Bioprospecting the Regional Diversity of Australian Wine Microbiota

Lisa Anne Hartmann

Bachelor of Science (Molecular Biology, Honours)

Department of Molecular & Biomedical Science

School of Biological Sciences

The University of Adelaide

February

Contents

Abstract

Like all plants, grapevines are host to a plethora of microorganisms. During the winemaking process, the microflora of the grape surface can be transferred into the fermentation environment. While the yeast *Saccharomyces cerevisiae* is the main driver of alcoholic fermentation during wine production, it has become apparent that other grape and cellar-derived microbes can have a significant impact on fermentation and the resulting wine. This impact is manifested in the sensorial differences between wines resulting from inoculated fermentations and spontaneous 'wild' fermentations and is attributed to the various secondary metabolites and enzymes these microorganisms produce.

Carbohydrate active enzymes can affect the composition of grape juice by altering the extractability and organoleptic properties of many wine compounds. Of these enzymes, glycoside hydrolases mediate hydrolysis of glycosides and are found ubiquitously through nature. Micro-organisms, especially those associated with plants, use glycoside hydrolases to afford access to plant resources by degrading the plant cell wall and similarly soil microbes utilise these enzymes during degradation and composting of plant material. During winemaking, glycoside hydrolases have the potential to affect the breakdown of complex sugars in plant cell walls, which can aid in juice extraction and clarification. Glycoside hydrolases can also mediate the release of volatile flavour and aroma compounds from glycosyl-linked, non-volatile precursors. These enzymes therefore represent an attractive target for biotechnological applications.

A combination of metagenomics and synthetic biology were used to explore the enzymatic potential of two grapevine derived environments, an Chardonnay grape must (CGM) and a mixed varietal grape marc (MVGM), with the aim of identifying novel enzymes with potential applications as winemaking adjuncts. The CGM was specifically selected as a source of fungal microbiota native to the grapevine, whilst the grape marc was chosen as a source of plant decomposing microorganisms from a glucose-limited environment. Whole genome sequencing and bioinformatic techniques were used to probe the genomic landscape of these two environments, identify enzymes of interest, assess the types of microbiota

present in the microbial communities and where possible assemble microbial genomes. The two environments contained very distinct microbial communities. As expected, the Chardonnay must community comprised a high portion of fungal genomes whereas the MVGM was dominated almost exclusively by bacteria.

Five candidate polygalacturonases, with potential applications in wine clarification, and seven β-glucosidases, with potential applications in wine aroma and flavour, were chosen for heterologous expression in *Pichia pastoris* and *Saccharomyces cerevisiae*. Two of the selected polygalacturonases exhibited significant levels of activity at wine pH and were tested for in-wine activity, using recombinant *S. cerevisiae* strains in fermentation experiments with synthetic grape must. With future testing in real grape juice, these enzymes and *S. cerevisiae* strains could be leveraged to improve juice extraction and clarification during winemaking.

Statement of Originality

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship.

Date: 17/02/2020

Acknowledgements

Chapter 1: Introduction and Literature review

A brief history of bioprospecting

"Bioprospecting" was coined in the 1980s, but its essence, the exploration of biological sources for resources of value, is an age-old concept. Although micro-organisms were only discovered in the 1680s (Gest, 2004), humans have been inadvertently harnessing their biochemical abilities to produce fermented beverages for millennia. The ability of yeast to catalyse fermentation was proven in the 1860s (Manchester, 1995) and the isolation of pure yeast cultures in 1888 (Hildebrand, 1938). Shortly thereafter, in 1890, these pure yeast cultures were first used to inoculate wine ferments, introducing stability and reliability to "modern" winemaking (Pretorius, 2000) by enhancing fermentation and reducing the influence of the native microbiota by competitive inhibition. Since then, isolation of novel *S. cerevisiae* strains has become a staple of oenological research in order to prompt better and tailored winemaking and more interesting wine. More recently, focus has shifted to non-Saccharomyces yeast as sources of diverse activities not found in *S. cerevisiae* strains.

Bioprospecting in a genomic age

The advent of next generation sequencing has broadened the scope of bioprospecting (Oulas et al., 2015). Primarily, this is a consequence of metagenomic sequencing (i.e. sequencing of genetic material recovered directly from environmental samples) which allows access to the genetic diversity of unculturable microbial communities, including a plethora of previously unidentified microorganisms. Sequencing based bioprospecting can be applied to any kind of microbial community from which genetic material can be obtained and has invigorated research in areas such as bioremediation, nanotechnology, drug discovery and development, biocatalyst and antimicrobial discovery, with applications to agriculture, medicine, food and pharmaceutical industries. However, its potential in a wine industry context is yet to be realised.

A metagenomic based approached to bioprospecting typically focuses on the identification of enzymes of interest from a microbial community (Prakash and Taylor, 2012). As much of the microbial DNA as possible is extracted from the given environment, is sequenced and

assembled, and sequences are then interrogated by homology-based search tools to identify proteins which show homology to proteins of interest. Once enzymes of interest have been identified, candidate sequences can be tailored to (e.g. codon bias) and expressed in a host organism and screened (Hess et al., 2011, Schmeisser et al., 2007). This metagenomic sequencing approach eliminates culturing dependences but is dependent on acquiring a sufficient amount of DNA, sequencing depth to produce a high-quality metagenome and prior knowledge of structure of the enzyme of interest. However, the sequencing data can also be re-analysed as knowledge of protein structure and functional predictions are improved and also mined for other purposes.

Bioprospecting the microbiota of spontaneous fermentations

Spontaneous ('wild') fermentations, which proceed without the inoculation of selected microbial starter cultures, have highly diverse microbial communities (Pinto et al., 2014). A large proportion of this diversity originates from the microbiota native to the grapevine, *Vitis vinifera,* and is supplemented by microorganisms present in the winery (Pinto et al., 2014). Moreover, the microbial communities across different grape fermentations have also been shown to differ and can vary according to region, climate, grape variety, age of the vineyard and levels of domestication (Bokulich et al., 2014, Zarraonaindia et al., 2015, Masneuf-Pomarede et al., 2015, Stefanini et al., 2016).

Through the production and secretion of secondary metabolites and a range of enzymes, microbial communities that are native to the grape surface can affect the outcomes of the winemaking process (Masneuf-Pomarede et al., 2015, Verginer et al., 2010). Consequently, the effects that grape microflora exert on wine has driven the bioprospecting of non-*Saccharomyces* yeasts from this environment. Research has primarily focused on the isolation of non-*Saccharomyces* yeasts from spontaneous fermentations and characterising their whole-cell attributes, with the aim of identifying alternate commercial starter cultures. Bioprospecting of non-*Saccharomyces* yeasts has been shown to be highly successful but is limited to species that can be cultured and which produce the desired activity under winemaking conditions. For example, recent work by Belda et al. (2016), reported relative

enzyme activities of yeast isolates from three winemaking regions. Fifteen different species were isolated and identified by partial sequencing of the 26S large subunit rRNA and screened for β-glucosidase, β-lyase, protease, cellulase, polygalacturonase, α-L-arabinofurnosidase and sulphite reductase activities. Enzyme activities of interest were identified in several yeasts under winemaking conditions, including *Hanseniaspora uvarum, Metschnikowia spp.* and *Lachancea thermotolerans*. However, the enzyme activities of the isolated microorganisms were highly variable, with high β-glucosidase and protease activities, but low cellulase, polygalacturonase and α-L-arabinofurnosidase (glycosidase) activities generally observed.

Bioprospecting of Grape Marc

An estimated 10 - 20% of grape mass is comprised of insoluble material such as grape seeds, stems and skins (Dávila et al., 2017). These materials remain after winemaking and are referred to as grape marc. Grape marc is complex, consisting of residual sugars, proteins, fats, polyphenols, an abundance of polysaccharides and other complex molecules (Bravo and Saura-Calixto, 1998). As such grape marc is colonized by microorganisms capable of producing extracellular lytic enzymes in order to obtain nutrients and in many cases these types of enzyme activities have relevance to industry. Despite the potential of grape marc as a source of microorganisms and genes encoding useful enzyme activities, it has been little targeted for bioprospecting, with only one publication describing the isolation of novel yeast strains for industrial applications (Favaro et al., 2013). Similar to bioprospecting of yeasts isolated from spontaneous fermentations, these yeasts were tested for pectinase, lipase, amylase, protease, xylanase and cellulase activities (Favaro et al., 2013).

Enzymes for winemaking

Enzymes are an alternate target to whole cells for bioprospecting. The global market for industrial enzymes has been estimated at $4.2 billion and is continuing to grow (Singh et al., 2016). The growth in the enzyme market is supported by concerns about sustainability, environmental friendliness, improving efficiency and reducing costs. Enzymes are a means to target and actuate specific chemical reactions without the use of chemical pollutants and

at milder conditions. Unlike the use of genetically modified organisms (GMOs), enzymes are viable under current regulation whilst still affording access to beneficial characteristics from heterologous sources. In winemaking, enzyme formulations are used to enhance wine organoleptic (affecting flavour or aroma) properties or aid in wine production. Relying on microorganisms to achieve the same end, as happens in traditional winemaking practice, carries the risk of wine spoilage and other undesirable outcomes. There is interest, therefore, in the development of novel enzyme formulations with improved winemaking attributes. However, these enzymes will only be useful if they are functional in winemaking conditions which are often challenging to enzyme activity. Wine temperature, pH, sugar concentrations, substrate competition, sulphur dioxide additions and numerous other variables, can all potentially impact on enzyme activity (Benucci et al., 2015). As an additional challenge the progression of fermentation changes these conditions substantially. At the beginning of fermentation, grape must (the mixture of grape juice, stems, seeds and skins) is a nutrient rich environment with a low pH, high osmotic pressure caused by the high sugar content, a mixture of simple sugars such as glucose and complex polysaccharides, soluble and insoluble components as well as various minerals and other compounds. As alcoholic fermentation proceeds the simple sugars are converted into ethanol and many parameters, such as pH, shift.

Temperature is often used to control fermentation rate and is well known to affect enzyme activity. Typically, winemaking temperatures are 20-30 °C for red wines and 7-15 °C for white wines (Reynolds et al., 2001). There is also an increasing preference for low-temperature winemaking due to the potential to reduce negative effects on flavour (Molina et al., 2007, Reynolds et al., 2001). Thus there is growing interest in developing enzymes that have low-temperature optima (De Santi et al., 2016).

Carbohydrate active enzymes

Carbohydrates are an important part of grape composition and are unevenly distributed between the cell layers of grape berries. These carbohydrates can affect extraction of various compounds into grape juice with a range of efficiencies depending on cell type and winemaking practice (Sacchi et al., 2005). Although there is variation in the exact

4

composition between cultivars (Ortega-Regules et al., 2007), grape berries can be separated into three main types of tissue: the skin, the pulp and seeds. Pulp cells within mature grape berry are large, having swollen during ripening. They represent the main storage areas of free sugars and organic acids and are the source of most of the constituents of the final grape juice (Vidal et al., 2001, Saulnier and Thibault, 1987). The walls of the grape pulp cells are primarily composed of pectin and cellulose and soften during ripening (Doco et al., 2003). Pulp cells are easier to lyse than skin and seed cells and consequently, their cell wall components are more likely to be extracted into the grape juice. Following juice extraction, fermentation and winemaking, the pectin and other plant cell wall carbohydrates that are released from the pulp cell walls, can interfere with clarification and therefore are targeted for removal during clarification processes.

In contrast to the sugars and acids that are released from pulp cells, a significant portion of the phenolic compounds (anthocyanins and tannins), which contribute to the aroma, colour and texture of wine, are extracted primarily from the skin cells of the grape (Sacchi et al., 2005). Skin cells are typically tightly packed, have sturdier cell walls than pulp cells and are protected by wax layers to diminish penetration by fungal pathogens (Gabler et al., 2003). As such, grape berry skin cells are more resistant to cell wall degrading enzymes than pulp cells and the integrity of the skin cell wall affects the extraction of the organoleptically beneficial compounds that they contain. However, similar compounds derived from grape seeds typically contribute to negative wine characteristics such as bitterness. Consequently, grape juice extraction methods need to enhance extraction from grape skins but also avoid negative contributions from the grape seeds.

Various enzymes are active on carbohydrates such as those found in plant cell walls. These vary by mechanism, target substrate or substrates, efficiency, and response to physiological conditions. For example, glycoside hydrolases (or glycosidases) facilitate hydrolytic cleavage of glycosyl bonds. Similarly, polysaccharide lyases depolymerise polysaccharides through β-eliminative cleavage (Yip and Withers, 2006) whereas lytic polysaccharide

monooxygenases use an oxidative mechanism to cleave polysaccharides (Hemsworth et al., 2013, Vaaje-Kolstad et al., 2010).

Glycoside hydrolases (GHs) are ubiquitous throughout nature, with as many types as there are glycoside substrates. However, *S. cerevisiae* generally exhibits low GH activity, although GH activity is commonly observed other yeasts associated with wine (Moreno-Arribas and Polo, 2005, Capozzi et al., 2015). Decaying plant material, such as grape marc, is also an important niche for GH active microorganisms, as has been demonstrated for biofuel development (Li et al., 2009, Li et al., 2011).

The mechanism of action by glycoside hydrolases is broadly conserved. Most GHs utilise a pair of amino acid residues which act as a nucleophile and proton donor pair to introduce water across the glycosyl bond (Davies and Henrissat, 1995). These amino acid residues can be supplied by a single monomeric unit or by multiple units in a more complex quaternary protein structure. There are also exceptions where the nucleophile is obtained from an alternate source (Amaya et al., 2004, Watts et al., 2003), a neighbouring group (Terwisscha van Scheltinga et al., 1995), or using an exogenous base (Burmeister et al., 2000). An alternate GH mechanism requiring NAD has also been observed (Yip et al., 2007). Despite a broadly conserved mechanism, these enzymes display highly variable substrate preferences, with some enzymes showing activity towards a wide number of substrates whilst others are highly specific (Henrissat, 1991, Henrissat and Davies, 1997, Davies and Henrissat, 1995). As such, GHs are classified into families by amino acid structure and not by substrate activity, many glycoside hydrolase families (GHFs) encompass multiple types of enzymes and specific enzyme activities often appear across multiple families.

Pectinolytic enzymes (pectinases), are a group of enzymes responsible for the decomposition of pectin, a polysaccharide comprised of a polygalacturonic backbone and variable branching, which is highly abundant in plant cell walls. Complete degradation of pectin requires a combination of several enzymatic activities: depolymerisation of the backbone structure by polygalacturonase-mediated hydrolysis or eliminative cleavage of

non-reducing residues by pectin lyases, de-esterification by pectin esterases and the removal of non-polygalacturonic acid branches by various other glycoside hydrolases. Polygalacturonases and pectin lyases have highly related beta-sheet protein structures with polygalacturonase typically containing an additional turn in the beta-sheet (Pickersgill et al., 1998, Petersen et al., 1997, McDonough et al., 2004).

There is evidence that pectin degradation aids in juice extraction, clarification and flavour release during winemaking (Alimardani-Theuil et al., 2011), however pectinolytic enzymes are primarily only used to aid in the clarification of grape and other fruit juices. Pectinolytic activity is also relevant in the production of biofuel and in other industries interested in the utilisation of plant cell wall biomass. The temperature sensitivity of fruit-based juices has directed studies into cold-active pectinases for applications in these industries (Adapa et al., 2014, Merin and Morata de Ambrosini, 2015, Sahay et al., 2013). In particular, Merin et al. (2015), identified pectinase activity in whole-cell assays of non-*Saccharomyces* yeasts that had the potential for low temperature winemaking, however the specific enzymes were not identified or isolated. There is some evidence that polygalacturonases may have reduced efficacy on the plant cell walls of Cabernet Sauvignon skin cells whilst pectin lyases appear to remain more effective (Gao et al., 2016). The reduced activity of polygalacturonases may be due to the accessibility of the polygalacturonic acid backbone and may be improved by use of lytic polysaccharide monooxygenases to provide increased access to the polymer (Vaaje-Kolstad et al., 2010, Forsberg et al., 2011, Quinlan et al., 2011). The activity of pectinolytic enzyme may also be affected by the higher levels of esterification in the grape skin cell walls compared to the pectin of the grape pulp cells.

While non-Saccharomyces yeasts are favoured as sources for polygalacturonase activity, *Saccharomyces cerevisiae* strains do contain genes for, and can express, polygalacturonases. However, these genes are subject to strong repression in the presence of glucose and other primary carbon sources, which is typical of the regulation of pectinolytic enzymes (Radoi et al., 2005, Merin and Morata de Ambrosini, 2015). The repression of polygalacturonase activity in the presence of glucose suggests that glucose-poor

environments, where pectin is a more significant carbon source, may be better reservoirs for isolating enzymes with pectinolytic activity.

Similar to pectolytic enzymes, other plant cell wall degrading enzymes can also aid in juice clarification. This has been demonstrated for xylanases, which hydrolyse xylan, a hemicellulose (Polizeli et al., 2005). However, like pectin, complete decomposition of hemicelluloses requires the activity of multiple enzymes. The most commonly studied examples are endo-1-4-β-xylanases, which degrade the main xylan backbone, and β-D-xylosidases, which degrade the subsequent smaller poly-xylooligosaccharides. Extracellular xylanase expression is highly abundant in filamentous fungi (Polizeli et al., 2005). However, xylanase activity has not been observed in studies of non-*Saccharomyces* yeast isolates from grape and wine (Favaro et al., 2013, Belda et al., 2016b). Interestingly, some glycosidases have shown xylanase activity in addition to other glycosidase activities (Gruninger et al., 2014).

Enzymes targeting organoleptic grape glycosides

Outside of carbohydrates, glycosides containing a non-sugar moiety are also of significant oenological interest. Glycosides are produced in plants throughout development and perform a variety of roles. These glycosides can be carried over into grape juice and wine where the aglycones (non-sugar moieties of the glycosides) can have important organoleptic properties when released from the glycoside partner. Hydrolysis of certain grape glycosides can therefore significantly affect the flavour and aroma properties of wine (Gil et al., 2005, Palomo et al., 2005, Palmeri and Spagna, 2007, Valcárcel and Palacios, 2008). For instance, geraniol glucoside can be hydrolysed to release geraniol which, as a volatile, can produce a floral aroma. However, acid hydrolysis is slow under winemaking conditions and can cause unwanted molecular rearrangements (Mateo and Jimenez, 2000). Hence, enzymatic hydrolysis by GHs is the preferred route.

Grape glycosides can occur with variable numbers of sugar moieties and release of the aglycone can either occur directly from removal of the adjacent sugar or progressively as the compound is degraded (Figure 1.1). Although cleavage from higher glycosides may

increase the availability of the aglycone for secondary hydrolysis, in most cases the final release from the attached glycone is required for the perception of any organoleptic properties.
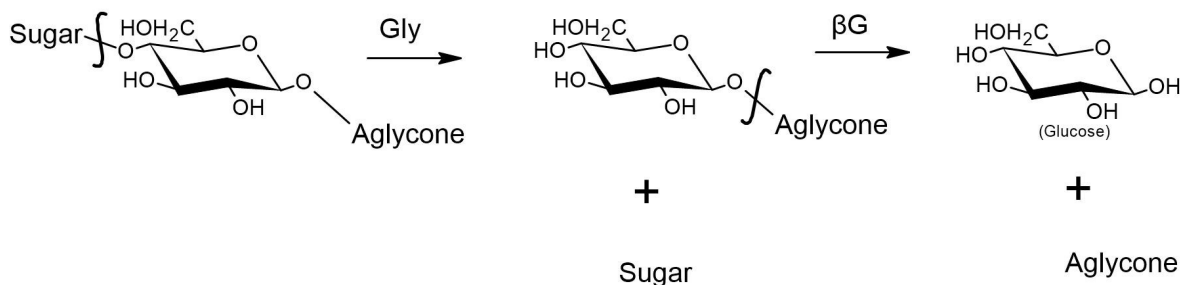
Figure 1.1: Sequential release of an aglycone from glycosidic complex by sequential enzymatic hydrolysis. 1) Initial cleavage by glycosidase (Gly) to release β-D-glucose-aglycone dimer. 2) Secondary cleavage of aglycone from β-D-glucose by β-glucosidase (βG).

In grapes, many organoleptic aglycones are linked to a glucose moiety (Williams et al., 1982, Voirin et al., 1990). β-glucosidases target terminal non-reducing residues of β-glucosides and β-glucosidase treatments have the potential to release flavour and aroma compounds in grape juice and in wines (Palmeri and Spagna, 2007, Gonzalez-Pombo et al., 2014, Riou et al., 1998, Sestelo et al., 2004). Previous studies have observed β-glucosidase activity both in wine yeasts and in lactic acid bacteria (Belda et al., 2016b, Michlmayr and Kneifel, 2014). Glucose can have variable effects on β-glucosidases activity. In some cases glucose inhibits β-glucosidase activity whilst other β-glucosidases are stimulated by the presence of glucose (De Giuseppe et al., 2014). Stimulation or at least glucose tolerance would be a beneficial trait for applications to winemaking.

Heterologous protein expression by *Saccharomyces cerevisiae* and *Pichia pastoris*

Synthetic biology methods are required to confirm the predicted function of the enzymes identified from metagenomic bioprospecting. There are numerous well-developed expression systems in both prokaryotic and eukaryotic hosts, which have been extensively reviewed and compared (Berlec and Strukelj, 2013, Unkles et al., 2014, Vieira Gomes et

al., 2018). In general, the heterologous expression of eukaryotic proteins often benefits from the use of eukaryotic host organisms, as this allows for eukaryotic-specific post-translational modifications to be performed, such as glycosylation and the formation of sulphite bridges when the protein is secreted.

The methylotrophic yeast *Pichia pastoris* allows for the production of high levels of heterologous protein with simplified protein purification protocols due to the combination of high cell culture density and a prolific secretory pathway (Cregg, 2008). The *P. pastoris* expression system also has well developed molecular protocols that utilise the methanol inducible promoters pAOX1 and pAOX2 to tightly control heterologous protein expression (Cregg, 2008).

S. cerevisiae has similar advantages to *P. pastoris* as an expression host, including the ability to secrete protein. *S. cerevisiae* does not grow to equivalent cell densities and has been noted to hyper-glycosylate proteins, which can be detrimental (Hoshida et al., 2013, Tang et al., 2016, Eckart and Bussineau, 1996). However, *S. cerevisiae* is well adapted to large-scale industrial conditions and for the production of various natural products (Billingsley et al., 2016). Significantly from a wine perspective, *S. cerevisiae* is inherently adapted to winemaking. This affords a potential for 'in-wine' fermentation experiments and negates the requirement to add protein to a grape fermentation.

Conclusions, thesis aims and methodology

Past studies of wine grape associated microbiota have been limited by the methodology of the time and potentially beneficial micro-organisms have not been investigated. However, recent developments in DNA sequencing and bioinformatics affords the opportunity for a deeper probe into microbial environments and into greater enzyme diversity.

This work aimed to identify and characterise novel enzymes for potential applications in winemaking from grape associated environments. A combination of metagenomic and synthetic biology techniques were used to identify candidate enzymes of interest from two wine grape derived environments: a pre-fermentation Chardonnay grape must and a composting mixed varietal grape marc.

From the 1.5 million predicted enzyme coding genes identified, seven putative β-glucosidases and five polygalacturonases, were selected for gene synthesis and heterologous expression in the yeasts *Pichia pastoris* and *Saccharomyces cerevisiae*. Enzymes of interest that were expressed successfully in *P. pastoris* were tested for putative activity and where applicable characterised for basic enzyme traits. Subsequently, two polygalacturonases were tested for activity in a wine-like environment using synthetic grape must and fermented by heterologous *S. cerevisiae* strains.

Chapter 2: Materials and Methods

2.1 Environmental samples and DNA extraction

Two grape derived environmental samples were collected for analysis: an unfermented organic certified 2016 Chardonnay grape must (Ngeringa Vineyards, SA) and a 2016 mixed varietal grape marc in the process of composting (Yangarra, SA).

Microorganisms were harvested from the Chardonnay grape must by centrifugation at 3000 rcf (Beckman-Coulter Allegra X-12D Centrifuge), washed with phosphate buffered saline (PBS: 0.02% (w/v) potassium chloride, 0.144% (w/v) sodium hydrogen phosphate, 0.024% (w/v) potassium dihydrogen phosphate, pH 7.2), and stored as pellets at –80 °C. DNA was extracted using ZR Soil Microbe MiniPrep Kit (Zymo Research) with an additional pre-lysis incubation at 65 °C for 10 minutes and beat beating steps performed at 8000 rpm on a Precellys Evolution Homogeniser (Bertin Technologies). DNA extracts were combined and concentrated using a Genomic DNA Clean and Concentrator -10 Kit (Zymo Research) and stored in elution buffer at -20 °C.

Subsamples of the grape marc were snap frozen with liquid nitrogen, ground into a moderately fine powder and stored at –80 °C in 50 ml tubes. DNA was extracted from 5 g of the frozen powder using ZR Soil Microbe DNA MidiPrep Kit (Zymo Research) with standard protocol and an additional elution step. DNA was stored in elution buffer at –20 °C.

DNA concentrations were determined by Qubit fluorometer (Invitrogen) and DNA shearing assessed on 1% (w/v) agarose.

2.2 Metagenomic and bioinformatics methods

2.2.1 Sequencing, taxonomic binning and assembly of Chardonnay and MVGM DNA

The 2016 Chardonnay must DNA was sequenced as a TruSeq nano library using a 2x250bp HiSeq 2500 Rapid run. The raw dataset was quality assessed by FastQC v0.11.5, trimmed by Trimmomatic (v0.36, parameters: PE, LEADING: 20, TRAILING: 20) and filtered against the Shiraz *Vitis vinifera* genome (Bowtie v1.1.2, SAMtools v1.9). Putative taxonomic assignments of reads were made using Centrifuge v1.0.3 (Kim et al., 2016) with the NCBI

NR protein database (downloaded 2018-04-06). The Chardonnay HiSeq data was assembled using Megahit (v1.0.3 with standard parameters) and assembly assessed using QUAST v4.3.

The 2016 mixed varietal grape marc (MVGM) metagenomic DNA was sequenced with Illumina and Oxford Nanopore Technologies (ONT). For the Illumina data DNA was prepared as a TruSeq library and run on an Illumina HiSeq X ten, to produce 150 bp paired end data. The MVGM HiSeq dataset was filtered to remove grape DNA sequences by alignment to the Shiraz *Vitis vinifera* genome (Bowtie v1.1.2, SAMtools v.1.9), quality filtered and trimmed by Trimmomatic (v0.36, parameters: PE, LEADING: 10, TRAILING: 10). Putative taxonomic assignment of reads was performed using Centrifuge (v1.03) with the NCBI NR protein database and standard parameters. The MVGM HiSeq dataset was assembled using Megahit (v1.1.1-2 using standard parameters).

For the ONT sequencing, MVGM metagenomic DNA was prepared as an SQK LSK108 library and sequenced on an Oxford Nanopore Technologies MinION using a FLO-MIN106 flow cell. The sequence dataset will be referred to as MVGM MinION. The MVGM MinION data was base called using Guppy v 3.2.4, quality filtered and filtered to remove grape vine sequences. Putative taxonomic assignments were made by Centrifuge v1.03 with the NCBI NR protein database using standard parameters. The MVGM MinION dataset was assembled by Meta-Flye (v2.5, parameters: --nano-raw $PATH/MVGM_ONT_data, --genome-size 2G, --meta). MVGM HiSeq and MVGM MinION assemblies were combined using Meta-Flye (v2.5, parameters: --subassemblies $PATH/MVGM_HiSeq_assembly $PATH/MVGM_ONT_assembly, --genome-size 2G, --meta). All assemblies were assessed by QUAST v4.3.

2.2.2 Contig binning and metagenome assembled genomes (MAGs)
Contigs larger than 5 kbp from the 2016 Chardonnay and 2016 MVGM assembles were given putative taxonomic assignment using Kaiju (v1.7.2). Bins were made using Metabat2 (v2.12.1) testing minimum contig sizes of 5 kbp and 10 kbp. Mmgenome2 (v2.0.7) was used to assess contig separation by GC, tetra-nucleotide, coverage and taxonomic assignment

and to bin the Chardonnay metagenome. Bins were assessed for completeness and contamination by CheckM (v1.0.7, default parameters), refined and filtered for >90% completeness and <10% contamination.

2.2.3 Gene predictions of 2016 Chardonnay metagenome and MVGM metagenome

Initial gene prediction of the 16-23 HiSeq Chardonnay metagenome was performed on the largest 5000 contigs using AUGUSTUS (v3.2.2, *Candida albicans* gene model). InterProScan (v5.22-61.0) was run on predicted sequences and gene families of interest collected. Gene families of interest were compared to the NCBI protein database by BLASTp (v2.7.1+) and predictions with exact matches were removed from further analysis. Gene predictions with unexpected sequence lengths relative to sizes of target of interest were flagged as potential gene model error i.e. truncations or elongation. Possible gene prediction errors were tested by comparison of *Candida albicans* gene model to alternate organism gene models, favouring models with closer taxonomic relation to gene prediction by BLASTp alignments, improved matches to domains identified by InterProScan and if necessary, a simpler model was selected. Multiple sequence alignment was performed on gene predictions belonging to families of interest by MUSCLE (v3.8.31) and the multiple sequence alignment clustered using PhyML (v20160207).

Subsequently, the full HiSeq Chardonnay metagenome was used in gene prediction with AUGUSTUS (v3.2.2, *Candida albicans* gene model) and predictions tested by InterProScan (v5.22-61.0, default parameters) and BLASTp of the NCBI NR protein database. Gene predictions were subsequently performed on the HiSeq MVGM metagenome using AUGUSTUS (v3.2.2, *Aspergillus fumigatus* gene model) and classified by InterProScan (5.22-61.0), gene families of interest compared to NCBI NR protein database by BLASTp (v2.7.1+). Corrections to enzymes of interest were made using the *E. coli* gene model in AUGUSTUS (v3.2.2). The gene predictions of target families of interested were combined and compared using a multiple sequence alignment produced by MUSCLE (v3.8.31) and clustered by PhyML (v20160207).

Three candidate GHF3 β-glucosidases were selected from the initial 5000 longest contigs Chardonnay Metagenome gene predictions, from separate clusters by MUSCLE (v3.8.31) alignment and PhyML (v20160207) clustering, for gene synthesis and subsequent gene expression experiments. Four candidate GHF1 β-glucosidases and five candidate GHF28 polygalacturonases were selected from these groups. SignalP (v4.0, default settings, eukaryotic, gram positive and gram negative bacterial models) was used to check sequences for signal peptides.

2.3 Molecular Techniques and Protocols, Recipes and microorganism strains

2.3.1 Plasmid miniprep

All plasmid DNA minipreps were performed from 5 ml overnight *E. coli* culture in LB containing 100 ug/ml ampicillin or 100 ug/ml zeocin at 37 °C. DNA was extracted using QIAprep Spin Miniprep Kit (QIAGEN) using standard protocol with elution buffer heated to 65 °C and eluted twice.

2.3.2 Gibson Cloning

Gibson cloning was performed using 2X Gibson Assembly Master Mix (NEB) according to the NEB standard protocols. A 1:3 ratio of vector to inserts with a total DNA mass ≤100 ng and made up to total volume of 10 ul. 10 ul of 2X Gibson Assembly was added and reaction mixture was incubated at 50 °C for 1 hour. 5 ul of reaction mixture was then used in transformations into *E. coli*.

2.3.4 *Saccharomyces cerevisiae* transformation protocol

S. cerevisiae competent cells were prepared from lab strain BY4742 and a *S. cerevisiae* wine isolate, AWRI1631, which were obtained from the AWRI culture collection and transformed by the Lithium acetate / ss carrier DNA / PEG method by Gietz and Schiestl, (2007). Competent cells were stored in 100 ul aliquots at -80 °C and thawed on ice before use.

Note: Plasmid DNA with HO homologous sites were linearized by restriction digest before transformation.

2.3.5 Yeast DNA extraction protocol

Harvested cells were resuspended in 250 ul Zymolyase solution (1.2 M sorbitol, 10 mM tris pH 8, 10 mM calcium chloride, 1% β-mercaptoethanol, 0.7 mg/ml zymolyase (20 U/mg)) and incubated for 30 minutes at 37 °C. 200 ul of lysis solution (50 mM tris pH 8, 50 mM EDTA, 1.2% (w/v) SDS) was added and mixed before the addition 100 ul of potassium acetate solution (5 M potassium acetate neutralized pH 5.5 with acetic acid). Precipitated protein was removed by centrifugation for 10 minutes at 16,110 rcf on an Eppendorf Centrifuge 5415D and supernatant transferred into fresh tube with 700 ul isopropanol. Nucleic acids were harvested by centrifugation for 5 minutes at 16,110 rcf and resuspended in 100 ul TE (10 mM Tris-HCl, 1 mM EDTA) and treated with RNase at 37 °C for 1 hour. DNA was precipitated with 10 ul sodium acetate and 100 ul isopropanol and pellet washed with 95% ethanol and then resuspended in 50 ul TE (10 mM Tris-HCl, 1 mM EDTA). DNA concentration was measured by using a Qubit fluorometer (Invitrogen).

2.3.6 Pichia pastoris electroporation

Electrocompetent cells were prepared from the *P. pastoris* strain BG11 which was obtained from ATUM. An overnight culture of BG11 in 10 ml YPD at 28 °C, was used to inoculate 50 ml YPD to OD_{600}= 0.15 - 0.2 and incubated at 28 °C, shaking, until reaching OD_{600} = 0.8 - 1.0. Cells were harvested at 3000 rcf on an Allegra X-12R Centrifuge (Beckman-Coulter) for 5 minutes and resuspended in 25 ml YPD/HEPES (20:1, 1 M HEPES, pH 8.0) and 650 ul 1 M DTT and incubated at 28 °C for 15 minutes without shaking. Cells were harvested at 3000 rpm for 5 minutes, 4 °C on an Allegra X-12R Centrifuge (Beckman-Coulter) and washed twice with 25 ml ice cold 1 M sorbitol and resuspended in a final volume of 500 ul of ice cold 1 M sorbitol. Cells were aliquoted (40 ul per reaction) and kept on ice immediately before use or stored at -80 °C.

DNA was linearized by restriction digest with *Sac*I-HF (NEB) at 37 °C for 1 hour and heat inactivated at 65 °C. DNA was then cleaned with QIAquick PCR purification kit (QIAGEN) using standard protocol.

40 ul of electrocompetent cells and 100 ng of plasmid DNA were incubated in pre-chilled 0.2 cm electroporation cuvette (Bio-Rad) on ice for minimum of 5 minutes. Cells were

exposed to an exponential decay pulse (2000 V for 5s) were applied using Gene Pulser Xcell II (Bio-Rad). 1 ml of ice-cold sorbitol was added immediately after electroporation and cells were incubated for 1 hour without shaking at 30 °C, then aliquots plated onto YPD-zeocin (50-200 ug/ml) and incubated for 2 - 3 days.

2.3.7 Methanol induction of pAOX1

1ml of an overnight YPD culture of appropriate *P. pastoris* strain was used to inoculate 50 ml of BMGY (1% (w/v) yeast extract, 2% (w/v) peptone, 13.4 g/L yeast nitrogen base without amino acids, 100 mM potassium phosphate pH 6.0, 0.4 mg/L biotin, 1% glycerol) and incubated at 28 °C, shaking. At 24 hours, cells were harvested at 3000 rcf for 5 minutes on an Allegra X-12R Centrifuge (Beckman-Coulter) and resuspended in 50 ml of BMM2 (0.2 M potassium phosphate, 13.4 g/L yeast nitrogen base without amino acids, 0.4 mg/L biotin, 0.5% (w/v) methanol) was added to induce protein production. 500 ul of methanol was then added every 24 hours for up to 96 hours to boost induction.

2.3.8 Protein Purification and Quantification

Supernatant and cells were harvested from yeast culture samples by centrifugation at 3000 rcf for 5 minutes on an Allegra X-12D Centrifuge (Beckman-Coulter). Protein lysates were prepared from cellular component by bead beating lysis in 1% SDS at 4000 rpm 3 x 30 s on Precellys Evolution Homogeniser with 0.1 mm zirconium beads. Cellular debris were removed by centrifugation. Protein was then extracted 100 ul of lysate or supernatant using the Methanol/Chloroform method. In brief, 400 ul of methanol, 100 ul of chloroform and 300 ul of water were added to the 100 ul of protein sample, vortexed and centrifuged at 16,110 rcf on an Eppendorf Centrifuge 5415 D for 3 minutes. The aqueous layer was removed with care not to disturb the protein pellet and 300 ul of methanol added, vortexed and centrifuged (16,110 rcf, Eppendorf Centrifuge 5415 D). Supernatant was discarded and protein pellet wash with 300 ul of methanol twice, dried and resuspended in resuspension buffer (0.1% SDS, 50 mM potassium phosphate, 1 mM DTT) or 25 ul 2 N sodium hydroxide for Lowry protein estimation.

Protein extracts were quantified using Lowry protein estimation and Bio-Rad protein dye concentrate (Bio-Rad) using standard protocol.

Lowry protein estimation: protein samples were extracted by Methanol/Chloroform method and resuspended in 25 ul of 2 N sodium hydroxide. 250 ul of fresh complex forming reagent were added (2% (w/v) sodium carbonate, 0.01% (w/v) copper sulphate pentahydrate, 0.02% (w/v) sodium potassium tartrate) and incubated at room temperature for 10 minutes. 25 ul of 1 N Folin–Ciocalteu reagent was added, vortexed and incubated at room temperature for 30-60 minutes. Absorbance was read at 750 nm. BSA concentrations from 0 to 1.5 mg/ml were used to create a standard curve.

2.3.9 Protein Gel Electrophoresis

Samples were diluted with NuPAGE LDS Sample Buffer (Invitrogen) with DTT added to a final concentration of 50 mM. Protein electrophoresis was performed using precast NuPAGE 10% Bis-Tris gels (Invitrogen) in MES SDS Running Buffer (Invitrogen, 50 mM MES, 50 mM Tris Base, 0.1% (w/v) SDS, 1 mM EDTA, pH 7.3) at 200 V for 35 minutes with biotinylated protein ladder (Cell Signalling Technology).

2.3.10 Coomassie staining

Coomassie stains were performed using Imperial Protein Stain (Invitrogen) according to the manufacturer instructions.

2.3.11 Western blotting

Proteins were transferred on a nitrocellulose membrane (0.45 um, Invitrogen) at 30 V for 1 hour in NuPAGE transfer buffer (Invitrogen) with methanol (10% per gel). Membranes were blocked with 20 ml SuperBlock (PBS) Blocking Buffer (Invitrogen) for 30 minutes. Membrane was immunoblotted with 1:1000 primary Anti-his mouse antibody (Cell Signalling Technology) for 1 hour at room temperature, washed 3 x 5 minutes with 20 ml of PBS Tween 20 and then immunoblotted with 1:2000 anti-mouse HRP conjugated and 1:2000 anti-biotin HRP conjugated antibody (Cell Signalling Technology) for 1 hour. Membrane was then washed 5 times with 20 ml PBS Tween 20 for 5 minutes and then visualised with 10 ml 1-Step Ultra TMB-Blotting solution (Invitrogen).

2.3.12 Media Recipes

LB (lysogeny buffer): 1% (w/v) tryptone, 0.5% (w/v) yeast extract, 1% (w/v) sodium chloride.

LB agar: 1% (w/v) tryptone, 0.5% (w/v) yeast extract, 1% (w/v) sodium chloride and 1% (w/v) agarose.

LB-Lennox: 1% (w/v) tryptone, 0.5% (w/v) yeast extract and 0.5% (w/v) sodium chloride.

LB-Lennox agar: 1% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.5% (w/v) sodium chloride and 1% (w/v) agarose.

YPD: 1% (w/v) yeast extract, 2% (w/v) peptone and 2% (w/v) glucose.

YPD agar: 1% (w/v) yeast extract, 2% (w/v) peptone, 2% (w/v) glucose and 2% (w/v) agarose.

10X Amino acid mix: 0.4 g/ml adenine sulfate, 0.2 g/ml uracil, 0.4 g/ml L-tryptophan, 0.2 g/ml L-histidine hydrochloride, 0.2 g/ml L-arginine hydrochloride, 0.3 g/ml L-lysine hydrochloride, 0.5 g/ml L-phenylalanine, 1.0 g/ml L-glutamic acid, 1.0 g/ml L-asparagine, 1.5 g/ml L-valine, 2.0 g/ml L-threonine, 3.75 g/ml L-serine, 0.2 g/ml L-methionine

SC Media: 6.7 g/L yeast nitrogen base without amino acids, 20 g/L glucose, 1X Amino acid mix

SC+ Media: 6.7 g/L yeast nitrogen base without amino acids, 100 g/L glucose, 1X Amino acid mix

2.3.13 Chemically Defined Media (Synthetic grape must)

100 g/L glucose, 100 g/L fructose, 0.2 g/L citric acid, 3 g/L malic acid, 2.5 g/L tartaric acid, 0.238 g/L dibasic potassium phosphate, 0.1 g/L monobasic sodium phosphate, 0.5 g/L magnesium sulphate, 0.183 g/L calcium chloride, 0.011 g/L boric acid, 0.84 g/L proline with 1 ml/L of Nitrogen stock mix, 1 ml/L Vitamin stock mix, 1 ml/L Trace elements stock and made to pH 3.5 with potassium hydroxide.

Nitrogen stock:

10.5 g/L L-alanine, *7.2 g/L* γ-aminobutyric acid, 27 g/L L-arginine, 0.4 g/L L-asparagine, 3 g/L L-aspartic acid, 0.4 L-citrulline, 6 g/L L-glutamic acid, 8.4 g/L L-glutamine, 0.4 g/L glycine, 1.2 g/L L-histidine, 1.2 g/L L-isoleucine, 1.2 g/L L-leucine, 0.4 g/L L-lysine, 0.4 g/L

L-methionine, 0.4 g/L L-ornithine monohydrochloride, 0.8 g/L L-phenylalanine, 5.4 g/L L-serine, 6 g/L L-threonine, 0.4 g/L L-tryptophan, 0.4 g/L L-tyrosine, 2.1 g/L L-valine and 1.2 g/L L-cysteine.

Vitamin stock:

0.5 g/L thiamine hydrochloride, 0.2 g/L Riboflavin, 1 g/L pyridoxine hydrochloride, 1 g/L calcium pantothenate, 1 g/L nicotinic acid, 10 g/L myo-Inositol, 0.05 g/L biotin, 0.05 g/L folic acid and 0.05 g/L 4-amino benzoic acid.

Trace elements stock:

3 g/L manganese sulphate, 1 g/L zinc chloride, 6 g/L iron sulphate heptahydrate, 1.5 g/L copper sulphate pentahydrate, 0.01 g/L potassium iodate, 0.03 g/L cobalt nitrate hexahydrate, 0.025 g/L sodium molybdate dihydride, 0.1 g/L lithium chloride, 0.05 g/L nickel sulphate hexahydrate and 0.7 g/L rubidium chloride.

2.3.14 Polygalacturonic acid Chemically Defined Media (Synthetic PG Must)

Made as 1.2X Chemically Defined Media (2.3.12) without citric acid and potassium phosphate and then made to 1X with addition of (10 g/L) polygalacturonic acid dissolved in citric acid-sodium phosphate buffer, pH 4.5, to final concentration of 2 g/L.

2.4 Expression of GHF3 β-glucosidases *in S. cerevisiae* and *P. pastoris*

2.4.1 Preparation of GHF3 plasmids for S. cerevisiae expression

Production of GH3-1, GH3-2 and GH3-3 was tested using two *S. cerevisiae* expression systems. pCVSα, a yeast integrating plasmid using *HO* homologous sites for genomic integration and a high copy 2 μm episome based expression system with pCV3S. Both systems utilise the *FBA1* promoter to drive protein expression and a modified alpha factor to facilitate protein secretion.

GH3-1, GH3-2 and GH3-3 were tested for native signal peptides by SignalP 4.0 and potential signal peptides were removed. The coding sequences for the three GHF3 candidates were codon optimised for expression in *S. cerevisiae* and synthesised (Genscript Biotech). Sequences were excised from the cloning plasmid (pUC57) by restriction digest with *Bse*RI and *Sca*I. Fragments were purified from agarose gel (1% (w/v)

agarose) with QIAquick Gel Extraction Kit (QIAGEN) using the standard protocol. The expression plasmid backbone, pCVSα, was digested with *Bbs*I and *Nhe*I and purified by QIAquick PCR Purification Kit (QIAGEN), using the standard protocol. Synthesized GHF3 variants were cloned into pCVSα by Gibson Assembly (2.3.2) to form pCVSα-GH3-1, pCVSα-GH3-2 and pCVSα-GH3-3. 5 ul of Gibson reaction mixtures were transformed into 10-beta Competent *E. coli* (NEB) according to High Efficiency Transformation Protocol (NEB) and selected on LB agar with 100 ug/ml ampicillin at 37 °C overnight. Plasmid construction (pCVSα-GH3-1, pCVSα-GH3-2 and pCVSα-GH3-3) was confirmed by restriction digest of plasmid DNA (*Nde*I: pCVSα-GH3-1. *Xba*I: pCVSα-GH3-2 and pCVSα-GH3-3) extracted from ampicillin resistant *E. coli* by miniprep (2.3.1).

The 2 μm episome plasmids were produced using the *FBA1* promoter, alpha-factor, GH3 and terminator sequence produced in pCVSα-GH3-1, pCVSα-GH3-2 and pCVSα-GH3-3 as the 2 μm episome (pCV3S) backbone plasmid does not contain the alpha factor sequence. The GH3 sequences were excised from pCVSα-GH3-1, pCVSα-GH3-2 and pCVSα-GH3-3 by restriction digest with *BamH*I and *Spe*I. GH3 sequences were purified from restriction digest by extraction from 1% (w/v) agarose gel using QIAquick Gel Extraction Kit using the standard protocol and ligated into pCV3S-KanMX (*Bam*HI and *Spe*I digested pCV3-pc4cl2) using T4 ligase (NEB, modified standard protocol with 1 hour at room temperature ligation step). Ligation reaction mixtures were transformed into 10-beta *E. coli* (NEB) using standard protocol and transformants were selected on LB agar with 100 ug/ml ampicillin. Plasmids (pCV3S-GH3-1, pCV3S-GH3-2 and pCV3S-GH3-3) were confirmed by *Pvu*I restriction digest of plasmid DNA extracted from ampicillin resistance *E. coli* by miniprep (2.3.1).

