcAG
GTCCAAGAaGTCCAGTCGTTTAACCAGACCcAGTGGAAACCACCcTCACC

>gene_44068 >isotig05812
CTtGATGCCTATATtGATtGTCCATGGAGAGAACAGGCACAATGgTGGGgCGAT
GCGAGAATTCAGGGGATAAATACTTATTATGCTTTTGGAGATATGCGTCTTCT

AAAGAGAGGAATAAAACAAGGAGGTCAATCTCAAATAGAAAATGGTCTTAC
TTACGGACATTTCCCAACAATAGCAGTTAATTGTATATTACCCGATTTtACATT
GACATGGATTGACACACACTTGGATTATTATCATTTTACGGGTGAAAAGAGTT
GATGAGAGAACAGTTTGAAAAATAAAAAAGCCATTAATTTTTTtGTTGATTTT
ATTTCAAAAAaTTATTTACTTGGTAGTATGCCTGAATTCTGGGTGTTTTTAGAT
TGGGCGCCTCTTTATAAGGAAGGATTTTCTGCAGTTTtCAATATGATGTTTTtA
AATTCTATAAGGACAATGTTAAAAaTATGCAGAATTTTGAGAAAATATGAAG
ATTTTAAATACTACGAAAaGATAGAGAAAATTGTtAAGAATAGAATAATAAAA
ATTTTCTGGGATAATAGGGAAAAAATTTTCTGGGATGGATATGATACTAAAA
ATAAAGAACAAATAAAAAGGATTTCTCAACACTCCCATTCTCTTGCTATCTTA
CTTGATTTAAAACCACGGTATCATAAAGATTGGATTGAGAATATTTtATTACC
ACCAATGAAaTTACCACCTTTACAACATGAAAAAaTAGTAGAATGTTCTCCAT
TTTTttATTTTtACACTGTTGAAGCGATTAAAaGAGTAGGGGggTATGAAAaTGA
AATAATAGAGTTTATAAAAAGAAGATGG

>gene_44261 >isotig05908
AGATTTTAGAAATTCCTTCAACCACAACGCTGAATCTTTCAATATCCTTTTTTG
CGTATTGTAATCTACGTAGATTATACCGAAACGTTTTGAGTATCCGTACGCCC
ATTCGAAGTTGTCCATCAAAGACCAAATGAAGTAACCTTTCAAATCAACACC
TGCATTGATTGCTTCAAGTGCTTTTtCAAAGTGCTTTTCCAAaTATTCAATTCGG
TAATTATCATGAACTCTTCCGTTTtCCAaTTTATCAGGTCCAGCCATCCCGTTCT
CTGTGATATAAAGTGGTAGTTTATaTCTTTCCTTCAGATAGACCAGCATATCA
AATAATCCTGctGGgTAgaTTtCCcATCCcATCTCcGTTTtGGGAAGgTCTCCCTGA
ACaTACGAAAATCCAAGaGGATTGTTCATATCAAAAACAaCAAGTGTTCTTGTg
TAATAATTCACaCCAAAGAaGTCTATAGGAGTCGAAATAaTaTtCAtATCGCTAT
CGGGAACTtGCAACCcTTTTTCcGTATAAAGTGCAACTGCTTCTTCGGGATATT
TtCcGAAAaCAACAGGGTCATGGAACCATGCaTTAACAAaCTTATCAACAAGAC
TTGCGACCAaGAAaCTTtCGGGTTTTgCatCGCCCGGTTctATTTtCaTCACAACgTT
GGTTAAGCCAACTTTtCCATCTTTtaC

>gene_44279 >isotig05916
CTTTtGAATGAATTGGGtATTTcTCCAGATTtCAGTTATGAACCAAGAAAaGAAT
TTGGTAGTATAAGGtATATTCACAAGACACTTCCAGAGGCAGAAGTATATTtC
GTTACCAATCAAGAACCTATTGAAATAGATATTAATTGTTCTTTCAGGGTTAA
GTGTAGATTTCCTGAGATATGGTTCCCTGAAACTGGAGAAATTAAGCGAGTG
GAAAGTTTTAAAGAAAAGGATGATTGCACAATTTTACCACTTTATTTAAGGCC
CTATGAATCTGTTTTTGTTGTTTTTCCAAAAGATAGTTCAAGTGAAAGGTGGA
AATCTTTTAATATTTATAGAAAacTtcAaaTAGAGAAAGAAATACCTATAAAAG
GTGAATGGAAAATTTCCTTTCCAGAAGGTTATGGAGCACCAAAaGAAATTATT
TtACCGGAACTTATATCTCTTtCAGAACATAGTAATGATGGGGTTCGTTTCTTTT
CTGGTAGTTTtGTTTATGAAAaGGAaTTTGAAaTACCGGAGGAaCTTTTtGATAA
AAaTAAGAGATATTTtCTTGATTTAGGgAGaGTTGAAaTAaTAGCAAAAATAaCA
CTTAATGGGAAaTATATTGGAACCCTATGGAAAATTCCATATTGTATAGATAT
AAGTGATGGGATAAGGATTGGAAAAAaCaTATTACAAATAAAAGTTACAAAT
CTTTGGGTAAATAGACTTATAGGTGAT

>gene_45023 >isotig06301
GGAAaGGAAGACGGTGTCCGAGAGGGCCtCGGCGAGGAGGTCTCGGCCTTCG
TAAGTGAGGGCGAAGGTCCACGATAGGTGGTtCAGGCCACGGACGATGAAGG
TGCACGAGTTGCGgTCCAGGCCTGCCCATTTGGAGATGGCCTCCACAGCCCCG
AGGGCAGAATCGCAAAGGCCGATTGCGTCGGAAAAGCCCTCGTCGTGGAGG
GCCTGGGTGACAAGGCCAGCAGGGTTCGTGAAGATGATGGGgAGGCAGCGG
GGgCTTAGgCGCTTGGCGAGGCTTACCTGTGCCAATACTTCAGGGATCGAGCG
CCAGGCAAaGGCGAAGCCTGCAGGACCGATTGTCTCTTGTCCGATGGTGCCG
AGGTCCAGGCATGCCCGCTCATCGTGGGCGCGGGCCTCCTCAAAGCCAGGGC
GGATGGTGAGGATGAGGGCCTCAGCGTCACTGAAGGCCTCTTCGGGGGTTTTt
gCGATTTGGATGGGAGGgAGGGTTCCCAGTTCTGAGGCGAGGGTGCGGGCGA
CAGgCAGGAGGAGGTGGACCCTTTCtGGCTTGACATCGTAGAGGGAGAGGAG
GGAGCAATGGCAAGTGGAGGCGTGGgCGCTTAGTGAGGCGAGCATGAGGGGg
AGgCGGGTGGAGCCtGCACCTGAGATGGCTATTTtCAT

>gene_45198 >isotig06385
CAGCCCTTTGAGTTTGACcACCAGTTGCTCGCTGGGCGATTGTTCCCTGAAAG
CCAAGACGGCTCCGCTGTCGGTTTCGGGGTCGTGAAATTGCCACACCGTCCA
ATCACCCAACGACCTCGCTTGCGGCGTCAGCAAAAaCCTCcGCTTCATCACGA
ACGGTCTCAACCGCTTGTATTCGgCGATGCACTTTTTCGCCCGCTCCCTCGTCG
CTTCGTCCCACTCAACAAGTCGCGGCGATATCCCGAAAGCGCCACCCATGCG
GCTCCGGAAATCGGCGAGGgTtGTCCCTTCGTGCACGACCCATGTGTTCAGGT
AGCGGGgCGGCAAAGCGAACCAaGCACCTGTCAAATGGAAGCGGACATTTGC
GGCGCGGTGGGTGAAATCGCTGAGCCAATAGGTGTGGACATGCTTCAGGATG
CCGAAATCAATTCTGTTGCCCCcGCTGGCGCAACCCTCAATGACCAGATGGGG
AAACTTCTCCCTCAAAAaCTCCTTGACAGCGTAAACGCCGAGGGTGTGGGCAT
AAGTCCcGTCGCCCTCTTGGTGTCCGTGCGTCGGGTCGTTGCAGCCCAAGCCG
TAGGCGATGTTGTAATCCCACTTGAGCCAATCAACaCCGTAGtCGGCTACAAC
CcTTGACAATCGCTCTTtGAACCaCTCAACGACTTGCGGgTTTCCGAAACaCAaG
TGAGC

>gene_45258 >isotig06420
GAAGGGTATGTGTTGTTCTCCGCTTAGAGCGGTGGGTGATGTGAAGATTTTTA
TTTCTTGCTTGAGTtCTtCAGGTACAGAtCCCTTTGGCAGGAATTTAAGGTAAGA
CTCTAAAGAAaGGAGTTTGATGGATTTGTTtGTCGACAGGTAATTTATCGCGTC
TGTTATCATCTTCAGCCTATCTTGATACCATTtCCATCCAAGTTCTGAGTCGGC
ACCTCCCcATGGTTCTCCATACCATTtCTtCTTCACCTCAAGGTTGTAATTGTAT
TtCCATGAACCGTTCCTTGGAGTATCACCTATGGGCATCTGACCGTAGAATTT
GTACATATCAATGGAAGCGGGGCTCAATTGCAAATCAAAGGGTGTTTTtGGTT
tCCACGATTTTGAATTCTGTTCAATCCACTTTTCGACGAATTGATAGCCATCGC
TTTtGTCGTAGAGAAATCTATTGAGCCAGATAGCGTGATTGAATCCGGCTACC
TGCCAGTCAACTTTTTTCATATCCAGTCCAAGAGCTTCGACCATTTCATGAAC
GCCGTAGTGTCCGTGGCAAAATCCCACTACCTTGATTTTTGAACAGCGATTTA
TCAGGTTTGTACCCTCGAAGACGGGATTTGCTGTTTGCAAAAGCCACGCATTT
GGACAGAGTTCCTCCATTTTTCTTGCCACGTCAACGAAAAaCTTCAACTGGTT
AAAGTT

>gene_45382 >isotig06485
GAAACAGTATtCACCTTAGGAAaTGGATATTTAGGAGTAAGAGGGGTAGAGG
AAGAAAaCTGGAACCAGGAAGAGGTTGGTCTATTTATTAATGGGGTTTACAA
TAGACCACCTGAGGATGTTACAGAATTAGTAAATCTTGGGAATCCCTTTGAA
GCAAATCTTtATATAGATAAAAagCCCTTTTCTTTtAATGAATACAAAAGGGTTC
TTCATTTAGAAAATGGAATTTTGGAAAGATACGTTTcTAATGAGAAGGTAGAA
GTAATTACAAGAAGGTTTGCATCAATGGCAGATAAAAATCTTTTATTTtGGGA
AATAGAGATCACTGCAAAGGAAAATGTAGAAATAGAGCTTGAATCATTGATA
AATACAGAAAGAAGATTGGATATGACCATCTTATTACTAAATTCAAGATGAA
AAATTTTtAaTTTGAAGACGAAGGAATCAAATaTTGAAATaGGCTTTCGTTtAGA
ACaTGAGGGAgAAAaTTTTAAtGAGGAATCTTTGATATTAGAGGAGAGGAAAA
aTTGGGAAAAAgATTTtCACACCTATTTAGAAAGAGGTAAGAAAGCAaGATTA
AAGAAaTACATATGGGTAACAGAAAaCAGAATAGATGGAGAAaaTCCCTTAGA
AaGATTAAAGAGG

>gene_45450 >isotig06526
GAGCCCTTTGGGCGgTTCTCGCCCCAAAGTCcGCCAGGCCTGGGcgAGATGGG
CCCGGAAGATCCGGGCAAAaTGCggagCAAgCGGGgTGtGGTTGCGcGagCTTCG
CCACCAgAaCCAATCGCTTCCctCCGCCACaTAAATaTGCTCAAGGGCTctCatgCG
CAGATCTTTGTcGgAaCAGCTTTCCACtGCaTCtcGGGCTTCGGCGAGGGCTTGCC
ACGCCAAaTTTTGCCGGGGCTCGCCGATCCAGGTGCGGAAATCCGCaTCGATC
CAGGAGCCTGACCAAATGGTGGGgaGTTCCCCGGcggCTCCGTAgCGCTCGATG
TACTCGCTAATTTTtATGGGgACGAAGTTTtcGgCGATGGCgCGGTAAAGagcgaG
GAGGAAAGgCCGGCCATTGCGCTCGTAATAATCCCaGgCATTTTCCCCGTCCA
TGGCGATgaGGATGAGGGGCGGGGCGCTTTCGTGCCAAGgaGAGGCagAAGCG
GCgaTTTTGGCGACTAagTCCGCcACcGCCTTCtCCcATGGCCATTTtCCGTACTCG
AAGCTAATgAGGTTGgAAaGGATTGTGTCCCGgaaGACCACGgTAATCTCccGAT
TTCCGAACCGGAGGCGCCAAGGgCG

>gene_45820 >isotig06740
cTCCGCACCGACCGgCTCCTCTGCCGGGTtgCCGCCGACGGCCGCATCGTCGTC
CTGGACGCCGACGGCGCGGTCCTCGCCGAGGAGGCGGCGCCGGTGGGCTGG
ACCGCCGACGGCCGGGCCGTCCTctGCTGGCGCATCCCCCCcGaCGCCCACTTC
TACGGCCTGGGGgAACGGgCGGCCCcGCTGGACCTGCGCGGCGGCGTCTACGT
CAACTGGAACaCCGACCCCCGcACCTACGGCCCCGGCGACGaCCCCCcTCtACC
TCTGCGTCCCcTTCCtGCTGACGCTCCATTCCGGGGGACGCGGgCCTTcGGCCT
GCTCCTGGCAAACTCCGCCCGCAGCCGGCTGgACCTGGGGCGgaCGGAGCCGG
AgACCGTCcGCATAGAGGTGGAGGACGAGCGGGTGgCGTACACCcTcTtCTTtgC
CCCc

>gene_45835 >isotig06750
ATTTtGGTCGCTTCCGGGAGGGACGAAAAaGAGGCAAGAAACCTTCTTGACCT
ACACAACCCCCAATACTCATCCGATGCCCGGCTTGACAATAGGGGGCTCATA
GGGgAGGTATTtCTTGAGAGCAGACCTCGTGGGgCACACATAAGGGATGTTTT
CATCCAGACATCCGTGCGGAAATGGGCTTTGAGTTtGGAGGTGGAAATTGAA

GGCGTAAAGcAACCCGGAGAGGCGGAAGTCGTAGCCCGTGTCCTCgaTGAAA
AaGGAaGAGAGGAGAAGCGGTTCACCTCAAAAGTGAGGTTGAGCGCCTCTCC
CcAACAGAAACTTTACCTCTCTTGGCATTGGAGAAATCCCcGTCTTTGGGATTA
TCGCCAGCCTAACCTTTACAAATtGCtCcTCCAAATCAAGGGCAAGGGAATAG
ACGATGAATGGGGTGGTATGTTCGGGTTCAGGGAGTTCTGGgTTGAaGGgAAG
CGcTtCTTCCTAAaTGGGACGGAGATTCGCCTTCGCCccAAGACtgCcgATGACCA
ATGGCAAACCTGTgCCGGCGTCTTGGAAGTCATaGAcGgTATGTtGGATGCTTA
CcTTGGGATGGGCTTTAACATTttagAACTTTGGCCTTGGAACCACGATGAGCGc
GGA

>gene_45889 >isotig06782
TTgaaaaTCaCGGTTATTGGTGGCGGAAGCgTCTAcACCCCTGAGCTTATCGAGGg
CTTCCTTTTGcGTtACAAGGAGCTCccAGTGGAAGAAaTTTGCCTCATGGATATC
AACTACGAAAGATTGCAAATAGTTGGGAAGTTTGCAGAACGCatGATTAAAGC
AGCAAAAGCTCCTATTCGCATTAAGCTAACAATGAgTTTGGAAGAAGCcTtGA
TGGGGGCGACTTTTGTtATCTCCCAAaTACGCGTTGGaGGCAtGGCTGCCCGCA
TCCTCGACGAGAAAATCCCTCTCCGATTCAATGTGATCGGTCAAGAAaCTATA
GGGCCCGGAGGCTTTGCCAaGGCCTTAcGCACAaTCCCTGTCATGCTTGaTATC
GCTAAAAAAaTGGAaGAACTTTGTCCAAAcGCCTTCCTTATCAaCTTCACCAAT
CCTTCCGGCATTATtACTGAAGCCATCCTCTCTCACACT

>gene_46259 >isotig06950
AAaTTGATtGAAGATATAGagAATAAACTTGAGATTTTTAGACCCTATGCAAAG
AAaTTtAAGGTtCATCTTATAGGtcAtGCCCATATTGATATGAACTGGTtaTGGAC
ATATGATGATACaGTTTCTGTTGTTTTAAGGgATTTTtCAACAATTaCAgATTtA

>gene_46259 >isotig07012
TTTCTTATATTTTGGTATGACATCACCTTCCCAGTAAAAaCCGAAaTTTAACTT
ATTTTCAAGAAaGTGAtATATGCTATATAAaCAGGACCTATTATTTGAACCAGC
GAAAaCAaGAAAATTTTGGTTATTTTTTTCAAAaGATTTGATTATAAAaCAATC
ATATTCAAAAAATTCTTGCAAaTATTTAATATCCTCTTCTGGAaGGATTTTCTT
TATTtCaCTGTCAtCCTTTGTCCCTaTAATTATtACATTTTtCTTTTTTtGaTATCTCA
TAACAGTGGGAATCTTAACTCCTGTGGaTTTTtCTAtGAAGGTTATCAATTCAG
AAGCAGAAAGTTTTTGAGCATAATTATTTTCATTACAAATATAAACCACTGC

>gene_46489 >isotig07152
ACCCGCTTCCACAGCATCCGATATATCAACCGAACCGTAAGGCCATTCAATC
CTCCCGCACTCAACGCCGTTAACGAAAaCAACCGCAtCCGTGCTCACCCGTTC
AAGAGTAAGCAGGATAGcccTCCCCTTCCAATGTGCCGGTATGCTCACTGTCTT
CTCGTAAAACGCCCTCCCGAGAGAGGAAAAGTCAATGCCTTCCCAAGCAGGA
CCTTCTCCCTTGCTTATTATGCCGGGAAGCCTCCAACTACCGTTCCAAGCCCc
GGGGACCCAGATACTGCCCCAacTTGATTGACTAACGGACtCcAAATCCTTTGC
AGGCAtGAATGACCATCGCCCGTTTAGGCAGATgCGGTCTCGCTTGgAGGAGA
TTGCCTC

>gene_46625 >isotig07235
GTCCCTGACGGCTACGGCGGATTCAACTCCCAATTTGAACTAACCCCTGATG
GAAGGATAATTTATCAGCCTTCGACCCCTCCCACACCCCcAACGGAGTTAAAT
TGGAAGGGGCCCGAAGCTGTTGAAGACGGTGTTATTTtCCGTTtCTATGAACCA
AACGCGCAGATAGTTTACCTCGCCGGGACTTTtAACAaCTGGGCTCCAACGGG
gCTCCcGATGTCAAAAACCAAaGACGGTATATGGGAAGTTAAGGTAAAACTTA
TCCCGGGTAATTATCAATACAAATTCGTTAtAAATGGGACTACCTGGAAAGAA
GACCCCACAAACCCTGCAAAAGTTGATGACGGATACGGCGGTTTTAATtCCGC
ATTCCGCcTTACCGATGATAACAAAaTAATCCTTGAAGctGTAACCCcTCAGAG
C

>gene_46781 >isotig07337
TCCAAAaTATTCcGAAAaTCTTAGTATAATTCCATCCCCATCCTCAGACCTTTTT
AAAGtGGAAATACAATATTGGGAgCATCCCAATTTATGAAGGTTTCCTTCGTC
GCCTTTCCTTAAAACCTTTGAAAATAAAGGTCTATTtAACTCTTCTGCAAaCTT
AATACAATCCCTCTTCCAATCACTACCATGCAATAAAATGGAATAGATAAAT
TTATGTTCTCCTAAGTCTGCATAGAAATCAGGAATAGCAGGAGACCTTAGAA
GAGTTAATCCTAAAACATTCTCCTTAATGCTATGTCCATGCTTCCCATTATTTA
ACAAGCTTATTCCAAAATCACCTTGAGAATAATTAATAAATCTTTGCACTGGA
AACTCAAAcTttGCCTTaTCCCAAGaGGTATTCTCaTGGgTAaGTTtaTAaTAAaTCC
cATAaGgTATCTCaTAaTGACAaaCATaTgAaTTGAtATTtAaTGGgAATAAaGCTCtC
AaCaTcACTCTCTTCTCATTCCA

>gene_46988 >isotig07465
ATAATAACTCCTCTATTCGGTGATATAAATGATGATGGAAAAGTTGATTCAG
AAGATTTAATAATTTTTGCTAAAAGTTTTGGAACTCAAATAGGTGATCCAAAT
TAtAATGagAAGTGTGATCTAAATGGTGATGGAAAAATTGATGCTGAAGATCT
TCTTCTTTTAGCTAACAaTTTTGGA

>gene_47005 >isotig07477
ATTATATAAATCATTAATATAGTTTgATACAATTTTATCAGCAAAGAAGTTTG
ATGCTGGATAAAAAaGTTCCATATTGTGTGCAATTGAAACTAAagAATCTTTGT
CCtCCTCTTTAATTATTTTAAATGcTTCCaTaTGGCTAAGAATCAAATTTCTtGCA
ACTTTTAAACCagTtttAAAaTTTTtAaTtCCAGgAGGCCACCACCcAAAAATATATC
CTAAaACTGCAAGAACAAGTGGTTCATTAATTGTTATCCA

>gene_47009 >isotig07480
ATCTATTATAGTTACAAtcTTAaaCCcTTCCCTATTCAAATCATTGATCATCTTTt
CAAAGTTTGGAAATCTTtCACagCTtGTagTAAATACCCTAAATCCCTCCATGTA
ATCTATATCAAGATAAATAACATCACAGGGAATTTTtCTCTCTCTAAAGTTTTT
AGCAATCTCCcTTACCTCTTCCTCTGaGGCaTAACTCCATCTGGACTGTTGATAT
CCCAATGCCCAAATAGGTGGAAGATagTATCTTCCTGTAAGATAGGTATACTC
CTCTAtcACATcTTtAGGAGTTGGTCCATAAAaGAaGTAATAGTCTAATTCTCCA
TTCTCTGCATAAAAaTaGTAGTATTtATTActCTCTTTgCCcATATCAAAaTAACTT
CTAAAGGTGTTGTCAAAGAATATTCCATAACTTTTTTTGGGATTCCAGGCAAT

gaaGAATGGATGGGACTGATACAAaGGATCTGTAGTAGGATAaTGgTTTGGATC
ATCAGTATTCCAATTTATTAATCTTTTtCccTTCTTGTTTAG

>gene_47799 >isotig08008
TtAAAATTtCCGGAAGAAttCctGtGGGgAgCCGCCACATCAGCGTACCAAATTGA
AGGTGCGTGGAACGAAGACGGAAAAGGACCGTCAATATGGgATATGTTTACC
CATACCCCAGGCAAAATCCGCGATGGATCCACGGGGGATGTGGCCTGCGATC
ATTATCACCGTTGGGCCGAGGACGTCGCTTTAATGAAAACGCTGGGCCTTAA
AGCTTATCGGTTCTCAATCTCTTGGCCCCGGATAATTCCTCAAGGTCGTGGGG
CATTGAATCAGAAAGGATtGgATTTTTAtAGCCGCCTTGTaGATGCCCTCTTAGA
AGCAAGCATCGCTCCCTTCGTAACT

>gene_47977 >isotig08136
TCTATTGGAATTACTTATTATtAGGTTTTTtATCTTTACATCTTGAACCTTACTA
TCCAAATGCCCTATtCTTACTCCTGCAGATCTTGTCTCCATAAAACAATCACTT
ATCATTACATTTTCACAAGGTTTATtCcAATTAGTTATTCCAGTTAGCGCTATA
CAATCATCTCCACATACAAACTTGCATTCTGAAATAGATACATCCCTGCATGA
GCATAGATgtACTCCATCACTATTTGGTATTCTAAGaTTAttgATAaTTtttATTCCC
CTTATTCTAATGCCTTCTGAGGAGTTTATAGAAATaGTCCAACAAGGAGAAtcT
TTAAAGGTtACATTCTCAATaGTTAAGTTCTTGCAATTATAGAAAAATATAGGT
TGATTAGGACGTTCCAATGgTTTtGCcTCaaTCTCCTCTATCTGTTCTGGagttaG

>gene_48127 >isotig08242
ACTTTAGACCGCGCTTTCTTTGAGATAGGCTCGGCTGAGGGAAGTGACTGGTT
CTGGTGGTTCGGTTCGGATCAGAACTCAGGGGATGATCGTTGGACCGATATT
ATGTTTAGAAGAACCCTTAAGAACGTTTACGAAATATTGGGACTTACTCCTCC
CTCTTTCTTGGACAAATCTATTTTGGAAGAGGCcaGTAAAatG

>gene_48326 >isotig08386
GATTTTATTCCAGGAGTTAGGgCtCAAaCAAcTCcTGAAaGTGCATGGGATGTAG
CTATaTTAGTAGAAGGATGGCAaaCaGaGCTTAAAaGTGCTGTAAAATCCGCAG
CTCCTGAAATGTATAAGGTATGTATTTtCCcAACAAAAGGaGTATCAGTAGAA
GGAAATACCATAGTAATTCCTGTTCCcAAAAAGaGTTTAGGAGGAGATTTtGA
GAAGGGATGGGGcTtCCAAGTATTCATCTTAGGACAAGAAGGcTtCCcAACTCA
AGATCCcgTCTCCTGCAGAATTAGAGAAGTAGTATCTACAtCCcAGCAATGGAG
ATTTGGTGGTGGAGATGATtACTAtGGAGATCCCAAtATCATAGATCTATTAGaT
TATGAaGGAATAAaTCAATTTGAGGTTCTAAGCAAGTATAAGAGT

>gene_48477 >isotig08502
AATCCATATGTAAGAgAACAGCTCGCCTTATCAGTTGGAGAGCTTGTATATGg
ATTAGGAGAAAGATTTGGTGCTTTTATTAAAAACGGACAAAGTATTGATATG
TGGAATGCAGATGCAGGAACAACATCTGATCAAACTTATAAAAaTATTCCCTT
TTATGTTACCAATAAAGGATATGGGGTATTTGTTAATCATCCAGAAAAAGTCT
CTTTTGAGATAGCCACTGAACATGTAGAAaGAGTACAGTTTAGTGTGCCAGGA
GAATATCTAAATTATTTTGTAATAGGaGGAAAAAACTTAAaGGAGGTACTGGA

TAATTATACCAATCTTAcTGGAAAACCTGCACTTCCCCCtGCATGGTCCTTtGG
ACTATGGCTCACTACCTCCTTTATAACAAaCTaCGAT

>gene_48822 >isotig08772
CACGGTTTTGCCTTTTGGACTTTGCCATTGGACCAAGCGTTTGAGAATGCCAT
CCCTGAGGTCAAGGgTGCGGCGATGTAACAAGATCTTTCCGTGGCATAAGGA
AAAAATTTCCTTATCAACATAAAGGCGGATAGCCGTCCAGTCaGgAGCGTTGA
cAATCTCAGAGTAATATAGTGGGACCTTGTCAtaaaGTCCGTTGATTAATGTTGC
TCTCTGCTCTCCTGGAAAaCCCTCTTCAAGCACCCCACGGGTACCAAGGTAGC
CGTTACCAACCGCAAAAATCGTTtCTGCTCGGCGTTGAGCTTtGGGGACAAATT
CaTCCTCTTTtATAATCcA

Appendix C: Nucleic Acid Sequences of the Unbleached Kraft Sample GHs

>gene_13844 >contig03967
GAGCCCTTTGGGAGGTTCTCGCCCCAAAGTGCGCCAGGCCTGGGCgAGATGG
GCCCGGAaGATCCGGGCAAAATGcGGGgCAAGCGGGgTGTGGTTGCGCGAGCT
TCGCCACCaGAACCAaTCGCTTCCcTCCGCCACATAAATATGCTCAAgcgCTtTC
ATGCGCAGATCTTtgTcGgcaCAGcTTTCcgCcGCgTCCCGGGCTTCGGcGAGGGCT
TGCCACGCCAaATTTtGtcGGGGCtCGCCGATCCAGGTgCGGAAATcaGCATCGAT
CCAAGAGCCTGACCAAATGGTGGGgAGTTCCCCGGTCACTCCGTAACGCTCG
ATGTACTCGCTAATTTTTATGGGGACGAAGTTTTCggCGATGGCGCGGTAAAG
AGCGAGGAGGAAAGGCCGgCCATTGCGCTCGTAATAATCCCAGgCATTTTCCC
CGTCCATGGCGATGAGGATGAGGGGTGGGGCGCTTTtGTGCCAAGGAGAGGC
aGAAGCGGCGATTTTGGCGACTAAGTCCGCCACCGCCTTCTCCCATGGCCATT
TTCCGTACTCGAAGCTAATGAGGTTGGA

>gene_14665 >isotig01314
CGGGAAgAtcACGGCGTAGCTCTTCCCGGACACATAGCTTCCGAGCATTTTGAC
CTGGTCATCCACGGTGGGcGCGAGGACATCGTAGAtCTGcGGCGCCCACATGG
GATCGGGgCAACCGCCGCCTGTCCACTCCCCGGCCgTCTTCCCAATGGGGCGG
ATATGGTTCGGACCaTAACCgTCCTGGGAACCCACGAGGACgTAGTGCCAGCC
aCGGATCTCTGGGACGAGGTCTTTGGGGATGGTGACGATGATCCGGCCGCGCT
TGGGGTCAGAGGCcACTTCCACAaGGAACGGGCCCTCGCCCGCGGCGGTCCAg
AGGTGCCGCCcATAACcGGGCCAACCCGCAaTCTTCAGcACGTAAtCCcACGGA
TGCtCTGGGTCGAACGCGACCTGGGCGGCCTTGCCCTCCTCaTGCACGTCGgTg
CGGCCACCGGGAGCGACGTCCAGGTAcACGAaGATGATGGGATGCGAAAAGC
CATGCGGgCCGTTCCAcGGGTTGGGAAGGGCCGGGAAGTCGAACGCCAGTTG
CCAGGCGTTGCCAGCATCATAGACCGCGTAGCGAACGAGATCGAAAAGCCC
GgCGCCTCCGTAAACACCTTgTTCAAGGGaTACACGTAAGTTCCGGgCCCAAA
GTCGTCGCCCTTGGGATCCTCCATGGCCCAgATCTCCGTCCCGCGGATGAGGG
TGGGGATGCGGGCAAAGAGGGGCCGCTCCGGAGCTTTCCCGAGAACCTCAgC
CcGGCCGCTCCAAAACCACGGCCACGGTGAGGCTCTGCCCGGgTTCCACCCCG
AGCTCCTCAAAGGgCACCATGAATTCCACCACTtCATCCACGGCGGCTTTGCG
AgAAAGGAGAGTTCGAACCGGCGAGGTTAAGCGCCATCCGCCCTTCCCATCC
GCGCCGTAGCGGAACACATACCCCGTTCCATCGGCcCGGATCTTCGtGAACTC
CAGCTCCACCGCGGAGGcAAACGCGAACCCAAGCTGCACCCCAGCGTGCCGG
GTGAGGATGTtcACcTTGTCGCTAGGCTTTCCGCTGGCATAGAGGACGAGGCG
GAGATcaGAGCCGACCAGCTCCCGCGCTGGCCTCGtAAGATCCgCCCGCACAT
AGAGGTAGGTtTCACCGTAACCCACGGCAAAGCTTTtGAT

