Directed evolution of wine-related lactic acid bacteria and characterisation of evolved isolates

by

Seipati Precious Tenyane

Dissertation presented for the degree of

Doctor of Philosophy (Agricultural Science)

at

Stellenbosch University

South African Grape and Wine Research Institute, Department of Viticulture and Oenology, Faculty of AgriSciences

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at are those of the author and are not necessarily to be attributed to the NRF.

Supervisor: Prof FF Bauer
Co-supervisor: Prof M du Toit
Co-supervisor: Dr D Rossouw

March 2020

Declaration

By submitting this dissertation electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated) that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: March 2020

Summary

Microorganisms form part of complex ecological networks, governed by either metabolic, physical or molecular processes that have positive, neutral or negative effects on microbial interactions. Understanding microbial interactions provides the opportunity to control and manipulate microbes for different biotechnological and industrial applications. For example, the production of beverages such as wine shows how microbial interactions can be controlled and manipulated to achieve desired outcomes. One example is the deliberate inoculation of lactic acid bacteria (LAB) such as *Oenococcus oeni* or *Lactobacillus plantarum* to inhibit the growth of spoilage bacteria by depleting available carbon sources such as L-malic acid in a process known as malolactic fermentation (MLF). Indeed, wine provides a good model to study microbial interactions because grape must is inhabited by multiple species of filamentous fungi, yeast, acetic acid bacteria (AAB) and LAB in an anthropogenic and relatively controlled environment.

In this study, I investigated the impact of the interaction between the wine yeast *Saccharomyces cerevisiae* and the LAB *L. plantarum*. Briefly, the impact of the yeast on the evolution of the bacteria was evaluated after 50 and 100 generations first phenotypically, followed by a genome-wide analysis to identify genetic targets of evolution. A serial transfer method was used for the directed evolution (DE) experiments, introducing bottlenecks and fluctuation between nutrient rich and poor environments after each transfer. This strategy results in a 'feast-and-famine' regime, which results in conflicting selective pressures, resembling what normally occurs in dynamic natural environments, which was important here to generate robust and resilient bacteria. Additionally, two yeast strains were used to investigate whether microbial interactions result in yeast-specific adaptations or generic adaptations. Therefore, the yeast strains were kept constant by discarding the yeast at the end of each DE cycle and re-inoculating the mother culture at the start of each DE cycle.

The data show yeast strain-specific phenotypes for isolates evolved for 50 generations. Genome-wide analysis showed that broadly targeted pathways are peptidoglycan biosynthesis and degradation, nucleic acid processing, and carbohydrate transport and metabolism in isolates evolved for 50 and 100 generations. These data show that yeast-driven DE results in yeast-specific phenotypic variations and high genetic diversity, but also in convergent evolution over time. The results obtained in this study suggest that yeast drive the evolution of bacteria by dominating the metabolic landscape, showing that strong competitive interactions promote positive selection in mixed species communities, and weak competitive interactions results in no adaptation. This work enriches our understanding of yeast-bacteria interactions over time. Moreover, an isolate that is superior to the parent strain in terms of growth and MLF was obtained, showing potential as a starter culture for winemaking.

Opsomming

Mikroörganismes maak deel uit van komplekse ekologiese netwerke wat deur metaboliese, fisiese of molekulêre prosesse beheer word, en dit het positiewe, neutrale of negatiewe effekte op mikrobiese interaksies. Insig in mikrobiese interaksies bied die geleentheid om mikrobes vir verskillende biotegnologiese en nywerheidstoepassings te kontroleer en te manipuleer. Die produksie van drinkgoed soos wyn toon byvoorbeeld hoe mikrobiese interaksies beheer en gemanipuleer kan word om die gewenste uitkomste te bereik. Een voorbeeld is die doelbewuste inenting van melksuurbakterieë (LAB) soos Oenococcus oeni of Lactobacillus plantarum om die groei van bederfbakterieë te belemmer deur beskikbare koolstofbronne soos L appelsuur in ’n proses genaamd malolaktiese fermentasie (MLF) te verarm. Wyn verskaf inderwaarheid ’n goeie model vir die bestudering van mikrobiese interaksies, aangesien daar verskeie spesies filamentagtige swamme, gis, asynsuurbakterieë (AAB) en LAB in ’n antropogeniese en relatief beheerde omgewing in druiwemos voorkom.

In hierdie studie het ek die impak van die wisselwerking tussen die wyngis Saccharomyces cerevisiae en die LAB L. plantarum ondersoek. Kortliks is die invloed van die gis op die evolusie van die bakterieë eers ná 50 en 100 generasies fenotipies geëvalueer, gevolg deur ’n genoomwye ontleding om genetiese teikens vir evolusie te identifiseer. ’n Reeksoordragmetode is vir die gerigte evolusie- (DE)-eksperimente gebruik, wat knelpunte en fluktuasie tussen voedingsryke en swak omgewings ná elke oordrag ingevoer het. Hierdie strategie het tot ’n “fees en hongersnood” regime gelei, met gevolglike teenstrydige selektiewe druk en voorkomste wat normaalweg in dinamiese natuurlike omgewings aangeneem word; hier belangrik vir die generering van robuuste en veerkragtige bakterieë. Daarbenewens is twee gisstamme gebruik om te vas te stel of mikrobiese interaksies gisspesifieke aanpassings of generiese aanpassings tot gevolg het. Daarom is die gisstamme konstant gehou deur die gis aan die einde van elke DE-siklus weg te gooi en die moederkultuur opnuut aan die begin van elke DE-siklus in te ent.

Die data dui daarop dat gisstam spesifieke fenotipes vir isolate oor 50 generasies heen ontwikkel het. Genoomwye ontledings toon die breedweg geteikende roetes omvat peptidoglikaanse biosintese en afbreking, nukleïensuurprosessering, asook koolhidraatvervoer en metabolisme in isolate, wat oor 50 en 100 generasies ontwikkel het. Hierdie data toon verder dat gisgedrewe DE tot gisspesifieke fenotipiese variasies en hoë genetiese diversiteit, ingesluit konvergente evolusie, oor tyd aanleiding gee. Die resultate wat in hierdie studie verkry is, dui daarop dat gis die evolusie van bakterieë dryf deur die metaboliese landskap te oorheers, wat wys dat sterk mededingende interaksies positiewe seleksie in gemengde spesiegemeenskappe aanmoedig, terwyl swak mededingende interaksies geen aanpassing tot gevolg het nie. Hierdie werk verryk ons begrip van gisbakterie interaksies oor tyd. Daarbenewens is ’n isolaat verkry wat beter as die ouerstam is sover dit groei en MLF betref, en oor die potensiaal beskik om as ’n aansitkultuur vir wynmaak te dien.

This dissertation is dedicated to my family

Science does not disprove the existence of God, it endorses it.

Biographical sketch

Seipati Precious Tenyane was born in Sebokeng, South Africa in 1993. She matriculated from Dinwiddie High School, Germiston, South Africa in 2010. She enrolled at the University of the Witwatersrand, Johannesburg in 2011 and obtained a Bachelor of Science degree in 2013. In 2014, she obtained her BSc Honours degree in Wine Biotechnology at the Institute for Wine Biotechnology, Stellenbosch University, where she enrolled for an MSc in Wine Biotechnology. In 2016, July, her MSc was upgraded to a PhD in Wine Biotechnology.

Acknowledgements

I wish to express my sincere gratitude and appreciation to the following persons and institutions:

- **Prof FF Bauer, Prof M du Toit,** and **Dr D Rossouw** for conceptualising this project, for sharing their expertise and knowledge, and for all the support and encouragement they provided these past few years. And for critically evaluating this dissertation.

- **Prof H Patterton,** and **Dr RN De Witt** for all the support, teaching and guidance on all the bioinformatics analyses.

- **Mr CJ Van Heerden** for critically evaluating Chapter 5 of this dissertation and for his encouragement and additional bioinformatics support.

- **Mrs W Kühn, Dr SC Fairbairn** and **Mrs AN du Toit** for assistance with metabolite analyses

- The **National Research Foundation (NRF), South African Grape and Wine Research Institute (SAGWRI), Wine Industry Network for Expertise and Technology (Winetech),** and **NRF** through **SARChI grant 83471** for financial assistance throughout my studies towards this PhD.

- **Mrs K Vergeer**, **Mrs EL De Villiers**, and **Ms N Radasi** for administrative support.

- **Dr RK Naidoo, Ms SC du Toit**, **Mr IJ Botma**, **Mr JR Smith**, and all my friends and colleagues at **SAGWRI** for their support and thought-provoking discussions.

- **Dr ME Setati** and **Dr YT Motlhalamme** for their encouragement, empathy and friendship which I will cherish for years to come.

- **Ms A Mahanjana** and **Mr SWC Ngara** for their love, encouragement and support. For being there for me even when I neglected them and was totally engrossed in writing this dissertation. I am eternally grateful to each of them.

- My mother, **MJ Tenyane** for being a strong woman who showed me that I can do absolutely anything through prayer and faith.

- My brother, **TK Tenyane**, and my sister **L Tenyane** for always lending an ear for me to vent. For their continuous support and encouragement.

- **The Lord, God Almighty** for giving me life, love, family, colleagues and for getting me here.

Preface

This dissertation is presented as a compilation of 6 chapters. Each chapter is introduced separately and is written according to the style of the journal South African Journal of Enology and Viticulture.

Chapter 1 **General introduction and project aims**

Chapter 2 **Literature review**

Impact of microbial interactions on the evolution of microbes: examples from the wine environment

Chapter 3 **Research results**

Biotic selection pressure by yeast strains leads to yeast strain-specific phenotypes in *Lactobacillus plantarum*

Chapter 4 **Research results**

Comparative genomics of yeast-driven evolved isolates of *Lactobacillus plantarum*

Chapter 5 **Research results**

Going too deep? Downstream sequencing assembly errors in whole genome sequencing using the Ion Proton System

Chapter 6 **General discussion and conclusions**

I hereby declare that, apart from the contributions made by others as stated below, I was responsible for performing all experimental work, including the directed evolution experiments, screening for different phenotypes, yeast-strain specificity screening, DNA extractions, some of the chemical analyses, data preparation and interpretation, bioinformatics analyses, interpreting and compiling the written work presented in this dissertation. Mr CJ Van Heerden was involved in the critical evaluation of the data and work presented in Chapter 5. Indexing and normalising the parental genome contig file for use on FreeBayes for variant calling was performed by Dr Riaan De Witt at the Centre for Bioinformatics and Computational Biology (Department of Biochemistry, Stellenbosch University). My supervisors Prof. FF Bauer and Dr D Rossouw conceptualised the study, and together with Prof. M du Toit, continuously provided an overall critical evaluation of the results and research.

Table of Contents

Chapter 1

General Introduction and Project Aims

Chapter 1: General introduction and project aims

1.1. Project rationale

Wine environments are inhabited by various microbial species originating from the vineyard, the cellar or introduced by human intervention (Ultee *et al.*, 2013; Kántor *et al.*, 2017; Steensels *et al.*, 2019). These environments are characterised by harsh parameters such as pH below 3.5 (Succi *et al.*, 2017), dynamic temperatures ranging between 12-35ºC (Henderson *et al.*, 2013), acidity > 5 g/l (Volschenk & Van Vuuren, 2006; Vilela, 2019), high sugar (>200 g/l) and low oxygen levels between 1 and 8 mg/l (du Toit *et al.*, 2006; Biyela *et al.*, 2009). Therefore, only species that can withstand these environments dominate and survive. The best adapted species include the eukaryotic yeast *Saccharomyces cerevisiae* and the lactic acid bacterium *Oenococcus oeni*, which tend to dominate alcoholic and malolactic fermentation (MLF) respectively (Saguir *et al.*, 2009; Albergaria & Arneborg, 2016). Recently *Lactobacillus plantarum* was also identified as a suitable starter culture - and has since been commercialised for MLF (Lerm *et al.*, 2011). The dominance of these species in wine is as a result of complex microbial interactions and extensive genetic adaptation within this environment (Novo *et al.*, 2009; Lorentzen & Lucas, 2019). Microbial interactions are generally classified in several ecological categories such as competition, inhibition and stimulation and have been studied extensively (Ivey *et al.*, 2013). In wine environments competition and inhibition between wine yeast and lactic acid bacteria (LAB) have been characterised, primarily from a metabolic perspective (see review by Balmaseda *et al.*, 2018). However, the specific physiological and molecular mechanisms that govern these interactions between yeast and LAB are not well understood.

1.2. Introduction

LAB, like yeast, are very important in winemaking as they add organoleptic complexity to wine, limit the likelihood of spoilage and reduce wine acidity (Fleet, 1984). LAB do this by producing L-lactic acid (perceived as a soft acid from a sensory perspective) and carbon dioxide from L-malic acid (perceived as a harsh acid) in a process known as MLF. Species of the genera *Pediococcus, Lactobacillus* and *Oenococcus* can conduct MLF (Lerm *et al.*, 2010). However, only some strains of *Lactobacillus plantarum* and *Oenococcus oeni* are available as commercial starter cultures (Lerm *et al.*, 2011). These strains are particularly acidophilic and are best able to proliferate in the harsh wine environment characterised by low pH (mostly <3.5), high ethanol (>10% v/v), oxygen levels below 8 mg/l, nutrient deficiency and sulphur dioxide (>30 mg/l). The latter is usually added by the winemaker but may also be produced by fermenting yeast (Garvie, 1967; Lonvaud-Funel, 1999; Lerm *et al.*, 2011; Wells & Osborne, 2011). Despite their ability to survive in this harsh environment, LAB frequently fail to complete MLF in both spontaneous and inoculated fermentations. Apart from physiochemical factors, the impact of inhibitory metabolites produced by

the fermenting yeast strains has been highlighted as one possible cause (Capucho & San Romão, 1994; Nehme *et al.*, 2008; Lerm *et al.*, 2010; Branco *et al.*, 2014; Rizk *et al.*, 2016, 2018). Data are however frequently contradictory, suggesting complex interactions between yeast and bacteria and multifactorial causes for fermentation failures. Such data also suggest that strain compatibility has a significant impact on the performance of LAB, i.e. that specific bacterial strains will perform better when paired with specific yeast strains (Lerm *et al.*, 2010; Liu *et al.*, 2016). However, the specific nature of these interactions is unknown, but may be described as competitive (Hoek *et al.*, 2016; du Plessis *et al.*, 2017; Janse van Rensburg, 2018), inhibitory (Zapparoli *et al.*, 2003; Bayrock & Ingledew, 2004; Osborne & Edwards, 2006; Dierings *et al.*, 2013; Rizk *et al.*, 2018) or stimulatory (Nehme *et al.*, 2008; Ponomarova *et al.*, 2017) response patterns.

During alcoholic fermentation the wine environment becomes progressively more inhospitable for all microorganisms, including bacteria, due to increases in concentrations of ethanol and other metabolites such as medium-chain fatty acids (Zapparoli *et al.*, 2009; Balmaseda *et al.*, 2018). In spontaneous fermentations, or fermentations inoculated with yeast only (currently the most common practice), yeast will dominate the alcoholic fermentation stage, in the presence of limited but viable bacterial populations (Alexandre *et al.*, 2004). In particular, populations of LAB will persist, and will start to degrade malic acid once alcoholic fermentation has ceased and yeast populations are in decline (Henick-Kling & Park, 1994). Data have shown that the ability of specific LAB strains to carry out this secondary fermentation in these conditions, at least in part, depends on the specific yeast strain – inoculated or not. These data suggest specific interactions between the two species even when the two steps of fermentation are separated in time (Cañas *et al.*, 2015; Liu *et al.*, 2016; Lasik-Kurdyś *et al.*, 2017).

Furthermore, many winemakers are co-inoculating yeast and LAB starter cultures to acclimatise the LAB to the environment earlier on and to increase the likelihood of complete MLF (Knoll *et al.*, 2012; Cañas *et al.*, 2015; Tristezza *et al.*, 2016; Versari *et al.*, 2016; Lasik-Kurdyś *et al.*, 2017). This brings yeast and LAB in direct contact at very high cell densities at the early stages of the process. These practices further increase the likelihood of interactions between the yeast and LAB (Yamasaki-Yashiki *et al.*, 2017). Indeed, it was shown that yeast and LAB compete for the same nutrients for growth (Volschenk *et al.*, 2003; Bayrock & Ingledew, 2004; du Plessis *et al.*, 2017; Balmaseda *et al.*, 2018; Janse van Rensburg, 2018).

Furthermore, Liu *et al.* (2016) investigated the exometabolomic profiles of MLF+ (yeast strains that promote MLF) and MLF- (yeast strains that inhibit MLF) yeast phenotypes. Their data suggest that yeast associated with sulphur-containing peptides tend to diminish the fermentative ability of LAB, while yeast that promote MLF release phenolic compounds, amino acids, peptides and carbohydrates (Liu *et al.*, 2016). In that study, however, only the by-products of the metabolism of

the yeast were reported, but the specific manner by which these compounds influence the yeast-LAB interactions is still lacking.

Some studies have sought to understand microbial interactions by establishing synthetic biological systems in the context of yeast-yeast (Shou *et al.*, 2007), yeast-algae (Hom & Murray, 2014; Naidoo *et al.*, 2019), yeast-bacteria (Zhou, Bottagisi, *et al.*, 2017; Zhou, Swamy, *et al.*, 2017; du Toit, 2018) and bacteria-bacteria interactions (Scott *et al.*, 2017). The data suggest that two microbial populations that may or may not co-exist naturally together can co-exist in an obligatory mutualism if each is made to depend solely on the metabolism of the other by exchange of essential nutrients such as amino acids. In these studies, mutualisms are based on reciprocal dependency for nutrients.

Contrary to the aforementioned studies, the current study sought to apply a directed evolution–based approach to investigate and characterise the interactions between strains of *S. cerevisiae* and *L. plantarum* (Fig. 1.1). Directed evolution (DE) is a tool frequently used to improve industrial phenotypes. It involves subjecting a microbial population to a selective pressure(s) over many generations to steer the population towards a desired phenotype, which will give the population a selective advantage over the parent population (Olson-Manning *et al.*, 2012). Several studies have applied directed evolution to improve LAB strains for industry-relevant phenotypes. In these studies, abiotic selective pressures (ethanol, sulphur-dioxide, pH, bacteriocins, etc.) are imposed on the organism of interest (Table 1.1). In contrast, this work involved the application of a biotic selective pressure in the form of actively fermenting wine yeast to evolve a strain of *L. plantarum.* The approach evaluates how the presence of another organism modulates the evolution of this bacteria. This is one of the first studies to apply such a strategy with the aim to investigate the underlying mechanisms involved in yeast-bacteria interactions and how these interactions may have shaped the evolution of the wine ecosystem.

Table 1.1 Sources of the application of directed evolution to improve strains of lactic acid bacteria for industrially-relevant phenotypes.

Species/strain	Selective pressure	Desired trait	Reference
Oenococcus oeni SB3	Ethanol (increased to 15% v/v)	High ethanol tolerance	(Betteridge *et al.*, 2018)
Oenococcus oeni A90	Increasing levels of pH (up to 3.5), ethanol (up to 15.1% v/v) and SO_2 (up to 26 mg/l)	Improved MLF	(Jiang *et al.*, 2018)
Lactococcus lactis strains	Bacteriocin Lcn972	Diverse phenotypes	(López-González *et al.*, 2018)

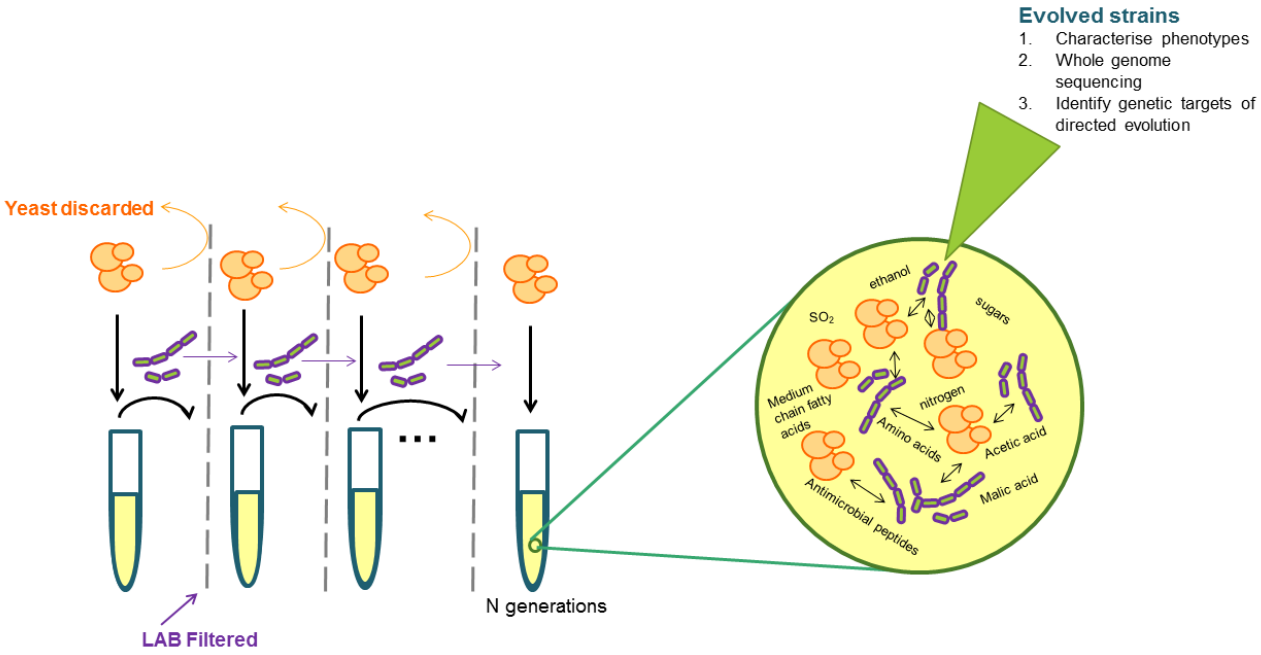

Figure 1.1 Representation of the directed evolution experiment for 50 and 100 generations. *Lactobacillus plantarum* and *S. cerevisiae* are co-inoculated in synthetic grape must, in triplicate. These species are therefore in close proximity and will share metabolic information. As has been shown in literature various interactions occur between yeast and bacteria and this includes nutrient competition (amino acids, organic acids), inhibition (production of growth-limiting metabolites i.e. ethanol, antimicrobial peptides and medium chain fatty acids), and stimulation (release of essential nutrients such as amino acids).

1.3. Project aims

This study is part of a bigger project at the South African Grape and Wine Research Institute (SAGWRI) that focuses on the application of synthetic microbial ecosystems to improve bioprocesses and adapt microorganisms for potential industry use. At the same time, these approaches are designed to better understand the mechanisms involved in interspecies interactions and the way these may have shaped evolutionary selection. The overarching aim of this study is to use a biotic selection driver (*S. cerevisiae*) to direct the evolution of *L. plantarum* and to investigate the underlying genomic mechanisms that govern microbial interactions. For this purpose, full genome sequences of evolved isolates were compared with the parental genome. Two commercial yeast strains were co-inoculated with a strain of *L. plantarum* over the course of MLF. The bacterial population was re-pitched into fresh media, together with the original mother culture of the yeast. In this set-up, the bacterial population could freely evolve, while the yeast remained unchanged. This strategy prevents a co-evolutionary 'arms race' where both species would co-evolve with one another (Zhou, Swamy, *et al.*, 2017; du Toit, 2018; Naidoo *et al.*, 2019). Two different yeast strains were utilised because different yeast strains can impact LAB differently during wine fermentations (Liu *et al.*, 2016). This approach will focus on the genetic targets and strategies of LAB adaptation to competitive yeast in particular. The objectives of this study are as follows:

i. to use directed evolution as a tool to improve strains of *L. plantarum* IWBT B063 to generate evolved isolates with increased fitness (in terms of growth and MLF) when used in combination with the 'driver' yeast:
   - to identify and characterise potentially evolved isolates and verify their 'fitness' experimentally across a variety of conditions;
   - to determine the specificity of the improved isolates and investigate whether they are better adapted to only the original 'driver' yeast, or to yeast in general.
   - to test the phenotypes in other stressful environments (i.e. real grape must) to validate the robustness of the interactions.

ii. to conduct whole genome sequencing of the evolved strains to identify the relevant genetic changes;
   - to assemble and annotate the genome and identify single nucleotide polymorphisms (SNPs);
   - to conduct an *in-silico* investigation to identify targets of directed evolution; and
   - to generate hypotheses and to carry out gene/enzyme functional analysis to elucidate which genes have been affected by our experimental design.

iii. to evaluate impact of deep and ultra-deep sequencing on LAB genome datasets
   - establish a coverage (depth) threshold for whole-genome sequencing of LAB using Ion Torrent sequencing

The outcomes of the research objectives are presented as follows in the thesis:

The directed evolution experiment of *L. plantarum* in co-culture with the *S. cerevisiae* strains EC1118® and Cross Evolution® was carried out in parallel, but the characterisation of bacterial colony isolates in the presence of either yeast was conducted separately. Characterisation of the colony isolates is presented in Chapter 3. A total of 5 evolved strains (colony isolates) were selected and tested in real grape must (Chapter 3). In Chapter 4, I evaluated whole-genome variation between each of the evolved isolates and the parent strain. In addition, 11 isolates were selected after 100 generations and sequenced to compare genome variation over time. Lastly, the impact of deep and ultra-deep sequencing of small genomes was evaluated (Chapter 5). This was done because coincidentally, I observed higher sequencing error of the parent strain with ultra-deep sequencing. Therefore, establishing a coverage (depth) threshold for whole-genome sequencing is important to reduce false positive data with downstream analyses (Desai *et al.*, 2013).

The research chapters in this thesis follow the literature review in Chapter 2 and the major outcomes of the study, new insights and hypotheses, and future research are discussed in Chapter 6.

References

Albergaria, H. & Arneborg, N., 2016. Dominance of *Saccharomyces cerevisiae* in alcoholic fermentation processes: role of physiological fitness and microbial interactions Appl. Microbiol. Biotechnol. 100, 5, 2035–2046.

Alexandre, H., Costello, P.J., et al., 2004. *Saccharomyces cerevisiae-Oenococcus oeni* interactions in wine: Current knowledge and perspectives Int. J. Food Microbiol. 93, 2, 141–154.

Balmaseda, A., Bordons, A., et al., 2018. Non-*Saccharomyces* in wine: Effect upon *Oenococcus oeni* and malolactic fermentation Front. Microbiol. 9, 1–8.

Bayrock, D.P. & Ingledew, W.M., 2004. Inhibition of yeast by lactic acid bacteria in continuous culture: Nutrient depletion and/or acid toxicity? J. Ind. Microbiol. Biotechnol. 31, 8, 362–368.

Betteridge, A.L., Sumby, K.M., et al., 2018. Application of directed evolution to develop ethanol tolerant *Oenococcus oeni* for more efficient malolactic fermentation Appl. Microbiol. Biotechnol. 102, 2, 921–932.

Biyela, B.N.E., du Toit, W.J., et al., 2009. The production of reduced-alcohol wines using Gluzyme Mono® 10.000 BG-treated grape juice South African J. Enol. Vitic. 30, 2, 124–132.

Branco, P., Francisco, D., et al., 2014. Identification of novel GAPDH-derived antimicrobial peptides secreted by *Saccharomyces cerevisiae* and involved in wine microbial interactions Appl. Microbiol. Biotechnol. 98, 2, 843–853.

Cañas, P.M.I., Romero, E.G., et al., 2015. Sequential inoculation versus co-inoculation in Cabernet Franc wine fermentation Food Sci. Technol. Int. 21, 3, 203–212.

Capucho, I. & San Romão, M. V., 1994. Effect of ethanol and fatty acids on malolactic activity of *Leuconostoc oenos* Appl. Microbiol. Biotechnol. 42, 2–3, 391–395.

Desai, A., Marwah, V.S., et al., 2013. Identification of optimum sequencing depth especially for de novo genome assembly of small genomes using next generation sequencing data PLoS One 8, 4, 60204.

Dierings, L.R., Braga, C.M., et al., 2013. Population dynamics of mixed cultures of yeast and lactic acid bacteria in cider conditions Brazilian Arch. Biol. Technol. 56, 5, 837–847.

du Plessis, H.W., du Toit, M., et al., 2017. Characterisation of non-*Saccharomyces* yeasts using different methodologies and evaluation of their compatibility with malolactic fermentation South African J. Enol. Vitic. 38, 1, 46–63.

du Toit, S.C., 2018. Coevolution of *Saccharomyces cerevisiae* and *Lactobacillus plantarum*: Engineering interspecies cooperation. Master thesis. Stellenbosch University.

du Toit, W.J., Marais, J., et al., 2006. Oxygen in must and wine: A review South African J. Enol. Vitic. 27, 1, 76–94.

Fleet, G.H., 1984. Evolution of yeasts and lactic acid bacteria during fermentation and storage of Bordeaux wines 48, 5, 1034–1038.

Garvie, E.I., 1967. *Leuconostoc oenos* sp.nov. J. Gen. Microbiol. 48, 3, 431–438.

Henderson, C.M., Zeno, W.F., et al., 2013. Fermentation temperature modulates phosphatidylethanolamine and phosphatidylinositol levels in the cell membrane of *Saccharomyces cerevisiae* Appl. Environ. Microbiol. 79, 17, 5345–5356.

Henick-Kling, T. & Park, Y.H., 1994. Considerations for the use of yeast and bacterial starter cultures: SO2 and timing of inoculation Am. J. Enol. Vitic. 45, 4, 464–469.

Hoek, T.A., Axelrod, K., et al., 2016. Resource availability modulates the cooperative and competitive nature of a microbial cross-feeding mutualism PLoS Biol. 14, 8, 1–17.

Hom, E.F.Y. & Murray, A.W., 2014. Niche engineering demonstrates a latent capacity for fungal-algal mutualism Science. 345, 6192, 94–98.

Ivey, M., Massel, M., et al., 2013. Microbial interactions in food fermentations Annu. Rev. Food Sci. Technol. 4, 141–162.

Janse van Rensburg, P.J., 2018. Investigating the influence of a wine yeast consortium on population dynamics, alcoholic and malolactic fermentation. Master thesis. Stellenbosch University.

Jiang, J., Sumby, K.M., et al., 2018. Directed evolution of *Oenococcus oeni* strains for more efficient malolactic fermentation in a multi-stressor wine environment Food Microbiol. 73, 150–159.

Kántor, A., Mareček, J., et al., 2017. Microorganisms of grape berries Proc. Latv. Acad. Sci. Sect. B Nat. Exact, Appl. Sci. 71, 6, 502–508.

Knoll, C., Fritsch, S., et al., 2012. Impact of different malolactic fermentation inoculation scenarios on Riesling wine aroma World J. Microbiol. Biotechnol. 28, 1143–1153.

Lasik-Kurdyś, M., Gumienna, M., et al., 2017. Influence of malolactic bacteria inoculation scenarios on the efficiency of the vinification process and the quality of grape wine from the Central European region Eur. Food Res. Technol. 243, 2163–2173.

Lerm, E., Engelbrecht, L., et al., 2010. Malolactic fermentation: The ABC's of MLF South African J. Enol. Vitic. 31, 2, 186–212.

Lerm, E., Engelbrecht, L., et al., 2011. Selection and characterisation of *Oenococcus oeni* and *Lactobacillus plantarum* South African wine isolates for use as malolactic fermentation starter cultures South African J. Enol. Vitic. 32, 2, 280–295.

Liu, Y., Forcisi, S., et al., 2016. New molecular evidence of wine yeast-bacteria interaction unraveled by non-targeted exometabolomic profiling Metabolomics 12, 69, 1–16.

Lonvaud-Funel, A., 1999. Lactic acid bacteria in the quality improvement and depreciation of wine Antonie van Leeuwenhoek, Int. J. Gen. Mol. Microbiol. 76, 317–331.

López-González, M.J., Escobedo, S., et al., 2018. Adaptive evolution of industrial *Lactococcus lactis* under cell envelope stress provides phenotypic diversity Front. Microbiol. 9, November, 1–17.

Lorentzen, M.P.G. & Lucas, P.M., 2019. Distribution of *Oenococcus oeni* populations in natural habitats Appl. Microbiol. Biotechnol. 103, 2937–2945.

Naidoo, R.K., Simpson, Z.F., et al., 2019. Nutrient exchange of carbon and nitrogen promotes the formation of stable mutualisms between *Chlorella sorokiniana* and *Saccharomyces cerevisiae* under engineered synthetic growth conditions Front. Microbiol. 10, 609.

Nehme, N., Mathieu, F., et al., 2008. Quantitative study of interactions between *Saccharomyces cerevisiae* and *Oenococcus oeni* strains J Ind Microbiol Biotechnol 35, 7, 685–693.

Novo, M., Bigey, F., et al., 2009. Eukaryote-to-eukaryote gene transfer events revealed by the genome sequence of the wine yeast *Saccharomyces cerevisiae* EC1118 Proc. Natl. Acad. Sci. U. S. A. 106, 38, 16333–16338.

Olson-Manning, C.F., Wagner, M.R., et al., 2012. Adaptive evolution Nat. Rev. Genet. 13, 1938, 867–877.

Osborne, J.P. & Edwards, C.G., 2006. Inhibition of malolactic fermentation by *Saccharomyces* during alcoholic fermentation under low- and high-nitrogen conditions: A study in synthetic media Aust. J. Grape Wine Res. 12, 1, 69–78.

Ponomarova, O., Gabrielli, N., et al., 2017. Yeast creates a niche for symbiotic lactic acid bacteria through nitrogen overflow Cell Syst. 5, 4, 345–357.e6.

Rizk, Z., El Rayess, Y., et al., 2016. Impact of inhibitory peptides released by *Saccharomyces cerevisiae* BDX on the malolactic fermentation performed by *Oenococcus oeni* Vitilactic F Int. J. Food Microbiol. 233, 90–96.

Rizk, Z., Rayess, Y. El, et al., 2018. Identification of multiple-derived peptides produced by *Saccharomyces cerevisiae* involved in malolactic fermentation inhibition FEMS Yeast Res. 18, 7, 80.

Saguir, F.M., Campos, I.E.L., et al., 2009. Identification of dominant lactic acid bacteria isolated from grape juices. Assessment of its biochemical activities relevant to flavor development in wine Int. J. Wine Res. 1, 175–185.

Scott, S.R., Din, M.O., et al., 2017. A stabilized microbial ecosystem of self-limiting bacteria using synthetic quorum-regulated lysis Nat. Microbiol. 2, 17083, 1–22.

Shou, W., Ram, S., et al., 2007. Synthetic cooperation in engineered yeast populations Proc. Natl. Acad. Sci. 104, 6, 1877–1882.

Steensels, J., Gallone, B., et al., 2019. Domestication of industrial microbes Curr. Biol. 29, R381–R393.

Succi, M., Pannella, G., et al., 2017. Sub-optimal pH preadaptation improves the survival of *Lactobacillus plantarum* strains and the malic acid consumption in wine-like medium Front. Microbiol. | www.frontiersin.org 8, 1–12.

Tristezza, M., di Feo, L., et al., 2016. Simultaneous inoculation of yeasts and lactic acid bacteria: Effects on fermentation dynamics and chemical composition of Negroamaro wine LWT - Food Sci. Technol.

Ultee, A., Wacker, A., et al., 2013. Microbial succession in spontaneously fermented grape must before, during and after stuck fermentation South African J. Enol. Vitic. 34, 1, 68–78.

Versari, A., Patrizi, C., et al., 2016. Effect of co-inoculation with yeast and bacteria on chemical and sensory characteristics of commercial Cabernet Franc red wine from Switzerland J. Chem. Technol. Biotechnol. 91, 4, 876–882.

Vilela, A., 2019. Use of nonconventional yeasts for modulating wine acidity Fermentation 5, 1, 27.

Volschenk, H., Van Vuuren, H.J.J., et al., 2003. Malo-ethanolic fermentation in *Saccharomyces* and Schizo*saccharomyces* Curr. Genet. 43, 379–391.

Wells, A. & Osborne, J.P., 2011. Production of SO_2 and SO_2 binding compounds by *Saccharomyces cerevisiae* during the alcoholic fermentation and the impact on wine lactic acid bacteria S. Afr J Enol Vitic 32, 2, 267–279.

Yamasaki-Yashiki, S., Sawada, H., et al., 2017. Analysis of gene expression profiles of Lactobacillus paracasei induced by direct contact with *Saccharomyces cerevisiae* through recognition of yeast mannan Biosci. Microbiota, Food Heal. 36, 1, 17–25.

Zapparoli, G., Torriani, S., et al., 2003. Interactions between *Saccharomyces* and *Oenococcus oeni* strains from Amarone wine affect malolactic fermentation and wine composition Vitis 42, 2, 107–108.

Zapparoli, G., Tosi, E., et al., 2009. Bacterial inoculation strategies for the achievement of malolactic fermentation in high-alcohol wines South African J. Enol. Vitic. 30, 1, 49–55.

Zhou, N., Bottagisi, S., et al., 2017. Yeast-bacteria competition induced new metabolic traits through large-scale genomic rearrangements in *Lachancea kluyveri* FEMS Yeast Res. 17, 6, 1–14.

Zhou, N., Swamy, K.B.S., et al., 2017. Coevolution with bacteria drives the evolution of aerobic fermentation in *Lachancea kluyveri* PLoS One 12, 3, 1–19.

Chapter 2

Impact of microbial interactions on the evolution of microbes: examples from the wine environment

Chapter 2: Impact of microbial interactions on the evolution of microbes: examples from the wine environment

2.1. Abstract

Natural grape must is home to numerous microbial species transferred from grape surfaces and cellar environments. These microbes are present at the onset of fermentation. Fermenting grape must is highly selective due to its low pH, poor nutrient status, high sugar, variable temperature, sulphur dioxide, and low oxygen levels. Moreover, the dominant fermenting yeast, *Saccharomyces cerevisiae,* produces metabolites that are often inhibitory to other yeast and bacteria species. Data show that in spite of this harsh environment the lactic acid bacteria *Oenococcus oeni* and *Lactobacillus plantarum* persist, and may carry out malolactic fermentation (MLF) after alcoholic fermentation (AF) has ceased. Both species today are used as starter cultures for MLF, driving the degradation of malic acid to lactic acid, which is desirable for various reasons including sensory improvements and microbial stability. Microbes in wine, however, do not passively coexist, but interact through various ecological mechanisms broadly described as either positive or negative. These microbial interactions have intrigued many researchers, increasing the amount of research on this topic. To date this research has focused mainly on the biochemical impact of these microbes on each other, with only a few studies investigating the long-term impact of microbial interactions on the molecular adaptation of individual species to the wine environment. This review explores the mechanisms by which microbes adapt to new environments and cohabitants and how microbial interactions play a significant role in shaping adaptations. Additionally, we provide examples in literature where scientists have taken advantage of microbial interactions to establish ecosystems for the development of 'superior' strains using principles from synthetic ecology and directed evolution. This review therefore aims to highlight the impact/role that biotic selection pressure may have in selecting microbes that rapidly adapt to dynamic environments such as wine.

2.2. Introduction

Grape must is inhabited by numerous microbial species which are normally found on the grape and cellar surfaces, including fungi, acetic acid bacteria (AAB), lactic acid bacteria (LAB) and yeasts (König & Fröhlich, 2009). These species are also present during alcoholic fermentation (AF). However, fermenting grape must is highly selective: It is characterised by high sugar levels (≥200 mg/l) (Tilloy *et al.*, 2014), a low pH between 2.75 and 3.5 (Liu, Jia, *et al.*, 2015), high levels of sulphur dioxide, and a wide range of temperatures ranging from 12 and 20°C for white and rosé wines (López-Malo *et al.*, 2015; García-Ríos *et al.*, 2016) to 25-35°C for red wines (Ganucci *et al.*, 2018). Oxygen levels are also low, between 1 and 8 mg/L (du Toit *et al.*, 2006; Morales *et al.*, 2015). Additionally, yeast dominate the wine landscape with *Saccharomyces cerevisiae* being the

main grape must fermenter. Yeast-derived metabolites such as antimicrobial peptides, medium chain fatty acids, succinic acid and ethanol increase the selectivity of the wine medium, therefore creating a niche for some microbial species and not others (Capucho & San Romão, 1994; Renouf *et al.*, 2007; Branco *et al.*, 2014; Liu, Jia, *et al.*, 2015; Rizk *et al.*, 2016).

The most prevalent bacteria in grape must and wine are LAB and AAB belonging to the genera *Enterococcus*, *Lactococcus, Pediococcus, Oenococcus, Lactobacillus, Acetobacter* and *Gluconacetobacter* (Renouf *et al.*, 2007; García-Ruiz *et al.*, 2012; Piao *et al.*, 2015). The presence of LAB has not always been appreciated in the production of wine and was often seen as spoilage of the final product (Bartowsky *et al.*, 2003). However, since Pasteur and Muller discovered the added benefit of having LAB in wines, many winemakers now conduct malolactic fermentation (MLF) on their red wines, white wines such as Chardonnay, and some sparkling wines (Semon *et al.*, 2001; Bartowsky *et al.*, 2003, 2015). Despite four genera having species more than capable of conducting MLF, only *Lactobacillus plantarum* and *Oenococcus oeni* have been commercialised as starter cultures (du Toit *et al.*, 2011; Lerm *et al.*, 2011; Iorizzo *et al.*, 2016). These species are often beneficial in wine as they reduce the potential for spoilage by utilising resources that would otherwise be available to spoilage bacteria such as the appropriately named AAB, which produce undesirable compounds (see review by Bartowsky *et al.*, 2009). *O. oeni* is the primary LAB in winemaking as it can withstand the harsh chemical wine environment. Amongst other mechanisms, the degradation of L-malic acid to L-lactic acid, MLF (Lonvaud-Funel, 1999), is in part responsible for this ability.

The microbes found in grape must and wine do not passively coexist but interact with each other (Braga *et al.*, 2016). Several types of ecological mechanisms by which microbes interact have been described, and these can be broadly classified as either negative or positive. Specifically, negative interactions are observed when species compete with or inhibit one another (Yurdugü & Bozoglu, 2002; Osborne & Edwards, 2006; Zapparoli *et al.*, 2009; Ding *et al.*, 2013; Lleixà *et al.*, 2016; Rizk *et al.*, 2016) and positive interactions when they share a mutual benefit or stimulate one another (Ponomarova *et al.*, 2017; du Toit, 2018; Sieuwerts *et al.*, 2018). These interactions occur between different yeast species (Albergaria *et al.*, 2010; Ciani & Comitini, 2015; Bagheri *et al.*, 2017, 2018; Benito *et al.*, 2017; Morrison-Whittle *et al.*, 2018; Rollero *et al.*, 2018a), between yeast and bacteria (reviewed by Balmaseda *et al.*, 2018), and between different bacterial species (see review by Liu *et al.*, 2017) (Figure 2.1). Thus far, the majority of work on the interactions between wine-related microbes has focused mainly on biochemical interactions (reviewed by Alexandre *et al.* (2004), Liu *et al.* (2017), and Balmaseda *et al.* (2018)). However, the underlying interaction mechanisms, and how they shape evolutionary selection of microbes in the wine environment, are not well understood. This is understandable given the complexities involved in studying both

ecological and evolutionary dynamics involved in microbial interactions (Harrington & Sanchez, 2014).

Recently, the evolutionary responses of wine yeast to biotic stresses were reviewed for the first time (Conacher *et al.*, 2019). To add to the discussions established in that review, we aim to highlight the role of microbial interactions in shaping the adaptation of wine microbes to the wine environment. Studies have shown that microbes use different mechanisms to adapt to dynamic environments and to establish symbiotic relationships (Goddard, 2008; Branco *et al.*, 2014; Williams *et al.*, 2015; Ponomarova *et al.*, 2017; Yamasaki-Yashiki *et al.*, 2017; Zhou, Bottagisi, *et al.*, 2017; Bechtner *et al.*, 2019; Xu *et al.*, 2019). These mechanisms include the modification of gene expression levels and/or the genome by horizontal gene transfer (HGT), and recombination through point mutations. The sections to follow highlight how continuous exposure of microbes to the wine environment and to other species in this environment has resulted in the selection of genetic changes that confer fitness advantages to some microbes over others.

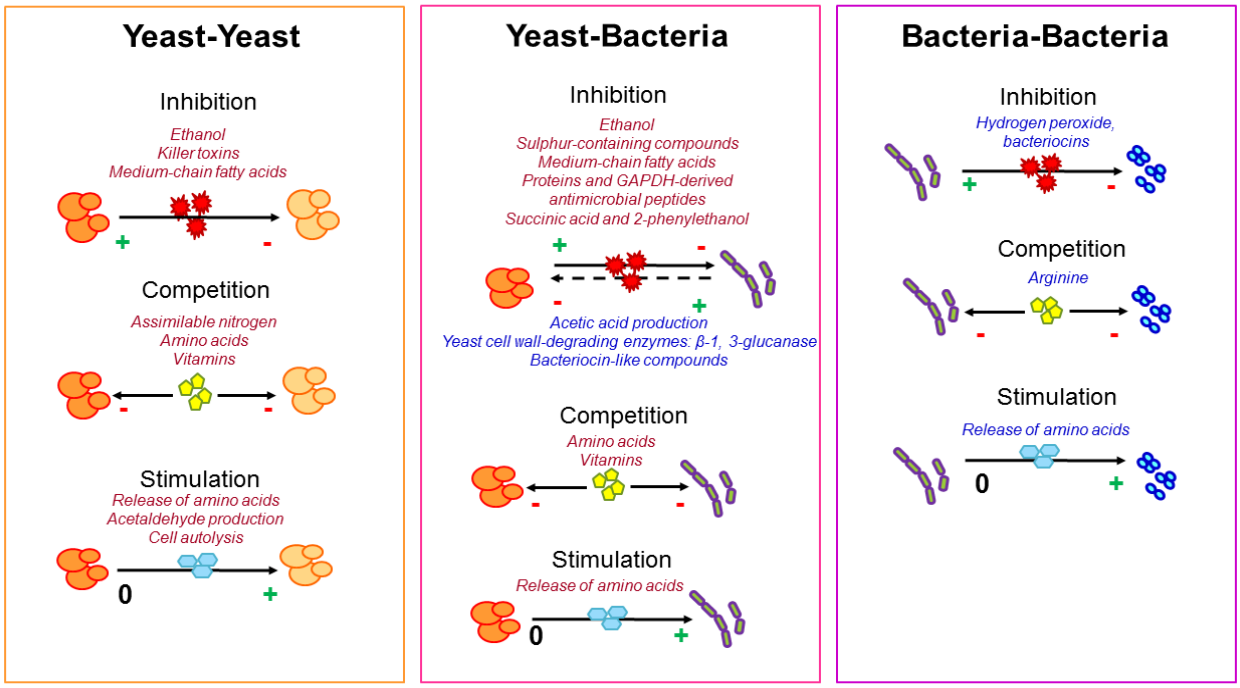

Figure 2.1 Summary of the types of microbial interactions occurring between wine microbes. Generally, these interactions are competitive, inhibitory or stimulatory. The specific metabolic compounds involved in these interactions are shown for yeast (red) and bacteria (blue).

2.3. Genetic mechanisms resulting in randomly selected adaptive genetic changes

Microbes are ubiquitous and are found in various challenging environments. To overcome the challenges (both abiotic and biotic) encountered microbes have to acquire adaptive mechanisms (see review by Bleuven & Landry, 2016). Mutations are a constant feature of all life and the driving

force of evolutionary adaptation (Hershberg, 2015). These mutations are random and most commonly neutral or deleterious. Genetic variation within a population is the basis for natural selection of individuals with increased fitness (Charlesworth et al., 2017). Neutral and deleterious mutations are more frequent than beneficial ones, but are usually removed from the population, in a process known as purifying selection, as they are naturally selected against by the prevailing conditions (Loewe & Hill, 2010). This means that the environment in which microbes find themselves selects for mutations that confer a fitness advantage, therefore increasing the beneficial mutation in the generations to come, granted these are inherited (Futuyma, 2009). It is the rare beneficial mutation(s) that drive adaptive evolution (Kirkpatrick & Peischl, 2012). There are several mechanisms by which the likelihood of microbes to adapt to challenging environments is increased. These mechanisms increase mutation frequency and genetic variation which increase the chances of encountering beneficial mutations: 1) point mutations, insertions and deletions, 2) genome reduction, 3) recombination and transposition, and 4) HGT and, although not a function of genetic variation, 5) modification of gene expression. Each of these are discussed in detail elsewhere (Hershberg, 2015; Bobay & Ochman, 2017; Husnik & McCutcheon, 2018; Rocha, 2018). HGT and gene expression modifications are the focus of this review. HGT is important here because this is one mechanism that explicitly shows the transfer of genes between species. In the case of wine yeast strains, it has been suggested that some HGTs have given *S. cerevisiae* a competitive advantage over other microbes specifically in the context of fermentation (Marsit *et al.*, 2016). Although, all these other mechanisms do provide some advantage to *S. cerevisiae* focusing on HGT provides more of an 'origins' story, and is of interest because it can be considered as part of the broader "interactions between species" topic, since HGT could be considered an ecosystem-based contribution to the adaptation of individual species to specific fermentation environments.

2.4. Competition between wine-related species

In grape must and wine, microbes are forced to share resources such as space, carbon sources, nitrogen-containing compounds, and vitamins. Species that are not efficient at utilising resources found in grape must and wine die out early on (Goddard, 2008; Rollero *et al.*, 2018a). This situation is often observed in the interactions between *Saccharomyces* and non-*Saccharomyces* yeasts during the early stages of fermentation where the former readily uses up the nutrients in the media, consequently producing ethanol that restricts the growth of some non-*Saccharomyces* species (Wang *et al.*, 2016; Rollero *et al.*, 2018a). *S. cerevisiae* are known to be Crabtree positive which means that this yeast prefers fermentation over cellular respiration in spite of the latter resulting in more units of energy (Dashko *et al.*, 2014). The Crabtree trait of *S. cerevisiae* was shown to be adaptive because it allows this species to have a competitive advantage over other yeast species. Specifically, *S. cerevisiae* isolates were shown to not only produce ethanol, which has detrimental effects on non-*Saccharomyces* yeast growth in wine, but also increase fermentation temperature,

increasing its fitness advantage and allowing it to dominate the fermentation environment in just 11 days (Goddard, 2008).