2.4.2 Production and protein expression of GHF3 recombinant *S. cerevisiae* strains

To aid in transformation and integration into the *P. pastoris* genome, the plasmid DNA of pCVSα-GH3-1, pCVSα-GH3-2 and pCVSα-GH3-3 was linearized before transformation by restriction digest with *Bsa*I (pCVSα-GH3-1) and *Nde*I (pCVSα-GH3-2 and pCVSα-GH3-3). Linear DNA was purified by QIAquick PCR purification kit (QIAGEN) by the standard

protocol. Linearized pCVSα-GH3-1, pCVSα-GH3-2 and pCVSα-GH3-3 plasmids, and circular pCV3S-GH3-1, pCV3S-GH3-2 and pCV3S-GH3-3 plasmids were transformed into *S. cerevisiae* strains, BY4742 and AWRI1631, using the *S. cerevisiae* transformation protocol (2.3.4).

To test for genomic integration colony PCR of the *HO* integration sites was performed on single isolates transformed with pCVSα-GH3-1, pCVSα-GH3-2 and pCVSα-GH3-3 plasmids. PCR was performed using OneTaq polymerase. Primer pairs: 5'HO1 F (5'-CGCCTTTGTCTTTTGCCTTTTTCA-3') and 5'HO1 CVS R (5'-TGCAGCACATCCCCCTTTCG-3'), and 3' HO1 CVS F (5'-AGGGTTCGCAAGTCCTGTTTCTATG-3') 3' HO1 R (5'-TGCTGTCGATTCGATACTAACGCC-3'). PCR conditions were: 94 °C for 30 s, 36 x (94 °C for 20 s, 54 °C (5' HO1) or 57 °C (3' HO1) for 20 s, 68 °C for 60 s), 68 °C for 5 minutes. Amplification products were visualised on 1% (w/v) agarose with RedSafe nucleic acid stain solution (iNtRON). Where possible a minimum of 16 single colonies per strain were tested from each transformation with 6 colonies tested in three separate rounds of PCR, unless positive transformants were found in earlier round.

In order to test the production of GHF3 recombinant protein of BY4742 transformed withs pCV3S-GH3-1, pCV3S-GH3-2 and pCV3S-GH3-3 plasmid, overnight cultures were grown in 5 ml YPD at 28 °C, on a rotating platform. 1 ml of the overnight cultures was used to inoculate 50 ml of SC+ media and grown at 28 °C, shaking for 24 hours. Samples were taken at 24 hours, cells harvested at 16,110 rcf for 2 minutes and supernatant retained for protein analysis. Proteins were isolated from supernatant samples taken at 24 hours by the methanol/chloroform method (2.3.8). Protein production was assessed by SDS-PAGE (2.3.9) and strained by Coomassie (2.3.10).

2.4.3 Production GHF3 plasmids for expression in *P. pastoris*

Yeast integrating plasmid expression of the GHF3 candidates in *P. pastoris* were produced from a modified version of the pD912-AKS expression plasmid (ATUM), referred to as simply "pD912" (Appendix A). pD912-AKS was modified to contain homologous regions to

pCVSα to allow Gibson assembly with either plasmid vector using the same gene fragment. pD912 (and pD912-AKS) contain a shortened alpha factor secretion signal to facilitate protein secretion and pAOX1 to facilitate methanol inducible protein expression and genomic integration. An N-terminal His tag variant of pD912-AKS was produced by cloning a synthetic DNA sequence (Gene fragment g-block Integrated DNA Technologies) containing a His tag into the plasmid by Gibson assembly (Appendix A) and used to produced N-terminal His tag variants of the GHF3 candidate enzymes. C-terminal His tag variants were produced by Gibson assembly with additional gene fragments containing the His tag coding sequence.

Synthesis of the *S. cerevisiae* codon-optimised coding sequences of GH3-1, GH3-2 and GH3-3 (the same synthesised sequences as 2.4.1) were excised from cloning plasmids (pUC57-GH3-1, pUC57-GH3-2 and pUC57-GH3-3) by restriction digest with *Bse*RI (NEB, standard protocols). Gene fragments were cloned into pD912, digested with *Bsa*I and *Nhe*I (NEB), and pD912-His, digested with *Bsg*I and *Nhe*I, by Gibson Assembly (2.3.2). 5 ul of Gibson reaction mixtures were transformed into 10-beta *E. coli* (NEB) using the High Efficiency Transformation Protocol (NEB) and selected on LB-Lennox agar with 100 ug/ml zeocin at 37 °C overnight. Plasmid DNA (pD912-GH3-1, pD912-GH3-2, pD912-GH3-3, pD912-His-GH3-1, pD912-His-GH3-2 and pD912-His-GH3-3) was extracted from zeocin resistant *E. coli* by DNA miniprep (2.3.1) and plasmid construction was confirmed by restriction digest with *Nde*I (pD912-GH3-1), *Xba*I (pD912-GH3-2 and pD912-GH3-3), *Sac*I (pD912-His-GH3-1) or *Nco*I (pD912-His-GH3-2 and pD912-His-GH3-3).

To produce the C-terminal His tagged variant pD9-12-GH3-2-His, pD912-GH3-2 was digested with *Nhe*I and the gene fragment Gibson assembled with the synthesised gene fragment GH3-2-His (Table 2.2). The C-terminal His tagged variants pD912-GH3-1 and pD912-GH3-3-His were produced by first excising the coding sequences from pUC57-GH3-1 and pUC57-GH3-3 by restriction digest with *Bse*RI and *Nhe*I (NEB, standard protocol). pD912-AKS was digested with *Bsa*I and *Nhe*I (NEB, standard protocol). GH3-1 and GH3-3 fragments, and the digested pD912 were then purified from 1% (w/v) agarose with QIAquick

Gel Extraction Kit (QIAGEN) using the standard protocol. GH3-1 and GH3-3 fragments were assembled with digested pD912 and G-blocks (Table 2.2), GH3-1-His and GH3-3-His, respectively, by Gibson Assembly (2.3.2), to form pD912-GH3-1-His and pD912-GH3-3-His. Reaction mixtures for the three plasmids (pD912-GH3-1-His, pD912-GH3-2-His and pD912-GH3-3-His) were transformed into *E. coli* and plasmid construction confirmed by restriction digest of plasmid DNA with *Xho*I.

Table 2.2: Synthesised DNA fragments (Gene fragments G-blocks from Integrated DNA technologies) for C-terminal His tag modification

Name	Sequence
GH3-1-His	5'- TGAATTTGGCCGGTAGTCCAGGTGGTTACGGTAAACCAAAGCCTAG ATATGTTAGAGGCCACCATCACCACCACCACTAGTGAGACCAGTGCGT CATTGCAAGTAGTCGGTTAAGGGGCGGCCGC – 3'
GH3-2-His	5'- TCTCAGCTGGTTTTTCCAGTGGTGACTTTATTGCACAAACAGAAATG AAGTTGTTAGGCCACCATCACCACCACCACTAGTGAGACCAGTGCGTC ATTGCAAGTATCGGTTAAGGGGCGGCCGC – 3'
GH3-3-His	5'- ACATCAGATTGACAGATAGATTACATATCCAACATGATCACAAGTGG AGAGGTTTAGGCCACCACCATCACCACCACTAGTGAGACCAGTGCGTC ATTGCAAGTAGTCGGTTAAGGGGCGGCCGC – 3'

2.4.4 Production and protein expression of GHF3 recombinant *P. pastoris* strains

The expression plasmids for the candidate GHF3 beta-glucosidases (pD912-GH3-GH3-1, pD912-GH3-GH3-2, pD912-GH3-GH3-3, pD912-His-GH3-GH3-1, pD912-His-GH3-GH3-2 and pD912-His-GH3-GH3-3, pD912-GH3-GH3-1-His, pD912-GH3-GH3-2-His and pD912-GH3-GH3-3-His) were transformed into the *P. pastoris* strain BG11 (ATUM) by electroporation (2.3.6). Transformants were selected on YPD agar with zeocin (50, 100, 150 and 200 ug/ml).

In order to confirm genomic integration of plasmid DNA in the pAOX1 locus, the integration sites were PCR amplified from genomic DNA of zeocin resistant isolates. Genomic DNA was extracted from overnight 5 ml YPD cultures (30 °C, rotating) using the Yeast DNA extraction protocol (2.3.5). The 5' and 3' genomic integration sites were PCR amplified with primer pairs 5' pAOX1 F (5'-TGCCCCCAAATCCAATGAGACT-3') and 5'pD912 R (5'-

CTGGCGGAAAATGGCAAAACA-3'), and 3' pD912 F (5'-GAGCGTCAGACCCCGTAGAAAAGA-3') and 3' pAOX1 R (5'-TGTATCCCCTCCTGTTGCGTTTG-3'). PCR conditions were: 94 °C for 30 s, 36 x (94 °C for 20 s, 55 °C (5' pAOX1) or 57 °C (3' pAOX1) for 15 s, 68 °C for 60 s), 68 °C for 5 minutes. OneTaq polymerase (NEB) and OneTaq standard reaction buffer were used.

Once plasmid integration was confirmed, protein expression of the recombinant strains was tested by induction of pAOX1 with methanol (2.3.7) for 96 hours with samples taken at 12 and 24 hours and then every 24 hours. Protein was precipitated by methanol/chloroform method (2.3.8), run on SDS-PAGE (2.3.9) and stained with Coomassie (2.3.10) or immunoblotted (2.3.11) as appropriate.

2.4.5 RT-qPCR

The induction of the pAOX1 promoter by methanol of the *P. pastoris* containing the C-terminal GHF3 β-glucosidases plasmids was tested by RT-qPCR. The C-terminal His tagged GHF3 β-glucosidases *P. pastoris* strains and BCAP8 control strain were inoculated into 50 ml BMGY (1% (w/v) yeast extract, 2% (w/v) peptone, 13.4 g/L yeast nitrogen base without amino acids, 100 mM potassium phosphate pH 6.0, 0.4 mg/L biotin, 1% (w/v) glycerol) with 1 ml of overnight YPD culture and incubated at 28 °C, shaking for 24 hours. An uninduced sample was taken at 24 hours and stored as pellet at -80 °C. Cells from remaining culture were harvested at 3000 rcf for 5 minutes with a Allegra X-12R centrifuge (Beckman-Coulter) and resuspended in 50 ml BMMY (1% (w/v) yeast extract, 2% (w/v) peptone, 13.4 g/L Yeast Nitrogen Base without amino acids, 100 mM potassium phosphate pH 6.0, 0.4 mg/L biotin, 0.5% (w/v) methanol) and incubated at 28 °C, shaking. Induced sample was taken after 6 hours incubation and stored as pellet at -80 °C. RNA was extracted from uninduced and induced cells using PureLink™ RNA Mini Kit (Life Technologies) with an additional bead beating step to lyse cells on a Bertin Precellys Evolution (8000 rpm, 4 x 20 s, 30 s pause). RNA was treated with 10 ul RNase free DNase (NEB) at 37 °C for 30 minutes. RT was performed using M-MuLV Reverse Transcriptase (NEB) according the standard protocol using 50 ng of RNA, 2 ul of d(t)$_{23}$ (50 µM), 1 ul of dNTP (10 mM) and water

to 10 ul heated to 65 °C for 5 minutes, then made to 20 ul with 2 ul of buffer, 1 ul of M-MuLV RT enzyme, RNase inhibitor and water and then reaction run at 42 °C for 1 hour. M-MuLV RT was heat inactivated at 65 °C for 20 minutes. cDNA was stored at –20 °C. qPCR reaction was prepared with 10 ng of cDNA with KAPA SYBR FAST Universal (Sigma). qPCR was run on BioRad machine. qPCR conditions were: 95 °C for 3 minutes, 40 x (95 °C for 3 s, 60 °C for 20 s).

Table 2.3: Primer pairs for RT-qPCR experiments.

Primer	Sequence
GH3-1 F	5' – CCTAATAACGGTGGTGGTGCTC – 3'
GH3-1 R	5' - AGGACACAAATAATCTGCGGTTCT – 3'
GH3-2 F	5' – CGCAAACGAACAAGAAACTCAA – 3'
GH3-2 R	5' - TCCATATCCAAACCACCATTAGC – 3'
GH3-3 F	5' – TAGCCGAAACCCCTGATGC – 3'
GH3-3 R	5' - GGACAAACCGTGACCGAATG – 3'
BCAP8 F	5' - GAGTTGACTTTGGGTGGTGTTGA – 3'
BCAP8 R	5' – GTTGTAAGCAGCAGTTGGGATGTA – 3'

2.4.6 Sanger sequencing

The 5' genomic integration region of *P. pastoris* transformed with pD912-GH3-1, pD912-GH3-2 and pD912-GH3-3 were PCR amplified with OneTaq and 5' pAOX1 F: 5'-TGCCCCCAAATCCAATGAGACT-3' and 5'pD912 R: 5'-CTGGCGGAAAATGGCAAAACA-3', T_a = 55 °C. PCR conditions were: 94 °C for 30 s, 36 x (94 °C for 20 s, 55 °C (5' pAOX1) or 57 °C (3' pAOX1) for 15 s, 68 °C for 60 s), 68 °C for 5 minutes. PCR products were purified using QIAquick PCR purification kit (Qiagen) using standard protocol and eluting into water.

PCR products were then Sanger sequenced with the 5' pAOX1 R primer by AGRF and sequences trimmed for low quality.

2.5 Expression of candidate GHF1 β-glucosidases in *P. pastoris*

2.5.1 Construction of GHF1 *P. pastoris* expression plasmids

Expression of GHF1 candidates was performed using a yeast integrating plasmid, pD912. pD912 uses pAOX1 for methanol induced protein expression and as a homologous region for genomic integration. pD912 also contains the modified alpha secretion signal to facilitate protein secretion. All four GHF1 candidate β-glucosidases were checked for signal peptides (SignalP 4.0), codon optimised for expression in *P. pastoris* and coding sequences were synthesised (TWIST Biosciences) with a C-terminal His tag and pD912 homologous sites for Gibson assembly (Gibson et al., 2009). Coding sequences were excised from pTWIST cloning plasmids (pTWIST-GH1-1, pTWIST-GH1-2, pTWIST-GH1-3 and pTWIST-GH1-4) by *Bse*RI (NEB) digest and then purified form 1% (w/v) agarose using a QIAquick Gel Extraction kit (QIAGEN). Coding sequences were cloned into *Nco*I and *Nhe*I digested pD912 by Gibson Assembly and transformed into 10-Beta *E. coli* (NEB) using standard protocols and LB agar (1% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.5% (w/v) sodium chloride) with zeocin (100 ug/ml zeocin). Plasmid DNA (pD912-GH1-1, pD912-GH1-2, pD912-GH1-3 and pD912-GH1-4) was extracted by miniprep (2.3.1). Plasmid DNA construction was confirmed by restriction digest with *Nde*I and *Xho*I (NEB) using standard protocols.

2.5.2 Production and protein expression by GHF1 recombinant *P. pastoris*

The *P. pastoris* expression plasmids, pD912-GH1-1, pD912-GH1-2, pD912-GH1-3 and pD912-GH1-4, were transformed into *P. pastoris* strain BG11 (ATUM) by electroporation as per the *P. pastoris* electroporation protocol (2.3.6). Transformant colonies were selected from YPD zeocin agar and re-streaked. Genomic DNA was extracted from 5 ml YPD cultures (overnight, 37 °C) per Yeast DNA extraction protocol (2.3.5). The 5' and 3' genomic integration sites were PCR amplified using OneTaq polymerase (NEB) and the OneTaq standard reaction buffer. 5' pAOX1 primer pair: 5' pAOX1 F: 5'-TGCCCCCAAATCCAATGAGACT-3' and 5'pD912 R: 5'-CTGGCGGAAAATGGCAAAACA-3', T_a = 55 °C. 3' pAOX1 primer pair: 3' pD912 F: 5'-GAGCGTCAGACCCCGTAGAAAAGA-3' and 3' pAOX1 R: 5'-TGTATCCCCTCCTGTTGCGTTTG-3', T_a = 57 °C. PCR conditions

were: 94°C for 30 s, 36 x (94 °C for 20 s, 55 °C (5' pAOX1) or 57 °C (3' pAOX1) for 15 s, 68 °C for 60 s), 68 °C for 5 minutes.

Recombinant *P. pastoris* strains were induced with methanol (2.3.7) for 96 hours with samples taken at 12 and 24 hours and then every 24 hours. Protein was extracted by methanol/chloroform method (2.3.8), run on SDS-PAGE (2.3.9) and immunoblotted (2.3.11).

2.5.4 β-glucosidase assays

Temperature optima of GHF1 β-glucosidases were determined with reactions of 50 ul of crude supernatant containing recombinant β-glucosidase with 150 ul of p-nitrophenyl glucopyranoside solution (5 mM p-nitrophenyl glucopyranoside, citric acid-sodium phosphate buffer pH 6.0) for 30 minutes at temperature 10 °C to 100 °C. Reaction were cooled on ice, 100 ul of 0.4 M sodium hydroxide was added and incubated for 15 minutes on ice. Absorbance measured at 400 nm with a FLUOstar Omega spectrophotometer (BMG LABTECH).

pH optima of GH1 β-glucosidases were determined with reactions of 50 ul crude supernatant containing recombinant β-glucosidase with 150 ul of p-nitrophenyl glucopyranoside solution (5 mM p-nitrophenyl glucopyranoside, citric acid-sodium phosphate buffer, pH 3 - 7) and incubated at 37 °C for 30 minutes. Reactions were cooled on ice, 100 ul of 0.4 M sodium hydroxide was added and incubated for 15 minutes on ice. Absorbance measured at 400 nm a BMG LABTECH FLUOstar Omega spectrophotometer.

Substrate kinetics were followed at 420 nm (on a BMG LABTECH FLUOstar Omega spectrophotometer) with 300 ul of reactions mixtures for a maximum of 30 minutes. Reaction mixture contained 10 ul of crude supernatant containing recombinant β-glucosidases, 290 ul reaction solution (500 mM sodium chloride, 100 mM HEPES pH 7.5, from 1 mM to 20 mM p-nitrophenyl glucopyranoside). Change of absorbance over time determined by slope of linear regression with MARS 3.32 R5 (BMG LABTECH) software.

2.6 Expression of GHF28 candidates in *P. pastoris*

2.6.1 Construction of GHF28 plasmids for *P. pastoris* expression

Coding sequences for GH28-1, GH28-2, GH28-3, GH28-4 and GH28-5 were codon optimised for *P. pastoris* expression and gene synthesised (TWIST Biosciences). Coding sequences were excised from the pTWIST cloning vector by restriction digest with *Bse*RI and then Gibson cloned into *Nco*I and *Nhe*I digested pD912 and transformed by heat shock into competent 10-Beta *E. coli* (NEB) and selected for on LB zeocin. Plasmid DNA was extracted from zeocin resistant colonies by miniprep (2.3.1). Plasmid DNA was identified by digest with a combination of restriction enzymes (NEB) as per standard protocols.

2.6.2 Construction and protein expression of recombinant GHF28 *P. pastoris* strains

The *P. pastoris* expression plasmids, pD912-GH28-1, pD912-GH28-2, pD912-GH28-3, pD912-GH28-4 and pD912-GH28-5, were transformed into *P. pastoris* strain BG11 (ATUM) by electroporation as per the electroporation protocol (2.3.6). Transformant colonies were selected from YPD zeocin agar and re-streaked. Genomic DNA was extracted from 5 ml YPD cultures (overnight, 37 °C) per Yeast DNA extraction protocol (2.3.5). The 5' and 3' pAOX1 genomic integration sites were PCR amplified with OneTaq polymerase (NEB). 5' pAOX1 primer pair: 5' pAOX1 F 5'-TGCCCCCAAATCCAATGAGACT-3' and 5'pD912 R 5'-CTGGCGGAAAATGGCAAAACA-3', T_a = 55 °C. 3' pAOX1 primer pair: 3' pD912 F 5'-GAGCGTCAGACCCCGTAGAAAAGA-3' and 3' pAOX1 R 5'-TGTATCCCCTCCTGTTGCGTTTG-3', T_a = 57 °C. PCR conditions were: 94 °C for 30 s, 36 x (94 °C for 20 s, 55 °C (5' pAOX1) or 57 °C (3' pAOX1) 15 s, 68 °C for 60 s), 68 °C for 5 minutes.

Confirmed recombinant *P. pastoris* strains were induced with methanol (2.3.7) for 96 hours with samples taken at 12 and 24 hours and then every 24 hours. Protein was extracted by methanol/chloroform method (2.3.8) and immunoblotted (2.3.11). Image analysis was performed using IMAGEJ.

2.6.3 Characterisation of recombinant GHF28 protein

The recombinant protein for the GHF28 candidates produced by *P. pastoris* transformed with pD912-GH28-1, pD912-GH28-2 and pD912-GH28-3 were tested for polygalacturonase

activity from the crude supernatant BMMY in which the cells were induced on polygalacturonic acid. Preliminary tests were performed with 0.25 ml of 2.5 mg/ml polygalacturonic acid and 0.25 ml of crude supernatant for 30 minutes at 37 °C. Polygalacturonase activity (i.e. production of reducing sugars from polygalacturonic acid) was measured in reaction mixtures using a modified di-nitrosalicylic assay (Miller, 1959). An equal volume of the DNS reagent (10 g/L di-nitrosalicylic acid, 300 g/L sodium potassium tartrate tetrahydrate, 16 g/L sodium hydroxide) was added to the polygalacturonase reaction mixtures, boiled for 15 minutes and then placed on ice. 50 ul of solutions were transferred to a 96 well plate (CLS3603, Corning), diluted with 150 ul of water and absorbance at 540 nm was measured with a Spectramax M2 fluorescence spectrophotometer. Samples were compared to a galacturonic acid standard curve, concentrations 0 - 1.75 mg/ml galacturonic acid monohydrate and boiled supernatant blanks, prepared under equivalent conditions. Enzyme activity (U) is defined as the μmol of galacturonic acid equivalent reducing sugars released per minute.

Subsequent to preliminary tests, the pH, temperature and substrate kinetics of GH28-1 and GH28-2 were tested. pH and temperature optima were determined using recombinant enzyme in crude supernatant in reaction mixtures with 2.5 mg/ml polygalacturonic acid in citric acid – sodium phosphate buffer (made from 0.2 M citric acid and 0.1 M sodium phosphate stocks). pH 3.0 to 7.0 and temperatures 10 °C to 90 °C were tested. For GH28-1, pH reactions were performed at 30 °C for 60 minutes, 50 ul of 2.5 mg/ml polygalacturonic acid and 50 ul of supernatant, and temperature reactions at pH 4.5 for 60 minutes, 50 ul 2.5 mg/ml polygalacturonic acid. For GH28-2 pH reactions were performed at 50 °C for 30 minutes, 95 ul of 2.5 mg/ml polygalacturonic acid and 5 ul of supernatant and temperature reactions at pH 4.5 for 30 minutes, 75 ul 2.5 mg/ml polygalacturonic acid with 25 ul GH28-2 supernatant. Boiled supernatant was used as a negative control and reactions mixtures were cooled on ice after the reaction period. Polygalacturonase activity was determined by addition of 100 ul of DNS reagent and boiled for 15 minutes. Samples compared to galacturonic acid standard curve with 50 ul of samples and standards diluted in 150 ul of

water on 96 well plate and absorbance measured at 540 nm using a FLUOstar Omega Spectrophotometer (BMG LABTECH). To determine substrate kinetics, recombinant GH28-2 supernatant solution (5 ul) was added to 50 ul of reaction buffer (0.2 M citric acid, 0.1 M sodium phosphate buffer, pH 4.5) containing different concentrations of polygalacturonic acid (0.5, 1, 1.5, 2, 2.5, 3, 4 and 5 mg/ml). Polygalacturonase activity was measured for 0 to 15 minutes to determine linear rate of reaction. 55 ul of DNS reagent was added and boiled for 15 minutes to facilitate redox reaction between reducing sugars and dinitrosalicyclic acid and then put on ice. 50 ul of the samples and galacturonic acid standards were diluted in 150 ul of water in a 96 well plate and absorbance at 540 nm measured on a FLUOstar Omega Spectrophotometer (BMG LABTECH). 5 ul of boiled supernatant was used as negative control for each polygalacturonic acid concentration. Enzyme velocity at each substrate concentration was determined from the slope of the linear regression of reducing sugar produced (equivalent μmol of galacturonic acid) over time using GraphPad Prism 8. Enzyme substrate kinetics were assessed by nonlinear regression of substrate concentration against enzyme velocity, curve fitting Michalis-Menten and Allosteric Sigmoidal models. Models were compared using inbuilt extra-sum-of-squares F test.

2.7 Protein expression of GH28-1 and GH28-2 by *S. cerevisiae*

2.7.1 *S. cerevisiae* expression systems for GH28-1 and GH28-2

Expression of GH28-1 and GH28-2 was tested in two expression systems. pCVSα is a yeast integrating plasmid with regions homologous to the *HO* locus for expression. The pCV3S plasmids are designed for 2 μm episomal based expression. Both systems use the *FBA1* promoter for constitutive protein expression and contain a modified alpha-factor sequence in the N-terminus to facilitate protein secretion. The coding sequences for GH28-1 and GH28-2 were excised from the pTWIST cloning plasmids by restriction digest with *Bse*RI (NEB) using the standard protocol. DNA fragments were cloned into *Bbs*I and *Nhe*I double digested pCVSα by Gibson Assembly (NEB) and 5 ul of reaction mixtures were transformed into 10-beta *E. coli* (NEB) using the High Efficiency Transformation Protocol (NEB). Transformants were selected on LB agar with 100 ug/ml ampicillin. Plasmid DNA was

isolated from ampicillin resistant isolates by miniprep (2.3.1). Construction of pCVSα-GH28-1 and pCVSα-GH28-2 plasmids was assessed by restriction digest with *Eco*RI (NEB, standard protocol).

As pCV3-NatR does not contain the alpha secretion signal the gene cassettes from the pCVSα-GH28-1 and pCVSα-GH28-2 were excised by restriction digest with *Bsg*I and *Xba*I (NEB, standard protocol). pCV3-NatR was also digested with *Bsg*I and *Xba*I (NEB, standard protocol). The desired plasmid DNA fragments were purified from 1% (w/v) agarose using the QIAquick Gel Extraction Kit (QIAGEN) using the standard protocol. Digested DNA fragments were ligated to form pCV3S-GH28-1 and pCV3S-GH28-2 using T4 ligase (NEB) at room temperature for 90 minutes. The T4 ligase was then heat inactivated at 65°C for 20 minutes. Ligation mixtures were transformed into 10-Beta *E. coli* (NEB) by heat shock using standard protocol and recovered on LB agar with 100 ug/ml ampicillin. Plasmid DNA was isolated by miniprep (2.3.1) and pCV3S-GH28-1 and pCV3S-GH28-2 were confirmed by *Xho*I (NEB) digest.

pCVSα-GH28-1, pCVSαs-GH28-2, pCV3S-GH28-1 and pCV3S-GH28-2 were transformed into AWRI1631 by the Lithium Acetate/ SS carrier DNA / PEG method (2.3.4) and recovered on YPD agar with zeocin (pCVSα) or 50 ug/ml clonat (pCV3S) for 3 days. Genomic DNA was extracted from the AWRI1631 transformants using the Yeast DNA extraction protocol (2.3.5).

To confirm genomic integration of pCVSα-GH28-1 and pCVSα-GH28-2 transformants, the 5' and 3' regions of the *HO* genomic integration PCR amplified using primer pairs 5' HO1 F (5' – CGCCTTTGTCTTTTGCCTTTTTCA – 3') and 5' HO1 CVS R (5' – TGCAGCACATCCCCCTTTCG – 3'), and 3' HO1 CVS F (5' – AGGGTTCGCAAGTCCTGTTTCTATG – 3') 3' HO1 R (5' – TGCTGTCGATTCGATACTAACGCC – 3'). PCR conditions were: 94 °C for 30 s, 36 x (94 °C for 20 s, 55 °C (5' HO1) or 57 °C (3' HO1) for 15 s, 68 °C for 60 s), 68 °C for 5 minutes. PCRs were performed using OneTaq Polymerase (NEB) using the OneTaq standard reaction buffer.

To confirm the transformation of pCV3S-GH28-1 and pCV3S-GH28-2 into *S. cerevisiae* AWRI 1631, regions of GH28-1 and GH28-2 loci were PCR amplified from *S. cerevisiae* transformants' DNA using OneTaq Polymerase (NEB) with the OneTaq standard reaction buffer. GH28-1 primers: GH28-1 F 5' – TGGTGGTTGGGGAGATGTTAGA – 3' and GH28-1 R 5' – TGGAGTTGGCAGCGGTTCT – 3'. GH28-2 primers: GH28-2 F 5' – AAAATCTTGGGTTGGCGGTC – 3' and GH28-2 R 5' – CAGTTTGGGCTAGATTGGGC – 3'. PCR reaction conditions were: 94 °C for 30 s, 36 x (94 °C for 20 s, 54 °C (GH28-1) or 57 °C (GH28-2) for 15 s, 68 °C for 30 s), 68 °C for 5 minutes. Where possible a minimum of 16 single colonies per strain were tested from each transformation with 6 colonies tested in three separate rounds of PCR, unless positive transformants were found in earlier round.

Confirmed transformants were used to ferment SC+ media (0.67% (w/v) yeast nitrogen base, 100 g/L glucose) for 48 hours. Protein from 5 ml supernatant at 48 hours was concentrated to approximately 1 ml by centricon-10 concentrators (Amicon) and protein extracted from 100 ul of concentrated sample by methanol-chloroform protein extraction. Protein precipitates were resuspended in 20 ul of 1 X SDS loading dye and immunoblotted (Western blot protocol 2.3.10).

2.7.2 Synthetic grape must (Chemically Defined Media) fermentations

S. cerevisiae strains AWRI1631, AWRI4241 (AWRI1631 transformed with pCV3-GH28-1) and AWRI4240 (AWRI1631 transformed with pCV3-GH28-2) were first inoculated into 5 ml of YPD (2% (w/v) peptone, 1% (w/v) yeast extract and 2% (w/v) glucose) in triplicate from separate plate colonies and cultured at 28 °C, rotating overnight. 1 ml of overnight cultures were used to inoculate 25 ml of 1:1 YPD and synthetic grape must (pH 3.5, Chemically Defined Media recipe 2.3.13.), and cultured for 24 hours, shaking at 28 °C, to acclimatise cells to the synthetic grape must. OD_{600} was measured at 24 hours post inoculation of 1:1 YPD and synthetic grape must cultures and then used to inoculate 100 ml of 100% synthetic grape must with 0.2% (w/v) Polygalacturonic acid (pH 3.5, Chemically Defined Medias with Polygalacturonic acid recipe 2.3.14). Fermentations were run at 22 °C, stirring at 250 rpm, until fermentation completion (0 g/L glucose/fructose detected) or stuck (minimal loss of sugar over 24 hours) with samples taken every 24 hours. 2 ug/ml of clonat was added

AWRI4241 and AWRI4240 cultures to ensure plasmid retention and added to ferments every 24 hours. Samples were used to monitor OD_{600}, glucose/fructose concentrations (2.8.3), galacturonic acid concentration (D-glucuronic/D-galacturonic Assay 2.8.4), unhydrolyzed polygalacturonic acid concentration (Ruthenium Red Assay 2.8.5).

Statistical analysis was performed on measurements of unhydrolyzed polygalacturonic acid concentrations taken over the course of fermentation using the GraphPad Prism 8 software. Each time point after the addition of *S. cerevisiae* to the ferment was analysed by Oneway ANOVA and multiple comparisons were performed using the Dunnett's test.

2.7.3 D-glucuronic Acid/D-galacturonic Assay
Galacturonic acid production was measured using the D-glucuronic Acid and D-galacturonic Acid Assay (Megazyme) using the microplate method (Megazyme). Concentrations of D-galacturonic acid were determined by comparison to galacturonic acid in synthetic grape must standard curve at concentrations 0 to 1.4 mg/ml in duplicate.

2.7.4 Ruthenium Red assay
Ruthenium red assays were used to test for endo-polygalacturonase activity, based on the original method by Ortiz et al. (2014) and the microplate method Torres et al. (2011). 80 ul aliquots were taken from synthetic grape must ferment samples and 200 ul of 1.125 mg/ml ruthenium red solution was added and vortexed for 30 s. 500 ul of 8 mM sodium hydroxide was then and to the solution, vortexed and centrifuged at 8000 rcf for 5 minutes (Eppendorf centrifuge 5415 D). 25 ul of the solution was added to a standard optical 96-well plate and diluted with 175 ul of water. Absorbance was read at 535 nm. Equivalent concentration of unhydrolyzed polygalacturonic acid was determined by comparison to polygalacturonic acid standard curve (concentrations: 0, 0.2, 0.4, 0.8, 1, 1.2, 1.4 and 2 mg/ml) in synthetic grape must. Results reported as mg/ml of hydrolysed polygalacturonic acid.

Chapter 3: Metagenomic assembly and analysis of a Chardonnay grape must (CGM) and a mixed varietal grape marc (MVGM)

Introduction

Metagenomics, the study of environmentally derived genetic material, has helped define the structure of complex microbial communities across a wide array of environmental sources. Metagenomics has revealed a vast number of microbial species that had previously been undiscovered, or at least inaccessible, due to an inability to culture them in a purified state (Schmeisser et al., 2007).

In winemaking, the alcoholic fermentation of grape juice is primarily driven by the yeast *Saccharomyces cerevisiae*. However, it is becoming increasingly apparent that other micro-organisms contribute to the fermentation process and shape the final sensory outcome of the wine (Pinto et al., 2014, Capozzi et al., 2015).

Grape marc is produced as a by-product of winemaking and contains the remnants of grape skins, seeds and stems. The microbes native to grape marc are not often considered for bioprospecting, however the presence of complex sugars and other substrates, as well as environmental factors such as pH and temperature, make grape marc a potential target of interesting microbial enzymatic activities (Favaro et al., 2013, Maragkoudakis et al., 2013, Ntougias et al., 2010). These enzymes could potentially be applied to winemaking and other applications such as biofuel (Dávila et al., 2017). Although grape marc has been used for bioprospecting (Favaro et al., 2013) and some exploration of bacterial metagenomics has been performed (Campanaro et al., 2014), there have been no studies using metagenomics based approaches to bioprospecting in grape marc previously.

This chapter describes the investigation of two grape derived metagenomes, an unfermented Chardonnay grape must (CGM) and a mixed varietal grape marc (MVGM) which had begun the composting process. This analysis considers the respective microbial communities of the two samples using metagenomic methods and the identification of genes of oenological interest.

Results and Discussion

Metagenomic profiling of a 2016 CGM and MVGM

Environmental DNA was extracted from the two samples. The first was an unfermented 2016 Chardonnay grape must (CGM), which was comprised of a mixture of grape juice, stems, seeds and skins that were in the process of being pressed. The second sample was a mixed varietal grape marc (MVGM), a winemaking by-product which contains all the remaining grape skins, seed and stems, which was being composted and had been turned once. Purified environmental DNA was then subjected to shotgun sequencing using short-read technologies. Final sequencing datasets were filtered for reads matching the grapevine genome which removed 39% and 0.15% of the reads from the CGM and MVGM metagenomes, respectively.

Taxonomic binning was performed on the filtered read sets to assess the microbial composition of the CGM and MVGM samples. Sequencing data was reference based binned against the NCBI protein database. At the kingdom level, there was a clear difference in the proportion of organisms populating the two environments (Figure 3.1). The CGM metagenome comprised a high portion of Fungi as well as a significant portion of *Viridiplantae*, which represents additional *V. vitis* sequences that escaped the bioinformatic filtering procedure (Figure 3.1). As these grapevine classifications are occurring after the reads were filtered against the Pinot noir PN00024 genome assembly these sequences could potentially be due to variation between the two varieties. Regardless, the CGM dataset largely conforms to previously studied metagenomes (Salvetti et al., 2016, Bokulich et al., 2014, Gayevskiy and Goddard, 2012). The metazoan reads present in the CGM metagenome consist primarily of matches to insects and spiders which are likely being harvested along with the grapes.

As opposed to the fungal-dominated CGM data, the MVGM dataset is mostly comprised of reads assigned to bacteria and has minimal remaining grapevine DNA. After separation, the grape marc is left in contact with the soil, likely facilitating colonisation of the MVGM by soil-derived microbiota, resulting in the high bacterial content of the MVGM metagenome. Additional compositional bias in the extracted DNA may have also been introduced through

the use of different protocols to extract the metagenomic DNA. The MVGM DNA was extracted with a larger quantity of sample but with gentler physical lysis than the CGM material. As such, organisms that are more difficult to lyse, such as filamentous fungi, may be under-represented in the MVGM sample. The reverse is true of the CGM sample. The CGM DNA were extracted using a harsher bead beating lysis and may therefore have lower quality DNA from cells lysed early in the process which may be underrepresented during sequencing.
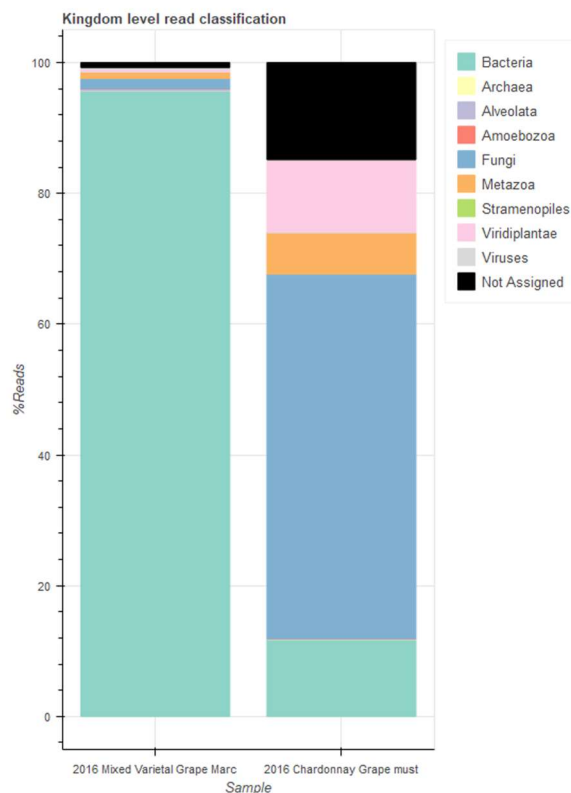

Figure 3.1: Taxonomic classification of DNA sequencing reads at the kingdom level of the 2016 Chardonnay grape must and 2016 mixed varietal grape marc metagenomic samples.

Both datasets have reads that could not be classified at any level. These unclassified reads can be attributed to gaps in the NCBI reference database and reads that are poorly identifiable due to sequence structures including repetitive elements and low-complexity sequences. Gaps in the NCBI reference database may also be causing misidentification, especially at genus and species level and in genomic regions that are highly conserved across species. Misidentification could be improved using more stringent parameters for the classification algorithms and more recent algorithms, but the resolution was sufficient for

subsequent analyses. However, some observations were taken at the genus level in order to avoid potential complications of species-level resolution (Table 3.1). The CGM metagenome had 56 genera that comprised at least 0.05% of reads. Many of the genera identified from the CGM data have been previously shown to be associated with grapevine. *Aureobasidium,* was the most common genus in the CGM dataset, with 0.85% of reads classified in the CGM metagenomic data being assigned to this genus. *V. vinifera* is known to host *Aureobasidium pullulans* (Belda et al., 2016a, Pinto et al., 2014) and *A. pullulans* is widespread across many environments, so it is logical for it to be present in this dataset. The majority of the remaining genera have been commonly identified from plants and grapes (Quaedvlieg et al., 2011, Piao et al., 2015, Barata et al., 2012, Puri et al., 2018), although *Mesorhizobium* and *Streptomyces* are more commonly found in vineyard soil than the grapevine itself (Martins et al., 2013).

126 genera were identified from the MVGM dataset with at least 0.05% of reads assigned to each taxon. The most abundant genera in the MVGM dataset were primarily comprised of soil and plant associated bacteria (Santos et al., 2007, Zarraonaindia et al., 2015). This is expected given the sample environment and contact with soil. Although no fungi were present in the top 20 most abundant read classifications, fungal genera were observed at lower abundances in the MVGM dataset. These fungi included many of the genera that were also identified in the Chardonnay metagenome, including *Fusarium* 0.20%, *Aspergillus* 0.11%, *Colletotrichum* 0.03%*, Penicillium* 0.02%, *Leptosphaeria 0.02%*, *Botrytis* 0.01%, *Bipolaris* 0.01% and *Saccharomyces* 0.01%.

Table 3.1: Top 20 most abundant bacterial and fungal genera in the CGM and MVGM metagenomic datasets.

CGM		MVGM	
Genus	% reads	Genus	% reads
Aureobasidium	0.85	*Pseudomonas*	1.70
Torulaspora	0.25	*Streptomyces*	1.43
Alternaria	0.21	*Stenotrophomonas*	0.71
Aspergillus	0.15	*Mycobacterium*	0.64
Cladosporium	0.15	*Burkholderia*	0.64
Saccharomyces	0.14	*Mesorhizobium*	0.55
Leptosphaeria	0.14	*Pseudoxanthomonas*	0.53
Botrytis	0.13	*Xanthomonas*	0.53
Paraburkholderia	0.13	*Rhodococcus*	0.51
Pseudomonas	0.10	*Bradyrhizobium*	0.39
Colletotrichum	0.10	*Micromonospora*	0.38
Fusarium	0.10	*Enterobacter*	0.38
Streptomyces	0.10	*Citrobacter*	0.36
Zymoseptoria	0.09	*Rhizobium*	0.35
Penicillium	0.08	*Acidovorax*	0.32
Mesorhizobium	0.08	*Flavobacterium*	0.29
Bipolaris	0.08	*Lysobacter*	0.28
Bradyrhizobium	0.07	*Myxococcus*	0.28
Burkholderia	0.07	*Sinorhizobium*	0.27
Verticillium	0.06	*Methylobacterium*	0.25

De novo assembly of CGM and MVGM metagenomes

The short sequencing reads (100-150 bp) produced by most next generation sequencing platforms do not span the length of full protein encoding genes and require assembly to produce contigs of sufficient length for full-length gene predictions. Metagenomic assembly is typically performed by *de novo* methods as reference-based approaches are generally not feasible as reference genomes for each member of the community are required. Prior to metagenomic assembly the datasets were filtered to remove the grape vine (*Vitis vinifera*) sequence data. This grapevine sequence data constituted a signification portion of the CGM dataset but was minimal in the MVGM dataset. The CGM and MVGM data were then assembled using the Megahit assembler (Table 3.2).

As is commonly observed in metagenomic assemblies, the CGM and MVGM metagenomes both contained a large number of contigs that span a wide range of assembled lengths (<1 kbp to 2 Mbp). The fragmented assemblies that are observed for metagenomes are a consequence of both a large effective genome size (i.e. the sum of all the organisms in the dataset), which leads to low average coverage and the variable abundances of individual species, which leads to both uneven coverage and very low coverage of many species. Short reads also prevent the resolution of repeat sequences, but this is common to all short-read assemblies not just those performed on metagenomes.

Table 3.2: Summary of CGM and MVGM DNA sequencing datasets and assembly.

	CGM	MVGM
Data type	Illumina HiSeq, 2x250	Illumina HiSeq, 2x150
# Paired Reads	134,259,496	454,567,563
% Reads aligned to *V. vinifera*	39%	0.15%
Assembler	Megahit v1.0.3	Megahit v1.1.1-2
N_{50} (bp)	1,111	3,395
# Contigs	5,697,851	4,260,285
# Contigs ≥ 1000 bp	421,837	973,312
Largest Contig (Mbp)	0.177	2.166
Total Length (Gbp)	3.033	5.097

The CGM metagenome had a lower N_{50} than the MVGM metagenome, a larger number of contigs and a shorter maximum contig length. These differences are partially due to coverage, as the MVGM had more data overall, which leads to better genome coverage and resolution. For example, from the CGM dataset, *Aureobasidium* is the most abundant genera identified, with approximately 0.85% of the Chardonnay reads classified to this genus. If all of these reads belonged to the species *A. pullulans* (29 Mbp genome), the average genome coverage would be approximately 12-fold. In comparison in the MVGM dataset, *Pseudonomas* (genome sizes 5 - 7 Mbp), represents 1.4% of the classified reads, leading to a genome coverage of approximately 268-fold. This difference is primarily a consequence of overall sequencing depth, although slightly inflated by the difference in genome sizes, the difference in percentage of reads classified and the assumption that these reads all belong to a single species. As these examples are of the most abundant genera of the CGM and MVGM datasets all other organisms in the genome can be expected to have lower coverage.

Oxford Nanopore Technology (ONT) long-read data was obtained for the MVGM dataset and was assembled in combination with the MVGM short-read assembly (Table 3.3). The MVGM long-read dataset was first assembled as a separate assembly and then assembled in combination with the MVGM short-read assembly, with small contigs removed, as if the assembled short-read data where its own long-read sequencing run. As the smaller contigs were removed from the MVGM, the overall length of the MVGM short-read assembly was reduced and the MVGM combined assembly also has decreased overall length. Although some of the improvement is due to the removal of short contigs, the MVGM combined assembly further improved contiguity, increasing the N_{50} to 24,477bp, and doubling the size of the assembly's largest contig. The increased N_{50} is partially due to the increased genome coverage but also scaffolding provided by the contigs produced from the long-read data. As error rates are lower the short-read data, it is expected that the quality of the ONT assembly may also have improved. The long-read data could have been used for gene prediction

without assembly as the contigs should span whole protein coding regions, but the higher error rates of long-read data are undesirable in gene predictions.

Table 3.3: Summary of MVGM assembly with Oxford Nanopore Technology MinION long-read data and MVGM short-read assembly.

	MVGM ONT assembly	MVGM combined assembly
Data type	Oxford Nanopore Technologies (ONT) MinION Data	MVGM ONT and Illumina HiSeq Assemblies
# Reads	1,162,438	-
Assembler	Meta-Flye v 2.5	Meta-Flye v 2.5
N_{50}	39,988	24,477
# Contigs	7,287	121,542
# Contig ≥ 1000bp	7,280	121,537
Largest Contig (Mbp)	6.764	4.915
Total Length (Gbp)	0.284	2.004

Binning of assembled contigs

Although some contigs from the CGM and MVGM metagenomic assemblies were less than 1 kbp, large contigs (e.g. 4.9 Mbp) were also produced and could potentially represent complete or near complete genomes (Tables 3.2 and 3.3). In order to extract complete genomes from the metagenomic assemblies of the CGM and MVGM samples, sequences were separated into bins based on tetranucleotide frequencies and abundance. Bin quality was filtered for a minimum of 90% genome completeness (percentage of core marker genes) and less than 10% contamination by Checkm. Putative taxonomic assignment of 'complete' bins was also performed by Checkm.