>gene_15250 >isotig01378
GgATTTCATATGTTGCATTGGCGATTGCTATaTCACACTGtGCCCAGAAccTGTG
ATACCTGAaCTCCGGTGCCTGtCctgCATTGTATGCTGAAACCAGCTTCTGCCaG
TCAGGATCCTGTTTGTACTTGCTCCTTATGTCTATAAACTTcaCTCcaCTCTTTAT
TACATCCCCGTTCGGCATCTTCCCTGTCCAGCCCGCTGGAaTGTATACCTCTTG
CTCAAAGAaCCTCTTGTAATCTCCTTCTTCTCAGGTGCCGACAAgCCcTtGTCaT
CCCTGTATAAGTTCCACATCCTgTCCAGAAGCTCTTTCGCCAAATTCTTCGCTG
CCTCATCAAATACTCCATACTTCTtcGTCGCCGCACTGTAGTACAAAAGcGCAT

TCGCAAGTGATGCAGTTATACCcAAATCTGTTCCATAATCCACTATCTACgTGC
AGATTTGGATTGCCCGTATATGTCCCATTCCATGTATCCGGCTGGCCgCTCCA
GTCAAGtGTCGACGGTATCGCAAATGTACCATCACTGTTCAGCTTCACTACAC
TCTTTACCCAgcTTaCCCACTTCTCAAGCAATGCTCCTGCATCCCTGTCTCCTGT
TACATAGTAGTACTCcGCTAcTCTttGCATCGACCATGCCTGGAaTCCAAACCAT
GTGTTGCTGCCAGGATCATGGTATACCGGgTTCGGCTCATATGCCATCCcaTAA
AATGTCGCTGTCCCcGCCGGATACTTCTCATATCTgCCATTCCATGAATTGGTC
GCTCCGCCTGCTATCGCTCCCTCAGCTGATTGCAGCCATCTGTAAAATTCTAT
CTGCCTCTTCAAGCTCTTCGCCCAGTCATTtGCTCCGTTCGGTGACTtCGGCTTC
ATATCACTATCATTAGCCAACGCCCATGCTGCCATCGGATTTtGATAaCCAAAG
TGCGCATGGCTACAACCTATCTTCCAGACCATGtCCCGTCAAGAGCTCTCCCC
ATGCATAGTACCATGACAGCAGATAGTGCGCACTGCCATATCCTGTTGctCCA
GCTGCGTACTTGTCCTGGCATCCCAATGGCTTGAAATACTTGTCAAACATCGC
ATACCTCAAATAGTCgCCCATCTTCGCTGCCTTCGCTACATACTGCTTATCTCA
TTGAGCTTGCCCTGTTGCTTTGCCCATTCCTTCGCCCAGTATGTCGCCTGTATT
GCCCTCGCATCTGCATCCGGTGCATCAGTATATCTCCATTGCTTCGCATAATT
CTGGTCCTTGATAAACAGATCCAAAAATCCATTtGGTCCGCCCCACTTAAATG
TCTCCCAGCTCGGATGTGGtACTGTCTCCCATACAGACTCCTCCGGCCCTCTCT
GGAATGTGTTGATAAATGATGCCCTGCTTACTCCATCTCCTCTCCTGCCATAT
CCATACCAGTTGTCTACATCCATCAACCAGTGCATCCCGTACATCATCTGTGT
ACCATATGTGTTAACCAGCTCACTGTGCAAAGGATCCTGCCCAACCGGTACG
TCAAACTCAAGCGGTGATGGATACTTGTCCGGTGTCTCCCATTCTCCCGCATA
TGTCGCTGGCTTGCTCGGATTGTATCCTCCCATCGGCTGATCCTCTGCTGACG
GTATCATATACTTCTCAAGTGTATCCCATGCCGTCTTGAACTTGCTCCAGTCA
CCTGTCAACTTCCCATATACCGCCTCAAGCCATACATAGTACGAAAATGCCTC
ACTCGTTGTCAAATGTCCATAATCAGGTGCCTCACATATCAACGTCTCCACTG
AATGATACGGTATCCCGTCCTGGTTAAAATAGCCGTTTGCCGGATCATGTATC
TTGTTCCACAACCACATAAATCTCTGCCCGTATACACCAAGTCCACTGCTTGA
TGTCGGTGTCGGAGTAGGAGTTGGGGTAATTGTTGCTGTCGGTGTTGGGGTA
GGTGTGGTTGTCGGTGTCGGTGTCGGGGTTGCCGTCGGCGTCGGTGCCGGTGT
CGGTGTCACTCCTCCCGCCGGCTTTGTACCCCTTCCCGGCTCTCTTCCCCAcAC
TATCGCATCATCTATATACCCTGTCACCTTCTCGTTCTCTATATAGCTTGTTAC
ACTTCTTATCGAATAGTCATTCGACTGATTGTAGCTGCCACTGCCCTTCTGTAT
GCTTAGCCTTATCTCCTTCGTGCTCTGCCcTGCTGCCAAAACACCTGCTCCACT
CTTAAACCCTACCTCTACATAATAGTCCGCTCCGCCTGCATTcGCCCCAAGCT
TCACAAACTTCACATCTATATACGCAGGATTTATGCTGCTCGCTACACTTACA
CTCTGtGTcGCCTCACCATCTATCGTGTACCAGTATCTTACCTTCACCCTGCTA
AGATCTACCGCCgTcGTCCCAaCATTCTCTATCTTTATCTTcGGATTCAATACAT
TCGTCGgGCTGCTtAaATTCCcATTCGCATACCATACCTTCAGCGCACCGTAACT
CTCACTCGGcGTCCAGTAACTGCCACCTaCCGGTGATGCCGTCGGTGCCGGTG
TTGGCGTTGCACTCACTGTcGGTGTCGGaGTCGCTGTCGGTGTAGGTGTTGGTG
TTGCTGTCGGTCTCGGTGTAGACGTTGGCGTTGCTGTCGGCGTCGGTGTTACT
GTTGCTGTCGGAGTTGGCGTTGGTGTTGCTGTGCTTGTCGGTGCAGGTGTCGC
TCCTCCCGGCTCCTGCCCCCATACCAGAACACCATCTACATACAGCGTTACCT
TCGTATTCTCTCCATAATTCGTCATGCTCTGCAACCATGACCAGTCGTCTGCCT
GATTGTAATTGCTCCAGTCATTCTTGTTAAACCTTACCTGTATATCCCCcGTGT

CCTTACCAGGCTGCAACTGCCCAGCTCCACTGCTAAATCCTACCTCTAAGTAA
TAATCCGCTCCACTTACAgCACTGCTtAGCTTCACaaaaTTGAATGTCACATTGCT
TGCACcTATCTGTGCCCAGTCACATACCGCACTCTGTGGCTTATCCCCATCCA
CTGTGTACCAGTATCTTATCTTAACCCTGCTAAGATCAACACTGCTGCTGCCT
CCATTCACTATCTTAAACCACGGCCTTATTGACGCTGCACTCGCACTGGTCTC
ATTGTTCTTGTAtATTtACCTTCAAaCCACTTCCCGTACCAGGTGtCCCtGTtGGTGT
CGGcGTCGGTGTCGGtGTCAcAGTAACCGTCACTGTCGGCGTCGGCGTTGGTGT
TGGAGTTGCTGTCACTGTCGGGCTTGGTGTTGGTGTTGATGTTGGTGCAGGTG
TGGATGTAATAGTTGCTGTTGGAGTCGGTGTCGGGCTGGAACCCGGTGTTGAT
GTTGGAGTTGCAGATGGCAAAaTATTTGAATCAGCAATTTCACCCCATTTTAT
ACCCAGGCCGTTTGTTGCAACAAATACTCGACCGTACACGCGTGGATCTCCA
GTTATGTCTGCATTTGCACATCCAAACTGATGTTTATCATCATTTATTCTTATC
CACGTTTTACCACAGTCGTCAGATCTAAATATACCTCTTACCCCATTAATTTTt
GCATAGGTATAAATTGCAGGATATGTTTTCCCTTCAGCAGGTTTACCAAATCC
TATACTAGCGGCATCCTCAACACCTGAAATTTTTACAAATGAATATCCACCAT
CAGTTGAGTGCCACATCCCATTATTACCAACCAGCCAGATATCACCTTCTATA
CCTGGAACAGCTTtGAAATTtCCtGaTATTGGTAAACCAGCAGCAGGaCTTTCGa
TAAaTGTTTTTCCTTtAtCAGCACTTATAtAAAATTtgCcATTcTTAATGCAtAAAaC
TTATTTGGATtACTCtGTCAGAATATACAaTAGCTTCTGAaGGTACATTTGCGCA
TTCAACCCATTTGTTTCCGTTATCTGTAGAATAGCATACTTTCGCTCCTTTTGG
TGCCCAAACAACAGCGCTGCCATCTGCAGCAGCTGCAACTGTACCACCTTCG
CTTGTTCCTTGAGGTTCTGTATTTCCCTGAAACCATGATTTTCCACCATCATAT
GAGAAaCCAATTCTGTTAGTATTTGGATTCCACTGTTTATCAaCATTCCCTACA
CGAACCATAAAATTTGGATTtAACTCAGCAAAATCAATATCAGTAGTAGTCCC
CATATTAGGCTGAACATATGTCCAATTAGGTGCTTTCTCCAAGTCTTCATGTC
TGAATCCCGCTATATCTCCCAGTGCACTGAaTAAATGTGGTCCAACTGGTGGg
CTtATCAGTGCCTGGACTGAAGTTTCTtCAaTTCCTATAGCCTTGACTTTAATGT
TGAtATTTTGCCCTTTGTCCCAATTtGTCAAATCATCGCATCCATACAaTGTAGC
TCCTGTTCCATAAAGCATTCTATCTGAATTAAATGGGTCTATTTCAAGTGTTC
CAACCATCCAGCCAAGTTTTGGGCTTACTTCAGGTGGAGTGGGATTAGTATTT
CCAAAGTTTAGCCAAGGAGCTGCAGAAaTATCCATGTTATAATGCAATGTTCT
ATTAGGATAGCCATTCCATTCCCAAATACATTTCCACGTTTCGCcACCATCGA
TACTTCTCCAAATaTAGGTATCTGGCCACCATGAGCTAAGAGCTGCTACCATC
ACTACTTTCGgaTtCTGGGcGTCTACAGCCAATCCACcATATCCAAAaTAaGTAT
CTTGAGCAGCCATTGGACTTATATTTTtCCATTCTCCGGTCTTAGTATTGTATC
TCCATACTTCACCATAATCACCaTTATATGGCCCCTGAGTGTTACTATAAGTTA
TATATAACATCCCATCAGAACTTAATTTTGCACGCTGAGGAAGTAAGCCAGTT
GGCTGCCCTGGTAATACTTG

>gene_15474 >isotig01433
CCTaTGCATCAGCTCCTGGAACGGCTTTAACTCGCCTGTCTCCGGgTCGTATGC
AACTATCTCTCCCTCCACTATTGCCTCATCAGCCTTCAAGTACTGTCTAGCCA
TCTCGACTACATCTGGGTATTGATGGGTTATATTCTCGAGTCTCCTTGAGAAG
ATCcATATCTTATCCCCCTTcTtATGTATCTGTGCTCTCTCACCGTCAAAGCATA
TTCGCACACCGCCACCCTGTGCCTCcGGGTACCAGTGTaTGTCAAGTACATCC
AATAACCTCTTCCCAAAaCTaTCCGATGCTTTCTTCATCTGCTCAAGGTACCAG

CTCAAAAaCCATCTGTGATTTCCTTTAACCTGATTCCAGTCAGGTGCATCCTGC
AATGTTAAATATCCCGCAAAACCATACGATGCAGGTCCAAAAACTTCTGCAT
CCGGATCAAGTGTCTTTATTACTTTCGCCAGCTCCACCGATTTATTTATCAATT
CACTGCAGGTTACCTTCTGTGGATGAATTCGCGGATGAGTAGTAaACCATAAG
TCCGGCTCGTTGTCAAGTATATATCCTTTAATTCCCGTTGCAGACGATGATCG
ACCATACTTATTAATCAGATAGTTAATAAACTcATCCATATATACATAGTTAT
CATTCACCTCAGGCTGCAATGACAGTGCACCATCTTTTTTAAACtTGACCTCAG
CCCATCTCGGCGACGGAGCTGTCTCTGACTCGCTCACTGTACCATTCCCATCC
TTTGCCACATAACCTGCCATCTGTAATGTGATGGCTGAATATGCATTTTGCTT
TATTGACTGCTCGTGAAATTTGCTTACAACAGCTGCTGGAACGTTCTTATCAT
TCCCTGTTATACCCATAATATAACACATATAATCATCGCTTGAATGATACCAG
TCACTCCCTGCATTGGACATATTGTTCTCCCAATTGTAACCCGTCAATCTGTTC
CCACCAAGTCGTCTTGCAGGGTGAACAACACCCTGGATATCCTGATTTGCTCC
ATAAaTaTACGGGCTTATCTTTGTtCTTCCAATGGACGTATCGATCGATATTTTA
AcATCgGGGGTTACCgCAGGTGTAGGAgTTGGGGTCACCgTCACAGTCGGGGTC
GGTGCCGGTGTCGGTGTCACTCCTCCCGCCGGCTTTGTAcCCCTTCCCGGCTCT
CTTCCCCAtACTATCGCACCATCTATATACCCTGTCACCTTCTCGTTCTCTATA
TAGCCTGTCGCACTTCTTAtcGAATAGTCATTTGACTGATTGTAGCTGCCACTG
CCCTTCTGTATGCTAAGCCTTATCTCCTTCGTGCTCTGCCcTGCTGCCAAAACA
CCTGCTCCACTCTTAAACCCTACCTCTACATAATAGTCCGCTCCGCCTGCATTc
GCCCCAAGCTTCACAAACTTCACATCTATATACGCAGGATTTATGCTGCTCGC
TACACTTACACTCTGtGTcGCCTCACCATCTATCGTGTACCAGTATCTTACCTTC
ACCCTGCTAAGATCTACCGCCgTcGTCCCAaCATTCTCTATCTTTATCTTcGGAT
TCAATACATTCGTCGgGCTGCTtAaATTCCcATTCGCATACCATACCTTCAGCGC
ACCGTAACTCTCACTCGGcGTCCAGTAACTGCCACCTaCCGGTGATGCCGTCG
GTGCCGGTGTTGGCGTTGCACTCACTGTcGGTGTCGGaGTCGcagTTGGTGTGgat
GTTGGTGTAGGTAcTtGAATAGCTGGcGGcGCTGGCAATGTTGAGTTGtCCGGca
CcACAGTTGGCTCAATCAGGCTCCAAAAGGCATATTTTGATTTATAGTTTTCA
TCAAATAACAACGGGTAATCACTTTTTCCAAAGTTTTGACTCAGCCATGAGTA
ATCATCCTTtAGTCCCCAGAATGTTACATTAGTTATTACATTATTGTACTTTTtA
AatAAATCgAaTAACTCCTTATATCTCATTGCCTGTTTTATCAGGAGATCTCTTG
GTGGCGTTGAaTAACTGGTACTCGAaCCCCACTGATAAAAACTCATATCAAGC
TCCGTAATGTGTATCTCcAATCCAGgTATAGAGCTGAACAaTTTtATGGTGTTCT
CTATCTCGCTAATCGAGGgCCAATCAAGATTTATaTGACTCTGCAATCCTATtC
CAtg

>gene_15773 >isotig01485
GCCGCGCACATATTGCATGAGCCAAGTGAGGGCAGGGCGCTCTGTGCCGTCC
GCCCTTAAGAGATAAGCATTTTCAACCCACATCTGACCCTGCACGTACCCCcA
CAACGTAATCCCGGCTACACTCGGGTGCTCCCATAGGATTGGAAACTTTTCTT
GGTAACGACGGAGTTGTGTTTGATCATCGCCTGTAATATCAAGTTCTGAAACG
TAGATCGGAAGACCAGGGGACGCCAAAATTTCTAGTGACTTAGACATAGTTG
CCACGTCAACATAATCCATCTCGAATGCATGGCATTGAATCCCTATTCCGTCG
ATAAGGTTTCGCTTTCGAAGCAATTCCACAAGTTCCACCAGTCTTTGCGCCTC
TTGTGGACTCGCTATAACTCCATAATCGTTGATAAGTAATTTGCTGTTAGGGA
aGTATTtCCTtGCAAGCTCGAAGgCGGTAATTACCCAATCCCAACCCGTTGaCCc

TTCCCCACCTAGCGCCTCTTTATaGAAaGgCGGGGAATGCAAGGGCTCGTTAA
CGACATCGATAAGATCGATAGCAGGGTACCGGGCAGCAATCGCCGCGAACC
ACTTCTCTATTTCAACACGCTGGTCGTCAGGGGAAAGGTTTAAAAGCCAAAG
GGGTTGTTGCTGTCCCCATATAAGGGTGTGAAATTTAAAAaTAAAGCCATGCT
GTATCGCATACTCATAAATtGCGTCGAGCAGCCCCCAAAACCAAAGGTCTTGA
GATAGTGCGATGGATCCCCACTTCCCTGCATTTTCAGGTGTTACCTGGTTCCA
ATATTCCGCAAACTTAGGCGGTTCTGTGTATCCACGAAAGATATTGCCTAAAA
ATTTTTCGTTGCT

>gene_23927 >isotig02375
TTGGTTCATTTTTTCTATTAATTCTTTCATCATTTCTTCAGTAAGCATTCCTCCG
CTAAAGGATATTAAGCTTCTAAGAGGCATATATCTCATCATTTGAGTAAACAT
TTCGTATATATCTTTATCGGTAGATGCTAAATTGGGGTTATTCTTTATAAATCC
TTCAGCTATAGGTTTAAATATTTCTCTTGCTATTGGATCACTCATAACATCCTC
TACAGTAGAATTGAGGTGGAACTTCTTTcTTTTtcTTATAGTTGGCTTTACGTAT
ACTACTTCTTTTAGTAGGaTGTCTTTTGAAGATTTACCTATTAATATTTCAAAT
TCACCTTCCTCTACGTACCAATCTCTCAATTCTACATCATAGTATGCAAAAGC
TCTCTTATCTAAAACAAAAGTTACTTCTTTTtCTTCTCCAGGATTAAGCTCTAT
CTTCTCAAAACCTTTAAGTTCCTTTTCGGGTCTgatcACACTgCTTTTTATATCTC
TGACGTATAATTGTACTATCTCTTTACCTTTCATATTCCCAACATTTTtAACCTT
TACTCTGACTGTTAGTGTATCTTTATCTGTTAATTCTTTCTTATctACTTGcAAAT
CTCTATACTCAAAATCCGTATAGCTTAAGCCAAAGCCAAAGGGGAAAAGgAC
ATCCATTTCTTTCTTATCATAATACCTATATCCTACAAATATACCTTCTCTATA
CTCcGCAGTATCGCCTTCTCCAGGATAGAATAAATATGATGGATTATCtCTTAG
CTTTTtGGGGAAAGTCTCAGGAAGTTTCCCTGAAGGGCTTACAACTCCAAATA
AAACGTCTGCAACTGCACCACCCCAAGCTTGTCCACCTCTATATGTCTCTAAC
AAACCTTTAACTTTATCTATCCATGGCATTTCTATAGGTGCCCCATTCGATAA
AACTACCACCAAATTACTATTTACCTTTGAAACTTCTTCAATCagTTTATTATG
GCTTTCTGGCATTCTCATATGAGctCTATCAAAGCCTTCTGATTCATAtCTTTCa
GgtAGTCCTGCAAACACTACTACTACaTTAGCCTTTTTTGcAATTTCCTTTGCTT
CAGAAATTAGATTATCATCAACctCGTCAGTTTCTACTTTGTAGCCATCTGCAT
ATAGAATTTGTGCTTTTCCTtCCACAATCTTTACAATTTCATCATAGGCATTGT
CTAtcATTGTGGGATTTACGTGGGCGCTTCCTCCACCTTGAATTTGGGGATTTT
TAGCAAATGCTCCTATAATGGCTATTGTTCCTtCTTTTTTAAGAGGAAGAATCT
TATCTTCATTCTTTAATAAGACAAAACATTCTCTTGCTAcTtCTCTTGCTATCTt
ATGATGTTCTTCTTTActATATGTGGCATTTTCCCTCTTaTTTtCTATTGCCTTAA
ATACAATTTTCAATATTCTTTCTACTGCCTTATCTAAAACTTCCTCAGATATTT
TTCCGCTCTTAACTGCGTCTATGATTTTtCTATCCCCAATTCCTCCATCATAGGg
CATTTGTAAATCTAAgCCTGCAAAAAGTCCCTTATCTCTCTCATTAaCaGCTCC
CCAGTCtGAAACTACAAAACCTTCAAATCCCCATTCTTCTTTtAATACATCATT
AAgAAGATACTTATTTTCtGAACagTATTCTCCATTtACTTTaTTgTAaGCACACAT
TACagTCCATGGcTTTCCTTCTTTAacaGctaTTTCgAAACTTGCAAGATAAaTTTCT
CTCAAaGTCcTTtCATCTATCACAGCATTAACtGTTAATCTTCTATGCTCTTGATT
GTTGGC

>gene_27008 >isotig02858
CAGATTTATATTACCGAAAaTGGTATGGCAGCcAGaGAtATtGAGACCaGAAAC
GGCgAGGtCCTTGATCTTGATAGAATTGAATTCATTAGACAGCATCTCGAAGA
AGTATGGAAGGCAATAATGGATGGAGTAAATGTAAGAGGATATTTCGTATGg
AGTTTACTTGATAACTTTGAGTGGAGTTATGGTTATGAGAagAGATTTGGAAT
TGTGAGGGTAAACTTTtCGAATCAGAGGCGAAGTATAAAACTTTCTGGACAGT
ATTATAGAGATGTAATA

>gene_27761 >isotig03038
GCTGGCATAGAGgACGAGgCGgAGATCCGAGCcGACCAGcTCCCGCGCTGGCC
TCGCAAGATCCGCCCGCACATAgAGgTAGGTCTCACcGTAACCCACGGCGAAG
CTTTTGATCTCCCCcGCGCCCGGgTAGACCGCGGCCTCGGCCcACTCCTCCGGG
gACGTGATTTTtCCATCCAGATtGGGTTTAAcTTCCCCCAAATTTGCACGGAAGG
GCACGATCAGgCGCAGGAAAGGgAAACGGGGATTtCTtCGTctGCATAGCCCGC
AGCGCGATAGGCaCCCACCAGATGGGCCTTGAACAGCCAGTCGAAGAGATCA
TCGgTCCCGGAGTCCTGGTCGTCGCCGTACCACCAGAACCAGTCCGAGCCCTC
cGCGGCGTAGAGGGCTTGAAGGGCCCGCGGATTGGGGTCGCGGGAGAACACG
GCCTCGCGGGCCTGGGCCAgAACCTTCCacGCCTCGTCCTCTTCGGgCTCGCCC
TGCCAAGTGGAGAGGTCaCCaGCCCACGAgCCGGTGGGAAGGCGGGAAAGGG
AGGCCTTGaCCCcATGCTGGGCCAGGAATTCCgcgggggTCACGGTCTTCACCCA
ATCCGCCTGAAGGAGGGCACGATAgAGAGcgcGCAGgAaCTCCCGGCCGTTATC
GGGGTACCCAGCCATGAACATCCAGTTCTCCCCGTCCAGGGCCACCACAAGga
CGTACTCTTCGGGATTAGGAAGGGCTTCCCAaTATCGGCGAAtCcGgTTCATGA
AATCCGCCACGGCTTCGGAAGTGGGCTTgtttCCGTAGGAGAAGgaGATctGGTC
GGAAATTGTGTGGTCGCGGAAGAAGATGGCGATGTTTCCCAGGCGCCACGGC
TGGGCCAGCTCCGCCGGGCTTGGCCCGCGGCCCAGGGCGGCGCTCAGGGTCC
ACTCATCGGCCACGGTCCAGAcgAAGCCGGCTTGCCCCAAAAGGGCCACGGC
TTTTTCGGACACGGCCTGCTCCGGCGGCCAAACaCCCACCGCCCTCTTCCCAA
AAAGCTTTTCGTGCTGCTCCTGCCCCAAGGAAAGCTGAGCCAGGATGTCCTCa
TCCCAtCCATATTCGGCGAGGATGGGAAGAATGGGGTGATAGAAGGGGCTGg
TGATGAGCTCCGCCCCGgCTTCCTGTGCGCGGCGGTAaTAgTCCACCACTTGGG
TTATCACCCGATGCTGGGCGCGGATCACCCGCACGATGTCCTCCGCtGTCCAG
CCCcGCTTTCCGCGCAGaTCcaGGAGgCCAAGCTCCCCGTGAAGCTCGGGagAG
ATTTGCCAAAGGAGGAAAAGCCCGGCAGCGTCGAGGAGCTCTTGGTCTGTTA
GGGCGCGCTGCCCTTTTAGGCTATTcAGCTCGGCGTAGCGgGgATCGTAATACT
TTCCCCCGGgCCGAAGCATGTACGGGTTGATCCAGAAAAACTGCTCCTGCATC
GCGGCGCGGATCTCGGGGGTGAGCTTTTCCGGCTCATGAATGAGGGTCCAAaT
CCAAAGAaGGTGGTTATCCACCGCGCCGATGAGCTCATAAAGGCCACCCCGC
GCCCGCTCCTCCTCCGTGATTTCCGCGTAGTCCAAAAGCTGCCAAAGGAGGC
TCGGCTGGAGGTTATACGTCACCTTGATCCCCGGGAACTCCAGCTGAATCcGC
GGGGAATCCACGTACTCCTGCACCGCGTGCACCCGCACCCACGGAAGCTCGT
ACTCGCCAGTGAGTCGGTTCCAGTAAAGGGGCTGATGCATATGCCAAACGAT
GGCTAGATTGATGGGGCCTTTTGC

>gene_28236 >isotig02995
TAAAaGATTCTCCTTGCTTTCTGctAaTATTGTTAGTAATTTTACAGCATTCATT
ACaTCATTTATATCCGCCATTTCATGGGGAGAATGTACGTATCTTGTAGGAAT
AGAAAGGGTTCCACTAACTATTCCTTCTCCCGTTGTCTGGATTACTGCTGCAT
TTGTTCCTCCCCTCAATAATATTTCTCTTTGATAGGGAATATTATTCTCCTCTG
CAATCctAATGAGTTTTTCAACAACCTTTCTATGGCTAATGGATGCACTGTCTTt
AATTTtGATGGCAGGTCCCTTTCCTAAAATAAGAGACATCCTTTTATTtCCcTTt
GGGGTATCTGAGtgggCAGTAaCATCTACAGCAATTGCCATGTCAGGTTTTAtAT
TGAAGGCTACGGTACTTGCCCCTaCAAGCCCCATTTCcTCTTGTACtGTAAAGG
TaGCGTAAaGATCATGATaGGgTTTTgCCTTcTtCAAAACCTCtATAATTACTGCaC
ACCCTAaTCTATCATCTAAGGaCTTTGATACAATCCTATCTCCTAAATCTATGA
AaGGAgCATAGTATCCTCCAaaGgTTCCTGGAGGAACCAACTTTTCTGCAGATT
CTTTAtCCTTTGCTCCAATATCCACGAAAaTTTtGTCAAaGGTTAAATCCTTCAT
GTTTTtGCTTAGCTCATCATAGGTTTCCCCcTCCACTCCGACAACTCCTAATAC
ATTGGGAAATTGAATGGTAGTACCTAATAAGTTGTAGGGAGAAACTCCTCCA
ATAGCTTCTATCCTTATGAATCCATTTTCATCTATGTAGGTGGCAATTACTCCA
ATCTCATCCATGTGGGCATCAAGAAGCAATTTCTCTCCACTTACTCCCTTTTTC
CacACTATTAAATTaCCTAATTTgTCTATCTCATAACCATCtATGAAGCCTTTTA
ACTCTTCCAATATTACTTCTCTTATTCTATCCTCTCTTCCACTAATAGAAGAGG
TTTGGGTAAGTTtCTCTATTAAATCTTTCAT

>gene_28379 >isotig03157
TACTTTAACTTTTAGTTGAATACTTTCATCTTCAGGATAAAACTCTAAAACAA
CTATAACATCTTCTGGGAACATCTGAACTACATGTTTTTtAATtACTATCATTG
AATCAGCCCTTACcTCAAAaTCATCCCAACcTCTTAAATATGGCGTCCATGAAT
CTCTAACAGGTCTAAAGTTGTTTGCAGTAACAACTTTTATGGCAGAAAGTTTT
GTACCATTAAATGCAATAGGGATTTTTAATTCAACTGGTAATCCTTTACCCAC
AAAAAGTCCTGTATCTAAAACTACAGGTTTTTtGTATTGAACTACCTCCAAAT
CGTAATCTGCACCTCTATCGAATTtAAaGGTTAAAACCGCAaTTGTCCCCCACT
CATCAGGTTTCAGAATAGATTTAAtATAGCTTGCCCTAATTGTTACTTTATTTC
TATCCACTAAATAaTCTTCTCCTCTTCTTAGCTCCTTACTTCCATTAAAGATTC
CCAAAAaGTTATTACCATTTAGGATTAACTCTACATTCTGGTCCTTTATCTGTT
CGTCACCTTTTATATATAATATCCCAGGATTAAGAAAaGAATTTGGTATCCCC
TTTGATGCATTCATTATTATTTTTATCTTCATATAATCCCGCCATTTTCTATTCC
ATCTATCATAGTTGTCATATCCATTATCCCAATAGAATGTTGCCATTTTGtAcTt
CTTAGCAGTCCTAATTAGATAATCAAAATaTAGCCATTCGTAAGGTCTATTTC
CATTAAATACGCCATACTCTCCAATTATAAATGCATAATTCTCAAAACCCTcA
ATAGCAGTTCTTATGTCTTTATCCATCTGCTCTATATCTCTCTGACTTCCCCAA
GTTTTtCTACCCCACCAATTAGCTACAAAATCCCAGGGAGAATAATAATGtACT
CCAATGATAATATAAGGATCCTTTGGTGCTCTAAAGTATTTCATTTTATAAGT
ATCAGTAGAAACAGGGGGAAGAACAACTAACCTTTTATCATTGAATCCACCT
GAATTTCTTATCACCTTAAgAATTCTTTCATTTACCTCATTTTGAATCTCTCCCT
TTTCCTGgTCACTAAAGCCATCATAGTCAGGctCGTTTATAGTTTCAAAAATTA
GTTTTTCAGAATAATCTTTAAATCTATCTGCAATCTGTtCCCATAATTTTTCAA
GCTTTCTCATTGTTTcTtCTTTTCTCTGCTTCATCTCCGTTTTTAGCCATCGCCAA
GAATCATGATGCACATTAATAATTACCCAAAGATCCCTTTCTAAGGCGTATCC

TACAACCTCTTCAACCCTATCCATCCACTCCTTATCAACCTTATAATTTGGAG
CAGGTCCCATATGATCAACCCATGTTACaGgAaTTCTTACACttctaaatcctgttctcttaat
atc