Additionally, some yeast strains have been shown to utilise L-malic acid (Benito *et al.*, 2017), which is an important carbon source for LAB, resulting in growth inhibition of LAB (Alexandre *et al.*, 2004). It becomes clear, therefore, that some microbes have a competitive advantage over others, for example, by evolving to be Crabtree positive as is the case in *S. cerevisiae* (Goddard, 2008); what is less clear is to what extent do interspecies interactions contribute to the evolution of microbial species in a population (or community). This section discusses HGT events in *S. cerevisiae* and wine LAB which have resulted in the dominance of these species in wine environments. Furthermore, the competitive interactions between different wine yeast and between yeast and bacteria is discussed.

2.4.1. Adaptation of *S. cerevisiae* by HGT to the wine environment

Data suggest that one of the mechanisms that has contributed to the evolution of wine strains of *S. cerevisiae* in particular is HGT. Indeed, it has been well established that some genes in *S. cerevisiae* have their origins in non-*Saccharomyces* yeast species and confer a competitive advantage for this yeast (Novo *et al.*, 2009; Borneman *et al.*, 2011; Marsit *et al.*, 2015). A decade ago Novo *et al.* (2009) demonstrated that a HGT event had occurred between non-*Saccharomyces* yeast and *S. cerevisiae* EC1118, a wine yeast strain. Particularly, they showed that EC1118 had 3 unique regions (A, B, and C) in its genome which encompassed genes that were essential for survival in the wine environment (Novo *et al.*, 2009). The genes of region B encode a C6 transcription factor, zinc-cluster transcription factor, nicotinic acid permease, a cell surface flocullin, and a 5-oxo-L-prolinase involved in the metabolism of carbon and nitrogen, stress responses and cellular transport (Novo *et al.*, 2009; Borneman *et al.*, 2011). Interestingly, genes in this region were not found in other *S. cerevisiae* strains (except RM11-1a) by a BLASTp analysis, but orthologs of these genes were observed in several non-*Saccharomyces* yeasts: *Lachancea kluyveri*, *Kluyveromyces thermotolerans* (*Lachancea thermotolerans*), *Candida guilliermondii*, *Pichia sorbitophila*, and *Zygosaccharomyces rouxii* (Novo *et al.*, 2009). PCR amplification of genes from region B, identified *Zygosaccharomyces bailii* CBS 680^{T} as the donor of these genes to EC1118, as both the genes and their organisation was conserved between the two strains (Novo *et al.*, 2009). The transfer of genes of region B from *Z. bailii* to *S. cerevisiae* EC1118 may confer a competitive advantage in low nitrogen, high sugar grape must and significantly contribute to the adaptation of this strain to fermenting must.

Recently, Marsit *et al.* (2015) investigated the origins of the unique region C of EC1118 found previously (Novo *et al.*, 2009). This region was suggested to have been donated by a yet unknown

yeast species to EC1118 (Novo *et al.*, 2009). Region C (65 kb) is found on the subtelomeric region of chromosome XV, with 19 genes, including *FSY1* (encoding a fructose symporter; (Galeote *et al.*, 2010)) and *FOT1/2* (encoding oligopeptide transporters; (Marsit *et al.*, 2016)). *FSY1* was highly expressed in high ethanol media conferring a competitive advantage in *S. cerevisiae* by increasing the efficiency by which this species transported residual fructose, that it could metabolise further (Galeote *et al.*, 2010), while the *FOT1/2* genes are essential in the uptake of oligopeptides in grape must (Marsit *et al.*, 2015). It was shown that *Torulaspora microellipsoides* is the source of region C found in EC1118 and other *S. cerevisiae* strains (Marsit *et al.*, 2015). However, this species is not a typical wine-related yeast, therefore, this genetic exchange event is speculative at this point with the working hypothesis being that a large (158 kb) genomic region was transferred from *T. microellipsoides* to *S. cerevisiae*, and then spread among various wine strains by the action of outcrossing (Marsit *et al.*, 2015). The yeast *T. microellipsoides* was previously isolated from fruit juices such as apple and in various other beverages (Kurtzman, 1998), however it is sporadic in wine environments which could suggest that it was introduced in these environments by human intervention.

2.4.2. Dominance of *Oenococcus oeni* and *Lactobacillus plantarum* in wine environments

O. oeni is the most widely used LAB species in the production of wine due to its high tolerance to acidic pH levels (<3.5) and high ethanol levels (9-16% v/v) and its low potential to produce off-flavours (Knoll *et al.*, 2011; Succi *et al.*, 2017; Sumby *et al.*, 2019). *Lb. plantarum* on the other hand is more sensitive to ethanol and should therefore be inoculated early during AF (van Bokhorst-van de Veen *et al.*, 2011). Some studies however show that ethanol tolerance of *Lb. plantarum* is strain specific and that there are high ethanol-tolerant strains (G-Alegría *et al.*, 2004; Berbegal *et al.*, 2016; Succi *et al.*, 2017). The adaptation of *O. oeni* to wine environments has been extensively studied (Maitre *et al.*, 2014; Bastard *et al.*, 2016; Dimopoulou *et al.*, 2016; Betteridge *et al.*, 2018; Jiang *et al.*, 2018; Collombel *et al.*, 2019), and molecular mechanisms which confer its robustness to fermenting wine have been presented and will be discussed briefly below (Bon *et al.*, 2009; Maitre *et al.*, 2014; Bastard *et al.*, 2016). However, only a few studies have attempted to link the adaptation of this species to biotic selective pressures such as other microbes (Favier *et al.*, 2012).

On a molecular level, plasmids and other mobile elements play a pivotal role in the evolution of LAB. Favier *et al.* (2012) discovered two plasmids in industrial *O. oeni* strains which carry genes (*tauE* and *oye*) relevant to survival in the wine environment: pOENI-1 (18.3-kb) and pOENI-1v2 (21.9-kb). These genes encode a sulphite exporter and a NADH:flavin oxidoreductase, respectively (Weinitschke *et al.*, 2007; Khairy *et al.*, 2016). The *tauE* gene is involved in the transport of sulphur-containing compounds such as sulphur-dioxide (SO_2) (Weinitschke *et al.*, 2007). In wine

SO_2 is often added to suppress the growth of potential spoilage microbes, however, this compound is also produced by fermenting yeast strains in the order of <30 mg/l, however, other strains can produce sulphur dioxide in excess of 100 mg/l (Eschenbruch, 1974). Free SO_2 at concentrations of at least 15 mg/l were shown to inhibit bacterial populations (Wells and Osborne, 2011). Moreover, yeast strains that inhibit malolactic activity were shown to produce sulphur-containing oligopeptides which may be the reason for the reduced MLF capability of LAB (Liu *et al.*, 2016). Therefore, it can be suggested that *O. oeni*, which are often in close proximity with such yeast, have acquired the *tauE* gene as a way to increase tolerance to sulphur-containing compounds, although numerous other genes may be involved (Favier *et al.*, 2012). In addition, the NADH:flavin oxidoreductase was shown to play a role in multiple biological functions such as oxidative stress response in *Bacillus subtilis* (Fitzpatrick *et al.*, 2003), but is poorly characterised in LAB (Valladares *et al.*, 2015). The presence of these genes may be important for the fitness of *O. oeni* during fermentation (Favier *et al.*, 2012), which gives this species a competitive advantage over other LAB in this environment.

On a genomic level, at least seven regions in the genome of *O. oeni* are thought to have been acquired by HGT (Margalef-Català, Felis, *et al.*, 2017). Only 1 of the 7 detected regions, region 3, has been shown to confer a fitness advantage to *O. oeni.* This region displays genetic markers for HGT from *Lactobacillus* species (Margalef-Català, Felis, *et al.*, 2017), including *Lb. plantarum* which is also found in close contact with *O. oeni* during grape must fermentation (G-Alegría *et al.*, 2004). The markers for region 3 include thioredoxin (*trx*), copper chaperone, and Crp/Fnr-like regulator (Margalef-Català, Felis, *et al.*, 2017). These genetic markers are involved in oxidative stress tolerance, temperature responses and regulation of molecular processes such as transcription (Serata *et al.*, 2012; Margalef-Català, Stefanelli, *et al.*, 2017) during wine fermentation and were shown to contribute to *O. oeni* fitness, which results in this species having a competitive advantage even in the presence of yeast (Comitini *et al.*, 2005).

The competitive ability of *Lb. plantarum* has thus far not been studied in wine, however, this species is known to be highly competitive in other environments (Jiang *et al.*, 2016). For example, there are numerous studies on the competitive inhibitory nature of this species through the production of bacteriocins (Calasso *et al.*, 2013; Gutiérrez-Cortés *et al.*, 2018), which inhibit other microbes allowing *Lb. plantarum* to thrive in its environment. In wine, it has been shown that *Lb. plantarum* strains can persist until the end of fermentation, and can tolerate high sugar, high ethanol and SO_2 (Brizuela *et al.*, 2019). These data suggest that this microbe has developed adaptive strategies that have allowed it to survive in wine fermentations and to be selected as a starter culture. Previously it was shown that *Lb. plantarum* genomes consist of a region known as the Lifestyle Island (Molenaar *et al.*, 2005) which harbours genes encoding proteins of carbohydrate metabolism. This region was shown to high hypermutability which allows *Lb.*

plantarum to rapidly evolve and adapt to changing environments (Kleerebezem *et al.*, 2003; Klaenhammer *et al.*, 2005; Molenaar *et al.*, 2005; Evanovich *et al.*, 2019) and possibly to mixed community environments.

2.4.3. Differential gene expression reveals competition between species

As was discussed in the previous section, *S. cerevisiae, O. oeni* and *Lb. plantarum* are well adapted to the wine environment and establish their dominance in wine fermentations by having genes that allow them to have a strong competitive advantage over other species. It is important then, to evaluate the expression of these genes in the context of microbial interactions. Moreover, it is necessary to understand how each of these species affect (or are affected by) the expression of genes in other species.

2.4.3.1. Yeast-yeast competition

Nitrogen sources are one of the main points of competition between wine yeast strains, with metabolites such as ammonium, γ-aminobutyric acid (GABA), arginine, and allantoin being among the readily consumed metabolites by yeast (Curiel *et al.*, 2017; Rollero *et al.*, 2018a). Curiel *et al.* (2017) showed that 3 hours after the co-inoculation of *S. cerevisiae* and *Torulaspora delbrueckii* allantoin (non-preferred nitrogen source) catabolism genes were upregulated which suggests the ability of *T. delbrueckii* to compete for nitrogen, resulting in *S. cerevisiae* requiring alternative sources of nitrogen. This finding is supported by the observations that allantoin catabolism genes were upregulated after 48 hours in single culture fermentations of *S. cerevisiae* when nitrogen sources were limited (Barbosa *et al.* 2015). Interestingly, when *S. cerevisiae* was co-cultured with other non-*Saccharomyces* yeast such as *Candida sake* or *Hanseniaspora uvaram* under aerobic conditions, allantoin catabolism genes were expressed, but not significantly (Curiel *et al.*, 2017). These data suggest that nitrogen resource usage is species-dependent as the intensity of gene expression of allantoin catabolism genes in *S. cerevisiae* differed in mixed cultures with different non-*Saccharomyces* yeast (Curiel *et al.*, 2017). In contrast, under anaerobic conditions *Hanseniaspora guilliermondii* did not show significant expression of these genes, but instead amino acid biosynthesis genes were overexpressed, indicating that in the absence of oxygen as a final electron acceptor, allantoin is not preferred (Cooper, 1984; Barbosa *et al.*, 2015; Tesnière *et al.*, 2019). These data show that gene expression of nitrogen catabolism is species (and strain) specific and is dependent on the conditions of the environment (Curiel *et al.*, 2017; Rollero *et al.*, 2018a; Tesnière *et al.*, 2019).

The wine environment is rich in fructose and glucose which are readily converted to ethanol, CO_2, and glycerol during alcoholic fermentation (AF) by *S. cerevisiae* (Walker & Stewart, 2016), a mechanism (among others) used to withstand osmotic stress (Babazadeh *et al.*, 2017) and used to

outcompete other microbes. The latter was observed between *S. cerevisiae* and *T. delbrueckii* (Tronchoni *et al.*, 2017). Tronchoni *et al.* (2017) discovered that genes involved in glycolysis, glucose uptake and the TCA cycle (*HXK1*, *GLK1*, *TDH1* to *TDH3*, *PGK1*, *GPM1*, *ENO1* and *ENO2*, *PDC6*, *PYK2*, *ADH3* and *ADH5*, and *ALD4*) were upregulated within the first 2 hours of co-cultivation with *T. delbrueckii* in *S. cerevisiae*, suggesting competitive interactions between the two yeast species for sugars. The rapid utilisation of the sugars shows that *S. cerevisiae* was the dominant yeast species, but also shows that *T. delbrueckii* directly influenced the biochemical responses in *S. cerevisiae,* emphasising the impact of biotic stresses

Other metabolites that are important in yeast-yeast competition, are vitamins. Thiamine is important for thiamine pyrophosphate biosynthesis, an essential cofactor for the conversion of pyruvate to acetaldehyde by pyruvate decarboxylase during fermentation (Hucker *et al.*, 2016). This step is important to restore redox imbalance incurred during glycolysis, which inadvertently impacts growth (Shi *et al.*, 2016). Therefore, the competition for thiamine will disadvantage at least one of the competing partners, as was observed between *S. cerevisiae* and *H. guilliermondii* (Barbosa *et al.*, 2015). In this study they observed the overexpression of *PHO*3 (encoding a constitutive acid phosphatase which is important for thiamine uptake (Sambuk *et al.*, 2011)), *THI20* (Trifunctional enzyme of thiamine biosynthesis) and *THI21* (Hydroxymethylpyrimidine kinase) after 24-48 hours, at which point glycerol production was increased. It was shown that in musts co-fermented with *S. cerevisiae* and non-*Saccharomyces* yeast glycerol production increases, when the must is deficient in thiamine (Bataillon *et al.*, 1996; Barbosa *et al.*, 2015; González-Royo *et al.*, 2015). The production of glycerol is important for non-*Saccharomyces* yeast for protection against ethanol and other stresses during AF (Tofalo *et al.*, 2016; Tronchoni *et al.*, 2017).

Furthermore, competition for biotin (vitamin H) was also observed between *S. cerevisiae* and *H. guilliermondii* (Barbosa *et al.*, 2015). After 48 and 96 hours of mixed fermentations, gene overexpression of *BIO3* (7, 8-diamino-pelargonic acid aminotransferase) and *BIO5* (putative transmembrane protein), which are involved in the biosynthesis of biotin, was observed for *S. cerevisiae*. This, like in the case of thiamine, indicates the importance of this vitamin in the competitive interactions between *S. cerevisiae* and non-*Saccharomyces* yeast (Barbosa *et al.*, 2015) and for yeast growth and fermentation capabilities (Duc *et al.*, 2017). Interestingly, the biotin biosynthesis genes found in *S. cerevisiae* are suggested to have been acquired by HGT from bacteria, based on comparative genome analysis (Hall *et al.*, 2005).

Recently it was shown that *S. cerevisiae* and *L. thermotolerans* compete for trace metals during wine fermentation (Shekhawat *et al.*, 2019). Iron and copper are essential transition metals necessary for cell survival and functioning, granted that in excess these metals are toxic to the cell

(Jo *et al.*, 2008). Iron and copper are important in many biosynthetic pathways including the TCA cycle because of their role as electron donors and acceptors and as cofactors in redox reactions (De Freitas *et al.*, 2003). Shekhawat *et al.* (2019) observed the upregulation of copper and iron uptake genes in *S. cerevisiae* in co-culture with *L. thermotolerans*. Among the upregulated genes was *FRE1, FRE2, ENB1, FIT2, CTR1* and *CTR3* which were shown earlier to be upregulated in iron- and copper-deficient environments (Shakoury-Elizeh *et al.*, 2010; Schlecht *et al.*, 2014). In *L. thermotolerans*, however, genes involved in iron transport and uptake were downregulated, suggesting that this yeast responds to iron-deficient environments by repressing iron catabolism or transport genes and utilising alternative mechanisms to survive, which is observed in other yeast species (Philpott *et al.*, 2012). The concentration of copper in the synthetic grape media used by Shekhawat *et a*l. (2019) was 0.11 µM. Copper depletion induces iron uptake in yeast, therefore at low concentrations iron metabolism genes are upregulated (Philpott & Protchenko, 2008). In natural grape must, it was found that copper and iron levels are above 0.07 mg/l and 2.93 mg/l, respectively (Kostić *et al.*, 2010). High concentrations of copper during wine fermentation were shown to inhibit growth and result in stuck fermentations (Sun *et al.*, 2019). However, copper-reducing starter cultures can be applied to wines high in copper (Capece *et al.*, 2018).

2.4.3.2. Yeast-bacteria competition

Yeast-bacteria competition is not well characterised in grape must fermentation, however, yeast and bacteria do (to some degree) utilise the same nutrients (Balmaseda *et al.*, 2018). As evidenced in the previous section, *S. cerevisiae* prefers taking up nitrogen sources (amino acids) from its environment as opposed to synthesising them (Curiel *et al.*, 2017; Rollero *et al.*, 2018a) which might result in less amino acids being available for the LAB partner. It is well known that LAB are fastidious and depend on the nutritional status of their environments for growth and survival (Terrade & Mira de Orduña, 2009). Therefore, depletion of nutrients during or after AF of yeast will have an impact on the growth of LAB (Costello *et al.*, 2003; Ivey *et al.*, 2013).

Recently it was shown that LAB compete for malic acid when in mixed communities with *Starmerella bacillaris, Issatchenkia orientalis, T. delbrueckii, Lachancea thermotolerans*, *Metschnikowia pulcherrima* and *Schizosaccharomyces pombe* (Seo *et al.*, 2007; du Plessis *et al.*, 2017; Janse van Rensburg, 2018). In another study, it was suggested that LAB use trace nutrients resulting in the inhibition of *S. cerevisiae* (Bayrock & Ingledew, 2004). The evolution of this competitive advantage between non-*Saccharomyces* yeast and LAB, or LAB to yeast has not yet been characterised. It is, however, interesting that most *S. cerevisiae* strains have poor malic acid degradation because the malic enzyme in most of these strains is located in the mitochondria and is not active under fermentative conditions (Volschenk *et al.*, 2003), but malic acid metabolism is pronounced in non-*Saccharomyces* yeast due to malate transporters and the presence of the malic

enzyme in the cytosol (Volschenk *et al.*, 2003; du Plessis *et al.*, 2017; Janse van Rensburg, 2018). This raises the question of whether wine-related *S. cerevisiae* strains lost the ability to actively degrade malic acid because of the domestication under fermentative conditions for thousands of years.

2.5. Inhibition between wine-related microbes

There are extensive reviews that detail the inhibitory metabolites that are produced by wine-related microbes in mixed community fermentations (Ciani *et al.*, 2016; Liu *et al.*, 2017; Balmaseda *et al.*, 2018). Thus far, only a few studies have evaluated the long-term impact of these metabolites on the evolution and adaptation of the affected species and strains to wine environments. In fact, most studies have investigated these metabolites in single culture experiments of sensitive strains (Miao *et al.*, 2018), excluding the potential impact of physical interactions. The transcriptomic responses to inhibitory metabolites in wine between interacting species have been previously investigated (Rossouw *et al.*, 2012; Maitre *et al.*, 2014; Margalef-Català *et al.*, 2016; Margalef-Català, Felis, *et al.*, 2017; Betteridge *et al.*, 2018; Jiang *et al.*, 2018; Miao *et al.*, 2018). The production of inhibitory metabolites is however not the only means by which inhibition occurs between wine-related species, as physical contact between species plays a significant role in this regard. In the following section the impact of cell-cell contact is discussed.

2.5.1. Inhibition between yeast species

Recently it was shown that in cell-cell interactions between *S. cerevisiae* and *L. thermotolerans*, the growth of the latter was significantly affected (Rossouw *et al.*, 2018; Petitgonnet *et al.*, 2019). Petitgonnet *et al.* (2019) attributed this negative interaction to the rapid consumption of oxygen and phytosterols by *L. thermotolerans* before the inoculation of *S. cerevisiae*, which resulted in the inability of the latter to produce unsaturated fatty acids and ergosterol (important for membrane biosynthesis and anaerobiosis (Duc *et al.*, 2017; Miao *et al.*, 2018)). Indeed Rossouw *et al.* (2018) observed similar results with *FLO5* and *FLO11* overexpressing *S. cerevisiae* strains in co-culture with *L. thermotolerans*. Moreover, it was observed that flocculation with *Wickerhamomyces anomalus* is detrimental to the growth of either *Hanseniaspora opuntiae* or the *FLO5* overexpressing *S. cerevisiae*, concluding that these genes play a significant role in the evolution of negative interactions between different yeast strains (Rossouw *et al.*, 2018).

Conversely, Shekhawat *et al.* (2019) observed the repression of *FLO10* and *FLO11* in *S. cerevisiae* and the upregulation of *FLO9* and *HSP12* in *L. thermotolerans* when in mixed fermentations, with no impact on growth. Additionally, to flocculation genes, the upregulation of *PAU* (seripauperin protein) genes was also observed. These proteins have been shown to play a

role in protecting cells in mixed species environments, mainly against *S. cerevisiae* killer toxins (Rivero *et al.*, 2015). Similarly, Tronchoni *et al.* (2017) observed the higher expression levels of *PAU* genes in *S. cerevisiae* cells co-cultured with *T. delbrueckii* during the early stage of fermentation. Taken together, these data emphasise that *FLO* genes and *PAU* genes are important cell wall role players in contact-based yeast-yeast interactions.

In an earlier study it was shown that numerically dominant viable *S. cerevisiae* cells are responsible for the growth inhibition of *L. thermotolerans* and *T. delbrueckii* (Nissen *et al.*, 2003). More specifically, when the two non-*Saccharomyces* yeast were grown in monoculture no growth limitations were observed, not even in the spent media of *S. cerevisiae*. However, as soon as *S. cerevisiae* was added the non-*Saccharomyces* yeast cells stopped growing. These data show that cell-cell contact between wine-related yeast species results in negative interactions under certain circumstances, independent of simple metabolic exchange. Although, the molecular mechanisms that drive these responses in non-*Saccharomyces* yeast were not evaluated by Nissen *et al.* (2003), it can be proposed that cell adhesion genes such as the *FLO* gene family played a role in the observed interactions as observed by Rossouw *et al.* (2018).

2.5.2. Cell-cell contact between yeast and LAB

Thus far no data on the impact of physical interactions between wine-related yeast and LAB have been published. However, several studies have investigated cell-cell interactions in yeast and LAB isolated from fukuyama pot vinegar (Furukawa *et al.*, 2010), cheese (Budinich *et al.*, 2011), human saliva (Kleerebezem *et al.*, 2003), and kefir grains (Garrote *et al.*, 2001). These data showed that yeast and LAB form aggregates mediated by yeast mannoproteins and bacterial DnaK (Furukawa *et al.*, 2011; Yamasaki-Yashiki *et al.*, 2017). Briefly, the heat shock protein DnaK recognises yeast mannoproteins, resulting in the adhesion of yeast and LAB (Katakura *et al.*, 2010). This mechanism of co-aggregation was shown to be pH, yeast and LAB strain-dependent (Pretzer *et al.*, 2005; Golowczyc *et al.*, 2009; Furukawa *et al.*, 2011). In these studies, the ecological impact of the co-aggregation on the yeast and bacteria were not reported, and warrant further study.

2.6. Stimulatory interactions in wine

2.6.1. Establishment of mutually stimulatory relationships between *S. cerevisiae* and LAB

For decades it was assumed that the rapid increase in LAB CFU/ml after AF was due to yeast cell death, resulting in the release of nutrients essential to LAB growth (Alexandre *et al.*, 2004). However, it has been shown that yeast release nutrients (amino acids) to the environment while alive (Ponomarova *et al.*, 2017; Bechtner *et al.*, 2019). The prior hypothesis should not be completely disregarded, because cell death of the yeast results in the release of nutrients such as

cell wall mannoproteins that LAB can use for growth (Ganan *et al.*, 2012; Dierings *et al.*, 2013), however, this would not be considered a direct interactive response between wine yeast and bacteria.

Recently, it was shown that in nitrogen-rich media *S. cerevisiae* actively releases amino acids, which coincidentally, were shown to be essential to *Lb. plantarum* and *Lactococcus lactis* survival (Ponomarova *et al.*, 2017). In fact, *S. cerevisiae* removes excess nitrogen sources to reduce nitrogen toxicity inside the cells. The specific amino acids released include serine, threonine, tryptophan, phenylalanine, and glutamine. These amino acids are essential for *Lb. plantarum* (Lee *et al.*, 2014; Ponomarova *et al.*, 2017; Botma, 2018) and either stimulatory or essential for *Lc. lactis* (Ponomarova *et al.*, 2017). Similar results were observed where *S. cerevisiae* released glutamine, glutamate, arginine, histidine, tryptophan, methionine and proline to the water kefir environment, thus supporting the growth of *Lb. hordei* (Bechtner *et al.*, 2019; Xu *et al.*, 2019). The usage of arginine by *Lb. hordei* resulted in the production of ammonium and an increase in pH which benefited yeast growth as well (Xu *et al.*, 2019).

Interestingly, it was shown that a mutual relationship is established between *Lc. lactis* and *S. cerevisiae* when lactose is the only carbon source because the yeast cannot metabolise this sugar (Ponomarova *et al.*, 2017). The study by Ponomarova *et al.* (2017) was the first to evidence the active release of amino acids by yeast to facilitate stimulatory interactions between yeast and LAB. Furthermore, the data show that the *TORC1* (target of rapamycin complex 1) pathway was induced (Kessi-Pérez *et al.*, 2019) and *NCR* (nitrogen catabolite repression) genes repressed in *S. cerevisiae* under nitrogen-rich conditions (Loewith, 2010; Curiel *et al.*, 2017), but when gene knockouts of the NCR genes were co-cultured with LAB, the stimulation interaction was abolished and competitive inhibition was observed. Similar trends were observed under nutrient-poor environments where NCR genes are expressed (Ponomarova *et al.*, 2017). Taken together, the data show the unpredictability of yeast-bacteria interactions in that these species can switch between stimulatory and inhibitory interactions as demanded by their environment.

Other examples of stimulation between *S. cerevisiae* and LAB were observed in co-cultivation experiments in water kefir and skimmed milk medium (Yamasaki-Yashiki *et al.*, 2017; Bechtner *et al.*, 2019; Xu *et al.*, 2019). There are currently no studies, to our knowledge, that have evaluated LAB genomic responses to *S. cerevisiae* in wine, yet the inverse has been explored (Rossouw *et al.*, 2012; Ponomarova *et al.*, 2017). In the works by Yamasaki-Yashiki *et al.* (2017), Bechtner *et al.* (2019), and Xu *et al.* (2019), genes involved in exopolysaccharide biosynthesis, citrate and amino acid metabolism, and carbohydrate metabolism were differentially expressed in *Lb. paracasei*, *Lb. nagelii*, and *Lb. hordei*, respectively. The expressed genes were shown to be involved in acid

stress tolerance in these studies (Yamasaki-Yashiki *et al.*, 2017; Bechtner *et al.*, 2019; Xu *et al.*, 2019). Although these studies were not performed in a wine environment, they demostrate the potential application of gene expression studies to investigate interactions between yeast and bacteria often found in close proximity in industrial food and beverage production.

2.7. Engineered microbial ecosystems to investigate evolutionary relationships between microbes

Microbial interactions in wine have almost exclusively been studied with a focus on physicochemical parameters as described above. Moreover, inhibitory interaction studies (with primary focus on *S. cerevisiae* metabolism) constitute most of the literature discussed in this review. Recently the impact of biotic stress between yeast species was reviewed, with the authors concluding that very limited data exist regarding specific adaptions to their ecological partners (Conacher *et al.*, 2019). The same observation applies to all other wine-related microbes.

However, a limited number of studies have considered microbial interactions within the ecosystem-based evolution and adaptation of microbial species. One possible approach to investigate such mechanisms that has rarely been used in the past is the application of the principles of synthetic ecology (Said & Or, 2017) and directed evolution (Cobb *et al.*, 2013) to engineer better adapted microbial ecosystems. The characterisation of the specific physiological and molecular adaptations of evolved strains provide the opportunity to indirectly and retrospectively link historical biotic selection to the genetic changes observed experimentally in the evolving population.

2.7.1. Synthetic ecology

Synthetic ecology is a relatively new field where simple ecological systems are engineered between at least two microbial species to create beneficial interactions including obligatory mutualisms (Dunham, 2007). The rationale behind the approach is that multiple species populations (consortia) can perform tasks that a single population would be unable to, or would normally take longer to perform. Also, microbial consortia tend to be more robust and resilient to environmental changes compared to single species populations (Zhang & Wang, 2016). The idea behind such ecosystem engineering is therefore to develop microbial ecosystems that can withstand environmental pressures while showing greater versatility and better performance (Said & Or, 2017).

2.7.2. Directed evolution (DE)

All organisms undergo random mutational gene sequence changes (translocations, indels, SNPs, etc.) during cell division or when exposed to mutagenic physicochemical factors such as UV light.

Due to the selective pressures imposed by their environment, including the other organisms with which they interact (Patel and Loeb, 2000; Bleuven and Landry, 2016), mutations that may improve the fitness of an organism will be retained and, over many generations, become established within a population due to the growth advantage and higher doubling rate of the mutated isolate under the selective conditions (Douglas J. Futuyma, 2009). Therefore, DE is the deliberate exposure of a microbial population to a particular selective pressure(s) over time, followed by the selection of a population or individual strains within this population with a 'fitness' advantage.

DE, experimental evolution (EE) or adaptive laboratory evolution (ALE) are all terms used to describe a technique borrowed from Darwin's theory of evolution by natural selection, which states that variation (phenotypic or genotypic) which confers a benefit to the population will be preserved in the progeny (Darwin, 1859; Bleuven & Landry, 2016). Recently other theories have been proposed which explain mechanisms involved in the adaptation of microorganisms to changing environments (Brooks *et al.*, 2011; Bleuven & Landry, 2016). To start off it has been hypothesised that many microorganisms use a bet-hedging strategy to adapt to fluctuating environments, where individuals in the population present different phenotypes specialised for different conditions (Brooks *et al.*, 2011). This strategy may be favoured by random genetic changes (genetic drift) highlighting the importance of an organism's genotype for adaptation to different environments (Agrawal & Whitlock, 2012; Bleuven & Landry, 2016). On the other hand, random genetic changes result in numerous deleterious mutations (mainly nonsynonymous) which have high fitness costs compared to neutral (synonymous) mutations (Rocha, 2018). However, purifying selection removes deleterious mutations out of the population to stop them spreading (Kuo *et al.*, 2009), especially in populations with large effective population sizes (N_e) (Charlesworth *et al.*, 2017). Moreover, a large N_e increases the likelihood of beneficial mutations fixing in the population (Ellegren, 2008).

In DE experiments the N_e often fluctuates because of bottlenecks incurred, for example, when cells are serially transferred between growth chambers (Van den Bergh *et al.*, 2018). After a bottleneck, genetic drift influences the evolution of microorganisms resulting in random mutations accumulating by chance (Lynch *et al.*, 2016), therefore increasing diversity and the likelihood of beneficial mutations accumulating. Selection of beneficial mutation, however, leads to a reduction in diversity and evolving populations converge (Van Cleve & Weissman, 2015; Wein & Dagan, 2019). This is often observed in rugged fitness landscapes consisting of a few fitness peaks representing fitness optima (Van den Bergh *et al.*, 2018). Rugged fitness landscapes are said to be common in experimental evolution studies (Van Cleve & Weissman, 2015) and result in increased diversity which may lead to individuals ending up on different fitness peaks i.e. display varying degrees of fitness. Alternatively, in fluctuating conditions such as those presented by DE

experiments, often populations evolving in parallel show high diversity on the genome level, but similarities at the phenotypic level (mostly due to selection), possibly reflecting the different paths or speeds taken to climb the same fitness peak as observed in similar growth conditions (Van den Bergh *et al.*, 2018). Application of DE represents what most likely occurs in natural environments, where microbes are faced with fluctuating environments which might result in many bottlenecks (Brooks *et al.*, 2011; Van den Bergh *et al.*, 2018). Nonetheless, microbes with adaptive changes are obtained (Van den Bergh *et al.*, 2018).

2.7.3. Examples for synthetic ecology systems

Recent work has employed both synthetic ecology and directed evolution using various techniques to establish specific microbial ecosystems which help increase our understanding of microbial interactions (Figure 2.2). There are two techniques that have been used to engineer evolving microbial ecosystems: 1) sacrificial sampling method, and 2) serial transfer method. Each of these are discussed below in the context of specific examples from literature.

2.7.3.1. Directed evolution using the sacrificial sampling method

Hom & Murray (2014) engineered a system between a strain of *S. cerevisiae* and *Chlamydomonas reinhardtii* with reciprocal exchange of carbon and nitrogen (Figure 2.2A). In this study, they found that the yeast strain, which was slow growing, completely depended on the algae, but *C. reinhardtii* disrupted the mutualism as soon as there was CO_2 from outside the system, and this alga outgrew the yeast (Hom & Murray, 2014). Similarly, Naidoo *et al.* (2019) showed that *S. cerevisiae* formed an obligate mutualism with *Chlorella sorokiniana* (isolated from winery wastewater) when mannose and nitrite were the only carbon and nitrogen sources for the yeast and alga, respectively (Figure 2.2A). The yeast breaks down mannose to CO_2 which the alga can use in photosynthesis, and the alga utilises nitrite to form ammonium which the yeast can use (Hom & Murray, 2014; Naidoo *et al.*, 2019). The obligate mutualism was disrupted when acetic acid was the sole carbon source in the medium (Naidoo *et al.*, 2019). Both these studies applied the sacrificial sampling technique where the co-cultures are inoculated simultaneously in several tubes and one was sacrificed for each sampling point (Hom & Murray, 2014; Naidoo *et al.*, 2019). This technique is limited because of the small volumes used, which limit the number of analyses that can be conducted on the samples (Naidoo *et al.*, 2019), and due to the high risk of human error during inoculation.

2.7.3.2. Serial transfer directed evolution experiments applied to wine-related species

It is noted that nutrient-exchange is at the forefront of engineered synthetic systems (Johns *et al.*, 2016), especially since many microbial species have a range of nutrient auxotrophies (e.g. vitamins, amino acids) (Zengler & Zaramela, 2018). Therefore, most synthetic ecology studies

focus on pairwise nutrient exchange as is seen in the studies by Hom & Murray (2014), Naidoo *et al.* (2019), and Shou *et al.* (2007). Shou *et al.* (2007) selected two yeast strains, each producing an amino acid required by the other (Figure 2.2B, 1a). In this work gene mutants of either lysine or adenine overproducing strains were generated by serially transferring a sample of the population to fresh media to establish an obligatory mutualism.

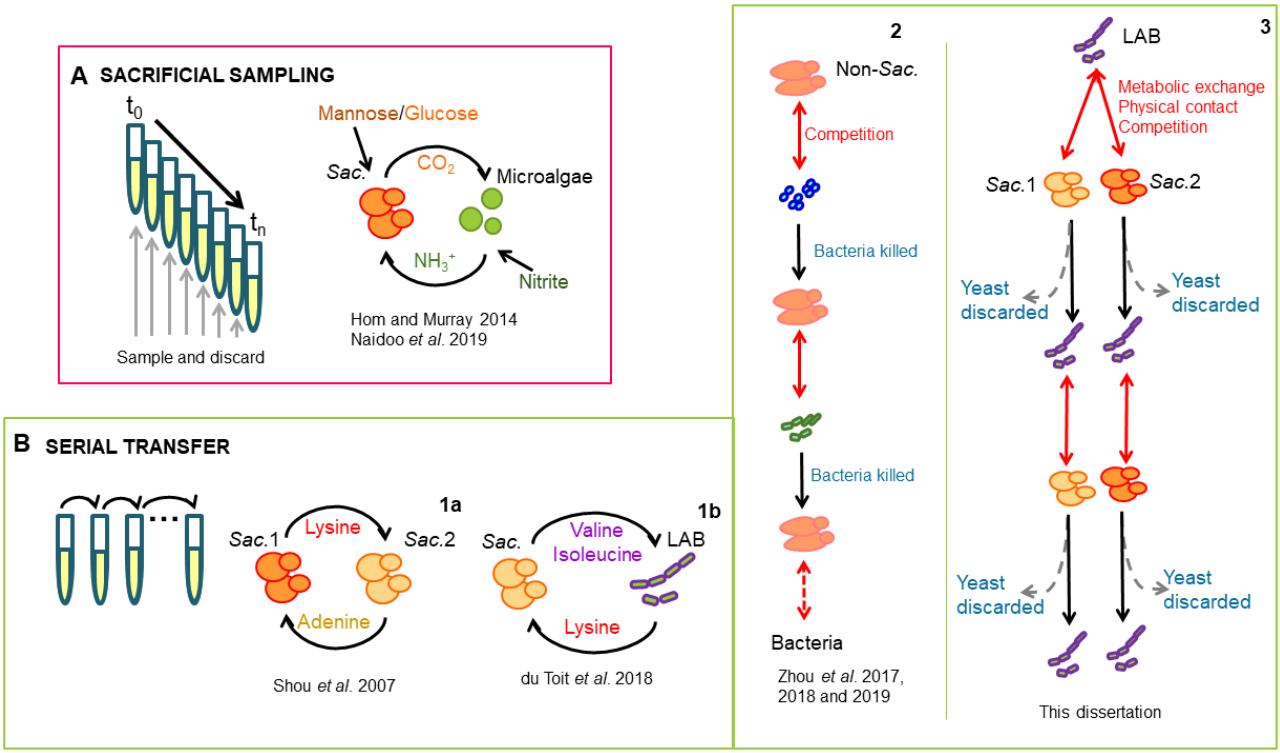

Figure 2.2 Summary of techniques used to study microbial interactions between wine-related species. A) Application of directed evolution using the sacrificial sampling method. A schematic showing the reciprocal nutrient exchange between yeast and microalgae (isolated from winery wastewater) is also shown. B) Application of directed evolution using the serial transfer method. 1a and 1b show the reciprocal exchange of amino acids. 2 and 3 show the application of a biotic selective driver in the evolution of either yeast or LAB, respectively.

Similarly, du Toit (2018) successfully applied this serial transfer technique to co-evolve strains of *S. cerevisiae* (lysine auxotroph) and *Lb. plantarum* (valine/isoleucine auxotroph) to be mutually dependent on each other (Figure 2.2B, 1b). The reciprocal exchange of nutrients was shown to be the driving force behind microbial interactions (du Toit, 2018). However, targeted nutrient exchange as seen in these studies limits the genomic responses of the interacting species, and thus is not a true reflection of what occurs in nature. Indeed, it was observed in the studies discussed above that the mutualisms that were established were soon abolished in the presence of an alternative nutrient source (Shou *et al.*, 2007; Hom & Murray, 2014; Naidoo *et al.*, 2019). Therefore, although more complex, it would be worthwhile to investigate microbial interactions in systems that mimic their natural environments as far as possible i.e. applications of consortia instead of pairwise comparisons. However, to study the evolution of mutualisms this system is still the most feasible.

Additionally, to have better control of population sizes and the effect of environmental fluctuations, the use of a continuous culture system would be more appropriate (Van den Bergh *et al.*, 2018).

Unlike the application of DE to engineer mutual relationships between un-related species, Morrison-Whittle *et al.*(2018) employed the serial transfer technique to establish new interactions between wine-related non-*Saccharomyces* yeast. *Candida glabrata* and *Pichia kudriavzevii*, evolved for ~65 generations, showed significantly reduced growth rates in co-culture compared to monoculture fermentations. Passive competition between the species resulted in the reduction of required resources to increase reproduction (Morrison-Whittle *et al.*, 2018). Often competition between species, as in the case of *S. cerevisiae* in winemaking, is viewed negatively because the populations of non-*Saccharomyces* yeast are reduced (Ciani *et al.*, 2016) and these yeasts have been shown to add desired aroma and flavour complexity to wine (Jolly *et al.*, 2014). However, the study by Morrison-Whittle *et al.* (2018) illustrates the use of co-evolution to establish competitive interactions between species, which resulted in increased aroma profiles of the tested wines. It must be kept in mind that often in the laboratory DE experiments are kept short and do not fully capture the long-term effects of these interactions, especially since microbial interactions are affected by environmental changes. Therefore, one should expect these microbial interactions to change over time and with changing environments (Ghoul & Mitri, 2016).

Recently, *L. kluyveri* was evaluated in co-culture experiments with various bacteria found in soil and on fruits (Zhou, Bottagisi, *et al.*, 2017; Zhou, Swamy, *et al.*, 2017; Zhou, Ishchuk, *et al.*, 2019). The data show the impact of employing a biotic selective pressure on the genome of *L. kluyveri* (Figure 2.2B, 2). Firstly they found that interactions with bacteria resulted in a Crabtree-like phenotype in *L. kluyveri* (Zhou, Swamy, *et al.*, 2017). Secondly, long-term competitive interactions between *L. kluyveri* and bacteria resulted in large-scale genomic rearrangements in the yeast, increasing its fermentative capabilities (Zhou, Bottagisi, *et al.*, 2017). It was shown that *S. cerevisiae* obtained its adaptation to fermenting environments through genome rearrangements (Zhou *et al.*, 2018) supporting previous reports (Novo *et al.*, 2009). Lastly, directed evolution of *L. kluyveri* under constant bacterial selective pressure yielded thermotolerant yeast strains (Zhou, Ishchuk, *et al.*, 2019). These data explicitly show the impact of biotic interactions and how they facilitate evolutionary changes within the target species. The directed evolution strategy employed (serially transferring evolving populations with bacteria as selective driver) is a stepping stone in microbial interaction research, which can be applied to numerous other species combinations in various laboratory environments and for various industrially relevant applications.

2.8. Conclusion and perspectives

Microbial interactions in various artificial and natural environments are a popular topic currently. This is not surprising given the advancements in technology for the investigation of microbial population dynamics. Next generation sequencing (NGS) is also a major player in assisting researchers to conduct comparative studies between microbial populations and species. Indeed, much research has focused on pairwise comparisons between interacting species, which is an important first step in mapping out the interactive landscape between species. The advent of synthetic ecology has opened a pathway for creativity in establishing systems that will allow the investigation of non-related species that would otherwise be difficult to evaluate. Moreover, the application of DE is only now beginning to gain traction to help researchers unravel the mechanisms involved in species adaptation to different environments and the mechanisms involved in interspecies interactions

HGT and large-scale genome rearrangements have played a significant role in establishing the dominance of *S. cerevisiae* and the two commercial wine LAB species (*O. oeni* and *Lb. plantarum*) in the wine fermentation ecosystem. Although not yet established between wine yeast and bacteria, data suggests that long-term co-evolution has been the driver for genomic diversity, at least between yeast species. Therefore, future work should explore in-depth the mechanisms that drive molecular evolution of microbes when in mixed microbial communities.

This review highlights that microbial interactions result in gene expression changes (Curiel *et al.*, 2017; Ponomarova *et al.*, 2017), and more importantly in genomic changes that confer competitive advantages (Novo *et al.*, 2009; Favier *et al.*, 2012; Marsit *et al.*, 2015). Therefore, in the chapters to follow I investigate the impact of biotic stress on the evolution of a LAB species, *Lb. plantarum*, by exposing it to yeast over numerous generations by DE (Figure 2.2B, 3). Therefore, I will demonstrate that in mixed microbial ecosystems, the evolution of a microbial population is not only influenced by its environment, but also by its interaction with other members in that environment. To simplify matters, I only explore a 'two-member' ecosystem between *S. cerevisiae* and *Lb. plantarum,* with a focus on the latter. The work that is presented in this dissertation is the first as far as I am aware to use *S. cerevisiae* as a selective driver to investigate the underlying molecular mechanisms that drive the molecular evolution of wine bacteria when interacting with yeast.

References

Agrawal, A.F. & Whitlock, M.C., 2012. Mutation load: The fitness of individuals in populations where deleterious alleles are abundant Annu. Rev. Ecol. Evol. Syst. 43, 1, 115–135.

Albergaria, H., Francisco, D., et al., 2010. *Saccharomyces cerevisiae* CCMI 885 secretes peptides that inhibit the growth of some non-*Saccharomyces* wine-related strains Appl. Microbiol. Biotechnol. 86, 3, 965–972.

Alexandre, H., Costello, P.J., et al., 2004. *Saccharomyces cerevisiae-Oenococcus oeni* interactions in wine: Current knowledge and perspectives Int. J. Food Microbiol. 93, 2, 141–154.

Alonso-del-Real, J., Lairón-Peris, M., et al., 2017. Effect of temperature on the prevalence of *Saccharomyces* non-*cerevisiae* species against a S. *cerevisiae* Wine strain in wine fermentation: Competition, physiological fitness, and influence in final wine composition Front. Microbiol. 8, 150.

Babazadeh, R., Lahtvee, P.J., et al., 2017. The yeast osmostress response is carbon source dependent Sci. Rep. 7, 990.

Bagheri, B., Bauer, F.F., et al., 2017. The impact of *Saccharomyces cerevisiae* on a wine yeast consortium in natural and inoculated fermentations Front. Microbiol. 8, 1988.

Bagheri, B., Zambelli, P., et al., 2018. Investigating the effect of selected non-*Saccharomyces* species on wine ecosystem function and major volatiles Front. Bioeng. Biotechnol. 6, 169.

Balmaseda, A., Bordons, A., et al., 2018. Non-*Saccharomyces* in wine: Effect upon *Oenococcus oeni* and malolactic fermentation Front. Microbiol. 9, 1–8.

Barbosa, C., Mendes-Faia, A., et al., 2015. Genomic expression program of *Saccharomyces cerevisiae* along a mixed-culture wine fermentation with *Hanseniaspora guilliermondii* Microb. Cell Fact. 14, 124.

Bartowsky, E.J., Costello, P.J., et al., 2009. Wine bacteria – friends and foes AWRI Rep. 24, 2, 18–20.

Bartowsky, E.J., Costello, P.J., et al., 2015. Emerging trends in the application of malolactic fermentation Aust. J. Grape Wine Res. 21, 663–669.

Bartowsky, E.J., Xia, D., et al., 2003. Spoilage of bottled red wine by acetic acid bacteria Lett. Appl. Microbiol. 36, 307–314.

Bastard, A., Coelho, C., et al., 2016. Effect of biofilm formation by *Oenococcus oeni* on malolactic fermentation and the release of aromatic compounds in wine Front. Microbiol. 7, 1–14.

Bataillon, M., Rico, A., et al., 1996. Early thiamine assimilation by yeasts under enological conditions: Impact on alcoholic fermentation kinetics J. Ferment. Bioeng. 82, 2, 145–150.

Bayrock, D.P. & Ingledew, W.M., 2004. Inhibition of yeast by lactic acid bacteria in continuous culture: Nutrient depletion and/or acid toxicity? J. Ind. Microbiol. Biotechnol. 31, 8, 362–368.

Bechtner, J., Xu, D., et al., 2019. Proteomic analysis of *Lactobacillus nagelii* in the presence of *Saccharomyces cerevisiae* isolated from water kefir and comparison with *Lactobacillus hordei* Front. Microbiol. 10, 325.

Benito, Á., Calderón, F., et al., 2017. The combined use of Schizo*saccharomyces* pombe and *Lachancea* thermotolerans - Effect on the anthocyanin wine composition Molecules 22, 5, 739.

Berbegal, C., Peña, N., et al., 2016. Technological properties of *Lactobacillus plantarum* strains isolated from grape must fermentation Food Microbiol. 57, 187–194.

Betteridge, A.L., Sumby, K.M., et al., 2018. Application of directed evolution to develop ethanol tolerant *Oenococcus oeni* for more efficient malolactic fermentation Appl. Microbiol. Biotechnol. 102, 2, 921–932.

Bleuven, C. & Landry, C.R., 2016. Molecular and cellular bases of adaptation to a changing environment in microorganisms Proc. R. Soc. B Biol. Sci. 283, 20161458.

Bobay, L.M. & Ochman, H., 2017. The evolution of bacterial genome architecture Front. Genet. 8, 72.

Bon, E., Delaherche, A., et al., 2009. *Oenococcus oeni* genome plasticity is associated with fitness Appl. Environ. Microbiol. 75, 7, 2079–2090.

Borneman, A.R., Desany, B.A., et al., 2011. Whole-genome comparison reveals novel genetic elements that characterize the genome of industrial strains of *Saccharomyces cerevisiae* PLoS Genet. 7, 2, e1001287.

Botma, I.J., 2018. *Lactobacillus plantarum*: amino acid utilization Stellenbosch University.

Braga, R.M., Dourado, M.N., et al., 2016. Microbial interactions: ecology in a molecular perspective Brazilian J. Microbiol. 47, Suppl 1, 86–98.

Branco, P., Francisco, D., et al., 2014. Identification of novel GAPDH-derived antimicrobial peptides secreted by *Saccharomyces cerevisiae* and involved in wine microbial interactions Appl. Microbiol. Biotechnol. 98, 2, 843–853.

Brizuela, N., Tymczyszyn, E.E., et al., 2019. *Lactobacillus plantarum* as a malolactic starter culture in winemaking: A new (old) player? Electron. J. Biotechnol. 38, 10–18.

Brooks, A.N., Turkarslan, S., et al., 2011. Adaptation of cells to new environments Wiley Interdiscip Rev Syst Biol Med 3, 5, 544–561.

Budinich, M.F., Perez-Díaz, I., et al., 2011. Growth of *Lactobacillus paracasei* ATCC 334 in a cheese model system: A biochemical approach J. Dairy Sci. 94, 11, 5263–5277.

Calasso, M., Cagno, R. Di, et al., 2013. Effects of the peptide pheromone plantaricin a and cocultivation with *Lactobacillus sanfranciscensis* DPPMA174 on the exoproteome and the adhesion capacity of *Lactobacillus plantarum* DC400 Appl. Environ. Microbiol. 79, 8, 2657–2669.