The contigs of the MVGM metagenome separated well by standard non-reference binning parameters, such as GC-content, tetranucleotide frequency, abundance and coverage, with bins becoming visually apparent on at contig thresholds ≥25 kbp (Figure 3.2). At the minimum contig size threshold of ≥25 kbp, there is also clearer separation of bacterial and

eukaryotic contigs and most unassigned contigs are removed. Binning was tested at multiple contig length thresholds. However, despite ≥25 kbp threshold producing more visually apparent bins, a lower threshold of ≥10 kbp produced the more bins which passed quality filtering from the MVGM dataset. 32 bins passed filtering at minimum contig size of ≥25 kbp, whereas 54 of 370 potential bins passed quality filtering at a minimum contig size of ≥10 kbp (Table 3.4). Of the 54 MVGM passing bins, no eukaryotic bins were obtained, and few taxonomic assignments were made below the family level. Taxonomic assignment of bins was made by Checkm based on taxonomy of marker genes. *Thermobifida fusca* is the only species level classification given to the MVGM bins but the genera *Bdellovibrio, Pedobacter, Asticcacaulis, Legionella, Planctomyces* were also observed. None of these genera appeared in the most abundant read classifications for the MVGM data. Further classification of these bins is required and reassembly of bin data may improve genome contiguity.
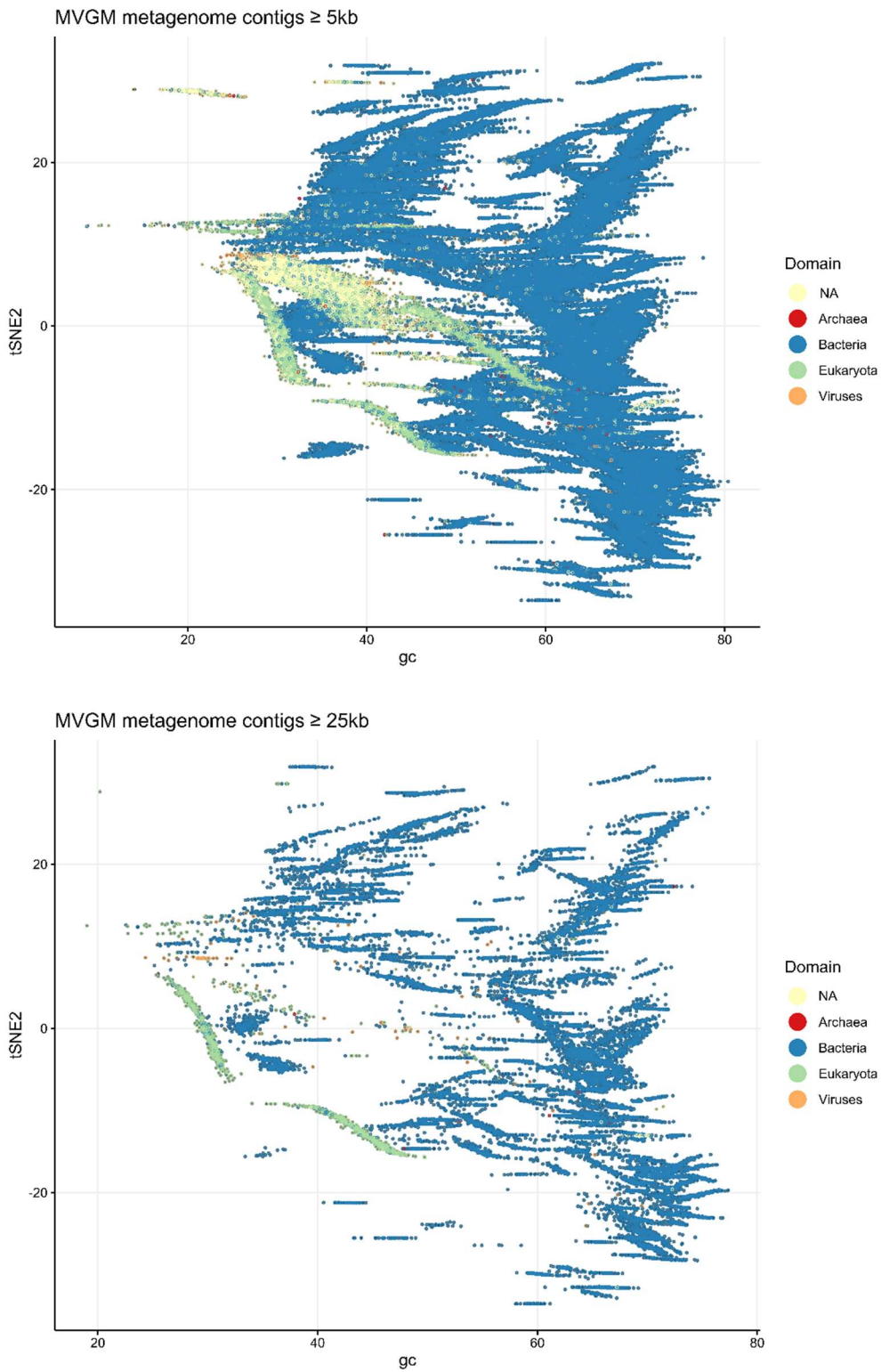
Figure 3.2: Separation of MVGM contigs by GC-content ('gc') and Barnes-Hut t-Distributed Stochastic Neighbour Embedding representations of tetranucleotide frequency ('tSNE2') at contig length thresholds 5 kbp and 25 kbp. NA = not assigned

Table 3.4 Bins produced from MVGM metagenome contigs with more than 90% of

essential genes and less than 10% contamination by other genomes. CM = completeness.

CT = Contamination

ID	Putative Taxonomy	CM	CT	ID	Putative Taxonomy	CM	CT
Bin 1	f__Planctomycetaceae	92.95	3.41	**Bin 28**	g__Legionella	90.2	2.19
Bin 2	f__Verrucomicrobiaceae	97.28	0	**Bin 29**	f__Flammeovirgaceae	98.6	2.44
Bin 3	f__Verrucomicrobiaceae	95.24	0.91	**Bin 30**	o__Burkholderiales	93.62	8.02
Bin 4	p__Chloroflexi	90.58	1.01	**Bin 31**	f__Chloroflexaceae	90.25	8.96
Bin 5	f__Planctomycetaceae	98.85	0	**Bin 32**	f__Flammeovirgaceae	92.92	0.3
Bin 6	f__Verrucomicrobiaceae	94.56	2.04	**Bin 33**	o__Actinomycetales	98.75	0
Bin 7	f__Myxococcaceae	97.11	0.7	**Bin 34**	o__Myxococcales	92.26	2.26
Bin 8	f__Sphingobacteriaceae	94.76	1.31	**Bin 35**	o__Cytophagales	93.44	0.3
Bin 9	f__Cryomorphaceae	95.38	1.89	**Bin 36**	o__Sphingobacteriales	98.52	0.99
Bin 10	f__Cytophagaceae	96.43	1.44	**Bin 37**	f__Rhizobiaceae	92.6	2.52
Bin 11	g__Bdellovibrio	99.1	3.55	**Bin 38**	g__Planctomyces	97.78	0
Bin 12	g__Pedobacter	93.62	5.26	**Bin 39**	o__Actinomycetales	93.64	3.67
Bin 13	f__Chitinophagaceae	90.84	0.99	**Bin 40**	c__Actinobacteria	91.64	2.14
Bin 14	p__Proteobacteria	98.16	1.34	**Bin 41**	c__Gammaproteobacteria	95.97	1
Bin 15	f__Deinococcaceae	92.16	0.85	**Bin 42**	o__Myxococcales	90.95	3.89
Bin 16	o__Sphingobacteriales	91.87	6.26	**Bin 43**	f__Myxococcaceae	93.04	4.41
Bin 17	f__Chitinophagaceae	95.57	1.48	**Bin 44**	f__Alcaligenaceae	92.02	2.76
Bin 18	f__Cyclobacteriaceae	91.55	0.79	**Bin 45**	f__Moraxellaceae	98.4	4.81
Bin 19	f__Microbacteriaceae	96.69	0	**Bin 46**	p__Chloroflexi	98.18	1.09
Bin 20	f__Chitinophagaceae	97.12	1.43	**Bin 47**	f__Oceanospirillaceae	95.13	2.53
Bin 21	k__Bacteria	97.8	2.2	**Bin 48**	o__Sphingomonadales	91.21	1
Bin 22	o__Rhizobiales	92.24	1.09	**Bin 49**	g__Planctomyces	97.78	4.17
Bin 23	g__Asticcacaulis	96	2.54	**Bin 50**	p__Bacteroidetes	90.17	2.38
Bin 24	f__Cytophagaceae	96.88	0.89	**Bin 51**	f__Chloroflexaceae	98.55	3.49
Bin 25	s__Thermobifida_fusca	97.72	1.62	**Bin 52**	f__Verrucomicrobiaceae	93.43	2.21
Bin 26	o__Verrucomicrobiales	97.3	2.7	**Bin 53**	o__Flavobacteriales	97.31	2.42
Bin 27	f__Planctomycetaceae	95.45	1.7	**Bin 54**	p__Bacteroidetes	95.48	2.32

Only 17 potential bins could be identified from the CGM metagenome by automated means and only 1 had both >90% completeness and <10% contamination. This poor result was caused by overall poor separation of the CGM contigs by tetranucleotide frequency and abundance. Similarly, the CGM contigs also poorly separate by GC-content and average sequencing coverage (Figure 3.3). At 10 kbp, clusters of contigs start to appear, including isolated clusters of eukaryotic sequences. Using manual methods to cluster the CGM contigs according to GC-content, tetranucleotide frequency and coverage, three bins (including an equivalent bin to the automated methods) were isolated from the CGM metagenome (Table 3.5). Additional bins that did not satisfy bin quality parameters were produced but these had high contamination which could only be reduced at the cost of marker gene completeness. This inverse relationship is caused by the limited number of large contigs and the poor overall separation of sequences. Despite the high proportions of fungal reads identified by reference-based classification of the CGM data and the appearance of at least three eukaryotic clusters at separation of contigs ≥ 10 kbp (Figure 3.3), no quality eukaryotic bins were obtained. Bin 3 was classified as *Mesorhizobium loti*, which corresponds well with *Mesorhizobium* as the 16th most abundant genus of read taxonomic classification of the CGM data. Similarly, bin 2 was identified as *Burkholderia* and the genus was present the top 20 most abundant CGM reads. Bin 1 was classified into the family *Xanthomonadaceae.* Four genera (*Strenotophomona*, *Pseudoxanthomonas*, *Xanthomonas* and *Lysobacter*) in the top 20 most abundant CGM reads belong to the *Xanthomonadaceae* family and bin 1 may be one of these. Further classification by predicting 16S sequence from the bin contigs may further elucidate taxon.
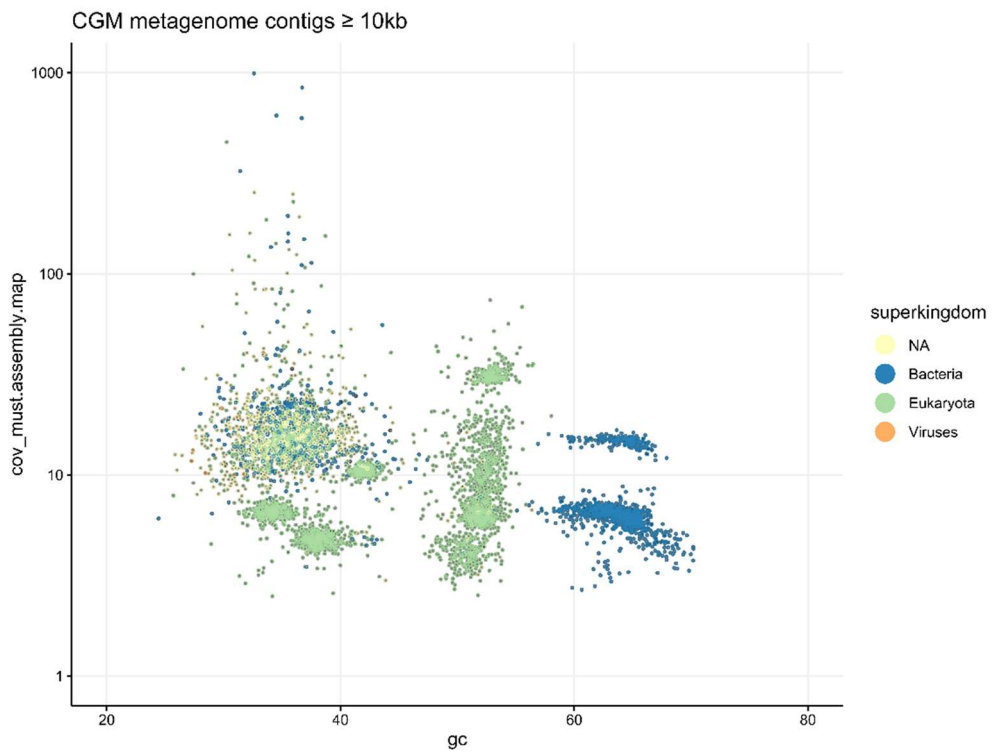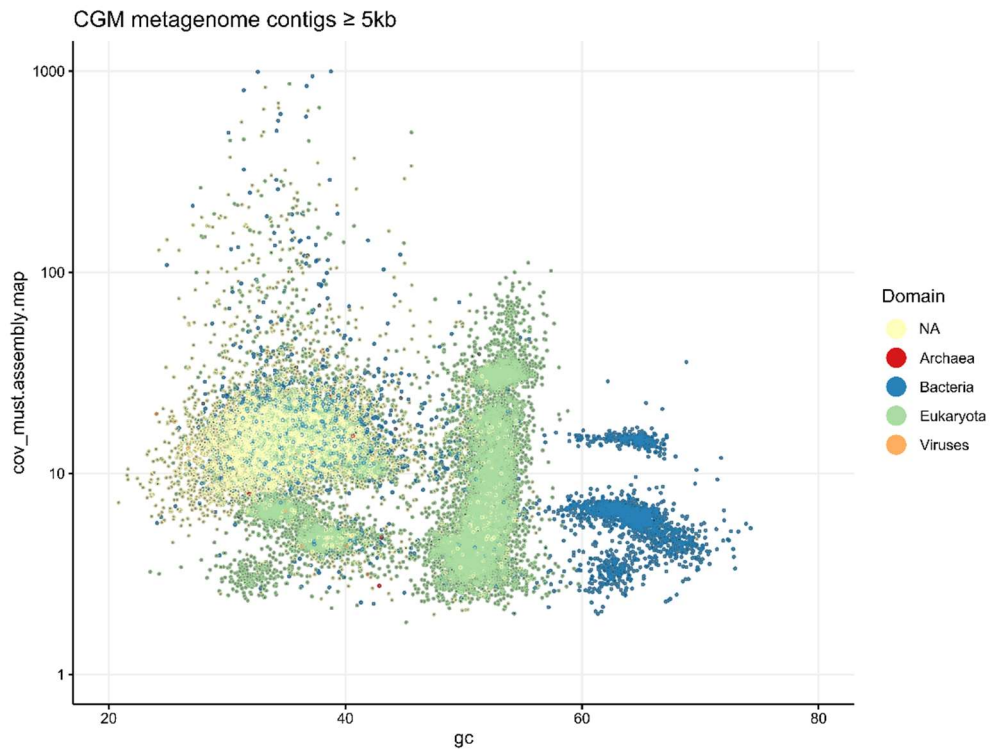
Figure 3.3: Separation of CGM metagenome contigs by GC-content ('gc') and coverage ('cov_must.assembly.map') at contig length threshold 5 kbp and 10 kbp. NA = not assigned

Table 3.5: Bins produced from CGM metagenome contigs with more than 90% of essential genes and less than 10% contamination by other genomes.

ID	Putative Taxonomy	Completeness	Contamination
Bin 1	f__Xanthomonadaceae	92.1	2.11
Bin 2	g__Burkholderia	100	9
Bin 3	s__Mesorhizobium_loti	92.72	4.48

Gene prediction of CGM and MVGM Metagenomes

In order to assess the potential enzyme repertoire of the CGM and MVGM metagenomic assemblies, gene prediction of the assembled contigs was performed, with functional annotation then performed on these predicted genes. *Ab Initio* gene prediction was performed on contigs of a minimum length 1 kbp for the CGM assembly using the *Candida albicans* gene model and a minimum length of 10 kbp with the *Aspergillus fumigatus* gene models for the MVGM (Table 3.6). This produced 984,472 gene predictions from the CGM metagenome, and 535,583 gene predictions form the MVGM metagenome. The minimum threshold was increased for the MVGM dataset as a 16% of gene predictions from the CGM gene predictions appeared to be truncated (<50 amino acids), which was undesirable. The two fungal models were therefore chosen to increase sensitivity for fungal genes which are of greater interest for future applications, *C. albicans* as a yeast model, and *A. fumigatus* as a filamentous fungus. However, in most cases the *A. fumigatus* gene models seem to be robust, with 75% of the MVGM gene predictions having a simple structure with a single ORF as would be more expected of bacterial genes. The remaining 25% would need to be further screened to ensure accuracy and testing with a bacterial model would be appropriate.

Putative functions were assigned to the gene predictions by InterProScan and BLASTp alignment to the NCBI non-redundant protein database. The gene predictions were assessed for matches to proteins in the NCBI NR protein database and gene predictions filtered for matches >90% and >80% amino acid identity (Table 3.6). 27% of the genes predicted from the CGM metagenome and 36% of the MVGM metagenome gene

predictions had matches >80% amino acid identity to the NCBI NR protein database. The presence of significant homologous matches to known proteins indicates that the metagenome assembly and gene prediction strategy was producing relevant gene predictions. Interestingly, a larger proportion of the MVGM gene predictions have putative signal peptides than in the CGM dataset when using the eukaryotic model for signal peptide prediction. This may be due to the influence of *A. fumigatus* gene models used for prediction of the MVGM dataset.

InterProScan was able to assign at least one match to its reference databases to 73% of all the CGM gene predictions and 95% of the MVGM predictions (Table 3.6). The remainder gene predictions could not be functionally annotated by InterProScan. These annotations were then mined for genes of interest.

Table 3.6: Summary of gene predictions for the CGM and 2016 MVGM metagenomes.

	2016 Chardonnay Grape must	**2016 Mixed Varietal Grape Marc**
Gene Model	*Candida albicans*	*Aspergillus fumigatus*
Total number of gene predictions (GPs)	984,472	535,583
Total number of GPs <50 amino acids	159,997(16%)	0 (0%)
GPs with >90% identity to NCBI NR protein database	177,548 (18%)	94,835 (17%)
GPs with >80% identity to NCBI NR protein database	264,456 (27%)	195,201 (36%)
GPs with putative signal peptide	16,163 (1.6%)	63, 674 (11%)
GPs with InterProScan matches	722,153 (73%)	506,586 (95%)

Functional assignment to gene predictions

Glycoside hydrolases (GHs), which mediate the hydrolysis of glycosidic compounds, have been a target of significant oenological interest. This is due to the prevalence of glycosides with significant oenological impact being present in the grape matrix, which poses both processing challenges and avenues for organoleptic manipulation during winemaking. However, there is also an increased interest in other categories of carbohydrate active enzymes and other non-carbohydrate acting enzymes, such as proteases for haze prevention and remediation. InterProScan was used to detect putative members of various carbohydrate active and hydrolytic enzyme groups from the CGM and MVGM gene prediction datasets. Results from both datasets were subsequently filtered for matches (>80% identity of the minimum alignment size) against the NCBI non redundant protein database, in order to reduce the overall number of matches, while enriching the final dataset for a greater proportion of novel enzyme sequences and presumably, novel enzyme properties.

In the CGM dataset, 6,050 gene predictions matching GH patterns were identified, approximately half of which had no significant matches (i.e. <80% amino acid identity) to proteins in the NCBI non redundant protein database (Table 3.7). The MVGM dataset contained 5,096 putative GHs, although only 33% with <80% amino acid identity to a protein in the NCBI protein database.

Polysaccharide lyases and lytic polysaccharide mono-oxygenases (LPMOs) are also involved in the degradation of carbohydrates. 88% the putative polysaccharide lyases identified in the MVGM data set had significant matches to existing NCBI proteins. However, most of the polysaccharide lyases from CGM data set (18%) did not have a match in the NCBI database (Table 3.7). A total of 30 (CGM) and 13 (MVGM) LPMOs were also identified, with 23 and 13 without matches in the NCBI protein database, respectively (Table 3.7).

A total of 12,685 putative proteases were identified from the two datasets. Of these 51% and 58% have matches to the NCBI protein database with <80% amino acid identity, for the CGM and MVGM proteins, respectively (Table 3.7).

Table 3.7: Carbohydrate active enzymes and proteases predicted in the2016 Chardonnay metagenome and 2016 MVGM metagenome. Genes predicted on contigs equal to or longer the 10 kbp.

Enzyme Family		CGM		MVGM	
		<80% identity	Total	<80% identity	Total
Glycoside	Family 1	110 (47%)	232	114 (55%)	207
Hydrolases	Family 3	448 (49%)	905	313 (53%)	585
	Family 28	49 (31%)	158	87 (72%)	120
	Total	3,061 (50%)	6050	1,658 (33%)	5096
Polysaccharide	Family 1	0	0	5 (100%)	5
lysates	Family 4	9 (16%)	57	9 (100%)	9
	Total	11 (18%)	61	37 (88%)	42
Lytic polysaccharide mono-oxygenases		23 (77%)	30	13 (27%)	48
Proteases	Aspartic	657 (49%)	1336	390 (66%)	595
	Total	1,967 (51%)	3850	5,105 (58%)	8835

β-glucosidases

β-glucosidases (EC 3.2.1.21) hydrolyse β-glucoside bonds between a glucose and a second molecule. While the sugar moiety causes the glycoside compound to be non-volatile, the secondary molecule can often be volatile as an aglycone. Liberation of volatile aglycones by β-glucosidases can therefore influence the organoleptic properties of fermented foods and beverages and are of particular interest for application in winemaking. Glycoside hydrolase families (GHFs) are separated by amino acid structure (Henrissat, 1991, Henrissat and Bairoch, 1993, Davies and Henrissat, 1995, Henrissat et al., 1995), as many of glycoside hydrolases are active on multiple substrates and cannot be effectively classified by substrate. There are currently 166 GHFs, although there are some protein sequences which have not yet been classified by this system. β-glucosidases are most commonly ascribed to either GHF1 and GHF3 but have also been identified in GHFs 2, 5, 9, 16, 30, 39 and 116. GHFs 1, 2, 5, 30 and 39 are also grouped in a glycoside hydrolase clan, GH-A, due to a shared $(\beta / \alpha)_8$ structure (Henrissat et al., 1995) and may share similar properties due to this similarity. In comparison, GHF3 proteins contain a $(\alpha/\beta)_8$ barrel domain connected via a linker region to a $(\alpha/\beta)_6$ sandwich domain (Harvey et al., 2000).

A total of 985 of the CGM and MCGM gene predictions were matched to either GHF1 or GHF3, with 47% of GHF1 and 55% of GHF3 gene predictions having <80% identity to any protein the NCBI protein database (Table 3.7). To compare the relatedness of the predicted metagenomic β-glucosidases from GHF1 and GHF3, outside of their conserved GHF structures, the amino acid sequences of the two families from both metagenomic datasets were aligned and clustered. The GHF1 sequences clustered in three main groups (Figure 3.4). Clusters A and B contained mostly MVGM sequences with only a small number of the CGM gene models. Cluster C displays higher primary amino acid sequence variation and contains a relatively even distribution of sequences from the MVGM and CGM gene models.

When the clusters are examined relative to the taxonomic assignment of each gene model, sequences from clusters A and B are comprised mainly of bacterial proteins, while cluster C has a blend of taxonomic assignments (Figure 3.5). Clustering of the sequences of bacterial origin makes sense as these sequences should be more evolutionary close to

each other than the eukaryotes. Similarly, the non-bacterial members of Cluster A are more divergent to the bacterial members. Cluster C is more complex with a blend of kingdom level taxonomic assignments, which do not separate evenly by evolutionary distance. This could be a sign of convergent evolution, horizontal gene transfer or due to the presence of different enzyme activities (and substrate preferences) in the GHF, such as β-xylanases and β-galactosidases instead of β-glucosidases, which could be expected to have more similar structures.

Compared to the GHF1 gene models, the GHF3 sequences were both more numerous and more divergent (Figure 3.6), with individual clusters diverging earlier. As seen with the GHF1 data, some GHF3 clusters were dominated by sequences from one dataset, while some contained a mixture of the two. The amino acid sequence variation of GHF3 is in part a function of the number of sequences.
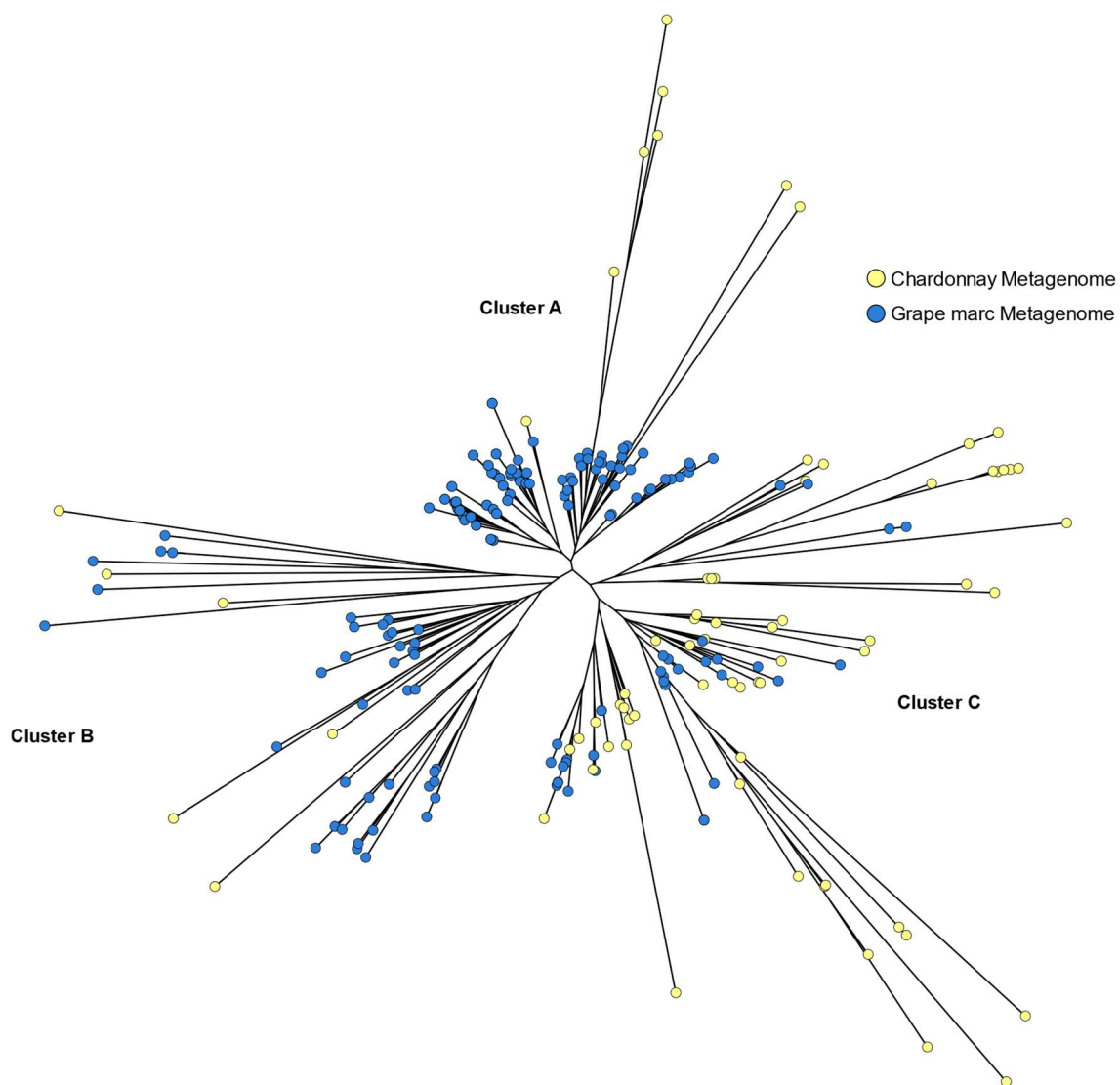
Figure 3.4: Clustering of gene predictions from the CGM and MVGM metagenomes matching GHF1 by amino acid sequence, following filtering to remove proteins with >80% identity to any protein in the NCBI nr protein database. Gene predictions are coloured by origin.
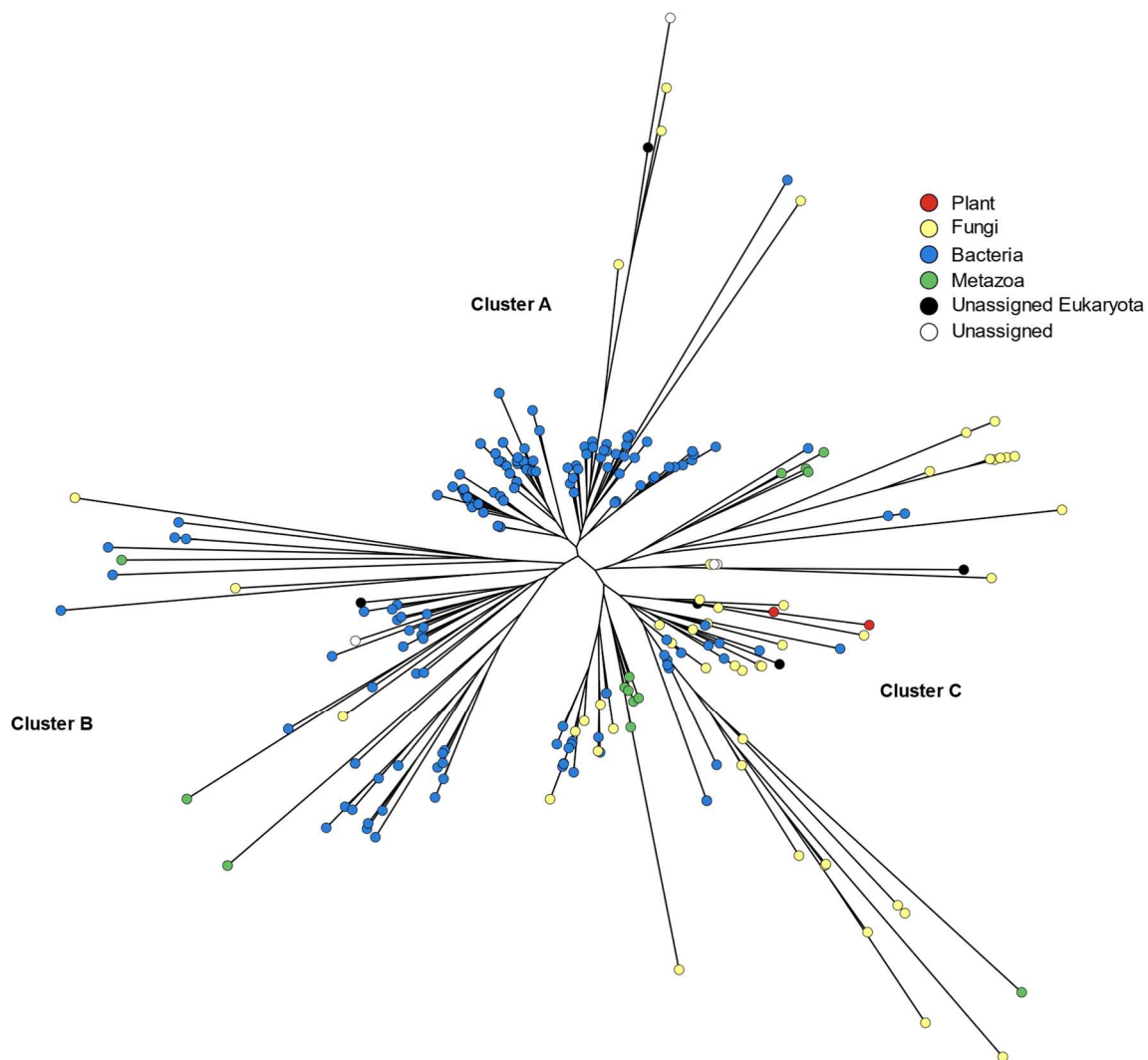
Figure 3.5: Taxonomic assignment of GHF1 gene predictions (filtered for identity to NCBI proteins) from the Chardonnay must and MVGM metagenomes, clustered by amino acid sequence.
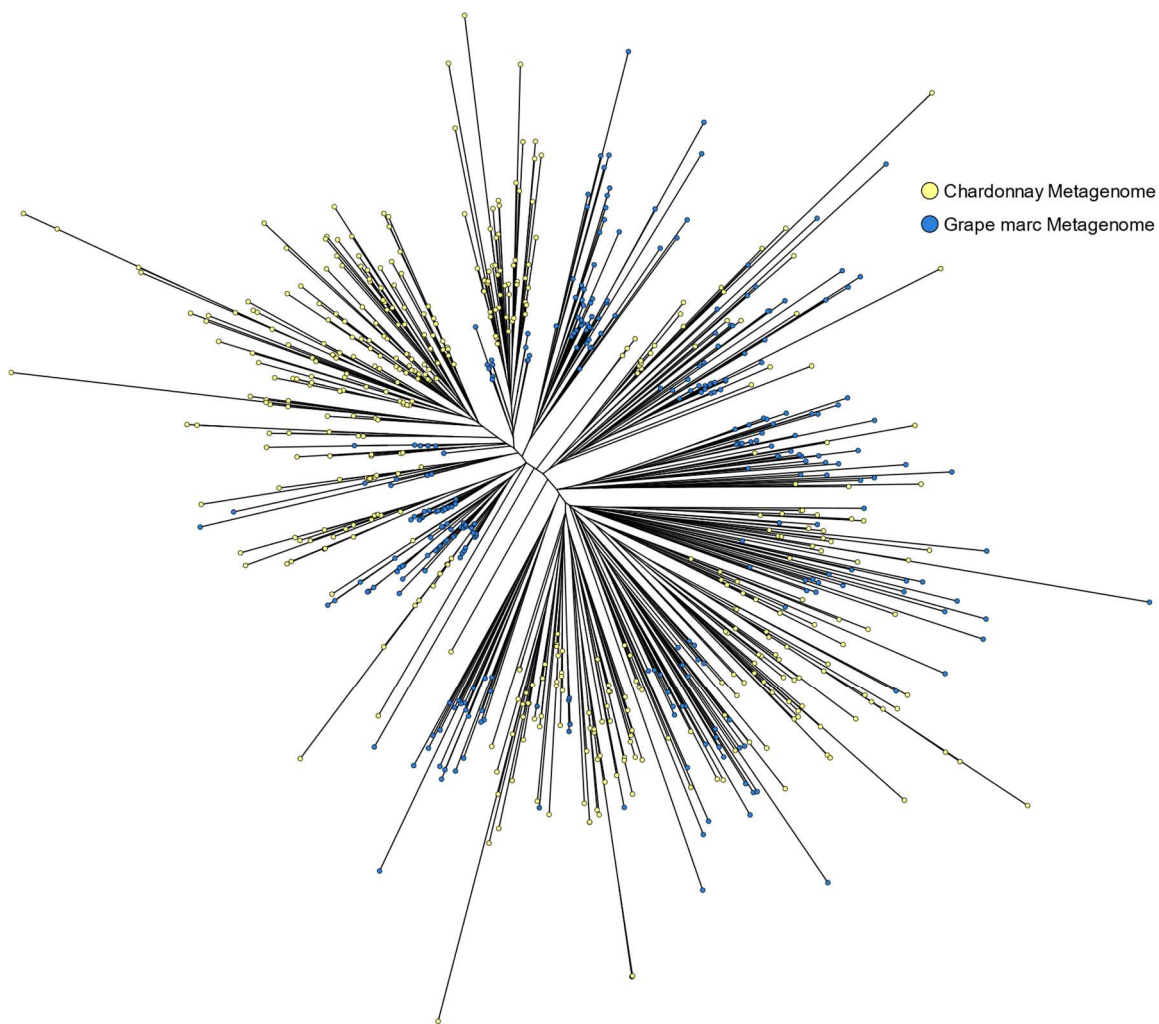
Figure 3.6: Clustering of gene predictions from the CGM and MVGM metagenomes matching GHF3 by amino acid sequence. Gene predictions were filtered to ensure <80% identity to any protein in the NCBI non-redundant protein database. Gene predictions coloured by origin.

Pectinases

Pectinases are a group of enzymes that are collectively responsible for the complete degradation of pectin, a major component of plant cell walls. The main structural unit of pectin is polygalacturonic acid, which itself is comprised of linked monomers of galacturonic acid. Polygalacturonic acid is hydrolysed by polygalacturonases (EC 3.2.1.15, EC 3.2.1.67), which reside solely with GHF28. GHF28 proteins have a (β)-helix structure that is common to proteins that act on pectin (Petersen et al., 1997, Pickersgill et al., 1998). Pectin can contain other non-galacturonic acid substituents and have corresponding glycoside hydrolases which act on these substrates. Many of these GHs, such as rhamnosidases, rhamnogalacturonases and xylogalacturonan hydrolases, are also found in GHF28.

A total of 136 gene models from the combined CGM and MVGM metagenomes are predicted to be members of GHF28 (after filtering for identity to NCBI proteins). Gene models from this family were compared by multiple sequence alignment and clustering (Figure 3.7). Almost all the sequences clustered separately by metagenome (Figure 3.7) and by taxonomic classification (Figure 3.8), with this overlap primarily driven by the different taxonomic composition of the CGM (fungal dominated) and MVGM (bacterial dominated) datasets. There are 12 sequences that do not cluster by taxa or by metagenomic origin. As mentioned with regards to clustering of the GHF1 sequences, these sequences could again be a sign of convergent evolution, horizontal gene transfer or a mixture of enzyme activities within a cluster.

In addition to GHF28 polygalacturonases, other pectin-active enzyme groups were identified. Pectin lyases terminally cleave polygalacturonic acid to produce galacturonic acid moieties and typically belong to polysaccharide lyase family (PLF) 1. No gene predictions from the CGM dataset were assigned to PLF1 (Table 3.7). However, five were identified in the MVGM dataset, which all lacked significant matches (>80% identity) to the NCBI protein database. Rhamnogalacturonan endolyase has related activity on the rhamnogalacturonan regions of pectin and belong to PLF4. A total of 57 CGM gene models were assigned to PLF4, but only nine lacked significant matches to the NCBI nr database. All nine of the MVGM gene models that were assigned to PLF4 also lacked significant matches to NCBI.
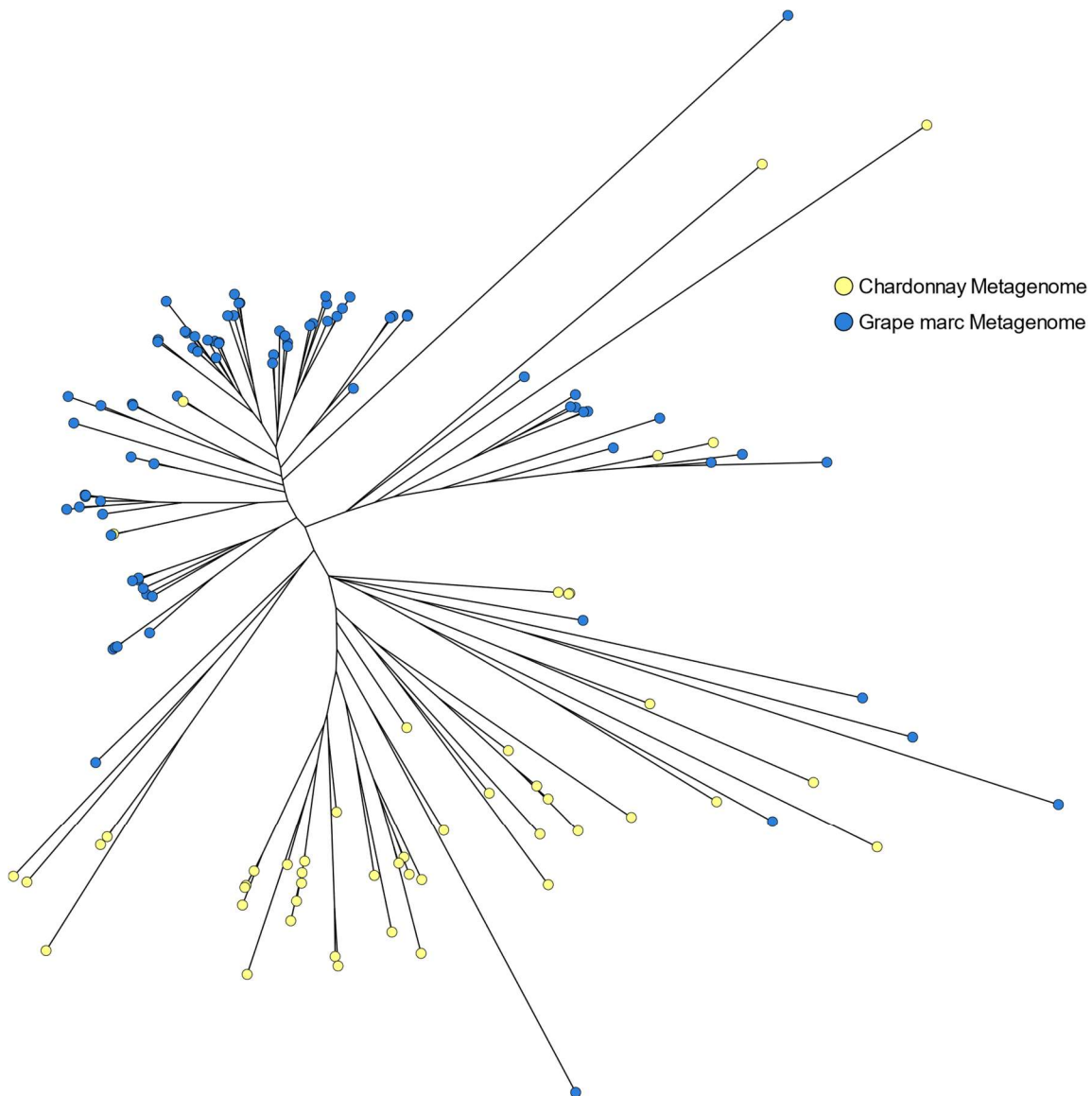
Figure 3.7: Clustering of gene predictions from the CGM and MVGM metagenomes matching glycoside hydrolase family 28 with <80% identity to any protein in the NCBI non-redundant protein database by amino acid sequence. Gene predictions coloured by origin.
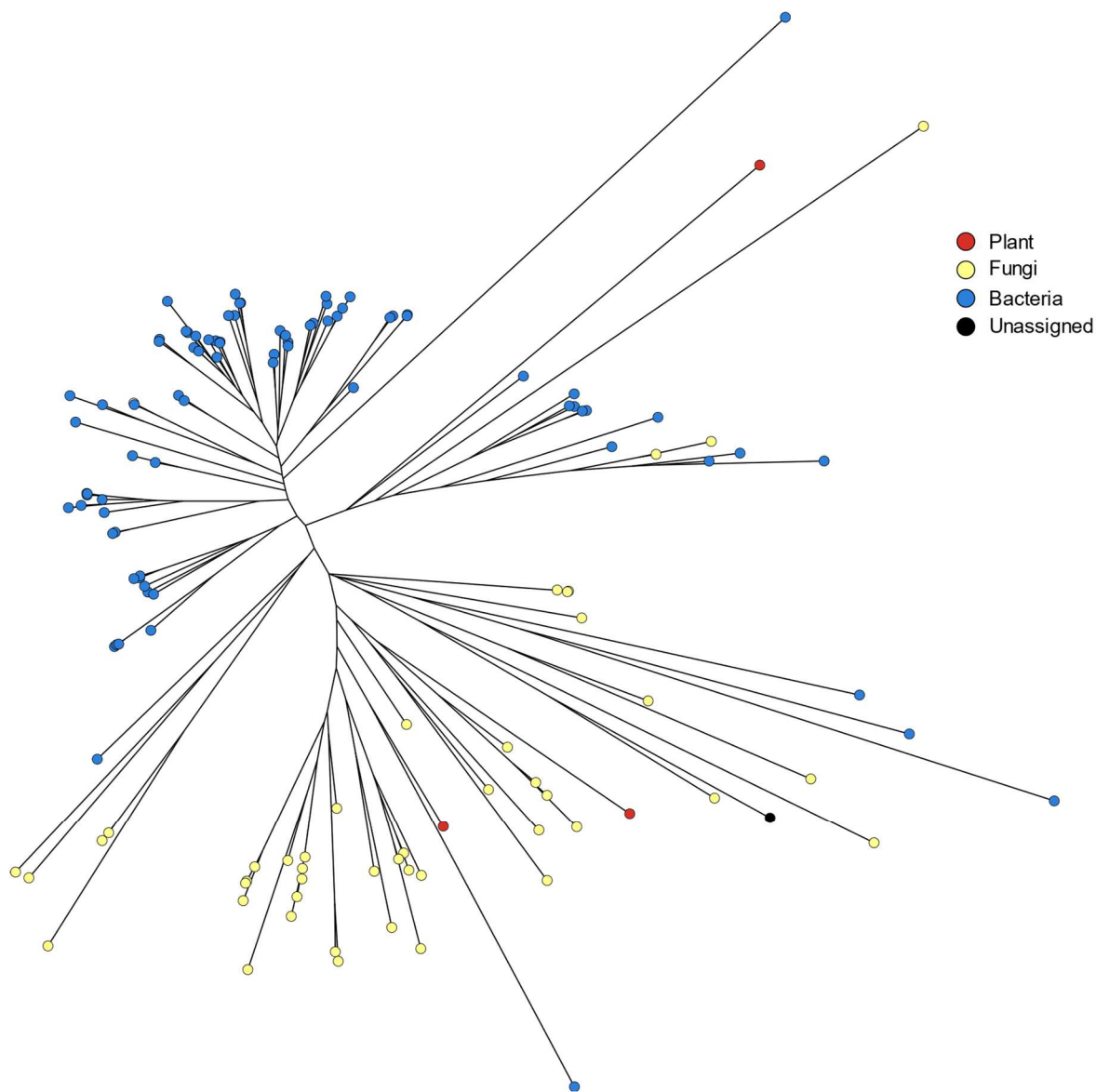
Figure 3.8: Clustering of gene predictions from the CGM and MVGM metagenomes matching glycoside hydrolase family 28 and <80% identity to any protein in the NCBI non-redundant protein database by amino acid sequence. Gene predictions coloured by putative taxonomic classification.