>gene_28603 >isotig03213
TAaTActAAATGGTAAACAGTAAgAGGTGGAATATTTATTTGGAAACTATTGTT
TTGAATATTATTTACCTCTCCCATTTTTCTTATTTCAGGAGATGAGTAGTCAAA
ACCATATATAACTGAAGTAGAATAGTTTTTGCTTGCATTATTAATAACTATAC
TTACTTTTCCTAATTTATCTAAATTTCTATTTATTATGATAAcaTGTAATTTATT
ATCAGATTCTCCGTGAATAGCTGAGTATACAGGTAAATTTTCAACATCtGTTG
TTTCtGCCTTTACACATGTATCTCCATaTTTTGAGCCCTTTCCATCATAGTTCAG
ATATATGTTATAGGCGCTTTTTGCGTAGGTTCCACTATCTCCCCATCTTGTAGC
AAGGTATACTCCATACTTtCcAAAAaTACcAAGCACATCTACAAGGGCAATTCC
ACCTGATATATCATTtCTTCcACCaTAATCAAATTCTGTAATGGCAAGTTTtGTT
CCTGGATAGTATGTGTTTATATCTTTtATTATATTtgTtAGTATAGGTAGATATTC
AGGAAACCATTGATTtATCCAACtGTTTTCTCCTGCTGTAaTTTGTCCCTTTTGtG
TAGTTTTATAAGTTGGATCCCATAAGGTTCTTGGGgCTTGCATTCTTGCATATC
TCATCTCTGGGCTTGtATtGTTTtCTCCATCAAAACATATTCTTACTCCATTTGG
GTCTCTTGCTTCTGGGTACCAATGGAGAGAAAGcACATCTAAAAGCCTCTTTC
CAGCTTTTTGAGACTCTTCTCTCATTTTTtCtAAGTAATAGCTTATAAACCATCT
ATGATTTccTtttAcacTggACCAATCAGgagCATTCTGCAGGCTAAGaaaTcccATGAA
TCCATAGGATTCGTATCCaaaTACTTCTGCATACGGATCAAT

>gene_28797 >isotig03261
ATATCTTAACCCAAAACCAAAAGGAAAAaTAGCCTTATCATccaTTTTTAGCTtC
CaaGAAAAGGaGAGTTTACCCTCAAAATtATACTCTCCCAATAAAACATCAGTT
ACACCTTCTCCCTCTGTACCTGGAAGCCATGCaGAAACCAAGGCAGAGGAAT
TTTcTAATATATCTTTTaAaTTTaCAGGTCTtCCTATGATaAGTATAGTTACAATC
TTtGGTTTTCTCTtAATTAaCTGGTTAAaTAACAATAAATCTtCTTCATCtATTTGA
GGAGTAACTCTATCCCCATAAAACTCTGCATATGGACTTTCtCCAATAACAAC
AATACATAAAtCATCTTTCtTtATTTTTTtAAGATCTTCTAtATCCTTtaTAAAAATt
ATTTCTGAAGttttAGACAATTTCCtCTTAAAAGCCTGTaAAATAGTTGTTCcTGGC
ATAATATTCCCCTTAATTCCCTGCCATGATAAAGTCCAACCTCCACATTGGGC
TCCTATGTCATCCGCTTTTTCACCTACAATAAATATTTTCTTTATATTCTTAGA
AAGAGGTAAAACCCCATCGTTCTTAAGtAAAACTAAACTTTTTCTAACAGCTT
CtTTAGCAACTTCTCTgTGTTCACTACTTCCAATTTCTTTGATAAAcTTTTTATTA
GCTATAGGCTTGTTGAAGAGATTTAATTCAAATTTtACTCTTAAAaTTCTCcTAA
CTGCATCATCTAtTCTTTcCATTTTTATTcTCTTTTTTAAAATTGCACTTCTCAGA
ACTGAAATAAACAAcTTAtAATTATCAGGGACCATAACCATATCGATACCCGA
ATTAATTGCAGTAACTACTTTATCCTCATAGCTTCCgTATAATTGATTTATTGc
TCCCCAGTCAGAAACTAAAAATCCCTTAAAATTTAATTCTTTTTTtAACACATC
GGTAaGTAAATATTTATTTGCATGCATTTTAATTCCATTCCAGCTGGAGAAAG
ATACCATAATAGATCTTGCTCCTACATTAATAACACCTATATATGGTTTTAAA
TGAATTTCTCTTAATTCCTTTTCAGAAATTCTTGTATCCCCTTGATCCAATAAT
CCATTTATTCCAGTACCAAACTTGGTCCCGCCGTCTCCTAAAAAATGCTTGGG
ACAAGAAAGAACACTATCCTCATCAGATAGATTATTACCCTGAAaCCCTTTCA

CCATTGCCATAGCTAATCTTGATACCAAATTTGGGtCTTCTGAAAAaCTTTCAT
AAGTTCTTCCCCATCTTACATCCTGCGCCACTGCAACACAGGGAGCAAAAGT
CCAATTTATTCCAACAGCAACACATtCTTTTGCTGTTATTCTACCTATTTTTtctA
CTAATTTTTCATCAAAAGtACAACCCAAACCAATATTATGAGGAAAAACAACa
GCAGAGTATACTTTTCCtACTCCATGTACAGCATCCACTCCaTATAGTAATGGA
ATTTTTAatcGAGtTTGCAAGGCAAATCTTTgAAAGGTTtCAaTAtATTTCAGCCA
ACTCTCTGGAGCCAAaTTAGCTGGACCTGAATTTCCCCcGCTtAATATTGATCC
TATAAAaTAATTCTtCACATCCTCTGGAGAACCCTCGATATAGGAACTGTCAAC
cTGAGTCATTTGTCCTATTTtCTCATCTAAACTCATAGCATTTAAGAGTTTCTCA
ACCTTTATATCTAAGCTATCCTG

>gene_28824 >isotig03266
AATTTtCTTGgAATTATACCATCTTATCTTTAAATTATCCTTTACATATTTTAAA
TCCTTTGAAGTTTTGAATCTAACCTCATAATCCTCTTTCGGATAAAGATGGAA
GTAATTATCTTCCCATTTTGCTCCATCAAGCTCTAAGGCTACAAATAATGCAG
TTACTTTTGAATTAAGCCTTATAACATACTCATcctttttGTTTCCTTATCTCATACT
CAACTTCTGCCTTAGGAAGCTTCAAATGTTTAAATTCCGCAAAGAAATACTCA
TTTCTATCTACAATCTCTCCATCCTTAATAAGCCTTACAGCAATAAACTCATT
ATAAACTTGGGGTTTATACCTAAAGGAGTAAACTTTTgTgCTACTATTTGCAG
GTATAGTACAATTAACGGTATAGCTTCTTTtCCTTCCGCCTCTAAAGCTTAAAA
CATCtATTTTAAGCTTTCCTTCAAAATCCTTTAAAGTATCATTTATtCCATAGAT
GATGATCTCTCCCTTATTTTCTTTAATTGTGACCTTTATATCTTTTAATGCCCTC
TTTAAAAAGTAaTAAGCAGGCTTCTCCCTCTTATAGTAGTCTATAATGCTCCAT
GAGACTACAGGCCAACAATCATTCCATTGCCACACTAAAGCTCCTGATGTATT
AAACTTATTACTTCTCCAATGTTCCATtCCCGTTTTcACTGCCAAaCCTTGTACT
ATCTGAGTAAGATATATATAATCTTCCATATCTGATGTAaTAGGGAAATGCTC
TGCCAAAAATCTTATGATCCTTTCTGTGCCATTTATCTGCTTATTATGAAATTC
CATCTCCATACTTTGAGGATgaTATTCATCAAGAGATGAGAGAAAATCAATAA
TAGTTTCcTtATGAGGAGGCGCCTGAAAGCCAAACTCACTTATAAACTTTCCAT
TATCTTTtAAGTAACCTTTATAATCTTGCCATCCTGCCCAAACTATCCAATTAT
GCCTgTCTCCTTCCTTTTCACTATTTGGATCTTtCCcTCCATAGGGACTACTTGG
CCAATAaGGTCTtGTTAAATCAAGCCTGGCACAAACATCAGGAAGAACCTTAT
GATAAATGGTCTCTCCCCAGAACTTTTCTCTTTCTCCCCACCAATGGGCATAA
AAGCCCcAATCATTCTCATTATTTCCGCACCAAATCACAATTGAAGGGTGATT
TCTCAACTTTTTtATGATGATCTCTGcTtCTTCTTCTACCTTCTTcaGAAATTCCTC
ATCCTCcGGATATTCCGCACAGGCAAACATAAAGTCCTGCCATACCATTATCC
CTAACTTATCACATAAATCATAGAAGATcTCATTCTCATAGATTCCTCCgCCCC
AAACCCTAAGCATATTAACTCCAATTTCCTTtGCCTTTAGAAGTAGCTCCTCAT
AATCTTTcTCTTTTAAGCGGGGgATAAaGGAGTCTGAAGGAATCCAATTGGCTC
CTTTTGCAAATATGGGGATATCATTTACATAAAAaGTAAAaCAACTTTCATCAT
TTTGATCTTGAGTTACTAAATCTACCTTCCTTATAGCAAACCTCTGCTCTTtCCT
ATCCAAGATCTCTCCCTTCTCATCTACTAAAaCCAACTGAAGAGTATAAATGG
GTTGATCTCCATATCCATTGGGATACCAAAGTTTTgGATTTTCTATAActAAtGG
TATTTTtACAATGATCCTTCcTTCTGGAAGAGTAAAtCTCAACTTTTTATCTAAA
ATAGGCTtATCTAAAGAAATTCTCAGTGGCACaTCTATAGTAGCATTCTCCTGG
AGCTCCATCTCAAGTTCcAAATtAATTTGGGCTTTTCCCTCCATGATCCCTACA

GGAATCCATATATCTCTAATCCTTGCCTTCCTCCAAGCCCTTATATAAACAGG
TCTCCATATGCCAGAGGTGGCAAGTCTTGGTCCCCAATCCCAACCAAAaGAaT
ACTGGGCTTTTCTtCCAAATACACGAGGTGCAAATTCTCCCCcTcTTAACtCCAC
TGGATAATTCTTAGCCTTTTCTTCCAATACCCTTTTAGGTgAGTGGAAAAGCA
CTTCTAATGTATTTTCTCCTTCTTTCACATAGGGCTTAATGTTGAAACTCCATG
GAATAAACATATTGTCAGTTTCTCCAACCTTTTGTCCATTTAAAAAGATaTCA
GAAAAAGTATCAAGCCCTTCAAAATAAAGCTCTAAATTGGGGAATTCAAGAA
ATTTTTTATCAAGATtAAAAGTTTTTCTATAGATCCAATCTTCctCTTCTACCCA
TCTAACCTCTAATTCATTTAGCCCAATAAAaGGATCAGGGATTAAATCATTCT
CTAATAGATCAAGATGCACACATCCTGGCACTTTTGCAGAATACCATTTAGA
ATCATCTTTCCTCTTAAAAAGCCATCCTCCATTAAGATCAAGCCACTCCAT

>gene_28956 >isotig03308
ATGGTAAAcATGTATTTTAATTTAGTTCTTCACTCCCATCTGCCCTATGTTAAG
AAAGCTGGAAGATGGCCCTTTGGAGAAGAATGGTTTTTTGAAGCAATGTTAG
AAACTTATATTCCTCTCTCAATGGTGTTTTATAACCTTAAAAaCAAAGGAGTTg
ATTTTCATATTACATTGGGGGTTACGCCTGTACTTGCAGAGCAATTATATGAT
CCTTACATGATCTATGAAACAGAAAAGTATATTGAGGAGAGGATATTATTAG
CAGAAGAAGATTATGAAAATTTTAAGAAAGAGGAAAAAGAAGCGGATTTAT
CAAAATTCTACATAGAATTTTATAAAAACATATTAAAAAACTTTAGGGAAAA
ATTCTCCAGGAATTTATTACAATTTTGGGGAGACTTACAAAGAGAGGGCAAC
TTAGAGATAATTACAAGTTCTGCTACCCATGCTTATCTACCATTATTAAAAAG
ACCTTCATCTATTTATGCACAGATTTTTACTGGAAAGGAAGCTTACAAAAGAC
ATTTTAaGACAGACCCTAAAGGTATTTGGTTACCAGAGTGCGCATAcAAACCT
GGATTAGAGAAGTTCCTTGAGGAATTAGGCTTGAAATATTTCTtcGTAGATAC
ACACTCTATTTTAGGGGGAGAAGCACTGAGCTATCCTGACTTTAAGAGGgTTG
ATATTGGAAGATCCATACTaaaaCCATATTATGTAgCGGATAGTAATGTGGTAG
TGTTTGGTAGGCATGAAAGAACAAGTATGCAAGTTTGGTCTGCAGAATGGGG
CTATCCTGGGGATGGGGTCTATAGAGAGTTCCATAAGAAGGCAGAAAAGTCA
GGCATACAaTATtGGagAATTGTATCAAAGAatATAGATTTAGGTAAAAAGGAG
CTTTACGATCCTGAGGTGGCAAGAAaTAGAGcTTtAGAACATGCAAAACATTTt
GTTGGTTtATTGGAGGAAGAAGGAAAGGAAAATGAAGGTATAaTAACTGCAA
TGTATGATACGGAACTTTTtGGGCATTGGTGGTTTGAAGGGATCATTTTCTTGG
AGAAGGTATTTGAGCTTATAAACGAGAGCAAAATTGTAAAAaGTATTTCTCCA
AGtAAATATTTAGAATTATACTCTCCTAAGGAGAAGATAAATATACCAGAATC
CtCTTGGGGTAGAGGAGGAAAACACGAGGTTTGGTTAAATAAAGATACGGAA
GGATTGTGGGAAAAAaTATATGAGGTAGAGGAAAAAaTGGAAGAAATAGCCA
TGATAGATGCAAATTCTTTATGGGAGGAGAAAGTTTTAAAGCAGATGGCAAG
AGAGAAACTTCTCTTAGAAAGTAGTGACTGGCCTTTTCTAATAACTACTGGAC
AAGCAAGAGATTATGGATATAATAGATTCCATCAACATCTTGATAATTTTAAT
AGgCTcTTtAGGGTTTTAAAAATTAAAGAGCCTAAAGTCGAAGAATTTTCTATG
CTGAAAAGTTTAGAGGAAAGAGATTCTTTATTCCCTTATATAGATTACAGGAT
TTTT

>gene_28957 >isotig03308
ATGAGTAAAAAGGTTTATTTTGCTATTGGACTACATAATCATCAACCTGTAGG
AAATTTTGATTTTGTAATGGAGCGAGCTTATAACTTATGTTATAAACCATTAA
TTGACTTTTTTATTAACAATCCCGATTTTATTTTCAGTATGCATTTTTCAGGTtA
TTTGTTCCTTTGGCTTGAAAAGAACCATCCAGAATATATTTCCTCTTTAAAAA
CtCTTGCACATAGGGGACAGATCGAATTCTTATCTGGAGGTTTTTATGAGCCA
ATTTTGTCAATTATTCCCGATGAAGATAAGATTCTCCAAATAAATAAAATGAA
TAAGTATATTTGGAATAAGTTTGGTCAAAAACCGAGAGGTCTTTGGATGGCG
GAAAGGGTGTGGGAGCCTCATTTAGTTAAATATTTGTCTATGGCAGAAATAG
AATATATAGTATTAGATGATGCTCATTTTCTATCCACAGGTTTGTTCCCTGAG
AATTTGTTTGGTTATTATATAATGGAGGAGCAAGGATACACATTAAAAGTTTT
CCCCATTAATATGAAGCTCAGATATCTTATtCCCTTTCACCATCCCGATGAGAT
TATAAATTATTTAAAAAACGCTACcTCTACAAGTGAAGATAAATTGGCTATTT
ATTTTGATGATGGAGAGAAATTTGGGCTTTGGCCCGATACATATAAAACAGT
CTATGAAGATCGATGGCTGGATAATTTTtGGCAGGCTTTGAAGGAAAATTCAG
AAGAAATAGGTCTTGTTCATTATGGAGATTATATAAaTAATTTtCCTCCTCTTG
gAaGGATTTATCTTCCTACCGCTTCTTACAGAGAGATGATGGAATGGGTTTTAT
TCCCAGAAGCTCAAAAAGTATTGGAAGAaCTgaCgAaTGAAGTTAAAGAGCAA
GGCAAGTGGGAtAAGTATTCTCCCTTTATCAGGGGAGGATTTTtCAGGAACTTT
TtGTCAAAGTATGATGAATCAAATCACATGCATAAAAGGATGTTGTATGTAAG
AAATAAGATTAAAAaTATAGAAGATAATGAGAAAAGAGAATTGGCAACAGA
AGAAATCTTGATGGCTCAAGCTAATGATGCTTATtGGCATGGAAtCTTTGgTGG
GCTTTATcTTCCTCATCTAAGATCTGCCATTTAcTCCCATATTATTAAAGCAGA
AAGTTTTATAGATTCTGATGAGATTTCTATTAAAAaTATTGATTTTGATTGTGA
TGGAAGGGAAGAGATAATTTTAgAATCAAAGTATTTTAATTTGTATTTTGCTC
CAAATTTAGGGGgAGGGGTTTTAGAGTGGGATTACAAACCTGCCTCTTTTAAT
TTGACCGATACACTTACAAGAAGGAAAGAGACATATCATGCTAAGTTATTCC
ATATAAAGGATGAAGAGGGGAAAGGAAaGACGATTCATGAGAGATGGGCAG
TGAAaGAGGAAGGTTTAGAAAATCTACTCTTTTACGATAGTCATAGGAGAATT
TCCTTTAAAGAATACTTTTTTGACACTGTACCTTCTATTATTTCTTTTTGGCAT
GgtAAGgAGAAGCCTTTATTTACATCCTTGGATGTACCTTTtACATGGAATATTG
ATAAAAATAAAGATTATGCGGTTTTAAATTTTAGTGGAGAAAAGGATGGAAT
AAAAatAGTTAAGAGCTTTATAGTTGATAACAATAAAAAaGATTTTGAAGTATT
TTATTCTTTAGAAAATAATGCTTCTCAGAGTAGAGAATTTGTATTTGCATGGG
AGATAGTATTAAACTTTTTAGCAGGAGAACATGAGGACTATtATTtCCAAGTG
GATGGAGAAAAGTaTTTGTTGAGGACGGTAGGAGAAAAGGAAATAGATTTTT
GGCAAATAAAGACTCCCATATTAAATTTGGACTGTAGTTTAGATAGAAAGAC
TCTTTGGTATAGATATCCagTGgAGACGGTTTCTCTTTCAGAAGAAGGTTTTGA
AAGGGTATACCAAGGAAGCTCTCTTATTCATTTGTATAATGTTTATCTAAGCC
CTAATGAAAAGTTTGTTGTAAGAGTGAAATTTACCGTC

>gene_29370 >isotig02058
AGGAACTAAATCTCCAATCCTTCCTTCTGCAAGAGCCATTACCTCATCCGCTC
CACCATAaTAAACAAGGATCTCATCATTCTCATCCAATAccTCTGTAGAGTCTT
TATACTTTGGTACAACCCcACAGGTAAAGACTACTTCAGGCACCCATCCCTCT
ACCTCATAatACTCCTCAGGCTCtAAGAtGGgAtATGgAGAGCGATATATTACCTT

ATCTGGATTTTCcAAAgAGGTAATCAcTAtACCTAAGGCATATGCCTTTTTCCcT
TTCCATTTtccTACCCCATGATAGATATGTAGcCAcCAAacTTAGCTTTAgAGGGG
GTGgCTCCTGCCCCtATTTTCTtCGCTATCCCAGtAGTTTGGACGTGGCaTTAAAA
AGGTTcttCCTaCAGtGgctCCTTTATATGGGaTCTCgTGGGTAAaTAAGATCTGCAT
ATCAGGAGGAATCCTATGGATCAAAGCTACCTTCTGAtGgTAtAGTTCTGGgAA
aaGATTTATAAAGTTTCCTTTATACTCCTTCTCTCCCAATTTTATATCATCTTTT
CTCTCCTCAACaGGGAATAAtACTGCATCCTTATCATCAAGCCCcTTAAAGATA
GGACCATTTCTCTCCCAAACCCTTTTCCACTCCTCATAGGATTTGAATtCCTTtA
CaCCCCTTAAAAAGTCATCTATATGGATCTTTGCCagaGAAAGCTGAATTAAAT
CTCCATAGGATGTATAAGTCATATATAGATAGTTGTTTATTAgaTAGATTCTTG
GGTCCTCTGTTCCtccAAGCTCCCTTAACATGCCTAAGTGCTTCAACTGCCAAT
CCTTTCTTTTCTGCAATCTCTTAGGAGATtCATACTCTTCTTctGGTCCAAAAaTT
GGGAAaGGAaGTCTTCCATCCTCCTTGTATCCATCTCTACTCCACCAAAGCCCA
ATtcTTGAGAAACCgTCCTCTCCCAATGCCCTGTAAAAAATATAGGTAACTCCg
TTTAATCTAATGGCGCCTGCATTGTAAACTAATCTCTCCCAATGAATCTTTTCT
CCCTTTACATCAaTGGTCAAGTCgTATCTTGCCTTAAATAAGGGATTgCCAATA
AAGCGCTTTAATTTTCCTCCTTGAGGATAACTCTTTGGATTTAGAATCTCATC
AAAGATGTATAAAAACTGACTTCCTACCCTCTCTGGTGAATGTCCATAGGCA
ATACCTTGATAATACTTTAATATATCCTCTCTTTCTTCCCCATTTCTCAATagCT
CTTTGACCTTATCTACTATGTTATCCTCATTATATCTTAGACTTTCAAGATACT
CATATCTTAACCTGCCTACATCCCTTACAAGAACTACCGCTCCTCCTGTAATT
ACCTGATAAAAAaGTACAGGATATTCTTCCTCCTTCTGATACAAAaTTACCAGT
TTTGCAGATTGAGTATACTTAACAATATCCTTCCAAGAGTTAACAACCcTAAA
ATCTATTTCAATATCCTTAAATTCCTCTCTTAATTTATTCTCCTCCTCCAAAGA
AAAGCAAATCCCCAATAAATATATGGATGGATCCTCCTTCAAgatCTTTATTTtA
CCTTTTAACGGaGAATAATCACCTAAAAGAATAATAATATTTTTAtCCTGGGG
AAGATTAAgcTCCCTTACTATCTCCTCCTTAGAATActCTTCTACATtATAATAG
GGGATTTCTACTTTATAAGTtCTTTTATCCTCCGATGGAACAAAGGATATGATG
GCGgACCAAAtAGATGAATAAAGGGCAgTATACTCTTCTCCCCAAGGATTTAG
GAGAATCTTTTTTGAACTAAATTTTTCTAATTCTTTTAGAATATTATCCCAATC
TAAATCTTTCTTCCCCAAAaCTAAAAAAaTACTGTATCCTTCTTCCCCCTTATA
ATCGAAaGAGACTTTAACTTTATACCCcTCTTTTTCCCACCAATCTAATATAGG
TACCAGATAATCTTCCTTACCTAAGACTAAAATTCCGCTCCCCTT

>gene_29387 >isotig02058
GAAAGAaTTTACATTGTTCCTGAGCCAAAGAGGATGAGTTTTTCTGGAAGATG
GTTTCCcTTTGATGGCTTTGAAAATTTACCAGATTTTATCTCAAGGgAGTTTAA
TATTCCAAAAGGAAGTtGGAGAATTGAAAAAaTAAaTAGAGAAGGGATAGGA
GTAGAAATAAAGGAAAAAaTTGTGAATATATGGGgAAACGAATATATTtGCTA
TGCTACTCTCATTCAAATCCTAATgCAAAAAAaGAaTATAATGCCTGAGGTAG
AGATTGAGGAGGAATTTTCATTCTCCTTTAGAGGATACCACTTGGATATTGCA
AGAGGAGGAGTACCAAaGGTAGAAACCcTAAAGAACTtACTAAAaTGGCTCTT
TTtATTAAAATATAACTATTTtGCGATATACTTTGAAGATCTTTTCCCATGGAG
AAAATACCcACAAATTGGAAGATTAAGAGGAAGAAtgACGGAAGAAGAGCTA
AAAGAGGtcATAAATTTTGGAAATCTTGGGATAGAGGTTTTttCCATCCtGGAGC
TTTGTGGACATATGGAGCATATCTtATCTcTTCCCGACTTTAtGACCTTTAGtGA

GTGGCATAGACCTCAAgAGGGATGCCTAAACGTTTCCAATGAATctGCAAGAT
CCTTTGCCTATGACCTActGAaGgaGGTTTTAGATTTCTTCCcATCAAAGCACAT
ATTAATAGGGGgcgATGAGACCTGGGCTTTGGGAAGAGGAAGGAGCCTGGAT
AAAAAAGGTATTTTTGAGGGACCATCCCTTTATGAGATGCATCATAGAAACA
TGGTGAATATTGTGAAGGAGAAAAATAAGgAACCCATCCTATGGGGAGATAT
GCTTaCAGGTATGTATCTtACAGAaGAGGAAAGAGAAAGATGGAAGGTGGCTT
TAGAGAGCGAAATATGGgATGAGGTGATAATTGCCAATTGGGACTATAGTGC
TAACTCCATAGAgCATTTTAGAAATAAGATAAATCTATTTGGgAAAAGAAAGG
CAAAGGAGCTTGCATGTCCAGgaCTTGCCAACTGGAATAGGTATTATCCAAAC
TTTGAcATGGCACTAACAAACcTGAAAAACTTCTTGCTCCCTGCAAAGGAGGA
AAAACTTCCAGGTTTTCTGGTTACTGCCTGGGGGGATGATGGTgAGGAGTGTT
TATTCTCCTTcTTAGACCCTCTAATTCTTGCCAGTATGGAAATAGCGGAAGGG
GATGgAAACTGGGAGGATAAATGGGTTGCCCTTACAGGAgAAGAGAAAGAAG
TTTTAGATTTAAGAAAGGAtTTTGGAAATACTATAATCTCAGAAACTTTaAAGC
ATGTTCTCttCAAAgATTTTACCTATAGGTATTACACAGAGATAGTAGCAAGTA
AAGCATCTGAGAGGAAGAAAACAGGGGACTTTTGGATGGACTATTATATGAG
CTTACTTGATAACTTAGcAAATaAGGaAAAATTaaTaGAAATTTATGaAAAGGTG
TATGaAAGGCaAaCAAACTTGaaTCTTCCTGAGGATcTATCCCtAATAAGGGATC
TAATTGGGATAGGCATTAATAGATTAAAGAATAAAAATAATTCAGCGGAGTT
TATTGCAGTCTCCTATAagTATAAAAAGCTATGGCTTTCtGAAAGGAAaGAAGA
AGGCTTAGAAAATATTATTACAAGGTTTTGGGgCTCTGCAGGAAGGgCGgACT
TA

>gene_29847 >isotig03523
AAAaCTCTTGTGGGAAGATACAAAGGGAAAGTTTATGCATGGgATGTGGTGA
AtGAAGCCATAGATCCATCTCAaCCcGATGGGTATAGAAGAAGTAAATGGTAT
GAAATAATCGGACCTGAATACATAGAAAAAGCATTTATATGGGCTCATGAGG
CAGATCCAGATGcAAAATTATTCTATAACGATTACAACACTGAAGATACTAA
GAAGAGAGAATTAATATATAAATTAGTGAAAAaTCTTAAAGATAAAGGTATC
CCAATACATGGAATTGGAATTCAAGGACATATAAGAATTGATTGGCCTGATG
TCTCTGAAATGGAGAAAACCATTCAACTTTTCAGTACAATACCTGGCATTGAA
ATACATATTACTGAACTTGACATGAGTATATACACTCAAGCAGGTATTGAGT
ATCCAAAACCACCAAGGAATTTAATGAtCGCTCAAGCATATAGATATAGAGC
CaTATTTGATATGCTTAAAAaGTATAAAAATGTAGTAACTAATGTAACTTTTTG
GGGATTAAAAGATGACTACTCATGGCTAACTATAAATAGAGGacGACTTGATT
ATCCTCTCTTATTTGATAAAGActACCAAGCAAAACTGGCATACTGGGCAttAG
TAGAACCTAAAGTGTTACCTCCACTAactCAACAGGGAGCTAtcGTAAAAGGAa
CTGCCATTGTTGATGGgaAaGAGGAt

>gene_29850 >isotig03524
AACGAATATTTgAaGATAAACGTtATAAATGATAAGATTATAAGGTTTCATTTT
TCAAAAAACAGAGAATGgAAGCATAATTACTCcTTTGCTATAGAAAAAGATTt
AATgTTAATGGATTTTtCTTTAGAAGAAAaGGAAAAaGAGATTCTTATCTTTAC
CTCATTACTAAAAaTATACATAGATAAAAaTAATGGAAAGATAAAAaTATATG
ATAAAGacaaTAATCTTATTCtCTCAGATTATGAAGACTTAGGTTAtaaaaaaTAGA
TAATAAGATTTACTGTTATAAGGAAaTTAGAGATGAAATAGCtttttcAGGCTTTG

GGGAAAgAACAGGAAGCTTAAATAAGAAAGGAGAAAGATTAATAAATTGGA
ATACTGACGAACCTCATCACTCACCTAAaGCGGATCCTCTCTATCAATCTCAtC
CCTTTTTtATtGCCTATAGTCCTACGCGAAGCTATGGATTATTCTTTGACAATAC
AAGTTTAAGTTATTTTGATATGGGAAaCgAAAaTGagAAGTACTGGtATTTTGCA
AGCGAGGgaGGGGAaTTAGACTATTATTTtATTTATGGACCAAcTCCTAAAgAG
GTAATAGAAGGATATACAAAACTTACTagAaGATATTATATGCCACCCATATG
GGCTTtAGGTTTtCAACAaTCGAGATGGAGCTATGAAAGTGAGGAAAAaGTAaG
AAATATAGCGAAGACTTTTAGAAAGAAAAATATTCCATGTGATGTTATCTAT
CTTGATATAGATTATATGGATGGGTATAGAGTATTCACCGTAAATAAAAAGA
AATTTCCAAACTTTGAAAAGATGGTAAAGGATTTAAAaGaGgAAGGGTTTAAA
ATAGTTTtAATTGTTGATCCTGgAaTTAAAAAaGACGAAagcTATGAgaTTTATAA
AGAAGGAATaGAgaAaGGATACtTTTGTAAAAGAAATggaGATGtGTATTTTGGA
TATGTGTGGCCAGGAAAATGTGCATTTtccTGATTTATtAGAAAAGATGTTAGA
TCTTGGTGGGGTGAAAAACTGAAAAAaTTAaTTAATCTTGgAATTAGTGGTATT
TGGAATGATATGAATGAACCATCTTCTCTTTCTAAATTGAGTATTATCTAATG
AGaTTACTTTTTTATTTCTTAAAACTAAAAGAACCTCCCTCTTTACCTAAACCT
TCAGAaTTCAACGCtAAAAaTTAAAGAGATAAaGAGAAAAaCTCTACcaTCAGA
TGTtATACATGGAGAAAATCAAGAATTTACTCATGCTGATATTCAtAATGCTTA
TGGGCTTCTTATGACAAAGgCAAGTTTCGAAGGATGGAaaAAGTATAGTGAtA
AAAGaCCtCtTAtTgTTTCAAGGGCAGGATTcTCTGGAATTCAAAAATATTCTGC
GGTATGGACAGGAGATAATAAAAGCAGTTGGGAACATCTatATATGAGTATTC
CTATGCtaCAAAATTTATCTTTATCTGGagTTccctttGTAGGaGCAGATGTGGGAGg
cTTTtGGagAGATtGtACtCCAGAACTTTTTaTAAGATGGATACagCTTGGagtATtCT
AtCCcTTTTTTAGAGTACATTCTGCCATGAACACAAAAtcACAAGAGCCTTGGA
GTTTTGGGgAAAAtGTAgAAAAAatAGCcAAAAAaTAtATTATTTTGAGatATAAA
CTTAtACCATATATCTACTcTCTTTTTTAtGAaGCAAGAAAG