Capece, A., Romaniello, R., et al., 2018. Yeast starter as a biotechnological tool for reducing copper content in wine Front. Microbiol. 8, 2632.

Capucho, I. & San Romão, M. V., 1994. Effect of ethanol and fatty acids on malolactic activity of *Leuconostoc oenos* Appl. Microbiol. Biotechnol. 42, 2–3, 391–395.

Charlesworth, D., Barton, N.H., et al., 2017. The sources of adaptive variation Proc. R. Soc. B Biol. Sci. 284, 20162864.

Ciani, M. & Comitini, F., 2015. Yeast interactions in multi-starter wine fermentation Curr. Opin. Food Sci. 1, 1–6.

Ciani, M., Capece, A., et al., 2016. Yeast interactions in inoculated wine fermentation Front. Microbiol. 7, 1–7.

Cobb, R.E., Chao, R., et al., 2013. Directed evolution: Past, present, and future AIChE J. 59, 5, 1432–1440.

Collombel, I., Campos, F.M., et al., 2019. Changes in the composition of the lactic acid bacteria behavior and the diversity of *Oenococcus oeni* isolated from red wines supplemented with selected grape phenolic compounds Fermentation 5, 1.

Comitini, F., Ferretti, R., et al., 2005. Interactions between *Saccharomyces cerevisiae* and malolactic bacteria: Preliminary characterization of a yeast proteinaceous compound(s) active against *Oenococcus oeni* J. Appl. Microbiol. 99, 105–111.

Conacher, C., Rossouw, D., et al., 2019. Peer pressure: Evolutionary responses to biotic pressures in wine yeasts FEMS Yeast Res. foz072.

Cooper, T.G., 1984. Allantoin degradation by *Saccharomyces cerevisiae* - A model system for gene regulation and metabolic integration Adv. Enzymol. Relat. Areas Mol. Biol. 56, 91–136.

Costello, P.J., Henschke, P.A., et al., 2003. Standardised methodology for testing malolactic bacteria and wine yeast compatibility Aust. J. Grape Wine Res. 9, 2, 127–137.

Curiel, J.A., Morales, P., et al., 2017. Different non-*Saccharomyces* yeast species stimulate nutrient consumption in S. *cerevisiae* mixed cultures Front. Microbiol. 8, 2121.

Darwin, C., 1859. On the origin of species by means of natural selection, or preservation of favoured races in the struggle for life. London, United Kingdom.

Dashko, S., Zhou, N., et al., 2014. Why, when, and how did yeast evolve alcoholic fermentation? FEMS Yeast Res. 14, 6, 826–832.

De Freitas, J., Wintz, H., et al., 2003. Yeast, a model organism for iron and copper metabolism studies in: BioMetals. Vol. 16 185–197.

Dierings, L.R., Braga, C.M., et al., 2013. Population dynamics of mixed cultures of yeast and lactic acid bacteria in cider conditions Brazilian Arch. Biol. Technol. 56, 5, 837–847.

Dimopoulou, M., Bardeau, T., et al., 2016. Exopolysaccharides produced by *Oenococcus oeni*: From genomic and phenotypic analysis to technological valorization Food Microbiol. 53, 10–17.

Ding, J., Bierma, J., et al., 2013. Acetic acid inhibits nutrient uptake in *Saccharomyces cerevisiae*: Auxotrophy confounds the use of yeast deletion libraries for strain improvement Appl. Microbiol. Biotechnol. 97, 16, 7405–7416.

du Plessis, H.W., du Toit, M., et al., 2017. Characterisation of non-*Saccharomyces* yeasts using different methodologies and evaluation of their compatibility with malolactic fermentation South African J. Enol. Vitic. 38, 1, 46–63.

du Toit, M., Engelbrecht, L., et al., 2011. *Lactobacillus*: The next generation of malolactic fermentation starter cultures-an overview Food Bioprocess Technol. 4, 6, 876–906.

du Toit, S.C., 2018. Coevolution of *Saccharomyces cerevisiae* and *Lactobacillus plantarum*: Engineering interspecies cooperation Stellenbosch University.

du Toit, W.J., Marais, J., et al., 2006. Oxygen in Must and Wine: A review South African J. Enol. Vitic. 27, 1, 76–94.

Duc, C., Pradal, M., et al., 2017. A set of nutrient limitations trigger yeast cell death in a nitrogen-dependent manner during wine alcoholic fermentation PLoS One 12, 9, e0184838.

Dunham, M.J., 2007. Synthetic ecology: A model system for cooperation Proc. Natl. Acad. Sci. 104, 6, 1741–1742.

Ellegren, H., 2008. Comparative genomics and the study of evolution by natural selection Mol. Ecol. 17, 21, 4586–4596.

Evanovich, E., De Souza Mendonça Mattos, P.J., et al., 2019. Comparative genomic analysis of *Lactobacillus plantarum*: An overview Int. J. Genomics 2019, 1–11.

Favier, M., Bilhère, E., et al., 2012. Identification of pOENI-1 and related plasmids in *Oenococcus oeni* strains performing the malolactic fermentation in wine PLoS One 7, 11, e49082.

Fitzpatrick, T.B., Amrhein, N., et al., 2003. Characterization of YqjM, an old yellow enzyme homolog from Bacillus subtilis involved in the oxidative stress response J. Biol. Chem. 278, 22, 19891–19897.

Furukawa, S., Nojima, N., et al., 2011. The importance of inter-species cell-cell co-aggregation between *Lactobacillus plantarum* ML11-11 and *Saccharomyces cerevisiae* BY4741 in mixed-species biofilm formation Biosci. Biotechnol. Biochem. 75, 8, 1430–1434.

Furukawa, S., Yoshida, K., et al., 2010. Mixed-species biofilm formation by direct cell-cell contact between brewing yeasts and lactic acid bacteria Biosci. Biotechnol. Biochem. 74, 11, 2316–2319.

G-Alegría, E., López, I., et al., 2004. High tolerance of wild *Lactobacillus plantarum* and *Oenococcus oeni* strains to lyophilisation and stress environmental conditions of acid pH and ethanol FEMS Microbiol. Lett. 230, 53–61.

Galeote, V., Novo, M., et al., 2010. FSY1, a horizontally transferred gene in the *Saccharomyces cerevisiae* EC1118 wine yeast strain, encodes a high-affinity fructose/H+ symporter Microbiology 156, 3754–3761.

Ganan, M., Carrascosa, A. V., et al., 2012. Effect of mannoproteins on the growth, gastrointestinal viability, and adherence to Caco-2 cells of lactic acid bacteria J. Food Sci. 77, 3, 176–180.

Ganucci, D., Guerrini, S., et al., 2018. Quantifying the effects of ethanol and temperature on the fitness advantage of predominant *Saccharomyces cerevisiae* strains occurring in spontaneous wine fermentations Front. Microbiol. 9, 1563.

García-Ríos, E., Ramos-Alonso, L., et al., 2016. Correlation between low temperature adaptation and oxidative stress in *Saccharomyces cerevisiae* Front. Microbiol. 7, 1199.

García-Ruiz, A., Cueva, C., et al., 2012. Antimicrobial phenolic extracts able to inhibit lactic acid bacteria growth and wine malolactic fermentation Food Control 28, 2, 212–219.

Garrote, G.L., Abraham, A.G., et al., 2001. Chemical and microbiological characterisation of kefir grains J. Dairy Res. 68, 4, 639–652.

Ghoul, M. & Mitri, S., 2016.

Goddard, M.R., 2008. Quantifying the complexities of *Saccharomyces cerevisiae*'s ecosystem engineering via fermentation Ecology 89, 8, 2077–2082.

Golowczyc, M.A., Mobili, P., et al., 2009. Interaction between *Lactobacillus* kefir and *Saccharomyces* lipolytica isolated from kefir grains: Evidence for lectin-like activity of bacterial surface proteins J. Dairy Res. 76, 1, 111–116.

González-Royo, E., Pascual, O., et al., 2015. Oenological consequences of sequential inoculation with non-*Saccharomyces* yeasts (*Torulaspora delbrueckii* or *Metschnikowia pulcherrima*) and *Saccharomyces cerevisiae* in base wine for sparkling wine production Eur. Food Res. Technol. 240, 5, 999–1012.

Gutiérrez-Cortés, C., Suarez, H., et al., 2018. Enhanced bacteriocin production by *Pediococcus pentosaceus* 147 in co-culture with *Lactobacillus plantarum* LE27 on cheese whey broth Front. Microbiol. 9, 2952.

Hall, C., Brachat, S., et al., 2005. Contribution of horizontal gene transfer to the evolution of *Saccharomyces cerevisiae* Eukaryot. Cell 4, 6, 1102–1115.

Harrington, K.I. & Sanchez, A., 2014. Eco-evolutionary dynamics of complex social strategies in microbial communities Commun. Integr. Biol. 7, e28230.

Hershberg, R., 2015. Mutation—the engine of evolution: Studying mutation and its role in the evolution of bacteria Cold Spring Harb. Perspect. Biol. 7, 9, 1–13.

Hom, E.F.Y. & Murray, A.W., 2014. Niche engineering demonstrates a latent capacity for fungal-algal mutualism Science. 345, 6192, 94–98.

Hucker, B., Wakeling, L., et al., 2016. Vitamins in brewing: Presence and influence of thiamine and riboflavin on wort fermentation J. Inst. Brew. 122, 126–137.

Husnik, F. & McCutcheon, J.P., 2018. Functional horizontal gene transfer from bacteria to eukaryotes Nat. Rev. Microbiol. 16, 2, 67–79.

Iorizzo, M., Testa, B., et al., 2016. Selection and technological potential of *Lactobacillus plantarum* bacteria suitable for wine malolactic fermentation and grape aroma release LWT - Food Sci. Technol. 73, 557–566.

Ivey, M., Massel, M., et al., 2013. Microbial Interactions in Food Fermentations Annu. Rev. Food Sci. Technol. 4, 141–162.

Janse van Rensburg, P.J., 2018. Investigating the influence of a wine yeast consortium on population dynamics, alcoholic and malolactic fermentation Stellenbosch University.

Jiang, J., Sumby, K.M., et al., 2018. Directed evolution of *Oenococcus oeni* strains for more efficient malolactic fermentation in a multi-stressor wine environment Food Microbiol. 73, 150–159.

Jiang, M., Zhang, F., et al., 2016. Evaluation of probiotic properties of *Lactobacillus plantarum* WLPL04 isolated from human breast milk J. Dairy Sci. 99, 3, 1736–1746.

Jo, W.J., Loguinov, A., et al., 2008. Identification of genes involved in the toxic response of *Saccharomyces cerevisiae* against iron and copper overload by parallel analysis of deletion mutants Toxicol. Sci. 101, 1, 140–151.

Jolly, N.P., Varela, C., et al., 2014.

Katakura, Y., Sano, R., et al., 2010. Lactic acid bacteria display on the cell surface cytosolic proteins that recognize yeast mannan Appl. Microbiol. Biotechnol. 86, 1, 319–326.

Kessi-Pérez, E.I., Salinas, F., et al., 2019. Indirect monitoring of TORC1 signalling pathway reveals molecular diversity among different yeast strains Yeast 36, 65–74.

Khairy, H., Wübbeler, J.H., et al., 2016. The NADH:flavin oxidoreductase Nox from *Rhodococcus erythropolis* MI2 is the key enzyme of 4,4′-dithiodibutyric acid degradation Lett. Appl. Microbiol. 63, 6, 434–441.

Kirkpatrick, M. & Peischl, S., 2012. Evolutionary rescue by beneficial mutations in environments that change in space and time Philos. Trans. R. Soc. B Biol. Sci. 368, 20120082.

Klaenhammer, T.R., Barrangou, R., et al., 2005. Genomic features of lactic acid bacteria effecting bioprocessing and health FEMS Microbiol. Rev. 29, 393–409.

Kleerebezem, M., Boekhorst, J., et al., 2003. Complete genome sequence of *Lactobacillus plantarum* WCFS1 Proc. Natl. Acad. Sci. U. S. A. 100, 4, 1990–1995.

Knoll, C., Fritsch, S., et al., 2011. Influence of pH and ethanol on malolactic fermentation and volatile aroma compound composition in white wines LWT - Food Sci. Technol. 44, 10, 2077–2086.

Konig, H. & Frohlich, J., 2009. Lactic Acid Bacteria In: H. Konig, J. Frohlich, & G. Unden (eds). Biol. Microorg. Grapes, Must Wine. Springer Berlin Heidelberg 3–29.

Kostić, D., Mitić, S., et al., 2010. The concentrations of Fe, Cu and Zn in selected wines from South-East Serbia J. Serbian Chem. Soc. 75, 12, 1701–1709.

Kuo, C.H., Moran, N.A., et al., 2009. The consequences of genetic drift for bacterial genome complexity Genome Res. 19, 8, 1450–1454.

Kurtzman, C.P., 1998. *Torulaspora* Lindner in: C.P. Kurtzman & J.W. Fell (eds). Yeasts A Taxon. Study. (4th ed.). Elsevier 404–408.

Lee, K., Kim, H.-J., et al., 2014. Amino acids analysis during lactic acid fermentation by single strain cultures of lactobacilli and mixed culture starter made from them African J. Biotechnol. 13, 28, 2867–2873.

Lerm, E., Engelbrecht, L., et al., 2011. Selection and characterisation of *Oenococcus oeni* and *Lactobacillus plantarum* South African wine isolates for use as malolactic fermentation starter cultures South African J. Enol. Vitic. 32, 2, 280–295.

Liu, X., Jia, B., et al., 2015. Effect of initial pH on growth characteristics and fermentation properties of *Saccharomyces cerevisiae* J. Food Sci. 80, 4, M800–M808.

Liu, Y., Forcisi, S., et al., 2016. New molecular evidence of wine yeast-bacteria interaction unravelled by non-targeted exometabolomic profiling Metabolomics 12, 69, 1–16.

Liu, Y., Rousseaux, S., et al., 2017. Wine microbiome: A dynamic world of microbial interactions Crit. Rev. Food Sci. Nutr. 57, 4, 856–873.

Lleixà, J., Manzano, M., et al., 2016. *Saccharomyces* and non-*Saccharomyces* competition during microvinification under different sugar and nitrogen conditions Front. Microbiol. 7, 1959.

Loewith, R., 2010. TORC1 signalling in budding yeast Enzymes 27, 147–175.

Lonvaud-Funel, A., 1999. Lactic acid bacteria in the quality improvement and depreciation of wine Antonie van Leeuwenhoek, Int. J. Gen. Mol. Microbiol. 76, 317–331.

López-Malo, M., García-Rios, E., et al., 2015. Evolutionary engineering of a wine yeast strain revealed a key role of inositol and mannoprotein metabolism during low-temperature fermentation BMC Genomics 16, 537.

Lynch, M., Ackerman, M.S., et al., 2016. Genetic drift, selection and the evolution of the mutation rate Nat. Rev. Genet. 17, 704–714.

Maitre, M., Weidmann, S., et al., 2014. Adaptation of the wine bacterium *Oenococcus oeni* to ethanol stress: Role of the small heat shock protein lo18 in membrane integrity Appl. Environ. Microbiol. 80, 10, 2973–2980.

Margalef-Català, M., Araque, I., et al., 2016. Transcriptomic and proteomic analysis of *Oenococcus oeni* adaptation to wine stress conditions Front. Microbiol. 7, 1554.

Margalef-Català, M., Felis, G.E., et al., 2017. Identification of variable genomic regions related to stress response in *Oenococcus oeni* Food Res. Int. 102, 625–638.

Margalef-Català, M., Stefanelli, E., et al., 2017. Variability in gene content and expression of the thioredoxin system in *Oenococcus oeni* Food Microbiol. 61, 23–32.

Marsit, S., Mena, A., et al., 2015. Evolutionary advantage conferred by an eukaryote-to-eukaryote gene transfer event in wine yeasts Mol. Biol. Evol. 32, 7, 1695–1707.

Marsit, S., Sanchez, I., et al., 2016. Horizontally acquired oligopeptide transporters favour adaptation of *Saccharomyces cerevisiae* wine yeast to oenological environment Environ. Microbiol. 18, 4, 1148–1161.

Miao, Y., Xiong, G., et al., 2018. Transcriptome profiling of *Issatchenkia orientalis* under ethanol stress AMB Express 8, 39.

Molenaar, D., Bringel, F., et al., 2005. Exploring *Lactobacillus plantarum* genome diversity by using microarrays J. Bacteriol. 187, 17, 6119–6127.

Morales, P., Rojas, V., et al., 2015. The impact of oxygen on the final alcohol content of wine fermented by a mixed starter culture Appl. Microbiol. Biotechnol. 99, 9, 3993–4003.

Morrison-Whittle, P., Lee, S.A., et al., 2018. Co-evolution as tool for diversifying flavor and aroma profiles of wines Front. Microbiol. 9, 910.

Naidoo, R.K., Simpson, Z.F., et al., 2019. Nutrient exchange of carbon and nitrogen promotes the formation of stable mutualisms between *Chlorella sorokiniana* and *Saccharomyces cerevisiae* under engineered synthetic growth conditions Front. Microbiol. 10, 609.

Nissen, P., Nielsen, D., et al., 2003. Viable *Saccharomyces cerevisiae* cells at high concentrations cause early growth arrest of non-*Saccharomyces* yeasts in mixed cultures by a cell - Cell contact-mediated mechanism Yeast 20, 4, 331–341.

Novo, M., Bigey, F., et al., 2009. Eukaryote-to-eukaryote gene transfer events revealed by the genome sequence of the wine yeast *Saccharomyces cerevisiae* EC1118 Proc. Natl. Acad. Sci. U. S. A. 106, 38, 16333–16338.

Osborne, J.P. & Edwards, C.G., 2006. Inhibition of malolactic fermentation by *Saccharomyces* during alcoholic fermentation under low- and high-nitrogen conditions: A study in synthetic media Aust. J. Grape Wine Res. 12, 1, 69–78.

Petitgonnet, C., Klein, G.L., et al., 2019. Influence of cell-cell contact between *L. thermotolerans* and *S. cerevisiae* on yeast interactions and the exo-metabolome Food Microbiol. 83, April, 122–133.

Philpott, C.C. & Protchenko, O., 2008. Response to iron deprivation in *Saccharomyces cerevisiae* Eukaryot. Cell 7, 1, 20–27.

Philpott, C.C., Leidgens, S., et al., 2012. Metabolic remodeling in iron-deficient fungi Biochim. Biophys. Acta - Mol. Cell Res. 1823, 1509–1520.

Piao, H., Hawley, E., et al., 2015. Insights into the bacterial community and its temporal succession during the fermentation of wine grapes Front. Microbiol. 6, 1–12.

Ponomarova, O., Gabrielli, N., et al., 2017. Yeast creates a niche for symbiotic lactic acid bacteria through nitrogen overflow Cell Syst. 5, 4, 345-357.e6.

Pretzer, G., Snel, J., et al., 2005. Biodiversity-based identification and functional characterization of the mannose-specific adhesin of *Lactobacillus plantarum* J. Bacteriol. 187, 17, 6128–6136.

Renouf, V., Claisse, O., et al., 2007. Inventory and monitoring of wine microbial consortia Appl. Microbiol. Biotechnol. 75, 149–164.

Rivero, D., Berná, L., et al., 2015. Hsp12p and PAU genes are involved in ecological interactions between natural yeast strains Environ. Microbiol. 17, 8, 3069–3081.

Rizk, Z., El Rayess, Y., et al., 2016. Impact of inhibitory peptides released by *Saccharomyces cerevisiae* BDX on the malolactic fermentation performed by *Oenococcus oeni* Vitilactic F Int. J. Food Microbiol. 233, 90–96.

Rocha, E.P.C., 2018. Neutral theory, microbial practice: Challenges in bacterial population genetics Mol. Biol. Evol. 35, 6, 1338–1347.

Rollero, S., Bloem, A., et al., 2018. Altered fermentation performances, growth, and metabolic footprints reveal competition for nutrients between yeast species inoculated in synthetic grape juice-like medium Front. Microbiol. 9, 196.

Rossouw, D., Du Toit, M., et al., 2012. The impact of co-inoculation with *Oenococcus oeni* on the transcriptome of *Saccharomyces cerevisiae* and on the flavour-active metabolite profiles during fermentation in synthetic must Food Microbiol. 29, 121–131.

Rossouw, D., Meiring, S.P., et al., 2018. Modifying *Saccharomyces cerevisiae* adhesion properties regulates yeast ecosystem dynamics mSphere 3, 5, e00383-18.

Said, S. Ben & Or, D., 2017. Synthetic microbial ecology: Engineering habitats for modular consortia Front. Microbiol. 8, 1–20.

Sambuk, E. V, Fizikova, A.Y., et al., 2011. Acid phosphatases of budding yeast as a model of choice for transcription regulation research Enzyme Res. 2011, 356093.

Schlecht, U., Suresh, S., et al., 2014. A functional screen for copper homeostasis genes identifies a pharmacologically tractable cellular system BMC Genomics 15, 263.

Semon, M.J., Edwards, C.G., et al., 2001. Inducing malolactic fermentation in Chardonnay musts and wines using different strains of *Oenococcus oeni* Aust. J. Grape Wine Res.

Seo, S.H., Rhee, C.H., et al., 2007. Degradation of malic acid by *Issatchenkia orientalis* KMBL 5774, an acidophilic yeast strain isolated from Korean grape wine pomace J. Microbiol. 45, 6, 521–527.

Serata, M., Iino, T., et al., 2012. Roles of thioredoxin and thioredoxin reductase in the resistance to oxidative stress in *Lactobacillus casei* Microbiology 158, 4, 953–962.

Shakoury-Elizeh, M., Protchenko, O., et al., 2010. Metabolic response to iron deficiency in *Saccharomyces cerevisiae* J. Biol. Chem. 285, 19, 14823–14833.

Shekhawat, K., Patterton, H., et al., 2019. RNA-seq based transcriptional analysis of *Saccharomyces cerevisiae* and *Lachancea* thermotolerans in mixed-culture fermentations under anaerobic conditions BMC Genomics 20, 1, 145.

Shi, X., Zou, Y., et al., 2016. Overexpression of a water-forming NADH oxidase improves the metabolism and stress tolerance of *Saccharomyces cerevisiae* in aerobic fermentation Front. Microbiol. 7, 1427.

Shou, W., Ram, S., et al., 2007. Synthetic cooperation in engineered yeast populations Proc. Natl. Acad. Sci. 104, 6, 1877–1882.

Sieuwerts, S., Bron, P.A., et al., 2018. Mutually stimulating interactions between lactic acid bacteria and *Saccharomyces cerevisiae* in sourdough fermentation LWT - Food Sci. Technol. 90, 201–206.

Succi, M., Pannella, G., et al., 2017. Sub-optimal pH Preadaptation improves the survival of *Lactobacillus plantarum* strains and the malic acid consumption in wine-like medium Front. Microbiol. | www.frontiersin.org 8, 1–12.

Sumby, K.M., Niimi, J., et al., 2019. Ethanol-tolerant lactic acid bacteria strains as a basis for efficient malolactic fermentation in wine: evaluation of experimentally evolved lactic acid bacteria and winery isolates Aust. J. Grape Wine Res. 25, 4, 404–413.

Sun, X., Liu, L., et al., 2019. Effect of high Cu 2+ stress on fermentation performance and copper biosorption of *Saccharomyces cerevisiae* during wine fermentation Food Sci. Technol. 39, 1, 19–26.

Terrade, N. & Mira de Orduña, R., 2009. Determination of the essential nutrient requirements of wine-related bacteria from the genera *Oenococcus* and *Lactobacillus*. Int. J. Food Microbiol. 133, 8–13.

Tesnière, C., Bessière, C., et al., 2019. Relief from nitrogen starvation entails quick unexpected down-regulation of glycolytic/ lipid metabolism genes in enological *Saccharomyces cerevisiae* PLoS One 14, 4, e0215870.

Tilloy, V., Ortiz-Julien, A., et al., 2014. Reduction of ethanol yield and improvement of glycerol formation by adaptive evolution of the wine yeast *Saccharomyces cerevisiae* under hyperosmotic conditions Appl. Environ. Microbiol. 80, 8, 2623–2632.

Tofalo, R., Patrignani, F., et al., 2016. Aroma profile of montepulciano d'abruzzo wine fermented by single and co-culture starters of autochthonous *Saccharomyces* and non-*Saccharomyces* yeasts Front. Microbiol. 7, 610.

Tronchoni, J., Curiel, J.A., et al., 2017. Early transcriptional response to biotic stress in mixed starter fermentations involving *Saccharomyces cerevisiae* and *Torulaspora delbrueckii* Int. J. Food Microbiol. 241, 60–68.

Valladares, R., Graves, C., et al., 2015. H_2O_2 production rate in *Lactobacillus johnsonii* is modulated via the interplay of a heterodimeric flavin oxidoreductase with a soluble 28 Kd PAS domain containing protein Front. Microbiol. 6, 1–14.

van Bokhorst-van de Veen, H., Abee, T., et al., 2011. Short- and long-term adaptation to ethanol stress and its cross-protective consequences in *Lactobacillus plantarum* Appl. Environ. Microbiol. 77, 15, 5247–5256.

Van Cleve, J. & Weissman, D.B., 2015. Measuring ruggedness in fitness landscapes Proc. Natl. Acad. Sci. U. S. A. 112, 24, 7345–7346.

Van den Bergh, B., Swings, T., et al., 2018. Experimental design, population dynamics, and diversity in microbial experimental evolution Microbiol. Mol. Biol. Rev. 82, 3, e00008-18.

Volschenk, H., Van Vuuren, H.J.J., et al., 2003. Malo-ethanolic fermentation in *Saccharomyces* and Schizo*saccharomyces* Curr. Genet. 43, 379–391.

Walker, G. & Stewart, G., 2016. *Saccharomyces cerevisiae* in the Production of Fermented Beverages Beverages 2, 4, 30.

Wang, C., Mas, A., et al., 2016. The Interaction between *Saccharomyces cerevisiae* and non-*Saccharomyces* yeast during alcoholic fermentation is species and strain specific Front. Microbiol. | www.frontiersin.org 7, 502.

Wein, T. & Dagan, T., 2019. The effect of population bottleneck size and selective regime on genetic diversity and evolvability in bacteria Genome Biol. Evol. 11, 11, 3283–3290.

Weinitschke, S., Denger, K., et al., 2007. The DUF81 protein *TauE* in *Cupriavidus necator* H16, a sulfite exporter in the metabolism of C2 sulfonates Microbiology 153, 9, 3055–3060.

Williams, K.M., Liu, P., et al., 2015. Evolution of ecological dominance of yeast species in high-sugar environments Evolution (N. Y). 69, 8, 2079–2093.

Xu, D., Behr, J., et al., 2019. Label-free quantitative proteomic analysis reveals the lifestyle of *Lactobacillus hordei* in the presence of *Saccharomyces cerevisiae* Int. J. Food Microbiol. 294, 18–26.

Yamasaki-Yashiki, S., Sawada, H., et al., 2017. Analysis of gene expression profiles of *Lactobacillus paracasei* induced by direct contact with *Saccharomyces cerevisiae* through recognition of yeast mannan Biosci. Microbiota, Food Heal. 36, 1, 17–25.

Yurdugü, S. & Bozoglu, F., 2002. Studies on an inhibitor produced by lactic acid bacteria of wines on the control of malolactic fermentation Eur. Food Res. Technol. 215, 1, 38–41.

Zapparoli, G., Tosi, E., et al., 2009. Bacterial inoculation strategies for the achievement of Malolactic fermentation in high-alcohol wines South African J. Enol. Vitic. 30, 1, 49–55.

Zengler, K. & Zaramela, L.S., 2018. The social network of microorganisms — how auxotrophies shape complex communities Physiol. Behav. 16, 6, 383–390.

Zhang, H. & Wang, X., 2016. Modular co-culture engineering, a new approach for metabolic engineering Metab. Eng. 37, 114–121.

Zhou, N., Bottagisi, S., et al., 2017. Yeast-bacteria competition induced new metabolic traits through large-scale genomic rearrangements in *Lachancea kluyveri* FEMS Yeast Res. 17, 6, 1–14.

Zhou, N., Ishchuk, O.P., et al., 2019. Improvement of thermotolerance in *Lachancea thermotolerans* using a bacterial selection pressure J. Ind. Microbiol. Biotechnol. 46, 2, 133–145.

Zhou, N., Katz, M., et al., 2018. Genome dynamics and evolution in yeasts: A long-term yeast-bacteria competition experiment PLoS One 13, 4, 1–16.

Zhou, N., Swamy, K.B.S., et al., 2017. Coevolution with bacteria drives the evolution of aerobic fermentation in *Lachancea kluyveri* PLoS One 12, 3, 1–19.

Chapter 3

Biotic selection pressure by yeast strains leads to yeast strain-specific phenotypes in *Lactobacillus plantarum*

Chapter 3: Biotic selection pressure by yeast strains leads to yeast strain-specific phenotypes in *Lactobacillus plantarum*

3.1. Abstract

In wine, *Saccharomyces cerevisiae* often dominates the fermentation landscape producing metabolites that may be inhibitory or supportive of lactic acid bacteria (LAB) which are important for malolactic fermentation (MLF). Both yeast and LAB interact by physical contact or metabolic exchange in wine, presenting an opportunity to study the interaction mechanisms between unrelated species in a well characterised ecosystem. In an effort to understand how the LAB adapt to this selective environment, a directed evolution approach to evolve a strain of *Lactobacillus plantarum*, by using two strains of *Saccharomyces cerevisiae* as selective drivers was implemented. In batch cultures, each yeast strain was co-inoculated with *L. plantarum,* and L-malic acid degradation was used to monitor the progression of MLF. Before completion of each round of MLF the bacteria were harvested by filtration and re-inoculated in fresh media, while the yeast was discarded, and a fresh culture of the original parental strain inoculated with the bacteria. After 50 generations, it was observed that the evolving population was degrading L-malic acid faster than the parent population – an indirect indication of growth and fitness as malic acid is the primary carbon source for LAB. At this point, colonies were randomly selected and characterised with regards to growth and L-malic acid degradation. The data revealed a range of phenotypes, with some evolved isolates showing generic improvement of growth and MLF, regardless of yeast strain used, while other isolates show yeast-specific improvement for either MLF, growth or both. The phenotypes were validated in natural grape must and the data show that in Chardonnay must most of the isolates sustained their respective behaviours, but the same was not true in Pinotage must. This suggests that additional complexity arises in microbial systems as interaction phenotypes are influenced by changes to the chemical matrix of interacting species. Improved interactions may exist across a range of conditions, but may not necessarily be stable in all types of natural grape must environments. Overall though, DE was applied successfully using yeast as selective driver for *L. plantarum* strain improvement. Moreover, I generated an isolate with superior L-malic acid degrading abilities to the parent strain, which is of commercial relevance.

3.2. Introduction

Saccharomyces cerevisiae and lactic acid bacteria (LAB) play separate roles during wine production, alcoholic fermentation (AF) and malolactic fermentation (MLF), respectively. The latter is the degradation of malic acid to lactic acid, which is important for stabilising the microbial load and increasing the flavour complexity of wines (Bartowsky *et al.*, 2004). In spontaneous fermentation LAB are usually suppressed during AF, but bacterial growth increases when the yeast

begins to die-off (Antalick *et al.*, 2013). With many winemakers opting to inoculate fermentations today, such as co-inoculating *S. cerevisiae* and LAB, the yeast and LAB will interact more directly both physically and by metabolic exchange (Ponomarova *et al.*, 2017; Bartle *et al.*, 2019). There have been numerous studies that investigated the impact of yeast on bacteria during fermentation, or the effect of LAB on *S. cerevisiae* (reviewed by Liu *et al.*, 2017). Research shows that *S. cerevisiae* may negatively impact some wine LAB and vice versa through the depletion of nutrients and the production of inhibitory metabolites such as ethanol, fatty acids, acetic acid, succinic acid, sulphur dioxide, antimicrobial peptides and enzymes such as β-glucanase (Guilloux-Benatier *et al.*, 2000; Yurdugü & Bozoglu, 2002; Branco *et al.*, 2014; Liu *et al.*, 2016; Rizk *et al.*, 2016, 2018). Yet, it was also reported that yeast can stimulate LAB growth by releasing essential amino acids to the environment (Ponomarova *et al.*, 2017) or by releasing acetaldehyde which LAB can use as a carbon source (Jussier *et al.*, 2006).

At present, yeast-bacteria interactions in wine are unpredictable and inoculating LAB starter cultures does not guarantee successful MLF. Sluggish or stuck fermentations are known to occur (Malherbe *et al.*, 2007). Apart from physicochemical parameters playing a role in this, inhibition of the bacteria by the inoculated yeast was shown to contribute to this industrial problem (Osborne & Edwards, 2006). It becomes important then to select yeast and bacteria with complementary phenotypes, however, these ‘compatible pairings’ of yeast and bacteria rarely deliver consistent outcomes across cultivars and vintages (Costello *et al.*, 2003). While several studies have focused on LAB strain selection for MLF starter culture development (Coucheney *et al.*, 2005; Lerm *et al.*, 2011; Iorizzo *et al.*, 2016; Brizuela *et al.*, 2019), little attention has generally been paid to LAB strain improvement (Betteridge *et al.*, 2018; Zhao *et al.*, 2019).

Strategies which have been applied to the improvement of individual LAB strains include gene recombination, gene shuffling, mutagenesis and directed evolution (DE) (Betteridge *et al.*, 2015; Zhao *et al.*, 2019). The latter is described as simply the genetic modification of microorganisms through controlled natural selection by exposing the organism(s) of interest to a selective pressure over numerous generations, which eventually results in fit populations and even individual isolates that outgrow and dominate the population structure due to their growth advantage under the particular selective conditions imposed (Cobb *et al.*, 2013). Through evolution, species adapt to their highly dynamic environments because they undergo gene sequence changes (mutations) over time due to spontaneous genetic changes which arise during ordinary cellular events such as DNA replication (Hershberg, 2015). While most of these mutations are deleterious or neutral to the cell, the use of a selective pressure offers the ability to select for the occasional beneficial mutation (Conrad *et al.*, 2011). LAB have a high genome plasticity and have high mutation rates due to a

poor mismatch repair system (Marcobal *et al.*, 2008; Bon *et al.*, 2009; Van Pijkeren & Britton, 2012) making this group ideal for improvement by DE.

In recent years DE was used successfully to improve strains of yeast using abiotic selective pressures such as oxygen, pH, low temperature, sugar saturation, nitrogen, ethanol, and fructose-limitation (McBryde *et al.*, 2006; Novo *et al.*, 2014; Tilloy *et al.*, 2014; López-Malo *et al.*, 2015). Only recently was DE applied to improve a wine-related LAB strain, when ethanol was used as the selective pressure to generate strains with high ethanol (up to 18%v/v) tolerance (Betteridge *et al.*, 2018), in addition, the same strain was evolved further in a multi-stressor environment with increasing concentrations of total SO_2, decreasing pH levels and increasing ethanol levels from 10.9% (v/v) to 16.5% (v/v) (Jiang *et al.*, 2018). Aside from the common practice of SO_2 addition in winemaking, ethanol and SO_2 are also inhibitory products of yeast metabolism. Some strains of *S. cerevisiae* have been shown to produce between around 20 and 60 mg/l SO_2 which impacts MLF (Wells & Osborne, 2011). These and many other yeast derived metabolites may be inhibitory or supportive to the growth and MLF of LAB, presenting an opportunity to study yeast-LAB evolutionary interactions.

This study is the first study, to our knowledge, to use a DE approach whereby the effect (whether positive or negative) of commercial yeast strains on the growth of the LAB is the selective driver for the evolution of a population of *L. plantarum* - i.e. uses a biotic selective pressure. Theoretically, this approach should ultimately yield *L. plantarum* isolates which are more resilient to the potentially inhibitory impacts of the driver yeast strains, and perhaps better able to generally co-exist and grow in the presence of *S. cerevisiae*. The hypothesis here is that the behaviour of the yeast which is driven by metabolic, molecular and physical responses to the fermentation media will supply the necessary evolutionary pressure for the LAB, resulting in isolates that are better adapted to the yeast. It is worth noting that the dynamic response of the yeast to its environment will result in an ever-changing selective pressure towards the bacteria and the principal selective pressures are therefore unknown, as is most likely the case in natural environments (Van den Bergh *et al.*, 2018). Therefore, here the only fixed constant is the yeast, which provides a platform to investigate the impact of wine yeast on the molecular evolution of LAB in fluctuating environments, which has thus far not been investigated.

3.3. Materials and Methods

3.3.1. Strains and growth conditions

The commercial *S. cerevisiae* strains Cross Evolution® (Lallemand SAS, Blagnac, France), EC1118® (Lallemand, Canada) and VIN13 (Anchor Yeast, South Africa) were used in this study.

The selective drivers used were Cross Evolution and EC1118, while VIN13 was used additionally to screen LAB isolates for yeast-specific phenotypes. Cross Evolution is a hybrid *S. cerevisiae* strain which has high ethanol (15% v/v) tolerance, low nitrogen requirements and can inhibit MLF (Lallemand). EC1118 can tolerate ethanol levels of up to 18% (v/v) and is known to produce high amounts (50 mg/l) of SO_2 in low nutrient conditions, which leads to inhibition of MLF (Tennessee Viticultural and Oenological Society). These strains were cultivated on yeast peptone dextrose (YPD) agar (Merck, Modderfontein, South Africa) for 3 days. Pre-cultures of each, prepared by a single colony inoculant in 5 ml YPD broth, incubated aerobically at 30 °C overnight on an orbital shaker. Subsequently, the cultures were used to inoculate 50 ml YPD broth and incubated at 30 ºC overnight on a shaker to obtain cells in mid-exponential phase, prior to inoculation in synthetic grape must (Rollero *et al.*, 2018b).

The *L. plantarum* IWBT B063 strain was taken from the culture collection at the South African Grape and Wine Research Institute (SAGWRI, Stellenbosch University, South Africa) and was used as the parental strain. All bacterial isolates were grown on MRS (De Man, Rogosa and Sharpe) agar (pH 5.2 adjusted with hydrochloric acid (HCl)) with 50 g/l MRS broth and 20 g/l bacteriological agar (Merck, Modderfontein, South Africa). The agar plates were incubated anaerobically using the Millipore Anaerocult® A cultivation jars and strips (Merck, Modderfontein, South Africa) at 30°C for 48 hours. All agar contained 50 mg/l Delvocid Instant (natamycin; DSM Food Specialties, Netherlands) to exclude yeast growth on the plates. Prior to inoculation in synthetic media, cultures were prepared as described above.

3.3.2. Evolution of bacteria by serial transfer

Synthetic grape must (SGM) that resembles standard grape juice (Rollero *et al.*, 2018b) was used with some modifications as shown in Table S3.1. Fermentations were carried out in 80 ml synthetic media, in spice jars fitted with S-shape fermentation airlocks. L-Malic acid degradation was determined by co-inoculating *L. plantarum* IWBT B063 ($2 \cdot 10^7$ CFU/ml) with either Cross Evolution or EC1118 ($1 \cdot 10^6$ CFU/ml) in SGM. Pure culture controls were also prepared. Sampling for L-malic acid concentrations and cell growth was performed on days 2, 4, 6, 8 and 11 to determine the appropriate time to harvest the bacterial cells for re-inoculation in fresh media for the DE experiments. After these initial fermentations it was observed that L-malic acid was completed by day 5 (data not shown). Therefore, it was decided that the yeast and bacteria would be allowed to be in physical contact for 3 extra days after the complete degradation of L-malic acid so that there would be more interaction time between the two species. Therefore, the bacterial cells were harvested after 8 days.

The parent strain was inoculated in the presence of either Cross Evolution or EC1118 at $2 \cdot 10^8$ CFU/ml in triplicate in SGM for all DE fermentations. Each replicate served as an independent evolutionary line (see Figure 3.1). The parent strain in co-culture with either yeast was used as a means of comparison for improvement in evolved populations. No monoculture controls were conducted for the DE experiments, because the conditions in which the DE occur are dynamic (due to yeast metabolism) and are not comparable to the conditions of the monoculture fermentations. For example, the SGM medium used has 230 g/l (Table S3.1) which the yeast utilises producing various metabolites (Walker & Stewart, 2016), changing the conditions of the environment for the evolving LAB. In monoculture fermentations the sugars remain ≤200 g/l (data not shown) and these necessary metabolic changes will not occur, rendering the environment incomparable. Moreover, the use of two different *S. cerevisiae* strains enables the comparison of yeast-specific impacts on the evolution of LAB, therefore, the two yeast serve as comparative controls for each other. Fermentation temperature was set at 20°C and pH at 3.3.

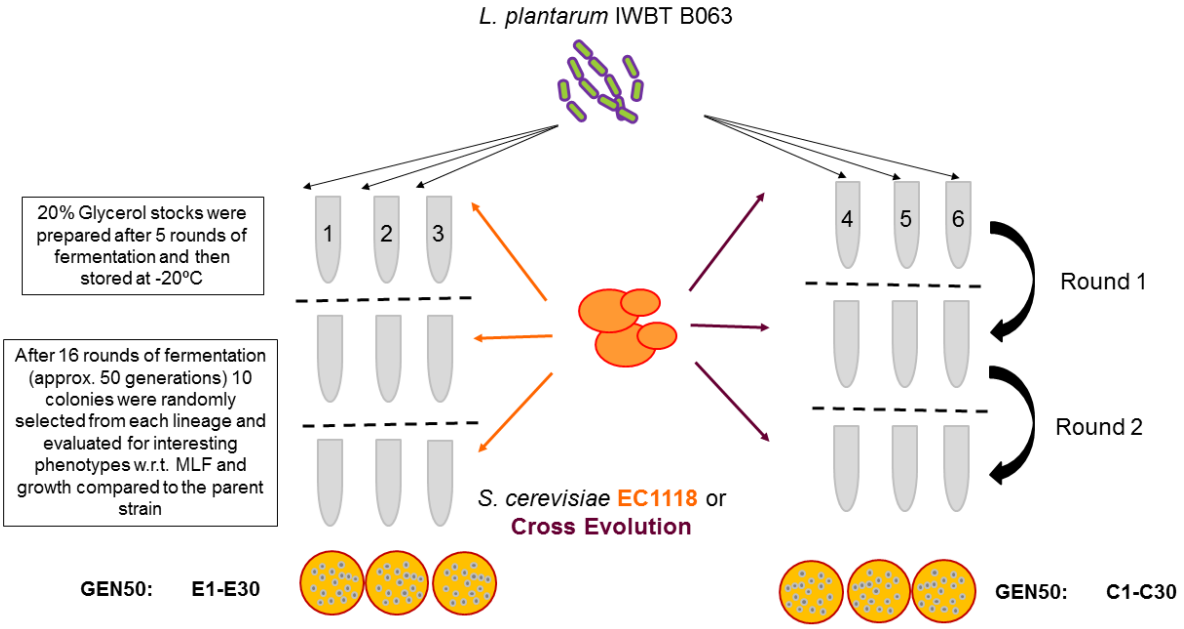

Figure 3.1 Graphic view of the DE experimental setup. Three spice jars labelled 1, 2, and 3 were inoculated with the parent bacteria and EC1118 whereas the other 3 spice jars labelled 4, 5, and 6 were inoculated with the parent bacteria and Cross Evolution. Each of the spice jars represent an independent evolutionary line. The bacteria were harvested at 16 rounds of fermentation at which point 30 colonies were randomly selected and characterised in duplicate micro-fermentations to select evolved isolates. The yeast was added at the beginning of every round and removed at the end of every round by filtration. Random colonies were numbered 1-30 with either an E (EC1118) prefix or a C (Cross Evolution) prefix.

The bacterial population was propagated for 16 rounds of L-malic acid degradation (approximately 50 generations) as shown in Figure 3.1. After each round the SGM was filtered using a Millipore filtration system (Merck™ All-Glass Filter Holder, 47mm, Merck, South Africa) and a Whatman® glass microfiber filter Multigrade GMF 150 with a 47mm diameter and 2µm pore size (Merck, South

Africa) to harvest bacterial cells from the mixed culture fermentations (Figure 3.1). The filtered media was supplemented with 50 mg/l Delvocid Instant (natamycin; DSM Food Specialties, Netherlands) to eliminate any yeast cells that might have been filtered through. The cells were used to inoculate the subsequent round of fermentation in co-inoculation with the original yeast strain populations. If the number of cells were less than 10^8 CFU/ml, the cells were cultured in MRS broth overnight. After every 5 rounds of MLF the bacterial cell cultures were harvested and 20% glycerol stocks were prepared and stored at -80 °C. The total number of cells at the end of each round of fermentation is shown in Figure 3.2.

3.3.3. Screening and selection of potentially evolved isolates

L-malic acid concentration during fermentation was measured enzymatically using an Arena 20XT photometric analyser (Thermo Fisher Scientific, Waltham, MA) and enzymatic assay kits (Enzytec™ Fluid L –malate, kit number: E5280 Roche, R-Biopharm, Germany). Sampling was performed at the beginning of MLF and then every second day for the duration of the fermentation to determine at which point the cells completely degrade L-malic acid.

After 16 rounds of DE in SGM (approx. 50 generations) 60 colonies of potentially evolved bacteria were randomly selected to determine their growth and L-malic acid degradation capabilities. Data from other studies has shown that faster L-malic acid degradation generally reflects better growth (Passos et al., 2003; Derkx et al., 2014), and L-malic acid presents an easy target compound for the purposes of screening a large number of interacting yeast and LAB. Moreover, malic acid is an important acid in winemaking (Volschenk & Van Vuuren, 2006) and biomass production under harsh wine conditions directly impacts MLF rate, (Berbegal *et al.*, 2017). Therefore, these two parameters were used as measures of 'fitness' for each evolved isolate. The initial screening was conducted in 15 ml fermentations (micro-fermentations) in duplicate (Figure 3.1). The bacteria were inoculated both in pure culture (control) and in co-culture with the original driver yeast strains. Control fermentations with the original parent strain were similarly set up and conducted in triplicate. Strains that degraded L-malic acid faster than the parental strain were then evaluated further in the larger scale fermentations (80 ml volumes) complete with cell enumeration. It is worth noting that both L-malic acid degradation and growth of the bacteria were a primary focus in this study. For logistical reasons only L-malic acid degradation was evaluated in the micro-fermentations using the assumption that faster degradation may likely translate to better growth (Passos *et al.*, 2003).

3.3.4. Fitness advantage and probability of selecting isolates with fitness benefits

To estimate fitness advantage of the evolving lines compared to the parent population in co-culture with either yeast driver the growth rates of either populations was calculated using the following

expression: $m = r_1 - r_2$, where r_1 is the growth rate of the evolving population and r_2 is the growth rate of the parent under the specific conditions in which the populations are evolved (Goddard, 2008; Salvadó *et al.*, 2011; García-Ríos *et al.*, 2014). To obtain m the growth rate was determined by $N_f = N_i e^{rt}$, where N_f and N_i are the final and initial cell densities (CFU/ml), t is the period under evaluation. The fitness percentage (w) is 100 ($exp(m) - 1$). To predict after how many generations would an individual with a beneficial mutation move from a frequency of 0.1% to 99.9% was calculated by the following equation: t = 1/m x ln $(p_t \times q_0)/(q_t \times p_0)$, where p_t and p_0 represent the final and initial frequencies of the evolving populations and q_t (1 - pt) and q_0 (1 - p_t) are the final and initial frequencies of the parent population (Goddard, 2008).

3.3.5. Yeast strain specificity screens

The isolates that showed improved growth or MLF as shown by the small and 'large-scale' fermentations were then selected and tested again for fitness in combination with another commercial yeast strain to determine if the isolates have a generic improvement towards all yeast strains or show improvement only when paired with the original 'driver' yeast. The yeast strains that were used included the two 'driver' yeast strains, Cross Evolution and EC1118, along with VIN13 (Anchor Yeast, South Africa). Therefore, isolates that were evolved with EC1118, for example, were tested for improved fitness by co-inoculation with either Cross Evolution or VIN13, and the same for isolates evolved with Cross Evolution, which were then co-inoculated with EC1118 or VIN13. All fermentations were conducted in modified SGM (Table S3.1; Rollero *et al.*, 2018) containing 5 g/l malic acid, 200 g/l total sugars and in 80 ml volumes at 20°C and pH 3.3. The isolates were tested in comparison to the parent and all fermentations were done in triplicate.

3.3.6. Evaluation of interaction phenotypes in grape must

Chardonnay and Pinotage grape must from the 2018 vintage obtained from Doolhof Wine Estate (Wellington, South Africa) were selected to evaluate the robustness of the interaction between the driver yeast strains and the evolved bacterial isolates. Table 3.1 shows the grape must composition of each of the wine cultivars. It should be noted that the composition of each of the grape musts was not altered in any way as the isolates were evolved in SGM and the grape must fermentation experiments were only to characterise the yeast-bacteria interactions in a natural grape must medium.

The bacteria were pre-cultured in MRS broth for 48 hours and washed twice in 0.9% saline solution prior to inoculation in the grape must, whereas the yeast strains were pre-cultured overnight in YPD broth and washed twice with 0.9% saline before inoculation in the grape must. Both yeast (with shaking) and bacteria were incubated at 30°C. The Chardonnay and Pinotage grape musts

were sterilised by thermovinification at 80°C to deactivate/kill the indigenous microbiota (Lisanti *et al.*, 2019). Aliquots of 2.5 L of the heat-treated grape must were stored at 4ºC until further use.

Table 3.1 Composition of natural grape must used in this study before thermovinification.

Parameters	Chardonnay (2018)		Pinotage (2018)	
	Before	After[a]	Before	After[a]
L-malic acid	5.53 g/l	5.36 g/l	2.52 g/l	2.30 g/l
Total sugars	193.21 g/l	194.78 g/l	267.44 g/l	279.12 g/l
Total acidity	7.98	–	4.92	–
Total SO2	23 mg/l	–	9 mg/l	–
pH	3.44	–	3.5	–
YAN	230.86 mg/l	–	152.92 mg/l	–

[a]L-malic acid and sugars were measured before inoculation, as these were followed up on throughout fermentation.