Conclusions

In order to investigate functional elements of the microbiota from grape derived environments, metagenomes from a fresh Chardonnay must and a composting grape marc were analysed. Taxonomic profiling of the two environments demonstrated two highly diverse communities, with the CGM dataset comprised mostly of fungal species, while the MVGM dataset was dominated by bacteria. The CGM metagenome was assembled from short-read data and is highly fragmented with a low N_{50} of 1,111 bp and a maximum contig size of 100 kbp. A combined assembly of short and long-read data was produced from the MVGM data, which produced an assembly with a N_{50} of 24,477 bp and a maximum contig size of 4.9 Mbp. Binning of the CGM metagenome contigs produced three bins of >90% completeness and <10% contamination, while the MVGM gave 54 bins that exceeded these parameters. Gene predictions from the two metagenomes produced over 1.5 million putative protein encoding gene models, of which 985 are putative β-glucosidases and 136 putative polygalacturonases with less than 80% identity to proteins in the NCBI nr protein database. These new putative enzymes are all divergent from each other by primary amino acid sequence and constitute a reservoir of potential diverse enzymatic activities for use in winemaking.

Chapter 4: Heterologous expression of putative β-glucosidases from GHF3 in *Pichia pastoris* and *Saccharomyces cerevisiae*

Introduction

As in most fruits, a diverse mix of glycoside conjugates accumulate in grape berries during the maturation process, where they typically occur at much higher final concentrations than their aglycone form (Huh et al., 2003, Vogtle et al., 2012, Maicas and Mateo, 2005). As the organoleptic properties of many aglycones (non-sugar moieties of glycosides) are inhibited by the sugar linkage, glycoside conjugates represent a potent reservoir of flavour and aroma compounds in fruit-derived products, such as wine. The exact type and abundance of glycosides are highly variable between grape varieties (Watanabe et al., 2015) and added complexity occurs through variation in both the type of sugar and the number of linked sugar moieties.

In wine, free monoterpenes are strongly associated with flavour and aroma. Grape terpene glycosides have been identified that incorporate rutinose, 6-O-α-l-arabinofuranosyl-β-d-glucopyranoside, 6-O-β-d-apiofuranosyl-β-d-glucopyranoside and/or β-d-glucopyranoside as sugar moieties (Williams et al., 1982, Voirin et al., 1990). As a result, the aglycone substituent is generally covalently linked to a β-glucosyl residue, which is the target substrate for β-glucosidases (EC 3.2.1.21) that hydrolyse the covalent bond and cleave the β-glucose residue from the associated compound. Thus, these enzymes can mediate the release of many aglycones from conjugated glucosides in wine. Enzymatic hydrolysis is also necessary in wine as spontaneous chemical hydrolysis of glycosides occurs slowly under winemaking conditions and often with undesirable molecular rearrangements.

The main wine yeast, *S. cerevisiae*, possesses several genes that encode β-glucosidases, which are involved in cellular processes including cell wall production (Nebreda et al., 1986, Larriba et al., 1993), sporulation (Muthukumar et al., 1993), vacuolar morphology and ergosterol-beta-glucoside catabolism (Watanabe et al., 2015) and some *S. cerevisiae* β-glucosidases are associated with the mitochondria (Huh et al., 2003, Vogtle et al., 2012).

However, despite these known enzyme activities, *S. cerevisiae* exhibits very limited β-glucosidase activity against grape-derived glycosides during the winemaking process (Mateo and Di Stefano, 1997, Aryan et al., 1987). Similarly, grape-derived β-glucosidases typically have very little activity during winemaking (Lecas et al., 1991). β-glucosidases are also produced by microorganisms associated with a plant host and during the decomposition of plant material and these are alternate source for β-glucosidase activity.

In order to investigate potential suitability of β-glucosidases derived from the grape-associated microbiota as a winemaking adjunct, three putative β-glucosidases from GHF3 were selected from the CGM gene models (Chapter 3), for heterologous production in the yeasts *Pichia pastoris* and *Saccharomyces cerevisiae*.

Results

Candidate GHF3 β-glucosidases

Three candidate proteins (GH3-1, GH3-2 and GH3-3) with homology to β-glucosidases of GHF3 were selected for protein expression from genes predicted from the Chardonnay Grape Must (CGM) metagenomic assembly (Table 4.1). These candidates were selected from a group of CGM gene predictions that were identified as potential β-glucosidases by homology against the Interpro protein structure databases (Table 4.1), appeared to have full-length ORFs and with sequences displayed no significant (>80% percent identity) matches to any protein in the NCBI nr protein database. However, as not all of these genes of interest could be tested, GH31, GH3-2 and GH3-3 were then selected for divergent primary amino acid sequences relative to each other (Table 4.1 and Figure 4.1) and as their highest matches in the NCBI database was to uncharacterised gene predictions. GH3-1 to a predicted hypothetical protein from *Pseudocercospora fijiensis*, the species responsible for leaf-spot disease in banana. GH3-2 to a hypothetical protein in *Aureobasidium pullulans*, a common yeast-like fungi and GH3-3 to a hypothetical protein from *Apiotrichum porosum*, a basidiomycetous yeast. All three are fungal species, with *P. fijensis* and *A. pullulans* belonging to the phylum Ascomycota and *A. porosum* to the phylum Basidiomycota. This

taxonomic separation correlates well with average amino acid identity between the proteins (Table 4.1 and Figure 4.1). The putative β-glucosidases were screened for potential signal peptides, which were subsequently removed prior to gene synthesis (Table 4.1). The final, screened coding sequences for the three putative β-glucosidases (GH3-1, GH3-2 and GH3-3) were codon optimised for *S. cerevisiae* and sourced as synthetic gene constructs. Protein expression was subsequently attempted in two expression hosts: *P. pastoris* and *S. cerevisiae* (Table 4.2), taking advantage of the yeasts' protein secretion machinery, *P. pastoris* high culture titres for high protein levels and *S. cerevisiae* fermentative characteristics.

Table 4.1: Summary of assessment and classification of selected putative β-glucosidases identified from 2016 Chardonnay metagenome. Signal peptides predicted by SignalP 4.1.

| Name	Size	Interpro domains	Closest NCBI match	Signal Peptide	% Identity to other candidates
GH3-1	892 aa, (96 kDa)	IPR017853 Glycoside hydrolase superfamily, IPR001764 Glycoside hydrolase, family 3 N-terminal, IPR002772 Glycoside hydrolase family 3 C-terminal domain, IPR026891 Fibronectin type III-like domain, IPR019800 Glycoside hydrolase, family 3, active site, GO:0004553 hydrolase activity, hydrolysing O-glycosyl compounds	98%, 72% Glycoside hydrolase family 3 protein (*Pseudocercospora fijiensis*) XP007920178.1	Yes, Amino acids 1-18. (Eukaryotic model)	40% identity to GH3-2 and 22% identity to GH3-3
GH3-2	778 aa, (83 kDa)	IPR001764 Glycoside hydrolase, family 3, N-terminal, IPR002772 Glycoside hydrolase family 3 C-terminal domain, IPR026891 Fibronectin type III-like domain, GO:0004553 hydrolase activity, hydrolysing O-glycosyl compounds	95%, 75% Hypothetical protein (Aureobasidium pullulans EXF-150) KEQ79228.1	No (Eukaryotic model)	40% to identity to GH3-1 and 21% identity to GH3-3
GH3-3	887 aa, (97 kDa)	IPR001764 Glycoside hydrolase, family 3, N-terminal, IPR002772 Glycoside hydrolase family 3 C-terminal domain, IPR037524 PA14/GLEYA domain, IPR026891 Fibronectin type III-like domain, Signal Peptide (Phobius), GO:0004553 hydrolase activity, hydrolysing O-glycosyl compounds	99%, 47% Hypothetical protein (Apiotrichum porosum) RSH84354.1	No (Eukaryotic model)	22% identity to GH3-1 and 21% identity to GH3-2
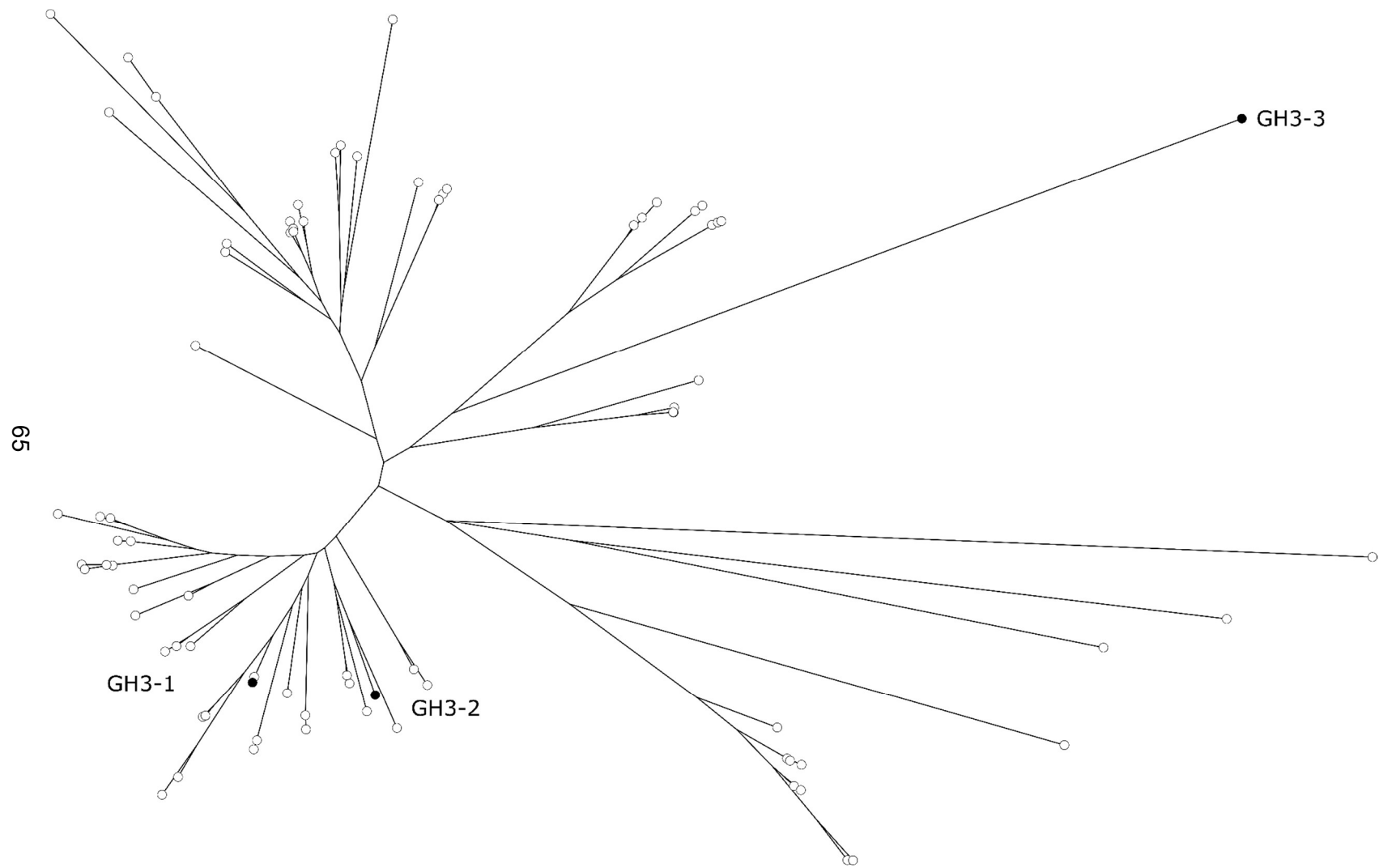
Figure 4.1: Genetic distance tree of GHF3 β-glucosidases candidates identified from the CGM metagenome gene models.

Table 4.2: Summary of expression system variants for putative β-glucosidases from glycoside hydrolase family 3: GH3-1, GH3-2 and GH3-3. All variants contain the alpha secretion factor at the n-terminus of coding sequence to target sequence for protein secretion.

Candidate Enzyme	Expression Host	Name of Variant	Tag	Genome Integration	Inducible/ Constitutive	Confirmed by PCR	Observable Protein
GH3-1	*P. pastoris*	pD912-GH3-1	No tag	pAOX1	pAOX1, Methanol Induction	Y	N
		pD912-His-GH3-1	N-terminal 6xHis	pAOX1	pAOX1, Methanol Induction	Y	N
		pD912-GH3-1-His	C-terminal 6xHis	pAOX1	pAOX1, Methanol Induction	Y	N
	S. cerevisiae	pCVSα-GH3-1	No tag	HO1	Constitutive	N	-
		pCVS3-GH3-1	No tag	Non-integrating, 2µ origin of replication	Constitutive	N	-
GH3-2	*P. pastoris*	pD912-GH3-2	No tag	pAOX1	pAOX1, Methanol Induction	Y	N
		pD912-His-GH3-2	N-terminal 6xHis	pAOX1	pAOX1, Methanol Induction	Y	N
		pD912- GH3-2-His	C-terminal 6xHis	pAOX1	pAOX1, Methanol Induction	Y	N
	S. cerevisiae	pCVSα- GH3-2	No tag	HO1	Constitutive	N	-
		pCVS3- GH3-2	No tag	Non-integrating, 2µ origin of replication	Constitutive	N	-
GH3-3	*P. pastoris*	pD912- GH3-3	No tag	pAOX1	pAOX1, Methanol Induction	Y	N
		pD912-His- GH3-3	N-terminal 6xHis	pAOX1	pAOX1, Methanol Induction	Y	N
		pD912- GH3-3-His	C-terminal 6xHis	pAOX1	pAOX1, Methanol Induction	Y	N
	S. cerevisiae	pCVSα- GH3-3	No tag	HO1	Constitutive	N	-
		pCVS3- GH3-3	No tag	Non-integrating, 2µ origin of replication	Constitutive	N	-

Saccharomyces cerevisiae expression systems for GHF3 enzymes

S. cerevisiae expression was tested via the use of two different expression plasmid variants, pCVSα, which integrates the heterologous expression cassette into the *HO* locus of the *S. cerevisiae* genome (pCVSα) and pCV3S which exists as a multi-copy 2 µm episome. All three expression variants made use of a modified alpha-factor leader sequence to facilitate protein secretion and placed the ORF of interest under the control of the *FBA1* promoter. The modified alpha-factor has amino acid deletions in comparison with the wild type *S. cerevisiae* alpha factor (Figure 4.2).
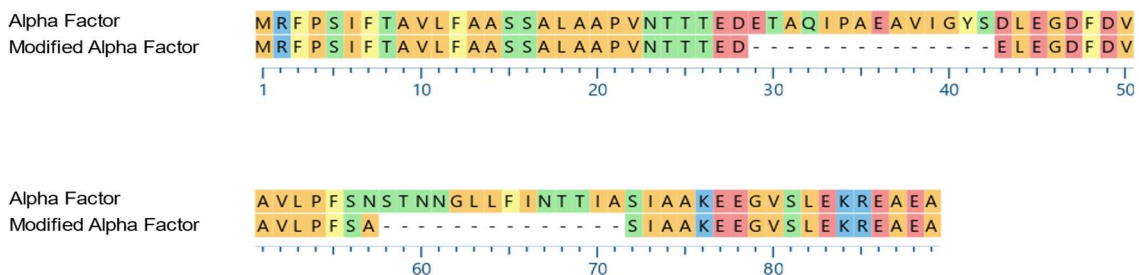

Figure 4.2: Comparison of amino acid sequence between *S. cerevisiae* alpha factor and the modified alpha factor.

The codon-optimised coding sequences for GH3-1, GH3-2 and GH3-3 were each cloned into pCVSα, pCVSα-His and pCV3S, with successful plasmid construction confirmed by restriction digest (Figure 4.3). The three resulting plasmids (pCVSα-GH3-1, pCVSα-GH3-2, pCVSα-GH3-3, pCV3S-GH3-1, pCV3S-GH3-2 and pCV3S-GH3-3) were transformed into both wine (AWRI1631) and laboratory (BY4742) strains of *S. cerevisiae*. However, despite colonies displaying initial resistance to the antibiotic used for positive transformant selection on plates, strains transformed with the expression plasmids destined for genomic integration (pCVSα-GH3-1, pCVSα-GH3-2 and pCVSα-GH3-3) were not viable upon further culturing. Furthermore, integration of the plasmid constructs at the *HO* locus could not be detected by colony PCR. The episomal *S. cerevisiae* plasmids (pCVS3-GH3-1, pCVS3-GH3-181776 and pCVS3-GH3-3) produced stable transformants in BY4742 background and were subsequently tested for protein production in synthetic complete medium with additional glucose. However, no heterologous protein was observed in the supernatant nor the lysate.

A)
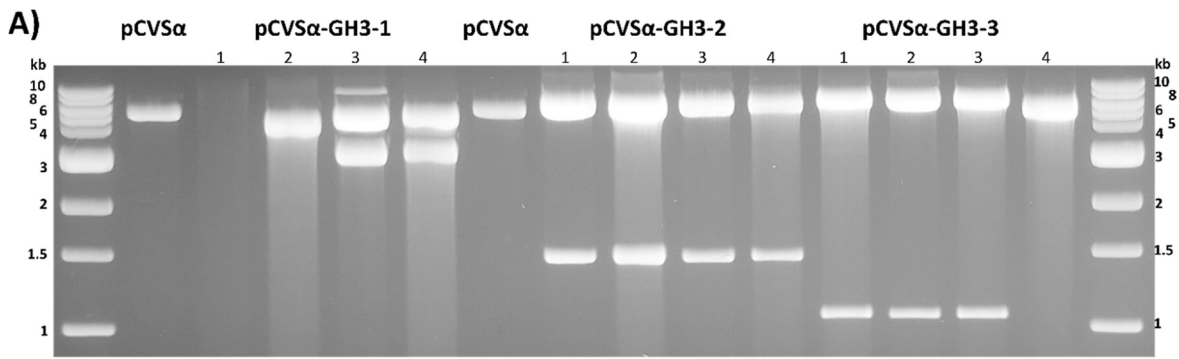
Expected DNA fragement sizes:

pCVSα	NdeI, XbaI	5455 bp
pCVSα-GH3-1	NdeI	3130 + 4928 bp
pCVSα-GH3-2	XbaI	6261bp + 1509 bp
pCVSα-GH3-3	XbaI	6978bp + 1119bp

B)
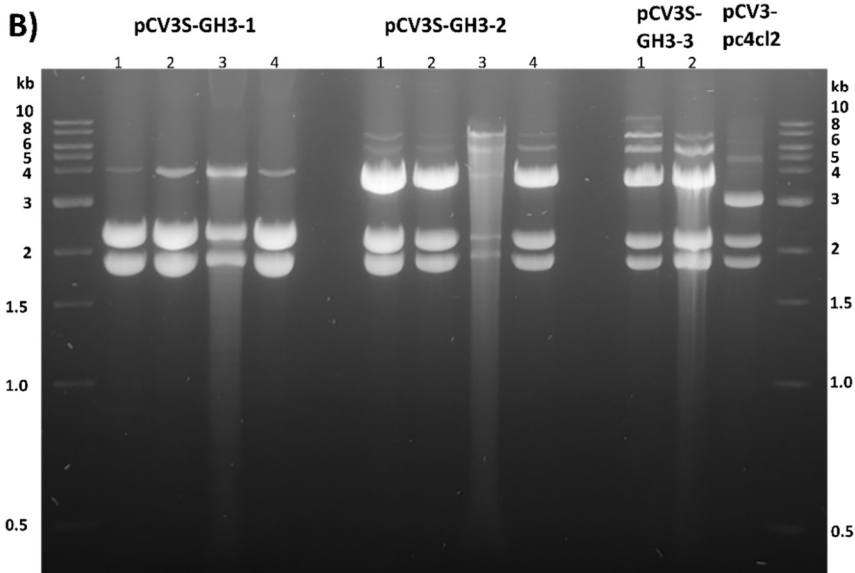
Expected DNA fragement sizes:

| pCV3-pc4cl2	PvuI	2274 + 2068 + 3368 bp
pCV3-GH3-1	PvuI	2274 + 2068 + 4581 bp
pCV3-GH3-2	PvuI	2274 + 2068 + 4293 bp
pCV3-GH3-3	PvuI	2274 + 2068 + 4620 bp

Figure 4.3: Restriction digests to confirm cloning of *S. cerevisiae* expression plasmids for the putative GHF3 β-glucosidases GH3-1, GH3-2 and GH3-3 in *E. coli*. A) The genome integrating plasmids pCVSα-GH3-1, pCVSα-GH3-2, pCVSα-GH3-3 using pCVSα as the backbone vector. B) The episomal plasmids pCV3S-GH3-1, pCV3S-GH3-2 and pCV3S-GH3-3 using the pCV3 backbone excised the plasmid pCV3-pc4Cl2. Bands visualised on 1% (w/v) agarose gel.

Pichia pastoris expression systems for GHF3 enzymes

Three expression plasmid variants were used for the assessment of protein expression in *P. pastoris*, pD912 (untagged), pD912-His (N-terminal His-tag) and pD912-ctHis (C-terminal His-tag). Expression of all three variants was controlled by the methanol-inducible promoter, pAOX1, which also acted as the site for homologous integration into the *P. pastoris* genome. As used in the *S. cerevisiae* plasmids, all of the *P. pastoris* plasmids made use of the modified alpha factor secretion signal. The nine different expression plasmids (pD912-GH3-1, pD912-GH3-2, pD912-GH3-3, pD912-His-GH3-1, pD912-His-GH3-2 and pD912-His-GH3-3, pD912-GH3-1-His, pD912-GH3-2-His and pD912-GH3-3-His) were constructed in *E. coli* and confirmed by restriction digest (Figure 4.4). Plasmids were then transformed into a slow methanol utilising *P. pastoris* strain (BG11). Positive transformants were identified by resistance to zeocin and correct integration confirmed via PCR of the target locus (Figures 4.5 to 4.7). As a positive control, the aspartic protease, BCAP8 (Van Sluyter et al., 2013), was also cloned and transformed into the *P. pastoris* expression plasmid (pD912-BCAP-His) and transformed into BG11 with PCR confirmation of zeocin resistant isolates.

The recombinant *P. pastoris* strains were tested for heterologous protein production over a 96 hr time course following methanol induction. Protein extracts from both the culture supernatant and cell lysates were assessed by protein electrophoresis and staining for the non-tagged variants and immunoblotting using an anti-His monoclonal antibody for the His tagged variants. However, no His-tagged protein was observed in of any of the strains.
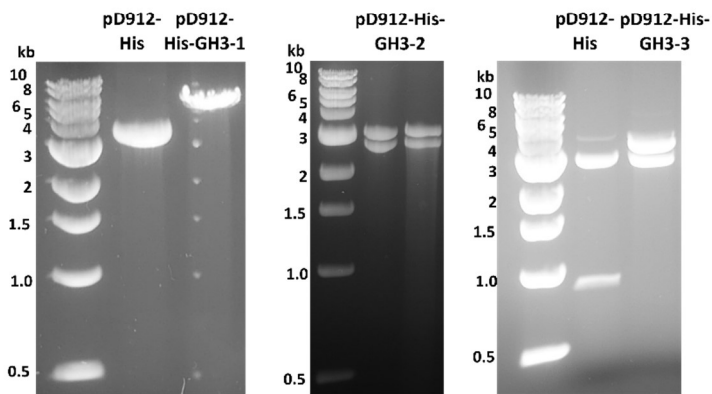
A) Untagged plasmid
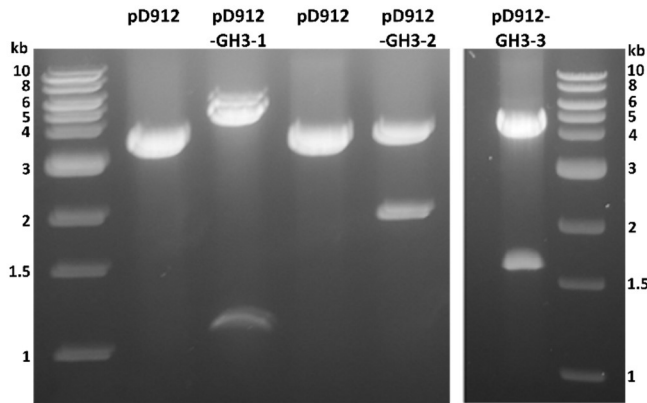
Expected DNA fragment sizes:

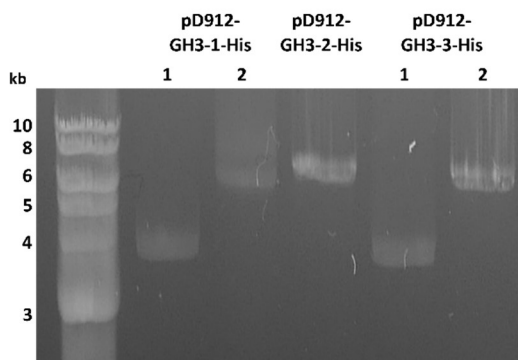|		
pD912	*NdeI , XbaI*	3786 bp
pD912-GH3-1	*NdeI*	1118 + 5246 bp
pD912-GH3-2	*XbaI*	2081 + 3995 bp
pD912-GH3-3	*XbaI*	1691 + 4712 bp

B) N-terminal His tagged

Expected DNA fragment sizes:

pD912-His	*SacI*	3692 bp
	NcoI	933 + 2852 bp
pD912-His-GH3-1	*SacI*	6386 bp
pD912-His-GH3-2	*NcoI*	3245 + 2852 bp
pD912-His-GH3-3	*NcoI*	3572 + 2852 bp

C) C-terminal His tagged

Expected DNA fragment sizes:

pD912	*XhoI*	3692 bp
pD912-GH3-1-His	*XhoI*	6409 bp
pD912-GH3-2-His	*XhoI*	6121 bp
pD912-GH3-3-His	*XhoI*	6448 bp

Figure 4.4: Restriction digest confirmation of cloning of GH3 β-glucosidases (GH3-1, GH3-2 and GH3-3) into *P. pastoris* expression plasmids pD912 (A), pD912-His (B) and pD912-ctHis (C) visualised on 1% (w/v) agarose gel.
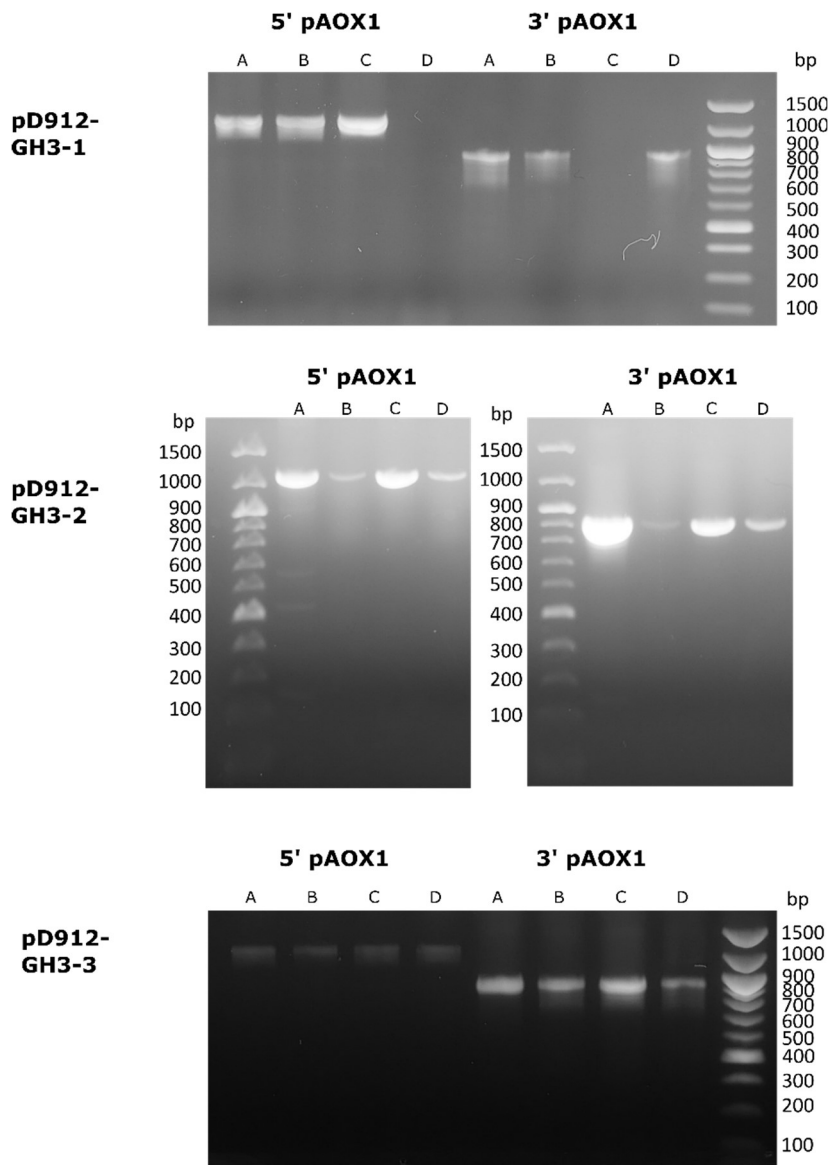
Figure 4.5: PCR confirmation of integration of expression plasmids (pD912-GH3-1, pD912-GH3-2, and pD912-His-GH3-3,) into *P. pastoris* strain BG11 at the pAOX1 locus. Fragments sizes: 1341 bp at 5' pAOX1 and 943 bp at 3' pAOX1 and visualised on 1% (w/v) agarose gel.

Figure 4.6: PCR confirmation of genomic integration of N-terminal His tagged expression plasmids (pD912-His-GH3-1, pD912-His-GH3-2 and pD912-GH3-3) into *P. pastoris* strain BG11 and visualised 1% (w/v) agarose gel. 5' pAOX1: 1341bp. 3' pAOX1: 943bp
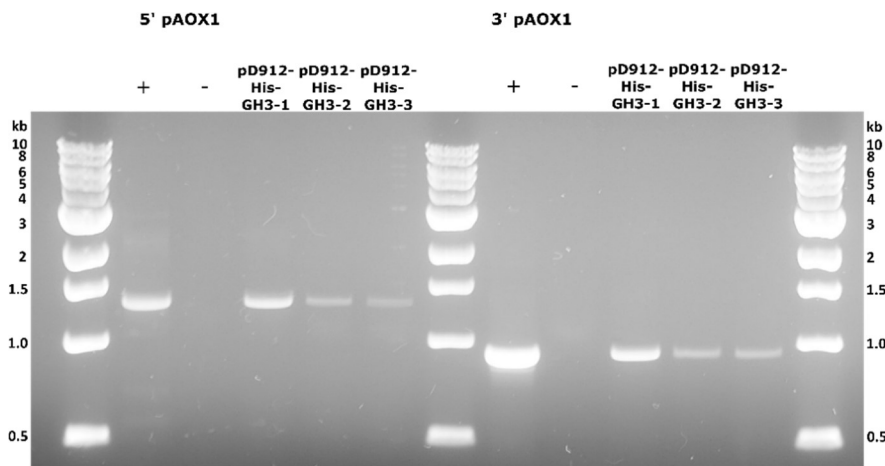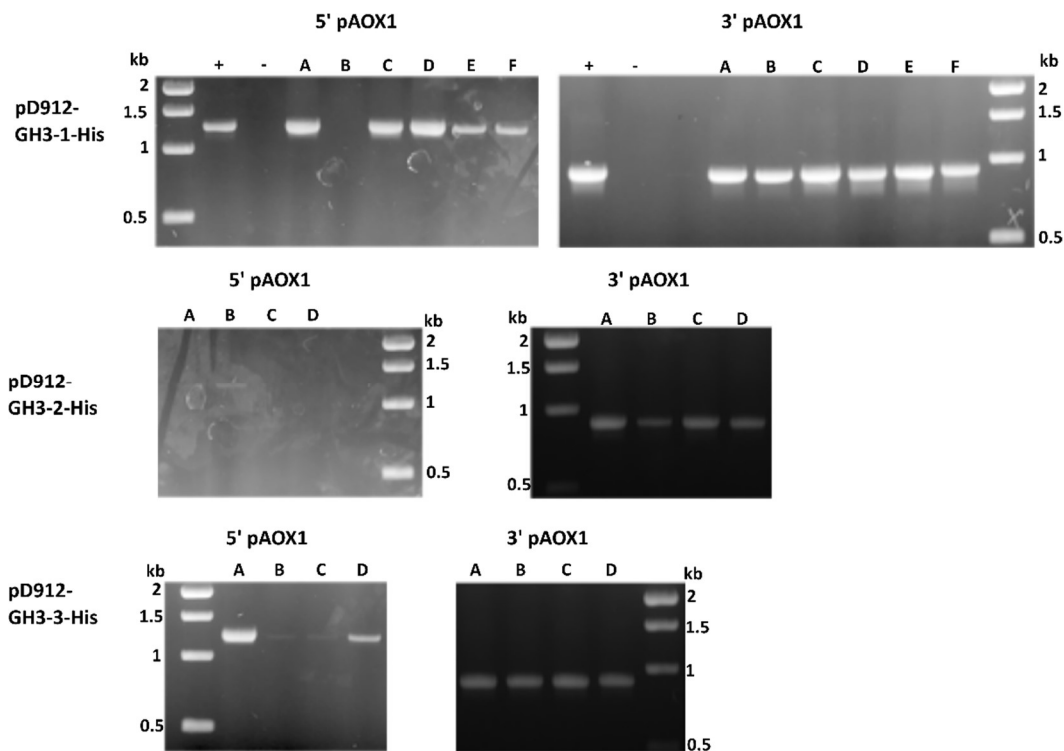

Figure 4.7 PCR confirmation of genomic integration of C-terminal His tagged expression plasmids (pD912-GH3-1-His, pD912-GH3-2-His and pD912-GH3-3-His) into the *P. pastoris* strain BG11. Fragments sizes: 1341 bp at 5' pAOX1 integration site and 943 bp at 3' pAOX1 integration site and visualised on 1% (w/v) agarose gel.

RT-qPCR

Successful induction of the *AOX1* promoter by methanol and production of specific mRNA was tested by RT-qPCR in the C-terminal His-tagged expression system for all three β-glucosidases (GH3-1, GH3-2 and GH3-3) and the BCAP8 control. Relative transcript levels were compared for each gene at just prior to, and 6 hrs after, the addition of methanol (Figure 4.8). Robust induction of mRNA expression was observed for the BCAP8 control demonstrating that the expression system itself was not fundamentally flawed. The GH3-2-His strain also appeared to be producing mRNA following methanol induction. However, the GH3-1-His and GH3-3-His *P. pastoris* strains exhibited no observable transcriptional response to methanol. In order to ensure the promoter regions had not been mutated, the 5' pAOX1 PCR product (used to determine genomic integration) was sequenced. Sequence data was consistent with the design and the sequence of the plasmid used to make the transformants (Figure 4.9).
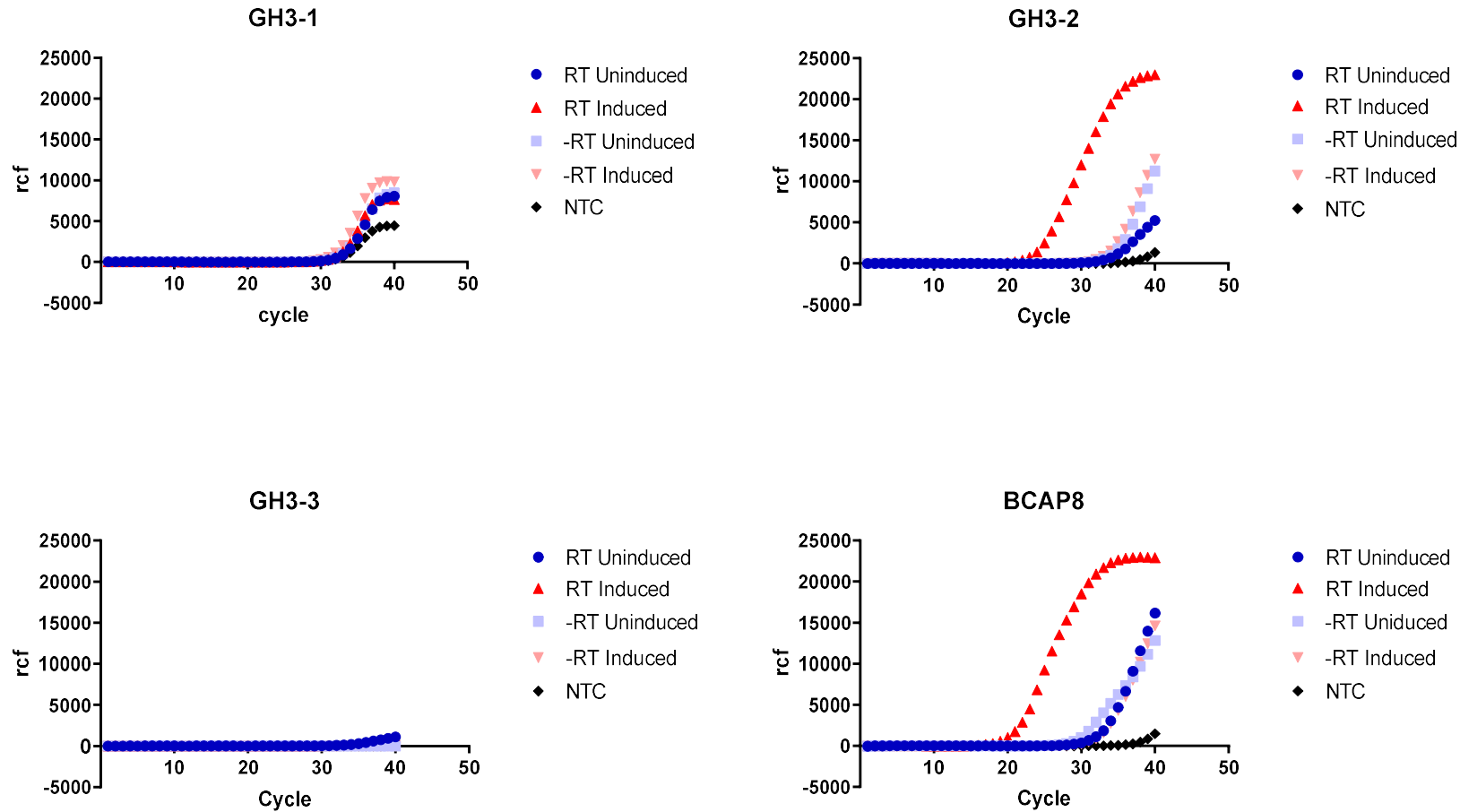
Figure 4.8: RT-qPCR amplification curves of GH3-1, GH3-2, GH3-3 and positive control (BCAP8) loci in uninduced (no methanol) and induced (methanol) *P. pastoris* strains containing expression plasmids for putative β-glucosidases, GH3-1, GH3-2 and GH3-3, and positive control (BCAP8). RT = mRNA was reverse transcribed. -RT= mRNA without reverse transcription.

Figure 4.9: 5' pAOX1 PCR product sequence of pD912-GH3-1-His, pD912-GH3-2-His and pD912-GH3-3-His expression strains, low quality bases trimmed.

Discussion

Heterologous protein expression of three GHF3 β-glucosidases (GH3-1, GH3-2 and GH3-3) was attempted in two different yeast expression hosts, *S. cerevisiae* and *P. pastoris*. Although there were no apparent issues regarding the assembly of the expression plasmids in *E. coli*, none of the expression cassettes were shown to produce detectable levels of protein in either yeast species. There are multiple levels at which protein production could be failing and although the reasons for failed expression are not clear, several avenues could be ruled out.

It is not expected that the expression of the β-glucosidase enzymes would be toxic to the hosts, as *P. pastoris* and *S. cerevisiae* have been used to express at least one other GHF3 β-glucosidase previously (Dan et al., 2000). However, this was achieved in *S. cerevisiae* using the inducible *GAL1* promoter which may indicate that transient expression may be more suitable than the constitutive expression provided by the *FBA1* promoter used in this study. However, the *GAL1* system would not be appropriate for 'in-wine' testing due to glucose inhibition and a different inducible promoter would be required.

In the case of the *S. cerevisiae* expression systems, there were two separate responses to transformation with expression plasmid. All plasmids containing the *HO* homologous sites for genomic integration (pCVSα-GH3-1, pCVSα-GH3-2 and pCVSα-GH3-3) rendered transformants that were non-viable upon sub-culturing in liquid media, despite initial antibiotic resistance on the transformation plates, suggesting poor selection and ineffective genomic integration. In comparison, cells transformed with the replicating plasmid (pCVS3-GH3-1, pCV3S-GH3-2 and pCV3S-GH3-21279) grew well in selective media but did not produce detectable levels of the heterologous protein.

The *HO* homologous regions used as genomic integration sites are approximately 150 bp in length and have been used previously with other gene cassettes (Lee et al., 2016). However, they have not been used in combination with the modified alpha factor as a secretion signal. As alpha factor is a native *S. cerevisiae* protein, it is possible the sequence (183 bp) is acting as an alternate homologous site and interfering with genomic integration

at the *HO* locus. If this is occurring, altering the peptide secretion signal may resolve the viability issues that were observed for these transformants.

The absence of detectable heterologous protein in the pCV3S series of transformants could have numerous causes, including repression of gene expression, unstable mRNA transcripts, protein degradation or loss of the plasmid. However, given the lower levels of secreted protein typically produced by *S. cerevisiae* relative to *Pichia*, the lack of protein could also simply be due to overall inadequate expression levels and not a specific problem with the expression system.

As *P. pastoris* is considered to be a more effective platform for the production of large amounts of secreted protein, the absence of protein expression in this system is more informative. The control plasmid (pD912-BCAP8-His) produced significant amounts of protein under the same culture conditions as the three β-glucosidase-expressing strains and confirmed that the expression system itself was functional. As genomic integration was confirmed for all *P. pastoris* plasmids by PCR and antibiotic resistance, issues also do not appear to be due to differences in genomic integration of the cassette.

Potential issues related to the production of mRNA were assessed using RT-qPCR, comparing mRNA levels of uninduced and methanol induced samples of the C-terminal His tagged variants (pD912-GH3-1-His, pD912-GH3-2-His and pD912-GH3-21279-His). The BCAP8 control was used to benchmark the expected transcriptional response of pAOX1 to methanol induction and relatively weak mRNA induction was observed for GH3-2. No mRNA was observed for either GH3-1 or GH3-3. This lack of mRNA template is likely the cause of GH3-1 and GH3-3 not being expressed. However, as no DNA mutations were observed in the promoter regions of these transformants which are identical to the control, the lack of specific mRNA is likely due to other post-transcriptional phenomena, such as transcript or protein instability.

The stability of the GH3 mRNAs may be adversely affected by the ORF size, as the glucosidases investigated in this study were large proteins (839 - 953 aa including the 61

aa alpha factor leader). In comparison, the BCAP8 positive control was only 381 aa (including the alpha factor leader). Alternate codon usage could also be used to potentially improve the transcript stability of GH3-1 and GH3-3. However, if transcription is being repressed, further investigation would be required. RNAseq experiments could be used to test for changes in cellular gene expression and potentially identify repression. However, this would not indicate why the β-glucosidases are causing an issue and is beyond the scope of this work. Simpler experiments would be to attempt protein expression in a different host organism such as *E. coli* or in vitro. Experiments could also attempt to express shorter protein fragments and detect specific problematic regions.

The third β-glucosidase, GH3-2, appears to have a functioning pAOX1 by RT-qPCR. As such, malfunction in protein expression is occurring post transcription; potentially during translation or protein processing such as folding and secretion. No products of degradation were observed indicating highly efficient degradation or very low overall expression. A protease deficient *P. pastoris* strain could be used to test for possible degradation.

Conclusions

Three putative β-glucosidases, GH3-1, GH3-2 and GH3-3, of yeast origin with homology to GHF3 were selected for heterologous expression in the yeasts *P. pastoris* and *S. cerevisiae*. However, expression attempts for these β-glucosidases failed in both hosts. Ineffective genomic integration appears to be preventing transformation with pCVSα plasmids into *S. cerevisiae*. However, it is not clear why the plasmid-based protein expression of pCV3S did not produce protein in *S. cerevisiae*. In comparison, *P. pastoris* expression issues appear to be enzyme specific. RT-qPCR results for the strains confirmed that mRNA transcript for GH3-1-His and GH3-3-His is not present 6 hours after induction of the AOX1 promoter with methanol and the lack of transcript is preventing production of these proteins. However, the GH3-2-His strain does appear to contain mRNA and protein production appears to be failing post-transcription. There are numerous avenues to investigate but in future work it would be expedient to continue experiments in *P. pastoris* over *S. cerevisiae*, as the issues appear to occur at later stages of protein production.

Chapter 5: Heterologous expression and characterisation of putative β-glucosidases from GHF1 in *Pichia pastoris*

Introduction

GHF1 β-glucosidases (EC 3.2.1.21) hydrolyse the covalent bonds of terminal non-reducing β-glucosyl residues, releasing a β-glucose and a second compound. The glycosylation and de-glycosylation of aglycones are involved in various biological roles but are significant to the food and beverage industries as glycosylation of organoleptic aglycones can inhibit sensorial properties and the glycosylated forms of aglycones tend to be more abundant. Consequently, glycosides are a reservoir of organoleptic compounds. In grapes, many of these aglycones are connected to glucose moieties (Williams et al., 1982, Voirin et al., 1990) and therefore β-glucosidases can mediate their release. In this way, β-glucosidases could potentially be used to modify and promote specific flavours in during winemaking.

Glucose tolerance is a significant property of GHF1 β-glucosidases (Riou et al., 1998, Perez-Pons et al., 1995, De Giuseppe et al., 2014) and is highly desirable for many applications including winemaking. GHF1 enzymes have a retaining mechanism (Withers et al., 1986) whereby stereochemistry of the anomer centre is retained after hydrolysis. As stereochemistry is retained, the structure of organoleptic compound is preserved.

Four β-glucosidases belonging to GHF1 were selected from the mixed varietal grape marc metagenome (Chapter Three) for heterologous expression in the yeast *Pichia pastoris* in order to investigate enzyme properties and assess potential suitability as a winemaking adjunct.

Results

Candidate β-glucosidases

Four candidate β-glucosidases (designated GH1-1, GH1-2, GH1-3 and GH1-4) with homology to β-glucosidases from GHF1 (Table 5.1) were selected for analysis from the gene predictions of the MVGM metagenome (Chapter 3). They contain no matches to the NCBI protein database (≥80% amino acid) and have divergent primary amino acid sequences relative to each other and other matching gene predictions (Figure 5.1). All candidates appear to cover the full length of InterPro GHF1 domain (IPR001360) or GHF1, β-glucosidase (IPR017736), and do not appear to be truncated. None of the candidate β-glucosidases were predicted to have signal peptides. GH1-1, GH1-2, GH1-3 and GH1-4 all appear to be of bacterial origin with their closest homologs being represented by bacterial proteins in the NCBI protein database (Table 5.1).