>gene_29901 >isotig03536
CCAATTTAAAGATAAGTCTTTTCCCATTTTAATATAAAGTTCTGAAATTCCAA
AAATGGGAGGAATTTCATATTCCAGCTCTTTCCAATTCTCTGTTTTGGGgaTAT
TTAAAGAAAATTCATTACTATTCCATACTAAAGTAATTTTATTCTCCTTAGGA
GAGAAGACTTCTAATGAGATTTTTCTTGAGGGTTTTGCAAATTCGaCATTTTTG
AaTAAAATCCAAGATTCATTTGAAAGGCTAAGTACATATGTTTCCTCAAAATC
TCTTTTGGTTTTtAACCTTATATCATTGTAGTCATCATAATTTATAGCCTTTGTT
TTtATGGTTAGGTTTCTATTAGGAATTTCTTCCCCTTCTATAAAGATTTTCTCTC
TTAGTCTAAGATCTTTAGAGGATGCACcAATTAAGATTTCATAATGTCCTTTTt
CTATGCAATATTTTtCTCTTGATACATCAAAAATGAAAAGTTCATtAACAGGAA
TAGTAAAaTTAACTCTTTTtCTTTCTCCtCTTTTtAAGAATATTCTCTTAAATCCT
TTtAATTGGAGCAGTGgTCTTTTtACTTTTGATTCTAAGCATCTTATATAAACTT
GGGGCACcTCaTCAGAATCcATATCTCCTTTATTTtCTAAATCAAACTCAATTTC
AATATTCTCTTCAGGGgTATATTTTTtCTTACTTAAGTGAAGATTAGAATATGA
AAATTCTGTATAAGATAAACCATGTCCAAAGGGGAAAAGTACTTCCCTATCA
AAGTACATGTATGTTCTTTTtCCTTTAATTATGTCGTAATCAGTAATGGGAGGA
AGATCGGAAGAGGATTTGTACCATGTCATATTTAGTCTTCCTGCCGGAGAATA
ATCTCCAAAGATAACATCCGCTATAGCATTTCCCATATCTTCTCCGCCATGAG
ATAAAAaTATTATTGAGGGAATATTATCCTTTGCCCAGCTTATTGCATAAGGA

TAACTACTTATTACCAATAATATTACTTTAGGATTTACCCTATATACTTCTTTT
ATAAGCTTCTCTTGGTGTTCTGGCAATACAATATCTGGTCTATCAACGTCTTCT
TTCCCATTAACAAAAGGATTATTTCCAACACATAGGATTACATAATCTACTTC
TTTAGCTAATTCACATGCTTTTTTAATTCCATCTTCTAACTTTTCTATAAAAAT
TTCTTCTAATGAATCACTAAAATCCTCTTTAACTTTTAATCTATTTTCTTCTGA
TACAAAGAGATACTTTCCATTCCAACTTTTTATACATAAAGAGTTATCTTTTA
AGGGTTCCATATTAAAAAGTTCTTTTACAAACCATCCTCTCACTTCTTCAGAG
CTTGCCCAAACATTTCCTGTATCATCTACAGTAACATATTTACCATTTGCTTCT
GATTGTAAACTTTTATTTCCCcAACCCcAaTCAATTagTaCAAAGGTCTCATTCTT
AGTTATTCtCTTTCCtCTTgCCCAGAGAGGACTTTCCAGAGCATCAATACTACTT
ATATAGGCGTtATTTTTTAAAGACTTGATAGCTATTCTATCAAAGGAGTCCCA
ATAtAATATTTCTACGTCTTTATTTATAAGCTTATTAATAATTCCCTGCAGAAC
AGTAACCTTATAAGGTAGAGTTCCACTGTAATGGTCTGTATATAGTTTATTTG
CcAAAGGACCTATAActGCTATCTTTTTAaCTTTATCCTTATTCAaGGgTAAAaTT
TTAGGATCATTTTTTAAAAGCACAATAGATTTTTGGGCAGAAATATAAGAAA
GCTTTAGATGATCTTTAtCACAAATCTTAGAAGgTGGCACATAAGAATAGGGA
TTTTTATCTTCTGTTTCAAACTCTCCTAACCTAAATCTTACTTTTAGGATATTT
CCAATGGCTCTATCAATATCAGCTTCTGTTATAAGATTTTTTtCTAAtgCTTCCT
TTGCAGACCCTATTACAAGATCTGGATCATCAGTGAAGCAGTCTATTCCTGCT
TTTAAGGCATAAGCCAGAGTTTCGTAATGATGCTCAAAGGTTTTATGCATAGT
TACTGTTTGACTAAAATCTCCAGCGTCTGTAACTACGAATCCAGAAAGACCCC
ATTTTTCCTtAACAATTTTCTTtACAATTGGGTtAATGATACAaGGAACTCCATTT
ACTTCATTGTACGCAGTCATAATACATGGTGCATTTCCCTTAGTAATTACTCTT
CTAAATACATCTAAATAGTATTCATACATGTTCCTGGGATCAATGGAGGCAG
AAAAGAAGGCTCTATCTTTTTCATTGTTATTGGCAAAAAAATGTTTtGGGgTTG
GGgCAGTCTTTAAATAAAAaGGATCATCTCCCATCATACCTCTGATATAAGCA
CTTGCCATTTCTCCTGCGagAAAaGGGTCCTCTCCATAACCTTCTTCTGTTCTTC
CCCATCGAGGATCTCTTACTAAATCTACAGTGGGAGCCCATAACATAAGTCCT
CCAATTTTTCCTCTCATGTAATAGTATGCTCTTGCCTCcTCTGATACTACTTTaC
CTATTTCTTCCATGAGATCTGTATCgAAGGTtGAACAAAGTCcAATtGGTTGAG
GAAAtACtGTAGCAACACCAAGCCAGGCTAcTCCATGAGCTGCTTCTCCTCCTA
CCCAAAATTCTCTAATTCCTAATCGGGGAATTGCTTTTTGTCGGGTAGG

>gene_30141 >isotig03897
AAGCTCCAAGCAACCGAaTTtGgTTGTGTCTCTCCAATTGTTGCCGACTTTATC
ATTCCAGGTCAATACCCCTACTCTTCTGCCTGTATGGTCAGCATCATTGACCT
GAATATCAAAACCTATAACCATTCCTTCGGAGAGTTTAGttgtCCTCATaTAAAC
CTGGGCTtCAACAACATACCCATTTGGAATTATTTTtGTTGCTGTgaTAAAGTTC
TTAGCATCACCATTTGTTCCAAATGTCTGAGTATTTTCATAGTTTACTCTGTAC
TGGACATCATCGTCTTCATAGTAAGgtGTCTTATgaTTGTTTTCATCAATAAATa
TTtCAATagagTCCTGctCCCAGGGATTTGTGTTTGCCTTGTTCAAAAGTGGaTCA
TAGACAACTGCATATACATAGATGCAATTTTCATCCCAcATCATCCTTGCTTTT
GCATATGCTGAATCATATACTGTGCCTGTTACTGTTACTaTtGTGTCAGTTACtat
ctcctgtgcatttttccatgcatcatccaattctgcgtcgatctttggcgttcccttctgtgctgaagcaaattttatactgtctgccattg
ttaaaattCCAtAActATCTGTTTCAAATaTTGTGaGTTtGTTCTATCATTCCAGCTGTA
CTGTTgCCcATTaTCAATtaCggCAATAgTCAAAaCcAatATTtGtGTTAactGttAAattG

TTGtctAAAAGCagCTTAActtCcATtGTATATCCATTTAATtGCTGtAACACTAcaTA
ATCTTTACaTAGCCAgTTTTTGGATTCATtCTTTgtGCCaTTtCTCgAAaTtGAAaCCc
agAAGTCATcAtCTTTTAACTCTGGCAACTTTCCATTGTCCTGATCAATGAAAAT
TAcAACCCTATCATTGTTAGCATCCACACTTGAaTCATTTACAGTGACATAAAG
ACAAaGCTCATTGCCACTCCACAGTGCCCTTACCTGAGCAATATCCTGtccTGC
ATCATTCAAAAtcGAAAGcGGAATGGTTCCTTTGTATTCTTtAtCCATAATCCCA
TCAATTCTtGGCTGTGCATtGTTtGCATATCCTTTATTtATaGCAACaGGcAACACC
GAAGGTTCTACAATAGCCCAGTAATTGTATTTtGCCTGGTaGTtACTAtCAAaTA
gCAGTGgCcAGTTATTTCTATCTTTTGACaGCCATGAATAATCAtCTTTTAGaCCC
CAAAGTGTAACATTtGTGATtCTGTcGCtGTGGCGCTTtAACAtttcaaaCAgctctttgaatt
ttaaaGCCTGtttaatCATAATATCCTGTGgTAaTGTgTCATAtcGCTGAcTGCTGCTTGT
ATATACGCTGATATCAAGCTCTGTTATTTGAATTtGAAGACCCGGTATTGTGCT
GAAcAAATTGATTGTATCTTCTATTtCTTTGACTGtAGgCGAATCCACATTAATA
TGTCCcTGCAGACCAACACCATCAATGAGACCTCTATCATGCAAAGCTTTGAC
CATGTtGTAAaTAAaCTGCcTCTtCTTTGGATTTTCAGTGCTGTAATCGTTGTaGA
AAAGTTTTGCGTTCGGGTCAGCTTCCCGTGCATACTGGAATGCAAGGATTATA
TACTCTGGTATTCcACCTGTTTCcGGAGTTGGTCCTAAAATTcTGTACCATtCAC
TTCTTCTATATCCATCCGGTtGattttCATCAAttGcTtCATTTACAACgTCCCATGCAt
aaaCTTTCCcctttATATCTTCCAaCAAGTgTCTGAATGTGTGTTTTTAACCTAtttCTC
ttAAAAGCTGCTTATCATCAGGGTTGTTtGGATCAAGTGgCGAACCGTCAGAaTG
CTGGAAAAACCAaTctGGAGTTTGTTGATGCCAAACCAGTGTATGACCTCTTAT
TCCtaTATTATTGCTCTGTGCAAAGTCAACATACTGGTCTGCAACACTGAAGTT
GAACTGCCCTTCTGTTTTTtGTATAGCATCGGGTTtCATCTCATTTTCAGCAGTA
ATACTGTTGAAATGCTTTAAAACCATTGctCTTTCcaTcGGGTTTTGgAGCACTTT
gTATGGTATTGCAACACCAATTGAAAAGTACTGGCTGTATTGCTGGCAAAGTG
ATGGAATCTCCCAcTCTGGCTGTTTTGCTGCAGTTGgAATCGTAATCGGGtTTTt
GTCtATGACTGTAAaGTCGTCAAGGTAGAaGTCAAGTGTTgcATTTGGTGATTCT
ACGTaGAAtATTAGCTGGGTtGCTGTCTgTGGCACTGTATATGAATTtGTTATCT
CTGTCAACCCAGATGGAACTGTTTGCTGATACTTTATAGAATCaTAATTTGTAg
TgcCATcCGCATTCTTCCTCTGCATTGTAAGGGTTATCTTTTgaTCACTTCCACTG
TTCTGATaTACCCATATGCTGAATTGATAATCCTTCCCCTTCTCAAGCAGGTTT
GTCATGTCTATCTGTGCACCCTGCCATGTAGATACCcTTCCCGATACATACAA
ACTCTTGCTACCTGAATGCGCTACTGTgTCAActACaGTAaTCTTtGCATCAGAA
CcTGTTCCCCTTGGCTGCCAGCCTTCAGTGCCACCGCTTTCAAAATTGCAAGA
CTTAACcACCCCGGATGATGCCGGTTGAGCAGATGTCAATATTATATTGTCTG
CGTAAAATTCAAGCGATCTGTTTGAAGGAACGCATATGAGAAGTTTTACACTc
TGGATGGGATTTGATGTTGTGACTGTAAACTTGCCAAaTATCTGTTTCCAGTA
CTGTGGCATAACcACCTTATCAGCTACCTGTATATACCTTTCTccgctgctatctttcacat
acgcataaactgaaaaTCTCTGCGGTTTTACATCATTGTGATAAACAAACAaTGACACA
GTGTACATGGtATTTaCCGAAACaCTTGATGTAACATCAaCAATtGCTCcATCCC
ATGCAGCACTTCTACCAGAGACTTTTAAaGAaTAAtcGCCCTCaTAAGCTATcTT
GTAAACTGTCTCAACAGTTGTGCTATTcccccTTGGACTCCAGcccTGAGTTGTTC
CAttttCaaaatcGAACtttACATAACTGgCtgcAGGATTTTGCtG

>gene_30428 >isotig00421
ttttacttcaatatcaaaattagaaaagatccaaccaaatttggCATCCTTTATATAAGGACTTCcAAACTC
TGTTCCTATTTCCcAATCTAAAaCAtACATGTTATCGTAATCGGTAACCTTAGAT
GAATATTTGgAAATTACATCCTTtAatTTTACCAAGAAAaGCTTTATAGGAAtCTT
AACTTCtCCTtCCTTTAtATtCTCTCTTGATTTAAAGGCAAtGTAATCCCATCCTCC
AGGTGAGAAGTAtATATCCCATATTAATtCTTTTtCTTTtCCaTTTAAAATGATAG
GAATCTtGGCAGTATCAACAATTTTCCCAGCcGGACCCAAATTATTGGcGTatAA
CCATACcATCATTtCAATATCCCCATTTGTTACAGTGTTtGGTTTTtCAGTCTTTG
TAATCCATGTTTCCATAGCAAAATTTATAGGAAGATTATTTTCGTACCATAGG
GAATATTTCATTACAAAATTGGCTTCTGGAAATTCTgAAACCTtCACAGGAAga
ggAAGCTTTGGAAGTGAAATACCGTGGGAAGACCATGGCTTGTATCCATAATA
AACTTCAGGGTATCCATGCACCCAGGAGCTTGGATTTCTAAGTACAATaTtCTT
TATATCCGCATAGTACTCTATGGCATCATCCTTTTCATAAAAAGCCATCCAAG
TTTCCCCCTCATAggTgGCAATGTTCCAGAAATTAAGCTCCATAGTAAGGGGA
AGATTTTTAGTCTTTATATCTCCAGAGGATGGCTTATTcAAAATAAAGGCATC
CTT

>gene_30429 >isotig00421
ATGGTCTCAAGGATAAAGCCATGGGGAAAAGTGGTaGgTTTtGAAAAAaGGGA
TAAATCACTTTTtGTCTTCGGAGAGGAGGAAGATTTAAGGATTGATGTGCTAT
CTGAAAGAACTATTAGATTTCATTATTCAAAAAATAAGAATTGGAcGCATAAT
TATTCCTTTGCCATAGAAAAATTTCCTCCTCTTAAAGATTTTTCTGTAACGGA
GAAGGATGGAGAAATAGTTCTTTCTACTGCCTTTTTAGTAGTAGCCATAAATA
AAGAGAAAGGAAAGATAAGTATTTTTGATAAAGAAGGTGATTTAATCCTTTC
TGATTATGAGAACTTAGGGTATGGAAGATTTAAGAATAAGGTATTTTCATAT
AAAGAATTGAGAAAtGaGGAGGCATTTTtAGGATTTGGAGAAAGGGTGGGAGG
TCTAAACAAGAAGGGAAAAAGaTTAATAAATTGGAATACTGATGATCCAAAT
CATTAtCCTAcTACAGATCCTTTGTATCAGTCCCATCCATTCTTcATTGCCTGGA
ATCCCAAAAAAaGTTaTGGAATATTctTTGACAACACCTTTAGAAGTTATTTTG
ATATgGgcAAAGAGAGTAATAAATACTACTATTTTtATGCAGAGAATGGAGAA
TTAGACTATTACTTCTTTTATGGACCAACTCCTAAGgATGTGATAGaGgagTATA
CCTATCTTACAGGAAGATACTATCTTCCACCTATTTGGGCATTGGGATATCAA
CAGTCCAGATGGAGTTATGCCTCAGAGGAAGAGGTAAGGGAGATTGCTAAA
AACTTTaGAGAGAGAAAAATTCCCTGTGATGTTATTTATCTTGAtATAGATTAC
ATGGAGGGATTTAGGGTATTTActACAAGCTGTGAAAGATTTCCAAaCTTTGAA
AAGATGATCAATGATTTGAATAGGGAAGGTTTTAAAATTGTAACTATAATAG
AtCCTGGtGTgaAAAAGGGATGTAAATTACGATGTTTATAAAGaggggTGgAGAA
GGATTATTTCTGTAGAAGATCTGATGGAACCGTATATATTGGACATGTTTGGC
CTGGAGAGTGCGCCTTTCCCGATTTTGTAAGGGAAGAGGTAAGGGTATGGTG
GGGAGAAAAaCAAAAAGCTTaattGgaaaaGGGaGTCTCTgggATATGGaaTGACAT
GAATGaGCCAGCATCTTttGAAAATGCCTCTACTTTTTGGGGGAAAACCCAAGA
TCCCGAAGAATCAAAGTACTATGTAAAGACCTTTCCTCAAGATGTGcTCCATG
GAAAAGGGGATAgaTTTACTCATGACGAGATCCACAATGTGTATGGTCTTTTG
ATGGCAAAGGCAAGCTTTGAAGGATGGAAGAGGGCAAGACCCAATGTACGC
CCCcTAATAATTACAAGGGCTGGCTTTTCTGGAGTGCAAAAATATTCATCAGT
ATGGACAGGAGATAATAAAAGTTGGTGGGAGCACCTTTATATGTCCTTTCCT

ATGCTCCAAAATCTTGGTATATCAGGGGTACCATTTATTGGGGCTGATGTAGG
AGGCTTTGGGGAGGACTGTACAGAAGAGCTTTTTATAAGATGGATTGAGGCA
GGCATTTtcTATCCCTTCTtCCGAAACCATTCCGCCATTAAtACAAGAAGGCAAG
AACCATGGTCTTTTtCCGaGGAGGCGGAgAAAATAGCAAGAAAATATATAACC
CTTAGGTATCAACTTATaCCTTACCTTTACTCTCTTTTCTGGGAAGCAAaGGaG
AAGGGTATTCCTCCTTTGAGGGCTCTAATCTTGGAATATCCAGAGGATAAGG
AAGCAATCCATAATGATGATGAGTTTATGCTTGGAGCaCATCTTTTAGTGGCG
CCCATATATAGGgAGGGGGCAAGAGCAAGACTTGTTtACCTTCCTCATGGAGA
ATGGTATAACTTCTGGACAGaGgAAAAaTTAGaGGGACCAGCTTATATATCTGT
ATCATCCCCTATAGAACTTATTCCtATTTTTGTAAAGGCGGGAACCATTCTTCC
TCTCTGGAAAAGCCAAAACTATGTAGGAGAAGAAAAACAGGATGTTTTAGAA
TTAAGGGTTTTCCCAGGAGAAGGAGAATTTATCTATTATGAGGATGATGGAG
TAaCcTGGGATTATGAAAAGGGAAAATTCAACcTGATAAAGTTCTCCTTGAAA
AaGGAAGGTAATTTAAAGATAGAGTATATTCATAAGGGTTATCAgTCAGAAA
GAAAAATATTTAGGATA

>gene_30476 >isotig03715
TAAAAATCTGTCTTCCATTGCTACCTCCTACCCTAAATATTGTTTTATTATATC
TTCAGAAATTCCATTATTCTCAACGAtATTTTtATAAATTTTAGCACTTTCTCTC
CATTTTCTTTCCAAAGTATTATAATCCATTTCaAAAaGCCCAAATCTCTGATAA
AAGCCTTCCGCCCACTCAAAATTATCCATGAGGGACCAATGTAAATATCCTTC
AACCTTTACcCCATCTTCAATAGCtcTATATAGCTGTAATAGATGGGATATGAG
GAATTTTGGTCTCTTATCATCCTTTGCATCCGATATACCATTTTCTGTAACAAT
GATGGGAATATGGGTCATCTCCcAAAaTCTTTtCAATACCCTATAAAAaCCTTCT
GGATAAATCTCCCATCCAAAATCAGAGCTTTCCACATTTTCAGGAACATACTG
CTTAATAAAATATGGAAAATCTTTAAGGGAAAAACCTACACGTACCCTTGTA
TAATAATTTATCCCCAAGTAATCTAATGTATTCTTTACATAAGGAATATCTTC
TTTTATAAAGGGAGATGAAAATTTTCCATTAATGATGGCATTTATAAAGGTGA
GATTATAAAGATTATCCAAAAACCTTTGGATTTTTCTATCTACATAGGAGTTT
TCTCTTATTGGATCAAAAaCCATAACATTATAAGCAATGCTTACTTTACTATCA
GGAATATTTTCATGAAGAATAGCATAGGCTTTTCCATGAGCCTTAACCATATT
ATTTAAtaCTTTCAACaTCTTAATTATACTTTTTtCttGAGGGGGAAACaTTCCCAT
AGCATAACTGGTAAAAGCATATGCATTAGGCTCATTAATAGTTATCCAATACT
TTACAAGATCTTTAAAAGATTTTGCAACTTTTTCTGTATACCTTAGAAAATAA
TCAATAGTCTTAGGATTTAACCATCCCCCCTcTTTCGATAtccATAAAGGATTTG
TAAAGTGATGAATAGTTACAAAGGGCTCGATATTATTCTCCCTTAGTCTAATA
ATTATTTTtCTATAaTGTTCAATGGCAGAATCATCAAAATaGCCAGGTCTTGGC
TCAATTCTACTCcaCTCTAaGGAAAAACGaAAGGCATtATTATTTAGCTCCTTCA
TCAACTTTATATCATCTTCATATCTATTCCAAAAATCACAAGCAATGCCTGAA
ACCTGACCATCTCTAACCTTTCCCTGTTTTTCAAACTCCCACCAATCATTATTC
GTATTATTTCCCTCTACCTGATGGGATGCCACAGCAGTTCCCCAAAAAAacCC
ATCGGGAAATTTATAAACCTTCAT

>gene_30590 >isotig03738
tttgtaatgtttatttgccttttcattccaaggctgagaaagaatagtttccactaaatttttgcttgttcaaagctctttgttcttggATC
ATCAGTTGTCATTAAGaTtAAGCTTGTTCTGTCTTCTTCTCTGAaTGCCTGTAGT
ATCCTCTCCATTCTAAGCCAACGAGGAATTAACACATAAAGCATTAAACGAG
AAGGAATTTTCTCAATCTTCTCCCTCTCTATTCCTTTTTTACTTACATAAACAG
GAATTTCCACAATCACATCATCGGGAATTCCTGGAATTGCGGAATTGTTAGG
AACATTTATAACCAACCTTCTTCTATTATCATTCACTATAGCATCTACAAAGG
GAATTATGGACTCCATACTTTCCACTACGGGAAATTCTTTtGTtACCAAAGCAG
AGAAATCTTTAGCTAATCTTTGCATAATCCCtAATCTAAATTTCAAGCTATAAA
GGTATAGACCCCATCCAATCTCAGAGTCAGGTCCCCCTGCAGgAAaCCATTTT
TCTTTtCATTTAGATCGGTATGGAACCACCAAGGAGAGACACTTCTtGGGGTAT
CTCcAATGGGgAAATAGCCcGTATAATTTATACATCTCTACACTGGCTGGAGA
AAGTTGGGTATATTCCCATGGAGATATCATGTAATCATaGgAATTCCAGAGCT
CAGGTGCCTTCTTTTCTATCCACTCATCTATTAAAGGATATGCATTTTTtCCTTT
GTATAGGAACTTTGTAAGAAAGATGCAGTGATTGAATCCTGCTACctGCTCTTC
CACCTCATTTcTATCCAATCCTAAAGCAcTTACAAATGGAACAATTCTAAATAT
CCaTGACAAAATCCTACCACTTTCAGACCAGTAGTTCTCAAAACTAaGTTTGTt
CccTtcAAAAACTGGATTAGATACctGCAAAAACCAaGCATTAGGGCTTATCTTC
TCCACATCCTTTGCTAAATCCATTATGAAATTTAGCTGTTCATAAGCGGCGAA
ACCACCATagTAATCACTTACCCTATCTCCAACTCCTCTgTAATATCCTAagCTcT
CTGCAAtCtCTCTTTCCTTCTCCATAAAATGATaCCcgCCgaCTTTtAcaGTgTTTAT
TACAAAATCTGCTCCATCAATGGCTTTTTTTCtATCTGTGGTCTTTTCtAATTTC
AAAactGTCCCTGTTTCTTTTATATACCTTTGAGCGAGATTATAAACGAGTTCTA
ACCTCTCCTCATCAaTATCCATAAaGGTGACagTACTTCCTGAAAGTCCCTTAA
ATACATTTAAATCTTGAATAAGTCTCATAGAGAAAGCTACACTTCCTGCACCA
ATAAtAGCAATcTtAaCCAT

>gene_31557 >isotig04097
ATTAAAGTTTACAATTtCCcAGAACGCCTTCTTTGGATTTTGATTTGCATCAAA
AATAAGAGGATAATCTTtCCTACCTCTCACAGGCCAAAAaTCAAGCCATGTAA
CTCCATCAGTTACGCCCCAAAATGTTACTCCACTAACATAATTTTtATTTCTCC
TTAATACTTCAAAAaTTCTCctGTAAACTTGAGCTTGAATCTCTAAACGATCCT
GAGGAGGATTTTTTGAGAAATTCGCATTCTCATATCTATCCCTATATATAGAA
ACATCAAGTTCTGTAATCTCAACCTTAACACCTAAAGAAGCAAATTTCTTAAT
TGATTCTTCAAGCATTTCAGGAgTAGGCCATTGTAAAAGCCAATGCCCCTGTA
TTCCTACACCATGAATTGGAATACCCTTTTCTTTtAATTtCTTAACCAACTGATA
TGCCTTATCCCTTTTTTATTGGTTCCTCTAAATTGTAATCATTGTAAAAgAGCAA
AGCATTTGGATCTGCCTCATGCGCCCAAATAAAaGCTTTCTCTATAAcTTCCTC
TCCTCCAATCTTATACCAAGGAGCATCTCTCAAAAATTCACTGGGATTATCAG
AAATTGCTTCATTTACCACATCCCAAACAGCtACTTTACCcTTATAATAACCAA
CCACcTTCATAATATGTTCCTTCAATCTCTGCAAAAGTACCTCTTTACTTACTG
GATTTCCgTTATCATCCCTAAAAACCCAAGCAGGCACCTGTTGATGCCAAACT
AATGTATGaCCcCTAACTCTCATCTTATTCTCCATAGCAAATTCCACAATTTTA
TCTGCAGGCTCAAACCTGTAAGTTGAGGGGgTAGGATGAATTATTTCCCATTT
CATTTGGTTCTCAGGagTCAAACTATTAAAATGTTTCTTCAAAAGCTCATCATA
AAACTCAATAGTTTCAGGACTTACTGCTGCACCAATAGGAAAGTAGTCTTTGT

AAACCTGCCAAAGAGAAGGAATCTCACTACCCATAACAACCTCCTTACTTGT
AAAAATTAAAGTTATAGAAAGGGATAACATGAGGATGAGTAAAATTTTAAA
GGCTCTCAT

>gene_31560 >isotig04097
TAATAAGACCTCCAaTCTTTCATCAGATTTAAACTTCTTTTCATTGATAAAAAT
CTCTTTAAGAGGAGGAAAGCCTCTAAGATTTATATAGAATTTTTtAACTCCAC
CTTGATATCCGGATTTGGTTTTCTCAATATCAATTCTCAAAACCCTttCTTCTAA
GTTAAATTTTATTTCATATAaATTgTACTCTCCATTTTTATAATTCATTGTCTTT
CCATCATCTTCATAATGAATATATTTTCCTTCTCTTCCCCCATATATCTCTATA
ATTTTACTTTCCTTCACATCTTCAGTATTTTTCCCCACTTCTTGCTTAACTATTA
TACTTCCTTCTTTTACAAACAATGGTATTCTATCTATtGGAGCATCTACCAAaT
AATAATTTTCCCcTtcATATTTActAAtCTGCCAATAATCATACCATACACCTTTG
GGAAGATAAACTTCTCTTTTGTTTTTAGAGGGAAGATAAATTGGCGCTACCAA
AAAGCTTTCTCCTAATAGAAATTCATCATATAGATTATGAGTATTTTCATCTtC
TTGATATTCAAAAACCAAAGGTCTCATTACAGGATAACCTTTTTTgCTTGCTAT
ATAAAATAAGTCATAAATATAGGGCAAAAACTCATATCTCATTTTTATATATT
TtCTTGCTATATTTTCGTATATTTCACCAAAAGCCCATGGTTCCTGATTCCTTGT
ATCCAAAGCAGAATGGTTTCTTAAAAaGGGAaTAAATATACCAGCTTGTAACC
ATCTTATAAATAACTCTCCATCGCAATCCCCAAcAAATCCTCCCACATCCGCT
CCAGCAAAaGGAACTCCAGAAAGTCCCATATTCATAATCATGGgAAAGGACA
TAAGTAAATGTTCATAcAAACTTCTATTATCTCCTGTCCATATTGCAGAGTATC
TCTGAATTCCaCTAAAACCGGATCTTGTTAGGATAAATGGTCTTTCATTTGATtt
ATATTTtAAAAGTGCATCTTTAGTTGCTATTGCCATATATAATCCATAAAGATT
ATGAACTTCTTTATGtAAAACTTTTTCACCATCCAAGATATgAaTATTGTCTTCA
GGCATCGTCAAATTTGGtGTATCAAACACTGATGGcTCATTCATATCATTCCAA
ATtCCATCTATACCgTCcTTTATAAAATCTATTAATTTTTCTCCCCACCAATCTC
TAACCCTTTTCTGGAAAAAATCAGGAAAACAGGCTTctCCAGGCCAAACTTTC
CcAACAAAAGTAaTACCGTATTtATCCTTTACAAAATAATCCCCTTCGATTCCct
CCTTATATACTTTATAGTCcGGATCTCTTTTTACCCCAGGATCTACAATAATTA
CCACCTTAAATCCTAATTTCTTTAATTCTTCCAACAGTTTTTTAGGTTCACTAA
ATCCTTCTTTATTCCAAGTGAAAACCCTATATCCATCCATATAATGAATATCA
AGGTAAATCACATCACAAGGAATATCCTTTTCTCTAAAGGTTTTCGCAATCTC
CAACACCTTTTCCGcaGgTTCATAACTGTATCTACTCTGCTGATAACCTAAAGC
CCATAAGGGGGgAagATCCATTCTtCCTGTTAAAAATGTATAGcTTtctACTACTT
CTTTtAtATTATCACcAAATATAAAATaGTAGTTCATTTGCCACcATCTGCACCA
AaGAAATAATACTCTTCACTCTTcTTTCCCATATCAAAaTACGTTCTGTAGCTAT
TATCAAAAaaTATTCcGTAaGTAAATTTTTACTTAATCCAaTAAAAAAGGAAaGG
ATGCaTAAAGAACATTTGTtGTTTTATTATGAGTTCTAAATTCATCAGTATTCC
ACATTaCAAAATCTTCACCTTTTTtATCTAAATAACCTGcttttCTCCAAATCCaTA
AAAATGATCATCTAAAAGCTTCTTATAaCAAAAAACcTtCTCATCAtcAAATTCT
ACCTCTTTATAATCTTCATTTATAAcTtCATCTTCTTTATTTAAAAAAGCAACAG
TTAGATCATCCTTATTTATTCTTACTTTTAATgCTGAGCTTTCCAAGGTTAGAA
ATTTCTCATCCTCTTTCACATCATACTTAGCAGGAGCATATTCTTtCcTTTCTaT
GGCTATTGTATCTTTTCTtCTCTCATCTTTATTTGTAAGGAAAATGTTTACAaTT