3.3.7. Statistical analysis

Microsoft Excel 2016 with XLSTAT version 16.0 add-in software was used for all data analysis. The means of the colony forming units (CFUs) and L-malic acid concentration throughout fermentation for the parent strain and each of the evolved isolates in both synthetic and natural grape must were analysed using a repeated measures two-way ANOVA with Fisher's least significant distance test.

3.4. Results

3.4.1. Evolving *L. plantarum* populations show increased biomass and L-malic acid degradation after 50 generations

In an effort to study microbial interactions during the conversion of grape must to wine, a two-species system was designed using strains of commercial *S. cerevisiae* (EC1118® and Cross Evolution®) and a non-commercial *L. plantarum* strain. The choice of *L. plantarum* was based on the availability of literature for this species confirming a range of beneficial and inhibitory interactions with yeast (Gobbetti, 1998; Liu *et al.*, 2016; Ponomarova *et al.*, 2017). Each yeast strain was co-inoculated with the bacterial strain to make up 2 sets of 3 evolving populations for the bacteria (Figure 3.1). In this set-up, it is assumed that the dynamic changes to the media undergoing AF and the physical presence of the yeast will provide selective pressure acting on the genome and on the phenotype of the bacteria.

To investigate the potential improvement of evolving *L. plantarum* populations compared to the parent population, L-malic acid degradation and cellular growth were initially measured on days 0,

5 and 8, which marked the start, middle and end of MLF, respectively. After 27 to 28 generations of bacterial growth in such conditions, physiological changes were already observed for all the evolving populations (Figure 3.2), with an increase in growth and a reduction in the time (from 8 days to 5 days) it took the evolving population of LAB to degrade L-malic acid. Therefore, sampling took place at days 0 and 3 from there onwards, with total contact time per batch extended to 5 days. The experiments were continued for at least 50 generations. Figure 3.2 shows that the evolving populations exhibit a decrease in biomass from the first round of the DE experiments (log 8.7 CFU/ml) to 8.6 CFU/ml) by round 10 (approximately 30 generations) and this increased to approximately log 9 CFU/ml by 50 generations. The sudden decrease is observed more so for the populations evolved with EC1118. The growth rates between the parent strain and the evolving populations (after 50 generations) in co-culture with either yeast strains were significantly different ($p < 0.0001$). The data show that the *L. plantarum* population had a fitness advantage of 0.35% ($m = 0.0035\ h^{-1}$) in both yeast-strain co-inoculant media. Although the fitness advantage is low, it would allow the evolved *L. plantarum* to move from a frequency of 0.1% to 99.9% in just 17 rounds of fermentation, which if the rate of growth remains constant over time would mean that the 100th population would have individuals with a high selective advantage.

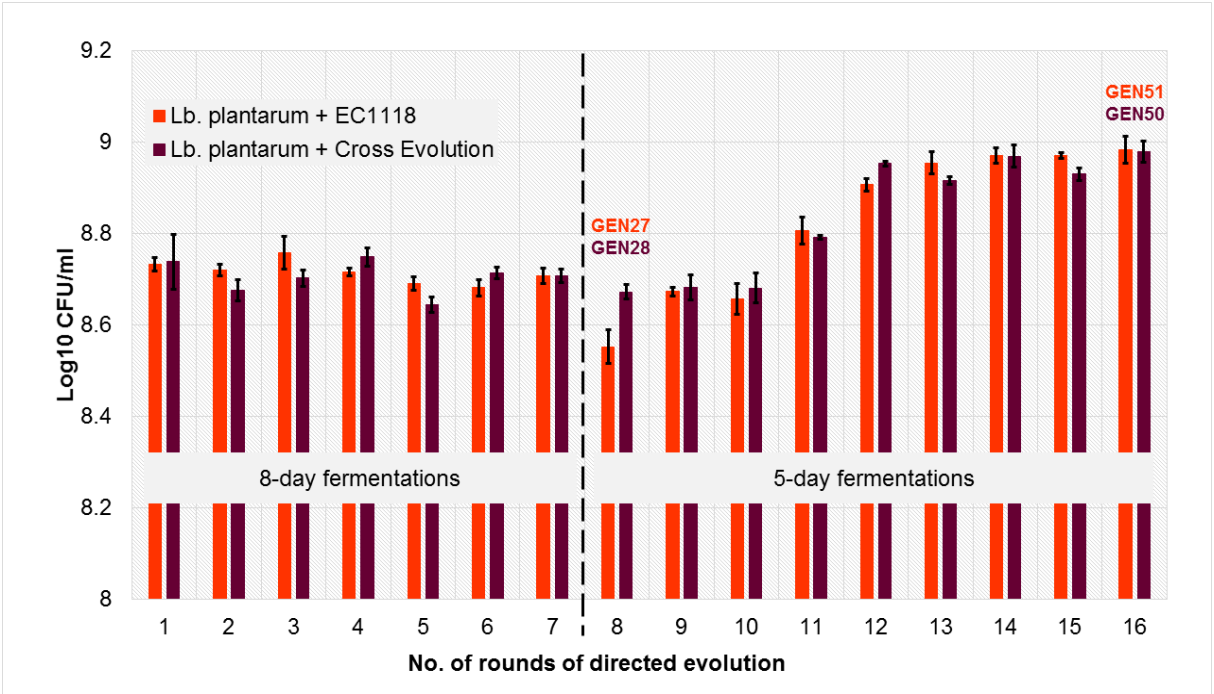

Figure 3.2 Total cell count at the end of each round of fermentation before the bacteria was separated from the yeast. The dotted line shows the separation between the 8- and 5-day fermentations (see text). The change in L-malic acid degradation was observed after 27 and 28 generations, for the respective driver yeast co-cultures (EC1118, orange; Cross Evolution, plum). At approx. GEN50 random isolates were selected for further characterisation. Data shows averages of at least 3 replicates and the error bars show standard deviations.

Sixty random colony isolates were screened, and the data showed that 23 isolates evolved with EC1118 completely degraded 3 g/l L-malic acid between 3 and 6 days, and only 12 isolates evolved with Cross Evolution degraded 0.9-1.4 g/l out of 3 g/l within the same time period (data not shown). Of the 23 isolates evolved with EC1118 10 isolates showing varying L-malic acid degradation patterns, and all 12 isolates co-cultured with Cross Evolution were selected for validation in larger volume fermentations.

3.4.2. EC1118-directed evolution promotes cell growth and Cross Evolution-directed evolution suppresses growth and L-malic acid degradation in the selected isolates

To confirm the individual phenotypes of each evolved isolate in terms of L-malic acid degradation, the micro-fermentations were up-scaled to 80 ml. Furthermore, in view of the differences between the evolved population as a whole and individual isolate in terms of the degradation of L-malic acid, cells were enumerated to generate viable cell count data. In the up-scaled experiments very different results were observed compared to the small volume screens. The reasons for the discrepancies between the small and larger volume fermentations could be due to various factors, such as continuous exposure to oxygen due to sampling of the fermentations daily, which could have slowed down the fermentation rates. *L. plantarum* is a facultative anaerobe (Kandler and Weiss, 1986), however, exposure to oxygen may sometimes delay its growth and fermentation capabilities (Papadimitriou et al. 2016). It is possible, also, that the bacteria were not getting sufficient nutrients as they formed sediments at the bottom of the 15 ml vials, and failure to completely re-suspend the cells may have restricted the majority of the cells from accessing nutrients in the media (Ferenci & Spira, 2007). Further investigation is required to better understand the impact of fermentation volume on the MLF capabilities of the evolved isolates.

The data shows that 6 of the 10 isolates (E14, E15, E20, E22, E26 and E29) co-inoculated with EC1118 degraded 3 g/l malic acid in 4 days, which was the same as the parent strain, and isolates E5, E17, E18 and E30 took 6 days to completely degrade 3 g/l L-malic acid (Table 3.2). The bacteria were inoculated at a log CFU/ml value of 7 (± 0.0-0.15) and the biomass of all isolates was between 8.1 (± 0.47) and 8.48 (± 0.02) log CFU/ml, respectively, by the end of the exponential phase. Only isolates E5, E15, E18, E22 and E26 were still growing at the end of fermentation with log CFU/ml values of 7.97 (± 0.03), 7.88 (± 0.32), 8.17 (± 0.04), 7.99 (± 0.16) and 8.08 (± 0.08), respectively (Table 3.2).

The isolates that were evolved with Cross Evolution as the selective driver generally showed poor L-malic acid degradation capabilities as only C4, C5, C14, C26 and the parent strain completely degraded L-malic acid by day 12, 4, 8, 14, and 6 respectively. All other isolates failed to completely degrade L-malic acid by day 14 (Table 3.2). Isolate C5 was the only isolate to degrade all the malic

acid 2 days before the parent strain, and it was the only isolate that grew to a log CFU/ml value of 8.21 (± 0.08), satisfying the assumption that fast MLF translates to faster growth (Derkx *et al.*, 2014). All the isolates that failed to degrade L-malic acid showed no growth throughout fermentation.

Table 3.2 Growth and MLF of isolates that were selected for further characterisation in synthetic grape must (SGM) with 3 g/l L-malic acid.

Isolates (lineage)	Initial cell count	Maximum cell count	cell count at end of fermentation	MLF duration (days)
		EC1118-directed evolution		
Parent	**7.01 (±0.15)**	**8.20 (±0.14)**	**7.92 (±0.04)**	**4**
E5 (1)	7.04 (±0.06)	8.40 (±0.02)	7.97 (±0.03)	6
E14 (2)	7.02 (±0.03)	8.39 (±0.00)	7.27 (±0.70)	4
E15 (2)	7.00 (±0.00)	8.34 (±0.11)	7.88 (±0.32)	4
E17 (2)	7.11 (±0.05)	8.29 (±0.20)	7.78 (±0.05)	6
E18 (2)	7.00 (±0.00)	8.11 (±0.47)	8.17 (±0.04)	6
E20 (2)	7.06 (±0.03)	8.19 (±0.15)	7.79 (±0.27)	4
E22 (3)	7.00 (±0.00)	8.48 (±0.02)	7.99 (±0.16)	4
E26 (3)	7.10 (±0.02)	8.25 (±0.01)	8.08 (±0.08)	4
E29 (3)	7.18 (±0.04)	8.37 (±0.12)	7.76 (±0.21)	4
E30 (3)	7.00 (±0.00)	8.24 (±0.05)	7.63 (±0.20)	6
		Cross Evolution-directed evolution		
Parent	**7.04 (±0.06)**	**7.29 (±0.84)**	**7.29 (±0.84)**	**6**
C4 (4)	7.12 (±0.02)	7.12 (±0.02)	6.68 (±0.84)	12-14
C5 (4)	7.06 (±0.02)	8.21 (±0.08)	7.18 (±0.00)	4
C7 (4)	7.10 (±0.02)	7.10 (±0.02)	na†	nc‡
C12 (5)	7.06 (±0.03)	7.29 (±0.27)	na†	nc‡
C14 (5)	7.06 (±0.01)	7.34 (±0.06)	6.61 (±0.22)	8
C20 (5)	7.09 (±0.01)	7.09 (±0.01)	na†	nc‡
C22 (6)	7.11 (±0.00)	7.20 (±0.12)	na†	nc‡
C24 (6)	7.04 (±0.06)	7.04 (±0.06)	na†	nc‡
C26 (6)	7.10 (±0.00)	7.45 (±0.52)	5.80 (±0.14)	14
C28 (6)	7.10 (±0.02)	7.19 (±0.02)	na†	nc‡
C29 (6)	7.11 (±0.00)	7.59 (±0.25)	na†	nc‡
C30 (6)	7.08 (±0.01)	7.08 (±0.01)	na†	nc‡

The data shows log CFU/ml (± standard deviation) of the bacterial isolates in synthetic grape must. Code: na† = not applicable, nc‡ = fermentation not completed by day 14 (i.e. isolates were selected because they performed better than the parent strain in 6-day micro-fermentations).

In summary, the data suggest that the E- isolates can reach higher cell counts than the parent strain, but degrade L-malic acid in the same 4-day time period. In contrast, the majority of the C-

isolates do not grow and fail to degrade all the malic acid in 14 days. These phenotypes were categorised into 4 groups (Figure 3.3): 1) 'good MLF/growth' for isolates that degrade all L-malic acid in 4 days and grow better than the parent (log CFU/ml > 8.2) when in co-culture with EC1118 (E14, E15, E22, E26, and E29) and or Cross Evolution (C5; log CFU/ml > 7.3). 2) 'Good MLF/ poor growth' for isolates that degraded all the L-malic acid, but do not show superior growth to the parent when in co-culture with EC1118 (E18 and E20) or Cross Evolution (C4), respectively. 3) 'Poor MLF/ good growth' for isolates that take more than 4 or 6 days to completely degrade L-malic acid, but grow better than the parent in co-culture with EC1118 (E5, E17, E30) or Cross Evolution (C14 and C26), respectively. Lastly, 4) 'poor MLF/growth' for isolates that could neither grow nor completely degrade L-malic acid regardless of co-cultured yeast strain (majority of C- isolates). Only isolate E15, E22 and E26 degraded all malic acid in the same time as the parent and were still growing thereafter. E22 and E26 were selected for further investigation. It is also worth mentioning that isolates E22 and E26 are both from lineage 3 (Figure 3.1). Isolates C5 and C26 (from lineage 4 and 6) were also selected for further study.

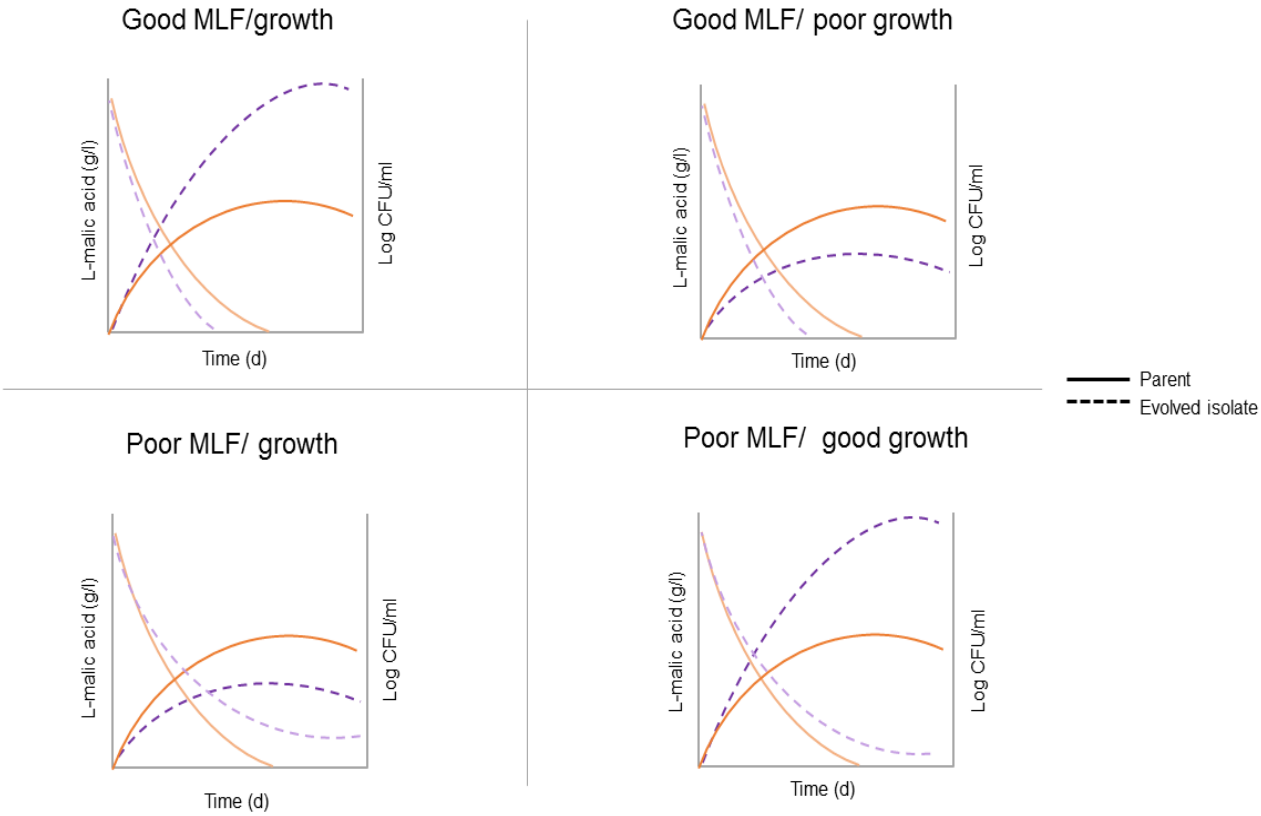

Figure 3.3 Graphical representation of the general phenotypes observed among different evolved isolates. The two y-axes represent residual L-malic acid and log CFU/ml, respectively. The x-axis represents the period of L-malic acid degradation in days. The parent strain is shown as solid lines and the evolved isolate as dashed lines. Specific details of each phenotype are described in the text.

3.4.3. Phenotypes are driver-yeast strain specific and are highly affected by the environment

The next step was to investigate the dependency of selected isolates on their 'driver' yeasts. Would the isolates show similar behaviours if paired with 'non-driver' yeast or even in the absence of yeast? Selected isolates were tested in SGM containing 5 g/l L-malic acid increasing contact time between the yeast and bacteria. Each of the selected isolates were inoculated as follows: 1) in monoculture; 2) in co-culture with EC1118; 3) Cross Evolution; and 4) VIN13. Two E-isolates, namely E22 and E26, showing good MLF/growth and 2 C-isolates showing good MLF/growth (C5) and poor MLF/good growth (C26) were selected.

The yeast strain, Cross Evolution suppressed the growth of the parent strain as well as the C-isolates ($p < 0.0001$). In contrast, the E-isolates show significant growth ($p < 0.001$) in the presence of EC1118 (Figure 3.4 and Figure 3.5). Interestingly, isolate E22 is the only strain that degraded L-malic acid only when in co-culture with yeast in general; in fact, this strain degraded all the malic acid 1-2 days earlier than the parent strain when in co-culture with its driver yeast EC1118 ($p = 0.06$), Cross Evolution ($p = 0.008$) and VIN13 ($p < 0.001$) (Figure 3.4B), while E26 performs similarly ($p > 0.05$) to the parent strain, regardless of yeast strain co-inoculant (Figure 3.4C). Notably, E22 was significantly poorer ($p < 0.001$) at degrading L-malic acid when in monoculture compared to the parent strain (Figure 3.4B). In contrast isolate E26 degraded significantly more L-malic acid between days 4 and 8 ($p < 0.001$) in monoculture fermentation, while the parent strain monoculture fermentation still had 1 g/l residual L-malic acid by day 8 (Figure 3.4D). The growth of E22 was only significant for days 3 and 8 compared to the parent strain in co-culture with EC1118 ($p < 0.001$, respectively) and VIN13 ($p = 0.000$ and $p = 0.005$, respectively). The significance in growth was also observed in co-culture with Cross Evolution between days 3 and 5 ($p = 0.002$ and $p = 0.014$), respectively compared to the parent strain (Figure 3.4A). Similarly, E26 grew significantly better between days 3, 5, and 8 when in co-culture with EC1118 ($p = 0.001$, $p = 0.027$, 0.000) and Cross Evolution ($p < 0.001$, $p = 0.000$, $p = 0.008$), respectively (Figure 3.4C). Isolate E26 showed no significant growth compared to the parent strain when co-cultured with VIN13 between days 3 ($p = 0.058$) and 8 ($p = 0.446$). Isolate E22 showed a significantly variable growth pattern compared to the parent strain ($p < 0.001$) and there was no significant difference between E26 and the parent strain ($p = 0.585$) in monoculture fermentations. This data suggests that the 'good MLF/growth' phenotype is maintained in E22 when in co-culture with all yeast strains used. Isolate E26 on the other hand has the 'poor MLF/good growth' phenotype when in coculture with both EC1118 and Cross Evolution. In monoculture fermentations E22 displayed a 'poor MLF' phenotype and the growth is inconclusive, while E26 displayed a 'good MLF/poor growth' phenotype.

Isolate C5 did not grow throughout fermentation. Cell numbers were reduced from 7.2 log CFU/ml at the start of fermentation to about 6.6 log CFU/ml in co-culture with Cross Evolution and VIN13 and 5.6 log CFU/ml in monoculture and co-culture with EC1118 by the end of fermentation (Figure 3.5A). Furthermore, this isolate had degraded 4 g/l of the total 5 g/l L-malic acid during the 8-day fermentation period when co-cultured with Cross Evolution ($p < 0.0001$) and when in monoculture ($p < 0.023$); and only 2-3 g/l L-malic acid when co-cultured with VIN13 ($p < 0.0001$) and EC1118 ($p < 0.0001$) (Figure 3.5B). Isolate C26 had significantly more residual L-malic acid between days 4 and 6 compared to the parent strain when in co-culture with all yeast strains used ($p < 0.0001$ for all strains). No significant difference was observed for the degradation of L-malic acid between C26 and the parent strain in monoculture fermentations ($p = 0.112$). Moreover, C26 shows significantly better growth to the parent strain in monoculture from day 3 ($p = 0.001$), day 5 ($p = 0.004$), and day 8 ($p = 0.018$). Similar results were observed when this isolate was co-inoculated with Cross Evolution ($p < 0.001$, $p = 0.000$, $p = 0.007$) and EC1118 ($p = 0.000$, $p = 0.001$, $p = 0.025$), respectively. Significant differences in growth were observed on days 3 ($p < 0.001$) and 8 ($p = 0.000$) only when grown together with VIN13. These data confirm the inhibitory effects of Cross Evolution (Scott Laboratories, 2018), evidenced by the poor growth of the parent strain. There could be a possible trade-off between survival and reproduction seeing that there is no growth, yet these isolates still utilise L-malic acid efficiently (at least that is the case for C26). The behaviour of each isolate is summarised in Table 3.3 and shows which isolate shows a significant difference to the parent strain in terms of growth and L-malic acid degradation in SGM.

Table 3.3 Summary of phenotypes of each of the evolved isolates after screening in co-culture with different yeast strains.

Yeast-strain treatment	Evolved isolates*			
	C5	**C26¥**	**E22**	**E26**
Cross Evolution	**--/--**	**--/--**	**+/+**	**+/0**
EC1118	**--/--**	**-/--**	**++/++**	**0/++**
VIN13	**--/--**	**--/-**	**++/++**	**0/0**
No yeast	**--/0**	**0/0**	**-/0**	**++/0**

*Data shows growth (left side of forward slash) and L-malic acid degradation (right side of forward slash) of evolved isolates when compared to the parent strain in co-inoculation with selected yeast strain or in single culture in SGM.

¥Although C26 shows poor growth and MLF compared to the parent, this strain grows significantly better than C5 which was evolved with the same driver yeast. (–: poor MLF/growth, +: good MLF/growth, 0: same MLF/growth as parent strain). Double signs show significant differences compared to the parent strain and single signs show non-significant differences.

3.4.4. Change of environment affects behaviour of evolved isolates

The growth and malolactic characteristics displayed by 2 of the 4 evolved isolates tested in section 3.4.3 above were evaluated in different conditions (chemically distinct grape musts). Isolate E22

and C26 were selected as they showed significant differences to the parent strain when paired with their driver yeast strains. Note that the parent strain was tested in Pinotage wine and was shown to have superior MLF properties compared to other LAB (Lerm *et al.*, 2011), therefore each of these isolates were tested in Pinotage grape must and Chardonnay grape must. The latter was selected because it is one of the white wines which undergo MLF and was used previously to investigate yeast-bacteria interactions (Liu *et al.*, 2016).

There was a difference in the results of the bacteria-yeast co-cultures in the two types of grape must (Table 3.4). This is evidenced by the growth kinetics of each of the isolates and the parent strain (data not shown). All tested isolates, including the parent strain struggled to grow in Pinotage grape must (from log 7.2 CFU/ml at day 0 to log 5.1 CFU/ml at day 8), but grew in the first 2 days in Chardonnay grape must (from log 7.2 CFU/ml at day 0 to log 8.2 at day 2) before declining to log 6.4 CFU/ml (data not shown). Isolate E22 degraded more (89.04%) L-malic acid when in co-culture with EC1118 in Chardonnay grape must (no statistical significance, $p = 0.72$) compared to the parent strain (72.86%) by day 12 of the fermentation (Table 3.4), supporting the data observed in the SGM (Figure 3.4B). Both E22 and the parent completely degrade L-malic acid in the Pinotage grape must by day 6, regardless of yeast strain co-inoculant. The 'newly' observed phenotype displayed by C26 in SGM ('poor MLF/growth') was evident in the Pinotage grape must (Table 3.4). C26 showed no significant difference to the parent strain in the Chardonnay grape must, regardless of yeast co-inoculant ($p = 0.057$ and $p = 0.063$, respectively), in terms of L-malic acid degradation (Table 3.4). These data show that C26 displays inconsistent phenotypes that are largely affected by change in growth conditions. Overall, the data suggest that the fermentation matrix drastically impacts the phenotypes of the isolates evolved with Cross Evolution i.e. C5 and C26.

Table 3.4 Summary of phenotypes of C26 and E22 in co-culture with different yeast strains in Chardonnay and Pinotage grape must. (–: poor MLF/growth, +: good MLF/growth, 0: same MLF/growth as parent strain). Double signs show significant differences compared to the parent strain and single signs show slight differences.

Yeast-strain treatment	C26		E22	
	Chardonnay	Pinotage	Chardonnay	Pinotage
Cross Evolution	**0/--**	**--/--**	**+/0**	**--/0**
EC1118	**0/--**	**--/--**	**++/0**	**--/0**

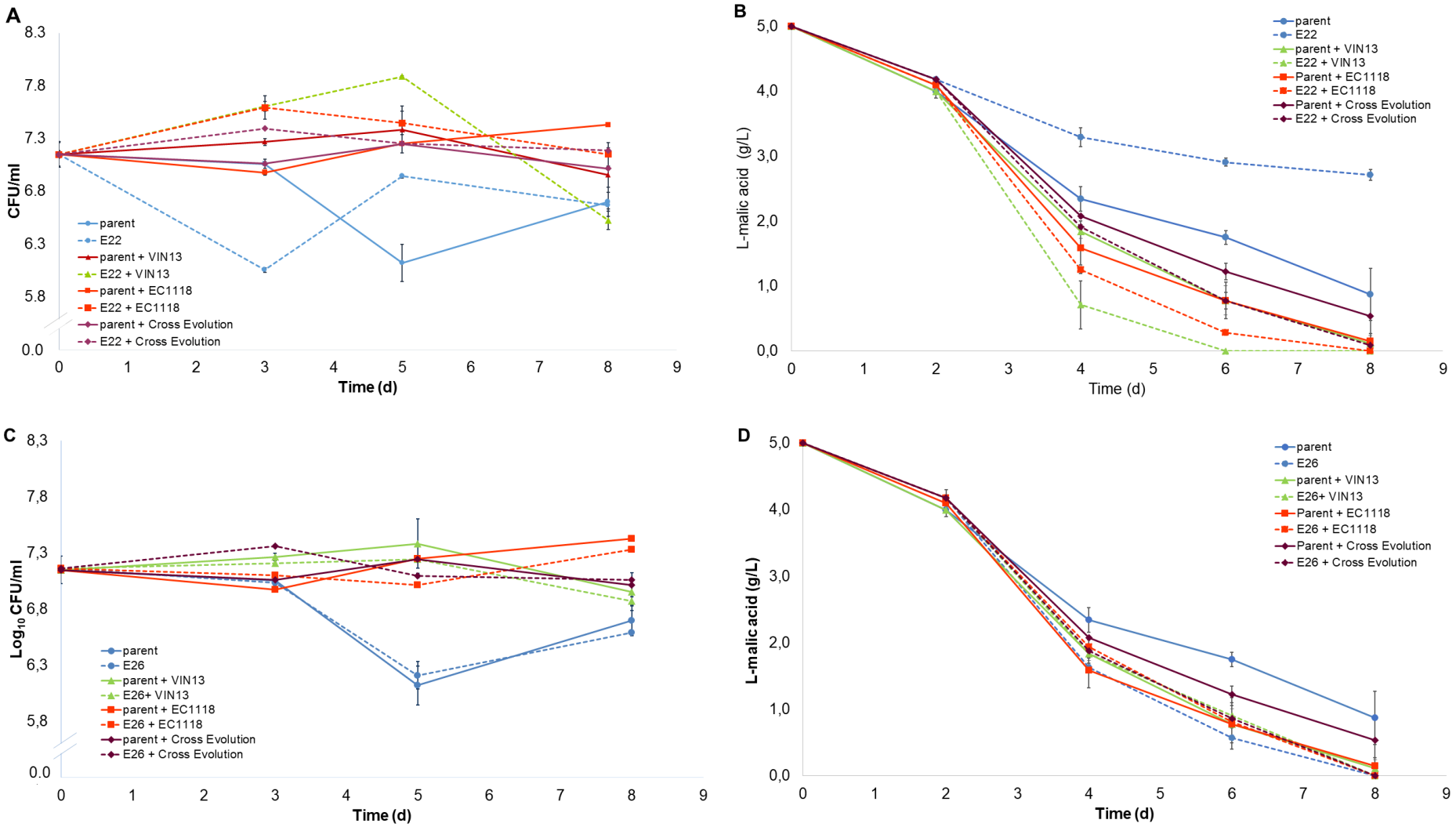

Figure 3.4 CFU/ml quantification of evolved isolates and parent strain and residual L-malic acid concentration throughout fermentation in SGM. Dotted lines represent the evolved isolates, while solid lines represent the parent strain. Isolate E22 data is shown in frames A and B, and isolate E26 data in frames C and D. Each of the bacteria was inoculated in monoculture (blue lines), in co-culture with EC1118 (orange lines), Cross Evolution (plum lines) and VIN13 (green lines). Data is averages of 3 biological repeats and error bars show the sample standard deviation.

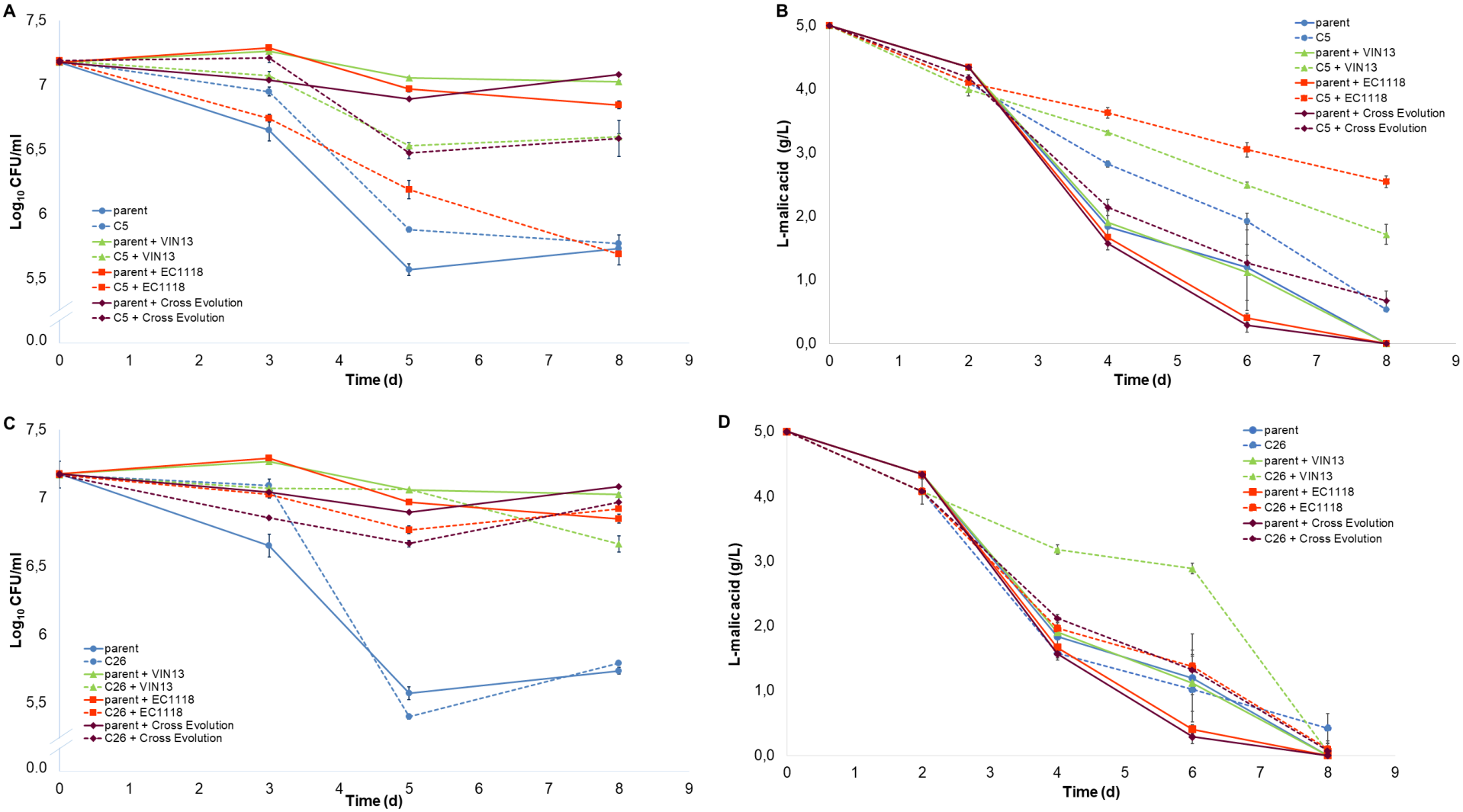

Figure 3.5 CFU/ml quantification of evolved isolates and parent strain and residual L-malic acid concentration throughout fermentation in SGM. Dotted lines represent the evolved isolates, while solid lines represent the parent strain. Isolate C5 data is shown in frames A and B, and isolate C26 data in frames C and D. Each of the bacteria was inoculated in monoculture (blue lines), in co-culture with EC1118 (orange lines), Cross Evolution (plum lines) and VIN13 (green lines). Data is averages of 3 biological repeats and error bars show the sample standard deviation.

3.5. Discussion

S. cerevisiae and *L. plantarum* both inhabit wine environments and competitive interactions may arise between these two species in favour of the yeast species (Kennes *et al.*, 1991). In support of this theory, the data show that yeast-driven DE using two different yeast strains results in yeast-strain specific phenotypes which may directly impact the fitness of evolved *L. plantarum* isolates. Specifically, DE in the presence of EC1118 increased growth and L-malic acid degradation of E22 and E26 when in co-culture. The data suggest that isolate E22 depends on the presence of yeast for growth and MLF, while yeast suppress the MLF capabilities of E26. Previous reports that some *S. cerevisiae* strains release essential nutrients to the media shared with LAB are supported by this data seeing that both isolates grew well in the presence of the yeast (Ponomarova *et al.*, 2017). It is interesting that E22 degraded L-malic acid faster than E26 when both isolates were evolved in the presence of EC1118 and within the same evolutionary line (fermentation flask). It is possible that E26 is not efficient at degrading L-malic acid because it was not essential for it to do so in the larger evolving population, perhaps the amount of L-malic acid it utilised was sufficient for its growth (West *et al.*, 2006) and that it behaves like a generalist (Martino *et al.*, 2016), contrary to E22 which grows under special conditions provided by yeast. .

In contrast, isolates evolved in the presence of Cross Evolution struggled when in co-culture with *S. cerevisiae* strains. The prolonged exposure of these isolates to the inhibitory yeast strain, Cross Evolution, has resulted in an overall poor growth and L-malic acid degradation phenotype. It is possible that these isolates are going through a fitness valley given their negative growth rates (Tadrowski *et al.*, 2018). Although, the parent strain was also suppressed by Cross Evolution its growth rate was close to zero. For these isolates to cross the fitness valley applying the DE experiments in slightly different environments may result in a rugged fitness landscape which provides different fitness peaks on which the mutations lying in fitness valley move up (Steinberg & Ostermeier, 2016). Inasmuch as the fitness advantage (*w*) between the populations evolved with EC1118 or Cross Evolution was the same (0.35%), individual isolates belonging to these populations such as E22 showed improved phenotypes that are adaptive in the presence of all yeast strains tested against. Therefore, it is safe to assume that some beneficial mutations existed prior to the start of the DE experiments and may perhaps be carried by more individuals than what was characterised in this study. Moreover, logistical limitations have prevented the analyses of more isolates, or even isolates from later generations which theoretically would have a high fitness advantage as per the prediction in the Results section. Therefore, characterising isolates from the 100^{th} generation would confirm the results presented here.

Although, this work is the first to present such data for wine LAB, it was shown that *Pseudomonas fluorescens,* a soil bacterium, had high specificity for strains of the same species which had a high

competitive advantage, but not to those with low competitive advantage (Zhao *et al.*, 2018). Moreover, Zhao *et al.* (2018) showed that stronger competitors reduced the relative fitness, possibly due to a reduction in the number of available mutations for selection, as a result of reduced population size. Similarly, our data suggests that EC1118 and Cross Evolution are strong competitors, but it was only in co-culture with Cross Evolution that significant reduction in population size was observed for the evolved isolates. Moreover, isolate C5 had a generally poor phenotype in the presence or absence of Cross Evolution. Similar data was observed for *P. fluorescens* evolved in the presence of *Pseudomonas putida*, providing a strong biotic pressure (Hall *et al.*, 2018).

The opposing phenotypes between isolates evolved with EC1118 and Cross Evolution suggest that the competition of these yeasts and the *L. plantarum* population may have resulted in increased selection pressure and reduction in population size, respectively (Hall *et al.*, 2018). The increase in selection pressure increases mutation which might be neutral, deleterious or adaptive (Osmond & de Mazancourt, 2013). In the case of E22 and E26 significant differences in growth are observed h compared to the parent strain when in co-culture with different yeast, but more so with its driver yeast, which suggest that these isolates may possess adaptive mutations which allow them to thrive over the parent population even beyond the environment in which they were evolved (grape must). Isolate E22 on its own does not show any adaptive changes, but rather deleterious changes as seen by the lack of growth compared to the parent strain, which emphasises the yeast-specific phenotypes.

It was suggested that if competing species have niches that partially overlap, the individuals in the target population which are furthest from the new niche will be suppressed by the stronger competitor, in this way increasing adaptation of the target population (Osmond & de Mazancourt, 2013). Yeast and bacteria compete for various metabolites such as glucose and vitamins (Nehme *et al.*, 2008; Jiang *et al.*, 2018), in this study the yeast and bacteria occupied partial niches because the yeast mainly utilises glucose and fructose as carbon sources and the bacteria utilises malic acid, although glucose regulates MLF (Kennes *et al.*, 1991; Betteridge *et al.*, 2015; Mendoza *et al.*, 2017). In this work, the L-malic acid concentration was monitored in each fermentation flask and serial dilutions were carried out to calculate CFU/ml. The degradation of L-malic acid of these evolving populations was compared to that of the parent strain under the same conditions (as described in the methods section). Populations that degraded L-malic acid faster than the parent population were assumed to have individuals with potentially improved phenotypes. Therefore, isolates were randomly selected from these populations. In such a system, isolates showing poorer growth and/or MLF were also selected at random. It is worth noting that in this study improved growth and L-malic acid degradation compared to the parent were the criteria for an ‘improved

phenotype'. One can assume that these isolates may have other traits that are beneficial to the population or the individual which we did not evaluate (West *et al.*, 2006).

The DE applied here utilised biotic selective pressures, which were the only constants in a changing environment, which was necessary for comparison of yeast-specific effects on the microbial evolution of *L. plantarum.* However, this resulted in fluctuating environments that were exacerbated further by the technique used to separate the evolving bacteria from the yeast strains which were added from the mother culture at the beginning of each cycle of DE. The serial transfer method used introduced bottlenecks which may have reduced genetic variation (Wein & Dagan, 2019) and reduced selection, especially in the populations evolved with Cross Evolution where the majority of isolates showed poor growth and MLF. Additionally, this method resulted in constant exposure of cells to alternating rich and poor nutrient conditions introducing another selective pressure (Van den Bergh *et al.*, 2018). Taken together, the use of biotic drivers in wine-like conditions, filtering the bacteria and the feast/starve regime introduced selective pressures that could not be measured, but were ideal here because they resemble what would normally occur in natural environments (Bleuven & Landry, 2016). Nonetheless, selection of isolates from a growth and malic acid degradation level alone is limiting, because other significant biological changes which do not drastically affect growth may have been missed. For example, Zhou *et al.* (2017) found that a *Lachancea kluyveri* evolved in the presence of different bacteria showed poor growth in 20 out of 95 carbon sources, but had acquired new traits which increased its fitness in various growth conditions. Therefore, further investigations are warranted to elucidate the depth of the impact of such strong selection pressures on evolved *L. plantarum* isolates.

3.6. Conclusion

In summary this work suggests that yeast-driven DE results in yeast-strain specific phenotypes after 50 generations, due to possibly strong competitive interactions between the yeast and bacteria. Moreover, Cross Evolution reduces population and possibly genetic variation which led to inconsistent data and generally no adaptation. This data supports similar studies exploring impacts of biotic stress on microbial evolution (Hall *et al.*, 2018; Zhao *et al.*, 2018), although this is the first to apply this to wine *L. plantarum* and *S. cerevisiae*. Allowing only the yeast to remain constant resulted in at least one isolate (E22) that showed adaptation to yeast in general, but performed significantly better in co-culture with its driver yeast strain (EC1118). The same isolate showed could neither grow nor degrade L-malic acid to the wine environment in monoculture, suggesting positive selection towards a yeast-specific phenotype. These data contribute to our understanding of yeast-bacteria interactions in wine, revealing that yeast drive the evolution of 'weaker' competitors in wine. The application of DE using the serial transfer technique has provided a system that can be applied to investigate microbial interactions under conditions that closely

resemble natural environments. However, this system introduces many bottlenecks, which must be reduced in future systems.

References

Antalick, G., Perello, M.C., et al., 2013. Co-inoculation with yeast and LAB under winery conditions: Modification of the aromatic profile of merlot wines South African J. Enol. Vitic. 34, 2, 223–232.

Bartle, L., Sumby, K., et al., 2019. The microbial challenge of winemaking: yeast-bacteria compatibility FEMS Yeast Res. 19, 4, 1–16.

Bartowsky, E.J., Henschke, P.A., et al., 2004. Chasing wine aroma. Does *Oenococcus oeni* have the potential to release aroma compounds from authentic grape precursors?

Berbegal, C., Spano, G., et al., 2017. Microbial resources and innovation in the wine production sector South African J. Enol. Vitic. 38, 2, 156–166.

Betteridge, A., Grbin, P., et al., 2015. Improving *Oenococcus oeni* to overcome challenges of wine malolactic fermentation Trends Biotechnol. 33, 9, 547–553.

Betteridge, A.L., Sumby, K.M., et al., 2018. Application of directed evolution to develop ethanol tolerant *Oenococcus oeni* for more efficient malolactic fermentation Appl. Microbiol. Biotechnol. 102, 2, 921–932.

Bleuven, C. & Landry, C.R., 2016. Molecular and cellular bases of adaptation to a changing environment in microorganisms Proc. R. Soc. B Biol. Sci. 283, 20161458.

Bon, E., Delaherche, A., et al., 2009. *Oenococcus oeni* genome plasticity is associated with fitness Appl. Environ. Microbiol. 75, 7, 2079–2090.

Branco, P., Francisco, D., et al., 2014. Identification of novel GAPDH-derived antimicrobial peptides secreted by *Saccharomyces cerevisiae* and involved in wine microbial interactions Appl. Microbiol. Biotechnol. 98, 2, 843–853.

Brizuela, N., Tymczyszyn, E.E., et al., 2019. *Lactobacillus plantarum* as a malolactic starter culture in winemaking: A new (old) player? Electron. J. Biotechnol. 38, 10–18.

Cobb, R.E., Chao, R., et al., 2013. Directed evolution: Past, present, and future AIChE J. 59, 5, 1432–1440.

Conrad, T.M., Lewis, N.E., et al., 2011. Microbial laboratory evolution in the era of genome-scale science Mol. Syst. Biol. 7, 509.

Costello, P.J., Henschke, P.A., et al., 2003. Standardised methodology for testing malolactic bacteria and wine yeast compatibility Aust. J. Grape Wine Res. 9, 2, 127–137.

Coucheney, F., Desroche, N., et al., 2005. A new approach for selection of *Oenococcus oeni* strains in order to produce malolactic starters Int. J. Food Microbiol. 105, 3, 463–470.

Derkx, P.M.F., Janzen, T., et al., 2014. The art of strain improvement of industrial lactic acid bacteria without the use of recombinant DNA technology Microb. Cell Fact. 13, 1–13.

Ferenci, T. & Spira, B., 2007. Variation in stress responses within a bacterial species and the indirect costs of stress resistance Ann. N. Y. Acad. Sci. 1113, 105–113.

García-Ríos, E., Gutiérrez, A., et al., 2014. The fitness advantage of commercial wine yeasts in relation to the nitrogen concentration, temperature, and ethanol content under microvinification conditions Appl. Environ. Microbiol. 80, 2, 704–713.

Gobbetti, M., 1998. The sourdough microflora: Interactions of lactic acid bacteria and yeasts Dtsch. Bäcker 9, 4, 209–215.

Goddard, M.R., 2008. Quantifying the complexities of *Saccharomyces cerevisiae*'s ecosystem engineering via fermentation Ecology 89, 8, 2077–2082.

Guilloux-Benatier, M., Pageault, O., et al., 2000. Lysis of yeast cells by *Oenococcus oeni* enzymes J. Ind. Microbiol. Biotechnol. 25, 4, 193–197.

Hall, J.P.J., Harrison, E., et al., 2018. Competitive species interactions constrain abiotic adaptation in a bacterial soil community Evol. Lett. 2, 6, 580–589.

Hershberg, R., 2015. Mutation—the engine of evolution: Studying mutation and its role in the evolution of bacteria Cold Spring Harb. Perspect. Biol. 7, 9, 1–13.

Iorizzo, M., Testa, B., et al., 2016. Selection and technological potential of *Lactobacillus plantarum* bacteria suitable for wine malolactic fermentation and grape aroma release LWT - Food Sci. Technol. 73, 557–566.

Jiang, J., Sumby, K.M., et al., 2018. Directed evolution of *Oenococcus oeni* strains for more efficient malolactic fermentation in a multi-stressor wine environment Food Microbiol. 73, 150–159.

Jussier, D., Morneau, A.D., et al., 2006. Effect of simultaneous inoculation with yeast and bacteria on fermentation kinetics and key wine parameters of cool-climate Chardonnay Appl. Environ. Microbiol. 72, 1, 221–227.

Kennes, C., Veiga, M.C., et al., 1991. Trophic relationships between *Saccharomyces cerevisiae* and *Lactobacillus plantarum* and their metabolism of glucose and citrate Appl. Environ. Microbiol. 57, 4, 1046–1051.

Lerm, E., Engelbrecht, L., et al., 2011. Selection and characterisation of *Oenococcus oeni* and *Lactobacillus plantarum* South African wine isolates for use as malolactic fermentation starter cultures South African J. Enol. Vitic. 32, 2, 280–295.

Lisanti, M.T., Blaiotta, G., et al., 2019. Alternative methods to SO_2 for microbiological stabilization of wine Compr. Rev. Food Sci. Food Saf. 18, 455–479.

Liu, Y., Forcisi, S., et al., 2016. New molecular evidence of wine yeast-bacteria interaction unravelled by non-targeted exometabolomic profiling Metabolomics 12, 69, 1–16.

Liu, Y., Rousseaux, S., et al., 2017. Wine microbiome: A dynamic world of microbial interactions Crit. Rev. Food Sci. Nutr. 57, 4, 856–873.

López-Malo, M., García-Rios, E., et al., 2015. Evolutionary engineering of a wine yeast strain revealed a key role of inositol and mannoprotein metabolism during low-temperature fermentation BMC Genomics 16, 537.

Malherbe, S., Bauer, F.F., et al., 2007. Understanding problem fermentations - A review South African J. Enol. Vitic. 28, 2, 169–186.

Marcobal, A.M., Sela, D.A., et al., 2008. Role of hypermutability in the evolution of the genus Oenococcus J. Bacteriol. 190, 2, 564–570.

Martino, M.E., Bayjanov, J.R., et al., 2016. Nomadic lifestyle of *Lactobacillus plantarum* revealed by comparative genomics of 54 strains isolated from different habitats Environ. Microbiol. 18, 12, 4974–4989.

McBryde, C., Gardner, J.M., et al., 2006. Generation of novel wine yeast strains by adaptive evolution Am. J. Enol. Vitic. 57, 4, 423–430.

Mendoza, S.N., Cañón, P.M., et al., 2017. Genome-scale reconstruction of the metabolic network in *Oenococcus oeni* to assess wine malolactic fermentation Front. Microbiol. 8, MAR, 534.

Nehme, N., Mathieu, F., et al., 2008. Quantitative study of interactions between *Saccharomyces cerevisiae* and *Oenococcus oeni* strains J Ind Microbiol Biotechnol 35, 7, 685–693.

Novo, M., Gonzalez, R., et al., 2014. Improved fermentation kinetics by wine yeast strains evolved under ethanol stress LWT - Food Sci. Technol. 58, 1, 166–172.

Osborne, J.P. & Edwards, C.G., 2006. Inhibition of malolactic fermentation by *Saccharomyces cerevisiae* during alcoholic fermentation under low- and high-nitrogen conditions: A study in synthetic media Aust. J. Grape Wine Res. 12, 1, 69–78.

Osmond, M.M. & de Mazancourt, C., 2013. How competition affects evolutionary rescue Philos. Trans. R. Soc. B Biol. Sci. 368, 20120085.

Passos, F. V., Fleming, H.P., et al., 2003. Effect of malic acid on the growth kinetics of *Lactobacillus plantarum* Appl. Microbiol. Biotechnol. 63, 2, 207–211.

Ponomarova, O., Gabrielli, N., et al., 2017. Yeast creates a niche for symbiotic lactic acid bacteria through nitrogen overflow Cell Syst. 5, 4, 345-357.e6.

Rizk, Z., El Rayess, Y., et al., 2016. Impact of inhibitory peptides released by *Saccharomyces cerevisiae* BDX on the malolactic fermentation performed by *Oenococcus oeni* Vitilactic F Int. J. Food Microbiol. 233, 90–96.