Table 5.1: Summary of assessment and classification of selected putative β-glucosidases identified from the Mixed Varietal Grape Marc (MVGM) metagenomes. Signal peptides predicted by SignalP 4.1.

| Name	Size	Glycosylation sites (Asn-Xaa-Ser/Thr)	Interpro domains	Closest NCBI match	Signal Peptide	Percentage Identity to other candidates
GH1-1	449 aa (51kDa)	No	Glycoside hydrolase, family 1, β-glucosidase (IPR017736), GO:0008422 β-glucosidase activity	100%, 66% β-glucosidase (Rheinheimera tuosuensis) WP_097110528.1	No, No. Bacterial gram + and gram – models.	GH1-2: 39% GH1-3: 22% GH1-4: 44%
GH1-2	455 aa (51kDa)	No	Glycoside hydrolase, family 1, β-glucosidase (IPR017736), GO:0008422 β-glucosidase activity	99%, 58% β-glucosidase (Sorangium cellulosum) WP_012240271.1	No, No. Bacterial gram + and gram – models.	GH1-1: 39% GH1-3: 24% GH1-4: 27%
GH1-3	422 aa (45kDa)	No	Glycoside hydrolase family 1 (IPR001360), GO:0004553 hydrolase activity, hydrolysing O-glycosyl compounds	93%, 48% glycoside hydrolase family 1 protein (Acidimicrobiia bacterium) RTL08589.1	No, No. Bacterial gram + and gram – models.	GH1-1: 22% GH1-2: 24% GH1-4: 21%
GH1-4	477 aa (54kDa)	Yes. Position: 424 amino acid. Sequence: NRSV	Glycoside hydrolase family 1 (IPR001360), GO:0004553 hydrolase activity, hydrolysing O-glycosyl compounds GO:0004553 hydrolase activity, hydrolysing O-glycosyl compounds	99%, 42% Hypothetical protein (Bdellovibrionales bacterium GWA2 29 15) OFZ13310.1	No, No. Bacterial gram + and gram – models.	GH1-1: 44% GH1-2: 27% GH1-3: 21%
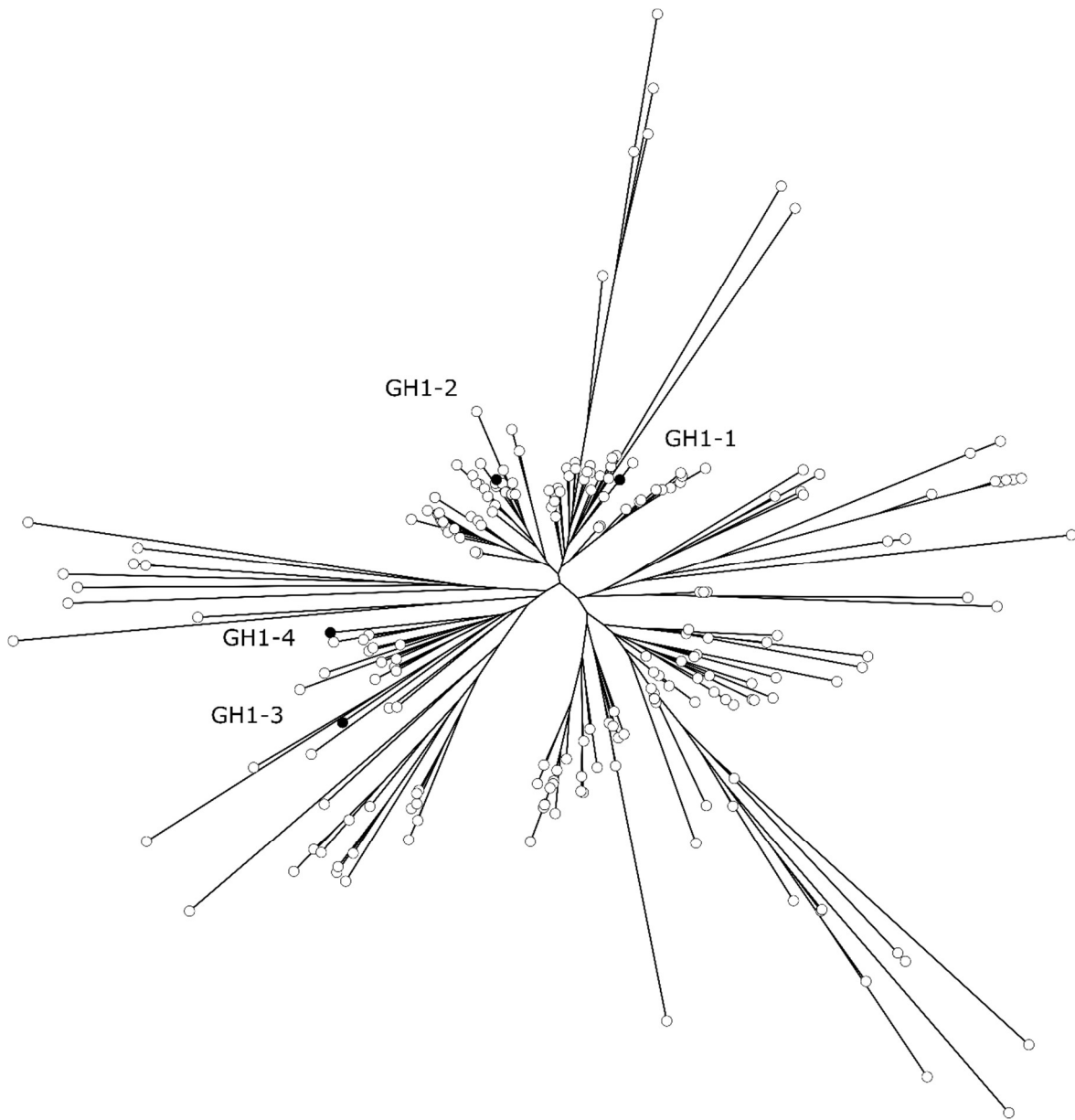
Figure 5.1: Distance in primary amino acid sequence of β-glucosidases candidates identified from CGM and MVGM metagenomes. Candidate GHF1 β-glucosidases in black.

Expression of GHF1 β-glucosidases, GH1-1, GH1-2, GH1-3 and GH1-4, in *Pichia pastoris*

The amino acid sequences of the four putative GHF1 β-glucosidases (GH1-1, GH1-2, GH1-3 and GH1-4) were codon optimised for *P. pastoris* expression and synthesised with a C-terminal 6xHis tag for detection by immunoblotting. These ORF sequences were cloned into the *P. pastoris* expression plasmid, pD912 using Gibson assembly (Gibson et al., 2009). Successful construction of the expression plasmids, pD912-GH1-1, pD912-GH1-2, pD912-GH1-3 and pD912-GH1-4, in *E. coli* was confirmed by restriction digest of purified plasmid DNA (Figure 5.2). The pD912 expression plasmid fuses the coding sequences for GH1-1, GH1-2, GH1-3 and GH1-4 to a modified alpha factor leader sequence to facilitate protein secretion. pAOX1 acts as both the homologous site for plasmid integration into the *P. pastoris* genome and also drives methanol induced expression of the ORF of interest.

Expected fragment sizes:

pD912	*Xho*I	576 + 3210 bp
pD912-GH1-1	*Xho*I	1895 + 3210 bp
pD912-GH1-2	*Xho*I	1913 + 3210 bp
pD912-GH1-3	*Xho*I	1811 + 3210 bp
pD912-GH1-4	*Xho*I	1979 + 3210 bp
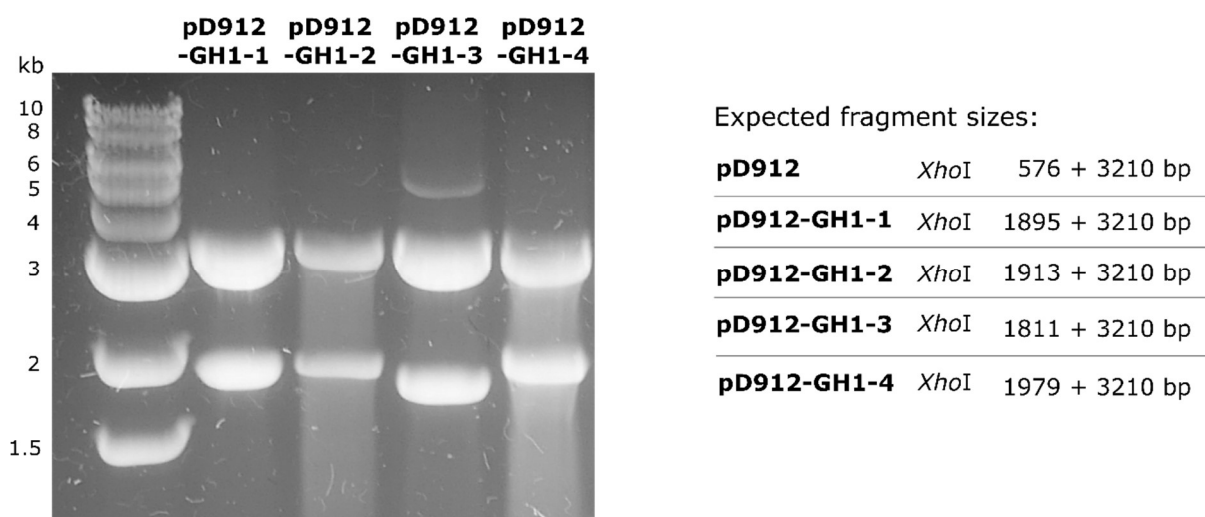
Figure 5.2: Preparation of C-terminally His-tagged gene expression cassettes for putative β-glucosidases (GH1-1, GH1-2, GH1-3 and GH1-4) in *P. pastoris* and visualised on 1% (w/v) agarose gel. Restriction digest of *P. pastoris* expression vectors: pD912-GH1-1, pD912-GH1-2, pD912-GH1-3 and pD912-GH1-4.

The expression plasmids were transformed into a slow methanol utilizing *P. pastoris* strain, BG11. Transformants were selected via zeocin resistance and successful genomic integration at pAOX1 was confirmed by PCR. All four candidate enzymes produced successfully integrated transformants (Figure 5.3).

Figure 5.3 PCR verification of genomic integration of expression plasmids (pD912-GH1-1, pD912-GH1-2, pD912-GH1-3 and pD912-GH1-4) into *P. pastoris* strain BG11 at pAOX1 and visualised on 1% (w/v) agarose gel.
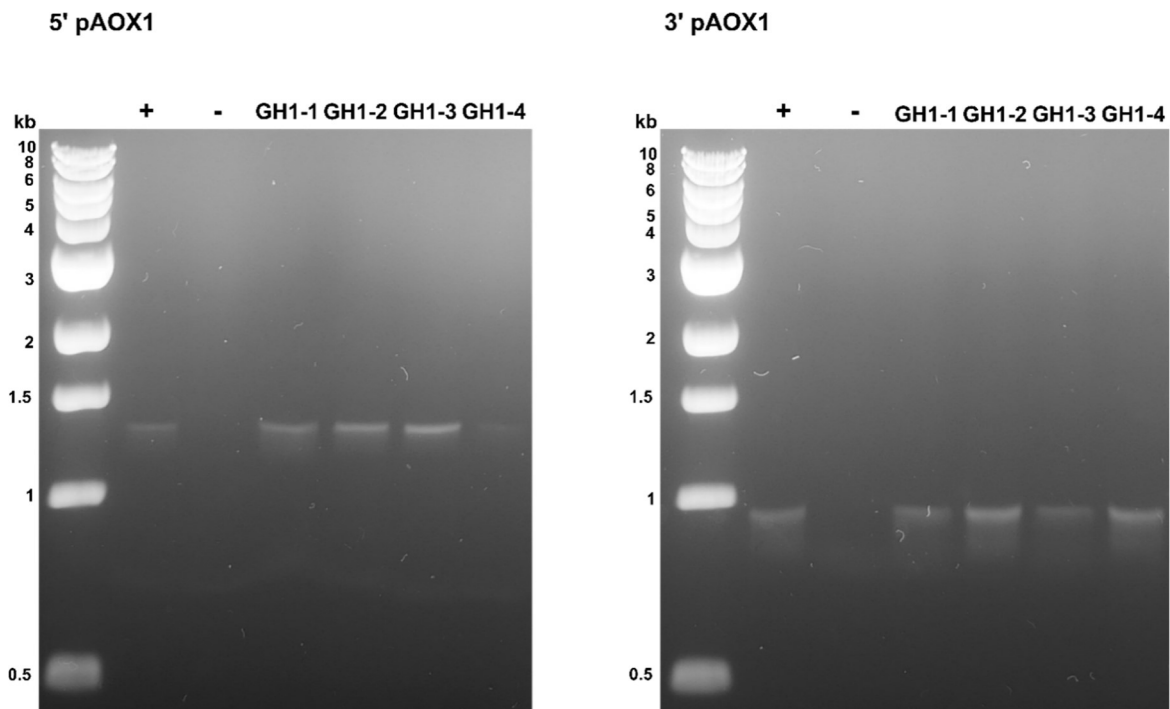
The resulting four recombinant *P. pastoris* strains for GH1-1, GH1-2, GH1-3 and GH1-4, respectively) were tested for protein expression by methanol induction over a 96 hr time course. Recombinant protein was detected via immunoblotting using an anti-His monoclonal antibody (Figures 5.4). Relative protein abundance was measured by the relative density of protein bands in immunoblot as determined by image analysis (Table 5.2). GH1-1, GH1-3 and GH1-4 were readily detected in culture supernatant. GH1-2 protein was not observed in supernatant (Figure 5.4). However, a protein band of the expected mass for GH1-2 was observed in the cell lysate which appeared to be methanol induced. GH1-1 and GH1-3 show highly abundant protein following 12 hr of methanol induction, however the intensity of these bands diminish as the induction period progresses and smaller protein bands appear (Figure 5.4 and Table 5.2). The abundance of GH1-1 at 12 hr was reduced to a tenth of its original intensity by 96 hr and little GH1-1 protein was observed in the cell lysate. GH1-3 expression reduces at 96 hr. GH1-3 also appeared in the cell lysate with an apparent increasing abundance over the 96 hr. GH1-4 accumulates over the 96 hr

induction period in both the supernatant and cell lysate with no degradation products detected. However, as immunoblot does not detect protein without the His epitope, there may be untagged products that are not observed. Optimum expression of GH1-1 occurs at 12 hr, GH1-3 at 24 hr and GH1-4 at between 72 hr and 96 hr (Table 5.2). GH1-1, GH1-3 and GH1-4 showed no obvious signs of glycosylation, with all of the observed protein bands occurring at the expected sizes based on amino acid sequence alone.
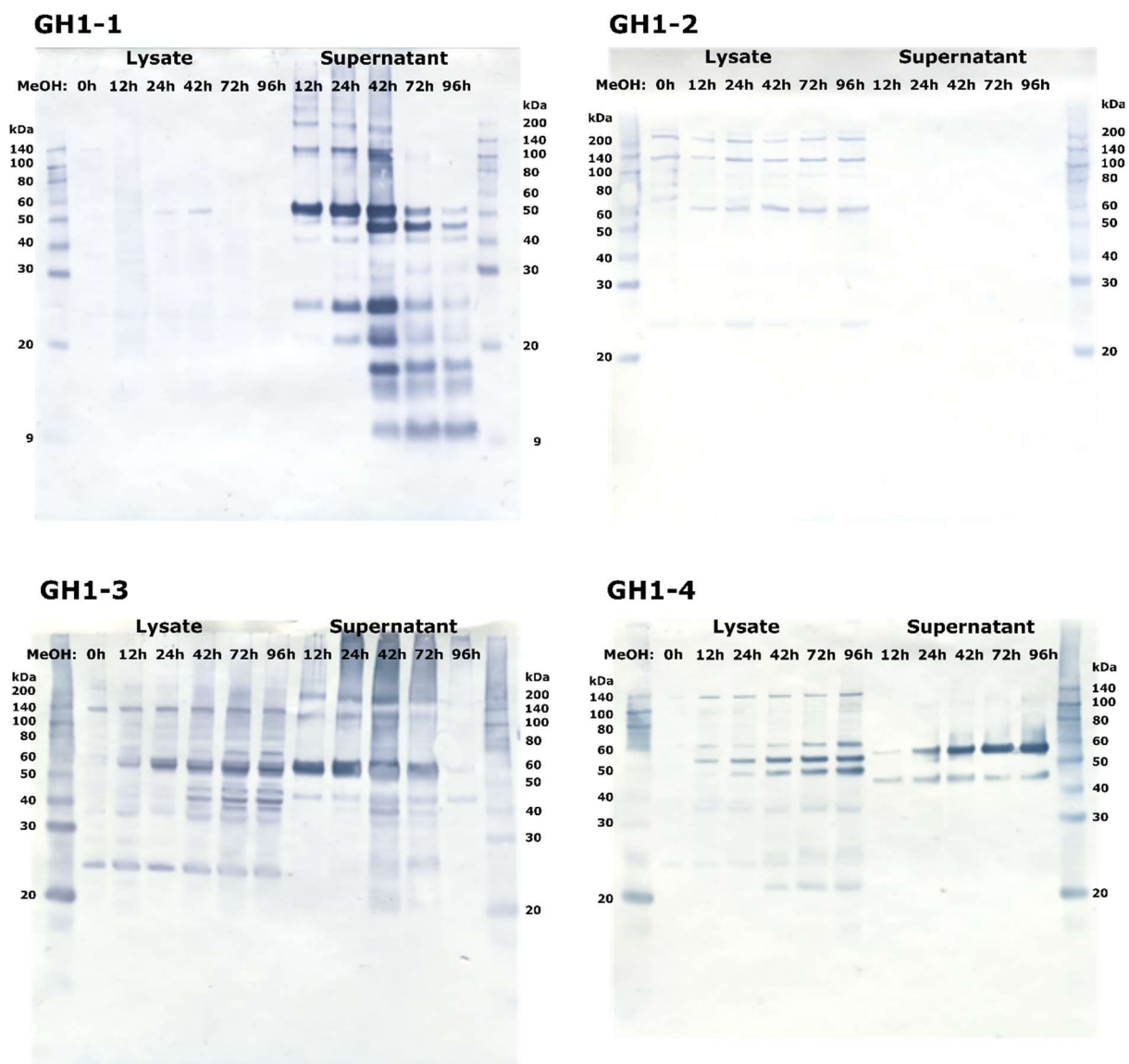

Figure 5.4: Protein expression of putative glycoside hydrolase family 1 β-glucosidases GH1-1 (at ~50kDa), GH1-2 (~60kDa), GH1-3 (~55kDa) and GH1-4 (~55kDa) by the recombinant *P. pastoris* strains. Visualised on nitrocellulose membrane immunoblotted with anti-His tag monoclonal antibody.

Table 5.2: Image analysis of GH1-1, GH1-2, GH1-3 and GH1-4 immunoblots over 96 hours of methanol induction. RD = relative density of recombinant protein. Optimum relative density in bold.

			12 hr	24 hr	48 hr	72 hr	96 hr
GH1-1	Supernatant	Area	22628	23181	13982	7417	2005
		RD	**1.00**	1.02	0.62	0.33	0.09
GH1-2	Lysate	Area	3172	8873	10971	9589	15750
		RD	1.00	2.80	3.46	3.02	**4.97**
GH1-3	Supernatant	Area	47480	48850	20708	32431	1434
		RD	1.00	**1.03**	0.44	0.68	0.03
	Lysate	Area	7732	28440	26726	29070	28672
		RD	1.00	3.68	3.46	**3.76**	3.71
GH1-4	Supernatant	Area	1532	16350	33061	31989	50747
		RD	1.00	10.67	21.58	20.88	**33.13**
	Lysate	Area	2913	6350	11395	13583	12369
		RD	1.00	2.18	3.91	**4.66**	4.25

Characterisation of GH1-1, GH1-3 and GH1-4 putative β-glucosidases

The β-glucosidase activity of heterologously expressed GH1-1, GH1-3 and GH1-4 was evaluated using p-nitrophenyl-β-D-glucopyranoside and recombinant protein in crude supernatant. Supernatants were harvested at the optimum expression period for each enzyme as determined from the immunoblotting experiments (Table 5.2). β-glucosidase activity was readily detected for all three enzymes and the temperature and pH optima for each enzyme were determined (Figure 5.5). GH1-3 and GH1-4 showed very similar patterns in response to temperature and pH and were most active at 30 °C and pH 6. However, GH1-3 appears to be more active per mg of total protein in the supernatant. In comparison, GH1-1 is most active at 20 °C and pH 7. All three enzymes are poorly tolerant of high temperatures but are more robust to varying pH levels.
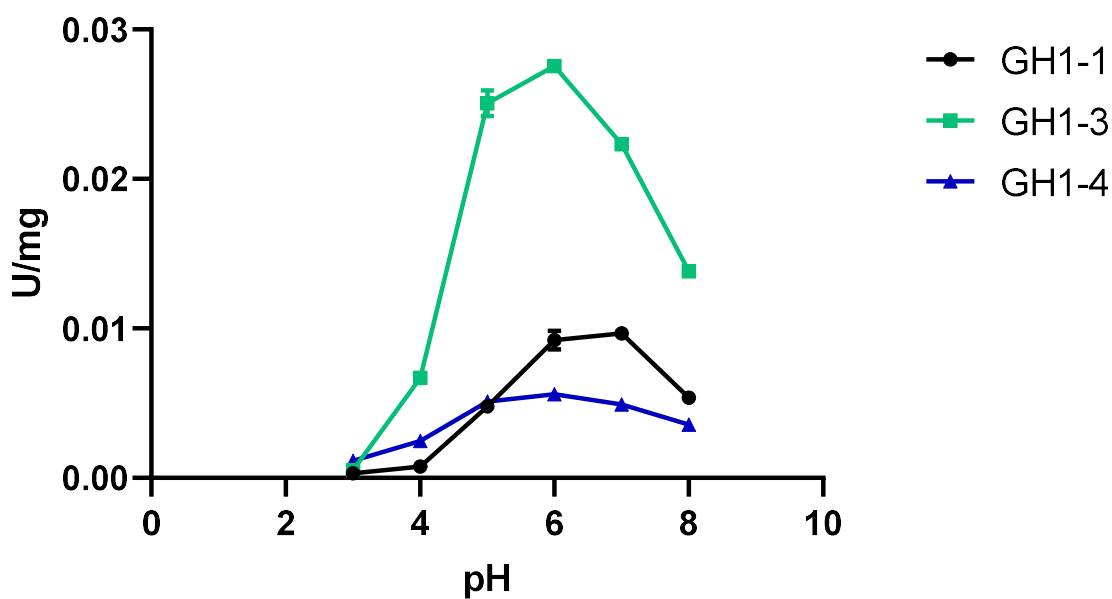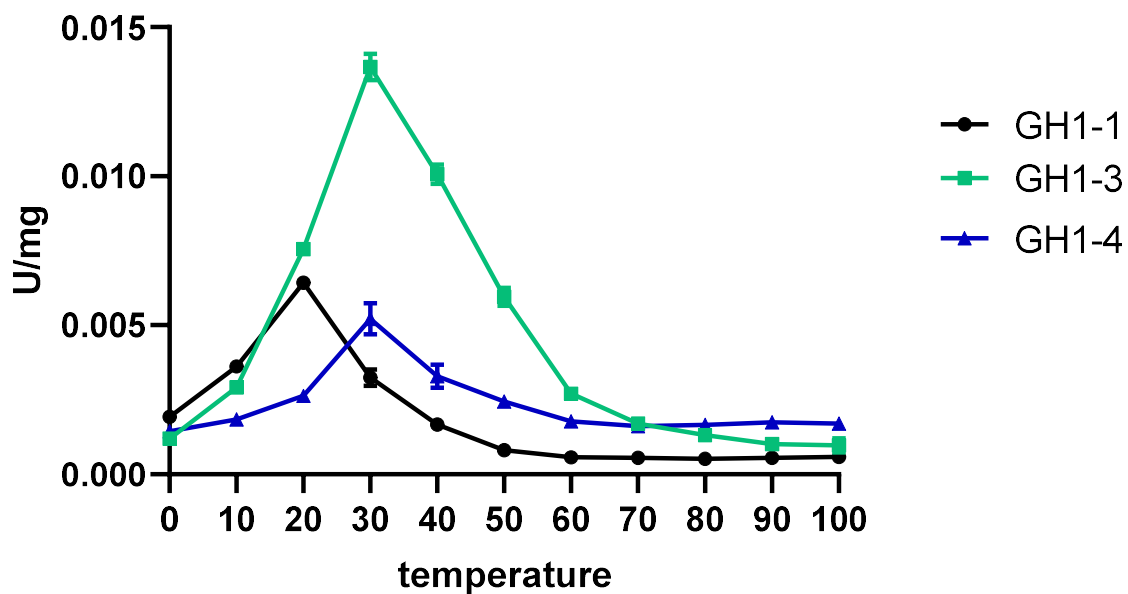
Figure 5.5: Response of putative β-glucosidases GH1-1, GH1-3 and GH1-4 activity on p-nitrophenyl glucopyranoside substrate to pH and temperature. pH test performed at 37 °C and temperature tests at pH 6.0. Activity (U) was defined as the µmol of β-glucose (equimolar with p-nitrophenol) released per minute.

The effect of different concentrations of the p-nitrophenyl substrate on enzyme velocity was evaluated in order to determine substrate affinity (K_m) and maximum velocity for each protein (Figures 5.6 to 5.8). Non-linear regressions were performed using Michaelis-Menten ($Y=V_{max}*X/(K_m+X)$) and Allosteric Sigmoidal ($Y=V_{max}*X^h/(K_{half}^h+X^h)$) models and best fit determined by the sum of squares F test. GH1-1 displayed allosteric sigmoidal enzyme kinetics, $R^2=0.9805$ (Figure 5.6). The maximum velocity of GH1-1 is 0.075 µmol/min/mg and its substrate concentration for half maximal enzyme velocity (K_{half}) is estimated at 149.8 mM. GH1-3 and GH1-4 both fit Michalis-Menten substrate kinetics. GH1-3 has a maximum enzyme velocity of 0.061 µmol/min/mg and a substrate affinity (K_m) of 5.5 mM. GH1-4 has a maximum enzyme velocity of 0.046 µmol/min.mg and Km of 7.2 mM.

GH1-1
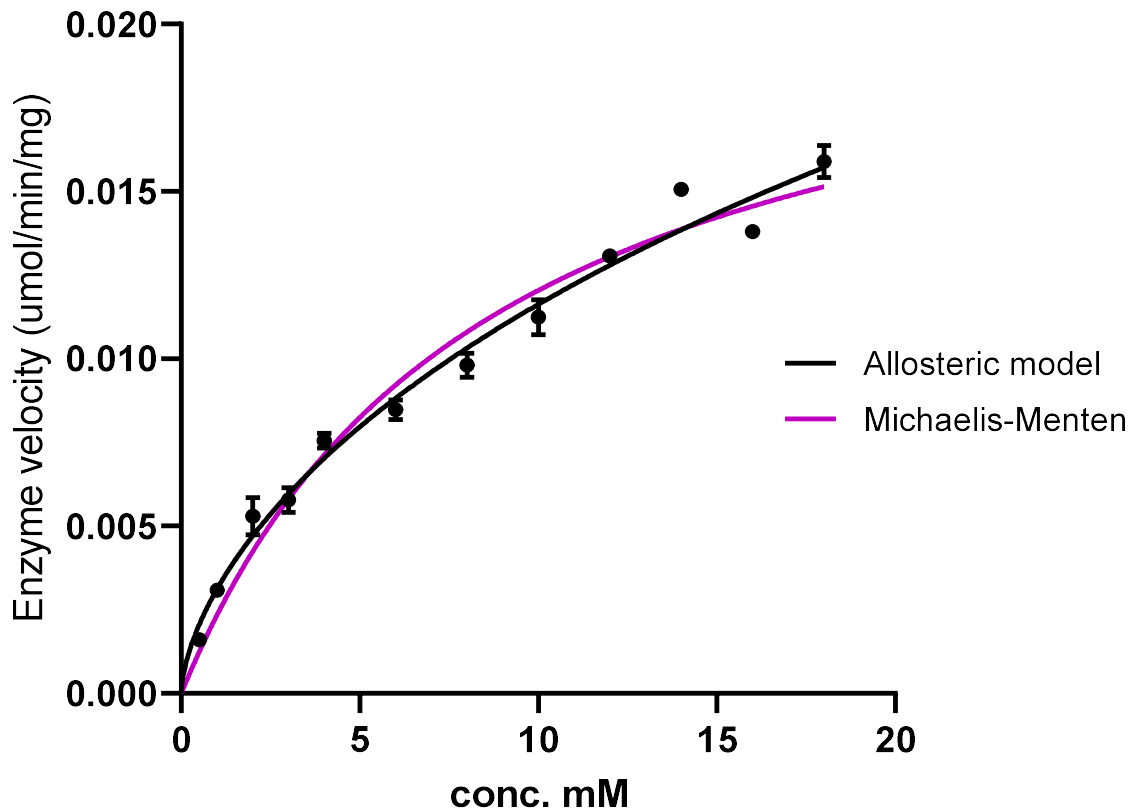
Allosteric sigmoidal Model	
V_{max}	0.075 μmol/min/mg
h	0.63
K_{half}	149.8 mM
K_{prime}	23.1 mM
Goodness of Fit	
R squared	0.9802
Constraints	
h	h > 0
K_{half}	K_{half} > 0

Figure 5.6: Substrate concentration and enzyme velocity kinetics of glycoside hydrolase family 1 β-glucosidase GH1-1 monitored on spectrophotometer at pH 7.5 and at 25 °C. Best fit curve matching allosteric sigmoidal non-linear regressions. Enzyme velocity as the number of μmols of p-nitrophenol released per minute. Error bars the standard deviation of replicates.

GH1-3
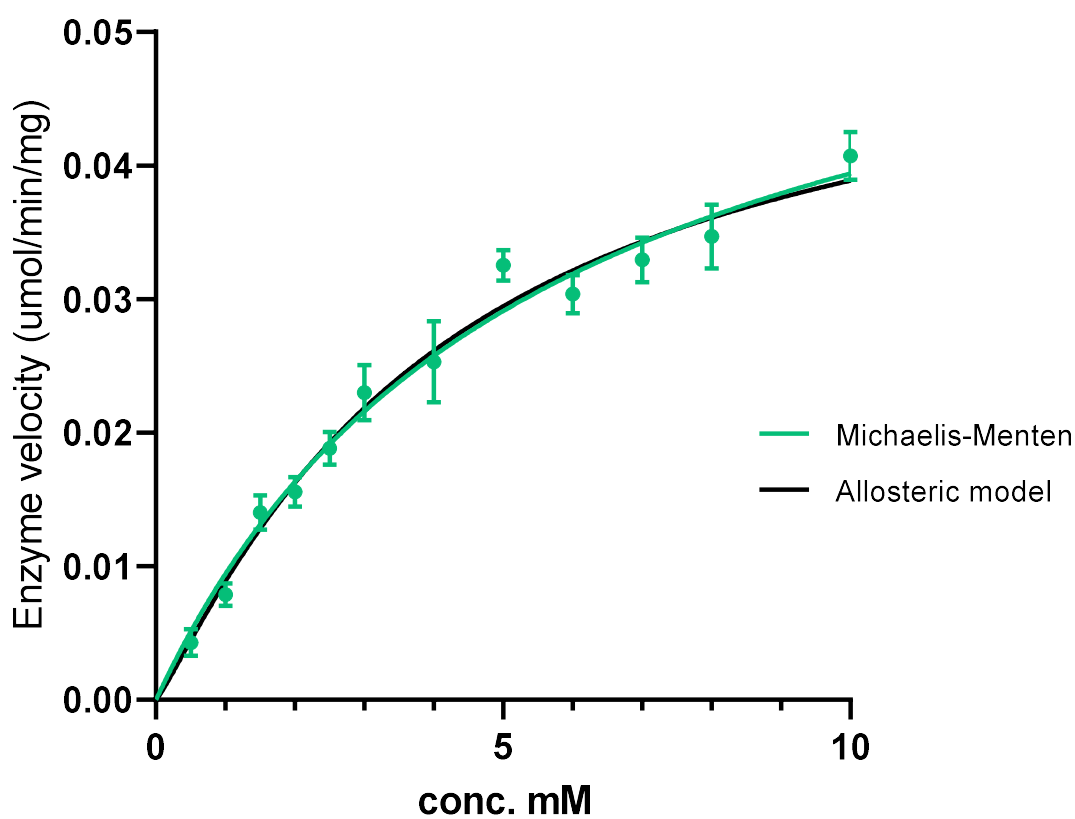
Michaelis-Menten	
V_{max}	0.061 µmol/min/mg
K_m	5.5 mM
Goodness of Fit	
R squared	0.9640
Constraints	
K_m	$K_m > 0$

Figure 5.7: Substrate concentration and enzyme velocity kinetics of glycoside hydrolase family 1 β-glucosidase GH1-2 monitored on spectrophotometer at pH 7.5 and at 25 °C, best fit curve matching allosteric sigmoidal non-linear regressions. Enzyme velocity as the number of µmols of p-nitrophenol released per minute. Error bars the standard deviation of replicates.

GH1-4
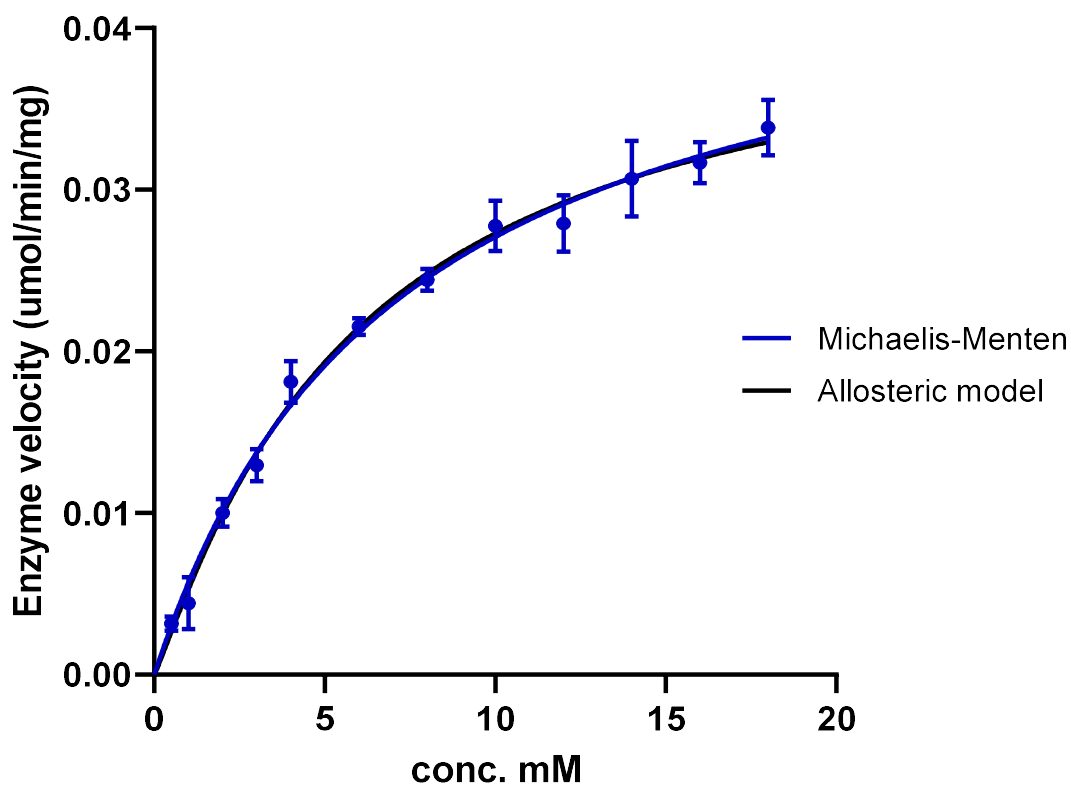
Michaelis-Menten	
Vmax	0.046 µmol/min/mg
Km	7.2 mM
Goodness of Fit	
R squared	0.9815
Constraints	
Km	Km > 0

Figure 5.8: Substrate concentration and enzyme velocity kinetics of glycoside hydrolase family 1 β-glucosidase GH1-4 monitored on spectrophotometer at pH 7.5 and at 25 °C, with Michaelis Menten non-linear regression curve fitted. Enzyme velocity as the number of µmols of p-nitrophenol released per minute. Error bars the standard deviation of replicates.

Discussion

The *Pichia pastoris* expression platform is well known for providing high levels of protein expression with simple purification protocols due to the yeast's highly active secretion pathway. Use of the methanol inducible alcohol oxidase promoters was also established to facilitate tightly regulated protein expression and to act as a locus for homologous recombination into *P. pastoris*. In this manner, pAOX1 was effective at integrating all four of the pD912 expression plasmids into the *P. pastoris* genome and three of the four candidate β-glucosidases (GH1-1, GH1-3 and GH1-4) expressed protein at high levels upon methanol induction. Protein of the final GHF1 candidate, GH1-2, was not detectable in the culture supernatant of the transformed strain. There was a protein band detected in the cell lysate that was close to the expected size of GH1-2, however, given this protein band was larger than the expected mass of GH1-2 and GH1-2 is not expected to be glycosylated, this is unlikely. An alternate explanation is that this band is a native protein with a 6xHis like motif producing a false negative in the immunoblotting. Isolation from the cell lysate and testing of the unknown band for β-glucosidase activity could potentially elucidate whether this protein has the expected enzyme identity.

There can be several reasons for failed protein production. However, given the other GHF1 β-glucosidase candidates were produced under the same expression system, it is improbable that transcription is not occurring upon methanol induction. More likely are post-transcriptional issues, such as poor transcript stability or proteolysis during processing. Transcript stability could be tested for by RT-qPCR and proteolysis reduced by testing expression in protease deficient strains.

As un-secreted GH1-3 and GH1-4 protein was observed in the cell lysates of the expression strains (Figure 5.4) improving protein secretion could also improve protein expression. This could be optimised by screening alternate secretion signal peptides. GH1-1 and GH1-3 are unstable over the 96 hr time course and future work to improve stabilisation of the protein would be necessary if the enzymes were targeted for commercial use. If this destabilisation

is the result of protease activity in *P. pastoris*, a protease deficient strain may be sufficient to ensure stability.

The β-glucosidase activity of GH1-1, GH1-3 and GH1-4 was evaluated against p-nitrophenyl-glucopyranoside, which forms glucose and p-nitrophenol when hydrolysed. In solution p-nitrophenol exists in equilibrium with the deprotonated form p-nitrophenolate which absorbs strongly at 400 to 420nm (Biggs, 1954) and this absorption was used to monitor the hydrolysis reaction. p-Nitrophenolate is the main form at pH ≥ 7.5 and as such, tests had to either be performed at pH ≥ 7.5 for direct monitoring or the pH had to be increased post reaction to force the formation of p-nitrophenolate, as was required for the pH and temperature tests. The substrate concentration kinetics were performed at pH 7.5 to allow for constant monitoring of the reaction. This pH is not optimal for GH1-1, GH1-3 or GH1-4, so it is expected that the observed maximum enzyme velocity may be increased if the experiments were performed at the optimum pH for each enzyme. This effect should however be limited, as the pH is not far outside of the optimal ranges for the three β-glucosidases and should also have no effect on the K_m and K_{half} estimates.

GH1-3 and GH1-4 are the most closely related by primary amino acid structure (Figure 5.1) and this is reflected in the response to pH and temperature which are almost identical except for overall activity per mg and K_m values. The three enzymes are suited to neutral to slightly acidic pH conditions and have significant decrease in activity below pH 5. GH1-1 and GH1-4 have almost no activity at pH 4 and GH1-3 loses its remaining activity by pH 3. This is a potential barrier to use in winemaking, especially in combination with low overall enzyme activity. GH1-1 has the lowest temperature optima at 20 °C, and GH1-3 and GH1-4 temperature optima is only moderately higher at 30 °C. These optima lie within wine fermentation temperature ranges.

Previously, GHF1 enzymes have been reported to function as dimers or more complex quaternary structures. As this could be a potential avenue for cooperativity non-Michaelis-Menten substrate kinetics of GH1-1 seem reasonable, although the precise cause is unknown.

93

β-glucosidase enzyme activity is of interest as a method to induce the release of organoleptic compounds from glycoside precursors in grapes by their addition during winemaking in order to modulate wine flavour. The β-glucosidases produced, GH1-1, GH1-3 and GH1-4, were poorly active in crude supernatant, with enzyme activities less than 0.1 U/mg observed consistently in all activity assays with p-nitrophenyl glucopyranoside. Purification of these enzyme from crude supernatant may improve activity. Other GHF1 β-glucosidases have demonstrated much high activities towards the p-nitrophenyl glucopyranoside substrate. However, as demonstrated in numerous studies (Biver et al., 2014, Lecas et al., 1991, Gruninger et al., 2014, Unno et al., 2014), the substrate affinity of β-glucosidases can be highly variable between different glucosides and can vary by orders of magnitude depending on the aglycone substituent. Therefore, it remains possible that GH1-1, GH1-3 and GH1-4 may have more viable levels of activity on alternate substrates, including grape flavour precursors . In future work, β-glucosidase activity by GH1-1, GH1-3 and GH1-4 on wine flavour precursors could by investigation by addition to grape fermentations or directly on purified wine flavour precursors in order to determine the specific enzyme kinetics. As GH1-1, GH1-3 and GH1-4, also had poor response to wine pH, testing of specific flavour precursors outside of wine-like such as with purified compounds may be more appropriate. Alternatively, if the structures of these enzymes could be resolved, the interactions between these β-glucosidases and various glycoside substrates could be modelled. If GH1-1, GH1-3 and GH1-1 have selectivity to either specific desirable or undesirable aglycones, such as in smoke tainted grapes, this may be more valuable than the poor suitability to the winemaking environment is a detriment as enzymes treatment could be targeted to specific flavours.

Conclusions

GH1-1, GH1-2, GH1-3 and GH1-4 were identified as potential glycoside hydrolase family 1 β-glucosidases from the MVGM gene predictions. GH1-1, GH1-3 and GH1-4 were successfully produced by *P. pastoris* and all demonstrated β-glucosidase activity on the p-nitrophenyl-glucopyranoside. GH1-1 and GH1-4 had similar responses to pH and temperature, with optimal activities at pH 6 and 30 °C. GH1-3 had optimal activity at pH 7 and 20 °C. However, all three showed significant loss of enzyme activity at pH 4, had low overall activity and poor substrate affinity. Future work should address overall activity as well as testing alternate substrates, such as purified flavour precursors which could potentially outweigh poor suitability to wine conditions, and as enzyme treatments during winemaking,.

Chapter 6: Heterologous expression of Polygalacturonases from GHF28 in *Pichia pastoris*

Introduction

Polygalacturonases (EC 3.2.1.15) hydrolyse polygalacturonic acid (homogalacturonan), the backbone structure of pectin, into small oligo-galacturonans and galacturonic acid. Pectin is a major component of plant cell walls and production of polygalacturonases by plants is important in fruit ripening (Prasanna et al., 2007, Gao et al., 2019, Ortega-Regules et al., 2007). Complete pectin degradation requires a mixture of activities including polygalacturonases, pectin lyases and pectin methylesterases. Microorganisms that inhabit the plant biome often produce pectinolytic enzymes to aid pathogenesis or to increase availability of nutrients (Girard et al., 2013, Armijo et al., 2016). Similarly, saprotrophic micro-organisms that degrade plant remains, such as those found in compost, produce these enzymes to breakdown the plant material (Zhou et al., 2017).

In industry, enzyme extracts from 'generally regarded as safe' fungal species, such as *Aspergillus niger*, are used to mediate the degradation of pectic substances and facilitate the clarification of fruit juices and to a lesser extent to enhance juice extraction (Jayani et al., 2005, Claus and Mojsov, 2018). However, these enzyme preparations have a history of off-target enzymatic activities (Claus and Mojsov, 2018, van Rensburg and Pretorius, 2000) and many enzymes are poorly adapted to winemaking conditions.

Pectin active glycoside hydrolases are typically classified as GHF28 and polygalacturonases have only ever been found from this enzyme family. GHF28 enzymes have been noted to have a distinct parallel (β)-helix architecture that is highly associated with pectinolytic enzymes (Petersen et al., 1997, Pickersgill et al., 1998) and typically use catalytic aspartic acid residues, as a nucleophile and proton donor pair to add water across the glycosyl bond and functionalise hydrolysis. These features make polygalacturonases an effective target for identification from predicted protein coding genes.

In order to assess potential suitability for winemaking applications, five putative polygalacturonases were selected from the gene predictions made from the CGM and MVGM metagenomes. The five candidates were targeted for heterologous protein expression in *Pichia pastoris*, and subsequently characterised.

Results

Heterologous expression of putative polygalacturonases in P. pastoris

Five putative polygalacturonases (GH28-1, GH28-2, GH28-3, GH28-4, and GH28-5) were selected from the genes predicted from the Chardonnay must and mixed varietal grape marc metagenomes (Chapter 3), with a mixture of bacterial and fungal origins (Table 6.1). The selected enzymes all contain homology to GHF28 and can be expected to be active on pectic substances. Candidate enzymes also had no significant (sequence identity <80%) matches to NCBI protein database, did not match to any characterised polygalacturonases in the NCBI protein database and cluster separately to each other by multiple sequence alignment and PhyML clustering (Figure 6.1). Potential signal peptides were detected in the amino acid sequences of GH28-1 and GH28-2 (Figure 6.2) and were removed from the coding sequences prior to gene synthesis.