TTATCAGTTATgAATCTAatcTCTATtCTTCTATTCCcTTTtcTAACTGATATGCTTT
CTATTTCTCTTATCAC

>gene_31680 >isotig04141
GGTTTCTCCATCACAGGCtAAAAGAAGGAGCCCCGgCGAAGAAGGgTCTCCCC
TTAAGGATTCCTTTATGAGTGTTTCTCCCATGAGGTcTCCATTCTTGAGCAtGT
CGCCGAAGgAGAGCTCCCCCGATAAGGCTCGGTGAAaGAAAAaTACCCATATG
TAaCGGCCATTtCCGAGAGATACCTTGTAAGGTCTACTTACGTCTAAGgACGCC
TCGTCTACCTTtAtCCATTTCCCATTTGGgATTTTCACTTTTTCACATTGATGTGC
TCCCAAGATGgTGAACTTTATtCCTTCTTCGGAAAGTATCTTTAGCGTCTTTAA
ATCTACCGCCGTTTCAGGAAGCCACATCCCTGCCGGCTTTCTACCGAAGGTGT
TCTCGAAGGACTTTATTCCCCATACAGTTTGGATTTTtATGTCTAATtCCTTtGC
GAGGGGCATTATTATATGACTGTGTGCTTGGGCTATGGCATTGCCGAAGCCG
AACTTTTTCTGGTTTTCTCCGTCTATTTCCTTTAAGATACTTATAAGTTCAGGA
TGGCTTTTCTTTATCCAAGAGATGAGGGTTGGGCCGAAGTTgAAGCTTAAATA
CTTGTAATTGTTCCATATCTCGTAAATCCTACCGTTTtCGTCCAATATCCTTGA
GGCCGTATTAGGTAGATAGCTTTC

>gene_31697 >isotig04150
ccaggagtatttgatATAAcccATGTagaaatgggaaagagcaaggatcaaatagtatttagagttaaaaTtAGGGG
aGAATTaGAAAaTCCATGGAaTtCTCCAATGGgTGTTtCCGTCCAAACTATAGAT
ATATATATTAATGATGGAaCTGgAAAGGTCCTCTATAACAAGGCATTACCGGG
ACGAAATGCAGAGAtACCTGAAGGGTGGAaCAAAGCTATATGGGCAGAAGGA
TGGTATCCTGCCTTTGTGGAACCAATTATTGATGAAAGGAAATTGACTCTAAA
GGATACTAAAGGGgTAGTTCAGGTTATGGCAGAACCAACGGAAaGAaTGATA
ATTATAACAGTGCCaGAAAAGATTTTGGGAACTCcTTCTCCCGATTGGAAGAT
TCTAATAATtCTTTgtGGACAAgAaGGgTACCCAAGaCCTGGAAGTTGGAGGGTA
AGGGAaGTTGAAGAAaTGGCTTCTCAATGGAGATTTGGCGgTGGCGATGACTT
cTATGGTgATCcTAAtaTTgttGATATGATTGTtccaCcAGgacAAAgaCAaGaagAtatcTT
t

>gene_32645 >isotig04527
CCCCTCTATACCATTTTCTTCAATCACCTTTTTATAGAAATATGCACTATCCTT
TAATATCCTATTCTGAGTTGAGTAATCGGTATATATTATTCcAAACcTTTtACTG
GTACCAAAaGCCCATTCAAAATTATCCATGAGAGaCCAGAcAAAATAGCCTCT
TAAaTTAACTCCATCTTTAATAGCCCTGTAACATTGATAAATATGTTCTCTTAA
ATAATCAATTCTCTTTTGATCCTCTACTCTAtcATCTTTtaGTtCATCTTTAAAAG
CTGCTCCATTTTCGGTAATTATTATGTTGTTAATATAGTCCTTTGATAATCTAA
CAAGAAGATCATATAAGCCTTCTGGATATACTTCCCAATTCATTtCCGTTTTTtC
gTTTGGTCCCTCAACATGTTtCAaTGGAaGGAAaGgATTagAcGGATCATACTTTA
CTATTGTTCTTGTATAATAATtAaCTCCCACAAAATCTATTGGAGTGgTTATAAT
CTTGAAGTCCTCTgAaTCATagctGAAGTTCCATTtAAATTTTGGAAGAAGATTTT
CTACcTCTTCTGGATATTtACCTTTAAATAATGGATCTAAAAACCaCCTATTGGT
ATACGCATCTTGCAACTTTGCACACtCtAAATCTTCTTCAGAATCTGATGCAGG
ATATACTGGACTTAGGTTTAAagCTATCCCTATTTCTCCTTTTACtCCATTATCA
TAAAAGGCTTGCACTGCCAATCCATGTGCTAATAAAAGGTtATGAGATGCAAT

AAAGgCTCCCTtCATATCTTTTTtCCCAGGGGCATGAACCCCAAatccATATCCAA
Gaaagcattaacccaaggctcatttag AGTAATCCATATTAtATCCCTGTTACCAAAtaCTTTAA
AAatAAaaGTtGGCATACTcTGCAAAATacTTCGCAGtATCtCTATtCAACCATCCA
CCTTtATCTTCTAATGCTTGTGGTAAATCCCAGTGATATAGAGTTATTACAGGC
TTTATATTATTTTTtATCAGagTATCAATTAACTGGCGATAAAAATTCAACCcTT
TATAATTAACAGATCCCTTTCCTTCTGGAAAGATTCTTGCCCATGAAATTGAA
AATCTATAACCTTTTAGACCAATAAATTTCATAAGTTCAACATCTTTTCTCCAT
AAaTgaTAATGATCaCAAGCTATATCACCATTTTGTGATTCAAAAAT

>gene_32767 >isotig04429
AACCTCTATTACCAAACCTTCAAAGGGAAGAAGATTGATCTCTCTATCtATTTT
TtCtCCTTCTTTCCTATGGgTTCCAATAAaTACTTTACCcTTATAATCAATTGATA
TTTtCTTttCTTtATCACTgAAaTTTAaTACAaTtAATTtCcTTtCctCTtCaTAAATCCTC
ATATAGGCTAATATAtCgTttATTCCAGGAaTTaCAAATTCAAGATcTCCAAGTTT
TAATGCTAAAGaTTcTCTTCTTTtatAAAttaaatttctgtaataatttaagatggaatt

>gene_33001 >isotig04516
AaGATAGATATAGAGACAGAgaTATTTTGGAATGAGAAGAGGGTGATGTTGAG
AGCTTTgTTCCCATTAAaTATCAATTCATATaTTTGTCATTATGAGATACCTTAT
GGGATTTATTATAAACTTACCCATGAGAATACCTCTTGGGAGAAGGCAAAAT
TTGAGTTTCCAGTGCAAAGATTTATTAATTATTCTCAAGGTGATTTTGGAATA
AGCTTGTTAAATAATGgAAagTTT

>gene_33199 >isotig04598
GTTGACTATACAaGgTACCTCTACTATAACATCCTTTGGTAATTCAGATATTGT
TCCATTATTAGGAATGTTCAAAAaTAATCTCTTTTCTTTATTATTAACTATGgC
ATTTATAAATAAAATCTGTTGTTCCCCACTCAATTTATCTTTAGAGAATTcAaT
AGGcAAAATTTCTGTTAAaTTAATATTAGGGTTTTCTAAAATCTGTTTTGCAAG
TTGAAGTCCACTTCTTCTTAAGGCTCTAAGTTGTTCATAAAATTTGGGTCTTTC
TAtttCATTATCTATACCACCAAATTTCCCAAGCCATTTTTTCTTTGTTTCAAGAT
TGTAaTGATATTTCCATGTGCCATTTCTTACCGTGTCTCCAATAGGGAGCATTC
CATAAAATTTGTACATATCTATGGCACAAGGACTAAGTTGTAAATCCCAGGG
ATCTTTTGGTtCCCATTCATGGGATTTTTCTTCTATCCATTGGTCCAGAATTTTG
tATGCGTCTTCGCCATTATATAAAAATCTATTTAGAAAAaTGCCATGATTTACT
CCTGCTATTTGCCagTCTACTTtATtGGGATCTAAGTTCAAAGTaTTtAtAACCTCA
TAAACCCCaTGAAAACCATGACATATTCCAATAaTTCTTACATCTGTTTCTCTG
CTTACTAATTGtGTAATTTCAAAtACTGGATTAGCGGTTTGAAGAAGATAAGCC
TTTGGAGATAaTTTaTcAaCGGTTTTtGCAATATcTAAaGAAAGTTTTAATTGGTT
AAAATTGgATAAAGTGTAATAATCAGAAACCATATTAAATTCTTGGCTATCTA
TACCTCTATAATATCCATGTTTCTCCCCTATTTCTCTCATtATTTCATACTGTAC
ATATCCATCTTCATGtCCTTTTGCTCTGTGTAAAGCTGTATTTATTACAAAACT
TGCAGAATCAATAGACTCTTCTAAACTTAGTGTTTTTTCGATTTTTacatc

>gene_33327 >isotig04663
CGCATCACCATATGGACTATAAGTCCAAAAaTCTCTTTTAGAAAATCCATTAA
TGAAGTCTTCCCACTTCTTAATATGATCAGGAATTGATGTATAGAAGTGAAAA
TCCTCTATATCAGTTTTTAAGTGAAAATTGGGAACACATGGAGAATTATCCAC
TACAAGCCTTGTGGGATCAAGATTTTTAACATAATCATACATCTCTTTGAGCC
ATTCTCTGtTTTCTCTAaTACTtAGATCTATTCCCCAGCTCTCATTTATAATCCCC
CAAATTACAATACATGGATGATTAAAaTCTCTTtCTATCATTTCTACTAaGGTTT
TtCTTGCTCTTTCTTTTGCTTTTTCTGTCAAATTTTCCCAATAAGGAAGTTCCTC
CCAAATTAATATCCCTATATGGTCCGCCCAATACAAATATTCAGGTAGAGGG
ACTTTTACATGAGTTCTTAGAAGATTTAAACCCATCTCCTTTGCCTTTTGAAGT
CGCTCTCTTATTTCCTCAGGAGATTTTGGTATATATATCTTTtCAGCATAGAAA
TCCTGATCCAAGGCTCCAAAAATATATATGGGCTTATTGTTTAGATATATTTC
TCCATCTCTTATCTCAATCTTTCTCATTCCAAaTtCTGtTTGAAACATATCTCTtA
ATATATTTCCTTTATATACTTtAACTTCTAAaTtATAAaGATTTGgAGCTTCCAaT
TCCCacAAAATTACTTCGGGAAATTCAAACTTAACAAtCTTTTTTTCTGAAGTT
ATTTCTACCTCtTTaTAATAaTCCcTtCcTAAAGgATCTATAaTTTtAagATTTAAcTT
AtAAtCCTTTTCTAAATCTTTAATGAAGATTTCAATTgTaGCACTTTTATCATCC
ACATTTGGAAATaCTTTTACTTTCTCTACATATGGAGcttCtACTTTTTCCAAATA
TACTTTTTG

>gene_34145 >isotig05157
TCCATCCgCTTTTTCAtaTTTATCTACTATTACGTCATCAGGAATTCCTGAAATT
ACACCTTTATTtGgTATaTTTAGAATAaGCcTTGCCTTATTtCCGTTAGTAATTGC
ATTGATAAAAaGAATCTGTTGCTCTCCACTTAATTTTtcTTtaGAAAATACTTCTG
GATcAATTTCAGTTAaTTTtatAGATGGATCCTTTTTCGCCTCTTCTGCTAATTtAa
TAaGTCTTtCTTTTTCAGCCCGCAGTTGATTATGAAATTTtGGTcTTTCAATTtCA
TT

>gene_34308 >isotig05287
AATTTTGGCATTACAGCCGAGGGTAGAAAATATCTCGTCTTTCAAAGGAGGA
CAGTAATTATACTCATCCCCTGTATCTCCTCCATCTTCAAATATATTAATATTC
TTGTATATTTCTTTCGTTTCTTTATCTTCTATTACCAAACTTCCATTGGACTGTA
TAAAAaCTTTTAAATATTCATTTTCCAGCTCTCCAAATCTTTTTGAAATATTGT
CAGGAAATATAAAGAGATATCTAGTGTTACATCTTTCTATTTTTAACATTCTG
AATCCTAAAGAGGGTAATTGAAGTTCTAAAGCAATGTTATATCTTTTtCCTTTT
tCcATATTTCTCAtATTTTTCTCATCCATATAAAaTTTGTtGTAATCTTC

>gene_34397 >isotig03913
ATGGCAAATATAAAAATTTCCTTTgTAGGGGCAGGcAGTGTCAGATATACATT
GAAATTGTTAGgTGATATTTTTAAAaCCAAaGAACTTTTTGGAGttACCATATCT
TtAATGGATATAGATGAAGAAAGACTTAATGCCACATATGTATTAGcTGAGAA
ATATAGAAAAGAGCTTTCagcagATGTAAAAATCGAAAAAACACTAAGTTTAG
AAGAGTCTATTGATTCTGCAAGTTTTGTAATAAATACAGCTTTACACAGAGCA
AAAGGaCATGAAGATGGATATGTACAGTATGAAATAaTGAGAGAAATAGGGG
AGAAACATGGATATTATAGAGGTATAGATAGCCAAGAATTTAATATGGTTTC
TGATTATTACACTTTATCcAATTTtAACCAATTAAAACTTtCTTtAgATATTGCAA

AACCGTtgaatAAATtATCTCCAAAGGCTTATCTTCTTCAAACCGCTAATCCAGTa
TTTGAAATTACaCAATTAGTAAGCaGAGAAACAGATGTAAGAATtATTGGAAT
ATGTCATGGTTTTCAtGGGGTTTATGAGGTTATAAAtACTTTGAACTTAGATCC
CAATAAaGTAGActGGCAAATAGCAGGAGTAAATCATGGCATTTTtCTAAATAG
ATTTTTATATAaTGGCGAAGACGCATaCAAAaTTCTGGACCAaTGGATAGAAGA
AAAATCCCATGAATGGGAaCCAAAAGATCCCTgGgATTTACAACTTAGTCCTT
GTGCCATAGATATGTACAAATTTTATGGAATGCTCCCTATTGGAGACACGGTA
AGAAATGGCACATGGAAATATCATTACAATCTTGAAACAAAGaAAAAATGGC
TTGGGAAAtTTGGTGGTATAGATAATGaaaTAGAAAGACCCAAATTTTATGAAC
AACTTAGAGCCTTAAGAAGAAGTGGACTTCAACTTGCAAAACAGATTTTAGA
AAACCCTAATATTAATTTAACAGAAAtTTTgCCTAttgAATTCTCTAAAGATAAA
TTGAGTGGGGAACAACAGATTTTATTTATAAATGCcATAGTTAATAATAAAGA
AAAGAGATTATTTTTGAACATTCCTAATAATGGAACAATATCTGAATTaCCAA
AGGATGTTATAGTAGAGGTACcttGTATAGTCAACAAAAaTGgTatATTCCCTGA
AAAAaTTGAaCCAGaTTTAACAGAGAgAATCAAAAAAaTGTATTTATTgCCGAG
AATATtAaGAAtGGAATGGGCACTtCATGCTTtCAAAaCTGgAGAtCGTAGAGTCT
tAGAaGAAATaTTAATAAGAGaTCcTCGgaCAAAGTCCTaTGAAcaGGTAAAAAaC
gTAATTGATGAAaTATTAaGTCTGCCTTTtAACGAaGAGATGAGAAAATAttatcaaa
aggaggagtaaaaaGTATGAGTAGATTTCAAACATTAGATAAAATTATTGATTGTGG
GGTT

>gene_34635 >isotig05525
TAACTTaGgAAAATAATCTTTATTTtCTTCTATCAATCTTTTTAAAACCTTCTCT
GCTAAAGAAAGAGAGGAGATTAATGGATGTTGAGCAAGAGCCCATAACGCC
TTTTTAAAAGAgCCTTCAACAGCGGATTCAATAGTTAATTCctCATAGGTTTtAA
CTGCTTGAATAATACCCcTTAcTTCTTTAGGCAAATTACCTATAACATACCTTT
GAATCCTTTCTCCTTCTATATAAACTGGAATTTCTACCACTACATTATGAGGA
AGATCAAAAATTGCTCCCTCATTAACGATGTTTACAATTTGAAAACCCTTCCT
AAGTCCAAGAAGATGAAAAATCAAGTTTACAGCAGATTTAGAGTATAAAGCT
CCTCCTCTTTTTtCTAATtCCTTAGGCTTTTCATCTAAATTAGGatCTCTATAAAG
CTTTAATAAATCTTCCTCAATTTTCATTAcTtCTTCTGCCCTTTTAGGCTTACTC
TTCATCTCCATAACCTTCTCTTCTCTAAAGTAGTAATATCTCAGATAATAATTG
GGGAACATGTTTAAATAATCCAAAACTTTTCTTTCCTCTTCAGAAAGATTTTC
TTtAATCTTaTTTAATGCCTCcTCTGTTCTATCTTTtCcTTTTACaTAAaTcTTtCTTA
CAAAaGtCAGATGATTAaGTCcAAAaTAaTCCATAAAtAtATCaTCCATTGaaaccccA
AAAaGCcTTGCAAATTGATTTTGGAAGTTtATAGCAACATtACATAaTCCTAAAA
CTTTaGCCTTTGTAtATTTtGATATTGCTTCTGTAATTAtACCTgAGGGATTAGCA
AAATTTATTAACcAAGCAGATGGAGAAAGCTCTTCTATCTTTTTGGCAATATT
TAGAGCTACAGGAATTGTTCTTAAAGCATtAGCAAAaCCaCCCaCTCCTGTAGT
TtCTTGCCCCAAAAGaTTAAATTCcAATGgTATtGTTTCAtCAAGTAAtCTTGCTcT
CTGTCCCccAACCcTTATCTGATTTATGACAAAGTCTGCATCTTCTATAGCCTG
CTCTAAGTGTGTAGtagtttttatattgatattagctttactctttcttgccaTtCTaatcAAAAAATTtGATA
CtATTtctAATCTTtCCTCaTTAaTatCCaTtAAaTatagaTCAATGGGGGATATCtCAGAa
GAAAGttCTATAAAACCcTCCATTAATTCGGGAgTGTATgTGCTtCCTCCTCCTAT
TACcGAGATCTtcAT

>gene_34676 >isotig05531
atcatatattccaaaactggaaCAAAaTtCCcAATAACTCCAAGCAATACCTCTCTTTtCAGC
AGACcTTGcgACAAAaGATGTCCATCTTACCCGGGATTCTAAATCTGCCTTTGA
GTATGCTCCAAATTCTCCCATATACAAAGGTACATTACCGTGCTCTTTTGCCC
AATCTACAGCAATaTCAAGCTCGAATTCTACAGTTctttttttCTTTAtCGGTACCCAT
CCACTTTACCCcAACAGGTGGAGCAGGGTTTaCCCATTCAGCCCCTTGATGAG
TAAATTGgAAGGGATTGTAATAATGGAATGTGACAaTTAAGTtCTTATCATCCA
ATGGAATTTCGAGCTCCTTTAGTTTATAAAGATTGTTCCAatCTGTAGGTCCAA
CTATTATCTTCCTTGTAGGATTAGTCTTTCTTATAATCTTTATTGCctCTTTTAG
ATATTGGTTCCACAGTTTgCTTGTtAAATTTTGGCAAGGTTCATTAAGAAGTTC
AAAGTATAACATGTCAGGATAATCTTTATAATGTTCTGCAATCtGTTCCCAAA
GCCTTAAAAATCTCTCCTTATGTTTTTCAGGCTCTTGCATAATCTCTTCATAAT
GGTGAATATTAACAATTATGTATAAATTATTCTTAAAAGCcTCATTTATAAGG
TGGTCTACCcTATCAAATATATACTGATTTATTTTATATGgCGGTTTATAATCA
GCATAGGCATtCCATCGCACtggAaGCCTTACATGATCAAAACCAGCCTCTTTA
ATTAGTTTAAAaTATTCATCTTTTATAACAGTCCCCCACTGACcTTCAAAGGGg
gCTTCTAAGGCATTTCCCAAGTTTACCCcACGTGCAATAGgTAA

>gene_34865 >isotig00376
AAaGTTGTTGTTGTGCCGCATACACACTGGGatAGaGAATGGtACCTGAgTTTTG
ATGAATTtCGCTTTCATCTAGTAGaGgCATtAGATCGGGTCCTTTCTCTtCTCGAA
GCTCACCCTCACTACAAATTTACCTTAGACGGTCAGGTGATCCCCCTTTTGGA
TtACCTCGAGGTGAAGCCGGAAAACGAGGCTGTGCTCCGCAAATACGTCGCG
GAAGGGAGGCTCCTCATCGGACCTTGGTATGTGCAGCCGGACGAATTTTtGGt
ATCGGGTgAGTCCTTGATACGAAACCTACTTTAtGGCACGCGGATTAGTCGAG
AGTTCGGTAATGTAATGTTGGAAGGTTACGTTCCCGATGCGTTCGGCCACATC
GCTCAGCTTCCCCAGATTCAAGGATTCGGTATTAActCAGCCTACGTCATGCgG
GGGgCGGACtCGGCGTGCGCGAAAGCAAACGGCCCGGATTTTGTCTGGCGTGC
TCCTGACGGCTCGGCCGTACTTTGTCACGTAATGGAGACGGGATATTGCAAC
GCCGACCAACTTTCGGCTGACCCCAAGAaGgTTCCTCCCCCcGTTGtAAGCCTG
aTAGAAGCTGGGCTTCTTAAACcTCaAaGTAATCCTCTCTTAGGgTTTTTGCAAG
AACTTTCtCGGCGAAGCCGTACTGAAGGGATCCTCCTTATGAACGGTTGCGAT
CACCGAGGCCCTCAAGAAGACctaCCagAGGTGGTaggACGAACTaAATGTGCGtT
TtCCTCATTTcaTGtTCATGATCGGTACCCTcaaCGACTACACTAGACTGCTTGAG
CAAACAAGGGACACACTTCCaGAGgTTCGCGGCGAATTTAGAACCcCCACAcG
GCACCCCATTCTTGCCGGAGTCCTTTCTACTCGAGTCTATCTAAAGCagGgCGA
ATCACCGcGTGGAAACGCTGCTTCTTAAGTAtGCAGAGCCCTTATCCGCGTTA
GCACGagTGCTTGGCGGaaaGaGTtAAGCATTtCCtTGaTTTAGCGTGGAAAACGC
TGCtTGAAAACCATGCCcAcGACTCGATTTGCGGTACGGGgATCGATTGTGTAC
ATCGGGAAATGGAATGTCGCTTtcTCcGGGcaGAaGCAaTTGGTAAAAAGaTAGT
ACATgCAAGTCTTCTtGCTCTtGCTCAAAAaTATATAGTagGGgAaGACGTtGTTTC
CGTTTTCGTTTTCAATCCTTGCCCGTGGGAGAGATTAGACGAGGTTTCCGCGG
ACATCCCTGCAGAGGCCGTTAAAGGGGCGCTCCTTGAtCATGcGGGCAGAGAA
ATTCCGTTCGTAGTCAAAGGCAAAAGCCTTGTATCTGAAAGAGTTCTTGGGGg
CGTTATTCATCATGAAAAAGCCACCATTACTTTCCCGGcAGCTCtccccgccTTTG
GCTTTGCTACATTTCGtATCATCCCTCGCCAACCGCAGAAgAAAGCCGAAaGCC

TCTTCgTAGACGACcataCATTAGAGAATGAGTTCTACAGAGTAGCAATCAGGG
AAGACGGGaCGTTTGATCTCTTGGACAAAATAACAGGTGTcACCTATCGTGGG
CTAAACtTTTTAGAGGACTCCGGGGACGCAGGTGaTGAGTATAACTTTTcgCCC
CCTGCTGAACAGgaTGTTATtACCTCGAgCGGTCTCcGAGGAGAAATAGAGCGc
GCTGAAGACGTCCCATGGAAAGGGTCGgTgaGGGTTAAACTACGacTGGTGGTC
CCAaGGAAGcTAGAcgCaCATCGAAAATCTCGTAGCAAAGagACAGTGGAACTA
CCAATTGTTTTTACGGTCTCCCtGAAGCGGGGAATAAGACGAGTGGAAATTGA
AgCtCaggTGGaGAACCGCGCCCGAGATcATCGGcttACGTGTAGCTTTTCCCACC
GGGATCGCCACGGAATCCTCCATTGCTGACGATGCCTTTTGGGTAATAGAGC
GGCCAACGCGcccaTcCCCTGGtGAAgaCgGgATTGAGGTTcCCCCcTGTTACACAT
CCGCAAAAGGCtTTCGTAGCGgTCGagAACGgaggAAaGGggTTTtGCCATCCTTA
AtcGAGGTCTTCCCGAATACGAAGTCacTCCAGATGGCACTaTCTATCTTACACT
tcTTCGGTGCGTGGGATGGCttCCCGAGACGACCTTaTTACGCGcCGAGGCCATG
CCGGCCCTCCCTATGAGACTCCGGAAGCGCaGTGctTtAGGCAtGCATCGCTTtG
AGTAtGCGATCTATACTTATTCAGGCACTTGGGAGAACTCCGGaGTTATACAA
GCAGCGCAGGAGTACTCtGCCCCTCCtGTGGCAATAGCTTTGAGAGGtGTGCTC
AAACCATGCTACGACAAGATCTTTCACGTAAAGCCGAATGCCCTTGTCTTAA
GCGCTGTgAAGCCCCCAGAAGAAGGGGATGGCCTTGTGGTGCGGGTCTACAA
TCCGACGCGCAAgGAaGTTATGGCCGAGCTCGAAACCGCTTTTCCtATaGGCGA
AGCTAAAGaAATaCGCCTTGATGAGTCTTCCCTTgCGGCgATggttccagcttcttcgcacct
cctgcgctttcccgtgcgctctggggaaaTAAAAACGTTAAGAATCAGCATTAGAAAa

>gene_35777 >isotig05829
tCCCAATCTATATCAATtCCATCTAAATTATATTCTTGTATAATTTCTAAaCAGC
TTTCaGCGAaGGTTTTTCGAGaCTCTTCAGTTAATGCCACATCGGAAAaCTCAC
CTGATAGGGTCCATCCGCCAATAGATATAAGTACCTTTACATCAAAaTCCTTT
TTgAaTTTATTTATAACCTTAATGTTTTCTTTAAATAAAtCcTCTTCGGTTGGATA
TGCAtGACCATTTTTTAtCTGAAAAAAGGCATAATTTAGATgtGTAATTTTATCC
CAAGGAATGCTTTCTATAGAaTAGCTTAGAAACTCCTCTTTAACAGACCATTC
AGGGAAATATGCCACAATTCtAAATTTTtCCAT

>gene_35778 >isotig05829
CTTGAGAAGAAAAGCAAAGGACTCgAAATCTTTTAGCCTTGGcaGTAAAATTC
CTATGTTCATCAGTTCGTCACCGTTATAGACCTTaTCTGTTCCATCGATTTTGT
ACTTTTtATCGGGATCTAAACCTTCGAGCTTAAGCCATTTGAAAGGGGGATTT
GGTTCTGCTAAGATTTGAACATATAAACCGAAAGCTTCatttttATCCTCTGTTAC
AAATATCCAGGCGCAAGCATTCTCTTCAAAGGGATTTATCAATCTATAAAAG
TctCCATGATGTATAATATGCTCTATTTTTTTATAAAATTCTATCTGCTTTTTAA
TAATTTCTTTTTCCTCTTGAgaTAATTTTGTCAGGTCTAATTCGTATCCAAAACT
ACCAGACATGGCAACATATCCACGGGTATCTAATGGAGTTATTCTACCTACTT
GATGGTTTGGTACaGCAGATACATGACAACTCATTGTAATAGGAGGATATACt
aTGCTTGTACCATATTGTATTTTCAATCTTTCAATGGCATCTGTATCATCACTG
GTCCAAATTTGAGGCATGAAATAtAAAAtCCcTGGATCAAATCTACCACCACCT
CCAGAGCATCCTTCAAATAAAACATGGGGAAATTCTTTTGTAAGAGTATCCA
TCATTTCGTATAAAGATAGTACATAACGATGAAAAaCCtCCTTTtGCCTTTCTG
GAGGTAGAGCCAAAGATCCAAcTtCTGTAaGGGGTCTaTTCATATCCCATTTTA

CATACTCTATAGgAGCACTTTTAAGAATTTCTCGCATCATTCGTAAAATTTCTT
CTCTCACATcTTTtCTtGTGATATCAAGAACATACTGGTTTCTGCATTGGGTTAG
GgTTCTATTTTtAACTtGTATACACCAATCGGGATGCTTTTTATATAACTCGCTA
TTAGGAGAAACCATCTCTGGTTCgAACCAGATTCCAAATTTCAGATTCATaTC
ATTAACTTTTCcTAAGCCATCAAGACCGCTTGGTATTTTCTTTTtATTAACGTAC
CAATCTCCTAAAGaGgAAGTATCATCATTTCTTTCTCCAAaCCATCCATCATCT
AAAACAAATAATTCAATTCCTAAATCTTTGGCTTCCTTTGCTAATTCTAATAA
TCTTTCTTCACTAATGTTAAAGTAAACTGCTCCAGGTATTTAAAATTATTGGT
CTTCTTCTATCTCTATAGTAACCTCTACAAAGATGTTTTCaTaTaGCTTATGAAA
AGTGCGAGACATccctcctaaaccattttcagAatatACCATAACTActtcaggagtttggaaactttctcc