Rizk, Z., Rayess, Y. El, et al., 2018. Identification of multiple-derived peptides produced by *Saccharomyces cerevisiae* involved in malolactic fermentation inhibition FEMS Yeast Res. 18, 7, 80.

Rollero, S., Bloem, A., et al., 2018. Fermentation performances and aroma production of non-conventional wine yeasts are influenced by nitrogen preferences FEMS Yeast Res. 18, 5, 55.

Salvadó, Z., Arroyo-lópez, F.N., et al., 2011. Quantifying the individual effects of ethanol and temperature on the fitness advantage of *Saccharomyces cerevisiae* Food Microbiol. 28, 1155–1161.

Scott Laboratories, 2018. Cross Evolution Yeast | Scott Laboratories Available at https://scottlab.com/cross-evolution-yeast-cross.

Steinberg, B. & Ostermeier, M., 2016. Environmental changes bridge evolutionary valleys Sci. Adv. 2, e1500921.

Tadrowski, A.C., Evans, M.R., et al., 2018. Phenotypic switching can speed up microbial evolution Sci. Rep. 8, 8941.

Tilloy, V., Ortiz-Julien, A., et al., 2014. Reduction of ethanol yield and improvement of glycerol formation by adaptive evolution of the wine yeast *Saccharomyces cerevisiae* under hyperosmotic conditions Appl. Environ. Microbiol. 80, 8, 2623–2632.

Van den Bergh, B., Swings, T., et al., 2018. Experimental design, population dynamics, and diversity in microbial experimental evolution Microbiol. Mol. Biol. Rev. 82, 3, e00008-18.

Van Pijkeren, J.P. & Britton, R.A., 2012. High efficiency recombineering in lactic acid bacteria Nucleic Acids Res. 40, 10, 1–13.

Volschenk, H. & Van Vuuren, H.J.J., 2006. Malic Acid in Wine: Origin, function and metabolism during vinification S Afr J Enol Vitic 27, 2, 123–136.

Wein, T. & Dagan, T., 2019. The effect of population bottleneck size and selective regime on genetic diversity and evolvability in bacteria Genome Biol. Evol. 11, 11, 3283–3290.

Wells, A. & Osborne, J.P., 2011. Production of SO_2 and SO_2 binding compounds by *Saccharomyces cerevisiae* during the alcoholic fermentation and the impact on wine lactic acid bacteria S. Afr J Enol Vitic 32, 2, 267–279.

West, S.A., Griffin, A.S., et al., 2006. Social evolution theory for microorganisms Nat. Rev. Microbiol. 4, 8, 597–607.

Yurdugü, S. & Bozoglu, F., 2002. Studies on an inhibitor produced by lactic acid bacteria of wines on the control of malolactic fermentation Eur. Food Res. Technol. 215, 1, 38–41.

Zhao, H., Liu, L., et al., 2019. Heterologous expression of argininosuccinate synthase from *Oenococcus oeni* enhances the acid resistance of *Lactobacillus plantarum* Front. Microbiol. 10, 1–13.

Zhao, X.F., Buckling, A., et al., 2018. Specific adaptation to strong competitors can offset the negative effects of population size reductions Proc. R. Soc. B Biol. Sci. 285, 20180007.

Zhou, D., Jiang, Z., et al., 2019. CRISPR/Cas9-assisted seamless genome editing in *Lactobacillus plantarum* and its application in N-acetylglucosamine production Appl. Environ. Microbiol. 16.

Zhou, N., Bottagisi, S., et al., 2017. Yeast-bacteria competition induced new metabolic traits through large-scale genomic rearrangements in *Lachancea kluyveri* FEMS Yeast Res. 17, 6, 1–14.

Supplementary material to Chapter 3

Supplementary tables

Table S3.1 Synthetic grape must (SGM) composition based on Rollero *et al.* (2018). %N represents the amount of nitrogen contributed by each of the amino acids to the must and it is shown here as a percentage of total nitrogen.

Carbon source	g/l	Vitamins	µg/l	Nitrogen sources	mg/l
Glucose	115	Myo-Inositol	100	NH_4Cl	460
Fructose	115	Pyridoxine-HCl	2	Ala	145.3
Acids		Nicotinic acid	2	Arg	374.4
KH Tartrate	2.5	Ca Pantothenate	1	Asp	44.5
L-Malic acid	3	Thiamine-HCl	0.5	Gln	505.3
Citric acid	0.2	PABA.K	0.2	Cys	13.09
Salts	**g/l**	Riboflavin	0.2	Glu	120.4
K2HPO4	1.14	Biotin	0.125	Gly	18.3
MgSO4.7H2O	1.23	Folic Acid	0.2	His	32.7
CaCl2.2H2O	0.44			Ile	32.7
Trace elements	**µg/l**	**Lipids/anaerobic factors**	**Mg/l**	Leu	48.4
MnCl2.4H2O	200	Ergosterol	15	Lys	17
ZnCl2	135	Oleic acid	5	Met	31.4
FeCl2	30	Tween80	0.5 ml	Phe	37.9
CuCl2	15			Pro	612.6
H3BO3	5			Ser	78.5
Co(NO3)2.6H2O	30			Thr	75.9
NaMoO4.2H2O	25			Trp	179.3
KIO3	10			Tyr	18.3
				Val	44.5

Chapter 4

Comparative genomics of yeast-driven evolved isolates of *Lactobacillus plantarum*

Chapter 4: Comparative genomics of yeast-driven evolved isolates of *Lactobacillus plantarum*

4.1. Abstract

The lactic acid bacteria taxonomic group presents a large genome diversity, and *Lactobacillus* species harbour some of the largest genomes within this group. These species occupy a wide spectrum of ecological niches, each niche characterised by many interactions among the different species found in this niche. Such interactions include competition for nutrients and growth inhibition due to the production of inhibitory metabolites. However, specific biotic selection pressures linked to inter-species interactions have rarely been investigated, although it can be assumed that such interactions have largely shaped the genome and gene function of each species. Here I describe the genome-wide differences between a parental strain and several evolved strains of *Lactobacillus plantarum* after a short-term (≈50 and 100 generations) directed evolution experiment using biotic selection pressure. Two strains of *Saccharomyces cerevisiae* were used as selective drivers through competitive interaction. Sixteen evolved strains, 8 each for the two *S. cerevisiae* strains, were sequenced *de novo*. The data show that phenotypic changes appear at least in part to be driver-yeast specific, but variant analysis suggests that mutations in specific genes were acquired independently. KEGG classification revealed that peptidoglycan biosynthesis and degradation, nucleic acid processing, and carbohydrate transport and metabolism- related genes showed the highest mutation frequency (>50%) in the majority of the isolates. The use of directed evolution in this study shows that the genomic differences do not fall into clearly separate groups but are specific to each evolved isolate. This suggests that biotic factor-driven evolution prefers genomic diversity within populations, while abiotic factor-driven evolution results in specific phenotypes and genotypes.

4.2. Introduction

Lactic acid bacteria (LAB) are Gram-positive, catalase negative, lactic acid producing and facultative anaerobic microorganisms, found in a wide spectrum of ecological niches (Melgar-Lalanne *et al.*, 2012). The ability of these organisms to occupy such diverse niches is due to their high genome plasticity and diversity (Siezen *et al.*, 2010; Wiedenbeck & Cohan, 2011; Dragosits & Mattanovich, 2013). Comparative genomics of complete LAB genome sequences showed that species in this group have genome sizes ranging between 1.3 Mb to 4.91 Mb and a low GC content (Wassenaar & Lukjancenko, 2014; Sun *et al.*, 2015). The species *Lactobacillus plantarum*, while showing variable genome sizes, is situated at the higher end of this scale, with the smallest reported genome size at the time of writing at 3,175,846 bp and the largest at 3,615,168 bp (Suryavanshi *et al.*, 2017; Jiang *et al.*, 2018). Most completely sequenced *L. plantarum* strains are

usually between 3.2 Mb and 3.5 Mb (Kant *et al.*, 2011; Siezen *et al.*, 2012; Wassenaar & Lukjancenko, 2014; Golneshin *et al.*, 2015; Liu, Wang, *et al.*, 2015; Huang *et al.*, 2018)

L. plantarum species occupy a wide spectrum of species-rich ecological niches such as the human and animal guts, dairy, food and beverages (juice and wine) (Saguir *et al.*, 2008; Lerm *et al.*, 2011; Vrancken *et al.*, 2011; Stefanovic *et al.*, 2017; Song *et al.*, 2018). These environments are characterised by numerous evolutionary relevant interactions between species such as competition for, or reciprocal exchange of nutrients (Bayrock & Ingledew, 2004; Ponomarova *et al.*, 2017; du Toit, 2018; Bechtner *et al.*, 2019). Such interactions between species can be assumed to have largely shaped the genome and gene function of each interacting species. However, the biotic selection pressures directly linked to inter-species interactions in LAB has not yet been reported. However, there are a few studies that link wine yeast (and bacteria) interactions to evolutionary change and adaptation. However, in these studies only yeast species were investigated. The data showed that competitive conditions between yeast and bacteria, can result in stronger selection for particular phenotypes in yeast (Zhou, Bottagisi, *et al.*, 2017; Zhou, Swamy, *et al.*, 2017; Morrison-Whittle *et al.*, 2018; Zhou *et al.*, 2018).

A small number of comparative genomic studies among LAB species have primarily evaluated the genomic differences that have occurred over evolutionary history (Makarova *et al.*, 2006; O'Sullivan *et al.*, 2009; Kant *et al.*, 2011; Borneman *et al.*, 2012; Sun *et al.*, 2015). No studies have evaluated whole genome variation for long (or short) -term evolution of LAB. Within the eubacteria, however, *Escherichia coli* has been the subject of several such studies which revealed that extensive genomic differences with significant phenotypic consequences can occur within a short time period, for example, under radiation and carbohydrate starvation stress (Herring *et al.*, 2006; Harris *et al.*, 2009).

It is known that, aside from abiotic selection pressures, adaptation is also driven by the response to dynamically changing environments created through continuous metabolic exchanges between the species in an environment (Helliwell *et al.*, 2018). In systems where the bulk of interaction-relevant metabolites is likely the product of a single dominating species, as is the case in wine fermentations (i.e. *S. cerevisiae*), other species, including wine LAB, must adapt by adjusting growth, biomass, and metabolic function under the selective pressures created by the dominant species (Jussier *et al.*, 2006; Nehme *et al.*, 2008; Muñoz *et al.*, 2014; Ciani *et al.*, 2016). In Chapter 3 of this dissertation, I applied this theory by designing directed evolution (DE) experiments where yeast strains were used as biotic selective drivers. These experiments were performed under yeast-specific conditions (i.e. the only difference between the treatments was the yeast strain used as the biotic driver -based selection pressure). This was done to determine

whether yeast-specific phenotypes and genotypes could be induced in the evolving bacteria, and what these adaptations would 'look like' on a genomic level. Therefore, here I aimed to determine the genetic targets that may have played a role in yeast strain-specific phenotypes, and have facilitated the adaptation of these isolates to the biotic selective pressure. In addition, this work applied *in silico* genome analyses to predict the impact of mutations on protein function and structure.

4.3. Materials and Methods

4.3.1. Bacterial cultures

Previously, a strain of *L. plantarum* IWBT B063 was evolved for 50 and 100 generations in co-culture with *S. cerevisiae* strains EC1118 and Cross Evolution in synthetic grape must (SGM) (see Chapter 3). Prior to inoculation in the SGM, the yeast strains were grown in 5 ml YPD broth and incubated overnight, with shaking, at 30ºC. The bacterial isolates were pre-cultured in MRS broth before inoculation in SGM medium. All bacterial isolates used were grown statically at 30ºC (Chapter 3). DE experiments were designed to allow for bacterial adaptation while maintaining the yeast in the same fixed parental 'evolutionary state'; hence the yeast was discarded and a fresh overnight mother culture was reintroduced with every round of fermentation (refer to Chapter 3). The fermentations were conducted at 20ºC. After 50 generations of DE experiments and successive screening, 2 isolates evolved with Cross Evolution (C5 and C26: G50C) and 3 isolates evolved with EC1118 (E5, E22 and E26: G50E) were selected for whole-genome sequencing.

To elucidate whether the phenotypic changes observed after 50 generations would also be present after 100 generations, 60 randomly selected isolates (30 each from EC1118-DE and Cross Evolution-DE) were screened in a 96-well microtitre plate during the screening experiments to indentify isolates with superior growth compared to the parent strain (Figure S4.1). Each microtitre plate was divided into blocks of 9 wells, with each replicate appearing once per block to minimize the effect of replicate position on the plate. Yeast-fermented SGM was used because previous data showed that the evolved isolates alter their behaviour based on the dynamic interaction of the yeast with its environment which affects the bacteria (Chapter 3). Each colony isolate was grown overnight in MRS broth at 30ºC and then acclimatised for 24 hours in 1 ml unfermented SGM at 20ºC prior to inoculation in 200 µl yeast-fermented SGM. The PowerWave™ Microplate Scanning Spectrophotometer (BioTek Instruments Inc, Vermon, USA) was used to measure optical density (OD) over 24 hours, with readings taken every hour. Isolates that showed improved biomass formation compared to the parent were selected and their ability to degrade L-malic acid was investigated in 15 ml co-culture fermentations. Isolates with improved MLF and growth compared to the parent strain were then selected for whole genome sequencing. These isolates are C33,

C35, C40, C43, C49 and C54 (G100C) which were evolved with Cross evolution, and E33, E42, E51, E59 and E60 (G100E) evolved in the presence of EC1118. The G100 isolates became available at the sequencing stage of the study and were not characterised for phenotypes beyond the initial screening process in 96-well plates (Figure S4.1) and microfermentations (Figure S4.2). Glycerol stocks (20%) were prepared for each isolate including the parent strain.

4.3.2. DNA extraction

Genomic DNA of the isolates and parent strain was prepared according to Lewington *et al.* (1987): a single colony of each of the parent and evolved isolates was inoculated in 5 ml MRS (Man, Regosa, and Sharp) broth and incubated overnight at 30°C. The 5 ml overnight cultures were harvested by centrifugation at 12000 rpm for 2 minutes, and the pellet washed with TE buffer. The pellet was resuspended in 435 µl solution A (1M Tris-HCl [pH 8.0], 1M EDTA [pH 8.0], 1M NaCl, 0.019 g/L sucrose) and 65 µl lysozyme, then incubated at 37°C for 90 minutes with inversion. After this period the homogenate was digested with 50 µl 10% sodium dodecyl sulphate (SDS) and 25 µl of Proteinase K (25 mg/ml), and then gently mixed by inversion several times. The tubes were subsequently incubated at 55°C for 60 minutes with mixing by gentle inversion. The genomic DNA was extracted twice in phenol: chloroform (1:1) and precipitated in 100% ice-cold ethanol and 0.3M sodium acetate at -20°C overnight. After centrifugation at 10000 rpm for 10 minutes at 4°C, the DNA pellet was washed with 70% ice cold ethanol and dissolved in 40 µl TE buffer (pH 8.0). RNAse was added (10 mg/ml, 1 µl RNAse/20 µl dissolved DNA) and the mixture was incubated at 37°C for 30 minutes. The clean-up and precipitation steps were repeated, and the DNA was subjected to quality control by 1% (v/v) agarose gel electrophoresis and quantified using the NanoDrop® ND- 1000 spectrophotometer (NanoDrop Technologies, Inc., Wilmington, USA).

4.3.3. NGS sequencing and genome assembly

The parent and each of the evolved isolates were sequenced at the Central Analytical Facilities: Next Generation Sequencing Unit (Stellenbosch University, South Africa). Parallel sequencing was performed on the Ion Proton™ System using the Ion S5™ Ion PI™ Hi-Q™ Sequencing 200 Kit according to the protocol (ThermoFisher Scientific, Waltham, MA, USA). The library preparation was performed from 100 ng gDNA using the Ion Plus Fragment Library Kit according to the protocol (ThermoFischer Scientific). The gDNA was fragmented on the Covaris S2 focused ultrasonicator (Covaris, Inc.; Woburn, MA, USA) using a 10% duty cycle, with 5% intensity, 100 cycles/burst and 40 sec treatment time. The fragmented gDNA was end-repaired in preparation for blunt-end ligation to Ion Xpress™ Barcode Adapters (ThermoFischer Scientific). The adapter-ligated, barcoded library was purified with Agencourt™ AMPure™ XP reagent (Beckmann Coulter). To select library fragments with 200 bp insert lengths, a 2% gel cassette was used with marker L on the PippenPrep (Sage Science Inc., Beverly, MA, USA) with a tight selection profile

for 270 bp fragments. Purified, size-selected library fragments were quantified using the Ion Universal Library Quantitation Kit (ThermoFischer Scientific). qPCR amplification was performed using the StepOnePlus™ Real-time PCR system (ThermoFisher Scientific).

Libraries were diluted to a target concentration of 100 pM. The diluted, barcoded, libraries were combined in equimolar amounts for template preparation using the Ion PI™ Hi-Q™ OT2 200 Kit (ThermoFisher Scientific). Seven microliters of the pooled library were diluted with 93 µl nuclease-free water for emulsion amplification using the Ion OneTouch™ 2 System (ThermoFischer Scientific) and subsequent template enrichment with MyOne Streptavidin C1 beads (ThermoFischer Scientific) on the Ion OneTouch™ ES (ThermoFischer Scientific); according to the manufacturer's protocol. Enriched, ion sphere particles were loaded onto an *Ion PI™ Chip* (v3). The Ion S5™ *Ion PI™ Hi-Q™ Sequencing 200 Kit* was used for sequencing according to the protocol. Flow space calibration and BaseCaller analyses were performed using default analysis parameters in the Torrent Suite Version 5.0.4 Software.

4.3.3.1. Sequence data assembly and annotation

All data analysis was performed on the Galaxy software framework, which is an open-source, web-based platform (Afgan *et al.*, 2018). The platform is equipped with various independent bioinformatics tools/modules that assist with sequence data analysis and characterization. Initially, sliding window trimming was performed on the reads to remove contaminant sequences such as primer sequences with a minimum quality score below 20 using Trimmomatic (Version 0.36.0). A quality control step was applied using *FastQC Read Quality Reports* and showed that the number of reads ranges from 3.8 million to 10 million with a sequencing coverage range of 205 to 509 for all isolates and the parent strain (Table 4.1).

In this work the parent strain is the reference genome to which all the evolved isolates were compared. Therefore, I used a reference-guided assembly method which bears similarities to the methods used by Schneeberger *et al.* (2011) and Lischer & Shimizu (2017). However, I mapped the reads against both *L. plantarum* LB2_1 (unpublished) and *L. plantarum* WCFS1 (Kleerebezem *et al.*, 2003; Siezen *et al.*, 2012) with Bowtie2 (Galaxy version 2.3.4.2) (Figure 4.1). These two strains were selected as references because WCFS1 is a well curated reference genome (GenBank accession no. AL935263), however this strain shows an overall alignment rate of 62.76% against the parent genome. Therefore, a second annotated genome was required, and LB2_1, which showed an overall alignment rate of 80.27% when aligned to the parent, was selected to complement WCFS1. Thereafter, a *de novo* consensus assembly of either aligned or unaligned reads was performed with SPAdes genome assembler (Galaxy version 1.0). The resulting contigs were merged using SPAdes to build a consensus parent genome which was used

as a reference genome for each of the evolved isolates (Figure 4.1). Two reference genomes were used due to the reference bias that is known to occur during assembly (Schneeberger *et al.*, 2009).

Table 4.1 Number of sequence reads and coverage for the parent and each of the evolved isolates.

Generations	Sample ID	Number of reads	Sequencing coverage
GEN0	Parent	6892053	354
GEN50	C5	8003988	437
	C26	6628856	364
	E5	5370954	295
	E22	5362501	296
	E26	3820882	205
GEN100	C33	6695452	338
	C35	6999436	359
	C40	6617488	338
	C43	8407678	427
	C49	7712994	387
	C54	7444805	376
	E33	7630887	374
	E42	7782842	390
	E51	6654963	338
	E59	10088289	509
	E60	6784845	336

A QUAST analysis was performed to generate summary statistics for the consensus genome assembly (Gurevich *et al.*, 2013). Using the consensus assembly technique reduced the number of contigs from 257 to 192 in the parent strain. Mauve multiple genome alignment (version 2.1.0a1) was used to reorder the contigs of the parent strain to that of WCFS1 which has a completely sequenced and annotated genome. The sequence alignment file obtained from Mauve was used to annotate the consensus genome using Prokka (Version 1.11.0) and default parameters were selected, with the exception of selecting 'Bacteria' in the 'Kingdom' tab. Prokka generated 11 files: a ".log", ".txt", ".err", ".tbl", ".fsa", ".sqn", ".fnn", ".faa", ".fna", ".gbk" and ".gff". The latter 5 files have gene DNA and amino acid sequences and the annotation of these genes and their positions on the genome, therefore these and the .tbl file (COG classification information) were used for downstream analyses.

The reads for each of the evolved isolates were mapped to the consensus parent genome obtained from Prokka (.fna) using Bowtie2 (Figure 4.1). Thereafter, the aligned reads were assembled using SPAdes (with default settings). Genome assembly quality was assessed using QUAST. Each of the assembled genomes were annotated using Prokka and the parent strain was used as the reference genome. Interactive Genome Viewer (IGV) was used to browse the genomes of all the isolates and strains used in this study (Robinson *et al.*, 2011).

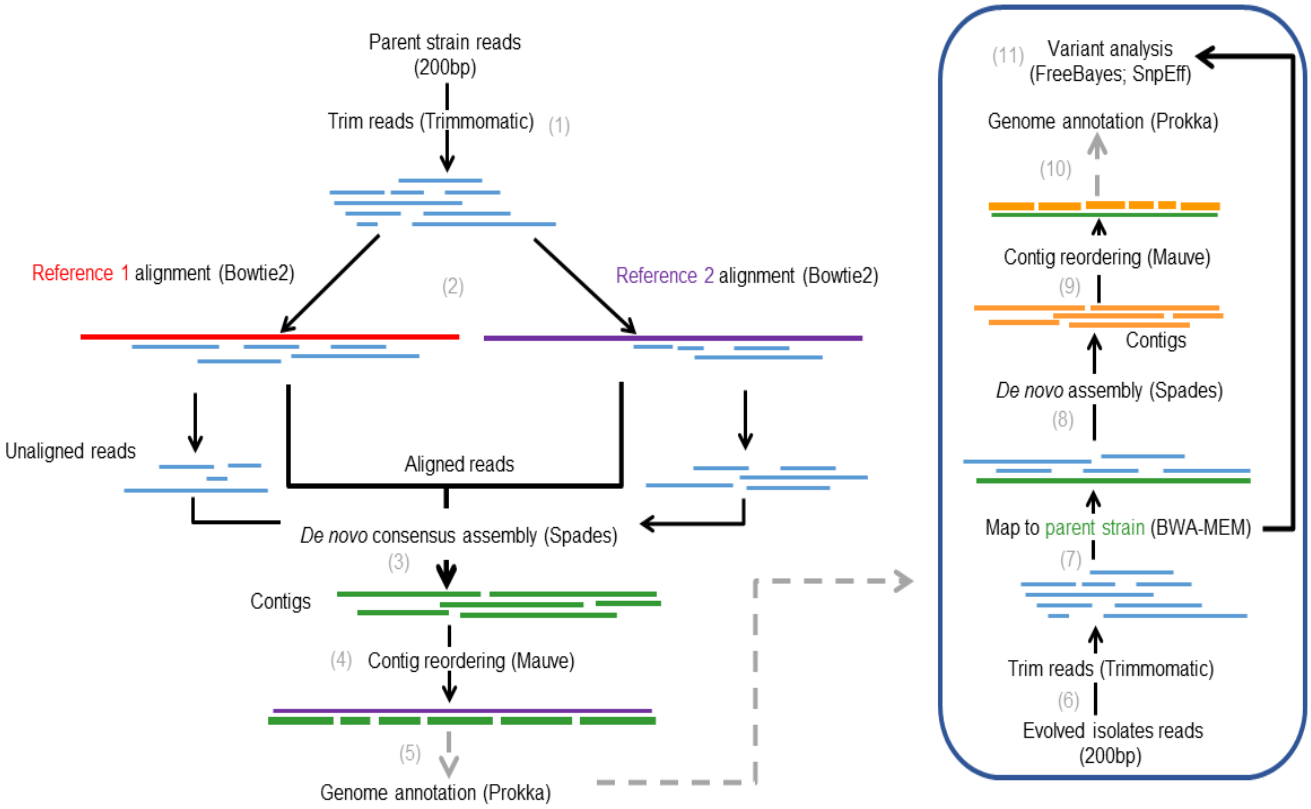

Figure 4.1 Depiction of reference-guided *de novo* assembly. The trimmed reads (1) of the parent strain were mapped against two references separately (2). First the aligned reads were *de novo* assembled, then the unaligned reads were *de novo* assembled. These assemblies were merged to form a consensus genome assembly (3). The resulting contigs were then reordered using Mauve (4). After a quality check the contigs were annotated using reference 2 (in this case *L. plantarum WCFS1*) (5). The reads from each of the evolved isolates were trimmed (6) and mapped to the annotated parent strain using BWA-MEM (7). All aligned reads were then *de novo* assembled (8) and the contigs were reordered based on the parent strain (9). After the quality check of the assembly, the contigs were used to annotate the genome (10) and finally call variants (11). Reference 1 (*L. plantarum* LB2_1) is shown in red, reference 2 is purple. All reads are shown in blue. Contigs are shown in green (parent) and orange (evolved isolates).

4.3.3.2. Identification of single nucleotide polymorphisms and indels

To summarise, a range of phenotypes were observed in each of the evolved isolates with regards to growth and malolactic fermentation capabilities, with the isolates evolved from EC1118 showing generally better growth and MLF than isolates evolved from Cross Evolution compared to the parent strain in co-culture with either yeast strain used (see Chapter 3). It is not clear what molecular mechanisms drove these observed "yeast-specific" phenotypes. Therefore, whole genome sequencing (WGS) was carried out on each of the selected evolved isolates to try and elucidate the specific targets of DE as per our experimental designed (i.e. using yeast strains as selective drivers of the evolution of bacteria). Firstly, BWA-MEM (Galaxy version 0.7.17.1) (Li & Durbin, 2009) was used to map the trimmed reads (.fastqsanger) for each isolate to the parent strain (.fna) (Figure 4.1). FreeBayes (Galaxy version 1.1.046-0) was used to call variants (Garrison

& Marth, 2012). The default settings were used except for a few minor changes: the parent genome (.fna) was used as the reference sequence, the variant calling was not limited to a set of regions, the expected mutation rate/pairwise nucleotide diversity was set at 0.001, ploidy was set to 1, left alignment of indels was turned on to increase variant call accuracy and to normalise variants (Tan *et al.* 2015). The resulting VCF data files were used in SnpEff (Galaxy version 4.3+T.galaxy1) to annotate variants and predict the effect of each on the corresponding sequence (coding and noncoding sequences) (Cingolani *et al.*, 2012). To use SnpEff I first had to build a SnpEff database. Due to the genomes having a large number of scaffolds (>100), I used the GFF method to build the database (Cingolani *et al.*, 2012). Here I used the parent .gff file from Prokka as well as the .fna file. These files contain coordinates of various features such as gene locations, as well as gene sequences. The built SnpEff database on Galaxy was subsequently used to run SnpEff and call variants using default settings.

Thereafter, the variants were first filtered based on the QUAL score (the probability score that a variant is present on particular allele locus) assigned by SnpEff. All variants with a QUAL value <20 were filtered out (Song *et al.*, 2016). Variants were then filtered based on a DP (sequence coverage/depth at locus) ≥15 as recommended by Song *et al.* (2016). Variants found in prophages and mobile genetic elements (such as insertion sequences and plasmids) were not considered for further analyses as it cannot be conclusively said that changes observed in these sequences are due to the interaction of the yeast and bacteria. Moreover, identical SNPs found in all isolates (selected from 6 independent evolutionary lines) were disregarded because it is more likely that they occurred before the start of the DE experiments, than it is that they are the result of the DE experiments. To identify key mutations, the mutation frequency of each mutation was calculated similarly to Pfeifer *et al.* (2017), using the AO/(AO+RO)*100 equation, where AO is the count of the observed allele variant in the evolved isolates (ALT) and RO is the count of the observed allele in the parent (Garrison & Marth, 2012). The mutations with the highest frequency were considered significant. Additionally, genes that showed the highest number of mutations were considered significant targets of DE. It is important to note that this study focused mainly on nonsynonymous mutations and frameshift mutations which are more likely to cause obvious changes in the encoded proteins.

4.3.4. *In silico* investigation of the probable impact of mutations on protein function

Probable genetic targets of DE were selected based on the number of mutations in a particular gene and also the frequency of these mutations across all evolved isolates. Therefore, the amino acid sequences of each of these genes (both parent and evolved isolate genes) were used in Phyre2, an online platform which detects sequence homology and predicts and analyses protein structures, functions and possible mutations (Kelley *et al.*, 2015). This was achieved by building 3D

models of target proteins and predicting ligand sites and variations in amino acid sequences based on comparisons with other published sequences (Kelley *et al.*, 2015). Additionally, ProFunc was used for the in-depth analysis of the predicted protein structures, especially in the event of gene targets that encode hypothetical proteins of unknown function (Laskowski *et al.*, 2005). Furthermore, Pfam (El-Gebali *et al.*, 2019) and Prosite (de Castro *et al.*, 2006) were used to detect conserved domains in each of the affected genes. Based on the results of these web-based services, *in silico* predictions were made as to the impact of the mutations on possible protein function and the evolved isolates.

4.4. Results and Discussion

In Chapter 3 of this dissertation I show that after only 50 generations of DE there were differences in the L-malic acid degradation capabilities of the isolates evolved with Cross evolution (G50C) and EC1118 (G50E) as selective drivers. Interestingly, the former all showed significantly poorer L-malic acid degradation and growth, while the latter showed significant improvement of both L-malic acid degradation and growth when compared to the parent strain (see Chapter 3). Here, 30 additional isolates from both Cross Evolution and EC1118 experiments after 100 generations of DE were selected. Eight and seven out of 30 isolates evolved with Cross Evolution and EC1118, respectively, showed improved biomass in comparison to the parent strain (Figure S4.1). Additionally, all these isolates completed MLF in SGM (Figure S4.2). Specifically, the 8 isolates evolved with Cross Evolution completely degraded 3 g/l L-malic acid in four to six days, whereas after 50 generation only four isolates could degrade L-malic acid within a 4- to 14-day range (Chapter 3). The isolates evolved with EC1118 degraded 3 g/l L-malic acid between 2 and 4 days, whereas isolates selected after 50 generations took 4 to 6 days (Chapter 3). These data show that L-malic acid degradation was improved for isolates evolved with either Cross Evolution or EC1118 over time and clearly show that the isolates evolved in the presence of EC1118 have significant improvement compared to those evolved in the presence of Cross Evolution, suggesting that the evolutionary pressure exerted by the two yeast strains leads to differing yeast strain-specific adaptations in the bacterial population. To gain a better understanding of the underlying molecular mechanisms which may have influenced the observed phenotypes, the genomes of each of the isolates after 50 and 100 generations of DE were sequenced, assembled, annotated and evaluated for variants and compared to the parent strain.

4.4.1. Global gene classification and function prediction

To start of this section describes the parent genome and how they relate to other *L, plantarum* strains and to the evolved isolates. Gene annotation was performed using Prokka: Prokka uses Prodigal to find genes and perform functional annotations using various databases including

NCBI's Clusters of Orthologous Groups (COGs) (Tatusov, 2000). To engage with a broad overview of the parent genome and that of the evolved isolates I compared all predicted protein coding genes and found that 35.28-36.25% had COG IDs (Table 4.4). In Figure 4.2 below, only the classification of the genes in the parent strain are shown because the overall differences between this strain and the evolved isolates were negligible. The majority (63.95%) of the predicted protein coding genes were poorly characterised due to the large number of 'hypothetical proteins' (Figure 4.2A). According to the COG classification, a higher proportion of the predicted genes were assigned to *amino acid transport and metabolism* (13.98%), *carbohydrate transport and metabolism* (11.66%), *ribosomal structure and biogenesis* (11.00%), and *transcription* (9.93%) (Figure 4.2B) which is in line with what was reported for *L. plantarum* WCFS1 (Kleerebezem *et al.*, 2003; Siezen *et al.*, 2012), *L. plantarum* E2C2 and E2C5 (Suryavanshi *et al.*, 2017), and over 20 *L. plantarum* strains obtained from NCBI (Evanovich *et al.*, 2019).

4.4.1.1. Metabolism

It is well known that *L. plantarum* species inhabit nutrient-rich environments and have adapted to these environments because they possess a plethora of enzymes that degrade complex compounds (Kleerebezem *et al.*, 2003). Fifty-three percent of the genes assigned to COG, when excluding those with poor characterisation, are involved in the metabolism and transport of various compounds (Figure 4.2B). To begin with there are 66 genes coding for amino acid and peptide transporters of which 4 (*ilvE*, *livH*, *livF* and *brnQ*) are specific for branched-chain amino acids (Table S4.1). This data is in support of previous reports on branched-chain amino acid auxotrophies in *L. plantarum* strains (Kleerebezem *et al.*, 2003; Botma, 2018; du Toit, 2018). Such a large number of amino acid transporters shows that *L. plantarum* utilizes nutrients found in its environment. Recently, it was shown that in a rich, multi-species environment with both yeast and LAB, *S. cerevisiae* strains tend to release amino acids to the environment to reduce excess nitrogen that might be potentially toxic to the cells, limiting growth(Hess *et al.*, 2006). This consequently provides the LAB with essential amino acids that they readily take up and use for growth (Ponomarova *et al.*, 2017; Bechtner *et al.*, 2019). These results are interesting because although many *L. plantarum* strains have complete sets of amino acid biosynthesis pathway genes (Kleerebezem *et al.*, 2003; Liu *et al.*, 2015), they are not necessarily able to synthesise all amino acids (Christiansen *et al.*, 2008). In this study it was observed that the isolates and parent strain used here have the complete set of genes for the biosynthesis of Ala, Arg, Asn, Asp, Cys, Glu, Gln, His, Met, Pro, Ser and Thr (data not shown). Data suggests that the inability of a bacterium to synthesize amino acids, although it has a full set of biosynthesis genes may be due to the amino acid biosynthetic pathway overlapping with another essential process for adaptation to protein-rich environments, or due to the synthesis of genes being energetically more expensive than assimilating them from the environment (Price *et al.*, 2018).

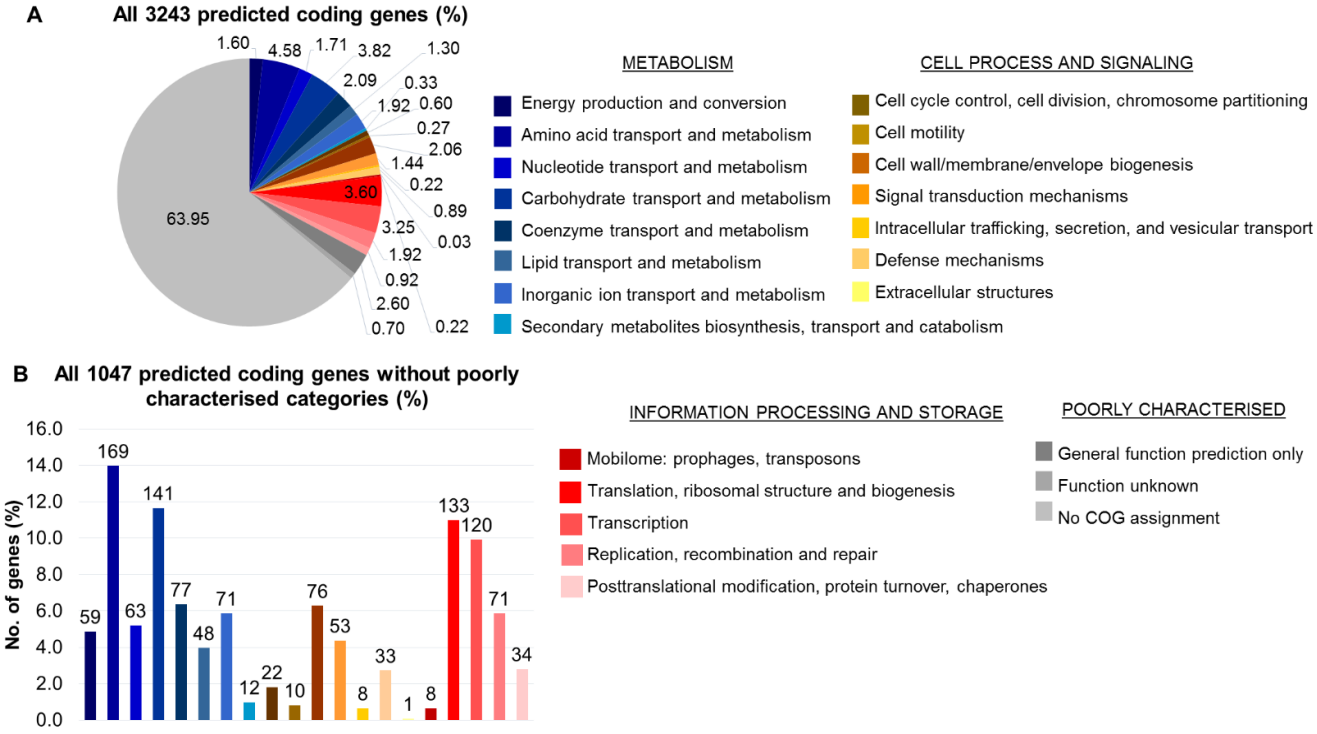

Figure 4.2 Classification of the parent strain (and evolved isolates) predicted protein genes into COG categories. A) All predicted protein genes including those with poor characterisation. B) Predicted protein genes categories excluding the poorly characterised categories. The data labels on top of each bar show the actual number of genes in the assigned category. The different colours represent different functional categories.

Approximately 200 genes were categorized under *Energy production and conversion* as well as *Carbohydrate transport and metabolism,* which supports previous reports that *L. plantarum* harbours genes involved in sugar metabolism and transport, as well as energy production which are found in a region known as 'The Lifestyle Island' (Molenaar *et al.*, 2005). According to Kleerebezem *et al.* (2003) this region is close to the origin of replication and the genes found in this region are suggested to have a high plasticity and may be important for adaptation to different environments (Molenaar *et al.*, 2005; Evanovich *et al.*, 2019). This explains the wide range of habitats that *L. plantarum* occupies, including the complex wine environment. In support of this, the parent strain as well as all the evolved isolates have 24 complete phosphotransferase (PTS) enzyme II system genes, but only isolates C5, E5, E33, E51, E59 and E60 have an additional *treP* gene encoding a PTS system trehalose-specific EIIBC component (Table S4.1). These results are comparable to those of *L. plantarum* WCFS1 which has 25 complete PTS systems (Kleerebezem *et al.*, 2003). In principle there were no other bacterial species in our experiments, therefore, it is unlikely that these isolates acquired the *treP* gene.

Normally, genome editing and gap filling by PCR are performed to ensure that a complete genome is attained (Schneeberger *et al.*, 2009; Lischer & Shimizu, 2017), however, that is beyond the scope of this work. With that in mind, the *treP* gene must be present in the parent as well, but was not annotated. The genes of the PTS enzyme II system are involved in the transport of sugars such as glucose, fructose, mannose, cellobiose, ascorbate and galactose (Table S4.1). Some of these hexoses and, to a lesser extent, pentoses (arabinose, xylose, ribose and rhamnose) and disaccharides (maltose, raffinose and trehalose) are found in grape must and wine (Unden & Zaunmüller, 2009). Therefore, the presence of PTS enzymes specific to these carbohydrates shows the ability of the parent strain (and the evolved isolates) to degrade these compounds in grape must and wine.

Stress tolerance is an important feature of microorganisms inhabiting highly dynamic environments. There are several important mechanisms that facilitate and control stress responses in LAB and these include, but are not limited to altering carbohydrate transport and fermentation pathways (Koponen *et al.*, 2012), decreasing L-lactic acid production (Heunis *et al.*, 2014) and inducing malolactic fermentation (especially in low pH, high sugar environments) (Filannino *et al.*, 2014). Several genes that play a role in each of these stress response mechanisms were identified in each of the evolved isolates and the parent strain (Table 4.2): the *ptsH* gene which encodes the phosphocarrier protein HPr and is involved in osmotic stress response, the *ccpA* gene which encodes a global transcriptional regulator involved in regulating the catabolism of carbon sources with the exception of glucose in glucose-rich media (Zhai *et al.*, 2014; Andreevskaya *et al.*, 2016; Choe *et al.*, 2017; Adu *et al.*, 2018), 8 genes encoding the F0F1 ATPase which regulates intracellular pH, 8 sodium-proton antiporters and 2 alkaline shock proteins which have all been implicated in acid, hydrogen peroxide and ethanol stress tolerance (Konings *et al.*, 1997; Mariela *et al.*, 2007; Parente *et al.*, 2010). Importantly, the malic enzyme gene (*mleS*) and the malate transporter gene *mleP* which are involved in L-malic acid decarboxylation to L-lactic acid were also identified. This deacidification reaction enhances the survival of *L. plantarum* in acidic environments such as wine (Konings *et al.*, 1997; Kleerebezem *et al.*, 2003; Spano & Massa, 2006; Liu *et al.*, 2015; Jia, Zhang, *et al.*, 2017).

Table 4.2 Genes implicated in stress tolerance in various LAB which have been found in the parent strain and each of the evolved isolates in this study.

Gene	Product	Stress	Reference
asp1	Alkaline shock protein 23	H_2O_2	(Mariela *et al.*, 2007)
asp2	Alkaline shock protein 23	H_2O_2	
ccpA	Catabolite control protein A	Multiple stresses	(Dressaire *et al.*, 2011; Zotta *et al.*, 2012; de Jong *et al.*, 2013)
citC	[Citrate [pro-3S]-lyase] ligase	Ethanol	(van Bokhorst-van de Veen *et al.*, 2011)
citD	Citrate lyase acyl carrier protein	Ethanol	
citE	Citrate lyase subunit beta	Ethanol	
citF	Citrate lyase alpha chain	Ethanol	
clpE	ATP-dependent Clp protease ATP-binding subunit ClpE	Oxidative, Acid	(Koponen *et al.*, 2012; Golomb *et al.*, 2016)
clpP	ATP-dependent Clp protease proteolytic subunit	Heat, Acid	(Ricciardi *et al.*, 2012; Golomb *et al.*, 2016)
clpX	ATP-dependent Clp protease ATP-binding subunit ClpX	Heat	(Ricciardi *et al.*, 2012)
cspL	Cold shock protein 2	Cold	(Chu-Ky *et al.*, 2005)
cspLA	Cold shock-like protein CspLA	Cold	
csp	Cold shock protein 1	Cold	
ctsR	Transcriptional regulator CtsR	Multiple stresses	(Ricciardi *et al.*, 2012; Golomb *et al.*, 2016; Huang *et al.*, 2016)
dnaK	Chaperone protein DnaK	Heat	(de Jong *et al.*, 2013)
dps	DNA protection during starvation protein	Ethanol, H_2O_2	(van Bokhorst-van de Veen *et al.*, 2011)
ftsH	ATP-dependent zinc metalloprotease FtsH	Multiple stresses	(Ricciardi *et al.*, 2012; Betteridge *et al.*, 2018)
groEL	60 kDa chaperonin	Ethanol, Heat, Acid	(Bastard *et al.*, 2016; Adu *et al.*, 2018)
groES	10 kDa chaperonin	Ethanol	(Adu *et al.*, 2018; Betteridge *et al.*, 2018)
grpE	heat shock protein GrpE	Ethanol, Heat	
hrcA	Heat-inducible transcription repressor HrcA	Multiple stresses	(Ricciardi *et al.*, 2012; Guidone *et al.*, 2015)
ptsH	Phosphocarrier protein HPr	Multiple stresses	(Andreevskaya *et al.*, 2016; Adu *et al.*, 2018)
L_02752	membrane-bound protease of the CAAX family	Ethanol	(van Bokhorst-van de Veen *et al.*, 2011)
nox	NADH oxidase	Oxidative	(Andreevskaya *et al.*, 2016)
npr	NADH peroxidase	Oxidative	(McLeod *et al.*, 2011)
pox5	Pyruvate oxidase	Oxidative, Heat	(Zotta *et al.*, 2012; Adu *et al.*, 2018)
trxB	Thioredoxin reductase	Oxidative	(Serata *et al.*, 2012)
acpP	Acyl carrier protein	Acid, osmotic	(McLeod *et al.*, 2011; Huang *et al.*, 2016)
plsX	Phosphate acyltransferase	Acid, Ethanol, Heat	(Huang *et al.*, 2016; Adu *et al.*, 2018)
fabZ	3-hydroxyacyl-[acyl-carrier-protein] dehydratase FabZ	Acid, Ethanol	(van Bokhorst-van de Veen *et al.*, 2011)
fabG	3-oxoacyl-[acyl-carrier-protein] reductase	Acid, Ethanol, Heat	(Huang *et al.*, 2016; Adu *et al.*, 2018)
fabF	3-oxoacyl-[acyl-carrier-protein] synthase 2	Acid, Ethanol	(van Bokhorst-van de Veen *et al.*, 2011)

Lastly, I identified 13 putative ABC transporters and several other sugar transporters (Table S4.1) which may be involved in carbohydrate transport (Zeng *et al.*, 2017) as well as the transport of other compounds such as metal ions, secondary metabolites, peptides and amino acids (Wilkens, 2015). ABC transporters are not well characterised in LAB, but their functions can be inferred from protein homology. For example, putative ABC transporters *yheS* and *ybiT* (Table S4.1) were shown to play a role in cold-stress protection, while *yheS* and *yknY* also protected *E. coli* against macrolide antibiotics (Phadtare & Inouye, 2004; Yamada *et al.*, 2012; Murina *et al.*, 2019). In wine LAB, it is most likely that these genes play a role in cold stress tolerance as MLF in some wines is conducted at temperatures ≤20ºC (Jussier & Ordun, 2006; Piao *et al.*, 2015). However, further investigation is required to elucidate the function of these putative ABC transporters in LAB.

4.4.1.2. Information processing and storage

In support of *L. plantarum* having a large number of genes dedicated to lifestyle adaptation (Molenaar *et al.*, 2005; Evanovich *et al.*, 2019), the genomes reported here have 29.7% genes assigned to information processing and storage (Figure 4.2B). This data shows that the majority (11%) of the genes encode genes related to *translation, ribosomal structure and biogenesis* and this includes tRNA synthases, 30S and 50S ribosomal proteins and tRNA methyltransferases. These data are similar to reports by Kant *et al.* (2011), which showed that 26% of the genes in the core genome of *Lactobacillus* species are attributed to the same COG class. Moreover, genes attributed to replication, recombination and repair made up 10% (in our case 6%), and those for transcription made up 7% of the core genome (Kant *et al.*, 2011), compared to the 9.93% in our data for this COG class (Figure 4.2B). A few stress-related genes were also detected and encode cold shock proteins (*cspLA*: Cold shock-like protein, *cspL*: Cold shock protein 2, *csp*: Cold shock protein 1), heat shock proteins (*hslR*: Heat shock protein 15, *groEL*: GroEL chaperonin, *groES*: GroES co-chaperonin, *grpE*: heat shock protein GrpE, *hrcA*: heat-inducible transcription repressor, *dnaK* and *dnaJ*: chaperone protein), and ATP-dependent proteases (*hslV*: ATP-dependent protease subunit, *clpP*: ATP-dependent Clp protease proteolytic subunit) (Table 4.2) (Kleerebezem *et al.*, 2003). These genes have been shown to be important for adaptation to various environments(van Bokhorst-van de Veen *et al.*, 2011; Bastard *et al.*, 2016; Golomb *et al.*, 2016; Adu *et al.*, 2018; Betteridge *et al.*, 2018). For example, Betteridge *et al.* (2018), showed that *clpX*, *groES* and *cite* were differentially expressed in ethanol-tolerant strains compared to the parent strain which was less tolerant to ethanol.

4.4.1.3. Cell process and signalling

It was found that 16.7% of 1047 genes were attributed to cell processing and signalling in the parent genome and that of the evolved isolates (Figure 4.2B). The genes in this category cover a broad range of processes including signal transduction mechanisms, cell cycle control, cell division, chromosome partitioning, and cell wall/membrane/envelope biogenesis. All of these processes are important for cell survival in various environments and for interactions with other cohabitants. The cell wall and cell membrane, for example, form a physical barrier that acts as the first line of defence against environmental stresses (Wang *et al.*, 2018) and also act as an interface for contact with other surrounding microbes (Yamasaki-Yashiki *et al.*, 2017). Genes encoding the secretion system components (apart from *secDF*) *secA*, *secE*, *secG*, *secY* were found in the strains used in this study. The same was reported for *L. plantarum* WCFS1 (Kleerebezem *et al.*, 2003), *L. plantarum* 5-2 (Liu *et al.*, 2015) and *L. plantarum* ZLP001 (Zhang *et al.*, 2018).

The genes that play a role in the interactions of *L. plantarum* with other species include those of the *oppABCDF* cluster which encodes the oligopeptide transport system (Bechtner *et al.*, 2019), and *epsH* which encodes a putative glycosyltransferase implicated in the production of exopolysaccharide for biofilm formation (Yamasaki-Yashiki *et al.*, 2017). Moreover, genes that have been reported to play a role in the interactions between yeast (*S. cerevisiae* specifically) and LAB were found in the genomes of the isolates studied here (Table 4.3).

Table 4.3 Significantly differentially expressed genes in LAB species in co-culture with *Saccharomyces cerevisiae* in water kefir, which are found in each of the evolved isolates and the parent strain in this study.