Table 6.1: Enzyme homology and classification of putative polygalacturonases (GH28-1, GH28-2, GH28-3, GH28-4 and GH28-5) identified from CGM and MVGM metagenomes

Name	Origin	Size	Interpro domains	Closest NCBI match	Percent identity to other candidates
GH28-1	CM	439 aa (49kDa)	IPR000743: Glycoside hydrolase, family 28. Signal Peptide 1-21. GO:0004650 Polygalacturonase activity.	100%, 75% hypothetical protein [*Hortaea werneckii*] RMZ19198.1	GH28-2: 28%, GH28-3: 22%, GH28-4: 24%, GH28-4: 23%
GH28-2	CM	436 aa (47kDa)	IPR000743: Glycoside hydrolase, family 28. Signal Peptide 1-17. GO:0004650 Polygalacturonase activity.	99%, 59% Probable exopolygalacturonase (*Phialocephala subalpina*) CZR64156.1	GH28-1: 28%, GH28-3: 26%, GH28-4: 27%, GH28-5: 22%
GH28-3	MVGM	422 aa (45kDa)	IPR000743: Glycoside hydrolase, family 28. GO:0004650 Polygalacturonase activity.	95%, 57% exopolygalacturonase (Verrucomicrobia bacterium) WP_09055898.1	GH28-1: 22%, GH28-2: 26%, GH28-4: 27%, GH28-5: 29%
GH28-4	MVGM	458 aa (51kDa)	IPR000743: Glycoside hydrolase, family 28. IPR006626: Parallel beta-helix repeat. GO:0004650 Polygalacturonase activity.	96%, 76% glycoside hydrolase family 28 protein (Asticcacaulis sp. YBE204) WP_023463568.1	GH28-1: 24%, GH28-2: 27%, GH28-3: 27%, GH28-5: 26%
GH28-5	MVGM	461 aa (51kDa)	IPR000743: Glycoside hydrolase, family 28. GO:0004650 Polygalacturaonase activity.	96%, 33% hypothetical protein (Ktedonobacteales bacterium Uno3) GCE13494.1	GH28-1: 23%, GH28-2: 22%, GH28-3: 29%, GH28-4: 26%
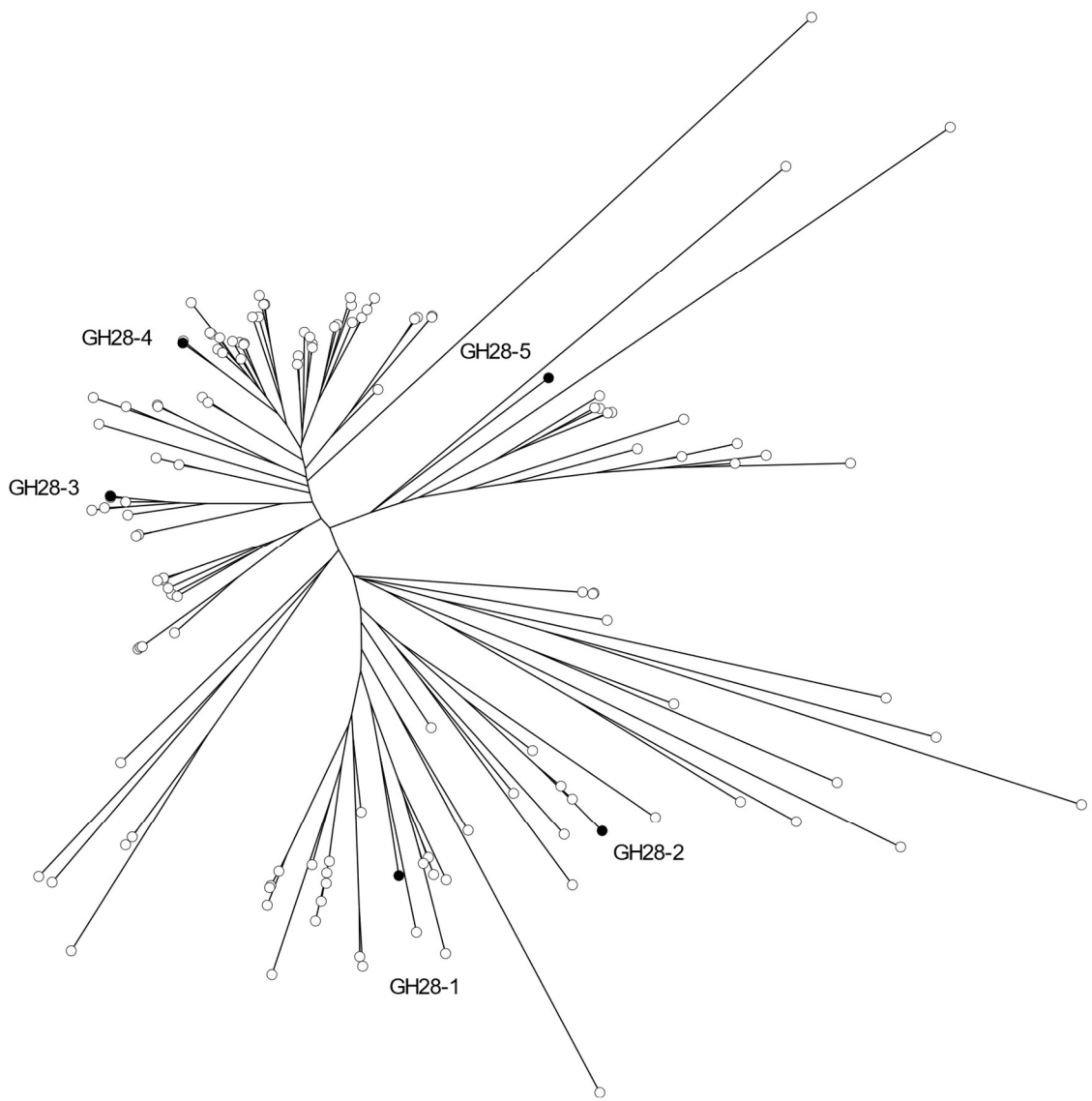
Figure 6.1: PhyML clustering of predicted proteins from the CGM and MVGM metagenomes containing GHF28 homologous domains with selected polygalacturonase candidates in black.
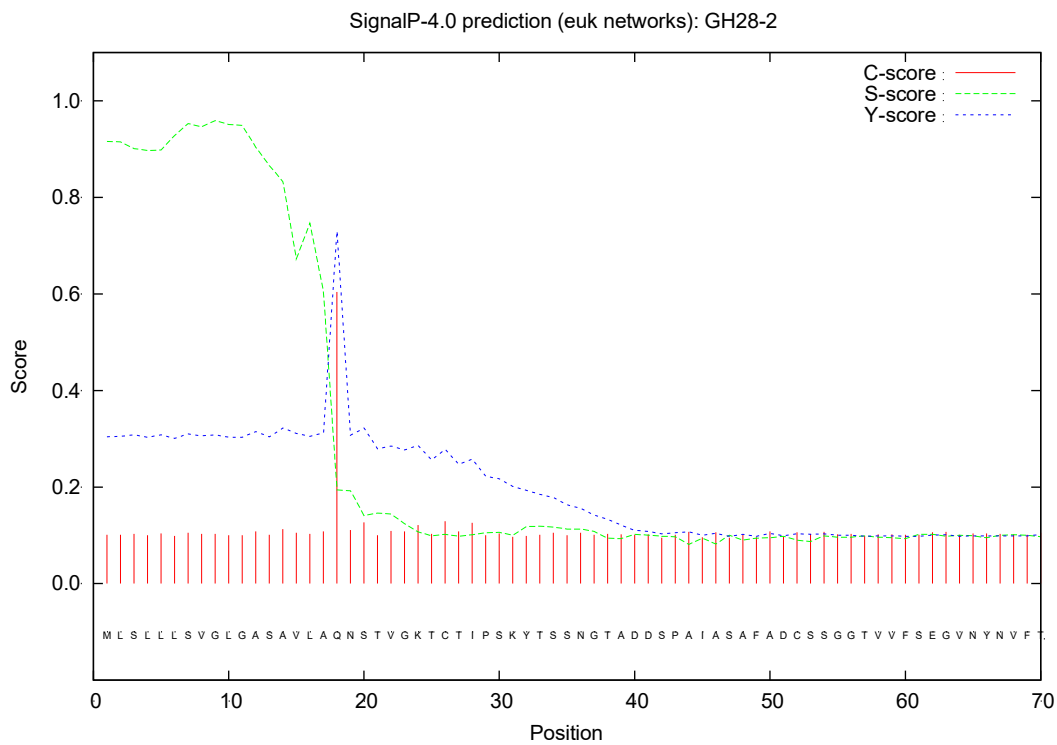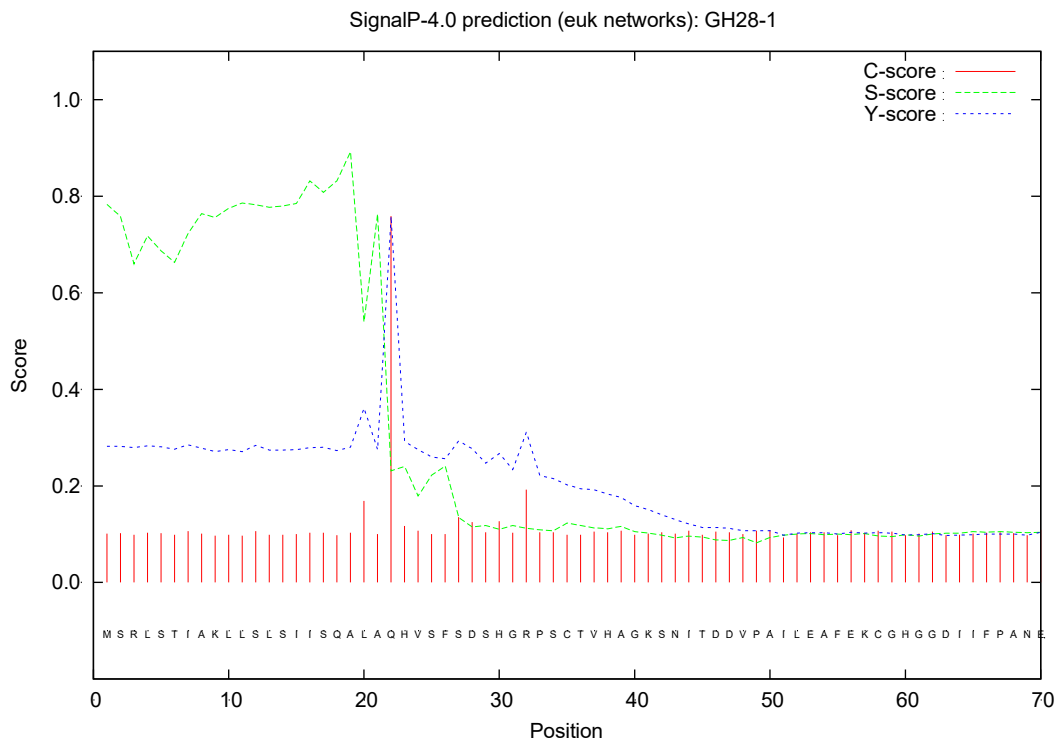
Figure 6.2: Signal peptide sites in GH28-1 and GH28-2 native amino acid sequences identified by SignalP 4.0 using the Eukaryotic model (euk networks) of signal peptides sites. Predicted cleavage sites are located at the intersection of C, S and Y-scores where total score is ≥ 0.5.

The coding sequences of the candidate polygalacturonases were optimised for expression in *P. pastoris* and the sequence for a 6xHis tag added to the C-terminus. The coding sequences were then cloned into the *P. pastoris* expression vector, pD912 using Gibson assembly (Gibson et al., 2009). Appropriate construction of the plasmids was confirmed by restriction digest (Figure 6.3). The *P. pastoris* expression vectors (pD912-GH28-1, pD912-GH28-2, pD912-GH28-3, pD912-GH28-4 and pD912-GH28-5) were transformed into a *P. pastoris* strain, BG11, with slow methanol utilisation. Genomic integration into the pAOX1 locus was confirmed by PCR of zeocin resistant isolates (Figure 6.4).
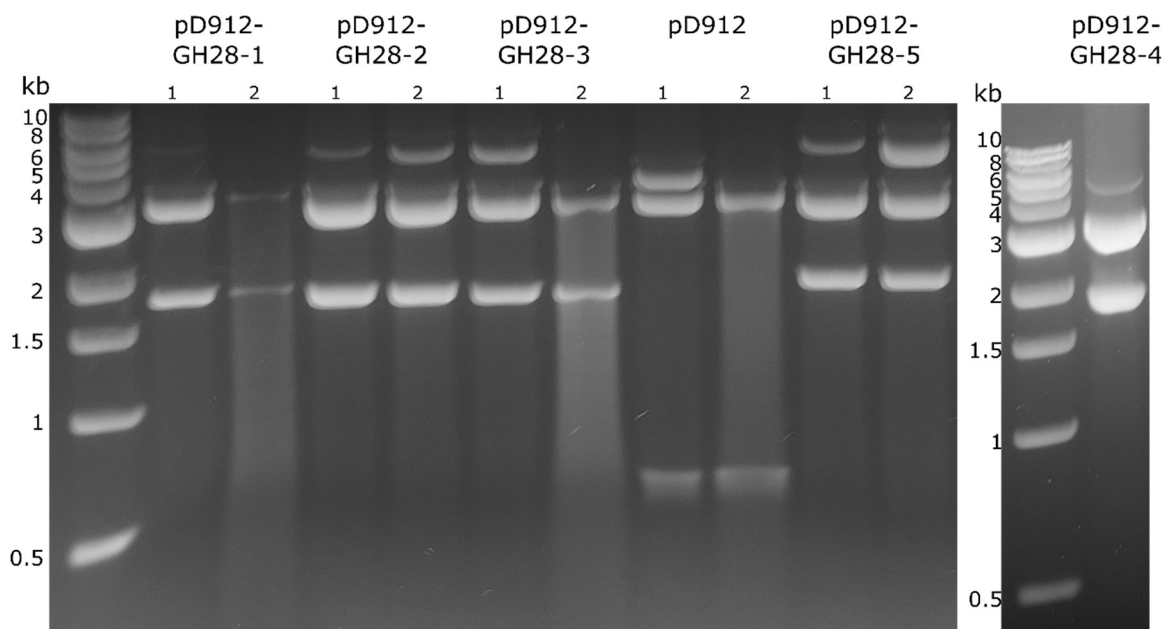
Figure 6.3: Confirmation of construction of *P. pastoris* expression plasmids, pD912-GH28-1, pD912-GH28-2, pD912-GH28-3, pD912-GH28-4 and pD912-GH28-5, by XhoI and NdeI double digests and visualised on 1% (w/v) agarose gel.
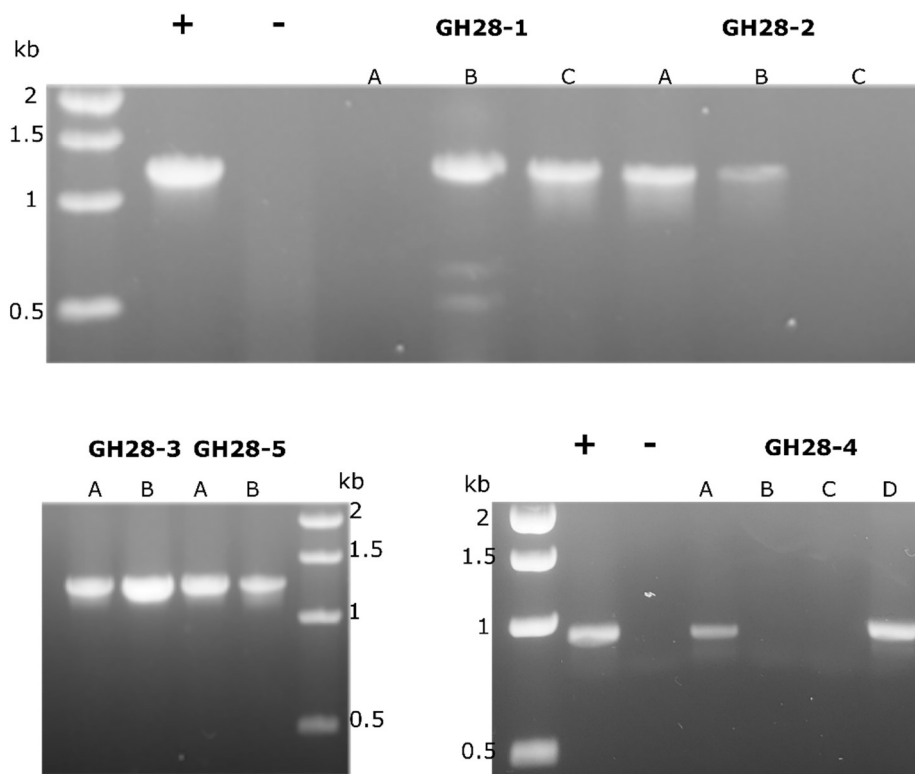
Figure 6.4: PCR of pAOX1 genomic integration sites of *P. pastoris* BG11 transformants with D912-GH28-1, pD912-GH28-2, pD912-GH28-3, pD912-GH28-4 and pD912-GH28-5. DNA products visualised on 1% (w/v) agarose gel. Expected amplicon sizes 1341 bp at 5' pAOX1 and 942 bp at 3' pAOX1.

Protein expression by recombinant *P. pastoris* strains

Once plasmid construction and genomic integration into the pAOX1 locus was confirmed, protein expression via methanol induction of the AOX1 promoter was assessed over a 96 hr time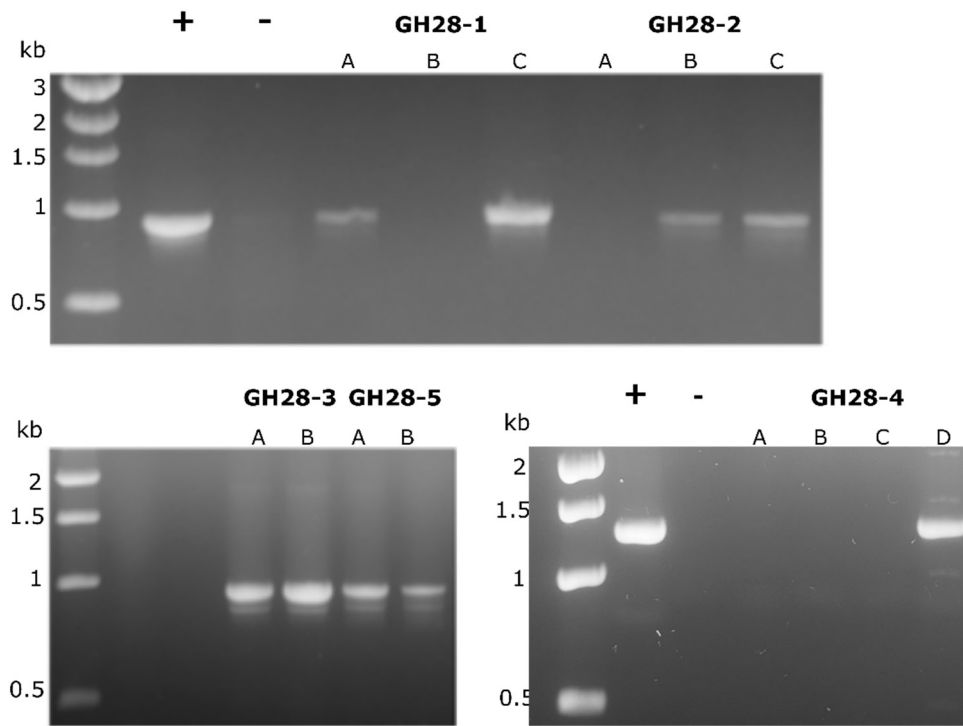 course using immunoblotting with an anti-His tag monoclonal antibody (Figure 6.5). The GH28-2 and GH28-3 proteins were produced at high levels (Figure 6.5 B and C), while lower levels of protein expression were observed for GH28-1 (Figure 6.5 A & D). However, no recombinant protein was detected in strains transformed with the GH28-4 and GH28-5 expression plasmids (Figures 6.5 D & E).

Immunoblotted protein bands for GH28-1 were observed at the expected mass (47 kDa) based upon the number of amino acids in the sequence. GH28-1 protein was not observed after 12 hours of induction, accumulates slightly from 24 hr to 96 hr.

The expected mass for GH28-2 is 48 kDa, however the protein in the recombinant strain appears to have a mass of approximately 80 kDa, suggesting that the protein is glycosylated. This observation correlates well with the presence of fourteen N-glycosylation consensus sites (Asn-X-Ser/Thr) in the GH28-2 amino acid sequence. GH28-2 appears to continue to accumulate in the supernatant over the full 96 hr time course (Table 6.2), increasing by 4.9-fold at 96 hr relative to 12 hr.

Two bands of similar but distinct sizes appear in the GH28-3 supernatants. These have a slightly higher mass than expected from translation of the ORF and may represent two different glycosylated forms of GH28-3. The intensity of both of these protein bands peaks at 48 hr and then decreases rapidly, with almost no full-sized protein observable after 96 hr. Smaller bands begin to appear after 12 hr and increase over time (Table 6.2). This data suggests that GH28-3 may be unstable in the culture supernatant.
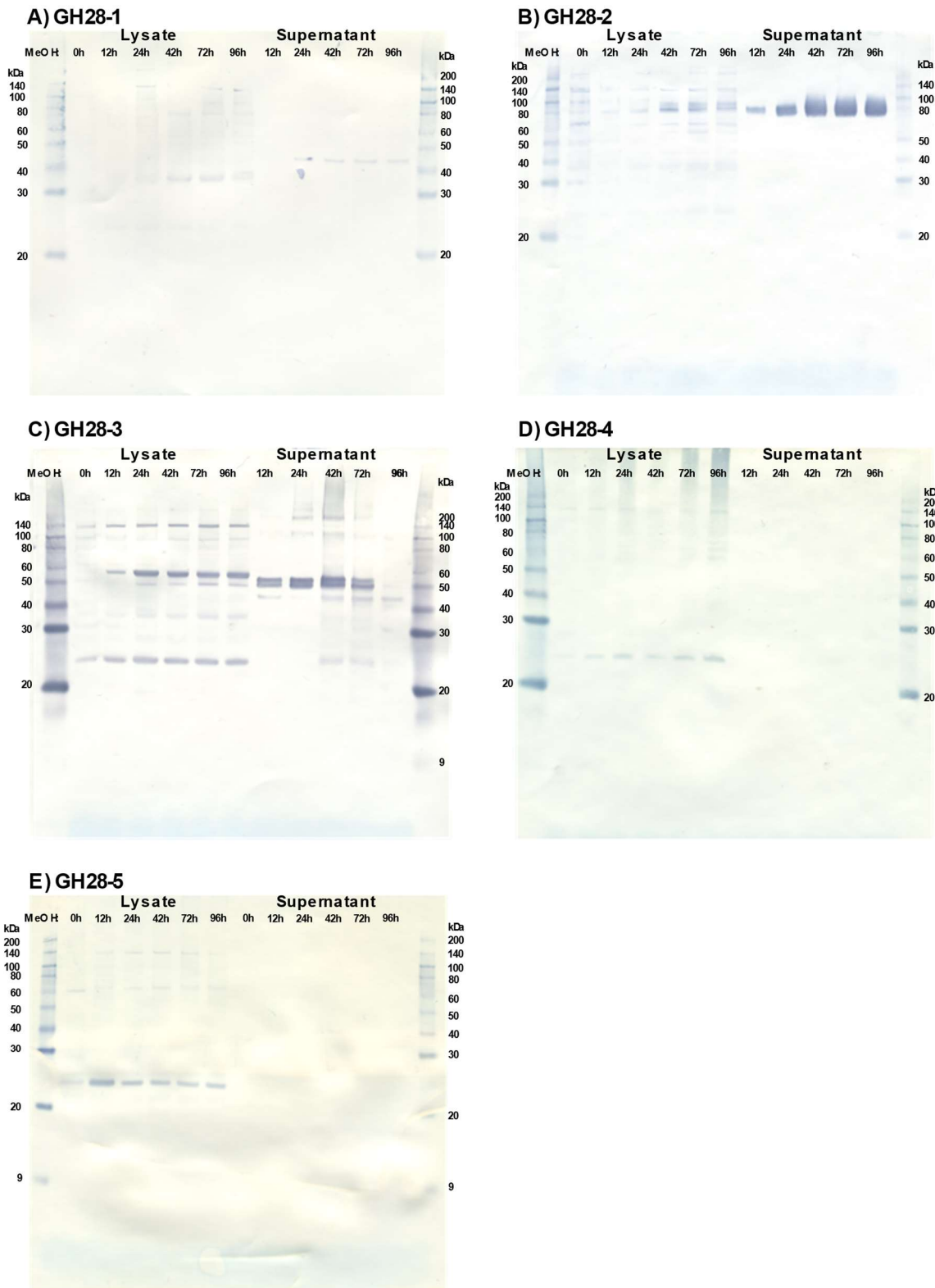
Figure 6.5: Protein expression post methanol induction of pAOX1 for putative polygalacturonases, GH28-1, GH28-2, GH28-3 and GH28-4, in *P. Pastoris.* Protein immunoblotted on nitrocellulose membrane by anti-His tag monoclonal antibody.

Table 6.2: ImageJ analysis of GH28-1, GH28-2 and GH28-3 induction in supernatants in western blot. RD =Relative density.

Time (hours)	GH28-1		GH28-2		GH28-3	
	Area	RD	Area	RD	Area	RD
12	0	0	8603	1.00	30865	1.00
24	5913	1.00	23404	2.72	41887	1.36
48	5761	0.97	39664	4.61	51485	**1.67**
72	7377	**1.25**	42443	**4.93**	30735	1.00
96	7269	1.23	42352	4.92	927	0.03

Enzymatic characterisation of heterologous GH28-1, GH28-2 and GH28-3

Crude supernatants of recombinant GH28-1, GH28-2 and GH28-3 were tested for polygalacturonase activity (polygalacturonic acid hydrolysis). Both GH28-1 and GH28-2 demonstrated measurable activity but no activity was detected for GH28-3. GH28-1 was shown to be less active (per milligram of protein mass) than GH28-2. The pH (Figure 6.6) and temperature (Figure 6.7) optima of GH28-1 and GH28-2 were also determined. Both enzymes are active over a large pH range. GH28-1 retained most of its activity across a pH range from pH 3.5 to pH 6.5, while GH28-2 activity remained high across a pH range from 4 to 6. GH28-1 was most active between pH 4.5 and pH 5.5. GH28-2 displayed optimal activity at approximately pH 5. GH28-1 retained more activity at low pH, with approximately 35% of maximum activity observed at pH 3.0, whereas GH28-2 retains 60% activity at pH 3.5, but very reduced activity (1.7%), at pH 3.2.

GH28-1 and GH28-2 display very different responses to temperature. GH28-1 was most active at 20 °C and retained most of its activity across the temperature range of 10 °C to 30 °C. In comparison, GH28-2 was most active at 60°C and retained 50% activity between 40 °C and 70 °C.

.
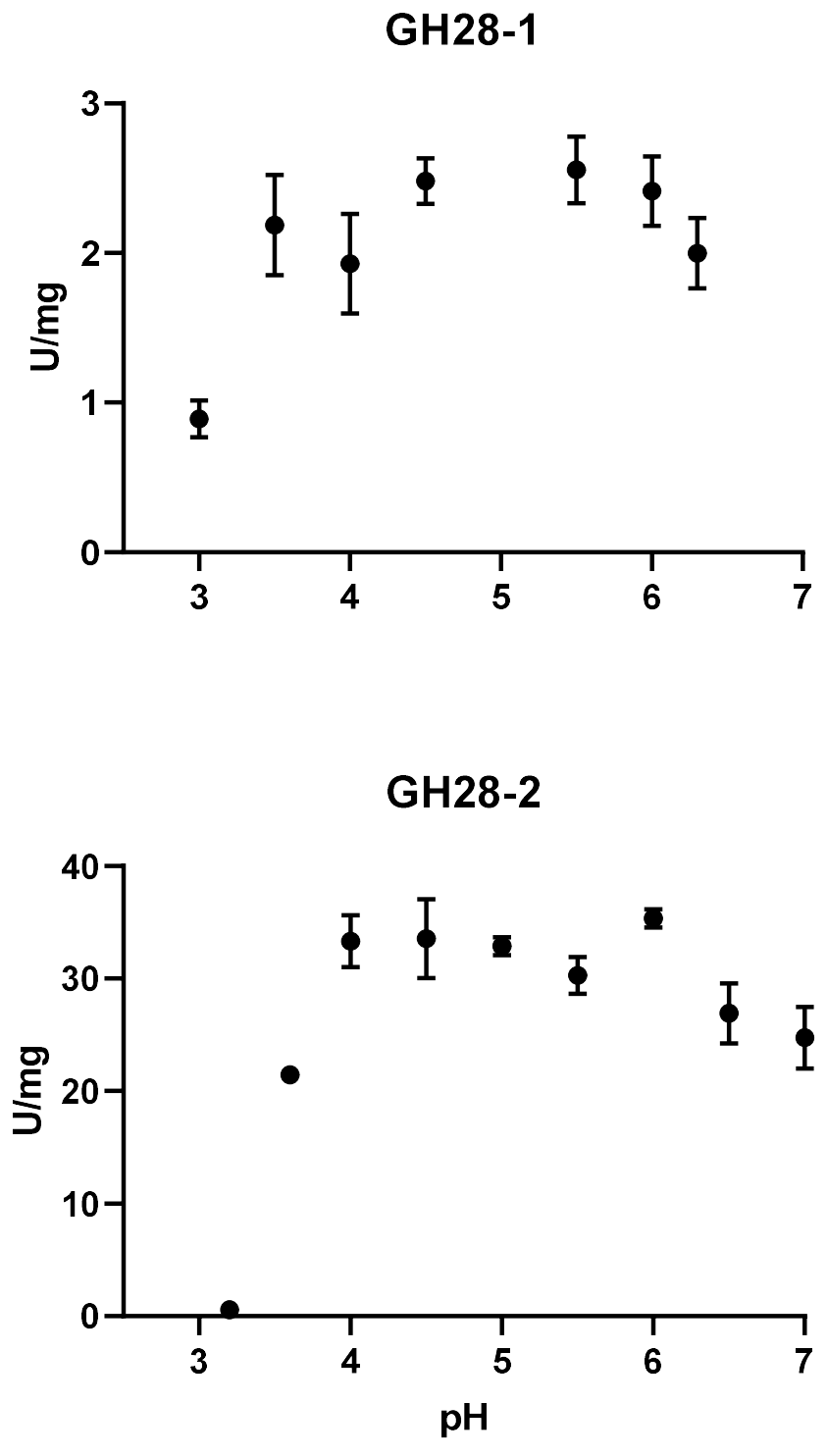
Figure 6.6: Response of *P. pastoris* expressed GH28-1 and GH28-2 polygalacturonase activity to pH.
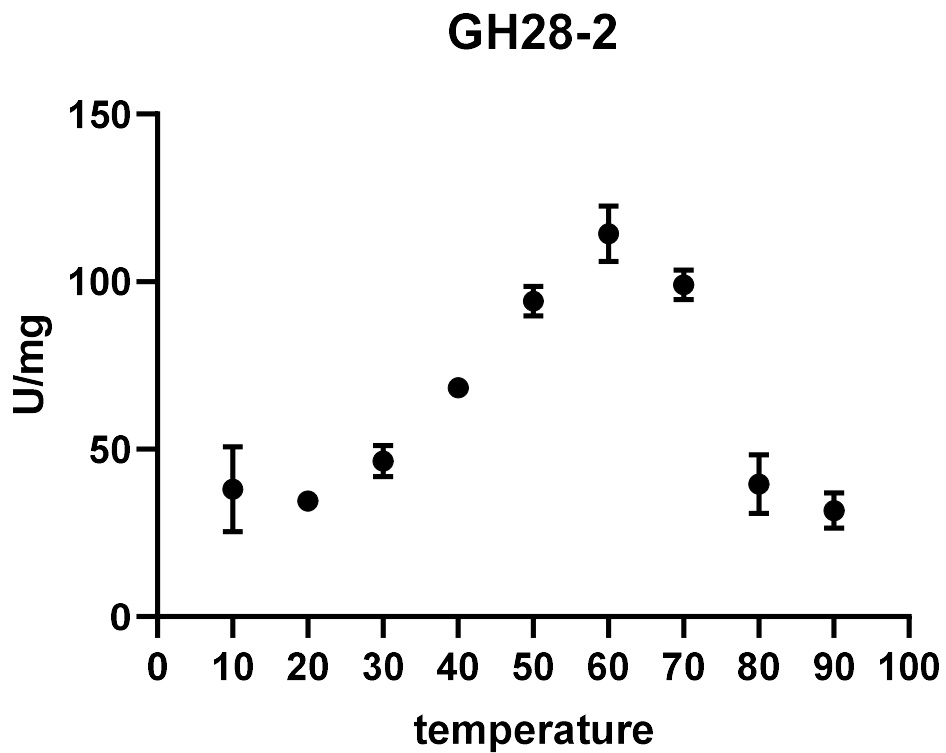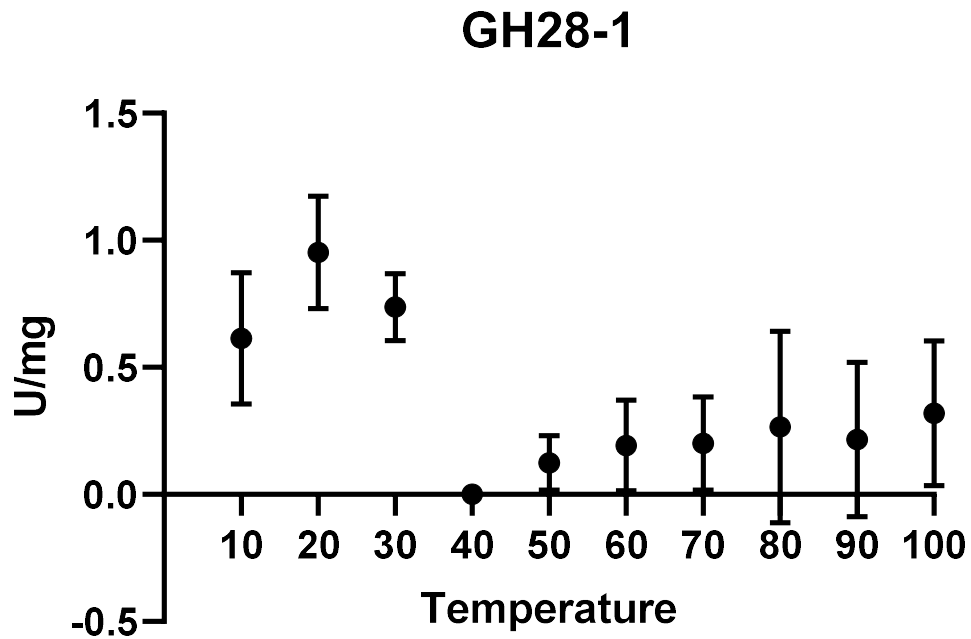
Figure 6.7: Response of *P. pastoris* expressed GH28-1 and GH28-2 polygalacturonase activity to temperature.

The change in velocity of GH28-2 according to increasing substrate concentrations was also determined. Non-linear regression of GH28-2 activity showed a better fit to the allosteric sigmoidal regression model than to standard Michalis-Menten kinetics (Figure 6.8). From the best fit of allosteric sigmoidal non-linear regression GH28-2 maximum enzyme velocity (V_{max}) is 137.6 µmol/min/mg, Hill's coefficient of 2.24, substrate concentration at half maximal enzyme velocity (K_{half}) is 1.11 mg/ml and K_{prime} is 1.27 mg/ml.

The GH28-1 substrate kinetics could not be determined due to low activity.

GH28-2 substrate kinetics

Allosteric sigmoidal Model	
V_{max}	137.6 µmol/min/mg
h	2.237
K_{half}	1.11 mg/ml
K_{prime}	1.27 mg/ml
Goodness of Fit	
R squared	0.8955
Constraints	
h	h > 0
K_{half}	K_{half} > 0
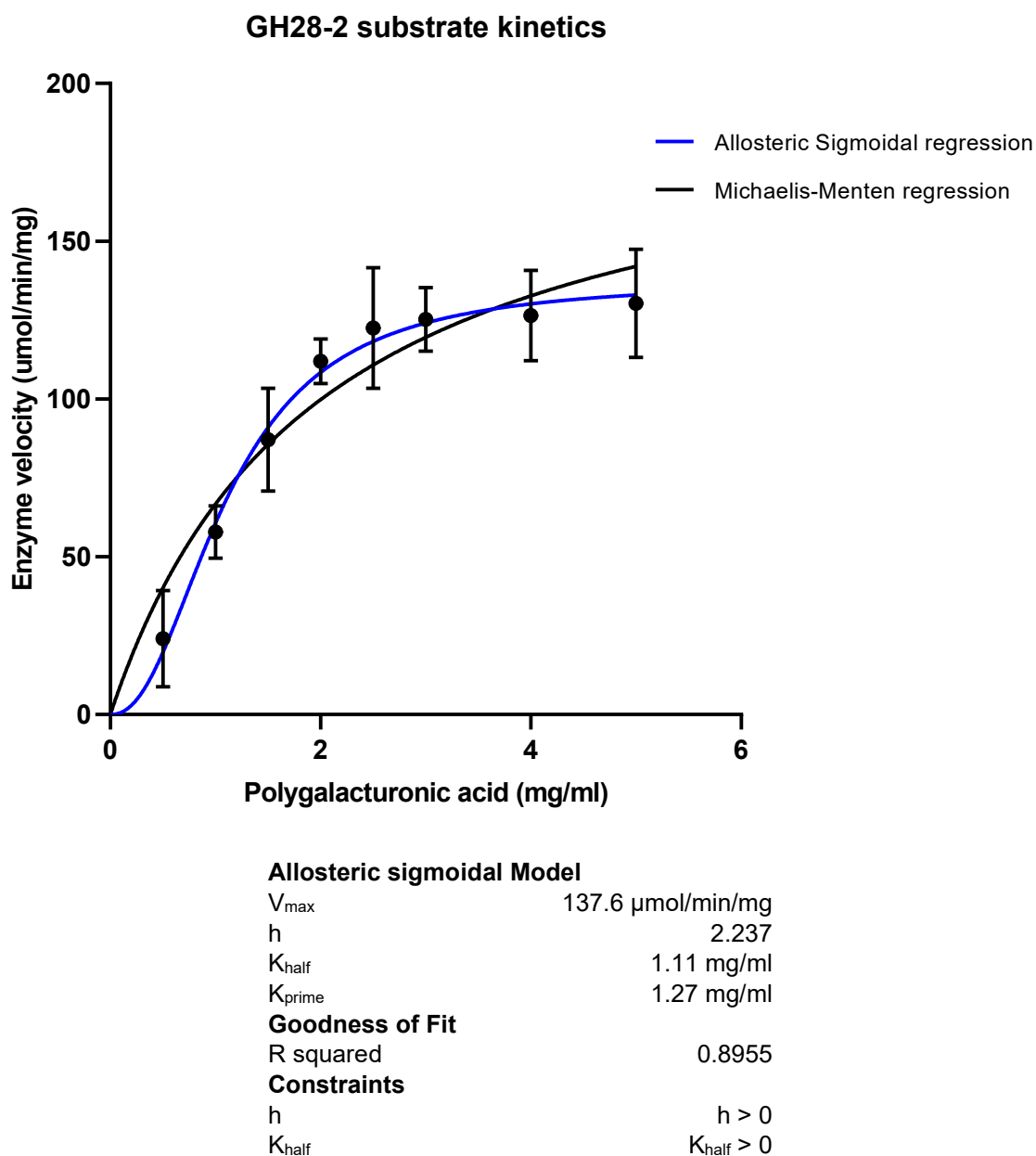
Figure 6.8: Polygalacturonic acid concentration and enzyme velocity kinetics of GH28-2.

Discussion

Three of the five putative polygalacturonases were successfully expressed in *P. pastoris*. The heterologous expression of GH28-2 in *P. pastoris* appears near optimal, with high levels of protein being secreted that accumulates over 96 hr and displays very little non-secreted protein in the lysate. This indicates efficient secretion by the modified alpha-factor signal peptide for this protein.

In comparison, the expression of GH28-1 and GH28-3 is suboptimal. GH28-1 has significantly lower levels of secreted protein than GH28-2. GH28-1 does not appear to be accumulating in the cell lysate and therefore secretion is not the limiting factor. There are also no obvious signs of protein degradation by immunoblot. There are two protein bands with mass close to the protein mass expected for GH28-3. It is not clear if the smaller band is a degraded version of the larger. Both these protein bands reduce in intensity over the 96 hr assay period, with smaller products accumulating as the full-length GH28-3 bands diminish. As such, the GH28-3 protein appears to be unstable in the supernatant and may be a target of proteolytic activity. The strains for all three enzymes have not been tested for overall proteolytic activity and testing the expression system in a protease deficient *P. pastoris* strain would be an appropriate next step to ensure if proteins are being targeted for degradation. mRNA production and stability have not been tested and are another source of potential issues. However, given lack of activity, improvements to GH28-3 may not be worth pursuing.

Improving protein production of GH28-2 would require increasing the overall amount of protein expression, potentially by increasing gene copy number and or introducing stronger promoter sequence. However, this would only be effective if secretion remains efficient and can keep pace with mRNA translation.

Polygalacturonase activity is the hydrolysis of alpha-1,4 glycosidic bonds between galacturonic acid residues, in either endo or exo configuration, depending on whether the depolymerisation occurs either randomly across the length of the molecule or terminally. Endo type activity is the preferred mechanism for winemaking applications as it has a more

immediate effect on polymer length and solubility and therefore more immediate effect on juice extraction and clarification. The closest matching NCBI protein to GH28-2 is a predicted exo-polygalacturonase (Table 6.1) which may indicate the enzyme is more likely to produce exo-polygalacturonase activity. However, neither endo- nor exo- type activity was predicted by InterProScan for GH28-1 nor GH28-2.

Endo- type of activity can be more difficult to characterise as there are a diverse range of product sizes and is complicated by uneven methylation and acetylation of polygalacturonic acid from naturally derived sources. However, regardless of endo- or exo- type activity, polygalacturonase hydrolysis produces additional reducing sugar residues which are typically used to measure activity. The assay used measures the amount of 3-amino,5-nitro salicylic acid produced by reduction of 3,5-Dinitrosalicylic acid (DNS) at high temperature in comparison to reduction by various levels of galacturonic acid. Consequently, enzyme activity (U) is defined as the number of μmols of galacturonic acid equivalents produced per minute and is underpinned by the assumption that reduction by a galacturonic acid molecule and a single oligo-galacturonan will be near equivalent under assay conditions.

Both GH28-1 and GH28-2 demonstrated polygalacturonase activity by the DNS assay and have potential as winemaking adjuncts. GH28-1 and GH28-2 differ significantly in activity per mg or protein, with GH28-2 being more than 10-fold more active under most pH and temperature conditions. However, GH28-1 outperforms GH28-2 at very low pH and is better suited to low temperatures. Wine grape juice can be expected to be below pH 4 and is often further reduced by winemakers through the addition of tartaric acid, to aid in preventing the growth of spoilage micro-organisms (Godden et al., 2015). GH28-2 retains 50% of activity at pH 3.5, but less than 2% by pH 3.2. This rapid loss of activity may be a barrier to its use in grape juices with pH levels below pH 3.5. GH28-1 also begins to show a decline in activity at low pH but this decline less significant at pH 3.5. Given, polygalacturonase activity is initiated by paired general acid and general base interactions which are supplied amino residues, loss of activity due to changes of pH may be due to changes to these residues. The steady decline in activity at low pH may be a result of excessive protonation at these

sites preventing function as a general acid. However, this is not reflected in GH28-1 and GH28-2 predicted isoelectric points (Table 6.1), of which GH28-1 is higher where it would be expected to be lower. If active site protonation is the main cause in loss of activity, activity is being preserved despite predicted net protein charge. Protein modelling and mutagenesis experiments could potentially elucidate this pH response.

Winemaking temperatures are generally between 20 °C and 30 °C for red wines and ≤15 °C for white wines, to which GH28-1 is better suited. However, although reduced, GH28-2 still retains significantly higher activity at these temperatures than GH28-1. Temperature does not appear to be a barrier to these enzymes for use in winemaking.

Numerous other polygalacturonases have been isolated from fungal sources for winemaking or other applications (Kant et al., 2013, Sahay et al., 2013, Trindade et al., 2016, Merin and Morata de Ambrosini, 2015, Cheng et al., 2016). Some of these have the potential to outperform GH28-1 and GH28-2 at pH 3.5 or at low temperatures or both (Merin and Morata de Ambrosini, 2015, Sahay et al., 2013, Trindade et al., 2016). Significant activity below pH 3.5 is far less common (Cheng et al., 2016, Kant et al., 2013). GH28-2 has similar temperature optima and response to low pH as the thermostable exo-polygalacturonase identified from *Rhizomucor pusillus* A13.36 (Trindade et al., 2016). Moreover, both enzymes demonstrate allosteric sigmoidal substrate kinetics. The underlying mechanism for these sigmoidal kinetics remains unresolved. Polygalacturonases are generally monomeric with a single binding site and as such, allosteric behaviour GH28-2 would be expected to be unlikely, as this type of enzyme behaviour would normally be caused by cooperativity between subunits. It is unknown whether this kinetic behaviour will have significant impact on polygalacturonase efficacy in winemaking applications. As Trindade and colleagues have not reported a maximum velocity (V_{max}) or half maximal enzyme velocity substrate concentration (K_{half}) it is not possible to compare the enzymes via these parameters.

Polygalacturonases have been reported with a range of substrate affinities (K_m), including a particularly low K_m of 0.083 mg/ml from an *Aspergillus niger* polygalacturonase (Kant et

al., 2013) and a much higher K_m of 5 mg/ml from a *Neurospora crassa* polygalacturonase (Polizeli et al., 1991). Although not identical, substrate affinity (K_m) and K_{half} both represent the substrate concentration at which an enzyme reaches its half maximal enzyme velocity and are somewhat comparable. With a K_{half} of 1.11 mg/ml, GH28-2 is similar to polygalacturonases identified from an Indian forest soil metagenome (K_m 1.685 and 1.542 mg/ml; (Sathya et al., 2014), and a polygalacturonase from *Penicillium oxalicum* CZ1028 (K_m 1.27 mg/ml; (Cheng et al., 2016).

In Shiraz wine grapes, the total mass of pectin ranges between 0.7% and 0.8% of total grape berry weight (Silacci and Morrison, 1990). In other grape cultivars pectin abundance ranges between 0.12% and 0.17% berry weight (Kawabata et al., 1974) and soluble pectin has been reported at between 0.09% and 0.28% in non-wine grape cultivars (Baker, 1997). Not all of the pectin is expected to be extracted into the grape juice, nor is all pectin comprised of homogalacturonan (the polygalacturonic acid portion). Although substrate levels will vary, it is expected that in most juices, homogalacturonan concentrations will be less than the K_{half} value of GH28-2 and enzyme velocity will be less than half maximal.

GH28-1 pH response is most similar to the polygalacturonase isolated from *A. niger* MTCC 3323 (Kant et al., 2013) but has vastly different temperature optima and lower activity.

As glycoside hydrolases GH28-1 and GH28-2 may exhibit activity on other substrates such as rhamnogalacturonans, xylogalacturonans, other grape cell wall derived polysaccharides or even grape glycosides. Ideally future experiments would test for possible side activities which may or may not be desirable in winemaking.

Although, GH28-1 and GH28-2 are not perfectly adapted to winemaking conditions, if the enzymes remain stable in grape juice and wine and are not inhibited by other compounds present during fermentation, the enzymes can have a long period to act throughout fermentations. *P. pastoris* expressed GH28-1 and GH28-2 have not been tested for long term stability at low pH nor inhibition by various potentially inhibiting wine grape compounds but is the basis for experiments in *S. cerevisiae* in Chapter 7.

Conclusions

Polygalacturonase activity is of oenological interest in order to aid in the breakdown of pectic substances in grape juice to aid in juice extraction and clarification, by softening plant cell walls and increased pectin solubility. Of the five putative polygalacturonases targeted for heterologous protein expression, GH28-1, GH28-2 and GH28-3 were successfully expressed. GH28-2 has strong expression in *P. pastoris* and accumulated over the period of induction. GH28-1 expression was significantly less prolific and GH28-3 expression was unstable. Both GH28-1 and GH28-2 demonstrated polygalacturonase activity, producing reducing sugar residues from polygalacturonic acid by the DNS. However, GH28-3 was not active. At optimum conditions (pH 5 and 60 °C), GH28-2 is more than 10-fold more active than GH28-1. GH28-1 has optimal activity between pH 4.5 - 5.5 and 20 °C but also retains much of its activity at wine pH. The response of GH28-1 to temperature and pH and the overall high levels of activity of GH28-2 make these polygalacturonases candidates for further experiments to test activity under wine-like conditions.