>gene_36357 >isotig06067
aaaatttttgttgcttttgtatggcatatgcatcagccctattataaggataggagaaggatattattATTCCCTTGGGTtA
GgCTTcATGGGATTAAAAaCTATTtCAaTATGGTGAATATATTGTATGATTATCc
AAAGATtAAacAAGTCTTTAaTTTGACACcAGTTCTTgtaatgcagttaaaagattatgtagaaaa
TAGAGCAAAAGACTTATGGTtGGAAAAAACTTTAATTCCTtCTACGGATCTTTC
TGAGGAAGATAAaGAATTtATTTTAGATAGATTTTTCTGGCTTAATTTTGATAA
GATGGCATCAATACATCCAAGGTATAATGAGTTATGGaGAAAAAAaCAGAAT
AAAGAGAGTTTTTCCTCGCAGGATTTCTTAGACTTACAAGTTTTATATAATCT
TTGTTGGTTTGATCcAGAATATAGGAATAAAGATGAGTTTTtAAACTATTtATG
GAAGAAaGGAAAAAaCTATACTGAGAGTGAAAAATACAGATTAGTaGAAAAA
CAATTTGAGCTCCTTGGTAAAgTCCTTAGTTTATaCAAAGATCTGCAAGATAA
AGGGCAAATAGAGGTtATTTTTTCTCCTTTTTATCATCCAATtcTTCCCCTCTtAA
TTGATACAGATtCTGCAAGGATAGCAACTCCTAATTtGCCACTACCTGCATTGC
CCTTTAGATATCCtgAGGATGCAAAATGCAAaTAAAAaTGGAATTGACTTTtATT
tGGAAAACTTTGGGgTTTACCCcAGGGGATGTGGCCcTCTGAaCAGgCAGTAAG
TCCAGAAGTCATAGAgATGgTTTCAGATTTTAATTTAaGAtGGATtATTACTGAT
GAGAAAATTcTCTTTAAAaCCTTAAAAAaGGATTtGgTTAGAGATTTAGACGGaT
aTTTAGTTGAACCgCATGTTTtATATAAaCCGTATATaGTtAACATAAatGgTaagaa
GACAagTGTGGTTTTTAGGgATAGTTTCTTgTCAGATAAAATAGGATTTACTTA
TATGCATCTTTCCCCAGAGGAGGCTaCAAAGGACTTATATCAAAGGTTgATAA
ATaTTAAAAATAGATTACCTAaTGATTTTCCTtACTTAGTgaCTATTGCATTGgAC
GGTGAGAaCTGCTGGGAATATTATGATAATGATGGAAGGgaTTTTTtAaGAAGC
TTTTATAGCTTACTATCcAATtCATcTGATaTAGAAACAaTTACTGTAGgtGAgtAT
TtAgagAAaTTTtCtCcAgTGgATGAATTGAACtAaTaTTttCActggttcttggtanatgcagatctta
ctacatggataggaGag

>gene_36443 >isotig06104
GCCAAAGGCGCAGATGTTGTGATTGTTGTTGTTGGTGACAAAGCAGGACTCA
GGCTTGACTGCACATCCGGTGAGTCAAGGGATAGAGCGTCTTTAAGACTTCC
GGGTGTTCAGGAGGATTTAGTAAAAGaGATTGTCAGTGTGAaTCCGAACACAG
TGGTTgTTCTGGTAAATGGAAGACCTGTTGCACTTGATTGGATTTTGGAGAAT
GTAAAAGCTGTTGTTGAGGCATGGTTCCCGGGCGAAGAAGGGGCAGAGGCTg
TATGCGATGTCCTGTTTGGGGATTACAATCCGGGTGGCAAACTTGCAATTTCC
TtCCCGCGTGATGTTGGTCAGGTtCCTGTTTACtATGGgCACAAACCATCCGGCG
GCAAATCATGCTGGCACGGgGATTATGTTGAGATGTCTTCAAAACCGCTTTTA

CCGTTtGGCTTTGgTCTTTCATACACAaCCTTTGAGTACAGGAATTTCTCAATTG
AAAAAGAAAaGATTGGCATGGATGAGAGCATAAAAGTATCTGTCGAAATTGA
GAAcACAGGAAATTAtGAAGGAGATGAAATTGTCCAGCTTTATACAAGAAAA
GAGGAATATCTtGTTACAAGACCTGTTAAAGAgCTGAAAGGTTATAAaAGAGT
TCATcTAAAACCGGGTGAGAaGAAAAAGGTTGTATTtGAaCtttaTCCTGACCtGTT
tGCCTTTTATGACTATGATATGAACAGGGTTGTAACACCGGGTATTGTTGAGG
TTATGATAGGTGCATCcTCcGAAGATaTtAAATTTACaGGTtCTTTTGAAATAAC
AGgtAGCAGGAaGgATGCAACAAAGATCAagCACTACTTCAGCAGaGTGCTTTG
CGAG

>gene_36604 >isotig06178
ACGGTGACGGGGTTATCCGGGTCTTGGGTCCATGCACCGTCCACGATAAACT
TGTATTCGTATTTACCCGGTGCCAACTtCAAAACAaTCcAAAAGGTTCCgTCTTC
CTCTCTTCTCATGGGATGGGATGtGGTAgACCAGTTATTGAAATCCCCcGCCAg
ATAGACAGATTTGGCCTGGGGATTatAGTATTCAAACcTAaCCCCATtAGGGGTT
gACTTTACcGcTTGTACTAAAAAaaTTCCcAAAAaaTGCCcATAAACAAC

>gene_36731 >isotig06241
AAAGAGCCATCAAACACGAGGATGGAAAAAATCGATgaTGTaTGGGTATTATT
CCAATATAATAATCCTGTACCAAGCTTTGACACATGGAATAGAACTCCAAAG
GGAAgAGAATCTATTtCTTTAGACGGAACTTGGAAATTTAAAataGATCCTAAT
GATATAGGTATACGAGATAAAGTGCCTTGGTATTCTCCAGAAATTGAAGACA
GTAGCTGGGATGAGTGTCTTGTTCCTTCTGTaTGGGACCTTTATAAGGATAAA
GaTGGAAAGgATTACAGGTTttATGATGGTAAGgTTTGGTACcGCAGAAAGTTTA
TtGTTCCAAAGGATTGGAaatATAAATTTGTTAAGGCAAATTTTCTTGgTATAAA
TTATAGGgCTAAAaTATGGATAAaCGGAAAAtctGTTGCCACCCATgAaGGaGGG
TATACTCCATTTTCTTTAGATATAAGTGATTTTTTAAATTATGGAAAAGAGAA
TGTTATTGCCATTCAGGTTGAAAGAAGAGTTTGGGGAAAAGTAAAGGaTGAA
GAAATACCGCCTGGTAATTTTGATTGGTGGCCATGGGGCGGAATACATAGAA
GTGTCTACTTAGAAGCATGCAATCAAATAAaTATTAGTAAGATCTTA

>gene_36826 >isotig06679
ATCCAAATCTATTCTTCCTGCAGAACCCCAAAATCTTGTTACTATATTTTCTAA
CCCCTCAGGTTTTCTTTCAGAAAGCCATAATCTTTTATACATATTGGCAAGCT
CTATAAAATCAGAGGCATTTACCTTTCCTTTTATTCTATTtATTCCTACTATTAA
AGCATTTCTTATAAATTCCAAaTCTTCAGGAAGACTAATATGAgAAACCCTTTC
CCATATTTCTTCATATCTTCTTAAAAGATTCTCTCTATCAGAAAGATTATCTAA
TATTTTAAGATAGTAATCTGCCCAATAATCTCCTATAGGTTTtCTTTTtGGAAG
AATTTTTCCTATCACTTCTGTATAATATCTAAATATAAAGTCTTTAAAAAGCA
CGTGTTTTATTGTTTCTGAAAAGATTGTTTTCCCAAACtCCTTTCTTGCTAAAaG
TATCTCTTtATTTTCTCCTGATAAAGCTATCCATTTTTCTtCCCAGTTAaTAcTCC
CTTCTGCAATTTCCATGCCTGCAAGAATAAGAGAATTCAGGAAaGAAAAAaGA
CATTCCTCCCCATCATCTCCCCATGCGGTAATTAAAAATCCAgAAaGTTTCTCT
TCTTTTGCAGGATTTAAAAAGTTTTTCAAGTTTGCAAGAGCAATATCAAAATT
AGGATAAAACCTATTCCAGTTAGAAAGTCCAGGACATGCCAGTTCTTTTTtCTt
CCTTTcTCCAAACATATTTATTTTAttcTTAAAATATTCCTGAGATTGCCCACTaT

AATCCCAATTAgCTATCATTACCTCATCCCAAATCTTACTTTCTAAAaCTATTT
TCCATCTTTCTCTTTCTTCCTCTCTAaGATACATTCCTGTAAGCATATCTCCCcA
AACTATAGgTTCTTTtCCACTTtCTTTAACCATCCTTATCATATTTTtgTGATGCA
TTtCGTATaGAgAAGGACCCTCAAAaaCTCCACTTTTATCTAAGCTTcTtCCCcTgc
caagagcccAAGTTTCATCTCCTCCAATATGAATATATTTTGAAGGGAAAAATTCT
AAAACTTCAGAGAGAAGATCATAGGAAAAATTTCTTCCCAGTTCATCGGAAA
CATCCAAGCACCCTTCTTGGGGTCTATGCCATTCACTAAATCTTCTAAATTCA
GGAAgAATTAAAATATTTTCCATGTGCCcACAGAGTTCTAAGgATGgAAACAC
TTCTATCCCAAGCTTTTTCGCATATTCTaTTATCTCTTTAAGCTcTTCCCCCTCT
AATCTTCCTCTCAGTTTTCCAATCTGAGGaTATTTTCTCCAAGGGAaTAGATCT
TCAAAaTATATAGCAAaGTAaTtATaTTTAAGAAgAAAGAGCcATTTTAAAAGCT
CTTTAAAAGTCTTTAAATTTGGAaCTCCCCcTCTGGCAATATCCAAaTGATACC
CCcTAAAAGAAAAaTTAAATTCTTcTTCAaTAgTtACcTCAGGAATAATATTTTtA
TTTTCCTTtATTAATtGAATAATAgTAGCAtGACATATATTTTCATCTCCCCAAata
TCAATTATTCCAtCCTTGATTTGAATTCCTATCCCTtCCTTTtcTTTtttGTTTATtCT
CCATTTTCCCTTTGGAATATTAAACtCCTTTGAAAGAAAATCTGGGAAATTCTC
AAATCCGTCAAAGGGGAACCATTTACCTGTAAACTCCAATTTTTtAGGCTCTG
GAACTACATAAATTTt

>gene_36926 >isotig06337
GtAATTCAAGTATtCTCTAAGCTCTTTaTtAAAAGgTTGTGTCAAAaTCTCGTCTA
AAACAGCAAGCACCTGTtCATAAGACTTAGTCCTAGGATCTCTTACtAATATCT
CCTCAAGtACTTTtAtATCTTTAGaGATAAACCCTTCTAAAGCCCATTCCATTCTt
AAAATtCTtGGCATTAAGTAGAAGATTTTtATCCTATGAGTAAGATCAGGCTCT
ATCTTTTCAGGCTTTATtCCGGTTTCATCTACCCATACAGGCACTTCTACCACT
ACATCGTCAGGAATACCTGATAGGGTACcTCTATTAGGAATATTCAATATTAA
TCTTGATTTGCCTTTTCCAGTAATACTTTTTATAAAATGTATATGCTGTTCACC
ACTTAATTTATCTTTATTAAATAACTCAGGCCATATCTCCATTAACTTCATACT
TTTATCTTTCTCAACCTCTTTAGCAAGTTCAATCATTCTTCTCCTACCTTTTCTT
AACTCCTCATGAAATTTAGGCCTTtCTATCTCATTATCAATACCTCCAAAcTCC
CATACCATCTTTTTAGTCTCAAGATTGtAGTTGtACTTCCATGTTCCATTCCTGC
aGGTATCCCCTATGGGgAGCATTCCaTAAAATTtATACATATCCATAaCCGCAG
GACTCATCTGAAGATCCCATGGATTCTTAGgcTCCCAATTTTGAGACTCCTTTT
CAATCCATTCATCAaGAAGaGG

>gene_36936 >isotig06342
ATGTACGTTCCAAAAGACCGGTACGTTTGGGACTTCTGGCTGATTACACATGC
CGGGCAATATCACCTGTTTCACCTTCAGgCCCCCCGAGGGCTCCCCGATCCGG
AGATGCGCCACGGCCTGGCCACCGTCGGACATGCCGTGTCAAAAGATCTGAT
TCACTGGCAGGACCTGGGTACAGCTTTAGGCCCAGGCCGGCCCGGAGAATGG
GATGACCGGGCAATCTGGACCGGcAGCGTCATCGAAAGGGACGgCcTTTtCtAC
aTGCTTTAtACCGGgaCCTGCcACgCcGAGCGCGGGAagaTCCAGAGGATCGGCCT
GGCCGTCTCCCGggACcTTCTGCACTGGGAAAAaCaCCCcAaGAaCCCGgTGCTC
GAGGCGGATAGAGGGTtGTACGAGGGTgCAGAGGAGTCTctCTTTGGGGAATT
GGCCTGGCGCGATCCTTACATCATCAagCACGAAGGCGTGTATTtCGCCCTGAT
TACGGCCAGGAGGAACAGGGGAGAtCCAAaGAAgCGGGGGTGCATAGCCACC

GCCTGGTCGGATGACTTGGTGCACTGGCGGATCGGCCcTCCCCTTAAGGTTCC
AGGAGGGTTtGCGCAAATGGAGGTCCCGCAGGTTATctACCTtgAaGACCGGTTT
TATCTATTGTTCTCTGCGGAGGcGGgCTGGGTTGCGGACGAGGcGTGCCCAAA
GGTGACCGGGACTTTCTACGCCACGgCTCCTGAGCTCTTCGGCCCCTATTCCG
TCCCTAAAGTCCTCCTAGGcGACCACCAAGGTTCGTTTTATGCGGCGAAGgtGA
TTCAGACcaTTCAAGGctCTTGGGTAGCCTTGGCCTGGGTTCGCACCCGCgaTGG
GAGCTTCGTAGGGGGgCTGTCGGATCCAaTACCGGTCCATTTCAAGGACGAT

>gene_37140 >isotig06445
ACCCAAGAGGGAAGgAAACTCGTGGTGCAGAGCGGAGGAGAGGTCATATGG
ATCGAGCCCTTTGGGCGAGATTGTCTCcGCTTCCGTTCCTCCgTCCGGGGGCGc
ATTGAGCCCCTCGATTGGACACTCcTCCCCCAGGAGGAcGTCGTACCaGAAAT
CACCATAACTCAAGAGGGAGCCACCATTCGGAATGGgAAAaTTGTtGCCGAGG
TCGATGCACGGGGgAgtGTCCGGTATAGGAAGCATACTGGAGAAGTGCTCCTT
GAGGAGCTCTGGATCGATGGAAGaGTTCTCAATGctGATCTTCTGAAAGCGaGA
AACTACAAGGCGAAAAGTTCTGACCTATTCGAAATCTCCCTTTACTTCAAAGC
CTACCCTGGGGAACACTTCTTTGGcATGGGGCAGTAcGCCAaCGATTGCCTTGA
TCTCAAAGGGTGTGTCTTGGAGCTCGCCCAGCGGAACACCCAGATCACCATC
CCcTTCCTTCTCTCTtCTCGAGGGTATGGATTCATTTGGAACACaCCCTCCAtcGG
CCGAGCAGAGCTTGTGAAAAaCCATACTATGTGGTTTTCCGCTGGATCTCGgC
AAATAGACTACCTtATCATCGCCGGAGACaCCCCTAAAGaGATTCTCCGAcGCT
AcgCcGAAATCACTGGgAAgCcACCCATGATTCCTGAGTGGGCgCtGGGATTCtgg
caatgtaagcttcggtaccgaacacaggaggaagtt

>gene_37349 >isotig06555
tccagcatatggaagtggcgagattacaggccatccttCCTCaTTAAaGAACATTTGATGAACtCTAA
CcTCATGGtGTTcTCCccTCTTaGGAAATCGAGTATGAAAAaTTAAAAAaTATTTA
TTtAACTTTGTATCGTAaTAAGCAGAATTATGTCCTGGTGATACATATCCATAA
ATCTCAATAGAGCCATAACCATATCCCGAAATATTACTTAAACTTGTATCAAA
GATAAAATTCCCCATCAATTTAACCCCATAATTGGAAATAGACTCAACATCA
AAATATTTGTAAGGAGGACACTTACAATCAAGCATATTTTtACCTTCTATATC
ATAGTATGGTCCATCAGGATTTTTGGACCTAACAACTCTTATATTGTATCCCC
CATCAGAAGCAAATCCACCAAAACTTAAAAATAAATAATAATAGTCTGTATG
AGGGCTATAAAGAATAAAGGGACCTTCTATTGGGCTATGGTTTCCCCCTAAT
AATTtcTtCCCATAACCCTGATTAGGATAAGGCATACCAGTTTCTGGATCCATA
CGAAGTATAAAAaTGCCACCAGAGTAAGAACCATATACCATCCACAGATTGC
CATTTTTATCATAAAATACATGAGGATCAACAACGTTTGGGTGTTTAGTCGGA
TCATATATTGAGCCATCTTCACTTGATTGTCCCCAcATTCCAGATTTCAAAATT
ATCCCTTTATTTTtATATGGTCCTTCAACATTTTCTGAGACTGCTACTCCTAATG
CAGAaCGTGGAGAATCTCCTTTACAAGCGCAATAATACATGTAAAACTTACCA
TTCTTTAGCTGAATAACATGTGGAGCCCATAATGTTTTCGTTtGTGCCCATTCA
AATGTATCcTTTAATtCCTCAaGAACATTaGGTATTAACTtATtCCCATTATAAACt
CCAGCATCTATTAACTCCCAaTgCATTAAATCTTtACTTTT

>gene_37391 >isotig06577
TCTCCCCTAAcTTTTTAtGAAATTCCATATGGAATAtATTTtAAACcAACCCATG
AAAATACTTCTTGGgAAAAGGCTAAATTTGAATTTCCAGCTCATAGATTTATT
AATCTTTcTCAGGGAGATTTTGGAgTAAGTATTTTAAATAATGGAAAATATGG
ACATAATGTAAAaGAAAaTATTATAGGATTAACTTTGCTAAGATCACCAGTCT
TACCTGACTTTTATGCAGACTTAGGGAAaCATGAATTTACATATTCTATACTTC
CTCATGGTGAAGATTGGAAATTTATAACAaTAAAATCCGCAGAAGAATTAAA
TAGACCTCTTTTTGCGAAAGAAATAAATAATGGAGAAAcAATAGAAGAATCT
ATTATAAAATGGGATTCTCTAAATATCTTAATAGGAGCATTAAAAAaGGGAG
AAGAGGAAGAAAGTATAATTCTAAGATTTGCAGAATATTTTAGTTGGGAAAA
aGAAaTTGAATTTATTTtCAATTTTCcTATAAAAGAaGTCTATGAATGTAATTTgT
TGGAAGAAAATCTTGAAAAAaTAAAaTTAGAAGGAAACAAAaTAAGAATAAA
AaTAAAACcTTATGAA

>gene_37408 >isotig06587
AAAaTTTACAATCTCCCAAAAaGCCTTCTTAGGATTTTGGTTTATATCAAAGAT
AAGAGgATAATCTTTTCTTCCTCTTACAGGCCAAAAaTCAAGCCAAGTAACAT
TGTCTGCTACGCCCcAAAAGgTTACTCCACTGATGTAaTTCTTATTTCTCCTTAA
tACCTCAAAAaTCCTTTtGTAAATCTCCGCTTGTTTTtCTAAaCGATCTTCAGGAG
GATTATCAGCAAAaTTTGCATTCTCATATCTATCTTTATATATAGAAACATCAA
GCTCTGTGATCTCTACTTTAATCCCTAAGGAAGCAAATTTCTTTATAGATTCTT
CCAACATCTCAGGAGTAGGCCAACCAAGGAGCCAGTGTCCTTGTATTCCTAC
ACCATGAATTGgAATTCCTTTTtCTTTTAATCTTTTAATCAATTGATATGCCTTA
TCTCTTTTTACAGGTTCTTCCAAGTTATAATCGTTATAAAAAAGTAAGGCATtG
GGGTTTGCCTCATGAGCCCAAaTAAAAGCTTTTTCAATTATCTCCTCTCCTCCA
ATTTTATACCATGGAGCATCTCTCAAAAATTCCTTAGGATTGTCAGATATAGC
CTCATTTACTACATCCCAAACAGCTACTTTTCCCTTATAATATCCTACTACTTT
CATTATATGCTCCCTTAATCTCTCTAAAaGGACTTTCTTGCTAACAAAATTTCC
CTTATCATCTTTAAAAACCcATGAAGGTATTTGTTGATGCCAAACTAAGGTAT
GACCTCTTACTCTCATATGGTTTTtCAAAGCAAaTTCTACTATTTtATCCGCAGG
ATCAAATCTATAAACTGAaGgAAAAGGATGAATAATCtCCCACTTCATCTGATT
TTCTGGTgTTAAACTaTTAAAATGTTTCTTTAAAAGCtCAtcaaaaagcactataGtttctgga
CTTactGCTGCAccaaTagggaaataatctttataaacctcacgaagagc

>gene_37470 >isotig00604
TACTTTAACTTTTAGTTGAATACTTTCATCTTCAGGATAAAACTCTAAAACAA
CTATAACATCTTCTGGGAACATCTGAACTACATGTTTTTtAATtACTATCATTG
AATCAGCCCTTACcTCAAAATCATCCCAACCTCTTAAATATGGCGTCCATGAA
TCTCTAACAGGTCTAAAGTTGTTTGCAGTAACAACTTTTATGGCAGAAAGTTT
TGTACCATTAAATGCAATAGGGATTTTTAATTCAACTGGTAATCCTTTACCCA
CAAAAAGTCCTGTATCTAAAACTACAGGTTTTTtGTATTGAACTACCTCCAAA
TCGTAATCTGCACCTCTATCGAATTtAAaGGTTAAAACCGCAaTTGTCCCCCAC
TCATCAGGTTTCAGAATAGATTTAAtATAGCTTGCCCTAATTGTTACTTTATTT
CTATCCACTAAATAATCTTCTCCTCTTCTTAGCTCCTTACTTCCATTAAAGATT
CCCAAAAaGTTATTACCATTTAGGATTAACTCTACATTCTGGTCCTTTATCTGT
TCGTCACCTTTTATATATAATATCCCAGGATTAAGAAAAGAATTTGGTATCCC

CTTTGATGCATTCATTATTATTTTTATCTTCATATAATCCCGCCATTTTCTATTC
CATCTATCATAGTTGTCATATCCATTATCCCAATAGAATGTTGCCATTTTGtAc
TtCTTAGCAGTCCTAATTAGATAATCAAAATaTAGCCATTCGTAAGGTCTATTT
CCATTAAATACGCCATACTCTCCAATTATAAATGCATAATTCTCAAAACCCTc
AATAGCAGTTCTTATGTCTTTATCCATCTGCTCTATATCTCTCTGACTTCCCCA
AGTTTTtCTACCCCACCAATTAGCTACAAAATCCCAGGGAGAATAATAATGtA
CTCCAATGATAATATAAGGATCCTTTGGTGCTCTAAAGTATTTCATTTTATAA
GTATCAGTAGAAACAGGGGGAAGAACAACTAACCTTTTATCATTGAATCCAC
CTGAATTTCTTATCACCTTAAgAATTCTTTCATTTACCTCATTTTGAATCTCTCC
CTTTTCCTGgTCACTAAAGCCATCATAGTCAGGctCGTTTATAGTTTCAAAAAT
TAGTTTTTCAGAATAATCTTTAAATCTATCTGCAATCTGTtCCCATAATTTTTC
AAGCTTTCTCATTGTTTcTtCTTTTCTCTGCTTCATCTCCGTTTTTAGCCATCGCC
AAGAATCATGATGCACATTAATAATTACCCAAAGATCCCTTTCTAAGGCGTAT
CCTACAACCTCTTCAACCCTATCCATCCACTCCTTATCAACcTTATAATTTGGA
GCAGGTCCCATATGATCAACCCATGTTACAGgAaTTCTTACACTtCTAAATCCT
gTTCttttAaTATcgtcaaaaatatatggCTCtGTCTTAGGATTACCCCAcGAGgTCTCATCAG
GAaTTGCATCTAAAGAGTTACCcAAATTCCAACCAGGATCCAtCTCTTCtACAG
CCTtAtGCGGATCAATTGGTgTAAACCTGCcTTTGTAGCTCATACcATACAC

>gene_37548 >isotig06663
CTGGACTACTTCTAAGCTTACATCAtCAAAGTATACTTTTCCtgTTTGAGGATTt
GCTGTTTTTCCAAGcTCGAAGGATAGaTTAACTACTGGATCAGTACcATTGTAA
GtAAATTGTAGTTCAAAAGTTTGCCAATCACTTGTTAATGAATAGGTACCGTC
TTTATAGAGAGCATATGTAGTTGgATTTAGGAATTTGACaTTTATATCTCTTGG
gTTATCTGCCTTtGCCTTGAAGGTTAGTTTATAGGTCTTTCCTGAAGAAAGTCC
AACCCATTGATTGAATTGTACATGCcAAGACTCCCATCCATAATTTTGTACAT
CCACcACTGCATATCCATTTTCTATACTTGCTGTAGCTGTTCCACCCATtCCGTA
GGTAGAACCATACCATAAGAACCACTCATCAGGaTTATTAGcTTGATCATTTA
CtAATGgATAATCAAaGGTACCGTTGTTTAtAGATTCCATGGTTACGCCTTTATA
AACTTTTACATATTTTACGTACATCTTTTGAGGAAATACTGTAGTATCATCTG
GATAGCCTGGCCAGTATCCTCCAACAGCCACATTCATTATAATAAAGAATGG
ATCATCAAAGACCCACTGACCATAGAATTCTATTTCTGGTCTTGTTATTTGGA
AGAACTTATTATCATCTACATACCAAGTAAGACCAATAGGATCCCACTCAAC
TGCAAATATATGGTAGTCTTGAGAGAAATCTCCATTGCTTAATATATAGTATG
AGCCTTTtCCATTAGCGCCAGAaTAaCCAGGTCCGTGAACTGTACCATATACCT
TATTGGGTTCATGACCAaGTAGCTCCATTACATCAATCTCTCCACATGCACAT
GCTGGCCAGGAAACTTGTGTTAtcttttCTCCTAACATCCATAAGGCAGGCCATAT
TCCCcTtCCTTTTGGAAGTTTTGCTTTGAATTCTATTCTTCCATAGGTAAAGTTA
AACTTGGTTTGAGTCTTCATTCTTGCAGAGGTATAATGGTAAGTCTCGggaGTT
CCATCGCCATTATCATCCGCTGTTACATCTTCCTTTCTTGCCTCTATCACGAGC
GCCTTTTCCGTAGGATCTATATATGCATTCTCTTTCTTATAATACTCccaTTCTC
CGTTACCCcAACCAGGAAGATTTGGACCaCCATTACcAATTTCAAAGCTCCATA
CATTTTCATCTATATTATCCATATTTTCAAAATCTTGTTGCCacgCtaCTTTATAC
CAAtATTTTGCACCCATTGCTGGTGGAGGAGCTGGcGGTtCTGCAGTACCAaCT
CTAATCAGTTTTACATTATCTATATAAATTGTACCATCATCTTTCCcTAATTGA
AATTCAAATCTTGCCTGCTCATCGgTTGGATTTGTCATAGTAAATTGGAAAGT

GTAAGTTTtCTTTTCGGTAGTAATACTAATTAAATATCCTCCAGATTGATCAGT
CTGACCTGAGGGACCAGGATTATATGCAGTCCATCCCTTATAACCATTGGCAC
CAACTTTTACTCCAATATTTCTATTTTTGCTTGCCcATGCATCAAaGCTTACCTT
AtAAaTTCCTAAGTATTCTACTTtAaTTGgTGATTGCAAAAGTTGTAcAGaCCAaG
AATTTGgACCTCcATTTgTtATATCTACTTTAAaTACaCCATTTTCtAcAACACCAT
TTgcAACTGCTCCATCTCCAgTTCtAAaTAgCCA

>gene_37856> isotig06836
ACCTGTACTCTCGTCGTAGAGaGTATCCAGGATACCGGCTTTGACTTCTTGCG
CTGCGGCGCGGAACAGCTCCTGCTCTTCTtCGTCGCCGAAGAGGTGGGCGAAG
CGCGCGGCCGCACGTAAAGCGGCATACACCGAGGCGCAGGTGTAGGTGTGG
ATGCCTCTCCGCTCCTCCCAAAGGTCATAGCTGGGCAAAGGCAGACGCGTCT
GCGGGTCgCGATAGGTGACcaTAAAGCGTGCTGCTGGCAGAATGAGGCTCTCG
TAGAGCGAGGcGATGAACTCCAGGTCATGGTAACGCTGATAGTGATACCACA
GTGCCCAGAGCACGGAAGCGGTGCTGTCCTCTTgAAAaGGGATTTCGtGCTGG
TCGCCGATAACCCAGGGATGCCACGaGgAGCCGACtgAaCCATCTGCGGAaTaTT
TGtGCATCAGCACCGGGCgATCCTTGGGCAGCACGCGGGCGCAGAaCTCGAAG
AAGCGACGCGTGGTGTCcGGATAGCCGATGGTGTcAAGcGCATAGGCGACGAT
GGCtCCATCgCGGGgCCACATGTAGCTGTAGTGCGCTCGGGCAGTGGTCATGA
TATCGGTGTCATTCGCCGCGATCACCGCTCCTCGGTCATCTATCTGtGTGCGCA
TGATCAGCAGGCTGcGCCGGAaCAGCTTTGCTACCGGTtCGGGCAGAGTCTGC
ACCAGTTCAGCGCCCTTCTCCGACCACCGTGCCCAGTAGTTTGCCGTGTCGCG
CATGATCTCGTCAAACCcTCTTTCCACCACGCGGTTGTGGAGCGCACGCACCT
GCTGGTAATCATGCCCGGCAGCAATCCAGTAACGGAGATACTCCTCCTCGCC
CGGAGACAGTAGCAGGCGGAAACCGATGGAGCTGTCTACCGAACC

>gene_38193 >isotig07037
AATGCcTTAAGAaCTACTCCAGTAGCTTTAAATATTGCTAAAAAAaTaGAAAaG
CTTTCtCCTAATGCcTGGTTAATAAATTTTGCAAATCCATCTGGAATTATAACG
GAGGCTATATCAAAGTATTCCAAGGTTAAAgTTTTAGGACTATGTAATGTACC
TATAAACTTTCAAAATTATTTTGCAAAAaTTTTAAATGTTCCTTTAGAAGAAGT
ATTTTTAGATTACTTTGGATTAAATCATTTGACCTTCGTAAAAAAAaTATTTGT
TAATGGAAAaGACAAGACAGAAGAAGCTTTTAATAAAaTAAAAGAaTTTCTcT
CTGAAGAAGAGATAAAAGTaTTAGATTATTTAAACATGTTCCCAAATTACTAT
TTAAGATATTTTTATTTTAGGGAagAAAAAGTAGAGGAATTAAAAAaTAAaCCC
AAAaGAGCTGAAgAGGTAATAAaGATTGAGGAGAACCTACtAAAACTTTATCA
AGATCCCAATTTAGATGAAAAGCCAAaGGAATTaGAAAAAAgAGgAGgtGCTTt
ATATtCTAAATCTGCaGTTAATCTAAtATTtCaTCTCTTaGGGTTAAAAAaGGGAT
TCcAAATTGTAAatATAaCAAATCATGGAGCAATTTTTGATCTTCCTTATAACtCt
GTAGTGGAAATTCCTGTATATaTAGAAGGAGAAAGAATTcAAAGATATATAaT
TGGAAATTTaCCcTTGAAAGTAAGGGGACTTATTCAGGCAGTAAAAaGCTATG
AAGAaTtaactatagaggcagcagtaGaaggctcctataaaaaagccatttgggcattggctcagcatccatta