Gene	Product	LAB species	Reference
adh	Alcohol dehydrogenase	*L. nagelii*	(Bechtner *et al.*, 2019)
aldC	Alpha-acetolactate decarboxylase	*L. nagelii*	(Bechtner *et al.*, 2019)
arcB	Ornithine carbamoyltransferase, catabolic	*L. nagelii*	(Bechtner *et al.*, 2019)
citEF	Citrate lyase	*L.hordei*	(Xu *et al.*, 2019)
epsH	Putative glycosyltransferase EpsH	*L. paracasei*	(Yamasaki-Yashiki *et al.*, 2017)
fabG	3-oxoacyl-[acyl-carrier-protein] reductase	*L. paracasei*	(Yamasaki-Yashiki *et al.*, 2017)
map	Methionine aminopeptidase	*L. nagelii*	(Bechtner *et al.*, 2019)
oppA	Oligopeptide-binding protein OppA	*L. hordei*	(Xu *et al.*, 2019)
oppB	Oligopeptide transport system permease protein OppB	*L. hordei*	(Xu *et al.*, 2019)
oppC	Oligopeptide transport system permease protein OppC	*L. hordei*	(Xu *et al.*, 2019)
oppD	Oligopeptide transport ATP-binding protein OppD	*L. hordei*	(Xu *et al.*, 2019)
oppF	Oligopeptide transport ATP-binding protein OppF	*L. hordei*/*L. nagelii*	(Bechtner *et al.*, 2019; Xu *et al.*, 2019)
ribE	Riboflavin synthase	*L. nagelii*	(Bechtner *et al.*, 2019)
ribH	6,7-dimethyl-8-ribityllumazine synthase	*L. nagelii*	(Bechtner *et al.*, 2019)
zwf	Glucose-6-phosphate 1-dehydrogenase	*L. nagelii*/*L. hordei*	(Bechtner *et al.*, 2019; Xu *et al.*, 2019)
adhE	Aldehyde-alcohol dehydrogenase	*L. hordei*	(Xu *et al.*, 2019)
accC	Biotin carboxylase	*L. hordei*	(Xu *et al.*, 2019)

These genes include *adh* and *adhE* which were downregulated in *L. hordei,* reducing the production of ethanol, with a concomitant increase in diacetyl and 2,3-butandiol (Xu *et al.*, 2019). In contrast, *adh* was upregulated in *L. nagelii,* increasing ethanol production, and reducing acetoin production (Bechtner *et al.*, 2019). These data show that molecular responses to the presence of yeast during fermentation are strain-dependent. Furthermore, the genes listed in Table 4.3 are good candidates for the investigation of yeast-bacteria interactions.

4.4.2. General features of *Lactobacillus plantarum* IWBT B063 and the evolved isolates

The number of total reads for each of these isolates and the sequencing coverage are shown in Table 4.1. Reference-guided genome assembly was performed for the parent strain using *L. plantarum* WCFS1 (Kleerebezem *et al.* 2003) and *L. plantarum* LB2_1 (unpublished) as reference genomes on SPAdes genome assembler (Galaxy version 3.12) and yielded 192 contigs (Table 4.4). The parent strain was then used as the reference genome for the assembly of the evolved isolates using SPAdes (Galaxy version 3.12). The number of contigs for the evolved isolates range between 134 and 175 (Table 4.4).

The total genome length of the parent and each of the evolved isolates is approximately 3.37 Mb and all have a 44.33% GC content (Table 4.4). Predicted genome sizes for the parent strain and all evolved isolates are within the range of previously reported *L. plantarum* genomes (Kleerebezem *et al.*, 2003; Zhang *et al.*, 2009; Golneshin *et al.*, 2015; Liu *et al.*, 2015; Wan *et al.*, 2017; Amoranto *et al.*, 2018; Learman *et al.*, 2019), however the genomes of the parent and evolved isolates are slightly larger than *L. plantarum* WCFS1 and smaller than the two reported draft genomes E2C2 and E2C5 (Suryavanshi *et al.*, 2017), which means that they are among the largest *L. plantarum* strains to be sequenced thus far. There are slight sequence differences between the genomes of each of the evolved isolates (and the parent) as seen by the small variation in predicted protein coding genes which range from 3237 (isolate C5, E5, E26 and E33) to 3253 (isolate E59) (Table 4.4). The number of predicted protein genes is comparable to that of published genomes which range from 2948 to 3361 (Zhang *et al.*, 2009; Siezen *et al.*, 2012; El Halfawy *et al.*, 2017; Inglin *et al.*, 2017; Jia *et al.*, 2017). The differences are either due to mis-annotation by Prokka, where some genes are not detected due to sequence breaks during contig assembly (Kremer *et al.*, 2016) or may be a result of gene duplications which are often observed with NGS sequence data (McElroy *et al.*, 2014). Forty-three to 44% of the predicted genes are annotated as 'hypothetical proteins' (Table 4.4) which is also in line with what has been reported in literature (Wassenaar and Lukjancenko, 2014). The genomes reported here contain 4-5 rRNAs and 64-68 tRNAs which are all within the range of previously published *L. plantarum* genomes (Kleerebezem *et al.*, 2003;

Siezen *et al.*, 2012; Golneshin *et al.*, 2015; Lamontanara *et al.*, 2015; El Halfawy *et al.*, 2017; Suryavanshi *et al.*, 2017; Amoranto *et al.*, 2018). The coding region for all the evolved isolates and the parent strain spans 82.6% of the genome.

Table 4.4 Global features of the *L. plantarum* IWBT B063 genome and each of the evolved isolates

Genome	Genome size (bp)	No. of contigs	GC content (%)	Predicted protein coding genes	Genes assigned to COG*	rRNA	tRNA	No. of hypothetical proteins
				GEN0				
Parent (B063)	3 371 245	192	44.33	3243	1169	4	68	1430
				GEN50				
C5	3 366 394	143	44.33	3 237	1 163	4	65	1 429
C26	3 367 885	145	44.33	3 238	1 175	4	64	1 431
E5	3 364 361	153	44.33	3 237	1 178	5	64	1 427
E22	3 365 592	154	44.33	3 245	1 180	5	65	1 434
E26	3 365 984	150	44.33	3 237	1 168	5	66	1 439
				GEN100				
C33	3 369 210	134	44.33	3 240	1 175	5	68	1 425
C35	3 370 876	149	44.33	3 243	1 172	4	67	1 432
C40	3 370 781	146	44.33	3 243	1 172	4	65	1 424
C43	3 370 333	145	44.33	3 243	1 172	5	68	1 430
C49	3 369 771	138	44.33	3 244	1 176	5	68	1 434
C54	3 370 454	139	44.33	3 242	1 173	5	68	1 429
E33	3 370 577	164	44.33	3 237	1 170	4	66	1 426
E42	3 371 676	175	44.33	3 248	1 177	4	68	1 437
E51	3 370 046	142	44.33	3 242	1 170	4	65	1 436
E59	3 371 082	152	44.33	3 253	1 178	4	67	1 434
E60	3 370 297	144	44.33	3 249	1 178	5	69	1 435

*NCBI's Clusters of Orthologous Groups

4.4.3. Identification of the genetic targets of directed evolution

Considering the nature of the likely selective pressures, stress tolerance genes (Table 4.2) and genes that play a role in yeast-LAB interactions would be good examples of genes expected to be affected by the different DE experiments. Here, mutations were analysed per mutation type. To identify these genetic targets FreeBayes variant detector (Galaxy Version 1.1.0.46-0) was used to call variants between the parent strain and each of the isolates, and SnpEff (Galaxy Version 4.3+T.galaxy1) was used to annotate and predict the effects that the variants might have on the affected genes. Only variants with a Phred score of ≥30 (represents 99.9% accuracy) were selected (Illumina, 2011) and then filtered based on the QUAL and DP scores as mentioned in the methodology section.

I identified a total of 67, 59, 54, 62 and 56 variants for isolates C5, C26, E5, E22 and E26, respectively from GEN50. The majority of the variants are upstream of specific genes (~ 200bp 5' end of gene; range from 18.9 to 42.4%), followed by missense (18.6 to 35.8%), synonymous (18 to 34.8%) and frameshift (13.1 to 21.2%). The distribution of these mutations is shown in Figure 4.3A.

Fifty-five to 114 variants were identified for the G100C isolates and 50 to 110 variants were identified for the G100E isolates. The majority (32.7 to 50.6%) of the mutations are upstream gene variants (Figure 4.3B). The second largest number of variants is that of synonymous gene variants (20.1 to 31.8%), followed by missense gene variants (23.5 to 29.8%) and frameshift gene variants (4 to 12%) as shown in Figure 4.3B. All variants are distributed over a total of 64 genes for G50C, 98 genes for G50E, 129 genes for G100C, and 128 genes for G100E (Figure 4.3C). This data shows that as expected, there was an accumulation of mutations over time. In our work, however, the evolved isolates were exposed to a continuously changing selective pressure i.e. the biotic selective pressure exerted by growing *S. cerevisiae* strains which also had to adjust to the selective fermentation medium. It is important to bear this in mind because such a situation is similar to the selective pressures that organisms face in nature, which often fluctuate and requires continuous adaptation to heterogeneous environments to ensure survival (Bleuven & Landry, 2016).

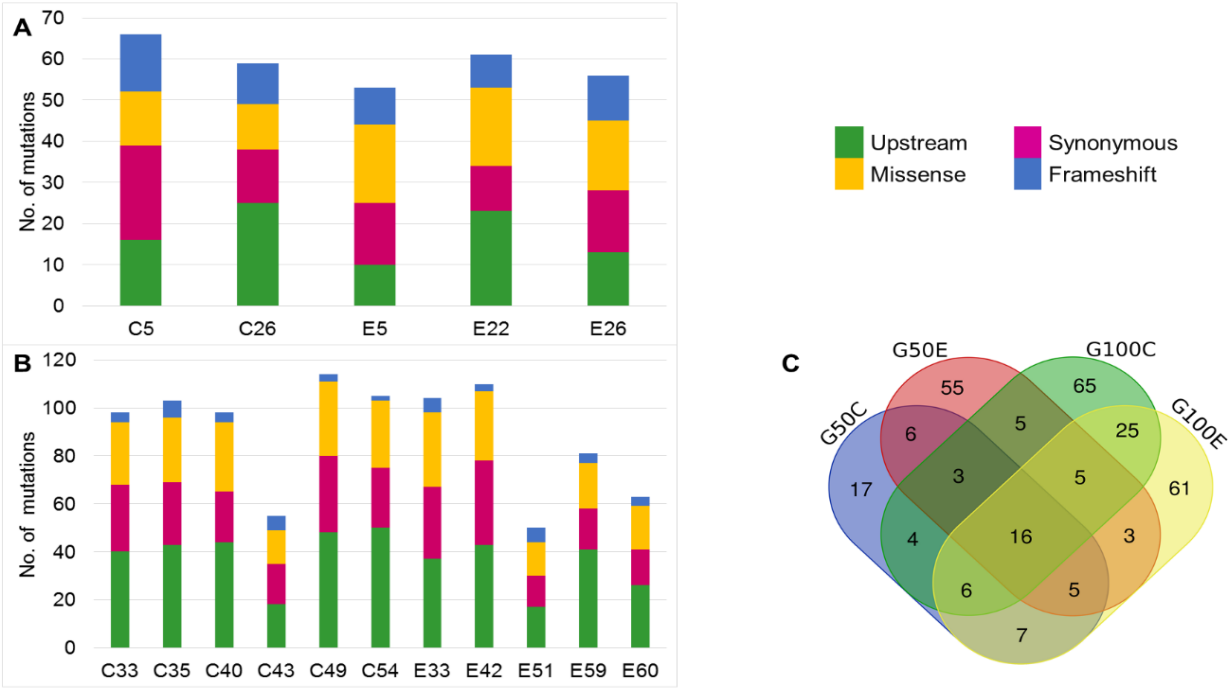

Figure 4.3 Distribution of the number of probable protein-altering mutations in all evolved isolates after (A) 50 and (B) 100 generations of DE as predicted by SnpEff. C) The total number of genes with predicted mutations for each group set. The Venn diagram was generated using the Bioinformatics & Evolutionary Genomics Venn diagram drawing tool that was accessed at http://bioinformatics.psb.ugent.be/webtools/Venn/.

4.4.3.1. The mutation spectrum and mutation frequency reveal potential genetic targets

Due to the large number of mutations in each isolate, I evaluated the mutation spectrum for each isolate. As shown in Figure 4.4, A→G (T→C) and G→A (C→T) transitions were the most common

mutations, contributing a range of 18-38% and 28-49% of all detected mutations, respectively. Transversion mutations (A→C (T→G); A→T (T→A); G→C (C→G); G→T (C→A)) were below 10% for all isolates (Figure 4.4). Notably, G→C (C→G) and A→C (T→G) transversions were not detected in isolates G50C, G50E, C43, E51 and E60 (Figure 4.4). These specific transversions were shown to be frequent in bacteria with a deficient nucleotide excision repair system (Murata-Kamiya *et al.*, 1999; Hershberg & Petrov, 2010; Baltz, 2014; Maharjan & Ferenci, 2015), with a highly expressed error-prone DNA polymerase (Maharjan & Ferenci, 2015) or in cells experiencing oxidative stress (Kino & Sugiyama, 2001). In this work, there is no evidence for the loss of repair functions or that the isolates experienced significant oxidative stress. Nonetheless, the consistent mutational bias towards transitions and not transversions for all evolved isolates suggests that the data presented here was not affected by any losses in mismatch repair functions. In support of this, the *mutT, mutH, mutL,* and *mutS* genes which are part of the mismatch repair system in various bacteria including *E. coli,* were shown to reduce the number of G→C (C→G) and A→C (T→G) transversions (Schaaper *et al.*, 1989; Machielsen *et al.*, 2010; Baltz, 2014) and the isolates used here possess both the *mutS* and *mutL* genes. Overall, there are no significant differences between all evolved isolates for the types of mutations ($p = 0.901$), and since transition mutations have been shown to be common in molecular evolution (mutational hypothesis), the yeast-driven DE appears to not have selected for any specific type of mutation (Jiang & Zhao, 2006; Hershberg & Petrov, 2010; Stoltzfus & Norris, 2016).

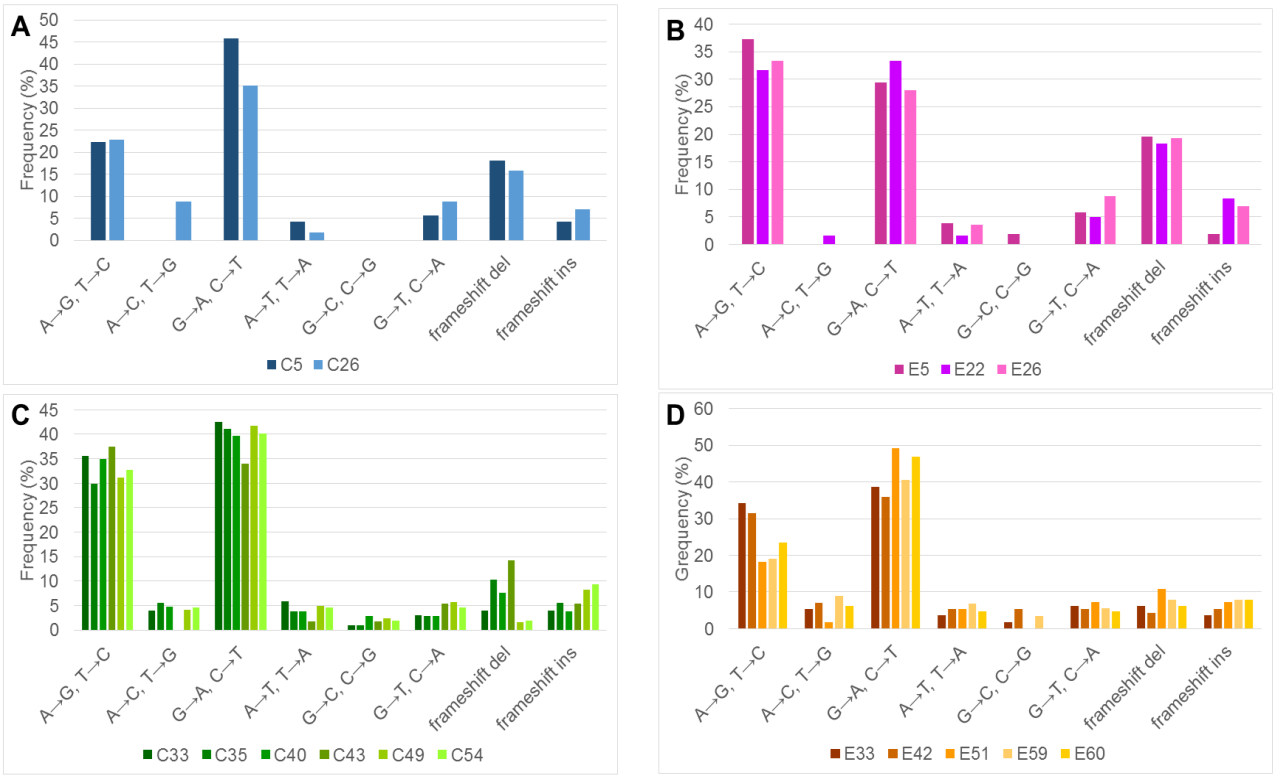

Figure 4.4 Mutation spectrum of accumulated mutations of each of the evolved isolates during the DE experiments carried out in synthetic grape must over 50 and 100 generations. The frequency of the mutations was calculated by count/total mutations*100.

To determine the number of times a mutation appeared on a particular allele the mutation frequency of all observed mutations was calculated for each variant as identified by FreeBayes and SnpEff. All affected genes and the frequencies of each mutation on the gene are shown in Table S4.2. Generally, there are more synonymous variants for G50C isolates (54.2-63.9%) compared to nonsynonymous variants (36.1-45.8%), as expected in molecular evolution (Dettman *et al.*, 2012) with a significance of $p < 0.0001$. In contrast, isolate E22 has significantly more (63.3%) nonsynonymous mutations than synonymous mutations, as determined by a one-tailed z-test ($p = 0.0002$). Generally, neutral and deleterious mutations occur more frequently than advantageous mutations and are often removed from the population in a purifying selection (Dettman *et al.*, 2012; Rocha, 2018). However, it is unlikely that all the mutations observed for all tested isolates confer a fitness advantage. It is more likely that these mutations are neutral or deleterious and are seen here due to the bottlenecks experienced during the DE experiments. Normally, bottlenecks allow such mutations to fix in the population (Bitbol & Schwab, 2014) which could explain why similar mutations are observed in isolates from both generation 50 and 100.

Interestingly, in the case of E22 the data suggests that there was strong positive selection towards protein-modifying mutations as this isolate had a larger number of nonsynonymous mutations compared to other isolates. This isolate was also the only isolate that consistently showed improved growth and L-malic acid degradation capabilities in SGM compared to all other isolates evolved for 50 generations (Chapter 3 of this dissertation). Although there were more nonsynonymous variants to synonymous variants for all other isolates evolved with EC1118, these differences were not significant (p-value = 0.807). Nonetheless, this data suggests that EC1118 exerted a stronger selection pressure, which may have increased the mutation rate and subsequent mutations, affecting the adaptation of the *L. plantarum* population. Overall, a 1:2 ratio was observed for synonymous to nonsynonymous variants after 50 generations, and a 1:1 ratio after 100 generations. This may be attributed to stronger positive selection at the early stages of evolution, suggesting that the evolving population was adjusting 'well' to the selective pressure by 100 generations, hence the 'weaker' positive selection (Dettman *et al.*, 2012; MacLean *et al.*, 2013).

In this work only nonsynonymous mutations were considered for all downstream analyses because these mutations cause obvious changes in the coded protein sequence and structure, and may affect fitness (Dettman *et al.*, 2012), defined here as the ability of isolates to utilise L-malic acid and grow in the environment in which they were evolved. Two isolates have been identified as 'fit' in Chapter 3, isolate E22 and E26. Isolates evolved for 100 generations must still be evaluated for fitness. Here, I attempted to identify genetic targets that may have resulted in the phenotypes observed for all tested isolates, including isolate E22 and isolate E26. Firstly, all the genes with

more than one mutation were considered potential targets of DE assuming that mutations occur one at a time, which would mean that such genes are probable evolutionary targets or may be mutation 'hotspots' (Wang *et al.*, 2018). Secondly, often in yeast-LAB interactions it is observed that interaction mechanisms and outcomes are usually strain specific (Edwards *et al.*, 1999; Zapparoli *et al.*, 2003; Muñoz *et al.*, 2014) and it is well to consider that the isolates tested did not all emerge from the same evolutionary line. Therefore, it is possible for some isolates to have unique mutations in genetic targets of DE, so if a gene target presented more than one mutation, but was affected in only one isolate, that gene was considered a potential target of DE as well. The gene mutation rate can be calculated as the number of genes divided by the total variant length which is the genome length divide by the number of variants. For example, the gene mutation rate for isolate C5 would be 0.064 variants/gene length and for C49 would be 0.111 variants/gene length. These data suggest that C49 is most likely to have a higher number of mutations per gene. Therefore, isolates with higher mutation rates are expected to have more changes which may increase the probability of obtaining mutations which may contribute to some of the observed phenotypes.

It was observed in Chapter 3 of this dissertation that isolate E22 was consistently faster at degrading L-malic acid when co-cultured with its driver yeast EC1118. This is evidence of these strain-specific phenotypes. Lastly, only the mutations that had a frequency ≥50% were considered and are shown in Table 4.5.

Of all the genes with nonsynonymous mutations shown in Table S4.2, the majority are genes of unknown function for all the evolved isolates (Figure S4.4). Genes involved in chromosome partitioning, transport, and peptidoglycan biosynthesis are affected in all isolates (Figure S4.4). The latter are the majority of genes with nonsynonymous mutations, which suggest some form of cell wall modification. Microscopy revealed that there are no obvious changes to cell morphology (data not shown), however it is likely that these isolates are gram variable. However, this needs further investigation. Moreover, the G50E genes involved in antimicrobial resistance, quorum sensing, two-component system, DNA replication, biosynthesis of amino acids, and ABC transporters are also affected (Figure S4.4). This data is similar to the data for G100C and G100E, which may suggest that the isolates evolved with EC1118, acquired mutations faster that those evolved with Cross Evolution. This may further suggest that in the presence of EC1118 there is a stronger selection pressure which increases mutation rate early on during adaptive evolution (Dettman *et al.*, 2012), especially in isolate E22. As calculated in Chapter 3, the populations evolved with either Cross Evolution or EC1118 had a fitness advantage of 0,35% which means it would take another 50 generations before a beneficial mutation reaches high frequency. The populations were inoculated at 2 x 10^8 which means the probability of any mutation being fixed in

the population is 5e-9, however, it would take only 2.5 generations to reach high frequency for a mutation with a 10% fitness advantage, which is a single round of the DE experiments. Should such a mutation exist it should already be observed in these isolates, however further investigation is necessary to pinpoint such a mutation.

Additionally, the similarity in the potential targets between the G50E and the G100E and G100C isolates suggests that these gene targets may be involved in the phenotypes observed in the yeast-inoculated SGM. This could explain why all the G50E, G100C, and G100E degrade all malic acid and grow better (in some cases) than the parent strain (Chapter 3; Figure S4.1 and S4.2). Another alternative explanation could be that the similarities in the mutations (such as A180T) observed for some of the genes shown in Table S4.2, i.e. *dnaI* (Primosomal protein) appearing in all G100C may be a result of the selection of a particular beneficial mutation and this particular one being carried along with in what is called genetic hitchhiking (Elena & Lenski, 2003). Such mutations may be neutral or deleterious and *in silico* analyses as was done here cannot confirm this. Therefore, experimentally testing these mutations will help shed some light on which mutations have played a role in the adaptation of the selected isolates and which contributed to the 'negative' phenotypes observed within the G50C isolates.

A total of 18 genes were identified as being the most affected (i.e. more than 1 mutation, with a frequency ≥50%, found in more than 1 isolate, unless stated otherwise). None of the genes listed in Table 4.2 and 4.3 were identified as being most affected, however, *nox*, *pox5*, *trxB*, *oppA*, *fabG*, and *acpS* were also affected in this study (Table S4.2). The genes *nox* and *pox5* have synonymous mutations, and all the others (except *oppA,* which was affected in G50E) were affected in G100C and G100E isolates (Table S4.2). These data highlight some of the targets presented in Table 4.2 and 4.3, but does not necessarily mean that they are involved in the observed phenotypes. Possibly the mutations in these genes are neutral and do not result in any changes. On the other hand, it could be argued that since these genes were involved in studies where their differential expression was investigated for the specified stresses and inoculation strategies (which varies from what was done here), they are not relevant for our experimental design. Furthermore, genes like *adh* and *adhE* are probably not affected because the SGM was not fermented to dryness, which would have increased ethanol concentrations in the media, putting pressure on the isolates to overexpress these genes (Papadimitriou et al., 2016). More than that, these genes were shown to be expressed generally in unstressed cells during fermentation, as opposed to during induced environmental stress such as starvation (Papadimitriou et al., 2016).

Of the genes that were mostly affected by mutations. 12 of the 18 genes were initially annotated as hypothetical proteins (Table S4.3). Therefore, a BLAST search was conducted and it was found

that only 7 of these genes are annotated as hypothetical or uncharacterized proteins with no functional annotation data available (L_03143, L_03120, L_03243, L_03214, L_03211, L_03242, L_03213) (Table 4.5). Therefore, these genes were not included in further analyses. According to the BLAST analysis, of the 11 remaining genes, five were identified as transmembrane protein encoding genes: *dap* (D-aminopeptidase) with 98.72% identity to serine-type D-Ala-D-Ala carboxypeptidase of *L. plantarum* WCFS1; L_00873 (hypothetical protein) which shares 98.98% identity with membrane protein PlnW of *L. plantarum* WCFS1; *spsB* (signal peptidase IB) with 99.49% identity to signal peptidase I of *L. plantarum* WCFS1; *menH* (2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate synthase) which shares 99.04% identity with alpha/beta hydrolase of *L. plantarum* WCFS1; and L_02713 (hypothetical protein) which shares 97.09% identity with the membrane-anchored cell surface SD repeat protein precursor of *L. plantarum* UCMA 3037 (Table 4.5).

One of the 11 genes was identified as an enzyme involved in DNA replication and chromosome separation: *smc* (Chromosome partition protein) with 99.39% identity to a hypothetical membrane protein of *L. plantarum* WCFS1 known as lp_1417 (Table 4.5). Two genes encoding RNA degrading enzymes are also among the genes with high mutation frequencies (Table 4.5). These include *yhaM* (encodes a 3'-5' exoribonuclease *yhaM*) and L_02894 (encoding a hypothetical protein) which share 100% and 98.96% identity with lp_1418 and lp_1556, each encoding a metal-dependent phosphohydrolase of the HD protein family in *L. plantarum* WCFS1, respectively. The remaining 3 genes encode a LysM domain containing protein (L_00288), a glycerate-2 kinase (*garK*) and a PTSEIIA domain-containing protein (*pts21A*), which share 58.49%, 98.46%, 100% identity with an extracellular transglycosylase (lp_0302), glycerate kinase (lp_3266), and PTS system, EIIA component (lp_2927) of *L. plantarum* WCFS1, respectively (Table 4.5).

4.4.5. Predicted impact on protein structure and function inferred from sequence homology

The next step was to analyse *in silico* each of the mutations to predict their impact on the target proteins. I used the amino acid sequences of each of the genes in Phyre2 (Kelley *et al.*, 2015) to predict protein crystal structure, and used ProFunc (Laskowski *et al.*, 2005) to determine active sites, binding sites, and sites of high sequence conservation. Additionally, Pfam (El-Gebali *et al.*, 2019) and Prosite (de Castro *et al.*, 2006) were used to determine conserved protein domains. In this way, I was able to determine whether the mutations occurred at significant sites that may lead to disruption of protein function.

4.4.5.1. Transmembrane proteins

D-aminopeptidase is a penicillin-binding (and degrading) protein which plays a significant role in the biosynthesis and modification of the peptidoglycan cell wall in bacteria (Bompard-Gilles *et al.*, 2000). This enzyme cleaves the terminal D-alanine residue of the pentapeptide chain of peptidoglycan, thus modifying it (Hung *et al.*, 2013). Due to its ability to remodel the peptidoglycan cell wall, this enzyme is suggested to play a role in cell shape regulation, where its inactivation resulted in cell morphology changes and hypersensitivity to detergents (Príncipe *et al.*, 2009). Moreover, it was shown that D-aminopeptidase plays a role in saline and hyperosmotic stress response in *Ochrobactrum* sp. 11a (Príncipe *et al.*, 2009). The *dap* gene is affected in isolates C5 and C26 (G50C), and isolates E51, E59 and E60 (G100E) where ALA109 is substituted for VAL. This mutation has a frequency of 100% (i.e. none of the parent alleles –in all the reads– at this position showed this variant) (Table 4.5 and Table S4.2). Isolate C5 was isolated from lineage 4 and C26 from lineage 6, both after 50 generations. Isolates E51, E59 and E60 were isolated from lineage 3 after 100 generations. These data suggest that this mutation occurred independently in each evolutionary line and that selection for this mutation was stronger in the Cross Evolution-DE experiments early on, and only later in the EC1118-DE experiments.

The 3D structure of the Dap protein (Figure 4.5A) shows that this protein has a β-lactamase domain (Pfam accession PF00144) from residue 73-386. This domain consists of an antiparallel β-sheet with five strands, flanked by 3 helices as was observed for *Ochrobactrum anthropi* (Bompard-Gilles *et al.*, 2000). Therefore, the mutation lies within this domain and on one of the five β-sheet strands (Figure 4.5A). β-Lactamases are essential for the breakdown of antibiotics such as penicillin, which breakdown the peptidoglycan cell wall (Jordan *et al.*, 2008). The predicted active sites for Dap are shown and these are S127-K130 and Y212-N214 (Figure 4.5A). Serine is the active residue (Bompard-Gilles *et al.*, 2000). It is unlikely that the mutation A109V will cause a functional change in the encoded protein, because these two amino acids have a close mutational distance i.e. they are only 1 nucleotide substitution away from each other and they belong to the same hydrophobic/uncharged group (Creixell *et al.*, 2012; Azad, 2018). Moreover, it has been shown that VAL and ALA have similar mutability (probability of mutation) and targetability (probability of a particular amino acid being the result of the mutation) ratios (Creixell *et al.*, 2012), which means that either amino acid is likely to be substituted for the other during evolution.

Another transmembrane protein affected during the DE experiments is the membrane protein PlnW (Table 4.5). This protein is involved in self-immunity against plantaracin (and other related bacteriocins) (Lages *et al.*, 2015). PlnW has 7 transmembrane helices as predicted by Phyre2 of which 4 make up the CPBP intramembrane metalloprotease domain (Pfam accession PF02517) that spans residue 123-221 (Figure 4.5B). This domain has three conserved motifs: FxxxH,

EExxxR and HxxxB at residue 175-179, 141-146, and 215-219, respectively (Figure 4.5B). These motifs are equivalent to those found in *L. plantarum* U10 which are suggested to make up the active site that is significant for self-immunity against self-produced bacteriocins such as plantaricin, pediocin, fermentcin and acidocin (Lages *et al.*, 2015). The gene that encodes the PlnW protein in isolate E5 (from lineage 1) has a mutation at M129V which lies within the conserved domain (Table 4.5). Similarly, to *dap* the substitution from MET to VAL is not surprising, because these amino acids have a short mutational distance and share similar physicochemical properties (Azad, 2018). Additionally, the mutation itself is outside of the 3 conserved motifs, which are critical for the self-immunity activity (Lages *et al.*, 2015).

It is not yet clear why proteins involved in antimicrobial activity are the targets for DE within a wine-like environment, but one hypothesis that can be offered is that these genes can be induced in the absence of antimicrobial compounds (Fraud & Poole, 2011) and that other stresses such as nutrient-limitation, oxidative stress, heat stress, cell envelope stress and acidity may confer antimicrobial resistance by activating genes involved in antimicrobial activity (Poole, 2012). This happens because antimicrobial compounds are active against actively growing cells; stressed cells do not grow as well and so these antibiotics do not work against them. In addition, mutants of proteins with a β-lactamase domain were shown to increase resistance to antimicrobial compounds in *E. coli* (Poole, 2004). Further investigation is warranted to elucidate the impact of these mutations on genes involved in antimicrobial activity.

Furthermore, there is a deletion of an ASN residue at position 158 of the signal peptidase IB gene (*spsB*) in isolate E26 (lineage 3; Table 4.5). This protein is a membrane bound endopeptidase whose function is to release the signal peptide from pre-proteins before and after translocation across the membrane (Van Roosmalen *et al.*, 2004). This protein has two SPase domains: Signal peptidases I signature 3 (Prosite accession PS00761) and Peptidase S24-like domain (Pfam accession PF00717) which spans residues 147-160 and 33-105, respectively (Figure 4.C). The active site residues include S38 and K83 which are equivalent to S90 and K145 of *E. coli* (Van Roosmalen *et al.*, 2004). The general structure of the SpsB protein consists of two antiparallel β-sheet domains (peptidase domains) (Figure 4.5C). The catalytic core domain is the signal peptidase I signature 3 domain where the active serine lies (Van Roosmalen *et al.*, 2004). The deletion resulted in the signal peptidase I signature 3 domain no longer being recognised by Prosite (de Castro *et al.*, 2006) or Profunc. This suggests that a deletion of a residue in this domain dramatically changes the structure of the domain, which might render it inactive. However, further investigation is necessary to confirm this. Moreover, this deletion as well as other mutations in the *spsB* gene (Table S4.2) may have rendered this gene a pseudogene with no actual functional consequence on the isolate itself (Lerat & Ochman, 2005).

menH encodes a 2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate (SHCHC) synthase (Figure 4.5D) which is an important enzyme in the menaquinone (vitamin K2) biosynthetic pathway (Jiang *et al.*, 2008). Menaquinone was shown to be important for adaptation to aerobic environments in *L. plantarum* WCFS1 (Brooijmans *et al.*, 2009) and *L. johnsonii* and *L. gasseri* (Maresca *et al.*, 2018). SHCHC synthase was hit with two missense mutations at position 135 where TYR was substituted for CYS and at position 190 where an ILE residue was substituted for a THR residue in isolate E22 and three isolates of the G100E group (E51, E59 and E60) (Table 4.5). Interestingly, all these isolates were isolated from lineage 3 after 50 and 100 generations, suggesting that this mutation occurred early during the DE experiment and that each of these isolates share an ancestor. It is important to note that isolate E26 is also from lineage 3, but neither of these mutations are present in this isolate. This suggests that E22 and E26 do not have the same parent.

The conversion of Y135C may be attributed to the selection of a less bulky amino acid (CYS) which is favoured by selection (Creixell *et al.*, 2012). On the other hand, the non-polar ILE is substituted for the polar THR, which may open up this residue to interact with water and other polar molecules (Azad, 2018). Furthermore, these mutations occurred in the second position of the amino acid codon, which suggests that these mutations are important and are a result of positive selection (Creixell *et al.*, 2012; Błażej *et al.*, 2018). These mutations fall within the hydrolase_4 domain (Pfam accession PF12146) which has a highly conserved catalytic site of the Ser-Asp-His motif (Jiang *et al.*, 2009; Sun *et al.*, 2014). This site is found at S169, D258, H245 and spans amino acid residue 87-310 in the isolates and strains used in this study (Figure 4.5D). Although, the I190T mutation exposes THR, the conformational change in the structure of MenH has not been drastically changed, and the active site residues are still clustered together in the active site (Figure 4.5D), which suggests that this domain has not been negatively affected.

Lastly, Phyre2 could not resolve the protein structure of L_02713; moreover, Pfam and Prosite found no hits for this protein. However, Profunc found that this protein has a bacterial Ig-like domain 2 (BID_2) using SMART (accession SM00635) (Schultz *et al.*, 1998). The BID_2 domain is found in many bacterial surface proteins such as intimin which is important for bacterial cell adhesion and carbohydrate recognition (Luo *et al.*, 2000). L_02713, like lp_1303a of *L. plantarum* WCFS1 is a large protein with 1671 residues. Unlike lp_1303a, however, L_02713 only has 522 Ser-Asp repeats instead of 1600 (Boekhorst *et al.*, 2006). It is this repeat that is involved in cell adhesion to host cells as was shown for the *Staphylococcus aureus* ClfB protein (Hartford *et al.*, 1997). Two missense mutations are observed at S1179P and L257F in isolates E5 and C40, respectively (Table 4.5). The mutation in E5 is within the Ser-Asp repeat. It is this mutation that is

likely to cause a change in the function of this protein, because SER (a polar amino acid) is substituted to a PRO residue which means that the mutation occurred on the first nucleotide of the codon, and may (to some extent) confer a positive change (Błażej *et al.*, 2018). However, due to the truncated nature of the L_02713 gene it is likely to be a pseudogene. The mutations in E5 and C40 may have occurred divergently, because these isolates are from lineage 1 and lineage 4, respectively.

Table 4.5 Significant mutations as determined by the frequency of the mutations in each group. Only the mutations with the highest mutation frequency are shown, for all other mutations see Table S4.2.

Gene name/ID	Product	G50C	Frequency (%)	G50E	Frequency (%)	G100C	Frequency (%)	G100E	Frequency (%)
L_03143	Hypothetical protein	ALL20ALF	58.5	ALL20ALF	60.9	L32F I93V GLY78GLN E73G	100.0 100.0 100.0 100.0	ALL20ALF	55.1
L_03120	Uncharacterised protein	V52	84.2	L223	58.1	L223	68.75	–	–
L_03243	Hypothetical protein	Q115	66.7	–	–	Q115	50	–	–
dap	D-aminopeptidase	A109V	100.0	–	–	–	–	A109V	100
L_00288	Extracellular transglycosylase	I165F	58.6	I165F	65.1	S190R	76.5	I165F	59.9
L_03214	Uncharacterized protein	–	–	I71V	51.9	I71V	55	I71V	53.4
L_03211	Uncharacterised protein	A25V	53.9	A25V	50.6	L28L A25V	57.1 36.3	L28L A25V	46.7 50.1
yhaM	3'-5' exoribonuclease *yhaM*	–	–	V120A T109A	100.0 100.0	–	–	–	–
L_00873	LysM domain-containing protein	–	–	M129V	100.0	–	–	–	–
spsB	Signal peptidase IB	–	–	N158	58.8	–	–	–	–
L_03242	Uncharacterized protein	–	–	K17T	51.4	K17T	51.0	K17T	53.1
menH	2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate synthase	–	–	Y135C	100.0	–	–	I190T	100
garK	Glycerate 2-kinase	–	–	T344A	100.0	–	–	–	–
L_03213	Uncharacterised protein	–	–	V48I	77.3	V48I	55.3	V48I	67.4
smc	Chromosome partition protein	–	–	R394C	100.0	–	–	–	–
L_02894	Metal-dependent phosphohydrolase, HD family	–	–	–	–	V96E	21.4	V96E	30
L_02713	Cell surface SD repeat protein precursor membrane-anchored	–	–	S1179P	100.0	L257F	100.0	–	–
pts21A	PTS system, EIIA component	–	–	–	–	–	–	V3A	100

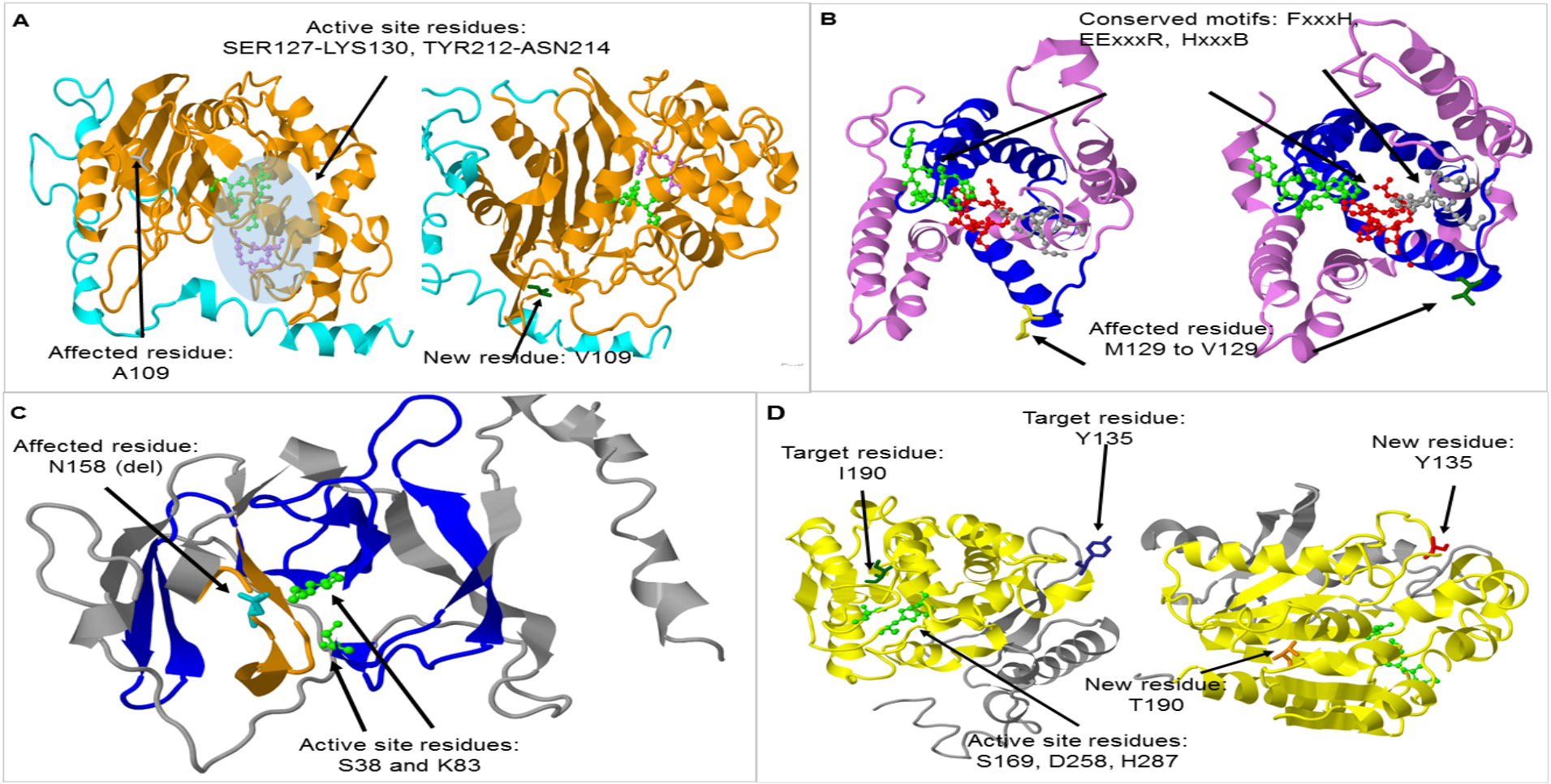

Figure 4.5 Ribbon structures of 4 predicted transmembrane proteins. A) Dap missense mutation at ALA109 in the parent (left) to V109 in the evolved isolates (right). The β-lactamase domain is shown (orange) with its active site motifs (S127-K130 and Y212-N214) shown as ball and stick structures coloured by motif. B) L_00873. On the left is the protein structure in the parent strain and on the right is the structure after the M129V missense mutation in the evolved isolate E5. The conserved domain, CPBP, is shown (blue) as well as the conserved residues (ball and stick structures coloured by motif). C) SpsB before the deletion at N158 (cyan). The two peptidase domains (peptidase 24-like domain and signal peptidase I signature 3) are coloured blue and yellow, respectively. The active site residues (Ser38 and Lys83) are depicted as ball and stick structures in green. D) MenH (left: parent; right: evolved isolate) with two missense mutations (Y135C and I190T). Active site residues are shown as stick structures (green). The structures were created with FirstGlance and (https://bioinformatics.org/firstglance/fgij/) and conserved domains were predicted using Pfam (El-Gebali *et al.*, 2019) and Prosite (de Castro *et al.*, 2006).

4.4.5.2. Nucleic acid processing proteins

The chromosome partition protein (Smc) is involved in the segregation of the chromosome during cell division, failure of which may result in anucleate cells (Nolivos & Sherratt, 2014). A single missense mutation is observed at R394C on the Smc protein in isolate E22 (Table 4.5). This mutation lies on one of the α-helices that form the coiled coil structure of the protein (Figure 4.6A). The Smc protein has four conserved ATPase motifs which form a part of the AAA_27 head domain (Figure 4.5A; Pfam accession PF13514). These are known as Walker-A (ATP-binding), Walker-B (ATP hydrolysis), and the C-motif and D-loop domain which stabalise ATP binding and are necessary for hydrolysis (Nolivos & Sherratt, 2014). The substitution from ARG to CYS, which have a long mutational distance, as observed in their physicochemical properties and the genetic code, may suggest that this protein was under positive selection (Creixell *et al.*, 2012; Azad, 2018; Błażej *et al.*, 2018), because the 'fitness' of E22 was not compromised, in fact this isolate showed significant growth and L-malic acid degradation compared to the parent strain (Chapter 3). If the mutation was null Smc would lose its functionality (Nolivos & Sherratt, 2014), resulting in cell death, which was not observed for isolate E22.

Furthermore, two genes encoding ribonucleases responsible for degrading RNA from the 3'-5' position, *yhaM* and L_02894, were affected (Figure 4.6B and C). YhaM was hit with mutations at T109A and V120A in isolates E5 (lineage 1) and E22 (lineage 3), respectively. These mutations occur in residues that have a short mutational distance (Creixell *et al.*, 2012), although the polar THR is converted to non-polar ALA. Both mutations lie within the putative helicase domain (Pfam accession PF07514) which is involved in hydrolysing phosphates during nucleic acid metabolism. In addition, this protein is suggested to have DNA-binding capabilities in *Bacillus subtilis* (Fang *et al.*, 2009). The nucleic acid binding protein domain (Pfam accession PF01336) spans residues 21-90 and there were no mutations on this domain (Figure 4.6B). While L_02894 has a missense mutation at V96E in isolate C35 (lineage 4), C54 (lineage 6), E42 (lineage 2) and E59 (lineage 3). No conserved domains could be identified for L_02894. However, three putative active sites were identified by Profunc and these are H43, H72 and D73 (Figure 4.6C). The growth of all the mutated isolates was not affected (Chapter 3 of this dissertation; TableS4.1) in spite of these mutations, suggesting that they likely did not impact strongly on the general enzyme activity. However, further investigation is required to elucidate the impact of the V96E mutation in the G100C and G100E isolates, since the substitution is of amino acids belonging to two biochemically different groups (Azad, 2018).

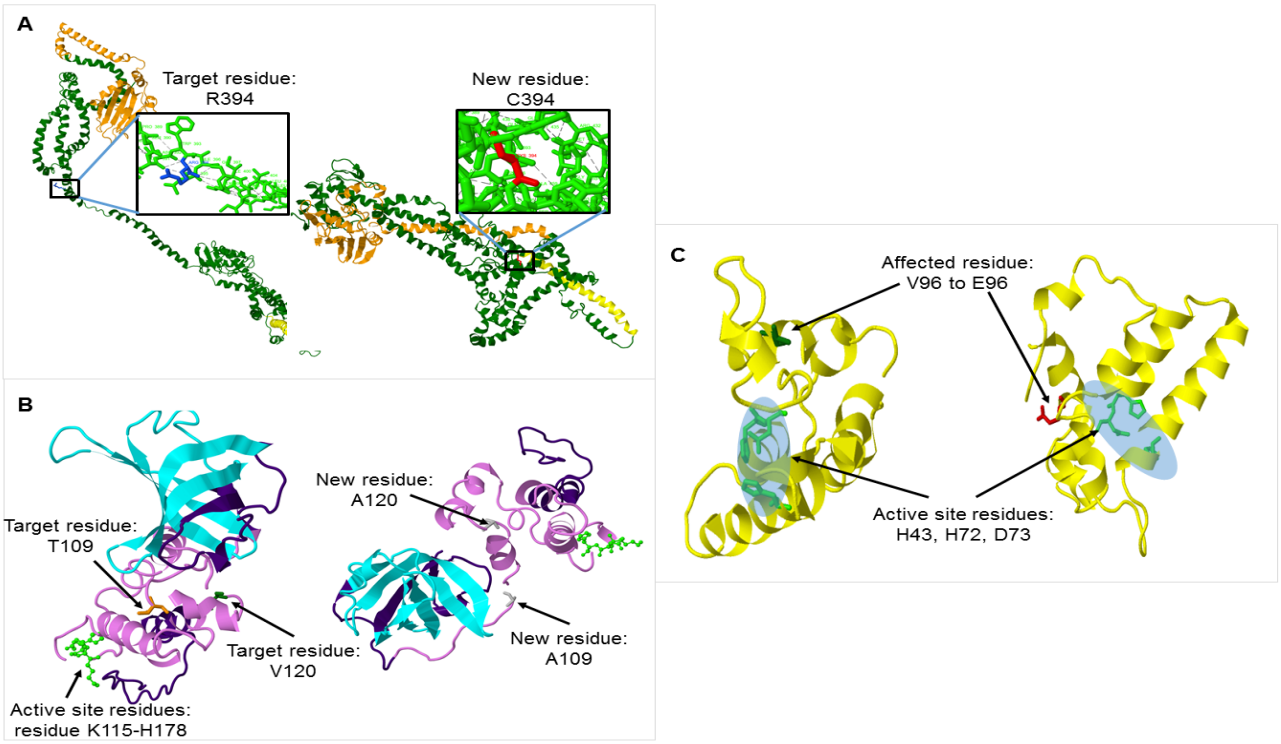

Figure 4.6 Predicted ribbon structures of the 3 mutated nucleic acid processing proteins. A) The Smc protein with the ATP-binding protein domain (AAA_27) shown in orange and the transcription regulation H-T-H motif shown in yellow. The two enlarged frames show the affected residue in the parent (left) and the substitution in the evolved isolate (right). B) For clarity the top and bottom view of the YhaM protein are shown for the parent (left) and the evolved isolate (right). The OB-fold nucleic acid binding domain is shown in cyan and the putative helicase domain (tral_2) is shown in violet. The active site residues of tral_2 are shown as ball and stick structures in green. The target residues as well as the new residues are shown and coloured by amino acid. C) L_02894 is shown along with the missense mutation (residues coloured by amino acid). The Profunc predicted active sites are shown as stick structures in green. The left structure is for the parent strain and the right structure is for the evolved isolate.