Chapter 7: Heterologous expression of Pectinase from GHF28 in *Saccharomyces cerevisiae*

Introduction

Despite many *S. cerevisiae* strains containing genes that encode polygalacturonases, *S. cerevisiae* typically displays little pectinolytic activity during the winemaking process (Eschstruth and Divol, 2011, Blanco et al., 1998). To overcome this lack of enzyme activity, crude extracts that display pectinase activity are commonly used, which typically contain a mixture of activities that can include polygalacturonase, pectin lyase and pectin methylesterase activities, in addition to other off-target activities. Some studies have demonstrated that these combined activities can alter the phenolics, flavour and colour of wine (Belda et al., 2016a, Sacchi et al., 2005). However, pectinase treatments are also commonly used during clarification (Nordestgaard, 2019).

Although improvements have been made, commercial enzyme preparations are not specifically adapted to winemaking conditions and can contain undesirable off-target activities (Kashyap et al., 2001, Claus and Mojsov, 2018). Wine is a complex and challenging environment in regard to enzyme activity, with low pH, high sugar concentrations, increasing concentrations of ethanol and numerous other potentially inhibitory substances, such the preservative sulphite. Reduced enzyme activity under these conditions is difficult to prevent or control. Classically, enzyme preparations are derived from fungal species which produce the desired secretory enzymes and have been selected to produce large titres of enzyme under industrial conditions, generally by solid state fermentation. Off-target activities typically arise from numerous other secretory enzymes that are also produced during the fermentation and which co-purify during the production of the crude extract.

Two polygalacturonases (GH28-1 and GH28-2) that were identified from the 2016 Chardonnay must metagenome (Chapter 3) were demonstrated to have at least partial retention of activity at wine pH and winemaking temperatures when expressed as secreted

proteins by the *P. pastoris* heterologous expression system (Chapters 6). On this basis, GH28-1 and GH28-2 were determined to have potential for application during wine fermentation. In order to test the two polygalacturonases under wine-like conditions, GH28-1 and GH28-2 were expressed in *S. cerevisiae,* to facilitate fermentation experiments and to facilitate enzyme introduction to the winemaking process without requiring enzyme additions.

Results

Heterologous production of GH28-1 and GH28-2 in *Saccharomyces cerevisiae*

Heterologous expression of GH28-1 and GH28-2 in *S. cerevisiae* was constructed to facilitate *de novo* production during wine fermentation. As *P. pastoris* and *S. cerevisiae* share highly similar codon usage (Sinclair and Choy, 2002), the *P. pastoris* codon optimised GH28-1 and GH28-2 coding sequences were used in preparing the *S. cerevisiae* expression plasmids. GH28-1 and GH28-2 coding sequences were cloned into the expression vector, pCVSα, by Gibson assembly (Gibson et al., 2009). Correct construction of pCVSα-GH28-1 and pCVSα-GH28-2 were confirmed by restriction digest (Figure 7.1 A). pCVSα contains two *HO* homologous sites to facilitate genomic integration. pCVSα-GH28-1 and pCVSα-GH28-2, contained GH28-1 and GH28-2 as a fusion with a modified alpha factor secretion signal to facilitate protein secretion. As pCV3-natR does not contain the alpha factor, the GH28-1 and GH28-2 expression cassettes were excised from pCVSα-GH28-1 and pCVSα-GH28-2 by restriction digest and cloned into pCV3-natR for episomal protein expression. Construction of pCV3S-GH28-1 and pCV3S-GH28-2 was confirmed by restriction digest (Figure 7.1 B).

Both the genome-integrating and 2 μm episome based expression plasmids (pCVSα-GH28-1, pCVSα-GH28-2, pCV3S-GH28-1 and pCV3S-GH28-2) were transformed into AWRI1631, a haploid derivative of an *S. cerevisiae* wine strain (Borneman et al., 2008). Genomic integration into the *HO* locus by pCVSα-GH28-1 and pCVSα-GH28-2 could not be confirmed by PCR. However, transformants containing pCV3S-GH28-1 (AWRI4241) and pCV3S-GH28-2 (AWRI4240) were confirmed (Figure 7.2).

The recombinant strains, AWRI4241 and AWRI4240, were used to ferment synthetic complete media with additional glucose (100 g/L). Low levels of GH28-2 and very low levels of GH28-1 were detected by Anti-His tag immunoblot of 25-fold concentrated supernatant of the new *S. cerevisiae* strains. (Figure 7.3).
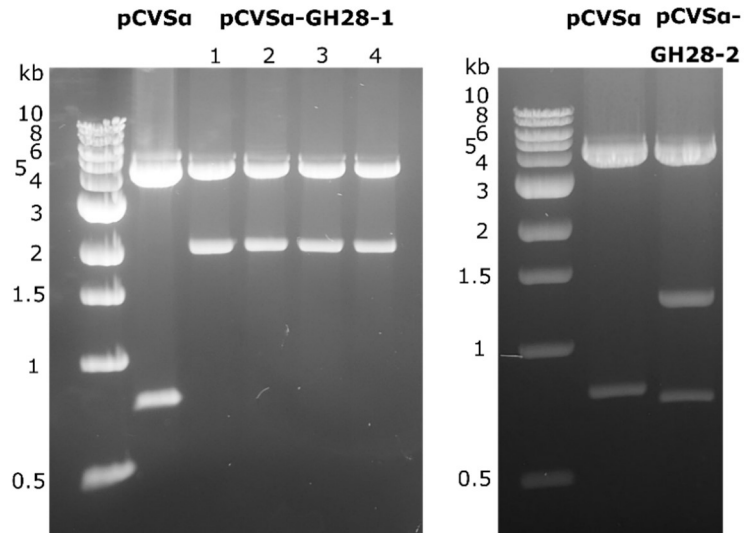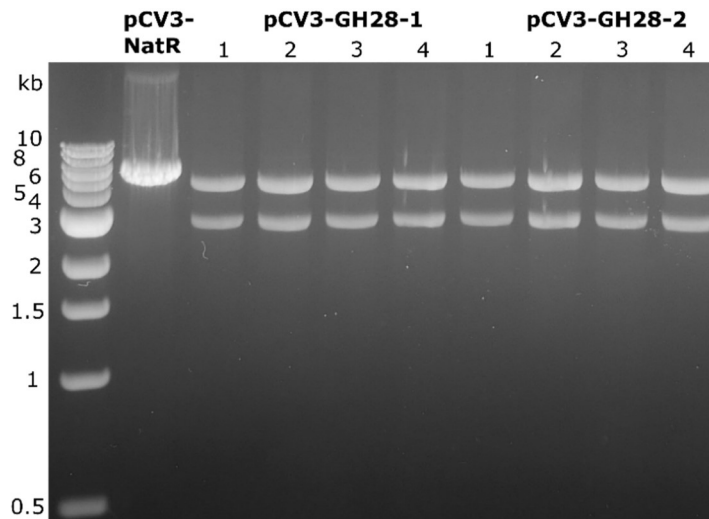
Figure 7.1: Confirmation of cloning of A) pCVSα-GH28-1 and pCVSα-GH28-2 and B) pCV3S-GH28-1 and pCV3S-GH28-2 plasmid DNA by *Eco*RI digest and visualised on 1% (w/v) agarose gel. Expected band sizes: pCVSα - 803bp + 4652bp, pCVSα-GH28-1 - 4652bp + 2057bp, pCVSα-GH28-2 - 4652bp + 1302bp + 758bp pCV3 – 5860bp, pCV3-GH28-1 – 2785bp + 4551bp, pCV3-GH28-2 – 2789bp + 4551bp.
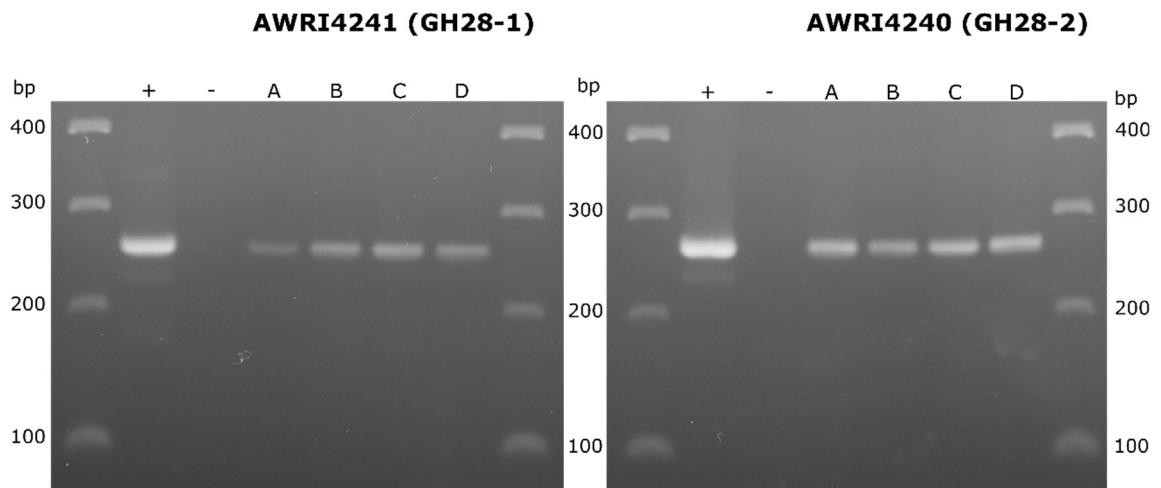
Figure 7.2: PCR amplification of GH28-1 and GH28-2 266 and 275bp amplicons in *S. cerevisiae* AWRI1631 transformed with pCV3S-GH28-1 and pCV3S-GH28-2 and visualised on 1% (w/v) agarose.
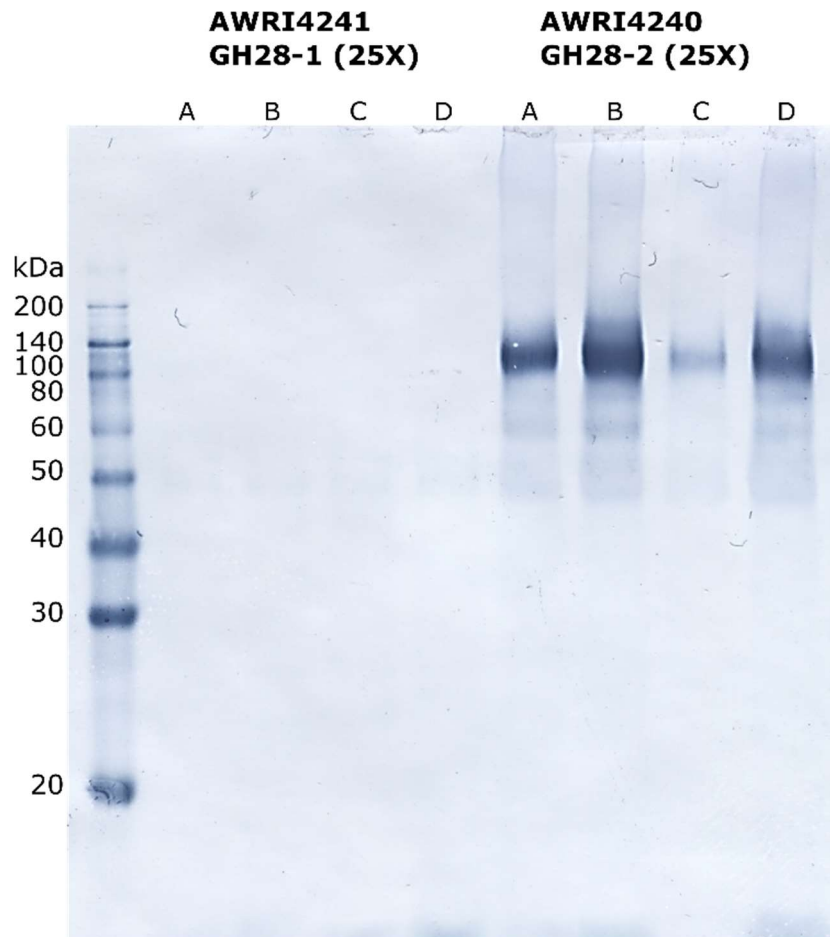
Figure 7.3: Anti-His tag immunoblotted nitrocellulose membrane visualising heterologous protein production of putative polygalacturonases GH28-1 and GH28-2 by recombinant *S. cerevisiae* strains, AWRI4241 and AWRI4241 in 25-fold concentrated supernatant.

Fermentation with recombinant GH28-1 and GH28-2 strains

The parental strain (AWRI1631), and the recombinant strains AWRI 4241 (GH28-1) and GH28-2 (AWRI4240), were used to ferment synthetic grape must that had been supplemented with polygalacturonic acid (2 g/L). AWRI1631 completed fermentation after 4 days post inoculation (Figure 7.4). However, the recombinant strains AWRI4241 and AWRI4240 displayed reduced fermentation rates. All three AWRI4241 replicates completed fermentation after 7 days and only one AWRI4240 replicate appeared to complete fermentation after 8 days. The remaining two AWRI4240 replicates reaching at 23 g/L and 40 g/L residual sugar and reduced fermentation rate at 8 days.

The production of galacturonic acid occurs from exo-polygalacturonase type cleavage of the monomeric sugar but can be achieved by both endo- and exo-polygalacturonases. Galacturonic acid levels were monitored via an enzymatic assay (Figure 7.5). There appeared to be a residual amount of galacturonic acid in the solution prior to the addition of yeast. This may have been produced from dissolving the polygalacturonic acid into the synthetic grape must. All three strains produced galacturonic acid in the first 12 hr of fermentation. AWRI4240 produced the highest concentration of galacturonic acid of 0.58 g/L, with AWRI4241 producing a maximum of 0.45 g/L galacturonic acid. In comparison, the AWRI1631 ferments reached a maximum of 0.37 g/L across replicates. On average, the recombinant strains AWRI4241 and AWRI4240 both showed higher concentrations of galacturonic acid than the parental strain. As fermentation progressed galacturonic acid levels reduced to undetectable levels within 4 days for all three strains.
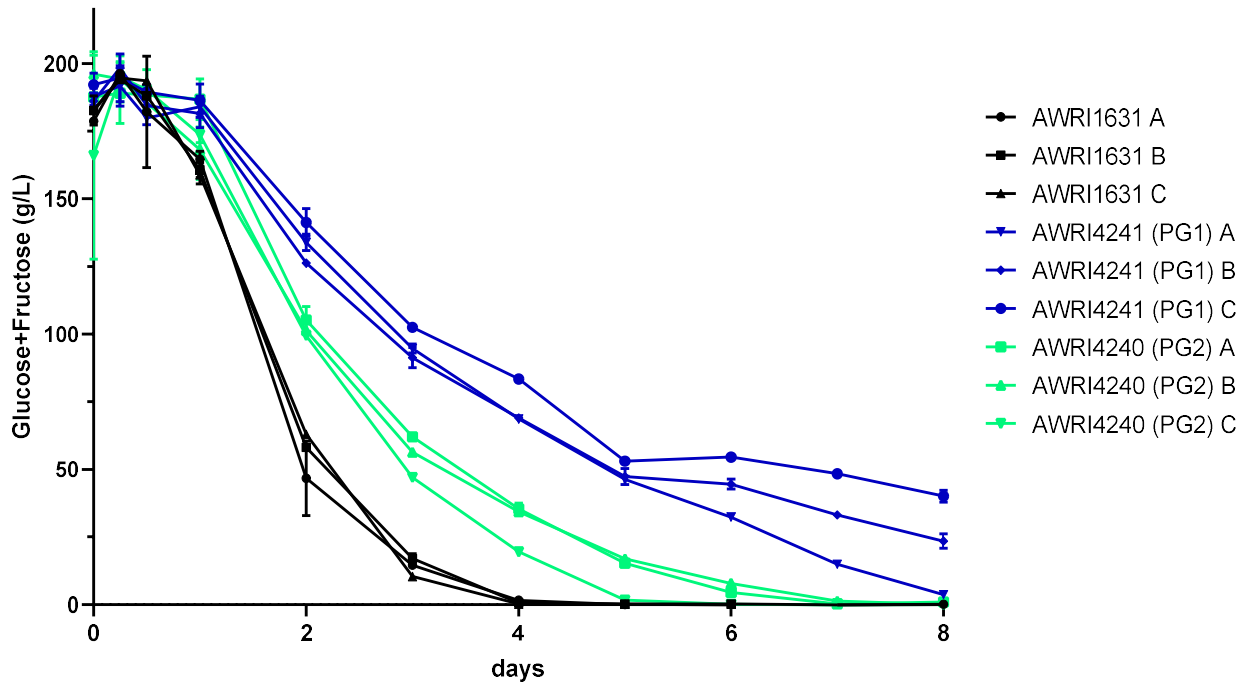
Figure 7.4: Sugar utilisation by AWRI1631 and recombinant polygalacturonase strains, AWRI4241 and AWRI4240 during fermentation of synthetic grape must at 22 °C.
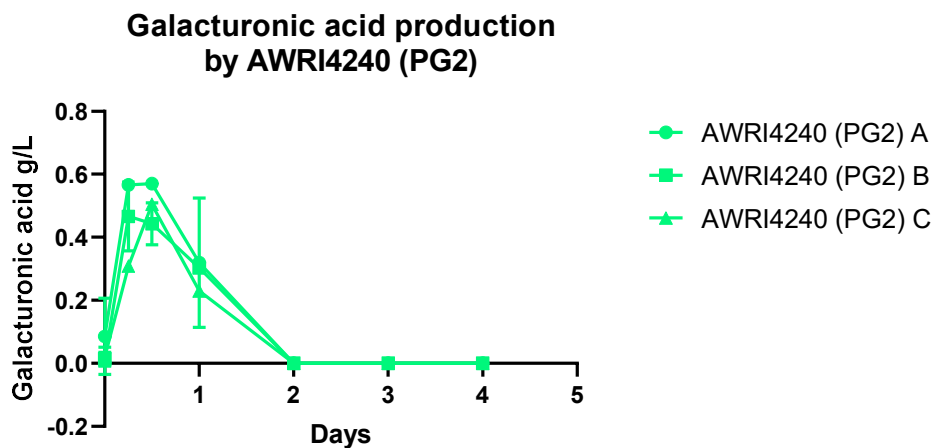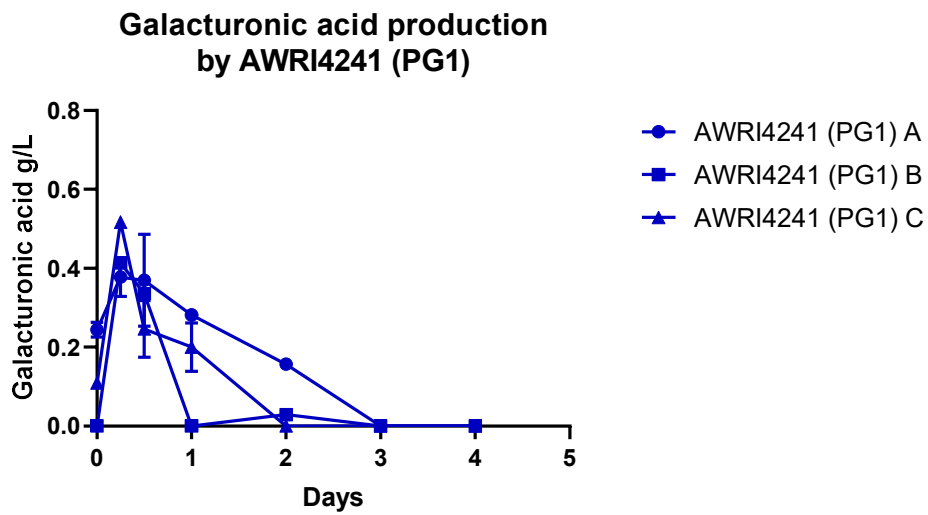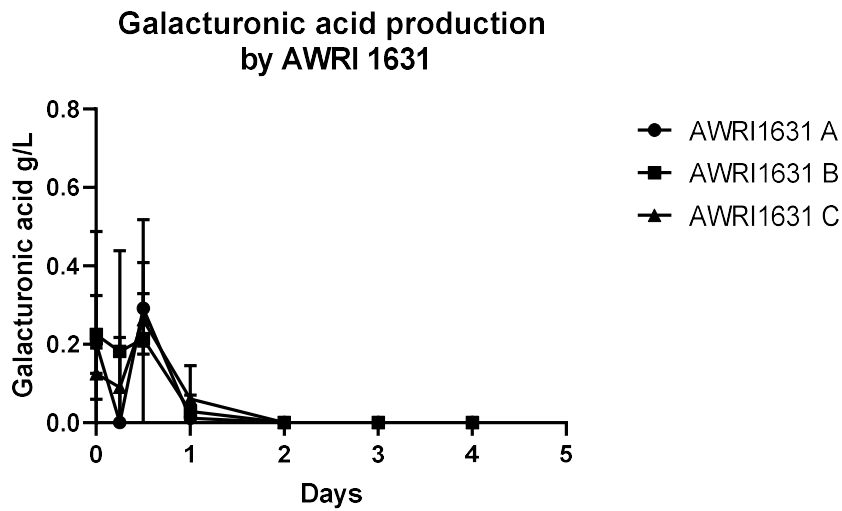
Figure 7.5: Galacturonic acid concentrations in synthetic grape must ferments by AWRI1631, AWRI4241 and AWRI4240 during ferementaiton of synthetic grape must at 22 °C.

Endo-polygalacturonase activity cannot be detected by measuring galacturonic acid concentrations. To measure de-polymerisation of polygalacturonic acid over the course of fermentation, a ruthenium red precipitation assay was used (Figure 7.6), although this method does not detect small galacturonic oligomers or complete pectin degradation. Both AWRI4241 and AWRI4240 had no unhydrolyzed polygalacturonic acid detected after 5 days. In comparison the control strain, AWRI1631, had no detectable unhydrolyzed polygalacturonic acid after 6 days. Oneway ANOVA was used to assess differences in mean unhydrolyzed polygalacturonic acid concentrations between the strains, AWRI1631, AWRI4241 and AWRI4240, during fermentation (Table 6.1). The differences between mean unhydrolyzed polygalacturonic acid concentrations for the three strains on days 2 and 5 of fermentation were sufficient ($p<0.05$) to conclude that these measurements were not from the sample population. As such Dunnett's tests were performed on the data for days 2 and 5 in order to compare the AWRI4241 and AWRI4240 to the control strain, AWRI1631 (Table 6.2). On both days 2 and 5 of fermentation significant difference ($p<0.05$) was observed in the mean concentrations of unhydrolyzed polygalacturonic acid between the control (AWRI1631) and AWRI4241 (produces GH28-1), and the control and AWRI4240 (produces GH28-2).
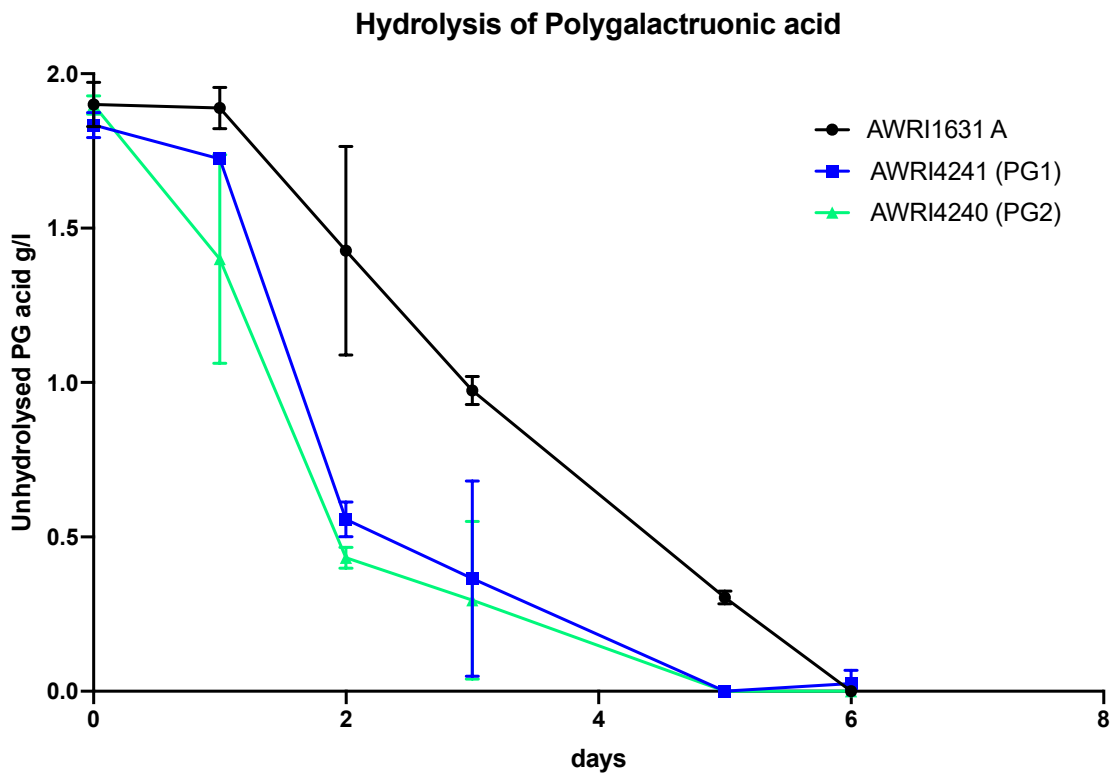
Figure 7.6: Concentration of unhydrolyzed polygalacturonic acid during fermentation of synthetic grape must by AWRI1631, AWRI4241 and AWRI4240 at 22 °C.

Table 7.1: Results of Oneway ANOVA performed on measurements of unhydrolyzed polygalacturonic acid concentrations on days 1, 2, 3 and 5 of synthetic grape must fermentations by AWRI1631, AWRI4241 and AWRI4240.

Day	F	P value	P value summary	R squared
1	1.713	0.258	ns	0.3634
2	18.13	0.0051	**	0.8788
3	4.727	0.0704	ns	0.6541
5	629.6	<0.0001	****	0.9953

Table 7.2: Comparison results by Dunnett's test on unhydrolyzed polygalacturonic acid measurements on Days 2 and 5 during synthetic grape must fermentations by control strain AWRI1631 and test strains AWRI4241 and AWRI4240,

Comparison	Day	Mean difference	CI 95%	Summary	Adjusted P value
AWRI1631-AWRI4241	2	0.8707	0.2704 to 1.471	*	0.0123
	5	0.3046	0.2762 to 0.3329	****	<0.0001
AWRI1631-AWRI4240	2	0.9946	0.4576 to 1.532	**	0.0044
	5	0.3046	0.2762 to 0.3330	****	<0.0001

Discussion

The production of recombinant GH28-1 and GH28-2 *S. cerevisiae* appears to be less prolific in *S. cerevisiae* than in *P. pastoris* (Chapter 7) with *S. cerevisiae* supernatants requiring concentration to detect protein. In addition to producing lower levels of protein, the 2µm episome based system is also expected to be less stable than the genome integrated expression produced in *P. pastoris* and is dependent on antibiotic selection pressure for plasmid maintenance. Despite this, the production of the two recombinant polygalacturonases shows some effect on polygalacturonic acid under wine-like conditions.

Red wines are typically fermented at between 20 °C and 30 ºC and whilst white wines are typically fermented at lower temperatures. The fermentations that were carried out with AWRI1631, AWRI4241 and AWRI4241 were performed at 22 ºC which better simulates red wine like fermentation conditions in terms of temperature. As a consequence of this temperature ferments progressed quickly. However, fermentation rate is reduced in the recombinant strains. This reduction in fermentation rate is probably be caused by the additional metabolic burden caused by the production of the recombinant protein which is being produced constitutively throughout fermentation. However, the antibiotic pressure used to ensure plasmid retention of AWRI4241 and AWRI4241 may also be having a delaying effect on fermentation. This would likely be reduced in a genomic based system which does not require this pressure.

All three strains demonstrated some activity on polygalacturonic acid under wine-like conditions (i.e. temperature, pH, high glucose and fructose concentrations and tolerance to other common chemical found in wine such as malic acid). Some strains of *S. cerevisiae* have a native polygalacturonase, *PGU1*, which is repressed by glucose (Fernandez-Gonzalez et al., 2004, Radoi et al., 2005). From sequencing data available for AWRI1631, the strain does not appear to contain *PGU1* nor any other polygalacturonases nor pectin lyases, which can also depolymerise polygalacturonic acid. AWRI1631 does have glycoside hydrolases predicted but these are not expected to be active on polygalacturonic acid. However, in order to preserve the polygalacturonic acid from hydrolysis during autoclave

processes, the polygalacturonic acid additions were not sterilised and added to filtered and concentrated synthetic grape must directly before fermentation. As such contaminant microbes were introduced in low abundance to the ferments. The affects by these species was expected to be limited by the inoculated *S. cerevisiae*, similar to inoculated grape fermentations, and to be equivalent across the different strains. These microorganisms are likely a contributing factor to polygalacturonase activity. Similarly, chemical hydrolysis of polygalacturonic acid could be a contributing a factor to galacturonic acid and polygalacturonic acid concentrations but were not tested.

Both AWRI4241 and AWRI4240 appear to have some enhancement in exo- and, potentially, endo- polygalacturonase activity compared to the parental strain which can be attributed to the activity of the recombinant polygalacturonases (GH28-1 and GH28-2). The concentrations of galacturonic acid in the media appeared to increase more significantly in the fermentations by AWRI4241 and AWRI4240 than the control strain, AWRI1631, over the first few days of fermentation (Figure 7.5). This suggests that the recombinant polygalacturonases GH28-1 and GH28-2 have exo-polygalacturonase activity.

However, although galacturonic acid levels increased during fermentation by all three *S. cerevisiae* strains, increases are modest and galacturonic acid concentrations reduce significantly after the initial increase. AWRI1631 ferments reached approximately 0 g/L of galacturonic acid after 1 day, ferments AWRI4240 after 2 days and AWRI4241 with 3 days (Figure 7.5). *S. cerevisiae* strains have not been demonstrated to metabolise galacturonic acid nor have any strains been identified with the genes required for galacturonic acid degradation. As the parent strain AWRI1631 also does not contain these genes, the three strains AWRI1631, AWRI4241 and AWRI4240 are unlikely to be removing galacturonic acid from the media by galacturonic acid degradation directly. However, it is possible that promiscuous enzyme activity could be contributing to the degradation of galacturonic acid. A previous study (Souffriau et al., 2012) has noted that various *S. cerevisiae* strains sequestered galacturonic acid from media, particularly at low pH, but not in concentrations higher than in the extracellular environment. It is likely some contaminant species are

capable of metabolising galacturonic acid and contributing to the reduction of galacturonic acid concentrations. Additionally, ethanol and other fermentation bioproducts may be having an inhibitory effect on the activity of the uronate dehydrogenase or preventing galacturonic acid availability to the enzyme. However, at least in terms of enzyme inhibition, this was expected to be minimal in the assay due to sample dilution (1/25) in enzyme buffer and tolerance to the compounds in the synthetic grape must in which the standard curve was produced. An alternate method, e.g. HPLC, would allow confirmation of the decrease in galacturonic acid concentration in these ferments.

In the ruthenium red method, endo type degradation of polygalacturonic acid (i.e. cleavage within the length of the polygalacturonic acid chain) is measured by the loss of co-precipitation of long chain polygalacturonic acid and ruthenium red (Torres et al., 2011). As polygalacturonic acid is depolymerised, shorter polygalacturonic acid oligomers are produced, and these oligomers have reduced capacity to precipitate with ruthenium red. As ruthenium red has strong absorbance at 535 nm and more ruthenium red remains in solution as the polygalacturonic acid chains become shorter, the absorbance of the solution increases. The linear range for this assay is between 0.1 and 2 g/L of unhydrolyzed polygalacturonic acid equivalents. However, shorter polygalacturonase oligomers may remain in solution as the point at which the shorter polygalacturonic acid oligomers completely lose the ability to precipitate ruthenium red is unknown i.e. polymer of five galacturonic acids might be unable to effectively precipitate ruthenium red. Therefore, the data from these assays cannot confirm if the polygalacturonic acid was completely degraded, especially given the absence of galacturonic acid detected in the media towards the end of fermentation. Visually, AWRI4241 and AWRI4240 appear to have enhanced endo-type polygalacturonase compared to AWRI1631 (Figure 7.6). However, statistical analysis only supports differences in the mean unhydrolyzed polygalacturonic acid concentrations between the control and the two strains on days 2 and 5 of fermentation. These results support that both GH28-1 and GH28-2 may have some endo-type polygalacturonase activity under fermentative conditions, but additional experiments are

required to confirm if putative activity is enhancing the rate of polygalacturonic acid degradation to a significant degree.

Despite being produced at very low levels, GH28-1 competes well with GH28-2 in activity. Although AWRI4240 (GH28-2 recombinant strain) produces higher maximum galacturonic acid concentrations, the two have similar rates of reduction of unhydrolyzed polygalacturonic acid. Previously (Chapter 7), GH28-1 was demonstrated be less active per mg of enzyme than GH28-2 but retained more relative activity at pH 3.5. Similarly, GH28-1 also had higher relative activity at low temperatures than GH28-2, which is most active at 60 ^{o}C. As the media had a pH 3.5 and fermentation was carried out at 22 ^{o}C, it is expected that GH28-1 is better suited to the conditions which is why it exhibits similar levels of activity to GH28-2. Improving GH28-2 response to these conditions and improving the overall activity and expression of GH28-1 will be necessary to enhance the *in situ* polygalacturonase activity. In a recent study, heterologous expression of protein on the surface membrane of *S. cerevisiae* has been demonstrated to improve enzyme activity under winemaking conditions (Zhang et al., 2019). This is a potential avenue to improve GH28-2 activity.

Although GH28-1 and GH28-2 demonstrate activity in a wine-like environment when expressed by *S. cerevisiae* in these experiments, they still need to be tested in wine grape fermentations. Purified polygalacturonic acid is less complex than grape pectin (Gao et al., 2016, Vidal et al., 2001) and grape pectin is recalcitrant to enzymatic activity. Pectin may not be completely dissolved in wine during clarification and less accessible to enzymes. Similarly, pectin will be less available to enzymes during juice extractions and as skins during fermentation. As such AWRI4241 and AWRI4240 may be less effective in grape fermentations and would need to be trialled for efficacy.

Conclusions

The introduction of recombinant GH28-1 and GH28-2 expression into the *S. cerevisiae* wine strain AWRI1631 improved degradation of polygalacturonic acid in wine-like media, synthetic grape must, during fermentation, with both recombinant strains generating increased production of galacturonic acid and a putative increase in the rate of polygalacturonic acid depolymerisation compared to the parental strain. Both GH28-1 and GH28-2 have demonstrated activity under wine-like conditions including as low pH, low temperature, high concentrations of glucose and fructose and tolerance to other common chemicals found in wine. However, the recombinant protein expression in *S. cerevisiae* was less efficient than had been established in *P. pastoris* (Chapter 7) and the strains show somewhat diminished fermentation rates compared to the parental strain. Improving heterologous protein expression and establishing a more stable expression system may allow increased polygalacturonase activity and expression stability but may not improve fermentative character. Further testing of these strains on grapes is required to assess suitability in a winemaking context.

Chapter 8: Final discussion and future directions

Enzymes are used in numerous industries worldwide to induce specific chemical reactions without the use of harsher processes and have been used as an aid in winemaking for decades (Aryan et al., 1987, van Rensburg and Pretorius, 2000). However, many enzyme preparations are non-specific, contain off-target activities, and are poorly adapted to winemaking conditions. The field of metagenomics has made it possible to access previously inaccessible microbiota and in combination with techniques in synthetic biology has greatly expanded the pool of enzymes available for investigations. Studies of wine microbiota have highlighted their importance in winemaking and on winemaking outcomes (Verginer et al., 2010, Maturano et al., 2012). Consequently, these organisms have great potential to contain wine active enzymes which are better adapted to the industry as the microbiota are better adapted to grape environments. If these enzymes could be applied to winemaking successfully, they have the potential to alter the production and even properties of wine and provide new tools to winemakers.

The intention of this project was to investigate the enzymatic potential of wine-related microbiota as a source of enzymes adapted to the wine environment. For this purpose, the properties of two grape related environments were investigated: a Chardonnay grape must and a mixed varietal grape marc that had begun the composting process. DNA was isolated from the two samples and subjected to bioinformatic analysis focusing on the discovery of novel enzymes. Glycoside hydrolases have several applications in winemaking, specifically the breakdown of plant cell wall polysaccharides (Louw et al., 2006) and the release of organoleptically active compounds from glycosyl linkage (Zhu et al., 2014). These enzymes also have a well understood catalytic function (Davies and Henrissat, 1995, Henrissat et al., 1995) on which to base functional predictions and were targeted for discovery and expression. Using molecular methods candidate β-glucosidases and polygalacturonases were expressed and tested for their putatively assigned functions.

This final chapter focuses on the overall outcomes of the project, the methods used, particularly the bioinformatic approaches taken, and possible future directions of this work.

Methods of metagenomics and gene prediction

A metagenomic approach was taken to maximise the number of organisms surveyed for potential enzymes of interest. This involved the isolation and shotgun sequencing of environmental DNA, metagenomic assembly, gene prediction, identification of enzymes of interest and then production and testing of these enzymes in laboratory experiments.

Although, these methods allowed access to non-culturable micro-organisms, there are limitations at every step of the process that introduced biases. For instance, the method of DNA extraction can itself be selective for different species. Sequencing quality, depth and types of sequencing data dictate the success of any genomic assembly and this is similar for metagenomic assemblies. For the CGM metagenome only, short-read data was used, whereas a mixture of long-read and short-read data was used in the final MVGM assembly. As a result of the higher sequencing depth and the scaffolding by long-reads, the combined MVGM metagenome is of a higher quality than the CGM counterpart. Although both metagenomes are fragmented compared to genome from a single isolate, fragmentation remains a characteristic for metagenomic assemblies despite significant improvements due to the availability of long-read data (Frank et al., 2016). In future metagenomic assemblies, it would be beneficial to be able to use this combination of short-read and long-read data.

Improved sequence depth and sequence length would reduce the difficulties of *de novo* assembly, which was necessary for metagenomic assembly as the genomes present in the data are largely unknown. Long read data also ensures a minimum length in lowly abundant species which are poorly captured even at higher sequencing depths. However, as error rates in long-read data remain high, short-read technologies are still required to improve accuracy in assemblies. As long-read technologies improve this may eventually no longer be an issue.

The taxonomic assignments of reads were made by the taxonomic classifier Centrifuge (Kim et al., 2016) using the NCBI protein database. These assignments assess the metagenomic community and relative microbial abundances which can impact on metagenomic assembly and assembly strategies. As a reference-based form of read

binning these classifications are highly dependent on both the reference database and the ability of the classifying software to accurately classify ambiguous reads to an appropriate level. This is particularly poignant for fungi and other eukaryotes which are often not considered in the design of taxonomic classification pipelines. Since this work, new entries have been made to the NCBI database (and are continuously) which could alter read classifications and new upgrades and pipelines have been produced. In particular, Kraken2 has been released (Wood et al., 2019) from the same group as the centrifuge pipeline and is of interest for future taxonomic classification work. No current pipelines have been developed for use with the NCBI whole genome shotgun database which could facilitate comparisons to other metagenomic data but is challenging due to database size. MASH (Ondov et al., 2016) is a potential avenue for metagenome comparisons by facilitating efficient compression of metagenomic data. Although, perhaps still not the entire NCBI whole genome shotgun database.

A non-reference-based binning method was used to identify and isolate whole genomes from the CGM and MVGM metagenomes with variable success. However, reference independent methods are typically modelled on bacterial genomes and were poorly efficient on the CGM data. Although this is partially due to the lesser quality of the CGM metagenome, poor separation is also a function of the high eukaryotic content. Recent developments in eukaryotic targeted binning methods (West et al., 2018) may be more effective at binning of this kind of metagenomic data.

Gene structure and patterns are variable between different organisms and gene models can be highly complex. Metagenomes by their nature contain a mixture of organisms and therefore of gene patterns. The software AUGUSTUS (Stanke et al., 2006) was used to predict coding genes from both metagenomes and is specialised for eukaryotic gene prediction, using *ab initio* methods combined with Hidden Markov models trained to various eukaryotic species. However, this software was developed for single genomes and does not differentiate between different species during classification which introduces errors into gene predictions. Other software have been designed for gene prediction from

metagenomes (Noguchi et al., 2008) to avoid these issues but are tailored to bacterial samples. An alternative approach would have been to separate the metagenomic contigs by binning and use gene models more specific to the species but is then more dependent on successful binning of data.

Additional efforts to reduce possible errors in gene prediction were performed on potential genes of interest identified by functional characterisation not on the entire dataset. Consequently, there are probable errors in the gene prediction datasets as a whole, but these should be reduced in the genes of interest. The functional characterisation of gene predictions was performed by InterProScan and BLAST to the NCBI protein database. InterPro contains various protein structural databases which are effective at predicting protein structure, but other databases and protein function predictors are available which could be used to similar effect. The genes of interest were flagged for possible truncations or extensions by comparison to other genes of their predicted type and tested with gene models to observe if their homology to NCBI proteins or InterPro domains improved.

The most significant limitation on the ability to predict function of putative enzyme from enzyme coding genes is the need for prior protein structure information of the target enzymes of interest, although in many cases this need only be at the amino acid level. This means that entirely new protein structures are generally not identified. Glycoside hydrolases were targeted as they are industrially relevant, but also a highly researched group of enzymes, which provides a solid foundation for functional prediction. A total of 11,146 putative glycoside hydrolases, of which 4,719 had <80% identity to genes from the NCBI nr protein database, were predicted. The 80% identity threshold was used in an effort to balance enzyme novelty (by removing known proteins) against potential gene prediction and classification errors, as manually curating 11,146 candidate enzymes individually was not feasible. However, 80% identity is still high and given the number of genes remaining in the enzyme's groups of interest, the maximum percent identity could have been lowered to remove more candidates.

Heterologous protein expression of enzymes of interest

Although, the chosen method of enzyme discovery has limitations, the approach was effective, with more putative enzymes identified than could be evaluated in this study. Gene synthesis is required in order to produce and test enzymes of interest identified from gene sequencing. Gene synthesis has advantages, such as tailoring to host organisms and better control of expression systems. However, this synthesis was also a bottleneck for the number of genes which could be tested. Gene synthesis and construction of expression systems was time consuming, is more costly than other methods and still required optimisation and troubleshooting. For instance, cDNA cloning directly from the samples could have been used to generate large numbers of candidates, increase throughput and avoid whole genome sequencing. However, this approach was also dependent on induction of desired activities and compatibility between the expression organism and the cDNA. Advancements in gene synthesis technology are reducing costs per gene so this is less of a concern, but time constraints still apply.

Compatibility between the candidate enzymes and the two expression hosts was an issue. The yeasts, *P. pastoris* and *S. cerevisiae,* were chosen as expression hosts for their enzyme secretory pathways, *P. pastoris'* demonstrated ability to produce large quantities of secreted protein and *S. cerevisiae's* fermentative properties for wine-like experiments. Moreover, there is cross compatibility between the two yeasts which simplified construct of expression plasmids. Three glycoside hydrolase family 3 (GHF3) β-glucosidases, four glycoside hydrolase family 1 (GHF1) β-glucosidases and 5 polygalacturonases were selected for expression. However, none of the GHF3 β-glucosidases were successfully expressed in *P. pastoris* or in *S. cerevisiae*. Three of four GHF1 β-glucosidases and three of five polygalacturonases were successfully expressed in *P. pastoris*. In general, failed expression in *P. pastoris* appeared to be occurring at the transcriptional and translation levels whereas *S. cerevisiae* expression plasmids integrated poorly at the *HO* locus. *S. cerevisiae* expression was established for the polygalacturonases GH28-1 and GH28-2 but was less efficient than in *P. pastoris* and was based on a 2 μm episome. Future experiments on these candidates would attempt to optimise protein expression of low expressing strains,

produce a stable genome integrated *S. cerevisiae* expression system for GH28-1 and GH28-2 and potentially the GHF1 β-glucosidases, and troubleshooting of *P. pastoris* expression of the non-expressing strains. A simple experiment to facilitate protein production in *P. pastoris* would be to transform the existing *P. pastoris* expression plasmids into a protease deficient strain. If the protease deficient strains failed, alternate codon sequences and short protein fragments may aid production of the non-producing candidates. *S. cerevisiae* optimisations would involve testing alternate integration sites and alternate secretion signals. Protein expression in alternate hosts such as *E. coli* could also be tested.

β-glucosidases

Predictions of enzyme functions were successful for the β-glucosidases: GH1-1, GH1-3 and GH1-4. When expressed by *P. pastoris,* all three candidates showed at least some level of activity on p-nitrophenyl-glucopyranoside. However, these three enzymes are poorly active under acidic conditions, which in combination with low overall activity precludes their use in winemaking and other potential applications. As discussed in Chapter 5, future work on these enzymes for winemaking applications would need to address the pH issue, overall activity and test activity on other substrates before wine and wine-like trials could be performed. Given β-glucosidases could easily be linked to the availability of carbon sources such as glucose, mutagenesis and adaptive evolution experiments could be an avenue to improve β-glucosidase activity.

Polygalacturonases

Polygalacturonase activity is of oenological interest in order to aid in the breakdown of pectic substances in grape juice to aid in juice extraction and clarification, by softening plant cell walls and increasing pectin solubility. The metagenome derived polygalacturonases GH28-1 and GH28-2 show low (GH28-1) to moderate (GH28-2) activity on polygalacturonic acid and were functional on polygalacturonic acid in synthetic grape must. As discussed in chapter 7, further testing in grape juice is required in order to validate these enzymes and recombinant *S. cerevisiae* for use in winemaking, and different expression strategies, such as cell surface protein expression, may be useful to improve protein expression and enzyme

properties. The addition of *P. pastoris* recombinant GH28-1 and GH28-2 to grape juice fermentations, similar to current usage of enzyme preparations, could also be trialled.

Three different assays were used to detect polygalacturonase activity; the uronate dehydrogenase (UDH) assay, which was used to detect the production of galacturonic acid; the di-nitrosalicylic acid (DNS) assay, which was used to detect activity by the production of reducing sugars; and the ruthenium red (RR) assay which detected reduction in the ability to precipitate ruthenium red by reduction of polygalacturonic acid polymer length. Although, these methods were sufficient for most of the experimental work, a HPLC or similar method would have improved fermentation analysis and should be the preferred avenue for future work in grape juice.