>gene_38194 >isotig07038
AAAAAGCAAGTTATTAaCACTCTCTATGCcAGATATGTCCgcAAAAAaCcAtCCG
TcGGAAGAAAAGGAGAGGTGAGCGTaTTTTATAGCGTTCAGGAGTTTTAGCAC
TCTTAccttttCcTCcttGCTtAcaTCTCTCTTTAAGTGCCTTTCCAAGTATTGGTCAAT
GGGCAGGTCcATTAAGGTATCCACAAAGTCCCAAAGGGCTTCTTCCACACTGT
AAAATACTTATCAGCTCGTGTAAGaGACTTCCTTTACCCTTGACCTTACCGCTT
CAAGTCCTTCTCTTAggggcTTTCTCCACTTTTGATGCCAGCCGGGAAaTCctCCC
gTAGaGCATCCGCAATCGCTTCTCCACCTTTCcacACCGTGAGGACAGCTCCAA
GATGTGTTTtCGTTtATGTctGTGGAAgTTCTTGGAGTGTTctATCGTAATACtcTtC
AAGGCTAaTAAAaTGTGgAGTTTTtAAAAaGGTAAGcAAGcCCcAtCTCaCCAAAC
TTTTTGTGGTGTCCAAAGGTCTCTCCGTCGGTAGCTATAAGagTAaGtCCTtCCC
TTGCATTTAGtctatCAAGCAAAAGCCTTGCATCCaCAAGCAGGTCTCCAAAGGC
TAcACCGTGAGAcAGCTCACCGTCATAAaCAAATATGTCTATGTGTCCCTCTTT
TGAGAAGTATCTCAAAAAGTTttcTTtGGTTTTtACCTGATGAGgAGCAAGGATT
ATATACTTTATTCCCTGTCTTATAAGTATGTCTATGGTCTCTTTATCCACCGCA
AGCTCAGGAAGCCAAAAACCTCTTGGTTTTCTtCCAAAAAA

>gene_38497 >isotig07217
TTCTTCTAAGCCATTTtCTTCAATTACTTtCTTGTAAAAATATGCACTGTCCTTT
AATATTCTTTTTtGAGTATCATAATCTACATATACTATTCCGAATCTTTTACTAT
ATCcAAATGCCCATTCAAAATTAtCTATTAATGTCCATATATAATATCCTTTTA
AGTCAACTCCATCTTTTATTGCTCTATAAGCTTGTTTTATATGTTCTTTTAAaTA
GTTTATtCTTTTTTCaTCCCTTATTCTTCCATCCTctACTATATCgTtAAAAGCAGc
GCCATTTtCTGTTATGTAGATGGTTTTAATaTACTCTTtAGCAAGTCTTactAAAaG
AtCATaTAATCcTtgAGGATATACTTCCcAaTCcaTTTCtgttttttcattgggaccaggaattctttttac
ccctaagaaaggatctgt

>gene_38682 >isotig07331
TTAAgCCGGGGgcTTCCTgaGCACGAGGtAGACGAAgaggggACtGTTTACTTGACC
cTTTtGcGcTgCgTAGGGTGGCTTtCGCGGGATGATCTTTCCACTAGACGCGGGC
ACgCcGGGCcGccttACCCcACTCCGgAaGCACAGTGccTtGGTCGgCATCGGTTTGA
GTACGCAATATTTACCTATCGCGGGTCGTGGATAAACACAAATTTATTGCGA
GTAGCACAGGAATTCTctCTTCCTCCACGAGGTTTTTtCGCCGAAAGCCACTGG
AAAGGGCCGCATTTTCTGCTAGCTTTGGATCCGCCTACGCTTATGTTGAGCGC
TTTtAAACCcGCGAATGAAGGCgCGGGCATAaTCgTGCGgATTTATAATCCCAC
ATCGGAAACGTTGGTCGGAAACAtcGTTTtCGGCTGGCCGGTAAAtGAGGTGTG
GGAAGCcaCCctAGATGAAAGTTTTATAAGAAAGCTCTCA

>gene_39002 >isotig07530
ACCAATTGCATTTGCTTTAAACCATGCACCCCATTGCGTTGGATTATTAGGAT
CCCAGTTATTTACCATATCTAAATATCCCACATAGCTTGAGTTGCCACACCAA
GACCAGCCAAACAAACCAATCTTATACTGTTTAGCATACTTTACTATTGAAGC
TtCATCAGGATCTCCATCAGTATGATTATATCCAAATTCGCCTATAACCAATG
GTAATCCCCTATCAACAAATGACTTTATATAACTTTCCACCTTATCCGCAGTA
TTGTATACACCGTACATATGAATCGAAAaTACCAGATTACGCAGTGGATCAGA
ATTGTATAtcGCTGTCGCATTGTCTCTCATAGTATAGCTCCAATCCTGTCCCCA

GTTaGGCGCATCAACCATTATTGTgTGctcAAAaCCCGCATTTCTtAATGCAATtA
TTGCATTTTtAGtATCATCtaCCCATTTCTGGTAATtATTATTGCcATATGGCTCAT
TtCCTATATTGATTATTaCAAAGTCTTCCTTACcATCAAGTaCACTTTTGATTtCC
TTCCagTActCtACTGCCTGTGCCAaGgAaCaTGCTGcTCCATC

>gene_39774 >isotig08057
AaCCACTACTGTATTAGGATTTACCTTTGCTACTTCTTCAATTAGTCTGTtATGA
TTTtCTGGCATCTTCATATGAGGTCTATCATATCCTTCAGACTCATATCTTTCTG
GAAGTCCAGCAAATATTACTGCTACTTTTGCCTTCTTTGCAATCTCTTTTGCCG
ATTCTATCAGTTTTTCATCTATATCATCAGTATCCAATTTATATCCATCTGCAT
ATAAAATTTCAGCTTTTCCTtCTGCTATCTTTACTATCTCATCATAGGCACTAT
CGAGCATGGTAGGATTAACATGAGAACTTCCTCCCCcTTGATATCTtGGGTTCT
TAGCAAAAGCTCCAATTAATGCAATACTCCCTTCcTTCTTTAATGGCAAAATA
TTATCTTCGTTTTtAAGAAGGACAAAaCaTtCCCTTGCAACTTCTGCAAGATTCA
AGTCATTAGCTGGAGCTATAAaGGAaGCATT

>gene_40132 >isotig08327
GATCATTGGAACAGGTATGAAGAAGATTTTGAGATTATTAAAGAAATGAACA
ATAATGCCTATCGCTTTTCTATTGAATGGAGTAGGATTGAGCCAAGTCCAGGA
aTATTTgATGGTtCTGCTATaGAACATTaTAAAAAGATGATTGTTtCTTtGAAAGa
GAGAGGAATAGAaCCTTTTGTAAcTTTACATCATTTtACtAATCCCATTtGGGTA
TCAAGAGAAGgTGGATGGTTAAATCCTAAAACCATTGATCATTTCGCAAAGT
ATGTGGAAAaGGTTGTAAGCTCATATAAAGATTTGGTGAAGTATTGGATAACC
ATTAATGAACCCAATGCCTATGCCCTTGTAAGCTATCTTATGGGCATGTTTCC
ACCTCAAGAGAAAAGTCTTTTtAAGATGATTACTGTCCTTAATAATATGGTAA
AGGCTCATGCACAGGGATATAAGATTATTCATAATTTAATtCCTGATGCCAAA
GTAAGTATTGCGTATAATGTTATGCTCTTTGATCCACAAGATCCTTCTTCCTTC
TTAGATAAAAaa

>gene_40145 >isotig08339
ACcGACCCgCGCgTCcGGGCCTGGTGGGGCGACCTGCACCGGgACcTCCTGGAG
GCCGGGGTGGACGGCATCTGGGACGACATGAACGAGCCCGCGCTCTtCGGGG
AGGACGGCGCGACGATCCCCGGCCCcGTCCGCCACGACCTGGAGGGCgccgggg
gCGACCACCGCGCGGCCCACAACCTCTACGGGCTGAACATGGCCCGCGCCAC
GGCGGAGGGACTGGCCCGCCTGGCCCCCGACCGCCGTCCCTTCGTCGTCACC
CGCTCCGGCTGGACGGGGgTTCAGCGCTACGCCGCgCACTGGACCGCCGACA
ACCGCTCCGACTGGGCCTCCCTCCGGCAGACGCTGCCCATGGTGCTCAACCT
GGGGCTCTCGGGCATCGCCTTCACCGGCTCCGACATCGGCGgcTTtgAGGGgTT
CgCCaCGGgCGAGCTCTTCACCcGCTGGCTccAGATgAGCgTCTTCTTCCCCTtcTG
TCGGGCCcaTACCTACTtCGCCAgCCCcGATCAGGAGCCCTGGTCgTGGGGGgAA
CCgtAC

>gene_40284 >isotig08439
AGGGGAATATTGAGCTTTTtCTTCATAAATGTCTTCGCCGATTATATCAaCATA
GTTGTCCCCcGGATACCATGCTGCATCCTgACCGTTCCATACCCAGATGAGGTT
GTTCAACTTGTGATAATTTACAAGCCTATCAAACATAAGTTTCCAGAGTTTGA

TATAGGGCTCGGGTCCTTTtGCCCCCcACCAGAACCAGCCACCTGAAGCTTCG
TGCAGCGGTCTGAAAAGAaCTGGGACTCCTTCTTTCTGAAGCCTCTTGAGCTG
CTCAGCAATTGCGTCTATATCTCTCAAAATGAGCTTATACTCAGTTGAATTAG
GATTGTCCATAGCCTTcTTTATATCAAATGTTGTTGCTTCTGTGTAAAATCCTC
TCCACCATTCTTtGCCCGgcTGGTCAATAaGCCCTGTTGGTGCATTCCAATGCCA
GCAGAAAGCAACAATTCCACCGCTCTTCcACCATTTTATTGCCtCATCTACATc
AGT

>gene_40458 >isotig08572
ATTTTTCAATGCCTCATTTACCACATGATCTACCCTATCAAAAAaTGCCTTTTC
AATAGTATATGGAGGATTCATATCTGCGTGAGCATTCCACCTTATAGGGATTC
TAACATGGGAAAATCCTGCCTCTTTTATAATTTTGAAATACTCATCCTTTATTA
CTACTCCCCATTCTCCCTCTCTTGGAGCTTCAAGAGCATTCCCCATATTTATAC
CTCGAAGAATGGGGAGCATCTCCAGTTTAAACTCATAGGAaGAaCAACCTGC

>gene_563 >contig00955
TTCAGAAGCAGGATGAAcaGGGGATAAGTCCAAAACGATGCCAATTTGAGCG
TACGGTTTTGTTAAGTCTCGAAGAACATCCACTGCAAGACCGTGAGCAAGAA
GAAGGTGATGAGCCACTTGTAATGCCACTTTTGGGTTTCGCAGACCGGGAGC
ATGCTCACCCCAGCCATGGGCTACCCAGGCCACAaCCCAAGGCTCATTGAGA
GTAGTCCAATAGATAACGCGGTCTCCCAAGGCCCGAGCTATAACAGCAACGT
AATCAGAGAAGTAGTAGGCTACATCACGATTAGCCCAGCCCCCGAGATTTTG
AAGCGCCTGAGGCAAGTCCCAATGATATAGAGTTACGAAGGGAGCGATGCTT
GCTTCTAAGAGGGCATCGACAAGGCGGCTGTAAAAATCCAGTCCTTTCTGAT
TCAATGCCCC

>gene_566 >isotig00957
GAACTGTGGAAGAAAATGATGATGAGCcGTTTTtATTTCCCGCCATAACCCTT
GGGCcACAGAAAAGGAaGgCACAAGAGGgTGAGCTAAAAGGGCTTGAAGGGC
GGTGCGTtCGTtGCCGGTTATtGCTGCCTCTATAGTAAGCTCTTCAAAGGCTTTA
ACAGCTTGGACCAAACCACGCACCTGCCGAGGTAAGCACGCAACTGGAAGA
GGATaGGCCCCATCTTGGTTTATACGGGCTGGTACTTCCACGACTGCATTATG
GGGAAGgTCAGGAACTGCTGATTTGTTTTGCACGTTTACAATGTGAATTTCGC
TTTTGTTGTTATAGATAGCACTGATTAAAGCcACAGCAGCCGTGGAATAGTGG
GCACCGCCGCGTTTTTCTAGAAGTGCGGGCTTTTGGgTGAGCGAGGGGTCCGC
ATAAAGCGtGAAAAGTTCCTTATCGATTGCCTGAACTTCTTCGCCTCGtGTTTT
CAAAGCGTGCTTTTGCTCCGCAAGGACTTCATCATGATGGTAATAGTAGCGG
AGGTAGTAACTTGGGATCAGaCCGAGCGTTTCTAAAAGGGCTGGCGAAAAaG
GAaGGACCCCCcGGtGGGCTTGTTCCAGGGCTACCGCGAAGGCGCGGTCATAT
ACATCTTCTCCCTCAATTTTTACGCCGCGAACCCAGCTCAGATGGTTTAGCCC
AATGTAGTCGAGCTCCACGCTTTGAGAATCAACCCCGAGCCAATTAGCAATC
ATGCGTGTCATCCCCAATGGAATATTGCAGAGCCCGACAGCACGAACGCGAG
TGTGAGAGAGGATGGCTTCAGTAATAATGCCGGAAGGATTGGTGAAGTTGAT
AAGGAAGGCGTTTGGACAAAGTTCTTCCATTTTTTTAGCGATATCAAGCATGA
CAGGGATTGTGCGTAAGGCCTTGGCAAAGCCTCCGGGCCCTATAGTTTCTTGA
CCGATCACATTGAATCGGAGAGGGATTTTCTCGTCGAGGATGCGGGCAGCCA

TGCCTCCAACGCGTATTTGGGAGATAACAAAAGTCGCCCCCATCAaGgCTTCT
TCCAAACTCATTGTTAGCTTAATGCGAATAGGAGCTTTTGCTGCTTTAATCAT
GCGTTCTGCAAACTTCCCAACTATTTGCAATCTTTCGTAGTTGATATCCATGA
GGCAAATTTCTTCCACTGGGAGCTCCTTGTAACGCAAAAGGAAGCCCTCGAT
AAGCTCAGGGgTATAGACGCTTCCgCCACCAATAACCGtGATcTTCAA

>gene_567 >isotig00957
GTTCTCTTTTATAAACTCGAAAAGTTCATCAAAAAGGCAAAAAGCTAGGCAA
GTCACTGTATCCGCcGCTCCATAGTAAATTGCAATACGCCCTGTTGGTGCGTC
ATATAGGGCAGCGGTAGGGAAAaCAACATTTGGCACATCGCCAACACATTCA
TAATATTCTCGCGGGctcAATAGATAAGGACGTGTGCGGTAAATTACTTTCCAT
GGTTCaTCCAAGTCTAAAATAGCaGCaCCAAAGCTATAACAACCGTTACATGA
TGTAAGAACACCGTGATAGAAAAGGAGCCATCCTTCTGTAGTTTCAATAGGC
ACAGGACCAGCACCAATTTTTGTAGACTGCCAGGAATTGGGGATAGGGGACA
TCACATGCCGGTGTTTGCCCCAGTGCACTAAGTCTGGGCTTTCAGAATAAAAA
aTGTCTCCAAAAGGTGTATGGCCTGGATCACTAGGGCGGCTTAACATtGCATA
CTTACCACGAATTTTaCGGGGAAAAAGCACGCCATTGCGGTTATATGGGAGA
AAGGCGTTCTCCAACTGGTAAAAATCGCGGAAATTATAAGTGTAAGCTAAAC
CTATAGTAGGGCCATAaTACCAGTTaCACCAAGTTACATAATACCGATCCTCA
AGCCAGGTCACCCTGGGATCGTAGCCATATACAAATTTTCCCACTTCAGGGTC
ATCACATATAAACTCGATTCGTTGAGGGTCGATTTCCCAATGGATCCCGTCcT
GGCTTCGGCCAGCGTGTAACTCCATGCTGCgTGTTGTTGTATCCACGCGGAAC
ACTCCGGCAAAGCCTTCTTGAAAGGGGACTATtGCACTATTAAAAATGCTATT
TGCcTTTGGGACAGCAATTCTAGTGATTATAGGATTACGTTTtGATCGCCATAC
AACATCGTAACAACCtCcTGGCCGATCTTCCCATGGGATATTAGGGAGCGCtGG
TCCCACTAGCCTTGCCTGCGGCAT

>gene_new >isotig01115
GACGGTAAATTTCACTCTTACAACAAACTTTTCATTAGGGCTTAGATAAACAT
TATACAAATGAATAAGAGAGCTTCCTTGGTATACCCTTTCAAAACCTTCTTCT
GAAAGAGAAACCGTCTCcActGGATATCTATACCAAAGAGTCTTTCTATCTAAA
CtACAGTCCAAATTTAATATGGGAGTCTTTATTTGCCAAAAATCTATTTCCTTT
TCTCCTACCGTCCTCAACAAAtACTTTTCTCCAtCCACTTGGaAATaATAGTCCT
CATGTTCTCCTGCTAAAAAGTTTAATACTATCTCCCATGCAAATACAAATTCT
CTACTCTGAGAAGCATTATTTTCTAAAGAATAAAATACTTCAAAATCTTTTTT
ATTGTTATCAACTATAAAGCTCTTAACTaTTTTtATTCCATCCTTTTCTCCACTA
AAATTTAAAACCGCATAATCTTTATTTTTATCAATATTCCATGTAAAaGGTAC
ATCCAAGGATGTAAATAAAGGCTTCTCcTTaCcATGCCAAAAAGAAATAATAG
AAGGTACAGTGTCAAAAAaGTATTCTTTAAAGGAAATTCTCCTATGACTATCG
TAAAaGAGTAGATTTTCTAAACCTTCCTCTTTCACTGCCCATCTCTCATGAATC
GTCTTTCCTTTCCCCTCTTCATCCTTTATATGGAATAACTTAGCATGATATGTC
TCTTTCCTTCTTGTAAGTGTATCGGTCAAATTAAAAGAGGCAGGTTTGTAATC
CCACTCTAAAACCCCTCCCCCTAAATTTGGAGCAAAaTACAAATTAAAATACT
TTGATTctAAAATTATCTCTTCCCTTCCATCACAATCAAAATCAATATTTTTAAT
AGAAATCTCATCAGAATCTATAAAACTTTCTGCTTTAATAATATGGGAgTAAA

TGGCAGATCTTAGATGAGGAAgATAAAGCCCACcAAAGaTTCCATGCCAATAa
GCATCATTAGCTTGAGCCATCAAGATTTCTTCTGTTGCCAATTCTCTTTTCtCA
TTATCTTCTAtATTTTTAATCTTATTTCTTACATACAACATCCTTTTATGCATGT
GATTTGATTCATCATACTTTGACaAAAAGTTCCTGAAAAATCCTCCCCTGATA
AAGGGAGAATACTTaTCCCACTTGCCTTGCTCTTTAACTTCATcGttcAGTtCTTC
CAATACTTTTTGAGCTTCTGGGAATAAAACCCATTCCATCATCTCTCTGTAAG
AAGCGGTAGGAAGATAAATCCTtCcAAGAGGAGGAAAATTATTtATATAATCT
CCATAATGAACAAGACCTATTTCTTCTGAATTTTCCTTCAAAGCCTGCCAAAA
aTTATCCAGCCATCGATCTTCATAGACTGtTTTATATGTATCGGGCCAAAGCCC
AAATTTCTCTCCATCATCAAAATAAATAGCCAATTTATCTTCACTTGTAGAGG
TAGCGTTTTTTAAATAATTTATAATCTCATCGGGATGGTGAAAGGGAaTAAGA
TATCTGAGCTTCATATTAATgGGGAAAACTTTTAATGTGTATCCTTGCTCCTCC
ATTATATAATAACCAAACAAATTCTCAGGGAACAAACCTGTGGATAGAAAAT
GAGCATCATCTAATACTATATATTCTATTtCTGCCATAGACAAATATTTAACTA
AATGAGGCTCCCACACCcTTTCCGCCATCCAAAGACCTCTCGGTTTTTGACCA
AACTTATTCCAAATATACTTATTCATTTTATTTATTTGGAGAATCTTATCTTCA
TCGGGAATAATTGACAAAATTGGCTCATAAAAACCTCCAGATAAGaATTCGA
TCTGTCCCCTATGTGCaAGaGtTTTTAAAGAGGAAATATATTCTGGATGGTTCT
TTTCAAGCCAAaGGAACAAATAACCTGAAAAATGCATACTGAAAATAAAATC
GGGATTGTTAATAAAAAAGTCAATTAATGGTTTATAACATAAGTTATAAGCT
CGCTCCATTACAAAATCAAAATTTCCTACAGGTTGATGATTATGTAGTCCAAT
AGCAAAATAAACCTTTTTACTCAT

Literature Cited

1. Adav, S. S., Ravindran, A., Cheow, E. S. H., & Sze, S. K. (2012). Quantitative proteomic analysis of secretome of microbial consortium during saw dust utilization. *Journal of proteomics*, *75*(18), 5590–5603. doi:10.1016/j.jprot.2012.08.011

2. Aksenova, H. Y. U., Rainey, F. A., Janssen, P. H., Zavarzin, G. A., & Morgan, H. W. (1992). *Spirochaeta thermophila* sp. nov. , an Obligately Anaerobic, Polysaccharolytic, Extremely Thermophilic Bacterium. *Internation Journal of Systematic Bacteriology*, *42*(1), 175–177.

3. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol. 215*(3):403-10.

4. Alvira, P., Tomás-Pejó, E., Ballesteros, M., & Negro, M. J. (2010). Pretreatment technologies for an efficient bioethanol production process based on enzymatic hydrolysis: A review. *Bioresource technology*, *101*(13), 4851–61. doi:10.1016/j.biortech.2009.11.093

5. Anbar, M., Bayer E. A. (2012) Approaches for improving thermostability characteristics in cellulases. *Methods Enzymol.* 510:261-71.

6. Ando, S., Ishida, H., Kosugi, Y., & Ishikawa, K. (2002). Hyperthermostable Endoglucanase from *Pyrococcus horikoshii*. *Applied and environmental microbiology*, *68*(1), 430–433. doi:10.1128/AEM.68.1.430

7. Angelov, A., Loderer, C., Pompei, S., & Liebl, W. (2011a). Novel family of carbohydrate-binding modules revealed by the genome sequence of *Spirochaeta thermophila* DSM 6192. *Applied and environmental microbiology*, *77*(15), 5483–9. doi:10.1128/AEM.00523-11

8. Arai, T., Kosugi, A., Chan, H., Koukiekolo, R., Yukawa, H., Inui, M., & Doi, R. H. (2006). Properties of cellulosomal family 9 cellulases from *Clostridium cellulovorans*. *Applied microbiology and biotechnology*, *71*(5), 654–60. doi:10.1007/s00253-005-0249-6

9. Aylward F. O., Burnum K. E., Scott J. J., Suen G., Tringe S. G., Adams S. M., Barry K. W., Nicora C. D., Piehowski P. D., Purvine S. O., Starrett G. J., Goodwin L. A., Smith R. D., Lipton M. S., Currie C. R. (2012). Metagenomic and metaproteomic insights into bacterial communities in leaf-cutter ant fungus gardens. *ISME J.* Mar 1. doi: 10.1038/ismej.2012.10.

10. Baldrian P., Kolařík M., Stursová M., Kopecký J., Valášková V., Větrovský T., Zifčáková L., Snajdr J., Rídl J., Vlček C., Vořišková J. (2012). Active and total microbial communities in forest soil are largely different and highly stratified during decomposition. *ISME J.* Feb;6(2):248-58. doi: 10.1038/ismej.2011.95. Epub 2011 Jul 21.

11. Balk, M., Weijma, J., & Stams, A. J. M. (2002). *Thermotoga lettingae* sp. nov., a novel thermophilic, methanol-degrading bacterium isolated from a thermophilic anaerobic reactor. *International journal of systematic and evolutionary microbiology*, *52*(Pt 4), 1361–8.

12. Barabote, R. D., Xie, G., Leu, D. H., Normand, P., Necsulea, A., Daubin, V., Médigue, C., et al. (2009). Complete genome of the cellulolytic thermophile *Acidothermus cellulolyticus* 11B provides insights into its ecophysiological and evolutionary adaptations. *Genome research*, *19*(6), 1033–43. doi:10.1101/gr.084848.108

13. Bauer, M. W., Driskill, L. E., Callen, W., Snead, A., Mathur, E. J., Kelly, R. M., & Snead, M. A. (1999). An Endoglucanase, EglA, from the Hyperthermophilic Archaeon *Pyrococcus furiosus* Hydrolyzes β-1,4 Bonds in Mixed-Linkage (1→3),(1→4)-β-d-Glucans and Cellulose. *Journal of Bacteriology*, *181*(1), 284–290.

14. Bayer, E. A., Shoham, Y., & Lamed, R. (2006). *The Prokaryotes*. (M. Dworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer, & E. Stackebrandt, Eds.) (pp. 578–617). Springer New York. doi:10.1007/0-387-30742-7

15. Beattie, A. J., Hay, M., Magnusson, B., de Nys, R., Smeathers, J., & Vincent, J. F. V. (2011). Ecology and bioprospecting. *Austral ecology*, *36*(3), 341–356. doi:10.1111/j.1442-9993.2010.02170.x

16. Beg, Q. K., Kapoor, M., Mahajan, L., Hoondal, G. S. (2001) Microbial xylanases and their industrial applications: a review. *Appl. Microbiol. Biotechnol*. 56:326-38.

17. Bhat, K. M., & Maheshwari, R. (1987). *Sporotrichum thermophile* Growth, Cellulose Degradation, and Cellulase Activity. *Applied and environmental microbiology*, *53*(9), 2175–82.

18. Bhat, M. K. (2000). Cellulases and related enzymes in biotechnology. *Biotechnol Adv*, 18, 355–383.

19. Bhattarai, K., Stalick, W. M., McKay, S., Geme, G., & Bhattarai, N. (2011a). Biofuel: an alternative to fossil fuel for alleviating world energy and economic crises. *Journal of environmental science and health. Part A, Toxic/hazardous substances & environmental engineering*, *46*(12), 1424–42. doi:10.1080/10934529.2011.607042

20. Blumer-Schuette, S. E., Kataeva, I., Westpheling, J., Adams, M. W., & Kelly, R. M. (2008). Extremely thermophilic microorganisms for biomass conversion: status and prospects. *Current opinion in biotechnology*, *19*(3), 210–7. doi:10.1016/j.copbio.2008.04.007

21. Bredholt, S., Sonne-hansen, J., Nie, P., Mathrani, M., & Ahring, B. K. (1999). *Caldicellulosiruptor kristjanssonii* sp. nov., a cel lulolytic, extremely

thermophilic, anaerobic bacterium. *Iternational Journal of Systematic Bacteriology*, (1 999), 991–996.

22. Brumm, P., Hermanson, S., Hochstein, B., Boyum, J., Hermersmann, N., Gowda, K., & Mead, D. (2011). Mining *Dictyoglomus turgidum* for enzymatically active carbohydrases. *Applied biochemistry and biotechnology*, *163*(2), 205–14. doi:10.1007/s12010-010-9029-6

23. Canfield, D. E., Rosing, M. T., & Bjerrum, C. (2006). Early anaerobic metabolisms. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *361*(1474), 1819–34; discussion 1835–6. doi:10.1098/rstb.2006.1906

24. Cannio, R., Di Prizito, N., Rossi, M., & Morana, A. (2004). A xylan-degrading strain of *Sulfolobus solfataricus*: isolation and characterization of the xylanase activity. *Extremophiles : life under extreme conditions*, *8*(2), 117–24. doi:10.1007/s00792-003-0370-3

25. Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., & Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic acids research*, *37*(Database issue), D233–8. doi:10.1093/nar/gkn663

26. Chamier, B., Lorenz, M. G., & W. Wackernagel, W. (1993). Natural transformation of *Acinetobacter calcoaceticus* by plasmid DNA adsorbed on sand and groundwater aquifer material. *Appl. Environ. Microbiol. 59*, 1662-1667.

27. Chandel, A. K., Rudravaram, R., Rao, L. V., Ravindra, P., & Narasu, M. L. (2007). Industrial enzymes in bioindustrial sector development: An Indian perspective. *Journal of Commercial Biotechnology*, *13*(4), 283–291. doi:10.1057/palgrave.jcb.3050065

28. Chandel, A. K., Chandrasekhar, G., Silva, M. B., & Silvério da Silva, S. (2012a). The realm of cellulases in biorefinery development. *Critical reviews in biotechnology*, *32*(3), 187–202. doi:10.3109/07388551.2011.595385

29. Chandra, G., Chater, K. F., & Bornemann, S. (2011). Unexpected and widespread connections between bacterial glycogen and trehalose metabolism. *Microbiology (Reading, England)*, *157*(Pt 6), 1565–72. doi:10.1099/mic.0.044263-0

30. Chhabra, S. R., Shockley, K. R., Conners, S. B., Scott, K. L., Wolfinger, R. D., & Kelly, R. M. (2003). Carbohydrate-induced differential gene expression patterns in the hyperthermophilic bacterium *Thermotoga maritima*. *The Journal of biological chemistry*, *278*(9), 7540–52. doi:10.1074/jbc.M211748200

31. Cohen, G. N., Barbe, V., Flament, D., Galperin, M., Heilig, R., Lecompte, O., Poch, O., et al. (2003). An integrated analysis of the genome of the hyperthermophilic archaeon *Pyrococcus abyssi*. *Molecular microbiology*, *47*(6), 1495–512.

32. Conners, S. B., Mongodin, E. F., Johnson, M. R., Montero, C. I., Nelson, K. E., & Kelly, R. M. (2006). Microbial biochemistry, physiology, and biotechnology of hyperthermophilic *Thermotoga* species. *FEMS microbiology reviews*, *30*(6), 872–905. doi:10.1111/j.1574-6976.2006.00039.x

33. Dai, X., Zhu, Y., Luo, Y., Song, L., Liu, D., Liu, L., Chen, F., et al. (2012). Metagenomic insights into the fibrolytic microbiome in yak rumen. *PloS one*, *7*(7), e40430. doi:10.1371/journal.pone.0040430

34. Demain, A. L., Newcomb, M., & Wu, J. H. D. (2005). Cellulase, *Clostridia*, and Ethanol. *Microbiology and Molecular Biology Reviews*, *69*(1), 124–154. doi:10.1128/MMBR.69.1

35. Elkins, J. G., Raman, B., & Keller, M. (2010). Engineered microbial systems for enhanced conversion of lignocellulosic biomass. *Current opinion in biotechnology*, *21*(5), 657–62. doi:10.1016/j.copbio.2010.05.008

36. Energy Information Administration (EIA), *Annual Energy Outlook 2012*, DOE/EIA-0383 (2012) (Washington, DC, July, 2012). http://www.eia.gov/oiaf/aeo/pdf/appa.pdf

37. Eriksson, K.-E. L., Bermek, H. (2009). Lignin, Lignocellulose, Ligninase. *Encyclopedia of Microbiology*. Elsevier Inc.

38. Feng, Y., Yu, Y., Wang, X., Qu, Y., Li, D., He, W., & Kim, B. H. (2011). Degradation of raw corn stover powder (RCSP) by an enriched microbial consortium and its community structure. *Bioresource technology*, *102*(2), 742–7. doi:10.1016/j.biortech.2010.08.074

39. Ferrer, M., Ghazi, A., Beloqui, A., Vieites, J. M., López-Cortés, N., Marín-Navarro, J., Nechitaylo, T. Y., et al. (2012). Functional metagenomics unveils a multifunctional glycosyl hydrolase from the family 43 catalysing the breakdown of plant polymers in the calf rumen. *PloS one*, *7*(6), e38134. doi:10.1371/journal.pone.0038134

40. Fiala G., Stetter K. O. (1986). *Pyrococcus furiosus* sp. nov. represents a novel genus of marine heterotrophic archaebacteria growing optimally at 100 degrees C. *Arch Microbio.,* 145:56-61.