4.4.5.3. Carbohydrate and sugar metabolism

The last 3 proteins with mutations are L_00288 (LysM domain-containing protein), GarK (glycerate 2-kinase), and Pts21A (PTSEIIA domain-containing protein) which are involved in the recognition and metabolism of carbohydrates and sugars. These compounds are essential for the sustenance, survival, signalling and adhesion of cells (Ganesan *et al.*, 2007; Wong *et al.*, 2014).

Firstly, the LysM domain-containing protein has missense mutations at I165F for isolates C5, E5, E26, C54, E33, E42 and E51 and at S190R in isolates C33 and C49 (Table 4.5). These isolates were isolated from lineages 1-5, although after 50 and 100 generations. These data suggest that each of these mutations occurred divergently as no two isolates from the same lineage and generation have either mutation. These mutations suggest a positive selection in isolates C33 and C49 due to the conversion of a polar amino acid (SER) to a larger and basic amino acid (ARG) (Creixell *et al.*, 2012; Azad, 2018), which has not affected their growth (Figure S4.1) or L-malic acid degradation (Figure S4.2). L_00288 has a 58.49% sequence identity to the *L. plantarum* WCFS1

lp_0302 (extracellular transglycosylase) protein. The major difference between these proteins is that lp_0302 has both a LysM domain (Pfam accession PF01476) and a transglycosylase domain (Pfam accession PF00912), while L_00288 only has the LysM domain (Figure S4.5). The LysM domain is a peptidoglycan-binding domain in bacteria (Buist *et al.*, 2008). This domain is important for cell division and binds *N*-acetylglucosamines that form the peptidoglycan structure. Therefore, this protein is an essential peptidoglycan hydrolase (Kleerebezem *et al.*, 2010). The mutations do not fall within the LysM domain as it spans residues 9-134 (Figure 4.7A). Although the transglycosylase domain is absent in L_00288, this domain has similar roles to the LysM domain, and is suggested to play a role in carbohydrate digestion, such as peptidoglycan (Davies & Henrissat, 1995). Moreover, it was shown that the extracellular transglycosylase of *L. plantarum* USM8613 is active against *Staphylococcus aureus* (Ong *et al.*, 2019). However, because transglycosylase domain is absent in the evolved isolates and the parent strain, these strains would not exhibit anti-staphylococcal activity. Moreover, it is likely this domain is absent because the parent strain is associated with wine (Lerm *et al.*, 2011) and could have lost this antimicrobial activity, in contrast. *L. plantarum* WCFS1 is an isolate of a strain found in human saliva, and was shown recently that *S. aureus* colonises the oral cavity (McCormack *et al.*, 2015), which suggests a possible evolutionary relationship between these bacteria resulting in the presence of genes that provide a competitive advantage for WCFS1.

Secondly, Glycerate-2 kinase (Pfam accession PF02595), which is involved in the conversion of glycerate to 3-phospho-glycerate in the glycolytic pathway (ATP dependent), has a single missense mutation at T344A in isolate E22 (Figure 4.6B). Glycerate kinase is involved in alanine production, serine/threonine metabolism and glycolysis (Smeianov *et al.*, 2007; Igamberdiev & Kleczkowski, 2018). No conserved domains were found for this protein, but Profunc predicted several active site residues: D41-G45, T46-S270, N271-K273, and H54-G56. The substitution of THR to ALA is worth noting as there is a conversion from a non-polar to a polar residue, however, this mutation is not entirely surprising as THR and ALA have a short mutational distance (i.e. 1 nucleotide mutation away from each other) (Creixell *et al.*, 2012). It is interesting to note that *garK* is mutated in isolate E22 only. Further investigation is necessary to evaluate the impact of the mutation on the functioning of this gene, and ultimately the physiology of this isolate, which consistently showed improved growth and L-malic acid degradation when in co-culture with driver yeast, EC1118 compared to the parent strain (Chapter 3 of this dissertation).

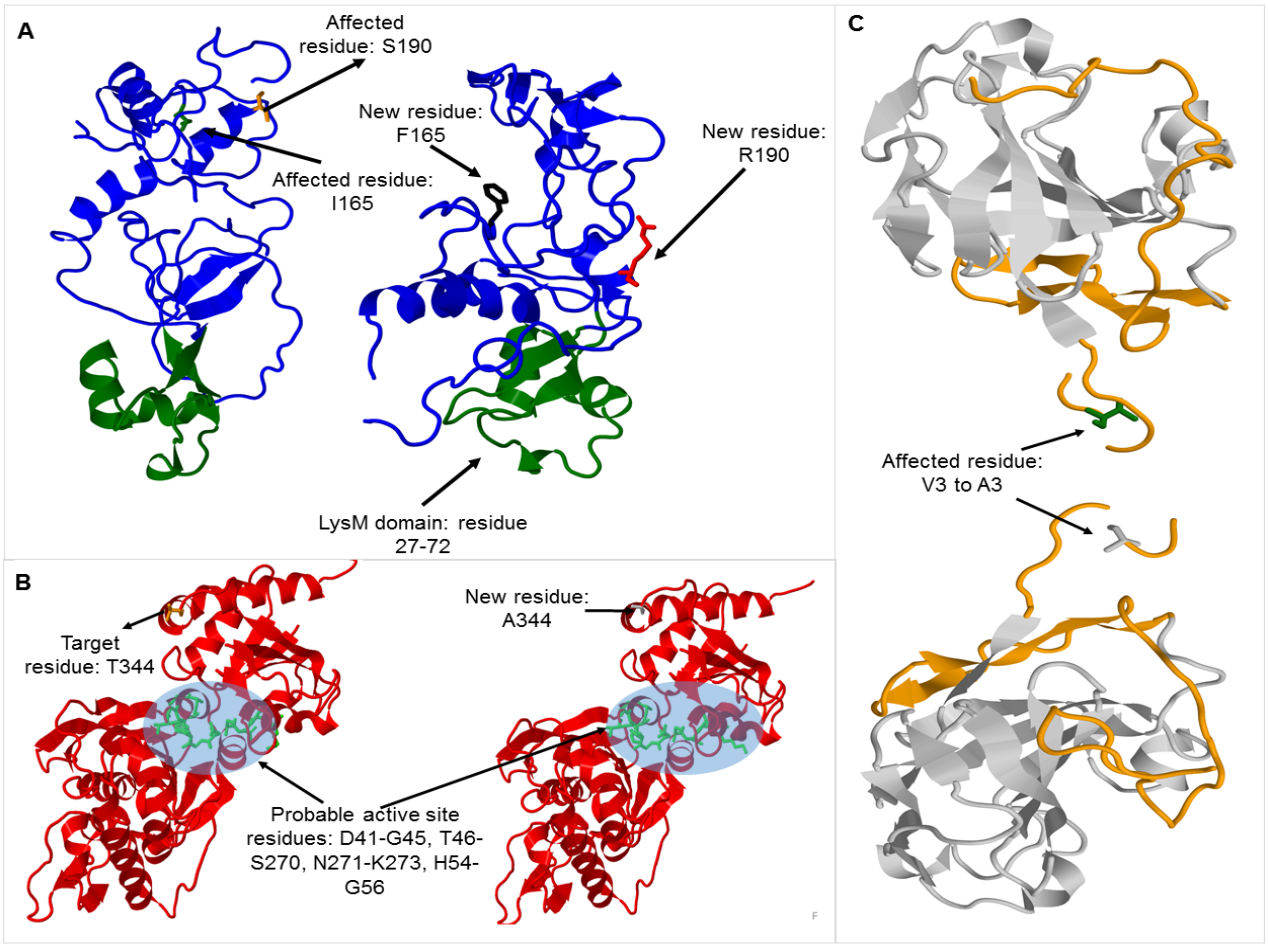

Figure 4.7 Ribbon structures of carbohydrate and sugar metabolism-related proteins affected by mutations. A) L_00288 with two missense mutations at I165F and S190R shown as stick structures and coloured by amino acid. The LysM domain is shown in green. B) GarK protein with putative active site residues shown as stick structures in light green. The missense mutation was at T344A and is shown as stick structures, coloured by amino acid. C) Pts21A with the PTS_EIIA_1 domain shown (grey) and the missense mutation (V3A) shown as stick structures, coloured by amino acids.

Lastly, a PTS system II protein has a mutation at position V3A (Table 4.5). This mutation only appears in isolate E33 (lineage 1). Pts21A harbours a PTS_EIIA_1 domain (pfam accession PF00358) (Figure 4.7C). The protein is probably involved in the transport and metabolism of sugars, although it is not evident which sugars, specifically. However, because the isolate was evolved in a medium high in glucose and fructose, it is likely that this protein is involved in their transport and metabolism as has been shown for *L. gasseri* ATCC 33323 (Francl *et al.*, 2010). Moreover, as mentioned in section 4.4.2.1 above, PTS enzyme II proteins are essential for survival in high sugar environments as they transport these sugars inside or outside the cell. Interestingly, these proteins are shown to work together with Hpr (phosphocarrier protein) and CcpA in discriminating against sugars other than glucose, which means that they induce glucose metabolism and suppress metabolism of other sugars in a medium high in glucose (Choe *et al.*, 2017). The mutation in the PTS system II protein does not affect the PTS_EIIA_1 domain which

suggests that the overall function of this protein in isolate E33 should be the same as in the parent strain and all other isolates.

4.5. Conclusion

In this work a strain of *L. plantarum* IWBT B063 was evolved for 50 and 100 generations in SGM, using *S. cerevisiae* strains as biotic selective drivers (discussed in Chapter 3 of this dissertation). All selected isolates showed a variety of phenotypes in terms of growth and MLF capabilities (see Chapter 3 of this dissertation; Figure S4.1 and S4.2). However, the molecular mechanisms that drove these phenotypic differences were not obvious. This study aimed to use *in silico* analyses to determine and compare the genome-wide genetic targets of yeast-driven evolution of *L. plantarum* isolates.

This work showcases the implications of continuous exposure to a biotic selective pressure. In this study it is observed that the use of a biotic selective pressure results in the mutation of cell wall and carbohydrate related genes (Figure S4.4, Table 4.5) which are all linked in some way to stress responses (Kleerebezem *et al.*, 2010; Wong *et al.*, 2014; Choe *et al.*, 2017). Moreover, cell wall related proteins such as the LysM domain-containing protein and D-aminopeptidase might have resulted in changes in peptidoglycan structure, which may have resulted in the isolates tested here displaying gram variable traits (Beveridge, 1990), which could be the subject of future work. Interestingly, genes encoding antimicrobial proteins were highlighted in the G50E isolates, but it was not clear what the link was to the DE experiments, warranting further investigation as the proteins D-aminopeptidase and membrane protein PlnW have not been implicated in yeast-bacteria interactions before. Nonetheless, similar proteins in other bacteria have been identified, and these confer β-lactam resistance, which is a physiological adaptation resulting from stress (Papadimitriou *et al.*, 2016).

Inasmuch as there were yeast-strain specific phenotypes among isolates evolved with Cross Evolution and EC1118, no yeast-strain specific genotypes could be identified in this work i.e. there are no genes that were exclusively affected in isolates evolved with Cross Evolution or EC1118 (Table S4.2). This is not entirely surprising because each ‘biological repeat’ in this study represented an independent evolutionary line, which follow random evolutionary landscapes (Van den Bergh *et al.*, 2018). However, there are genes that are affected only in specific isolates, such as E22 or C5, which have the ‘good growth/MLF’ and ‘poor growth/MLF’ phenotypes, respectively. These genes include *garK*, and *smc* as discussed in this chapter, and *lacM* (Beta galactosidase small subunit) in Isolate E22 (Table S4.2), and L_03243 (hypothetical protein) and lp_2817 (Transcription regulator of multidrug-efflux transporter) in isolate C5 (Table 4.5 and Table S4.2). These isolates showed improved and poor growth and MLF capabilities, respectively (refer to

Chapter 3). It is possible that the mutation of carbohydrate-related genes in E22 have provided this isolate with a competitive advantage over the parent strain, as E22 can probably utilise other carbon sources, which might explain its improved growth even after MLF was completed (Chapter 3). This hypothesis warrants further experimental validation to account for these phenotypes.

Finally, I show, for the first time, that mutations in peptidoglycan biosynthesis and degradation genes, nucleic acid processing genes and carbohydrate transport and metabolism genes are affected by DE, although it cannot be conclusively said that this was due to specific yeast drivers. Moreover, the majority of the affected genes are currently not well annotated in *L. plantarum* and sequence homology analysis was necessary to elucidate their potential functions. Additionally, affected genes were not specific for isolates evolved with either yeast, which suggests that a biotic factor-driven evolution prefers a large genetic diversity within populations and that in as much as random genes were mutated, similar pathways are involved. However, further levels of analyses such as transcriptomics or proteomics may help further characterise the impact of yeast-driven DE on the evolution of *L. plantarum*.

References

Adu, K.T., Wilson, R., et al., 2018. Proteomic analysis of *Lactobacillus* casei GCRL163 cell-free extracts reveals a SecB homolog and other biomarkers of prolonged heat stress PLoS One 13, 10, 1–28.

Afgan, E., Baker, D., et al., 2018. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update Nucleic Acids Res. 46, W1, W537–W544.

Andreevskaya, M., Johansson, P., et al., 2016. *Lactobacillus oligofermentans* glucose, ribose and xylose transcriptomes show higher similarity between glucose and xylose catabolism-induced responses in the early exponential growth phase BMC Genomics 17, 1, 1–18.

Bastard, A., Coelho, C., et al., 2016. Effect of biofilm formation by *Oenococcus oeni* on malolactic fermentation and the release of aromatic compounds in wine Front. Microbiol. 7, 1–14.

Bayrock, D.P. & Ingledew, W.M., 2004. Inhibition of yeast by lactic acid bacteria in continuous culture: Nutrient depletion and/or acid toxicity? J. Ind. Microbiol. Biotechnol. 31, 8, 362–368.

Bechtner, J., Xu, D., et al., 2019. Proteomic analysis of *Lactobacillus* nagelii in the presence of *Saccharomyces cerevisiae* isolated from water kefir and comparison with *Lactobacillus hordei* Front. Microbiol. 10, 325

Betteridge, A.L., Sumby, K.M., et al., 2018. Application of directed evolution to develop ethanol tolerant *Oenococcus oeni* for more efficient malolactic fermentation Appl. Microbiol. Biotechnol. 102, 2, 921–932.

Beveridge, T.J., 1990. Mechanism of gram variability in select bacteria J. Bacteriol. 172, 3, 1609–1620.

Błażej, P., Wnętrzak, M., et al., 2018. Optimization of the standard genetic code according to three codon positions using an evolutionary algorithm PLoS One 13, 8, 1–32.

Bleuven, C. & Landry, C.R., 2016. Molecular and cellular bases of adaptation to a changing environment in microorganisms Proc. R. Soc. B Biol. Sci. 283, 20161458.

Boekhorst, J., Wels, M., et al., 2006. The predicted secretome of *Lactobacillus plantarum* WCFS1 sheds light on interactions with its environment Microbiology 152, 11, 3175–3183.

Bompard-Gilles, C., Remaut, H., et al., 2000. Crystal structure of a D-aminopeptidase from Ochrobactrum anthropi, a new member of the “penicillin-recognizing enzyme” family Structure 8, 9, 971–980.

Borneman, A.R., McCarthy, J.M., et al., 2012. Comparative analysis of the *Oenococcus oeni* pan genome reveals genetic diversity in industrially-relevant pathways BMC Genomics 13, 373, 1–13.

Brooijmans, R.J.W., De Vos, W.M., et al., 2009. *Lactobacillus plantarum* WCFS1 electron transport chains Appl. Environ. Microbiol. 75, 11, 3580–3585.

Buist, G., Steen, A., et al., 2008. LysM, a widely distributed protein motif for binding to (peptido)glycans Mol. Microbiol. 68, 4, 838–847.

Choe, M., Park, Y.H., et al., 2017. The general PTS component HPr determines the preference for glucose over mannitol Sci. Rep. 7, 43431, 1–11.

Ciani, M., Capece, A., et al., 2016. Yeast interactions in inoculated wine fermentation Front. Microbiol. 7, 1–7.

Cingolani, P., Platts, A., et al., 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3 Fly (Austin). 6, 2, 80–92.

Creixell, P., Schoof, E.M., et al., 2012. Mutational properties of amino acid residues: Implications for evolvability of phosphorylatable residues Philos. Trans. R. Soc. B Biol. Sci. 367, 1602, 2584–2593.

Davies, G.J. & Henrissat, B., 1995. Structures and mechanisms of glycosyl hydrolases Structure 3, 853–859.

de Castro, E., Sigrist, C.J.A., et al., 2006. ScanProsite: Detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins Nucleic Acids Res. 34, 362–365.

Dettman, J.R., Rodrigue, N., et al., 2012. Evolutionary insight from whole-genome sequencing of experimentally evolved microbes Mol. Ecol. 21, 9, 2058–2077.

Dragosits, M. & Mattanovich, D., 2013. Adaptive laboratory evolution – principles and applications for biotechnology TL - 12 Microb. Cell Fact. 12, 64, 1–17.

du Toit, S.C., 2018. Coevolution of *Saccharomyces cerevisiae* and *Lactobacillus plantarum*: Engineering interspecies cooperation by Stellenbosch University.

Edwards, C.G., Reynolds, A.G., et al., 1999. Implication of acetic acid in the induction of slow/stuck grape juice fermentations and inhibition of yeast by *Lactobacillus* sp. Am. J. Enol. Vitic. 50, 2, 204–210.

El-Gebali, S., Mistry, J., et al., 2019. The Pfam protein families database in 2019 Nucleic Acids Res. 47, D427–D432.

El Halfawy, N.M., El-Naggar, M.Y., et al., 2017. Complete genome sequence of *Lactobacillus plantarum* 10CH, a potential probiotic lactic acid bacterium with potent antimicrobial activity Genome Announc. 5, 48, 1–2.

Evanovich, E., De Souza Mendonça Mattos, P.J., et al., 2019. Comparative genomic analysis of *Lactobacillus plantarum*: An overview Int. J. Genomics 2019, 1–11.

Fang, M., Zeisberg, W.M., et al., 2009. Degradation of nanoRNA is performed by multiple redundant RNases in Bacillus subtilis Nucleic Acids Res. 37, 15, 5114–5125.

Filannino, P., Cardinali, G., et al., 2014. Metabolic responses of *Lactobacillus plantarum* strains during fermentation and storage of vegetable and fruit juices Appl. Environ. Microbiol. 80, 7, 2206–2215.

Francl, A.L., Thongaram, T., et al., 2010. The PTS transporters of *Lactobacillus gasseri* ATCC 33323 BMC Microbiol. 10, 77, 1–13.

Fraud, S. & Poole, K., 2011. Oxidative stress induction of the MexXY multidrug efflux genes and promotion of aminoglycoside resistance development in Pseudomonas aeruginosa Antimicrob. Agents Chemother. 55, 3, 1068–1074.

Ganesan, B., Stuart, M.R., et al., 2007. Carbohydrate starvation causes a metabolically active but nonculturable state in *Lactococcus lactis* Appl. Environ. Microbiol. 73, 8, 2498–2512.

Garrison, E. & Marth, G. Haplotype-based variant detection from short-read sequencing. Preprint at https://arxiv.org/abs/1207.3907 (2012).

Golneshin, A., Adetutu, E., et al., 2015. Complete genome sequence of *Lactobacillus plantarum* strain B21, a bacteriocin-producing strain isolated from Vietnamese fermented sausage Nem Chua Genome Announc. 3, 2, 1–2.

Golomb, B.L., Hirao, L.A., et al., 2016. Gene expression of *Lactobacillus plantarum* and the commensal microbiota in the ileum of healthy and early SIV-infected rhesus macaques Sci. Rep. 6, 1–10.

Gurevich, A., Saveliev, V., et al., 2013. QUAST: Quality assessment tool for genome assemblies Bioinformatics 29, 8, 1072–1075.

Harris, D.R., Pollock, S. V., et al., 2009. Directed evolution of ionizing radiation resistance in Escherichia coli J. Bacteriol. 191, 16, 5240–5252.

Hartford, O., Francois, P., et al., 1997. The dipeptide repeat region of the fibrinogen-binding protein (clumping factor) is required for functional expression of the fibrinogen-binding domain on the Staphylococcus aureus cell surface Mol. Microbiol. 25, 6, 1065–1076.

Helliwell, K.E., Pandhal, J., et al., 2018. Quantitative proteomics of a B12-dependent alga grown in coculture

with bacteria reveals metabolic trade-offs required for mutualism New Phytol. 217, 2, 599–612.

Herring, C.D., Raghunathan, A., et al., 2006. Comparative genome sequencing of Escherichia coli allows observation of bacterial evolution on a laboratory timescale Nat. Genet. 38, 12, 1406–1412.

Heunis, T., Deane, S., et al., 2014. Proteomic profiling of the acid stress response in *Lactobacillus plantarum* 423 J. Proteome Res. 13, 9, 4028–4039.

Huang, M.L., Huang, J.Y., et al., 2018. Complete genome sequence of *Lactobacillus pentosus* SLC13, isolated from mustard pickles, a potential probiotic strain with antimicrobial activity against foodborne pathogenic microorganisms Gut Pathog. 10, 1, 1–6.

Hung, W.C., Jane, W.N., et al., 2013. Association of a D-alanyl-D-alanine carboxypeptidase gene with the formation of aberrantly shaped cells during the induction of viable but nonculturable Vibrio parahaemolyticus Appl. Environ. Microbiol. 79, 23, 7305–7312.

Igamberdiev, A.U. & Kleczkowski, L.A., 2018. The glycerate and phosphorylated pathways of serine synthesis in plants: The branches of plant glycolysis linking carbon and nitrogen metabolism Front. Plant Sci. 9, 1–12.

Inglin, R.C., Meile, L., et al., 2017. Complete and assembled genome sequence of *Lactobacillus plantarum* RI-113 isolated from salami Genome 5, 16, 1.

Jia, F.F., Zhang, L.J., et al., 2017. Complete genome sequence of bacteriocin-producing *Lactobacillus plantarum* KLDS1.0391, a probiotic strain with gastrointestinal tract resistance and adhesion to the intestinal epithelial cells Genomics 109, 5–6, 432–437.

Jiang, M., Chen, X., et al., 2008. Identification and characterization of (1R,6R)-2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate synthase in the menaquinone biosynthesis of Escherichia coli Biochemistry 47, 11, 3426–3434.

Jiang, Y., Zhang, J., et al., 2018. Complete genome sequencing of exopolysaccharide-producing *Lactobacillus plantarum* k25 provides genetic evidence for the probiotic functionality and cold endurance capacity of the strain Biosci. Biotechnol. Biochem. 82, 7, 1225–1233.

Jordan, S., Hutchings, M.I., et al., 2008. Cell envelope stress response in Gram-positive bacteria FEMS Microbiol. Rev. 32, 1, 107–146.

Jussier, D., Morneau, A.D., et al., 2006. Effect of simultaneous inoculation with yeast and bacteria on fermentation kinetics and key wine parameters of cool-climate Chardonnay Appl. Environ. Microbiol. 72, 1, 221–227.

Kant, R., Blom, J., et al., 2011. Comparative genomics of *Lactobacillus* Microb. Biotechnol. 4, 3, 323–332.

Kelley, L.A., Mezulis, S., et al., 2015. The Phyre2 web portal for protein modelling, prediction and analysis Nat. Protoc. 10, 6, 845–858.

Kleerebezem, M., Boekhorst, J., et al., 2003. Complete genome sequence of *Lactobacillus plantarum* WCFS1 Proc. Natl. Acad. Sci. U. S. A. 100, 4, 1990–1995.

Kleerebezem, M., Hols, P., et al., 2010. The extracellular biology of the lactobacilli FEMS Microbiol. Rev. 34, 2, 199–230.

Konings, W.N., Lolkema, J.S., et al., 1997. The role of transport processes in survival of lactic acid bacteria. Energy transduction and multidrug resistance in: Antonie van Leeuwenhoek, Int. J. Gen. Mol. Microbiol. Vol. 71. Kluwer Academic Publishers 117–128.

Koponen, J., Laakso, K., et al., 2012. Effect of acid stress on protein expression and phosphorylation in *Lactobacillus rhamnosus* GG J. Proteomics 75, 4, 1357–1374.

Kumar, S., Stecher, G., et al., 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets Mol. Biol. Evol. 33, 7, 1870–1874.

Lages, A.C., Mustopa, A.Z., et al., 2015. Cloning and expression of Plantaricin W produced by *Lactobacillus plantarum* U10 Isolate from “Tempoyak” Indonesian fermented food as immunity protein in *Lactococcus lactis* Appl. Biochem. Biotechnol. 177, 4, 909–922.

La Sarre, B., McCully, A.L., et al., 2017. Microbial mutualism dynamics governed by dose-dependent toxicity of cross-fed nutrients ISME J. 11, 2, 337–348.

Laskowski, R.A., Watson, J.D., et al., 2005. ProFunc: A server for predicting protein function from 3D structure Nucleic Acids Res. 33, SUPPL. 2, 89–93.

Lerat, E. & Ochman, H., 2005. Recognizing the pseudogenes in bacterial genomes Nucleic Acids Res. 33, 10, 3125–3132.

Lerm, E., Engelbrecht, L., et al., 2011. Selection and characterisation of *Oenococcus oeni* and *Lactobacillus plantarum* South African wine isolates for use as malolactic fermentation starter cultures South African J.

Enol. Vitic. 32, 2, 280–295.

Lewington, J., Greenaway, S.D., et al., 1987. Rapid small-scale preparation of bacterial genomic DNA, suitable for cloning and hybridization analysis Lett. Appl. Microbiol. 5, 3, 51–53.

Li, H. & Durbin, R., 2009. Fast and accurate short read alignment with Burrows–Wheeler transform Bioinformatics 25, 14, 1754–1760.

Lischer, H.E.L. & Shimizu, K.K., 2017. Reference-guided de novo assembly approach improves genome reconstruction for related species BMC Bioinformatics 18, 1, 1–12.

Liu, C.J., Wang, R., et al., 2015. Complete genome sequences and comparative genome analysis of *Lactobacillus plantarum* strain 5-2 isolated from fermented soybean Genomics 106, 6, 404–411.

Luo, Y., Frey, E.A., et al., 2000. Crystal structure of enteropathogenic Escherichia coli intimin-receptor complex Nature 405, 1073–1077.

Makarova et al., K.S., 2006. Comparative genomics of the lactic acid bacteria. Proc. Natl. Acad. Sci. U. S. A. 103, 42, 15611–15616.

Maresca, D., Zotta, T., et al., 2018. Adaptation to aerobic environment of *Lactobacillus johnsonii*/*gasseri* strains Front. Microbiol. 9, 1–11.

Mariela, L.M., Molenaar, D., et al., 2007. Thioredoxin reductase is a key factor in the oxidative stress response of *Lactobacillus plantarum* WCFS1 Microb. Cell Fact. 6, 29.

MacLean, R.C., Torres-Barceló, C., et al., 2013. Evaluating evolutionary models of stress-induced mutagenesis in bacteria Nat. Rev. Genet. 14, 3, 221–227.

McCormack, M.G., Smith, A.J., et al., 2015. Staphylococcus aureus and the oral cavity: An overlooked source of carriage and infection? Am. J. Infect. Control 43, 35–37.

McElroy, K., Thomas, T., et al., 2014. Deep sequencing of evolving pathogen populations: applications, errors, and bioinformatic solutions Microb. Inform. Exp. 4, 1, 1.

Melgar-Lalanne, G., Rivera-Espinoza, Y., et al., 2012. *Lactobacillus plantarum*: An overview with emphasis in biochemical and healthy properties in: "*Lactobacillus* Classif. Uses Heal. Implic. 1–34.

Muñoz, V., Beccaria, B., et al., 2014. Simultaneous and successive inoculations of yeasts and lactic acid bacteria on the fermentation of an unsulfited Tannat grape must Brazilian J. Microbiol. 45, 1, 59–66.

Murina, V., Kasari, M., et al., 2019. ABCF ATPases involved in protein synthesis, ribosome assembly and antibiotic resistance: Structural and functional diversification across the tree of life J. Mol. Biol. 431, 3568–3590.

Nehme, N., Mathieu, F., et al., 2008. Quantitative study of interactions between *Saccharomyces cerevisiae* and *Oenococcus oeni* strains J Ind Microbiol Biotechnol 35, 7, 685–693.

Nolivos, S. & Sherratt, D., 2014. The bacterial chromosome: Architecture and action of bacterial SMC and SMC-like complexes FEMS Microbiol. Rev. 38, 3, 380–392.

O'Sullivan, O., O'Callaghan, J., et al., 2009. Comparative genomics of lactic acid bacteria reveals a niche-specific gene set BMC Microbiol. 9, 1–9.

Ong, J.S., Taylor, T.D., et al., 2019. Extracellular transglycosylase and glyceraldehyde-3-phosphate dehydrogenase attributed to the anti-staphylococcal activity of *Lactobacillus plantarum* USM8613 J. Biotechnol. 300, 20–31.

Papadimitriou, K., Alegria, A., et al., 2016. Stress physiology of lactic acid bacteria Microbiol. Mol. Biol. Rev. 80, 3, 837–888.

Parente, E., Ciocia, F., et al., 2010. Diversity of stress tolerance in *Lactobacillus plantarum*, *Lactobacillus pentosus* and *Lactobacillus paraplantarum*: A multivariate screening study Int. J. Food Microbiol. 144, 2, 270–279.

Pfeifer, E., Gätgens, C., et al., 2017. Adaptive laboratory evolution of Corynebacterium glutamicum towards higher growth rates on glucose minimal medium Sci. Rep. 7, 1, 1–14.

Phadtare, S. & Inouye, M., 2004. Genome-wide transcriptional analysis of the cold shock response in wild-type and cold-sensitive, quadruple-csp-deletion strains of Escherichia coli J. Bacteriol. 186, 20, 7007–7014.

Piao, H., Hawley, E., et al., 2015. Insights into the bacterial community and its temporal succession during the fermentation of wine grapes Front. Microbiol. 6, 1–12.

Ponomarova, O., Gabrielli, N., et al., 2017. Yeast creates a niche for symbiotic lactic acid bacteria through nitrogen overflow Cell Syst. 5, 4, 345–357.

Poole, K., 2004. Resistance to β-lactam antibiotics Cell. Mol. Life Sci. 61, 17, 2200–2223.

Poole, K., 2012. Bacterial Stress Responses as Determinants of Antimicrobial Resistance J. Antimicribial

Chemother. 67, 2069–2089.

Price, M.N., Zane, G.M., et al., 2018. Filling gaps in bacterial amino acid biosynthesis pathways with high-throughput genetics PLoS Genet. 14, 1, 1–23.

Príncipe, A., Jofré, E., et al., 2009. Role of a serine-type d-alanyl-d-alanine carboxypeptidase on the survival of Ochrobactrum sp. 11a under ionic and hyperosmotic stress FEMS Microbiol. Lett. 295, 2, 261–273.

Reller, L.B., Weinstein, M.P., et al., 2007. Detection and identification of microorganisms by Gene Amplification and Sequencing Clin. Infect. Dis. 44, 8, 1108–1114.

Robinson, J.T., Thorvaldsdóttir, H., et al., 2011. Integrative genomics viewer Nat. Biotechnol. 29, 1, 24–26.

Rocha, E.P.C., 2018. Neutral theory, microbial practice: Challenges in bacterial population genetics Mol. Biol. Evol. 35, 6, 1338–1347.

Saguir, F.M., Loto Campos, I.E., et al., 2008. Utilization of amino acids and dipeptides by *Lactobacillus plantarum* from orange in nutritionally stressed conditions J. Appl. Microbiol. 104, 6, 1597–1604.

Schneeberger, K., Hagmann, J., et al., 2009. Simultaneous alignment of short reads against multiple genomes Genome Biol. 10, 98, 1–12.

Schneeberger, K., Ossowski, S., et al., 2011. Reference-guided assembly of four diverse Arabidopsis thaliana genomes Proc. Natl. Acad. Sci. U. S. A. 108, 25, 10249–10254.

Schultz, J., Milpetz, F., et al., 1998. SMART, a simple modular architecture research tool: Identification of signaling domains (computer Proc. Natl. Acad. Sci. USA 95, 5857–5864.

Siezen, R.J., Tzeneva, V.A., et al., 2010. Phenotypic and genomic diversity of *Lactobacillus plantarum* strains isolated from various environmental niches Environ. Microbiol. 12, 3, 758–773.

Siezen, R.J., Francke, C., et al., 2012. Complete resequencing and reannotation of the *Lactobacillus plantarum* WCFS1 genome J. Bacteriol. 194, 1, 195–196.

Smeianov, V. V, Wechter, P., et al., 2007. Comparative high-density microarray analysis of gene expression during growth of *Lactobacillus helveticus* in milk versus rich culture medium Appl. Environ. Microbiol. 73, 8, 2661–2672.

Song, K., Li, L., et al., 2016. Coverage recommendation for genotyping analysis of highly heterologous species using next-generation sequencing technology Sci. Rep. 6, 35736, 1–7.

Song, Y., He, Q., et al., 2018. Genomic variations in probiotic *Lactobacillus plantarum* P-8 in the human and rat gut Front. Microbiol. 9, 893, 1–11.

Stefanovic, E., Fitzgerald, G., et al., 2017. Advances in the genomics and metabolomics of dairy lactobacilli: A review Food Microbiol. 61, 33–49.

Sun, Z., Harris, H.M.B., et al., 2015. Expanding the biotechnology potential of lactobacilli through comparative genomics of 213 strains and associated genera Nat. Commun. 6, 8322, 1–13.

Suryavanshi, M. V, Paul, D., et al., 2017. Draft genome sequence of *Lactobacillus plantarum* strains E2C2 and E2C5 isolated from human stool culture Stand. Genomic Sci. 12, 15, 1–9.

Thompson, J.D., Higgins, D.G., et al., 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice Nucleic Acids Res. 22, 22, 4673–4680.

Unden, G. & Zaunmüller, T., 2009. Metabolism of sugars and organic acids by lactic acid bacteria from wine and must in: H. König, G. Unden, & J. Fröhlich (eds). Biol. Microorg. Grapes, Must Wine. Springer Berlin Heidelberg 135–147.

Van Roosmalen, M.L., Geukens, N., et al., 2004. Type I signal peptidases of Gram-positive bacteria Biochim. Biophys. Acta - Mol. Cell Res. 1694, 279–297.

Vrancken, G., De Vuyst, L., et al., 2011. Adaptation of *Lactobacillus plantarum* IMDO 130201, a wheat sourdough isolate, to growth in wheat sourdough simulation medium at different pH values through differential gene expression Appl. Environ. Microbiol. 77, 10, 3406–3412.

Wassenaar, T.M. & Lukjancenko, O., 2014. Comparative genomics of *Lactobacillus* and other LAB In: Lact. Acid Bact. Biodivers. Taxon. Vol. 9781444333. John Wiley & Sons, Ltd, Chichester, UK 55–69.

Wiedenbeck, J. & Cohan, F.M., 2011. Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches FEMS Microbiol. Rev. 35, 5, 957–976.

Wong, J.E.M.M., Alsarraf, H.M.A.B., et al., 2014. Cooperative binding of LysM domains determines the carbohydrate affinity of a bacterial endopeptidase protein FEBS J. 281, 4, 1196–1208.

Xu, D., Behr, J., et al., 2019. Label-free quantitative proteomic analysis reveals the lifestyle of *Lactobacillus hordei* in the presence of *Saccharomyces cerevisiae* Int. J. Food Microbiol. 294, 18–26.

Yamada, Y., Tikhonova, E.B., et al., 2012. YknWXYZ is an unusual four-component transporter with a role in

protection against sporulation-delaying-protein-induced killing of Bacillus subtilis J. Bacteriol. 194, 16, 4386–4394.

Zapparoli, G., Torriani, S., et al., 2003. Interactions between Saccharomyces and *Oenococcus oeni* strains from Amarone wine affect malolactic fermentation and wine composition Vitis 42, 2, 107–108.

Zhai, Z., Douillard, F.P., et al., 2014. Proteomic characterization of the acid tolerance response in *Lactobacillus delbrueckii* subsp. bulgaricus CAUH1 and functional identification of a novel acid stress-related transcriptional regulator Ldb0677 Environ. Microbiol. 16, 6, 1524–1537.

Zhang, Z.-Y., Liu, C., et al., 2009. Complete Genome Sequence of *Lactobacillus plantarum* JDM1 J. Bacteriol. 191, 15, 5020–5021.

Zhou, N., Bottagisi, S., et al., 2017. Yeast-bacteria competition induced new metabolic traits through large-scale genomic rearrangements in *Lachancea kluyveri* FEMS Yeast Res. 17, 6, 1–14.

Zhou, N., Swamy, K.B.S., et al., 2017. Coevolution with bacteria drives the evolution of aerobic fermentation in *Lachancea kluyveri* PLoS One 12, 3, 1–19.

Zhou, N., Katz, M., et al., 2018. Genome dynamics and evolution in yeasts: A long-term yeast-bacteria competition experiment PLoS One 13, 4, 1–16.

Supplementary data to Chapter 4

Supplementary tables

Table S4.1 Genes involved in the metabolism and transport of sugars, amino acids and metal ions as found in the parent strain and each of the evolved isolates. *Gene products implicated in yeast-bacteria interactions.

Gene name/ID	Product	Substrate	Reference
	Sugar-specific PTS enzyme II system		
ulaC	Ascorbate-specific PTS system EIIA component	Ascorbate	(Linares *et al.*, 2011)
ulaA	Ascorbate-specific PTS system EIIC component	Ascorbate	
bglF	PTS system beta-glucoside-specific EIIBCA component	β-Glucosides	(Monderer-Rothkoff & Amster-Choder, 2007)
celA	PTS system cellobiose-specific EIIB component	Cellobiose	(Michlmayr & Kneifel, 2014)
L_02857	PTS system EIIBC component	Non-specific	
fruA	PTS system fructose-specific EIIABC component*	Fructose	(Barrangou *et al.*, 2006)
gatB	PTS system galactitol-specific EIIB component	Galactitol, Galactose, Lactose	(Francl *et al.*, 2010; Cibrario *et al.*, 2016)
gatC	PTS system galactitol-specific EIIC component	Galactitol, Galactose, Lactose	
srlB	PTS system glucitol/sorbitol-specific EIIA component	Glucitol/Sorbitol	(Jia, Pang, *et al.*, 2017)
crr	PTS system glucose-specific EIIA component	Glucose	(Choe *et al.*, 2017)
ptsG	PTS system glucose-specific EIICBA component	Glucose	
glcB	PTS system glucoside-specific EIICBA component	Uncharacterised	
manX	PTS system mannose-specific EIIAB component*	Mannose	(Andreevskaya *et al.*, 2016)
manZ	PTS system mannose-specific EIID component*	Mannose	
manP	PTS system mannose-specific EIIBCA component*	Mannose	(Heravi & Altenbuchner, 2014)
mtlF	Mannitol-specific phosphotransferase enzyme IIA component	Maltose	
chbA	PTS system N,N'-diacetylchitobiose-specific EIIA component	N,N'-diacetylchitobiose	(Keyhani *et al.*, 2000)
chbC	PTS system N,N'-diacetylchitobiose-specific EIIC component	N,N'-diacetylchitobiose	
nagE	PTS system N-acetylglucosamine-specific EIICBA component	N-acetylglucosamine	(Francl *et al.*, 2010)
gmuC	PTS system oligo-beta-mannoside-specific EIIC component	Mannobiose	(Kanmani *et al.*, 2018)
sorA	PTS system sorbose-specific EIIC component*	Sorbose	
treP	PTS system trehalose-specific EIIBC component	Trehalose	(Golomb *et al.*, 2016)
licC	Lichenan permease IIC component	Cellobiose	(Heo *et al.*, 2018)
licA	Lichenan-specific phosphotransferase enzyme IIA component	Cellobiose	
licB	Lichenan-specific phosphotransferase enzyme IIB component	Cellobiose	
	Putative ABC transporters		

Table S4.1 continued

Gene name/ID	Product	Substrate	Reference
L_00578	putative ABC transporter ATP-binding protein	–	
L_01196	putative ABC transporter ATP-binding protein	–	
L_01197	putative ABC transporter ATP-binding protein	–	
L_02335	putative ABC transporter ATP-binding protein	–	
L_02337	putative ABC transporter ATP-binding protein	–	
L_01287	putative ABC transporter ATP-binding protein	–	
L_02681	putative ABC transporter ATP-binding protein	–	
L_01772	putative ABC transporter ATP-binding protein	–	
nosF	putative ABC transporter ATP-binding protein NosF	Copper	(Ulrike Honisch & Zumft, 2003)
ybiT	putative ABC transporter ATP-binding protein YbiT	–	
yheS	putative ABC transporter ATP-binding protein YheS	–	
yknY	putative ABC transporter ATP-binding protein YknY	–	
yxlF	putative ABC transporter ATP-binding protein YxlF	–	
	Other sugar transporters		
glgE	Alpha-1,4-glucan:maltose-1-phosphate maltosyltransferase	Maltose	(Syson *et al.*, 2011)
sotB	sugar efflux transporter	Melibiose/lactose	(Condemine, 2000)
maa	Maltose O-acetyltransferase	Maltose	(Fontana *et al.*, 2019)
frcA	Fructose import ATP-binding protein FrcA*	Fructose	(Lambert *et al.*, 2001)
glcU	Glucose uptake protein GlcU	Glucose	(Smeianov *et al.*, 2007)
lacS	Lactose permease	Lactose/Galactose	(Deutscher *et al.*, 2014)
araQ	L-arabinose transport system permease protein AraQ	Arabinose	(Arzamasov *et al.*, 2018)
arnC	Undecaprenyl-phosphate 4-deoxy-4-formamido-L-arabinose transferase	UDP-4-amino-4-deoxy-L- arabinose	(Breazeale *et al.*, 2002)
malK	Maltose/maltodextrin import ATP-binding protein MalK	Maltose/maltodextrin	(Wilkens, 2015)
rbsD	D-ribose pyranase	Ribose	(Andreevskaya *et al.*, 2016)
sugC	Trehalose import ATP-binding protein SugC	Trehalose	(Kanmani *et al.*, 2018)
	Amino acid and ion transporters		
glnH	ABC transporter glutamine-binding protein GlnH	Glutamine	
glnQ	Glutamine transport ATP-binding protein GlnQ	Glutamine	(Tanaka *et al.*, 2018)
glnP	putative glutamine ABC transporter permease protein GlnP	Glutamine	
glnM	putative glutamine ABC transporter permease protein GlnM	Uncharacterised	
mtsA	Metal ABC transporter substrate-binding lipoprotein	Manganese/Iron	(McLeod *et al.*, 2011)
psaA	Manganese ABC transporter substrate-binding lipoprotein	Manganese/Zinc	(Tanaka *et al.*, 2018)
fecE	Fe(3+) dicitrate transport ATP-binding protein FecE	Iron	(Davies *et al.*, 2009)

Table S4.1 continued

Gene name/ID	Product	Substrate	Reference
livH	High-affinity branched-chain amino acid transport system permease protein	BCAA	
ilvE	Branched-chain-amino-acid aminotransferase	BCAA	(Santiago *et al.*, 2012)
brnQ	Branched-chain amino acid transport system 2 carrier protein	BCAA	(Trip *et al.*, 2013)
glmS	Glutamine--fructose-6-phosphate aminotransferase [isomerizing]	Alanine/Aspartate/Glutamate	(Bearne, 1996)
serP2	DL-alanine permease SerP2	DL-Alanine/DL-Serine/Glycine	(Noens & Lolkema, 2015)
cycA	D-serine/D-alanine/glycine transporter	D-Serine/D-Alanine/Glycine	(Marlinghaus *et al.*, 2016)
braC	Leucine-, isoleucine-, valine-, threonine-, and alanine-binding protein	L-alanine/Threonine/BCAA	(Hoshino & Kose, 1989)
artM	Arginine transport ATP-binding protein ArtM	Arginine	(Wüthrich *et al.*, 2018)
artQ	Arginine transport system permease protein ArtQ	Arginine	(Walshaw *et al.*, 1997)
yitJ	Bifunctional homocysteine S-methyltransferase/5,10-methylenetetrahydrofolate reductase	Cysteine/Methionine	(Liu *et al.*, 2012)
lysP	Lysine-specific permease	Lysine	(Trip *et al.*, 2013)
metI	D-methionine transport system permease protein MetI	Methionine	(Willett *et al.*, 2015)
metN	Methionine import ATP-binding protein MetN	DL-Methionine/Methionine sulfoxide	(Wüthrich *et al.*, 2018)
metP	Methionine import system permease protein MetP	DL-Methionine/Methionine sulfoxide	
metQ	Methionine-binding lipoprotein MetQ	DL-Methionine	(Tanaka *et al.*, 2018)
potA	Spermidine/putrescine import ATP-binding protein PotA	Ornithine	
potB	Spermidine/putrescine transport system permease protein PotB	Ornithine	(Van De Guchte *et al.*, 2006)
potD	Spermidine/putrescine-binding periplasmic protein	Ornithine	
serC	Phosphoserine aminotransferase	Serine	(Koo *et al.*, 2017)
cysE	Serine O-acetyltransferase	Serine/Cysteine	(Sperandio *et al.*, 2005)
czcD	Cadmium, cobalt and zinc/H(+)-K(+) antiporter	Cobalt/Zinc/Cadmium	(Papadimitriou *et al.*, 2016)
znuC	High-affinity zinc uptake system ATP-binding protein ZnuC	Zinc	(Tanaka *et al.*, 2018)
znuB	High-affinity zinc uptake system membrane protein ZnuB	Zinc	
copA	Copper-exporting P-type ATPase	Copper/Silver	(Rensing *et al.*, 2000)
nhaK	Sodium, potassium, lithium and rubidium/H(+) antiporter	Sodium/Potassium/Lithium/Rubidium	(Fujisawa *et al.*, 2005)
ktrA	Ktr system potassium uptake protein A	Potassium	(Holtmann *et al.*, 2003)

Table S4.2 List of all the genes with variants identified in all evolved isolates. Significant variants are those that were counted as appearing more often in the alternate allele (isolate) than in the reference allele (parent). Moreover, genes that had the highest number of mutations were also considered relevant. In the event that the same mutation appears in more than one isolate, the average count was used to calculate frequency. All genes highlighted in red have nonsynonymous and/or frameshift variants with a frequency >50%.