Future Directions

There are multiple natural offshoots to this project. With over 1.5 million protein encoding genes predicted from the two metagenomes produced in this work, there are numerous potential enzymes which can still be investigated and the sequencing data itself can be further mined. Numerous polygalacturonases or β-glucosidases with the potential to alter cell wall structure and modulate wine flavour were identified but not investigated. Given the two polygalacturonases of fungal origin (GH28-1 and GH28-2) had function under wine-like conditions further characterisations of these polygalacturonases and the bacterial β-glucosidases could be explored. In particular, the activity displayed by the two polygalacturonases, GH28-1 and GH28-2, needs to be confirmed by testing on grape pectin and on grape must and for any potential use as a commercial product, the purification and stability of GH28-1 and GH28-2 needs to be tested as well as the ability to scale up enzyme production. This may require testing of alternate expression systems and expression constructs to optimise enzyme production. The β-glucosidases which showed activity to the colorimetric substrate could be tested for specific activities on wine flavour precursors with the potential to select for specific flavours in wine.

Alternate enzyme activities could also be valuable. In a recent study of grape cell wall degrading enzymes (Gao et al., 2016) pectin lyases and rhamnogalacturonan lyases were

highlighted as a potentially more important component of cell wall degradation than polygalacturonases. In this work, five putative pectin lyases from polysaccharide lyase family 1 and 18 putative rhamnogalacturonan lyases from polysaccharide lyase family 4 were identified from the two metagenomes and could be suitable future candidates for protein expression and characterisations. Similarly, lytic polysaccharide mono-oxygenases (LPMOs) show promising activity on recalcitrant polysaccharides (Hemsworth et al., 2013) and 36 putative LPMOs were identified. Mono-oxygenases typically function with a redox partner, provided either by a second enzyme, such as a cellulose dehydrogenase, or chemical electron donor (Loose et al., 2016). As the chemical compounds which can be legally added to wine fermentations are limited, a dual protein expression system to express candidate LPMOs and an enzyme electron donor would be ideal. Multiple partner proteins may need to be tested. Numerous proteases were also identified and may also be targeted for applications to preventing and remediating wine haze (Van Sluyter et al., 2015). If viable enzymes from the different groups are produced, synergistic experiments could allow the production of enhanced enzyme preparations that effectively combine the different enzymes, particularly those involved in cell wall degradation, and be targeted to different cell wall structures.

Genes for the biosynthesis of various metabolites are often grouped into clusters in both bacteria and eukaryotes (van der Lee and Medema, 2016, Chen et al., 2019). Recently, metagenomic sequencing data has been used to identify biosynthetic gene clusters (BGCs) and the related potential for metabolite production (Donia et al., 2014, Aleti et al., 2019) but these techniques have not yet been applied wine microorganisms. During wine grape juice fermentation, the production of secondary metabolites is significant as it affects the organoleptic properties of wine and can introduce distinct flavours outside of the compounds found in the grapes (Verginer et al., 2010, Cordente et al., 2012). Given the importance of secondary metabolites in wine, the investigation of wine biosynthetic clusters is highly relevant to industry and the identification of biosynthetic gene clusters (BGCs) are a potential avenue for future research using the CGM and MVGM metagenomes.

Appendix A

Construction of pD912 and pD912-HIS

The modified version of pD912-AKS (ATUM), "pD912", with homologous regions to pCVSα

for simplified cloning and the His tagged variant, "pD912-His", were constructed by Gibson

assembly of synthesised DNA fragments (Integrated DNA Technology) "pD912-insert" and

"pD-His" (Table A.1) into the linear pD912-AKS (provided as linear fragment by ATUM).

Gibson reaction mixture was transformed into 10-Beta E. coli (NEB) using the standard heat

shock protocol and selected on LB (1% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.5% (w/v)

sodium chloride) with 100 µg/ml zeocin. Successful plasmid construction confirmed by

restriction digest of plasmid DNA extracted from zeocin resistant colonies.

Table A.1: DNA sequence of synthesised DNA of GNA inserts to produce pD912-AKS variants.

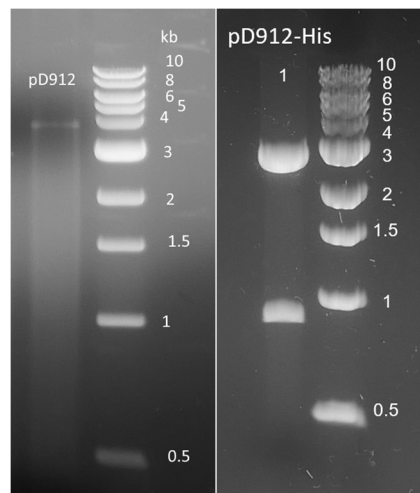| Name	Sequence
pD912-insert	5'- GGGGTATCTCTCGAGAAAAGAGAGGCTGAAGCCATGGGTCGAGAC CGGAGAACCTCTAGCTGCTGACCGTTATTCTACTGGGCTAGCGGTTAA GGGGCGGCCGCTCAAGAGGATGTCAGAATGCC-3'
pD-His	5'-CAGCATTGCTGCTAAAGAAGAAGGGGTATCTCTCGAGAAAAGAGAG GCTGAAGCCCACCATCATCATCATCATGCCATGGCGTCTTCCTGCACA TTCTATCTGCTAGCGGTTAAGGGGCGGCCGC-3'

Figure A.1: Restriction digests of plasmid DNA to confirm pD912 and D912-His construction

and visualised on 1% (w/v) agarose gel. Expected band sizes: pD912 with *Nhe*I and pD912-

His with *Nco*I 2852bp + 933bp

Bibliography

ADAPA, V., RAMYA, L. N., PULICHERLA, K. K. & RAO, K. R. 2014. Cold active pectinases: advancing the food industry to the next generation. *Appl Biochem Biotechnol,* 172**,** 2324-37.

ALETI, G., BAKER, J. L., TANG, X., ALVAREZ, R., DINIS, M., TRAN, N. C., MELNIK, A. V., ZHONG, C., ERNST, M., DORRESTEIN, P. C. & EDLUND, A. 2019. Identification of the Bacterial Biosynthetic Gene Clusters of the Oral Microbiome Illuminates the Unexplored Social Language of Bacteria during Health and Disease. *mBio,* 10**,** 1-19.

ALIMARDANI-THEUIL, P., GAINVORS-CLAISSE, A. & DUCHIRON, F. 2011. Yeasts: An attractive source of pectinases-From gene expression to potential applications: A review. *Process Biochem,* 46**,** 1525-1537.

AMAYA, M. F., WATTS, A. G., DAMAGER, I., WEHENKEL, A., NGUYEN, T., BUSCHIAZZO, A., PARIS, G., FRASCH, A. C., WITHERS, S. G. & ALZARI, P. M. 2004. Structural insights into the catalytic mechanism of *Trypanosoma cruzi trans*-sialidase. *Structure,* 12**,** 775-84.

ARMIJO, G., SCHLECHTER, R., AGURTO, M., MUNOZ, D., NUNEZ, C. & ARCE-JOHNSON, P. 2016. Grapevine Pathogenic Microorganisms: Understanding Infection Strategies and Host Response Scenarios. *Front Plant Sci,* 7**,** 1-18.

ARYAN, A. P., WILSON, B., STRAUSS, C. R. & WILLIAMS, P. J. 1987. The Properties of Glycosidases of *Vitis vinifera* and a Comparison of Their β-Glucosidase Activity with that of Exogenous Enzymes. An Assessment of Possible Applications in Enology. *American Journal of Enology and Viticulture,* 38**,** 182-188.

BAKER, R. A. 1997. Reassessment of some fruit and vegetable pectin levels. *Journal of Food Science,* 62**,** 225-229.

BARATA, A., MALFEITO-FERREIRA, M. & LOUREIRO, V. 2012. The microbial ecology of wine grape berries. *Int J Food Microbiol,* 153**,** 243-59.

BELDA, I., CONCHILLO, L. B., RUIZ, J., NAVASCUES, E., MARQUINA, D. & SANTOS, A. 2016a. Selection and use of pectinolytic yeasts for improving clarification and phenolic extraction in winemaking. *Int J Food Microbiol,* 223**,** 1-8.

BELDA, I., RUIZ, J., ALASTRUEY-IZQUIERDO, A., NAVASCUES, E., MARQUINA, D. & SANTOS, A. 2016b. Unraveling the Enzymatic Basis of Wine "Flavorome": A Phylo-Functional Study of Wine Related Yeast Species. *Front Microbiol,* 7**,** 1-13.

BENUCCI, I., ESTI, M. & LIBURDI, K. 2015. Effect of wine inhibitors on the proteolytic activity of papain from Carica papaya L. latex. *Biotechnol Prog,* 31**,** 48-54.

BERLEC, A. & STRUKELJ, B. 2013. Current state and recent advances in biopharmaceutical production in *Escherichia coli*, yeasts and mammalian cells. *J Ind Microbiol Biotechnol,* 40**,** 257-74.

BIGGS, A. I. 1954. A spectrophotometric determination of the dissociation constants of p-nitrophenol and papaverine. *Transactions of the Faraday Society,* 50**,** 800-802.

BILLINGSLEY, J. M., DENICOLA, A. B. & TANG, Y. 2016. Technology development for natural product biosynthesis in *Saccharomyces cerevisiae*. *Curr Opin Biotechnol,* 42**,** 74-83.

BIVER, S., STROOBANTS, A., PORTETELLE, D. & VANDENBOL, M. 2014. Two promising alkaline b-glucosidases isolated by functional metagenomics from agricultural soil, including one showing high tolerance towards harsh detergents, oxidants and glucose. *J Ind Microbiol Biotechnol,* 41**,** 479-88.

BLANCO, P., SIEIRO, C., REBOREDO, N. M. & VILLA, T. G. 1998. Cloning, molecular characterization, and expression of an endo-polygalacturonase-encoding gene from *Saccharomyces cerevisiae* IM1-8b. *Fems Microbiology Letters,* 164**,** 249-255.

BOKULICH, N. A., THORNGATE, J. H., RICHARDSON, P. M. & MILLS, D. A. 2014. Microbial biogeography of wine grapes is conditioned by cultivar, vintage, and climate. *Proc Natl Acad Sci U S A,* 111**,** E139-48.

BORNEMAN, A. R., FORGAN, A. H., PRETORIUS, I. S. & CHAMBERS, P. J. 2008. Comparative genome analysis of a *Saccharomyces cerevisiae* wine strain. *FEMS Yeast Res,* 8**,** 1185-95.

BRAVO, L. & SAURA-CALIXTO, F. 1998. Characterization of dietary fiber and the in vitro indigestible fraction of grape pomace. *American Journal of Enology and Viticulture,* 49**,** 135-141.

BURMEISTER, W. P., COTTAZ, S., ROLLIN, P., VASELLA, A. & HENRISSAT, B. 2000. High resolution X-ray crystallography shows that ascorbate is a cofactor for myrosinase and substitutes for the function of the catalytic base. *J Biol Chem,* 275**,** 39385-93.

CAMPANARO, S., TREU, L., VENDRAMIN, V., BOVO, B., GIACOMINI, A. & CORICH, V. 2014. Metagenomic analysis of the microbial community in fermented grape marc reveals that Lactobacillus fabifermentans is one of the dominant species: insights into its genome structure. *Appl Microbiol Biotechnol,* 98**,** 6015-37.

CAPOZZI, V., GAROFALO, C., CHIRIATTI, M. A., GRIECO, F. & SPANO, G. 2015. Microbial terroir and food innovation: The case of yeast biodiversity in wine. *Microbiol Res,* 181**,** 75-83.

CHEN, R., WONG, H. L. & BURNS, B. P. 2019. New Approaches to Detect Biosynthetic Gene Clusters in the Environment. *Medicines (Basel),* 6**,** 1-11.

CHENG, Z., CHEN, D., LU, B., WEI, Y., XIAN, L., LI, Y., LUO, Z. & HUANG, R. 2016. A Novel Acid-Stable Endo-Polygalacturonase from *Penicillium oxalicum* CZ1028: Purification, Characterization, and Application in the Beverage Industry. *J Microbiol Biotechnol,* 26**,** 989-98.

CLAUS, H. & MOJSOV, K. 2018. Enzymes for Wine Fermentation: Current and Perspective Applications. *Fermentation,* 4**,** 1-19.

CORDENTE, A. G., CURTIN, C. D., VARELA, C. & PRETORIUS, I. S. 2012. Flavour-active wine yeasts. *Appl Microbiol Biotechnol,* 96**,** 601-18.

CREGG, J. M. 2008. *Pichia Protocols, Second Edition*, Humana Press Inc., Totowa, NJ.

DAN, S., MARTON, I., DEKEL, M., BRAVDO, B. A., HE, S., WITHERS, S. G. & SHOSEYOV, O. 2000. Cloning, expression, characterization, and nucleophile identification of family 3, Aspergillus niger β-glucosidase. *J Biol Chem,* 275**,** 4973-80.

DAVIES, G. & HENRISSAT, B. 1995. Structures and mechanisms of glycosyl hydrolases. *Structure,* 3**,** 853-9.

DÁVILA, I., ROBLES, E., EGÜÉS, I., LABIDI, J. & GULLÓN, P. 2017. The Biorefinery Concept for the Industrial Valorization of Grape Processing By-Products. *In:* GALANAKIS, C. M. (ed.) *Handbook of Grape Processing By-Products.* Academic Press.

DE GIUSEPPE, P. O., SOUZA TDE, A., SOUZA, F. H., ZANPHORLIN, L. M., MACHADO, C. B., WARD, R. J., JORGE, J. A., FURRIEL RDOS, P. & MURAKAMI, M. T. 2014. Structural basis for glucose tolerance in GH1 β-glucosidases. *Acta Crystallogr D Biol Crystallogr,* 70**,** 1631-9.

DE SANTI, C., ALTERMARK, B., DE PASCALE, D. & WILLASSEN, N. P. 2016. Bioprospecting around Arctic islands: Marine bacteria as rich source of biocatalysts. *J Basic Microbiol,* 56**,** 238-53.

DOCO, T., WILLIAMS, P., PAULY, M., O'NEILL, M. A. & PELLERIN, P. 2003. Polysaccharides from grape berry cell walls. Part II. Structural characterization of the xyloglucan polysaccharides. *Carbohydrate Polymers,* 53**,** 253-261.

DONIA, M. S., CIMERMANCIC, P., SCHULZE, C. J., WIELAND BROWN, L. C., MARTIN, J., MITREVA, M., CLARDY, J., LININGTON, R. G. & FISCHBACH, M. A. 2014. A systematic analysis of biosynthetic gene clusters in the human microbiome reveals a common family of antibiotics. *Cell,* 158**,** 1402-1414.

ECKART, M. R. & BUSSINEAU, C. M. 1996. Quality and authenticity of heterologous proteins synthesized in yeast. *Curr Opin Biotechnol,* 7**,** 525-30.

ESCHSTRUTH, A. & DIVOL, B. 2011. Comparative characterization of endo-polygalacturonase (Pgu1) from *Saccharomyces cerevisiae* and *Saccharomyces paradoxus* under winemaking conditions. *Appl Microbiol Biotechnol,* 91**,** 623-34.

FAVARO, L., BASAGLIA, M., TRENTO, A., VAN RENSBURG, E., GARCIA-APARICIO, M., VAN ZYL, W. H. & CASELLA, S. 2013. Exploring grape marc as trove for new

thermotolerant and inhibitor-tolerant *Saccharomyces cerevisiae* strains for second-generation bioethanol production. *Biotechnology for Biofuels,* 6**,** 1-14.

FERNANDEZ-GONZALEZ, M., UBEDA, J. F., VASUDEVAN, T. G., CORDERO OTERO, R. R. & BRIONES, A. I. 2004. Evaluation of polygalacturonase activity in *Saccharomyces cerevisiae* wine strains. *FEMS Microbiol Lett,* 237**,** 261-6.

FORSBERG, Z., VAAJE-KOLSTAD, G., WESTERENG, B., BUNAES, A. C., STENSTROM, Y., MACKENZIE, A., SORLIE, M., HORN, S. J. & EIJSINK, V. G. 2011. Cleavage of cellulose by a CBM33 protein. *Protein Sci,* 20**,** 1479-83.

FRANK, J. A., PAN, Y., TOOMING-KLUNDERUD, A., EIJSINK, V. G. H., MCHARDY, A. C., NEDERBRAGT, A. J. & POPE, P. B. 2016. Improved metagenome assemblies and taxonomic binning using long-read circular consensus sequence data. *Sci Rep,* 6**,** 1-10.

GABLER, F. M., SMILANICK, J. L., MANSOUR, M., RAMMING, D. W. & MACKEY, B. E. 2003. Correlations of Morphological, Anatomical, and Chemical Features of Grape Berries with Resistance to Botrytis cinerea. *Phytopathology,* 93**,** 1263-73.

GAO, Y., FANGEL, J. U., WILLATS, W. G. T., VIVIER, M. A. & MOORE, J. P. 2016. Dissecting the polysaccharide-rich grape cell wall matrix using recombinant pectinases during winemaking. *Carbohydr Polym,* 152**,** 510-519.

GAO, Y., ZIETSMAN, A. J. J., VIVIER, M. A. & MOORE, J. P. 2019. Deconstructing Wine Grape Cell Walls with Enzymes During Winemaking: New Insights from Glycan Microarray Technology. *Molecules,* 24**,** 1-19.

GAYEVSKIY, V. & GODDARD, M. R. 2012. Geographic delineations of yeast communities and populations associated with vines and wines in New Zealand. *ISME J,* 6**,** 1281-90.

GEST, H. 2004. The discovery of microorganisms by Robert Hooke and Antoni Van Leeuwenhoek, fellows of the Royal Society. *Notes Rec R Soc Lond,* 58**,** 187-201.

GIBSON, D. G., YOUNG, L., CHUANG, R. Y., VENTER, J. C., HUTCHISON, C. A., 3RD & SMITH, H. O. 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods,* 6**,** 343-5.

GIETZ, R. D. & SCHIESTL, R. H. 2007. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat Protoc,* 2**,** 31-34.

GIL, J. V., MANZANARES, P., GENOVES, S., VALLES, S. & GONZALEZ-CANDELAS, L. 2005. Over-production of the major exoglucanase of Saccharomyces cerevisiae leads to an increase in the aroma of wine. *Int J Food Microbiol,* 103**,** 57-68.

GIRARD, V., DIERYCKX, C., JOB, C. & JOB, D. 2013. Secretomes: the fungal strike force. *Proteomics,* 13**,** 597-608.

GODDEN, P., WILKES, E. & JOHNSON, D. 2015. Trends in the composition of Australian wine 1984-2014. *Australian Journal of Grape and Wine Research,* 21**,** 741-753.

GONZALEZ-POMBO, P., FARINA, L., CARRAU, F., BATISTA-VIERA, F. & BRENA, B. M. 2014. Aroma enhancement in wines using co-immobilized *Aspergillus niger* glycosidases. *Food Chem,* 143**,** 185-91.

GRUNINGER, R. J., GONG, X., FORSTER, R. J. & MCALLISTER, T. A. 2014. Biochemical and kinetic characterization of the multifunctional b-glucosidase/b-xylosidase/a-arabinosidase, Bgxa1. *Appl Microbiol Biotechnol,* 98**,** 3003-12.

HARVEY, A. J., HRMOVA, M., DE GORI, R., VARGHESE, J. N. & FINCHER, G. B. 2000. Comparative modeling of the three-dimensional structures of family 3 glycoside hydrolases. *Proteins,* 41**,** 257-69.

HEMSWORTH, G. R., DAVIES, G. J. & WALTON, P. H. 2013. Recent insights into copper-containing lytic polysaccharide mono-oxygenases. *Curr Opin Struct Biol,* 23**,** 660-8.

HENRISSAT, B. 1991. A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem J,* 280**,** 309-16.

HENRISSAT, B. & BAIROCH, A. 1993. New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem J,* 293**,** 781-8.

HENRISSAT, B., CALLEBAUT, I., FABREGA, S., LEHN, P., MORNON, J. P. & DAVIES, G. 1995. Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. *Proc. Natl. Acad. Sci. U.S.A.,* 92**,** 7090-4.

HENRISSAT, B. & DAVIES, G. 1997. Structural and sequence-based classification of glycoside hydrolases. *Curr Opin Struct Biol,* 7**,** 637-44.

HESS, M., SCZYRBA, A., EGAN, R., KIM, T. W., CHOKHAWALA, H., SCHROTH, G., LUO, S., CLARK, D. S., CHEN, F., ZHANG, T., MACKIE, R. I., PENNACCHIO, L. A., TRINGE, S. G., VISEL, A., WOYKE, T., WANG, Z. & RUBIN, E. M. 2011. Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science,* 331**,** 463-7.

HILDEBRAND, E. M. 1938. Techniques for the isolation of single microorganisms. *The Botanical Review,* 4**,** 627-664.

HOSHIDA, H., FUJITA, T., CHA-AIM, K. & AKADA, R. 2013. N-Glycosylation deficiency enhanced heterologous production of a *Bacillus licheniformis* thermostable a-amylase in *Saccharomyces cerevisiae*. *Appl Microbiol Biotechnol,* 97**,** 5473-82.

HUH, W. K., FALVO, J. V., GERKE, L. C., CARROLL, A. S., HOWSON, R. W., WEISSMAN, J. S. & O'SHEA, E. K. 2003. Global analysis of protein localization in budding yeast. *Nature,* 425**,** 686-91.

JAYANI, R. S., SAXENA, S. & GUPTA, R. 2005. Microbial pectinolytic enzymes: A review. *Process Biochemistry,* 40**,** 2931-2944.

KANT, S., VOHRA, A. & GUPTA, R. 2013. Purification and physicochemical properties of polygalacturonase from *Aspergillus niger* MTCC 3323. *Protein Expr Purif,* 87**,** 11-6.

KASHYAP, D. R., VOHRA, P. K., CHOPRA, S. & TEWARI, R. 2001. Applications of pectinases in the commercial sector: a review. *Bioresour Technol,* 77**,** 215-27.

KAWABATA, A., SAWAYAMA, S. & URYU, K. 1974. A Study on the Contents of Pectic Substances in Fruits, Vegetable fruits and Nuts. *The Japanese Journal of Nutrition and Dietetics,* 32**,** 9-18.

KIM, D., SONG, L., BREITWIESER, F. P. & SALZBERG, S. L. 2016. Centrifuge: rapid and sensitive classification of metagenomic sequences. *Genome Res,* 26**,** 1721-1729.

LARRIBA, G., BASCO, R. D., ANDALUZ, E. & LUNA-ARIAS, J. P. 1993. Yeast exoglucanases. Where redundancy implies necessity. *Arch Med Res,* 24**,** 293-9.

LECAS, M., GUNATA, Z. Y., SAPIS, J.-C. & BAYONOVE, C. L. 1991. Purification and partial characterization of β-glucosidase from grape. *Phytochemistry,* 30**,** 451-454.

LEE, D., LLOYD, N. D., PRETORIUS, I. S. & BORNEMAN, A. R. 2016. Heterologous production of raspberry ketone in the wine yeast Saccharomyces cerevisiae via pathway engineering and synthetic enzyme fusion. *Microb Cell Fact,* 15**,** 1-7.

LI, L. L., MCCORKLE, S. R., MONCHY, S., TAGHAVI, S. & VAN DER LELIE, D. 2009. Bioprospecting metagenomes: glycosyl hydrolases for converting biomass. *Biotechnol Biofuels,* 2**,** 1-11.

LI, L. L., TAGHAVI, S., MCCORKLE, S. M., ZHANG, Y. B., BLEWITT, M. G., BRUNECKY, R., ADNEY, W. S., HIMMEL, M. E., BRUMM, P., DRINKWATER, C., MEAD, D. A., TRINGE, S. G. & LELIE, D. 2011. Bioprospecting metagenomics of decaying wood: mining for new glycoside hydrolases. *Biotechnol Biofuels,* 4**,** 1-13.

LOOSE, J. S., FORSBERG, Z., KRACHER, D., SCHEIBLBRANDNER, S., LUDWIG, R., EIJSINK, V. G. & VAAJE-KOLSTAD, G. 2016. Activation of bacterial lytic polysaccharide monooxygenases with cellobiose dehydrogenase. *Protein Sci,* 25**,** 2175-2186.

LOUW, C., LA GRANGE, D., PRETORIUS, I. S. & VAN RENSBURG, P. 2006. The effect of polysaccharide-degrading wine yeast transformants on the efficiency of wine processing and wine flavour. *J Biotechnol,* 125**,** 447-61.

MAICAS, S. & MATEO, J. J. 2005. Hydrolysis of terpenyl glycosides in grape juice and other fruit juices: a review. *Appl Microbiol Biotechnol,* 67**,** 322-35.

MANCHESTER, K. L. 1995. Louis Pasteur (1822-1895)--chance and the prepared mind. *Trends Biotechnol,* 13**,** 511-5.

MARAGKOUDAKIS, P. A., NARDI, T., BOVO, B., D'ANDREA, M., HOWELL, K. S., GIACOMINI, A. & CORICH, V. 2013. Biodiversity, dynamics and ecology of bacterial community during grape marc storage for the production of grappa. *Int J Food Microbiol,* 162**,** 143-51.

MARTINS, G., LAUGA, B., MIOT-SERTIER, C., MERCIER, A., LONVAUD, A., SOULAS, M. L., SOULAS, G. & MASNEUF-POMARE`DE, I. 2013. Characterization of

epiphytic bacterial communities from grapes, leaves, bark and soil of grapevine plants grown, and their relations. *PLoS One,* 8**,** 1-9.

MASNEUF-POMAREDE, I., BELY, M., MARULLO, P. & ALBERTIN, W. 2015. The Genetics of Non-conventional Wine Yeasts: Current Knowledge and Future Challenges. *Front Microbiol,* 6**,** 1-15.

MATEO, J. J. & DI STEFANO, R. 1997. Description of the β-glucosidase activity of wine yeasts. *Food Microbiology,* 14**,** 583-591.

MATEO, J. J. & JIMENEZ, M. 2000. Monoterpenes in grape juice and wines. *J Chromatogr A,* 881**,** 557-67.

MATURANO, Y. P., RODRIGUEZ ASSAF, L. A., TORO, M. E., NALLY, M. C., VALLEJO, M., CASTELLANOS DE FIGUEROA, L. I., COMBINA, M. & VAZQUEZ, F. 2012. Multi-enzyme production by pure and mixed cultures of *Saccharomyces* and non-*Saccharomyces* yeasts during wine fermentation. *Int J Food Microbiol,* 155**,** 43-50.

MCDONOUGH, M. A., KADIRVELRAJ, R., HARRIS, P., POULSEN, J.-C. N. & LARSEN, S. 2004. Rhamnogalacturonan lyase reveals a unique three-domain modular structure for polysaccharide lyase family 4. *FEBS letters,* 565**,** 188-194.

MERIN, M. G. & MORATA DE AMBROSINI, V. I. 2015. Highly cold-active pectinases under wine-like conditions from non-*Saccharomyces* yeasts for enzymatic production during winemaking. *Lett Appl Microbiol,* 60**,** 467-74.

MICHLMAYR, H. & KNEIFEL, W. 2014. β-Glucosidase activities of lactic acid bacteria: mechanisms, impact on fermented food and human health. *FEMS Microbiol Lett,* 352**,** 1-10.

MILLER, G. L. 1959. Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar. *Analytical Chemistry,* 31**,** 426-428.

MOLINA, A. M., SWIEGERS, J. H., VARELA, C., PRETORIUS, I. S. & AGOSIN, E. 2007. Influence of wine fermentation temperature on the synthesis of yeast-derived volatile aroma compounds. *Appl Microbiol Biotechnol,* 77**,** 675-87.

MORENO-ARRIBAS, M. V. & POLO, M. C. 2005. Winemaking biochemistry and microbiology: current knowledge and future trends. *Crit Rev Food Sci Nutr,* 45**,** 265-86.

MUTHUKUMAR, G., SUHNG, S. H., MAGEE, P. T., JEWELL, R. D. & PRIMERANO, D. A. 1993. The *Saccharomyces cerevisiae* SPR1 gene encodes a sporulation-specific exo-1,3-b-glucanase which contributes to ascospore thermoresistance. *J Bacteriol,* 175**,** 386-94.

NEBREDA, A. R., VILLA, T. G., VILLANUEVA, J. R. & DEL REY, F. 1986. Cloning of genes related to exo-β-glucanase production in *Saccharomyces cerevisiae*: characterization of an exo-β-glucanase structural gene. *Gene,* 47**,** 245-259.

NOGUCHI, H., TANIGUCHI, T. & ITOH, T. 2008. MetaGeneAnnotator: detecting species-specific patterns of ribosomal binding site for precise gene prediction in anonymous prokaryotic and phage genomes. *DNA Res,* 15**,** 387-96.

NORDESTGAARD, S. 2019. *AWRI Vineryard and Winery Practices Survey*.

NTOUGIAS, S., KAVROULAKIS, N., PAPADOPOULOU, K. K., EHALIOTIS, C. & ZERVAKIS, G. I. 2010. Characterization of cultivated fungi isolated from grape marc wastes through the use of amplified rDNA restriction analysis and sequencing. *J Microbiol,* 48**,** 297-306.

ONDOV, B. D., TREANGEN, T. J., MELSTED, P., MALLONEE, A. B., BERGMAN, N. H., KOREN, S. & PHILLIPPY, A. M. 2016. Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol,* 17**,** 132.

ORTEGA-REGULES, A., ROS-GARCÍA, J. M., BAUTISTA-ORTÍN, A. B., LÓPEZ-ROCA, J. M. & GÓMEZ-PLAZA, E. 2007. Differences in morphology and composition of skin and pulp cell walls from grapes (Vitis vinifera L.): technological implications. *European Food Research and Technology,* 227**,** 223-231.

ORTIZ, G. E., GUITART, M. E., ALBERTO, E., FERNANDEZ LAHORE, H. M. & BLASCO, M. 2014. Microplate assay for endo-polygalacturonase activity determination based on ruthenium red method. *Anal Biochem,* 454**,** 33-5.

OULAS, A., PAVLOUDI, C., POLYMENAKOU, P., PAVLOPOULOS, G. A., PAPANIKOLAOU, N., KOTOULAS, G., ARVANITIDIS, C. & ILIOPOULOS, I. 2015.

Metagenomics: tools and insights for analyzing next-generation sequencing data derived from biodiversity studies. *Bioinform Biol Insights,* 9**,** 75-88.

PALMERI, R. & SPAGNA, G. 2007. β-Glucosidase in cellular and acellular form for winemaking application. *Enzyme and Microbial Technology,* 40**,** 382-389.

PALOMO, E., HIDALGO, M., GONZALEZVINAS, M. & PEREZCOELLO, M. 2005. Aroma enhancement in wines from different grape varieties using exogenous glycosidases. *Food Chemistry,* 92**,** 627-635.

PEREZ-PONS, J. A., REBORDOSA, X. & QUEROL, E. 1995. Properties of a novel glucose-enhanced beta-glucosidase purified from *Streptomyces* sp. (ATCC 11238). *Biochim Biophys Acta,* 1251**,** 145-53.

PETERSEN, T. N., KAUPPINEN, S. & LARSEN, S. 1997. The crystal structure of rhamnogalacturonase A from *Aspergillus aculeatus*: a right-handed parallel b helix. *Structure,* 5**,** 533-44.

PIAO, H., HAWLEY, E., KOPF, S., DESCENZO, R., SEALOCK, S., HENICK-KLING, T. & HESS, M. 2015. Insights into the bacterial community and its temporal succession during the fermentation of wine grapes. *Front Microbiol,* 6**,** 1-12.

PICKERSGILL, R., SMITH, D., WORBOYS, K. & JENKINS, J. 1998. Crystal structure of polygalacturonase from *Erwinia carotovora* ssp. *carotovora*. *J Biol Chem,* 273**,** 24660-4.

PINTO, C., PINHO, D., SOUSA, S., PINHEIRO, M., EGAS, C. & GOMES, A. C. 2014. Unravelling the diversity of grapevine microbiome. *PLoS One,* 9**,** e85622.

POLIZELI, M. L., RIZZATTI, A. C., MONTI, R., TERENZI, H. F., JORGE, J. A. & AMORIM, D. S. 2005. Xylanases from fungi: properties and industrial applications. *Appl Microbiol Biotechnol,* 67**,** 577-91.

POLIZELI, M. L. T. M., JORGE, J. A. & TERENZI, H. F. 1991. Pectinase production by *Neurospora crassa*: purification and biochemical characterization of extracellular polygalacturonase activity. *J Gen Microbiol,* 137**,** 1815-23.

PRAKASH, T. & TAYLOR, T. D. 2012. Functional assignment of metagenomic data: challenges and applications. *Brief Bioinform,* 13**,** 711-27.

PRASANNA, V., PRABHA, T. N. & THARANATHAN, R. N. 2007. Fruit ripening phenomena-an overview. *Crit Rev Food Sci Nutr,* 47**,** 1-19.

PRETORIUS, I. S. 2000. Tailoring wine yeast for the new millennium: novel approaches to the ancient art of winemaking. *Yeast,* 16**,** 675-729.

PURI, A., PADDA, K. P. & CHANWAY, C. P. 2018. Evidence of endophytic diazotrophic bacteria in lodgepole pine and hybrid white spruce trees growing in soils with different nutrient statuses in the West Chilcotin region of British Columbia, Canada. *Forest Ecol Manag,* 430**,** 558-565.

QUAEDVLIEG, W., KEMA, G. H., GROENEWALD, J. Z., VERKLEY, G. J., SEIFBARGHI, S., RAZAVI, M., MIRZADI GOHARI, A., MEHRABI, R. & CROUS, P. W. 2011. *Zymoseptoria* gen. nov.: a new genus to accommodate *Septoria*-like species occurring on graminicolous hosts. *Persoonia,* 26**,** 57-69.

QUINLAN, R. J., SWEENEY, M. D., LO LEGGIO, L., OTTEN, H., POULSEN, J. C., JOHANSEN, K. S., KROGH, K. B., JORGENSEN, C. I., TOVBORG, M., ANTHONSEN, A., TRYFONA, T., WALTER, C. P., DUPREE, P., XU, F., DAVIES, G. J. & WALTON, P. H. 2011. Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components. *Proc. Natl. Acad. Sci. U.S.A.,* 108**,** 15079-84.

RADOI, F., KISHIDA, M. & KAWASAKI, H. 2005. Endo-polygalacturonase in *Saccharomyces* wine yeasts: effect of carbon source on enzyme production. *FEMS Yeast Res,* 5**,** 663-8.

REYNOLDS, A., CLIFF, M., GIRARD, B. & KOPP, T. G. 2001. Influence of fermentation temperature on composition and sensory properties of Semillon and Shiraz wines. *Am J Enol Viticult,* 52**,** 235-240.

RIOU, C., SALMON, J. M., VALLIER, M. J., GÜNATA, Z. & BARRE, P. 1998. Purification, Characterization, and Substrate Specificity of a Novel Highly Glucose-Tolerant β-Glucosidase from *Aspergillus oryzae*. *Appl Environ Microbiol,* 64**,** 3607-14.

SACCHI, K. L., BISSON, L. F. & ADAMS, D. O. 2005. A review of the effect of winemaking techniques on phenolic extraction in red wines. *American Journal of Enology and Viticulture,* 56**,** 197-206.

SAHAY, S., HAMID, B., SINGH, P., RANJAN, K., CHAUHAN, D., RANA, R. S. & CHAURSE, V. K. 2013. Evaluation of pectinolytic activities for oenological uses from psychrotrophic yeasts. *Lett Appl Microbiol,* 57**,** 115-21.

SALVETTI, E., CAMPANARO, S., CAMPEDELLI, I., FRACCHETTI, F., GOBBI, A., TORNIELLI, G. B., TORRIANI, S. & FELIS, G. E. 2016. Whole-Metagenome-Sequencing-Based Community Profiles of *Vitis vinifera* L. cv. Corvina Berries Withered in Two Post-harvest Conditions. *Front Microbiol,* 7**,** 1-17.

SANTOS, M., DIÁNEZ, F., DEL VALLE, M. G. & TELLO, J. C. 2007. Grape marc compost: microbial studies and suppression of soil-borne mycosis in vegetable seedlings. *World Journal of Microbiology and Biotechnology,* 24**,** 1493-1505.

SATHYA, T. A., JACOB, A. M. & KHAN, M. 2014. Cloning and molecular modelling of pectin degrading glycosyl hydrolase of family 28 from soil metagenomic library. *Mol Biol Rep,* 41**,** 2645-56.

SAULNIER, L. & THIBAULT, J. F. 1987. Extraction and Characterization of Pectic Substances from Pulp of Grape Berries. *Carbohydrate Polymers,* 7**,** 329-343.

SCHMEISSER, C., STEELE, H. & STREIT, W. R. 2007. Metagenomics, biotechnology with non-culturable microbes. *Appl Microbiol Biotechnol,* 75**,** 955-62.

SESTELO, A. B. F., POZA, M. & VILLA, T. G. 2004. β-Glucosidase Activity in a *Lactobacillus plantarum* Wine Strain. *World Journal of Microbiology and Biotechnology,* 20**,** 633-637.

SILACCI, M. W. & MORRISON, J. C. 1990. Changes in Pectin Content of Cabernet-Sauvignon Grape Berries during Maturation. *American Journal of Enology and Viticulture,* 41**,** 111-115.

SINCLAIR, G. & CHOY, F. Y. 2002. Synonymous codon usage bias and the expression of human glucocerebrosidase in the methylotrophic yeast, *Pichia pastoris. Protein Expr Purif,* 26**,** 96-105.

SINGH, R., KUMAR, M., MITTAL, A. & MEHTA, P. K. 2016. Microbial enzymes: industrial progress in 21st century. *3 Biotech,* 6**,** 1-15.

SOUFFRIAU, B., DEN ABT, T. & THEVELEIN, J. M. 2012. Evidence for rapid uptake of D-galacturonic acid in the yeast *Saccharomyces cerevisiae* by a channel-type transport system. *FEBS Lett,* 586**,** 2494-9.

STANKE, M., KELLER, O., GUNDUZ, I., HAYES, A., WAACK, S. & MORGENSTERN, B. 2006. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res,* 34**,** 435-9.

STEFANINI, I., ALBANESE, D., CAVAZZA, A., FRANCIOSI, E., DE FILIPPO, C., DONATI, C. & CAVALIERI, D. 2016. Dynamic changes in microbiota and mycobiota during spontaneous 'Vino Santo Trentino' fermentation. *Microb Biotechnol,* 9**,** 195-208.

TANG, H., WANG, S., WANG, J., SONG, M., XU, M., ZHANG, M., SHEN, Y., HOU, J. & BAO, X. 2016. N-hypermannose glycosylation disruption enhances recombinant protein production by regulating secretory pathway and cell wall integrity in *Saccharomyces cerevisiae. Sci Rep,* 6**,** 1-13.

TERWISSCHA VAN SCHELTINGA, A. C., ARMAND, S., KALK, K. H., ISOGAI, A., HENRISSAT, B. & DIJKSTRA, B. W. 1995. Stereochemistry of chitin hydrolysis by a plant chitinase/lysozyme and X-ray structure of a complex with allosamidin: evidence for substrate assisted catalysis. *Biochemistry,* 34**,** 15619-23.

TORRES, S., SAYAGO, J. E., ORDONEZ, R. M. & ISLA, M. I. 2011. A colorimetric method to quantify endo-polygalacturonase activity. *Enzyme Microb Technol,* 48**,** 123-8.

TRINDADE, L. V., DESAGIACOMO, C., POLIZELI, M. L., DAMASIO, A. R., LIMA, A. M., GOMES, E. & BONILLA-RODRIGUEZ, G. O. 2016. Biochemical Characterization, Thermal Stability, and Partial Sequence of a Novel Exo-Polygalacturonase from the Thermophilic Fungus *Rhizomucor pusillus* A13.36 Obtained by Submerged Cultivation. *Biomed Res Int,* 2016**,** 8653583.

UNKLES, S. E., VALIANTE, V., MATTERN, D. J. & BRAKHAGE, A. A. 2014. Synthetic biology tools for bioprospecting of natural products in eukaryotes. *Chem Biol,* 21**,** 502-508.

UNNO, T., IDE, K., YAZAKI, T., TANAKA, Y., NAKAKUKI, T. & OKADA, G. 2014. High Recovery Purification and Some Properties of a β-Glucosidase from *Aspergillus niger*. *Bioscience, Biotechnology, and Biochemistry,* 57**,** 2172-2173.

VAAJE-KOLSTAD, G., WESTERENG, B., HORN, S. J., LIU, Z., ZHAI, H., SORLIE, M. & EIJSINK, V. G. 2010. An oxidative enzyme boosting the enzymatic conversion of recalcitrant polysaccharides. *Science,* 330**,** 219-22.

VALCÁRCEL, M. C. & PALACIOS, V. 2008. Influence of `Novarom G' Pectinase β-glycosidase Enzyme on the Wine Aroma of four White Varieties. *Food Science and Technology International,* 14**,** 95-102.

VAN DER LEE, T. A. & MEDEMA, M. H. 2016. Computational strategies for genome-based natural product discovery and engineering in fungi. *Fungal Genet Biol,* 89**,** 29-36.

VAN RENSBURG, P. & PRETORIUS, I. S. 2000. Enzymes in Winemaking: Harnessing Natural Catalysts for Efficient Biotransformations - A Review. *S. Afr. J. Enol. Vitic.,* 21**,** 52-73.

VAN SLUYTER, S. C., MCRAE, J. M., FALCONER, R. J., SMITH, P. A., BACIC, A., WATERS, E. J. & MARANGON, M. 2015. Wine protein haze: Mechanisms of formation and advances in prevention. *Journal of Agricultural and Food Chemistry,* 63**,** 4020-4030.

VAN SLUYTER, S. C., WARNOCK, N. I., SCHMIDT, S., ANDERSON, P., VAN KAN, J. A., BACIC, A. & WATERS, E. J. 2013. Aspartic acid protease from *Botrytis cinerea* removes haze-forming proteins during white winemaking. *J Agric Food Chem,* 61**,** 9705-11.

VERGINER, M., LEITNER, E. & BERG, G. 2010. Production of volatile metabolites by grape-associated microorganisms. *J Agric Food Chem,* 58**,** 8344-50.

VIDAL, S., WILLIAMS, P., O'NEILL, M. A. & PELLERIN, P. 2001. Polysaccharides from grape berry cell walls. Part I: tissue distribution and structural characterization of the pectic polysaccharides. *Carbohydrate Polymers,* 45**,** 315-323.

VIEIRA GOMES, A. M., SOUZA CARMO, T., SILVA CARVALHO, L., MENDONCA BAHIA, F. & PARACHIN, N. S. 2018. Comparison of Yeasts as Hosts for Recombinant Protein Production. *Microorganisms,* 6**,** 1-23.

VOGTLE, F. N., BURKHART, J. M., RAO, S., GERBETH, C., HINRICHS, J., MARTINOU, J. C., CHACINSKA, A., SICKMANN, A., ZAHEDI, R. P. & MEISINGER, C. 2012. Intermembrane space proteome of yeast mitochondria. *Mol Cell Proteomics,* 11**,** 1840-52.

VOIRIN, S., BAUMES, R. & BAYONOVE, C. 1990. Synthesis and n.m.r. spectral properties of grape monoterpenyl glycosides. *Carbohydrate Research,* 207**,** 39-56.

WATANABE, T., TANI, M., ISHIBASHI, Y., ENDO, I., OKINO, N. & ITO, M. 2015. Ergosteryl-b-glucosidase (Egh1) involved in sterylglucoside catabolism and vacuole formation in *Saccharomyces cerevisiae*. *Glycobiology,* 25**,** 1079-89.

WATTS, A. G., DAMAGER, I., AMAYA, M. L., BUSCHIAZZO, A., ALZARI, P., FRASCH, A. C. & WITHERS, S. G. 2003. *Trypanosoma cruzi trans*-sialidase operates through a covalent sialyl-enzyme intermediate: tyrosine is the catalytic nucleophile. *J Am Chem Soc,* 125**,** 7532-3.

WEST, P. T., PROBST, A. J., GRIGORIEV, I. V., THOMAS, B. C. & BANFIELD, J. F. 2018. Genome-reconstruction for eukaryotes from complex natural microbial communities. *Genome Res,* 28**,** 569-580.

WILLIAMS, P. J., STRAUSS, C. R., WILSON, B. & MASSYWESTROPP, R. A. 1982. Novel Monoterpene Disaccharide Glycosides of Vitis-Vinifera Grapes and Wines. *Phytochemistry,* 21**,** 2013-2020.

WITHERS, S. G., DOMBROSKI, D., BERVEN, L. A., KILBURN, D. G., MILLER, R. C., JR., WARREN, R. A. & GILKES, N. R. 1986. Direct ^{1}H n.m.r. determination of the stereochemical course of hydrolyses catalysed by glucanase components of the cellulase complex. *Biochem Biophys Res Commun,* 139**,** 487-94.

WOOD, D. E., LU, J. & LANGMEAD, B. 2019. Improved metagenomic analysis with Kraken 2. *Genome Biol,* 20**,** 1-12.

YIP, V. L., THOMPSON, J. & WITHERS, S. G. 2007. Mechanism of GlvA from *Bacillus subtilis*: a detailed kinetic analysis of a 6-phospho-alpha-glucosidase from glycoside hydrolase family 4. *Biochemistry,* 46**,** 9840-52.

YIP, V. L. & WITHERS, S. G. 2006. Breakdown of oligosaccharides by the process of elimination. *Curr Opin Chem Biol,* 10**,** 147-55.

ZARRAONAINDIA, I., OWENS, S. M., WEISENHORN, P., WEST, K., HAMPTON-MARCELL, J., LAX, S., BOKULICH, N. A., MILLS, D. A., MARTIN, G., TAGHAVI, S., VAN DER LELIE, D. & GILBERT, J. A. 2015. The soil microbiome influences grapevine-associated microbiota. *mBio,* 6**,** 1-10.

ZHANG, Y., MIN, Z., QIN, Y., YE, D. Q., SONG, Y. Y. & LIU, Y. L. 2019. Efficient Display of *Aspergillus niger* b-Glucosidase on *Saccharomyces cerevisiae* Cell Wall for Aroma Enhancement in Wine. *J Agric Food Chem,* 67**,** 5169-5176.

ZHOU, M., GUO, P., WANG, T., GAO, L., YIN, H., CAI, C., GU, J. & LU, X. 2017. Metagenomic mining pectinolytic microbes and enzymes from an apple pomace-adapted compost microbial community. *Biotechnol Biofuels,* 10**,** 1-15.

ZHU, F. M., DU, B. & LI, J. 2014. Aroma enhancement and enzymolysis regulation of grape wine using b-glycosidase. *Food Sci Nutr,* 2**,** 139-45.