41. Freier, D., Mothershed, C. P., & Wiegel, J. (1988). Characterization of *Clostridium thermocellum* JW20. *Applied and environmental microbiology*, *54*(1), 204–211.

42. Frostegård, A., Courtois, S., Ramisse, V., Clerc, S., Bernillon, D., Le Gall, F., Jeannin, P., et al. (1999). Quantification of bias related to the extraction of DNA directly from soils. *Applied and environmental microbiology*, *65*(12), 5409–20.

43. Frostegård, A., Petersen, S. O., Bååth, E., Nielsen, T. H. (1997). Dynamics of a microbial community associated with manure hot spots as revealed by phospholipid fatty acid analyses. *Appl Environ Microbiol.63*(6):2224-31.

44. Fukui, T., Atomi, H., Kanai, T., Matsumi, R., Fujiwara, S., & Imanaka, T. (2005). Complete genome sequence of the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1 and comparison with *Pyrococcus* genomes. *Genome research*, *15*(3), 352–63. doi:10.1101/gr.3003105

45. García-Pedrajas, M. D., Bainbridge, B. W., Heale, J. B., Pérez-Artés, E., & Jiménez-Díaz, R. M. (1999). A Simple PCR-based Method for the Detection of the Chickpea-wilt Pathogen *Fusarium oxysporum* f.sp. ciceris in Artificial and Natural Soils. *European Journal of Plant Pathology*. *105*(3), 251-259.

46. Garcia Sanchez, R., Karhumaa, K., Fonseca, C., Sànchez Nogué, V., Almeida, J. R., Larsson, C. U., Bengtsson, O., et al. (2010). Improved xylose and arabinose utilization by an industrial recombinant *Saccharomyces cerevisiae* strain using evolutionary engineering. *Biotechnology for biofuels*, *3*, 13. doi:10.1186/1754-6834-3-13

47. Gibbs M. D., Reeves R. A., Farrington G. K., Anderson P., Williams D. P., Bergquist P. L. (2000). Multidomain and multifunctional glycosyl hydrolases from the extreme thermophile *Caldicellulosiruptor* isolate Tok7B.1. *Curr Microbiol.* May;40(5):333-40.

48. Gilbert M., Yaguchi M., Watson D. C., Wong K. K., Breuil C., Saddler J. N. (1993). A comparison of two xylanases from the thermophilic fungi *Thielavia terrestris* and *Thermoascus crustaceus*. *Appl Microbiol Biotechnol*. Dec;40(4):508-14.

49. Gomes I., Gomes J., Gomes D. J., Steiner W. (2000). Simultaneous production of high activities of thermostable endoglucanase and beta-glucosidase by the wild thermophilic fungus *Thermoascus aurantiacus*. *Appl Microbiol Biotechnol*. Apr;53(4):461-8.

50. Gomez, L. D., Steele-King, C. G., & McQueen-Mason, S. J. (2008). Sustainable liquid biofuels from biomass: the writing's on the walls. *The New phytologist*, *178*(3), 473–85. doi:10.1111/j.1469-8137.2008.02422.x

51. González, J. M., Masuchi, Y., Robb, F. T., Ammerman, J. W., Maeder, D. L., Yanagibayashi, M., Tamaoka, J., et al. (1998). *Pyrococcus horikoshii* sp. nov., a hyperthermophilic archaeon isolated from a hydrothermal vent at the Okinawa Trough. *Extremophiles : life under extreme conditions*, *2*(2), 123–30.

52. Graham, J. E., Clark, M. E., Nadler, D. C., Huffer, S., Chokhawala, H., Rowland, S. E., Blanch, H. W., et al. (2011). Identification and characterization of a multidomain hyperthermophilic cellulase from an archaeal enrichment. *Nature communications*, *2*, 375. doi:10.1038/ncomms1373

53. Guyen, N. H. N., Aruset, L. M., Engwetwanit, T. U., Huantong, W. M., Arnpicharnchai, P. H., Hampreda, V. C., Anapongpipat, S. T., et al. (2012). Identification and Characterization of a Cellulase-Encoding Gene from the Buffalo Rumen Metagenomic Library. *Bioscience, Biotechnology, and Biochemistry, 76*(6), 1075–1084. doi:10.1271/bbb.110786

54. Hamilton-Brehm, S. D., Mosher, J. J., Vishnivetskaya, T., Podar, M., Carroll, S., Allman, S., Phelps, T. J., et al. (2010). *Caldicellulosiruptor obsidiansis* sp. nov., an anaerobic, extremely thermophilic, cellulolytic bacterium isolated from Obsidian Pool, Yellowstone National Park. *Applied and environmental microbiology, 76*(4), 1014–20. doi:10.1128/AEM.01903-09

55. Hamilton-Brehm S. D., Vishnivetskaya T. A., Allman S. L., Mielenz J. R., Elkins J. G. (2012). Anaerobic high-throughput cultivation method for isolation of thermophiles using biomass-derived substrates. *Methods Mol Biol*.;908:153-68.

56. Harmsen, P., Huijgen, W., Bermudez Lopez, L. M., & Bakker, R. R. C. (2010). *Literature Review of Physical and Chemical Pretreatment Processes for Lignocellulosic Biomass* (pp. 1–49) (ECN-E-10-013).

57. Healy F. G., Ray R. M., Aldrich H. C., Wilkie A. C., Ingram L. O., Shanmugam K. T. (1995). Direct isolation of functional genes encoding cellulases from the microbial consortia in a thermophilic, anaerobic digester maintained on lignocellulose. *Appl Microbiol Biotechnol*. Aug-Sep;43(4):667-74.

58. Henderson-Maclennan, N. K., Papp, J. C., Talbot, C. C., McCabe, E. R. B., & Presson, A. P. (2010). Pathway analysis software: annotation errors and solutions. *Molecular genetics and metabolism, 101*(2-3), 134–40. doi:10.1016/j.ymgme.2010.06.005

59. Hongoh, Y. (2011). Toward the functional analysis of uncultivable, symbiotic microorganisms in the termite gut. *Cellular and molecular life sciences : CMLS, 68*(8), 1311–25. doi:10.1007/s00018-011-0648-z

60. Hubbert, M. K. (1956). Nuclear Energy and the Fossil Fuels 'Drilling and Production Practice. Spring Meeting of the Southern District. Division of Production. American Petroleum Institute. San Antonio, Texas: Shell Development Company. pp. 22–27. http://www.hubbertpeak.com/hubbert/1956/1956.pdf.

61. Huber R., Langworthy T. A., Konig H., Thomm M., Woese C. R., Sletyr U. B., Stetter K. O. (1986). *Thermotoga maritima* sp. nov. represents and new genus of unique extremely thermophilic eubacteria growing up to 90 -C. *Arch Microbiol.* 144:324-333.

62. Ingram, L., Gomez, P., Lai, X., Moniruzzaman, M., Wood, B., Yomano, L., & York, S. (1998). Metabolic engineering of bacteria for ethanol production. *Biotechnology and bioengineering*, *58*(2-3), 204–14.

63. Izquierdo, J., Sizova, M. V., & Lynd, L. R. (2010). Diversity of bacteria and glycosyl hydrolase family 48 genes in cellulolytic consortia enriched from thermophilic biocompost. *Applied and environmental microbiology*, *76*(11), 3545–53. doi:10.1128/AEM.02689-09

64. Jensen, E. B. (2009). *Conversion of renewables: Cellulosic bioethanol - R&D status.* (pp. 1–20).

65. Jeon, E., Hyeon, J.-E., Suh, D. J., Suh, Y.-W., Kim, S. W., Song, K. H., & Han, S. O. (2009). Production of cellulosic ethanol in *Saccharomyces cerevisiae* heterologous expressing *Clostridium thermocellum* endoglucanase and *Saccharomycopsis fibuligera* beta-glucosidase genes. *Molecules and cells*, *28*(4), 369–73. doi:10.1007/s10059-009-0131-y

66. Jiang, C.-J., Chen, G., Huang, J., Huang, Q., Jin, K., Shen, P.-H., Li, J.-F., et al. (2011). A novel β-glucosidase with lipolytic activity. *Folia Microbiol*. *56*, 563–570.

67. Kanokratana, P., Uengwetwanit, T., Rattanachomsri, U., Bunterngsook, B., Nimchua, T., Tangphatsornruang, S., Plengvidhya, V., et al. (2011). Insights into the phylogeny and metabolic potential of a primary tropical peat swamp forest microbial community by metagenomic analysis. *Microbial ecology*, *61*(3), 518–28. doi:10.1007/s00248-010-9766-7

68. Khanna, M., & Stotzky, G. (1992). Transformation of *Bacillus subtilis* by DNA bound on montmorillonite and effect of DNase on the transforming ability of bound DNA. *Appl. Environ. Microbiol. 58,* 1930-1939.

69. Knauf, M., & Moniruzzaman, M. (2004). Lignocellulosic biomass processing: A perspective. *International sugar journal*, *106*(1263), 147–150.

70. Kashima, Y., Mori, K., Fukada, H., & Ishikawa, K. (2005). Analysis of the function of a hyperthermophilic endoglucanase from *Pyrococcus horikoshii* that hydrolyzes crystalline cellulose. *Extremophiles : life under extreme conditions*, *9*(1), 37–43. doi:10.1007/s00792-004-0418-z

71. Kasting, J. F. (1993). Earth's Early Atmosphere. *Science*, *259*(5097), 920–926.

72. Lane, D. J. (1991). 16S/23S rRNA sequencing. In: *Nucleic acid techniques in bacterial systematics.* (E. Stackebrandt, M. Goodfellow, Eds.) (pp. 115-175). John Wiley and Sons, New York, NY.

73. Langston, J., Shaghasi, T., Abbate, E., Xu, F., Vlasenko, E., & Sweeney, M. D. (2011). Oxidoreductive cellulose depolymerization by the enzymes cellobiose dehydrogenase and glycoside hydrolase 61. *Applied and environmental microbiology*, *77*(19), 7007–15. doi:10.1128/AEM.05815-11

74. Lee, J. (1997). Biological conversion of lignocellulosic biomass to ethanol. *Journal of biotechnology*, *56*(1), 1–24.

75. Lee, K. C.-Y., Dunfield, P. F., Morgan, X. C., Crowe, M. a, Houghton, K. M., Vyssotski, M., Ryan, J. L. J., et al. (2011). *Chthonomonas calidirosea* gen. nov., sp. nov., an aerobic, pigmented, thermophilic micro-organism of a novel bacterial class, *Chthonomonadetes* classis nov., of the newly described phylum Armatimonadetes originally designated candidate division OP10. *International journal of systematic and evolutionary microbiology*, *61*(Pt 10), 2482–90. doi:10.1099/ijs.0.027235-0

76. López, M. J., Nichols, N. N., Dien, B. S., Moreno, J., & Bothast, R. J. (2004). Isolation of microorganisms for biological detoxification of lignocellulosic hydrolysates. *Applied microbiology and biotechnology*, *64*(1), 125–31. doi:10.1007/s00253-003-1401-9

77. Lorenz, M. G., Reipschliger, K., & Wackernagel, W. (1992). Plasmid transformation of naturally competent *Acinetobacter calcoaceticus* in nonsterile soil extract and groundwater. *Arch. Microbiol*. *157*, 355-360.

78. Lothier, J., Lasseur, B., Le Roy, K., Van Laere, A., Prud'homme, M.-P., Barre, P., Van den Ende, W., et al. (2007). Cloning, gene mapping, and functional analysis of a fructan 1-exohydrolase (1-FEH) from Lolium perenne implicated in fructan synthesis rather than in fructan mobilization. *Journal of experimental botany*, *58*(8), 1969–83. doi:10.1093/jxb/erm053

79. Lu, S., Li, L., & Zhou, G. (2010). Genetic modification of wood quality for second-generation biofuel production. *GM crops*, *1*(4), 230–6. doi:10.4161/gmcr.1.4.13486

80. Lu, Y., Zhang, Y.-H. P., & Lynd, L. R. (2006). Enzyme-microbe synergy during cellulose hydrolysis by *Clostridium thermocellum*. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(44), 16165–9. doi:10.1073/pnas.0605381103

81. Lukashin, A. V., & Borodovsky, M. (1998). GeneMark.hmm: new solutions for gene finding. *Nucleic acids research*, *26*(4), 1107–15.

82. Lynd, L. R., Weimer, P. J., Zyl, W. H. V., Isak, S., & Pretorius, I. S. (2002). Microbial Cellulose Utilization : Fundamentals and Biotechnology. *Microbiology and Molecular Biology Reviews*, *66*(3), 506–577. doi:10.1128/MMBR.66.3.506

83. Maki, M., Leung, K. T., & Qin, W. (2009). The prospects of cellulase-producing bacteria for the bioconversion of lignocellulosic biomass. *International journal of biological sciences*, *5*(5), 500–16.

84. Marchler-Bauer, A., Lu, S., Anderson, J. B., Chitsaz, F., Derbyshire, M. K., DeWeese-Scott, C., Fong, J. H., Geer, L. Y., Geer, R. C., Gonzales, N. R., Gwadz, M., Hurwitz, D. I., Jackson, J. D., Ke, Z., Lanczycki, C. J., Lu, F.,

Marchler, G. H., Mullokandov, M., Omelchenko, M. V., Robertson, C. L., Song, J. S., Thanki, N., Yamashita, R. A., Zhang, D., Zhang, N., Zheng, C., Bryant, S. H. (2011). CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res. 39*(Database issue):D225-9.

85. Margeot, A., Hahn-Hagerdal, B., Edlund, M., Slade, R., & Monot, F. (2009). New improvements for lignocellulosic ethanol. *Current opinion in biotechnology*, *20*(3), 372–80. doi:10.1016/j.copbio.2009.05.009

86. Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., et al. (2006). Genome Sequencing in Open Microfabricated High Density Picoliter Reactors. *Nature*, *437*(7057), 376–380.

87. Mohanty, B. K., & Kushner, S. R. (1999). Analysis of the function *of Escherichia coli* poly(A) polymerase I in RNA metabolism. *Molecular microbiology*, *34*(5), 1094–108.

88. Morana, A., Paris, O., Maurelli, L., Rossi, M., & Cannio, R. (2007). Gene cloning and expression in *Escherichia coli* of a bi-functional beta-D-xylosidase/alpha-L-arabinosidase from *Sulfolobus solfataricus* involved in xylan degradation. *Extremophiles : life under extreme conditions*, *11*(1), 123–32. doi:10.1007/s00792-006-0020-7

89. Murphy, D. J., & Hall, C. S. (2011). Energy return on investment, peak oil, and the end of economic growth. *Annals of the New York Academy of Sciences*, *1219*, 52–72. doi:10.1111/j.1749-6632.2010.05940.x

90. Mutanda, T., Ramesh, D., Karthikeyan, S., Kumari, S., Anandraj, A., & Bux, F. (2011). Bioprospecting for hyper-lipid producing microalgal strains for sustainable biofuel production. *Bioresource technology*, *102*(1), 57–70. doi:10.1016/j.biortech.2010.06.077

91. Nelson, K. E., Clayton, R., Gill, S. R., Gwinn, M. L., Dodson, R. J., Haft, D. H., Hickey, E. K., et al. (1999). Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*. *Nature*, *399*(6734), 323–9. doi:10.1038/20601

92. Nimchua, T., Thongaram, T., Uengwetwanit, T., Pongpattanakitshote, S., & Eurwilaichitr, L. (2012). Metagenomic analysis of novel lignocellulose-degrading enzymes from higher termite guts inhabiting microbes. *Journal of microbiology and biotechnology*, *22*(4), 462–9.

93. Ogram, A., Sayler, G. S., Gustin, D., & Lewis. R. J. (1988). DNA adsorption to soils and sediments. *Environ. Sci. Technol. 22*, 982-984.

94. Okeke, B. C., & Lu, J. (2011). Characterization of a defined cellulolytic and xylanolytic bacterial consortium for bioprocessing of cellulose and hemicelluloses. *Applied biochemistry and biotechnology*, *163*(7), 869–81. doi:10.1007/s12010-010-9091-0

95. Olson, D. G., Tripathi, S., Giannone, R. J., Lo, J., Caiazza, N. C., Hogsett, D., Hettich, R. L., et al. (2010). Deletion of the Cel48S cellulase from *Clostridium thermocellum*. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(41), 17727–32. doi:10.1073/pnas.1003584107

96. Osei-Poku, J., Mbogo, C. M., Palmer, W. J., & Jiggins, F. M. (2012). Deep sequencing reveals extensive variation in the gut microbiota of wild mosquitoes from Kenya. *Molecular ecology*, *21*(20), 5138–50. doi:10.1111/j.1365-294X.2012.05759.x

97. Paget, E., Simonet, L. J., & Monrozier, P. (1992). Adsorption of DNA on clay minerals: protection against DNase I and influence on gene transfer. *FEMS Microbiol. Lett. 97,* 31-40.

98. Parawira, W., Tekere M. (2010). Biotechnological strategies to overcome inhibitors in lignocellulose hydrolysates for ethanol production: review. *Crit Rev Biotechnol. 31*(1), 20-31.

99. Pedersen, M., Nissen, S., Mitarai, N., Lo Svenningsen, S., Sneppen, K., Pedersen, S. (2011). The functional half-life of an mRNA depends on the ribosome spacing in an early coding region. *J Mol Biol*. *407*(1), 35-44.

100. Peixoto, R. S., Vermelho, B., & Rosado, S. (2011). Petroleum-degrading enzymes: bioremediation and new prospects. *Enzyme research*, *2011*, 475193. doi:10.4061/2011/475193

101. Perevalova, A., Svetlichny, V., Kublanov, I. V., Chernyh, N., Kostrikina, N., Tourova, T. P., Kuznetsov, B. B., et al. (2005). *Desulfurococcus fermentans* sp. nov., a novel hyperthermophilic archaeon from a Kamchatka hot spring, and emended description of the genus *Desulfurococcus*. *International journal of systematic and evolutionary microbiology*, *55*(Pt 3), 995–9. doi:10.1099/ijs.0.63378-0

102. Pereyra, L. P., Hiibel, S. R., Prieto Riquelme, M. V., Reardon, K. F., & Pruden, A. (2010). Detection and quantification of functional genes of cellulose-degrading, fermentative, and sulfate-reducing bacteria and methanogenic archaea. *Applied and environmental microbiology*, *76*(7), 2192–202. doi:10.1128/AEM.01285-09

103. Pietramellara, G., Ascher, J., Ceccherini, M.T., Nannipieri, P., Wenderoth. (2007). Adsorption of pure and dirty bacterial DNA on clay minerals and their transformation frequency. *Biol. Fertil Soils. 43*, 731-739.

104. Podosokorskaya, O., Kublanov, I. V., Reysenbach, L., Kolganova, T. V., & Bonch-Osmolovskaya, E. (2011). *Thermosipho affectus* sp. nov., a thermophilic, anaerobic, cellulolytic bacterium isolated from a Mid-Atlantic Ridge

hydrothermal vent. *International journal of systematic and evolutionary microbiology*, *61*(Pt 5), 1160–4. doi:10.1099/ijs.0.025197-0

105. Podosokorskaya, O., Merkel, A. Y., Kolganova, T. V., Chernyh, N., Miroshnichenko, M. L., Bonch-Osmolovskaya, E., & Kublanov, I. V. (2011). *Fervidobacterium riparium* sp. nov., a thermophilic anaerobic cellulolytic bacterium isolated from a hot spring. *International journal of systematic and evolutionary microbiology*, *61*(Pt 11), 2697–701. doi:10.1099/ijs.0.026070-0

106. Qi, M., Wang, P., O'Toole, N., Barboza, P. S., Ungerfeld, E., Leigh, M. B., Selinger, L. B., et al. (2011). Snapshot of the eukaryotic gene expression in muskoxen rumen--a metatranscriptomic approach. *PloS one*, *6*(5), e20521. doi:10.1371/journal.pone.0020521

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

108. Rainey F.A., Donnison A.M., Janssen P.H., Saul D., Rodrigo A., Bergquist P.L., Daniel R.M., Stackebrandt E., Morgan H.W. (1994). Description of *Caldicellulosiruptor saccharolyticus* gen. nov., sp. nov: an obligately anaerobic, extremely thermophilic, cellulolytic bacterium. *FEMS Microbiol Lett.* Jul 15;120(3):263-6.

109. Rastogi, G., Bhalla, A., Adhikari, A., Bischoff, K. M., Hughes, S. R., Christopher, L. P., & Sani, R. K. (2010). Characterization of thermostable cellulases produced by *Bacillus* and *Geobacillus* strains. *Bioresource technology*, *101*(22), 8798–806. doi:10.1016/j.biortech.2010.06.001

110. Raven, P. H., Evert, R. F., Eichhorn, S. E. (2003). Biology of Plants, Sixth Edition. W. H. Freeman and Company / Worth Publishers, New York, NY.

111. Reese, E. T., Levinsons, H. S., Downing, M. (1950). Quartermaster culture collection. *Farlowia.* 4, 45-86.

112. Ravot, G., Magot, M., Fardeau, M. L., Patel, B. K., Prensier, G., Egan, Garcia, J. L., et al. (1995). *Thermotoga elfii* sp. nov., a novel thermophilic bacterium from an African oil-producing well. *International journal of systematic bacteriology*, *45*(2), 308–14.

113. Romanowski, G., Lorenz, M. G., & Wackernagel, W. (1991). Adsorption of plasmid DNA to mineral surfaces and protection against DNase I. *Appl. Environ. Microbiol. 57*, 1057-1061.

114. Romanowski, G., Lorenz, M. G., Wackernagel W. (1993). Use of polymerase chain reaction and electroporation of Escherichia coli to monitor the persistence of extracellular plasmid DNA introduced into natural soils. *Appl Environ Microbiol. 59*(10), 3438-46.

115. Rosgaard, L., Pedersen, S., Langston, J., Akerhielm, D., Cherry, J. R., & Meyer, A. S. (2007). Evaluation of minimal *Trichoderma reesei* cellulase mixtures on differently pretreated Barley straw substrates. *Biotechnology progress*, *23*(6), 1270–6. doi:10.1021/bp070329p

116. Scheller, H. V., & Ulvskov, P. (2010). Hemicelluloses. *Annual review of plant biology*, *61*, 263–89. doi:10.1146/annurev-arplant-042809-112315

117. Selinger, D. W., Saxena, R. M., Cheung, K. J., Church, G. M., & Rosenow, C. (2003). Global RNA half-life analysis in *Escherichia coli* reveals positional patterns of transcript degradation. *Genome research*, *13*(2), 216–23. doi:10.1101/gr.912603

118. Sensen, C. W., Charlebois, R. L., Chow, C., Clausen, I. G., Curtis, B., Doolittle, W. F., Duguet, M., et al. (1998). Completing the sequence of the *Sulfolobus solfataricus* P2 genome. *Extremophiles : life under extreme conditions*, *2*(3), 305–12.

119. Singhania, R. R., Sukumaran, R. K., Patel, A. K., Larroche, C., & Pandey, A. (2010). Advancement and comparative profiles in the production technologies using solid-state and submerged fermentation for microbial cellulases. *Enzyme and Microbial Technology*, *46*(7), 541–549. doi:10.1016/j.enzmictec.2010.03.010

120. Steinberg, C. E. W., Meinelt, T., Timofeyev, M. A., Bittner, M., & Menzel, R. (2008). Humic Substances Part 2: Interactions with Organisms. *Env Sci Pollut Res*, *15*(2), 128–135. doi:http://dx.doi.org/10.1065/espr2007.07.434

121. Stott, M. B., Crowe, M. a, Mountain, B. W., Smirnova, A. V., Hou, S., Alam, M., & Dunfield, P. F. (2008). Isolation of novel bacteria, including a candidate division, from geothermal soils in New Zealand. *Environmental microbiology*, *10*(8), 2030–41. doi:10.1111/j.1462-2920.2008.01621.x

122. Suen, G., Scott, J. J., Aylward, F. O., Adams, S. M., Tringe, S. G., Pinto-Tomás, A. a, Foster, C. E., et al. (2010). An insect herbivore microbiome with high plant biomass-degrading capacity. *PLoS genetics*, *6*(9). doi:10.1371/journal.pgen.1001129

123. Sukharnikov, L. O., Cantwell, B. J., Podar, M., & Zhulin, I. B. (2011). Cellulases: ambiguous nonhomologous enzymes in a genomic perspective. *Trends in biotechnology*, *29*(10), 473–9. doi:10.1016/j.tibtech.2011.04.008

124. Svetlichnyi, V.A., & Svetlichnii, T.P. (1988). *Dictyoglomus turgidus* sp. nov., a new extremely thermophilic eubacterium isolated from hot springs of the Uzon volcano caldera. *Mikrobiologiya*, 57, 435–441.

125. Svetlichnyi V.A., Svetlichnaya T.P., Chernykh N.A., Zavarzin G.A. (1990). *Anaerocellum thermophilum* gen. nov sp. nov: an extremely thermophilic cellulolytic eubacterium isolated from hot springs in the Valley of Geysers. *Microbiology*, 59:598-604.

126. Taherzadeh, M. J., & Karimi, K. (2007). Enzyme-based hydrolysis processes for ethanol from lignocellulosic materials: a review. *BioResources*, *2*(4), 707–738.

127. Tai, S.-K., Lin, H.-P. P., Kuo, J., & Liu, J.-K. (2004). Isolation and characterization of a cellulolytic *Geobacillus thermoleovorans* T4 strain from sugar refinery wastewater. *Extremophiles : life under extreme conditions*, *8*(5), 345–9. doi:10.1007/s00792-004-0395-2

128. Takada-Hoshino, Y., Matsumoto, N. (2004). An improved DNA extraction method using skim milk from soils that strongly absorb DNA. *Microbes Environ. 19*(1), 13-19.

129. Takahata, Y., Hoaki, T., & Maruyama, T. (2001). *Thermotoga petrophila* sp. nov. and *Thermotoga naphthophila* sp. nov., two hyperthermophilic bacteria from the Kubiki oil reservoir in Niigata, Japan. *International Journal of Systematic and Evolutionary Microbiology, 51*, 1901–1909.

130. Takayama, K., Kjelleberg, S. (2000). The role of RNA stability during bacterial stress responses and starvation. *Environ Microbiol*. *2*(4), 355-65.

131. Tang, Y., Zhu, L., Zhang, W., Shang, X., & Jiang, J. (2012). Integrated process of starch ethanol and cellulosic lactic acid for ethanol and lactic acid production. *Applied microbiology and biotechnology*. doi:10.1007/s00253-012-4461-x

132. Tartar, A., Wheeler, M. M., Zhou, X., Coy, M. R., Boucias, D. G., & Scharf, M. E. (2009). Parallel metatranscriptome analyses of host and symbiont gene expression in the gut of the termite *Reticulitermes flavipes*. *Biotechnology for biofuels*, *2*(i), 25. doi:10.1186/1754-6834-2-25

133. Te'o V. S., Saul D. J., Bergquist P. L. (1995). celA, another gene coding for a multidomain cellulase from the extreme thermophile *Caldocellum saccharolyticum*. *Appl Microbiol Biotechnol.* May-Jun;43(2):291-6.

134. Tejirian, A., & Xu, F. (2011). Inhibition of enzymatic cellulolysis by phenolic compounds. *Enzyme and microbial technology*, *48*(3), 239–47. doi:10.1016/j.enzmictec.2010.11.004

135. Tolonen, A. C., Chilaka, A. C., & Church, G. M. (2009). Targeted gene inactivation in *Clostridium phytofermentans* shows that cellulose degradation requires the family 9 hydrolase Cphy3367. *Molecular microbiology*, *74*(6), 1300–13. doi:10.1111/j.1365-2958.2009.06890.x

136. Touzel, J. P., Donohue, M. O., Debeire, P., Samain, E., & Breton, C. (2000). *Thermobacillus xylanilyticus* gen. nov., sp. nov., a new aerobic thermophilic xylan-degrading bacterium isolated from farm soil. *International Journal of Systematic and Evolutionary Microbiology*, *50*, 315–320.

137. Tuncer, M., Kuru, a, Isikli, M., Sahin, N., & Celenk, F. G. (2004). Optimization of extracellular endoxylanase, endoglucanase and peroxidase production by *Streptomyces* sp. F2621 isolated in Turkey. *Journal of applied microbiology, 97*(4), 783–91. doi:10.1111/j.1365-2672.2004.02361.x

138. Ueno Y., Haruta S., Ishii M., Igarashi Y. (2001). Microbial community in anaerobic hydrogen-producing microflora enriched from sludge compost. *Appl Microbiol Biotechnol*. Nov;57(4):555-62.

139. Van Dyk, J. S., & Pletschke, B. I. (2012). A review of lignocellulose bioconversion using enzymatic hydrolysis and synergistic cooperation between enzymes-Factors affecting enzymes, conversion and synergy. *Biotechnology advances*. doi:10.1016/j.biotechadv.2012.03.002

140. Van Ooteghem S. A., Jones A., Van Der Lelie D., Dong B., Mahajan D. (2004). H_2 production and carbon utilization by *Thermotoga neapolitana* under anaerobic and microaerobic growth conditions. *Biotechnol Lett.* 26:1223-1232.

141. Vazana, Y., Moraïs, S., Barak, Y., Lamed, R., & Bayer, E. (2010). Interplay between *Clostridium thermocellum* family 48 and family 9 cellulases in cellulosomal versus noncellulosomal states. *Applied and environmental microbiology, 76*(10), 3236–43. doi:10.1128/AEM.00009-10

142. Volossiouk, T., Robb, E. J., Nazar, R. N. (1995). Direct DNA extraction for PCR-mediated assays of soil organisms. *Appl Environ Microbiol. 61*(11), 3972-6.

143. Wang, P., Qi, M., Barboza, P., Leigh, M. B., Ungerfeld, E., Selinger, L. B., Mcallister, T. A., et al. (2011). Isolation of high-quality total RNA from rumen anaerobic bacteria and fungi, and subsequent detection of glycoside hydrolases. *Can. J. Microbiol. 598*, 590–598. doi:10.1139/W11-048

144. Webb, E.C., 1992. Enzyme nomenclature: recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the nomenclature and classification of enzymes. Academic Press, New York.

145. Wei, H., Xu, Q., Taylor, L. E., Baker, J. O., Tucker, M. P., & Ding, S.-Y. (2009). Natural paradigms of plant cell wall degradation. *Current opinion in biotechnology, 20*(3), 330–8. doi:10.1016/j.copbio.2009.05.008

146. Yankson, K. K., Steck, T. R. (2009). Strategy for extracting DNA from clay soil and detecting a specific target sequence via selective enrichment and real-time (quantitative) PCR amplification. *Appl Environ Microbiol. 75*(18), 6017-6021.

147. Zhang, W., Lou, K., & Li, G. (2010). Expression and characterization of the Dictyoglomus thermophilum Rt46B.1 xylanase gene (xynB) in Bacillus subtilis. *Applied biochemistry and biotechnology, 160*(5), 1484–95. doi:10.1007/s12010-009-8634-8

148. Zou, Z.-Z., Yu, H.-L., Li, C.-X., Zhou, X.-W., Hayashi, C., Sun, J., Liu, B.-H., et al. (2012). A new thermostable β-glucosidase mined from Dictyoglomus thermophilum: properties and performance in octyl glucoside synthesis at high temperatures. *Bioresource technology*, *118*, 425–30. doi:10.1016/j.biortech.2012.04.040