Isolates C5 and C26 which were evolved for 50 generations in the presence of Cross Evolution

Gene name	Product	Mutation type	Mutation (base change) [REF/ALT]	Mutation (amino acid change)	Alternate allele observation count (AO)	Reference allele observation count (RO)	Frequency (%)
A0U96_14875	Uncharacterized protein	Frameshift gene variants	aaa/aa	K82	41	27	60.29
apu	Amylopullulanase	Frameshift gene variants	ggc/gc	G102	10	22	31.25
AVR82_05345	Integral membrane protein	Upstream gene variants	GCC/GC		26	9	74.29
AYO51_00615	Hypothetical protein	Missense gene variants	acaacg/Ataact	TT2IT	57	59	49.14
AYO51_00615	Hypothetical protein	Missense gene variants	Ctt/Ttt	L32F	85.5	64	57.19
AYO51_00615	Hypothetical protein	Missense gene variants	gcattgctt/gcGTTGTtt	ALL20ALF	96.5	68.5	58.48
AYO51_00615	Hypothetical protein	Synonymous gene variants	ggC/ggT	G35	99	86	53.51
AYO51_00615	Hypothetical protein	Upstream gene variants	C/T		75	71	51.37
AYO51_02915	Uncharacterized protein	Upstream gene variants	TAA/TA		25	19	56.82
AYO51_04665	Putative transcriptional regulator	Synonymous gene variants	tgT/tgC	C141	141.5	0	100.00
BIZ32_05665	L-lysine permease	Frameshift gene variants	tta/ta	L79	21.5	14	60.56
BIZ32_07310	Uncharacterised protein	Frameshift gene variants	cta/ca	L223	30.5	13	70.11
BIZ32_07310	Uncharacterised protein	Frameshift gene variants	gtg/tg	V52	16	3	84.21
BIZ32_11210	PTS lactose/cellobiose transporter subunit IIA	Frameshift gene variants	acc/ac	T76	10	23	30.30
BIZ32_12025	Uncharacterized protein	Upstream gene variants	CAAAAAT/CAAAAT		14	2	87.50
BV299_01050	VOC family virulence protein	Upstream gene variants	C/A		41.5	9	82.18
C0682_13070	DUF916 and DUF3324 domain-containing protein	Missense gene variants	gCg/gTg	A196V	28.5	0	100.00
C6Y10_13250	Hypothetical protein	Synonymous gene variants	ttT/ttC	F125	62	72	46.27
C6Y10_13250	Hypothetical protein	Upstream gene variants	AA/ACA		9	1	90.00
carB	Carbomoyl-phosphae synthase large chain	Synonymous gene variants	ggA/ggG	G16	31.5	0	100.00
CFI62_15645	LysM domain-containing protein	Upstream gene variants	G/A		14.5	7	67.44
clsA	Major cardiolipin synthase ClsA	Frameshift gene variants	gcc/gc	A202	17	8	68.00
CP352_04270	Hypothetical protein	Frameshift	caa/ca	Q115	10	5	66.67

Table S4.2 continued

Gene name	Product	Mutation type	Mutation (base change) [REF/ALT]	Mutation (amino acid change)	Alternate allele observation count (AO)	Reference allele observation count (RO)	Frequency (%)
CP352_04270	Hypothetical protein	Synonymous gene variants	tgT/tgC	C236	66	5	92.96
csbC	**Putative metabolite transport protein**	**Frameshift variant**	**atga/ata**	**M1**	**22**	**6**	**78.57**
CUR48_05825	**PAP2 family protein**	**Missense gene variants**	**gCg/gTg**	**A26V**	**58**	**0**	**100.00**
dap	**D-aminopeptidase**	**Missense gene variants**	**gCg/gTg**	**A109V**	**68.5**	**0**	**100.00**
dap	D-aminopeptidase	Synonymous gene variants	gcT/gcC	A251	17	0	100.00
ddl	**D-alanine--D-alanine ligase**	**Missense gene variants**	**gCc/gTc**	**A173V**	**19.5**	**0**	**100.00**
def	Peptide deformylase	Frameshift gene variants	gaa/ga	E77	6	19	24.00
EQG53_13400	Helix-turn-helix domain-containing protein	Synonymous gene variants	atC/atT	I68	45	38	54.22
EQG53_13400	Helix-turn-helix domain-containing protein	Synonymous gene variants	aaA/aaG	K26	28	19	59.57
LAP9492_01753	2-haloacrylate reductase	Upstream gene variants	ATTTTTA/ATTTTA		32	17	65.31
lp_0302	**Extracellular transglycosylase**	**Missense gene variants**	**Atc/Ttc**	**I165F**	**51**	**36**	**58.62**
lp_0302	Extracellular transglycosylase	Synonymous gene variants	aaG/aaA	K25	22	8	73.33
lp_0302	Extracellular transglycosylase	Upstream gene variants	GTT/GTTT		14	7	66.67
lp_0302	Extracellular transglycosylase	Upstream gene variants	T/C		30	10	75.00
lp_0302	Extracellular transglycosylase	Upstream gene variants	G/A		27	3	90.00
lp_0723	Putative ABC transporter ATP-binding protein	Upstream gene variants	C/T		43	31	58.11
lp_0723	Putative ABC transporter ATP-binding protein	Upstream gene variants	A/T		50	0	100.00
lp_0723	Putative ABC transporter ATP-binding protein	Upstream gene variants	G/A		45	0	100.00
lp_1982	**N-acetylmuramoyl-L-alanine amidase**	**Frameshift gene variants**	**gtt/gt**	**V127**	**13**	**7**	**65.00**
lp_2817	**Transcription regulator of multidrug-efflux transporter**	**Frameshift gene variants**	**gatttt/gattt**	**DF70**	**6**	**3**	**66.67**
lp_2847	Extracellular transglycosylase, with LysM peptidoglycan binding domain	Upstream gene variants	C/T		26	15	63.41
lp_3323	Lipase/esterase, SGNH or GDSL hydrolase family	Upstream gene variants	CA		30.5	0	100.00
lp_3491	FMN-binding protein	Synonymous gene variants	cgT/cgC	R101	572	0	100.00
Lp19_2747	Uncharacterized protein	Synonymous gene variants	gtA/gtG	V41	125	86	59.24
Lp19_2747	Uncharacterized protein	Synonymous gene variants	caT/caC	H33	132	86	60.55

Table S4.2 continued

Gene name	Product	Mutation type	Mutation (base change) [REF/ALT]	Mutation (amino acid change)	Alternate allele observation count (AO)	Reference allele observation count (RO)	Frequency (%)
Lp19_2747	Uncharacterized protein	Synonymous gene variants	ccG/ccA	P77	95	55	63.33
Lp19_2747	Uncharacterized protein	Synonymous gene variants	agggtt/agAGTC	RV73	93.5	53	63.82
LPJSA22_00247	**PTS beta-glucoside transporter subunit IIABC**	**Missense gene variants**	**Gca/Aca**	**A271T**	**11**	**4**	**73.33**
LpLQ80_06975	Serine/threonine phosphatase stp	Synonymous gene variants	gcC/gcT	A104	15	0	100.00
LpLQ80_09265	Uncharacterized protein	Upstream gene variants	C/T		133	112	54.29
malP	Maltose phosphorylase	Frameshift gene variants	tac/tc	Y693	6	20	23.08
menH	**2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate synthase**	**Missense gene variants**	**aTc/aCc**	**I190T**	**109**	**0**	**100.00**
mutL	DNA mismatch repair protein	Upstream gene variants	A/C		40	0	100.00
Nizo2802_2173	**Uncharacterised protein**	**Missense gene variants**	**gCc/gTc**	**A25V**	**96**	**82**	**53.93**
Nizo2802_2173	Uncharacterised protein	Synonymous gene variants	ccC/ccT	P19	80	83	49.08
Nizo2802_2173	Uncharacterised protein	Synonymous gene variants	ttT/ttC	F40	118	112	51.30
Nizo2802_2173	Uncharacterised protein	Missense gene **variants**	Gta/Ata	V48I	60	27.5	68.57
Nizo2802_2173	Uncharacterised protein	Upstream gene variants	C/T		26	0	100.00
Nizo2802_2173	Uncharacterised protein	Upstream gene variants	G/A		31	0	100.00
pfkA	ATP-dependent-6-phosphofructokinase	Synonymous gene variants	ggT/ggC	G212	98	0	100.00
repX	Replication protein RepA	Upstream gene variants	CAAAAAG/CAAAAAAG		18	15	54.55
rpmE2	**50S ribosomal protein L31 type B**	**Frameshift gene variants**	**tta/ta**	**L11**	**29**	**26**	**52.73**
sdr	cell surface SD repeat protein precursor, membrane-anchored	Synonymous gene variants	agT/agC	S2757	16	9	64.00
sdr	cell surface SD repeat protein precursor, membrane-anchored	Synonymous gene variants	tcG/tcA	S1921	17	0	100.00
sdr	cell surface SD repeat protein precursor, membrane-anchored	Synonymous gene variants	agT/agC	S1929	17	0	100.00
sdr	cell surface SD repeat protein precursor, membrane-anchored	Synonymous gene variants	tcA/tcT	S1927	17	0	100.00
uvrC	UvrABC system protein C	Upstream gene variants	C/A		16	4	80.00
xerC	**Tyrosine recombinase *xerC***	**Frameshift gene variants**	**atg/Aatg**	**M298N**	**20**	**14.5**	**57.97**
yhaM	3'-5' exoribonuclease *yhaM*	Synonymous gene variants	ggG/ggA	G213	35	0	100.00
yheS	Putative ABC transporter ATP-binding protein	Upstream gene variants	C/T		42	29	59.15
yheS	Putative ABC transporter ATP-binding protein	Upstream gene variants	G/A		42	0	100.00

Table S4.2 continued

Gene name	Product	Mutation type	Mutation (base change) [REF/ALT]	Mutation (amino acid change)	Alternate allele observation count (AO)	Reference allele observation count (RO)	Frequency (%)
ytcD	Putative HTH-type transcriptional regulator	Upstream gene variants	ATTTTTA/ATTTTA		9	5	64.29
Isolates E5, E22 and E26, evolved for 50 generations in the presence of EC1118							
Gene name	**Product**	**Mutation type**	**Mutation (base change) [REF/ALT]**	**Mutation (amino acid change)**	**Alternate allele observation count (AO)**	**Reference allele observation count (RO)**	**Frequency (%)**
A0U96_14875	**Uncharacterized protein**	**Frameshift gene variants**	**aaa/aa**	**K82**	**28.00**	**18.00**	**60.87**
addB	**ATP-dependent helicase/deoxyribonucleic subunit B**	**Missense gene variants**	**gTc/gCc**	**V72A**	**78.00**	**0.00**	**100.00**
AVR82_06480	**NADP oxidoreductase**	**Missense gene variants**	**cTc/cCc**	**L160P**	**65.00**	**0.00**	**100.00**
AVR82_12850	**Putative phosphotransferase YtmP**	**Missense gene variants**	**Acg/Gcg**	**T215A**	**35.00**	**0.00**	**100.00**
AVR82_13025	**Hypothetical protein**	**Missense gene variants**	**Tcg/Ccg**	**S228P**	**61.00**	**0.00**	**100.00**
AYO51_00615	**Hypothetical protein**	**Missense gene variants**	**gcattgctt/gcGTTGTtt**	**ALL20ALF**	**79.50**	**51.00**	**60.92**
AYO51_00615	**Hypothetical protein**	**Missense gene variants**	**Ctt/Ttt**	**L32F**	**64.00**	**61.50**	**51.00**
AYO51_00615	Hypothetical protein	Synonymous gene variants	ggC/ggT	G35	73.00	57.00	56.15
AYO51_14370	Nickel transporter	Synonymous gene variants	ctA/ctG	L87	41.00	0.00	100.00
BIZ32_01220	**ATP-dependent helicase/deoxyribonuclease subunit B**	**Missense gene variants**	**gGc/gAc**	**G605D**	**34.00**	**0.00**	**100.00**
BIZ32_07310	**Membrane protein**	**Frameshift gene variants**	**cta/ca**	**L223**	**18.00**	**13.00**	**58.06**
BV241_15255	Hypothetical protein	Synonymous gene variants	cgT/cgC	R38	64.00	0.00	100.00
BV299_01050	VOC family virulence protein	Upstream gene variants	C/A		34.50	2.50	93.24
C0682_11360	**HlyC/CorC family transporter**	**Frameshift gene variants**	**aaa/aa**	**K200**	**15.00**	**6.00**	**71.43**
C0682_13070	**DUF916 and DUF3324 domain-containing protein**	**Frameshift gene variants**	**aat/at**	**N168**	**22.00**	**7.50**	**74.58**
C0682_14530	**GNAT family N-acetyltransferase**	**Missense gene variants**	**tAt/tGt**	**Y7C**	**37.00**	**0.00**	**100.00**
C6Y09_01835	**Uncharacterized protein**	**Missense gene variants**	**Cgg/Tgg**	**R45W**	**271.00**	**0.00**	**100.00**
cah	Carbonic anhydrase	Upstream gene variants	A/G		36.00	0.00	100.00
cbf1	**3'-5' exoribonuclease *yhaM***	**Missense gene variants**	**gTc/gCc**	**V120A**	**40.00**	**0.00**	**100.00**
cbf1	**3'-5' exoribonuclease *yhaM***	**Missense gene variants**	**Acc/Gcc**	**T109A**	**58.00**	**0.00**	**100.00**

Table S4.2 continued

Gene name	Product	Mutation type	Mutation (base change) [REF/ALT]	Mutation (amino acid change)	Alternate allele observation count (AO)	Reference allele observation count (RO)	Frequency (%)
CFI62_03005	Hypothetical protein	Synonymous gene variants	acG/acA	T146	18.00	0.00	100.00
CFI62_15645	LysM domain-containing protein	Missense gene variants	Atg/Gtg	M129V	24.00	0.00	100.00
CFI62_15645	LysM domain-containing protein	Upstream gene variants	T/C		11.00	5.00	68.75
cls	Cardiolipin synthase	Missense gene variants	Acg/Gcg	T224A	28.00	0.00	100.00
ctpE	Putative cation-transporting ATPase E	Missense gene variants	gAg/gGg	E87G	24.00	0.00	100.00
CUR48_14360	DUF536 domain-containing protein	Upstream gene variants	CTTTTTTG/CTTTTTG		17.00	7.00	70.83
cytX	Putative hydroxymethylpyrimidine transporter CytX	Missense gene variants	tAc/tGc	Y342C	60.00	0.00	100.00
czcD1	Cadmium, cobalt and zinc/H( )-K( ) antiporter	Missense gene variants	Gtt/Ttt	V181F	27.00	0.00	100.00
dacB	D-alanyl-D-alanine carboxypeptidase	Missense gene variants	Gat/Aat	D213N	48.00	0.00	100.00
EQG53_13400	Helix-turn-helix domain-containing protein	Missense gene variants	Act/Gct	T40A	11.00	4.00	73.33
EQG53_13400	Helix-turn-helix domain-containing protein	Synonymous gene variants	aaA/aaG	K26	21.00	13.00	61.76
EQG53_13400	Helix-turn-helix domain-containing protein	Synonymous gene variants	atC/atT	I68	24.50	14.00	63.64
EQG58_08925	Uncharacterized protein	Downstream gene variants	C/A		17.00	4.00	80.95
EQJ86_05285	Uncharacterized protein	Upstream gene variants	TAA/TA		10.00	6.00	62.50
FC81_GL002113	Uncharacterized protein	Missense gene variants	gCg/gTg	A140V	482.00	2.00	99.59
fum	Fumarate hydratase class II	Synonymous gene variants	gcG/gcA	A81	22.00	0.00	100.00
garK	Glycerate 2-kinase	Missense gene variants	Acc/Gcc	T344A	58.00	0.00	100.00
hpk2	Two-component system histidine protein kinase sensor protein	Missense gene variants	gTc/gCc	V451A	43.00	0.00	100.00
IV39_GL002953	Acetyltransferase, GNAT family	Missense gene variants	gTg/gCg	V58A	20.00	0.00	100.00
lacM	Beta galactosidase small subunit	Missense gene variants	gAc/gGc	D602G	24.00	0.00	100.00
LAP9434_02799	Uncharacterized protein	Missense gene variants	gCg/gTg	A24V	106.00	0.00	100.00
lp_0287	Hypothetical protein	Synonymous gene variants	ctG/ctA	L7	65.00	0.00	100.00
lp_0302	Extracellular transglycosylase	Missense gene variants	Gct/Tct, Atc/Ttc	I165F	28.00	15.00	65.12
lp_0302	Extracellular transglycosylase	Synonymous gene variants	aaC/aaT	N25	12.00	3.00	80.00

Table S4.2 continued

Gene name	Product	Mutation type	Mutation (base change) [REF/ALT]	Mutation (amino acid change)	Alternate allele observation count (AO)	Reference allele observation count (RO)	Frequency (%)
lp_0302	Extracellular transglycosylase	Upstream gene variants	G/A		32.00	2.67	92.31
lp_0302	Extracellular transglycosylase	Upstream gene variants	T/C		38.00	16.00	70.37
lp_0960	Hypothetical protein	Missense gene variants	Gtt/Ttt	V351F	30.00	0.00	100.00
lp_1982	N-acetylmuramoyl-L-alanine amidase	Missense gene variants	Tcg/Ccg	S1179P	24.00	0.00	100.00
lp_2534	Protein DegV	Upstream gene variants	GTC/GC		12.50	8.50	59.52
lp_2926	Uncharacterized protein	Upstream gene variants	TAA/TA		19.00	7.00	73.08
lp_3323	Lipase/esterase, SGNH or GDSL hydrolase family	Upstream gene variants	TA/CA, C/A				#DIV/0!
Lp19_1149	Uncharacterized protein	Synonymous gene variants	Ctg/Ttg	L108	60.00	0.00	100.00
Lp19_2642	DL-alanine permease SerP2	Missense gene variants	Gct/Cct	A110P	72.00	0.00	100.00
Lp19_2747	Uncharacterized protein	Missense gene variants	Att/Gtt	I71V	67.00	62.00	51.94
Lp19_2747	Uncharacterized protein	Synonymous gene variants	ccG/ccA	P77	63.00	45.00	58.33
Lp19_2747	Uncharacterized protein	Synonymous gene variants	agggtt/agAGTC	RV73	61.00	42.00	59.22
Lp19_2747	Uncharacterized protein	Synonymous gene variants	caT/caC	H33	88.00	88.00	50.00
Lp19_2747	Uncharacterized protein	Synonymous gene variants	ttT/ttC	F125	50.00	44.00	53.19
Lp19_2747	Uncharacterized protein	Synonymous gene variants	tcG/tcA	S50	64.00	46.00	58.18
Lp19_2747	Uncharacterized protein	Synonymous gene variants	ttA/ttG	L11	37.00	47.00	44.05
Lp19_2747	Uncharacterized protein	Upstream gene variants	ATT/AT		19.00	11.00	63.33
LPJSA22_00096	L-threonine 3-dehydrogenase	Synonymous gene variants	ttA/ttG	L104	66.00	0.00	100.00
LPJSA22_01163	Uncharacterized protein	Frameshift gene variants	cgc/cg	R187	29.00	22.00	56.86
LPJSA22_02223	Uncharacterized protein	Synonymous gene variants	atC/atT	I421	66.00	0.00	100.00
LpLQ80_09265	Uncharacterized protein	Missense gene variants	aAa/aCa	K17T	38.00	36.00	51.35
LpLQ80_09265	Uncharacterized protein	Upstream gene variants	C/T		84.00	72.00	53.85
LpLQ80_13370	Cell surface protein	Upstream gene variants	TAAAAAAT/TAAAAAT		12.00	5.00	70.59
LpLQ80_14210	Hypothetical protein	Missense gene variants	gCc/gTc	A90V	57.00	31.00	64.77
luxS	S-ribosylhomocysteine lyase	Missense gene variants	gAc/gGc	D73G	17.00	0.00	100.00
menH	2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate synthase	Missense gene variants	tAc/tGc	Y135C	50.00	0.00	100.00

Table S4.2 continued

Gene name	Product	Mutation type	Mutation (base change) [REF/ALT]	Mutation (amino acid change)	Alternate allele observation count (AO)	Reference allele observation count (RO)	Frequency (%)
mscL	**Large-conductance mechanosensitive channel**	**Missense gene variants**	**gAc/gGc**	**D36G**	**18.00**	**0.00**	**100.00**
Nizo1839_1169	ATP-dependent Clp protease ATP-binding subunit	Synonymous gene variants	ggT/ggC	G237	33.00	0.00	100.00
Nizo2802_0857	**Uncharacterised protein**	**Missense gene variants**	**Gta/Ata**	**V48I**	**33.00**	**9.67**	**77.34**
Nizo2802_0857	Uncharacterised protein	Synonymous gene variants	ttT/ttC	F17	18.00	19.00	48.65
Nizo2802_0857	Uncharacterised protein	Upstream gene variants	G/A		53.67	31.00	63.39
Nizo2802_2173	**Uncharacterised protein**	**Missense gene variants**	**Ctt/Ttt, gCc/gTc**	**A25V**	**79.00**	**77.00**	**50.64**
Nizo2802_2173	Uncharacterised protein	Synonymous gene variants	ttT/ttC	F40	76.00	71.50	51.53
norG	HTH-type transcriptional regulator NorG	Synonymous gene variants	ttA/ttG	L193	37.00	0.00	100.00
oppA	**Periplasmic oligopeptide-binding protein**	**Missense gene variants**	**Acc/Gcc**	**T116A**	**46.00**	**0.00**	**100.00**
pgmB1	**Beta phosphoglucomutase**	**Missense gene variants**	**gAc/gCc**	**D16A**	**24.00**	**0.00**	**100.00**
pip	Proline iminopeptidase	Upstream gene variants	T/G		30.00	23.50	56.07
prfC	**Peptide chain release factor 3**	**Missense gene variants**	**cGg/cAg**	**R83Q**	**26.00**	**0.00**	**100.00**
repX	Replication protein RepA	Upstream gene variants	CAAAAAG/CAAAAAAG		9.00	8.67	50.94
sdr	cell surface SD repeat protein precursor, membrane-anchored	Synonymous gene variants	agC/agT	S2769	16.00	19.00	45.71
sdr	cell surface SD repeat protein precursor, membrane-anchored	Synonymous gene variants	tcG/tcT	S2777	12.00	3.00	80.00
sdr	cell surface SD repeat protein precursor, membrane-anchored	Synonymous gene variants	agT/agC	S2757	18.00	1.00	94.74
sdr	cell surface SD repeat protein precursor, membrane-anchored	Synonymous gene variants	gaT/gaC	D2774	18.00	5.00	78.26
sdr	cell surface SD repeat protein precursor, membrane-anchored	Synonymous gene variants	tcG/tcT	S2777	15.00	3.00	83.33
sdr	cell surface SD repeat protein precursor, membrane-anchored	Synonymous gene variants	tcG/tcT	S2765	31.00	9.00	77.50
smc	**Chromosome partition protein**	**Missense gene variants**	**Cgc/Tgc**	**R394C**	**59.00**	**0.00**	**100.00**
spsB	**Signal peptidase IB**	**Frameshift gene variants**	**aat/at**	**N158**	**20.00**	**14.00**	**58.82**
suhb	**Inositol-1-monophosphate**	**Missense gene variants**	**gTg/gCg**	**V226A**	**23.00**	**0.00**	**100.00**
tagG	Techoic acid translocation permease protein	Upstream gene variants	CAA/CA		4.00	11.00	26.67
tsaD	**tRNA N6-adenosine threonylcarbamoyltransferase**	**Missense gene variants**	**gTc/gCc**	**V185A**	**89.00**	**0.00**	**100.00**
xerC	**Tyrosine recombinase *xerC***	**Frameshift**	**atg/Aatg**	**M298N**	**11.33**	**9.33**	**54.84**

Table S4.2 continued

Gene name	Product	Mutation type	Mutation (base change) [REF/ALT]	Mutation (amino acid change)	Alternate allele observation count (AO)	Reference allele observation count (RO)	Frequency (%)
xrtG	exosortase family protein XrtG	Upstream gene variants	C/T		7.00	15.00	31.82
ydfG	**4-carboxymuconolactone decarboxylase**	**Missense gene variants**	**gAc/gGc**	**D88G**	**41.00**	**0.00**	**100.00**
yfbR	**5'-deoxynucleotidase**	**Missense gene variants**	**Tca/Cca**	**S121P**	**35.00**	**0.00**	**100.00**
yheS	Putative ABC transporter ATP-binding protein	Upstream gene variants	A/T		26.00	0.00	100.00
yheS	Putative ABC transporter ATP-binding protein	Upstream gene variants	G/A		22.67	0.00	100.00
yheS	Putative ABC transporter ATP-binding protein	Upstream gene variants	C/T		27.00	14.67	64.80

Isolates C33, C35, C40, C43, C49 and C54 evolved for 100 generations in the presence of Cross Evolution

Gene name	Product	Mutation type	Mutation (base change) [REF/ALT]	Mutation (amino acid change)	Alternate allele observation count (AO)	Reference allele observation count (RO)	Frequency (%)
acpS	Holo-[acyl-carrier-protein] synthase	Upstream gene variants	C/A		12.00	9.00	57.14
araQ	**L-arabinose transport system permease protein**	**Frameshift gene variants**	**ttt/tt**	**F120**	**10.00**	**6.00**	**62.50**
arnC	**Undecaprenyl-phosphate 4-deoxy-4-formamido-L-arabinose transferase**	**Missense gene variants**	**tAt/tGt**	**Y444C**	**42.40**	**0.00**	**100.00**
aroE	Shikimate dehydrogenase (NADP(+))	Synonymous gene variants	gcG/gcA	A255	89.50	0.00	100.00
AVR82_04290	Aquaporin	Upstream gene variants	A/G		101.00	2.00	98.06
AVR82_07550	Cytoskeleton protein RodZ	Synonymous gene variants	cgC/cgT	R106	82.17	0.00	100.00
AVR82_08035	Glucose uptake protein GlcU	Upstream gene variants	T/C		12.50	10.00	55.56
AVR82_13025	Hypothetical protein	Synonymous gene variants	agT/agC	S359	24.80	0.00	100.00
AVR82_13160	Cation efflux protein	Upstream gene variants	G/T		57.00	0.00	100.00
AVR82_16315	**N-acetylmuramoyl-L-alanine amidase**	**Missense gene variants**	**Cca/Tca**	**P123S**	**279.00**	**0.00**	**100.00**
AYO51_00615	**Hypothetical protein**	**Missense gene variants**	**Ctt/Ttt**	**L32F**	**140.00**	**0.00**	**100.00**
AYO51_00615	**Hypothetical protein**	**Missense gene variants**	**Ccc/Tcc**	**I93V**	**20.00**	**0.00**	**100.00**
AYO51_00615	**Hypothetical protein**	**Missense gene variants**	**ggcctgtat/ggTCTGAat**	**GLY78GLN**	**82.00**	**0.00**	**100.00**
AYO51_00615	**Hypothetical protein**	**Missense gene variants**	**gag/Ggt**	**E73G**	**96.00**	**0.00**	**100.00**
AYO51_00615	**Hypothetical protein**	**Missense gene variants**	**gcattgctt/gcGTTGTtt**	**ALL20ALF**	**146.25**	**80.50**	**64.50**
AYO51_00615	Hypothetical protein	Synonymous gene variants1	ggC/ggT	G35	166.00	0.00	100.00

Table S4.2 continued

Gene name	Product	Mutation type	Mutation (base change) [REF/ALT]	Mutation (amino acid change)	Alternate allele observation count (AO)	Reference allele observation count (RO)	Frequency (%)
AYO51_00615	Hypothetical protein	Upstream gene variants	C/T		130.67	51.33	71.79
AYO51_01645	Uncharacterized protein	Upstream gene variants	GTTTTTTTC/GTTTTTTTTC		23.67	11.00	68.27
bglA	Aryl-phospho-beta-D-glucosidase BglA	Missense gene variants	aTt/aCt	I67T	129.00	0.00	100.00
bglF	Protein-N(Pi)-phosphohistidine--sugar phosphotransferase	Upstream gene variants	T/C		56.80	0.00	100.00
bglP	PTS system, trehalose-specific IIBC component	Missense gene variants	Gca/Aca	A271T	16.75	5.75	74.44
BIZ32_05665	L-lysine permease	Frameshift gene variants	tta/ta	L79	8.00	9.00	47.06
BIZ32_07310	Membrane protein	Frameshift gene variants	cta/ca	L223	11.00	5.00	68.75
BIZ32_07310	Membrane protein	Frameshift gene variants	gtg/tg	V52	18.00	17.00	51.43
C0682_05590	Glycosyl transferase family 1	Synonymous gene variants	ctG/ctA	L424	93.00	0.00	100.00
C0682_05720	HTH-type transcriptional regulator AdhR	Upstream gene variants	C/T		17.80	1.00	94.68
C0682_11360	HlyC/CorC family transporter	Frameshift gene variants	aaa/aa	K200	51.00	50.00	50.50
C0682_12350	Putative serine/threonine protein kinase	Frameshift gene variants	ttt/tt	F208	12.00	7.00	63.16
C0682_12600	DUF2712 domain-containing protein	Frameshift gene variants	tca/tcAa	S26S	25.00	13.50	64.94
cadA	Cadmium-translocating P-type ATPase	Downstream gene variants	A/G		95.83	0.00	100.00
cbf1	3'-5' exoribonuclease *yhaM*	Synonymous gene variants	ggG/ggA	G213	39.75	0.00	100.00
CFI62_10080	TetR/AcrR family transcriptional regulator	Missense gene variants	gAt/gGt	D52G	65.80	0.00	100.00
CFI62_15645	LysM domain-containing protein	Upstream gene variants	C/T		28.00	8.00	77.78
CFI62_15645	LysM domain-containing protein	Upstream gene variants	T/C		23.00	10.25	69.17
CFI62_15645	LysM domain-containing protein	Upstream gene variants	G/A		21.50	11.75	64.66
CFI62_15645	LysM domain-containing protein	Upstream gene variants	TAA/TA		5.00	17.00	22.73
cls	Cardiolipin synthase	Upstream gene variants	ATT/ATTT		33.33	24.67	57.47
clsA	Major cardiolipin synthase ClsA	Frameshift gene variants	gcc/gc	A202	32.00	13.00	71.11
CP352_04270	Hypothetical protein	Frameshift gene variants	caa/ca	Q115	8.00	8.00	50.00
crtN	Dehydrosqualene desaturase	Missense gene variants	Tct/Cct	S339P	64.20	0.00	100.00
CUR48_07570	Uncharacterized protein	Upstream gene variants	T/C		85.17	0.00	100.00
CUR48_14360	DUF536 domain-containing protein	Upstream gene variants	CTTTTTG/CTTTTG		45.00	43.50	50.85
cydD	Cytochrome D ABC transporter, ATP-binding and permease protein	Frameshift gene variants	ttg/tg	L536	25.00	21.00	54.35
ddl	D-alanine--D-alanine ligase	Missense	gCc/gTc	A173V	37.50	0.00	100.00

Table S4.2 continued

Gene name	Product	Mutation type	Mutation (base change) [REF/ALT]	Mutation (amino acid change)	Alternate allele observation count (AO)	Reference allele observation count (RO)	Frequency (%)
dnaI	**Primosomal protein DnaI**	**Missense gene variants**	**Gcg/Acg**	**A180T**	**66.83**	**0.00**	**100.00**
ecfA2	**Energy-coupling factor transporter ATP-binding protein EcfA2**	**Missense gene variants**	**tCa/tTa**	**S142L**	**41.00**	**0.00**	**100.00**
EQG58_16085	Carbohydrate ABC transporter permease	Upstream gene variants	CAA/CAAA		20.00	21.00	48.78
EQJ92_04985	Glucose-starvation inducible protein B	Upstream gene variants	A/T		67.20	0.20	99.70
fabG1	3-oxoacyl-[acyl-carrier protein] reductase	Upstream gene variants	C/T		19.00	0.00	100.00
hicD2	L-2-hydroxyisocaproate dehydrogenase	Upstream gene variants	T/C		93.00	0.00	100.00
hisA	**1-(5-phosphoribosyl)-5-[(5-phosphoribosylamino) methylideneamino] imidazole-4-carboxamide isomerase**	**Missense gene variants**	**gTc/gCc**	**V49A**	**64.00**	**0.00**	**100.00**
hsrA_2	**Putative transport protein HsrA**	**Missense gene variants**	**Ggc/Agc**	**G140S**	**68.20**	**0.40**	**99.42**
IV39_GL001561	Uncharacterized protein	Upstream gene variants	T/C		63.50	0.50	99.22
IV39_GL002161	Integral membrane protein	Upstream gene variants	A/G		33.00	0.00	100.00
IV39_GL002658	Tryptophanyl-tRNA synthetase II	Synonymous gene variants	ttA/ttG	L31	94.67	0.00	100.00
lp_0287	Hypothetical protein	Upstream gene variants	T/C		99.20	0.00	100.00
lp_0302	**Extracellular transglycosylase**	**Missense gene variants**	**agT/agA**	**S190R**	**13.00**	**4.00**	**76.47**
lp_0302	**Extracellular transglycosylase**	**Missense gene variants**	**Atc/Ttc**	**I165F**	**50.33**	**32.33**	**60.89**
lp_0302	Extracellular transglycosylase	Synonymous gene variants	aaC/aaT	N25	18.33	7.67	70.51
lp_0302	Extracellular transglycosylase	Synonymous gene variants	aaG/aaA	K2	35.00	18.67	65.22
lp_0348	Multidrug-efflux transporter, major facilitator superfamily (MFS), EmrB family	Synonymous gene variants	atC/atT	I155	56.40	0.20	99.65
lp_0348	Multidrug-efflux transporter, major facilitator superfamily (MFS), EmrB family	Upstream gene variants	CTTTTTTC/CTTTTTC		15.00	6.50	69.77
lp_0533	**Transport protein, QueT family**	**Missense gene variants**	**Cgc/Tgc**	**R100C**	**100.67**	**0.00**	**100.00**
lp_0691	Cytokinin riboside 5'-monophosphate phosphoribohydrolase	Upstream gene variants	TG/TGCG		16.50	9.75	62.86
lp_0862	Uncharacterized protein	Synonymous gene variants	cgG/cgA	R65	88.50	0.00	100.00
lp_0988	Extracellular lipoprotein, Asp-rich	Upstream gene variants	CAAAAAG/CAAAAG		13.00	3.00	81.25
lp_1336	Bifunctional protein: ABC transporter, ATP-binding protein transcription regulator, LytR family	Synonymous gene variants	ttG/ttA	L285	87.00	0.00	100.00
lp_1556	Metal-dependent phosphohydrolase, HD family	Missense gene variants	gTg/gAg	V96E	6.00	22.00	21.43
lp_1556	Metal-dependent phosphohydrolase, HD family	Synonymous gene variants	gtG/gtA (C49)	V96 (C49)	9.00	10.00	47.37

Table S4.2 continued

lp_1982	**N-acetylmuramoyl-L-alanine amidase**	**Frameshift**	**gtt/gt**	**V127**	**16.00**	**9.00**	**64.00**

Gene name	Product	Mutation type	Mutation (base change) [REF/ALT]	Mutation (amino acid change)	Alternate allele observation count (AO)	Reference allele observation count (RO)	Frequency (%)
lp_1982	N-acetylmuramoyl-L-alanine amidase	Synonymous gene variants	tcT/tcG	S1507	15.00	12.00	55.56
lp_2488g	**Uncharacterized protein**	**Missense gene variants**	**gAt/gGt**	**D15G**	**43.00**	**0.00**	**100.00**
lp_2745	Nucleotide-binding protein, universal stress protein UspA family	Upstream gene variants	ACG/AG		11.00	8.00	57.89
lp_2768	Transport protein, major facilitator superfamily (MFS)	Upstream gene variants	TAC/TC		3.00	11.00	21.43
lp_2797	Uncharacterized protein	Upstream gene variants	G/A		20.00	0.00	100.00
lp_2799	**Amino acid transport protein**	**Missense gene variants**	**Gca/Aca**	**A241T**	**88.33**	**0.00**	**100.00**
lp_2925	Cell surface protein, LPXTG-motif cell wall anchor	Upstream gene variants	T/C		20.00	0.00	100.00
lp_3098	**NAD-dependent epimerase/dehydratase family protein**	**Missense gene variants**	**Ggt/Agt**	**G253S**	**166.00**	**0.00**	**100.00**
lp_3259	**Zinc-dependent proteinase**	**Missense gene variants**	**Gaa/Aaa**	**E192K**	**47.00**	**0.00**	**100.00**
Lp19_1149	Uncharacterized protein	Upstream gene variants	AA/ACA		46.00	4.00	92.00
Lp19_2747	**Uncharacterized protein**	**Missense gene variants**	**Att/Gtt**	**I71V**	**143.00**	**117.00**	**55.00**
LPJSA22_00247	**PTS beta-glucoside transporter subunit IIABC**	**Missense gene variants**	**aGc/aCc**	**S279T**	**75.00**	**0.00**	**100.00**
LPJSA22_02902	Uncharacterized protein	Synonymous gene variants	ttA/ttG	L411	47.50	0.00	100.00
LpLQ80_06975	Serine/threonine phosphatase stp	Synonymous gene variants	gcC/gcT	A104	17.00	0.00	100.00
LpLQ80_06975	Serine/threonine phosphatase stp	Synonymous gene variants	gcC/gcT	A104	17.50	0.00	100.00
LpLQ80_09265	**Uncharacterized protein**	**Missense gene variants**	**aAa/aCa**	**K17T**	**98.00**	**94.00**	**51.04**
LpLQ80_13370	Cell surface protein	Upstream gene variants	ATTTTTA/ATTTTTTA		13.00	9.00	59.09
lysP	**Lysine transport protein**	**Missense gene variants**	**Gtg/Atg**	**V207M**	**24.80**	**0.00**	**100.00**
mae	Malic enzyme, NAD-dependent	Synonymous gene variants	agC/agT	S112	97.20	0.00	100.00
mutL	DNA mismatch repair protein	Upstream gene variants	A/C		73.20	0.00	100.00
nagA	N-acetylglucosamine-6-phosphate deacetylase	Frameshift gene variants	aaa/aa	K346	15.00	44.00	25.42
ndh2	NADH dehydrogenase, membrane-anchored	Synonymous gene variants	tcC/tcT	S600	125.50	0.00	100.00
nha2	Na(+)/H(+) antiporter	Frameshift gene variants	atc/at	I87	11.00	19.00	36.67
nhaK_4	Sodium, potassium, lithium & rubidium/H(+)	Upstream gene variants	T/C		49.50	0.00	100.00

Table S4.2 continued

Nizo2802_0857	Uncharacterised protein	Missense	Gta/Ata	V48I	80.83	31.00	72.28

Gene name	Product	Mutation type	Mutation (base change) [REF/ALT]	Mutation (amino acid change)	Alternate allele observation count (AO)	Reference allele observation count (RO)	Frequency (%)
Nizo2802_2173	Uncharacterised protein	Missense gene variants	gCc/gTc	A25V	183.50	111.83	62.13
nox2	NADH oxidase	Synonymous gene variants	tcT/tcC	S176	22.50	0.00	100.00
nrdH	Glutaredoxin-like protein *nrdH*	Upstream gene variants	TAA/TA		14.00	7.00	66.67
pepD2	Dipeptidase A	Synonymous gene variants	ggT/ggC	G172	143.00	0.00	100.00
pglH	GalNAc-alpha-(1->4)-GalNAc-alpha-(1->3)-diNAcBac-PP-undecaprenol alpha-1,4-N-acetyl-D-galactosaminyl transferase	Missense gene variants	cGg/cAg	R44Q	27.33	0.00	100.00
pncB	Nicotinate phosphoribosyltransferase	Missense gene variants	tCg/tTg	S260L	65.50	0.00	100.00
ppx	Exopolyphosphatase	Frameshift gene variants	gaa/ga	E35	37.00	20.00	64.91
prfB	Peptide chain release factor 2	Missense gene variants	Cat/Tat	H314Y	54.50	0.50	99.09
pts21A	PTS system, EIIA component	Upstream gene variants	TAA/TA		43.00	26.50	61.87
purK1	N5-carboxyaminoimidazole ribonucleotide synthase	Missense gene variants	Ggt/Agt	G942S	58.00	0.00	100.00
recR	Recombination protein RecR	Synonymous gene variants	ggT/ggC	G178	49.50	0.00	100.00
rsmB	16S rRNA (Cytosine(967)-C(5))-methyltransferase	Synonymous gene variants	cgG/cgA	R40	15.67	0.00	100.00
slmA	Nucleoid occlusion factor SlmA	Upstream gene variants	ATT/ATTT		58.00	44.00	56.86
smc	Chromosome partition protein	Upstream gene variants	G/C		17.50	0.00	100.00
tarL	Techoic acid poly(ribitol-phosphate) polymerase	Frameshift gene variants	tta/ttTa	L422F	26.67	19.67	57.55
thiD	Phosphomethylpyrimidine kinase & hydroxymethylpyrimidine kinase	Missense gene variants	Gta/Ata	V21I	46.50	0.00	100.00
thrA	Homoserine dehydrogenase	Synonymous gene variants	gaT/gaC	D202	65.67	0.00	100.00
topA	DNA topoisomerase 1	Missense gene variants	Gcg/Acg	A608T	34.50	0.00	100.00
trxA1	Thioredoxin	Missense gene variants	aAg/aCg	K108T	23.00	0.00	100.00
udk	Uridine kinase	Missense gene variants	Cgg/Tgg	R154W	83.00	0.00	100.00
uvrA	UvrABC system protein A	Missense gene variants	Ggg/Agg	G745R	103.00	0.00	100.00
uvrC	UvrABC system protein C	Upstream gene variants	C/A		23.40	1.40	94.35
wapA	Cell surface protein, LPXTG-motif cell wall anchor	Synonymous gene variants	gaC/gaT	D414	15.00	0.00	100.00
wapA	Cell surface protein, LPXTG-motif cell wall anchor	Synonymous gene variants	gaC/gaT	D399	15.00	1.00	93.75

Table S4.2 continued

xrtG	exosortase family protein XrtG	Upstream gene variants	A/C		25.40	0.00	100.00

Gene name	Product	Mutation type	Mutation (base change) [REF/ALT]	Mutation (amino acid change)	Alternate allele observation count (AO)	Reference allele observation count (RO)	Frequency (%)
xylB_3	Xylulose kinase	Upstream gene variants	G/A		17.00	9.00	65.38
yheS	Putative ABC transporter ATP-binding protein	Upstream gene variants	G/A		60.00	0.00	100.00
yjem	**Glutamate/gamma-aminobutyrate family transporter YjeM**	**Missense gene variants**	**gGa/gAa**	**G24E**	**40.00**	**0.00**	**100.00**
ytcD	**Putative HTH-type transcriptional regulator**	**Missense gene variants**	**gCg/gTg**	**A842V**	**46.00**	**0.00**	**100.00**

Isolates E33, E42, E51, E59 and E60 evolved for 100 generations in the presence of EC1118

Gene name	Product	Mutation type	Mutation (base change) [REF/ALT]	Mutation (amino acid change)	Alternate allele observation count (AO)	Reference allele observation count (RO)	Frequency (%)
A0A165S4L8	Uncharacterized protein	Synonymous gene variants	acA/acG	T157	183.50	0.00	100.00
addB	**ATP-dependent helicase/deoxyribonuclease subunit B**	**Missense gene variants**	**ttT/ttA**	**F500L**	**16.00**	**14.00**	**53.33**
aspAT	**Aminotransferase**	**Missense gene variants**	**Acg/Gcg**	**T162A**	**127.67**	**0.33**	**99.74**
AVR82_04325	**Cell surface protein**	**Missense gene variants**	**gAt/gGt**	**D86G**	**133.00**	**1.00**	**99.25**
AVR82_11000	**MurR/RpiR family transcriptional regulator**	**Frameshift gene variants**	**ttt/tt**	**F88**	**17.00**	**8.00**	**68.00**
AYO51_00615	**Hypothetical protein**	**Missense gene variants**	**gcattgctt/gcGTTGTtt**	**ALL20ALF**	**145.00**	**118.00**	**55.13**
AYO51_00615	**Hypothetical protein**	**Missense gene variants**	**Ctt/Ttt**	**L32F**	**143.00**	**140.00**	**50.53**
AYO51_00615	Hypothetical protein	Synonymous gene variants	ggC/ggT	G35	114.00	87.00	56.72
AYO51_04795	**Uncharacterized protein**	**Missense gene variants**	**Tat/Cat**	**Y5H**	**26.00**	**0.00**	**100.00**
bglA	Aryl-phospho-beta-D-glucosidase BglA	Frameshift gene variants	gcc/gc	A382	17.00	49.00	25.76
bglP	**PTS system, trehalose-specific IIBC component**	**Missense gene variants**	**Gca/Aca**	**A271T**	**21.25**	**8.75**	**70.83**
C0682_00850	Uncharacterized protein	Synonymous gene variants	ggT/ggC	G221	184.50	0.00	100.00
C0682_11360	**HlyC/CorC family transporter**	**Frameshift gene variants**	**aaa/aa**	**K200**	**38.00**	**24.00**	**61.29**
C0682_12350	**Serine/threonine protein kinase**	**Missense gene variants**	**Ccg/Tcg**	**P122S**	**223.00**	**0.00**	**100.00**
C0682_12600	**DUF2712 domain-containing protein**	**Frameshift gene variants**	**tca/tcAa**	**S26S**	**28.00**	**21.00**	**57.14**
C0682_13070	**DUF916 and DUF3324 domain-containing**	**Missense gene variant**	**gCg/gTg**	**A196V**	**55.00**	**0.00**	**100.00**

Table S4.2 continued

Gene name	Product	Mutation type	Mutation (base change) [REF/ALT]	Mutation (amino acid change)	Alternate allele observation count (AO)	Reference allele observation count (RO)	Frequency (%)
carB	Carbomoyl-phosphae synthase large chain	Synonymous gene variants	ggA/ggG	G16	60.00	0.00	100.00
cbf1	3'-5' exoribonuclease *yhaM*	Synonymous gene variants	ggG/ggA	G213	43.25	0.00	100.00
cbh	Choloylglycine hydrolase	Synonymous gene variants	ggA/ggG	G157	123.00	0.00	100.00
CFI62_15645	LysM domain-containing protein	Upstream gene variants	G/A		16.00	9.00	64.00
CFI62_15645	LysM domain-containing protein	Upstream gene variants	T/C	T189	22.00	8.50	72.13
cls	**Cardiolipin synthase**	**Missense gene variants**	**gAa/gGa**	**E218G**	**26.00**	**0.00**	**100.00**
cls	Cardiolipin synthase	Upstream gene variants	ATT/ATTT		33.75	31.75	51.53
clsA	**Major cardiolipin synthase ClsA**	**Missense gene variants**	**cGc/cAc**	**R189H**	**158.00**	**0.00**	**100.00**
clsA	**Major cardiolipin synthase ClsA**	**Frameshift gene variants**	**gcc/gc**	**A202**	**17.00**	**4.50**	**79.07**
CUR48_01495	**Uncharacterized protein**	**Missense gene variants**	**cTa/cCa**	**L103P**	**137.00**	**0.00**	**100.00**
CUR48_02310	**Choice-of-anchor A family protein**	**Missense gene variants**	**Atg/Gtg**	**M260V**	**23.00**	**0.00**	**100.00**
CUR48_02430	**MFS transporter**	**Missense gene variants**	**aTc/aCc**	**I206T**	**132.00**	**0.00**	**100.00**
CUR48_05135	**DUF806 domain-containing protein**	**Missense gene variants**	**tAc/tGc**	**Y52C**	**101.00**	**0.00**	**100.00**
CUR48_05825	**PAP2 family protein**	**Missense gene variants**	**gCg/gTg**	**A26V**	**110.00**	**0.00**	**100.00**
CUR48_07430	**DNA translocase FtsK**	**Frameshift gene variants**	**ttg/tg**	**L181**	**11.00**	**10.00**	**52.38**
CUR48_10370	**Gfo/Idh/MocA family oxidoreductase**	**Missense gene variants**	**cTg/cCg**	**L43P**	**102.50**	**0.00**	**100.00**
dap	**D-aminopeptidase**	**Missense gene variants**	**gCg/gTg**	**A109V**	**95.00**	**0.00**	**100.00**
ddl	**D-alanine--D-alanine ligase**	**Missense gene variants**	**gCc/gTc**	**A173V**	**44.00**	**0.00**	**100.00**
def	Peptide deformylase	Frameshift gene variants	gaa/ga	E77	6.00	14.00	30.00
dhaA	Alpha/beta hydrolase	Upstream gene variants	ATT/AT		21.00	17.00	55.26
dsdA	Probable D-serine dehydratase	Missense gene variants	tTg/tGg	L56W	5.00	13.00	27.78
ecfT	**Energy-coupling factor transporter transmembrane protein EcfT**	**Missense gene variants**	**aCt/aTt**	**T195I**	**149.00**	**52.00**	**74.13**
EQG53_13400	Helix-turn-helix domain-containing protein	Synonymous gene variants	atC/atT	I68	52.60	32.60	61.74
EQG58_16085	Carbohydrate ABC transporter permease	Upstream gene variants	CAA/CAAA		60.00	65.00	48.00
estA	Acetyl esterase (Promiscuous)	Frameshift gene variants	ttt/tt	F250	12.00	41.00	22.64
fabG1	3-oxoacyl-[acyl-carrier protein] reductase	Upstream gene variants	C/T		29.00	1.00	96.67

Table S4.2 continued

Gene name	Product	Mutation type	Mutation (base change) [REF/ALT]	Mutation (amino acid change)	Alternate allele observation count (AO)	Reference allele observation count (RO)	Frequency (%)
fus	Elongation factor G	Upstream gene variants	A/G		35.00	0.00	100.00
glnH	Glutamine ABC transporter, substrate binding protein	Upstream gene variants	A/G		59.00	0.00	100.00
glpK1	**Glycerol kinase 1**	**Missense gene variants**	**tAc/tGc**	**Y230C**	**66.50**	**0.00**	**100.00**
IV39_GL002393	**Lipoate--protein ligase**	**Missense gene variants**	**cGc/cAc**	**R122H**	**113.00**	**0.00**	**100.00**
larC	Pyridinium-3,5-bisthiocarboxylic acid mononucleotide nickel insertion protein	Synonymous gene variants	ggC/ggT	G19	90.50	0.00	100.00
lmrA	Multidrug resistance ABC transporter ATP-binding and permease protein	Synonymous gene variants	ggC/ggT	G308	190.50	0.00	100.00
lp_0083	**Transcription regulator, ArsR family**	**Missense gene variants**	**gAt/gGt**	**D62G**	**83.00**	**0.00**	**100.00**
lp_0118	DNA-binding protein with HIRAN domain	Synonymous gene variants	ctA/ctG	L157	59.00	0.00	100.00
lp_0209	RNA-binding protein, DUF814 family	Frameshift gene variants	accttt/actt	TF259	4.00	15.00	21.05
lp_0302	**Extracellular transglycosylase**	**Missense gene variants**	**Atc/Ttc**	**I165F**	**67.67**	**45.33**	**59.88**
lp_0302	Extracellular transglycosylase	Synonymous gene variants	aaG/aaA	K2	26.67	18.00	59.70
lp_0348	Multidrug-efflux transporter, major facilitator superfamily (MFS), EmrB family	Upstream gene variants	CTTTTTTC/CTTTTTC		20.00	7.00	74.07
lp_0383	Hypothetical membrane protein	Upstream gene variants	G/T		82.00	0.00	100.00
lp_0475	Hydrolase, HAD superfamily, Cof family	Upstream gene variants	A/G		71.00	0.00	100.00
lp_0691	Cytokinin riboside 5'-monophosphate phosphoribohydrolase	Upstream gene variants	TG/TGCG		28.00	9.00	75.68
lp_1556	Metal-dependent phosphohydrolase, HD family	Missense gene variants	gTg/gAg	V96E	9.00	21.00	30.00
lp_1915	Lipoprotein	Upstream gene variants	T/C		137.00	0.00	100.00
lp_2745	Nucleotide-binding protein, universal stress protein UspA family	Upstream gene variants	ACG/AG		9.00	6.00	60.00
lp_2817	**Transcription regulator of multidrug-efflux transporter**	**Frameshift gene variants**	**gat/ga**	**D70**	**42.00**	**31.00**	**57.53**
lp_2929	**Diguanylate cyclase/phosphodiesterase, GGDEF domain**	**Missense gene variants**	**gAt/gGt**	**D127G**	**97.00**	**0.50**	**99.49**
lp_3042	Multidrug ABC transporter, ATP-binding and permease protein	Synonymous gene variants	cgC/cgT	R195	115.00	0.00	100.00
lp_3411	**Extracellular protein, DUF1002 family**	**Missense gene variants**	**Acc/Gcc**	**T237A**	**16.00**	**0.00**	**100.00**
lp_3425	**Uncharacterized protein**	**Missense gene variants**	**gAt/gGt**	**D59G**	**224.00**	**0.00**	**100.00**
Lp19_2747	**Uncharacterized protein**	**Missense gene variants**	**Att/Gtt**	**I71V**	**167.00**	**146.00**	**53.35**
LpLQ80_06975	Serine/threonine phosphatase stp	Synonymous	gcC/gcT	A104	25.67	0.00	100.00

Table S4.2 continued

Gene name	Product	Mutation type	Mutation (base change) [REF/ALT]	Mutation (amino acid change)	Alternate allele observation count (AO)	Reference allele observation count (RO)	Frequency (%)
LpLQ80_09265	Uncharacterized protein	Upstream gene variants	C/T		144.00	151.00	48.81
menH	**2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate synthase**	**Missense gene variants**	**aTc/aCc**	**I190T**	**121.67**	**0.00**	**100.00**
morA	**Morphine 6-dehydrogenase**	**Missense gene variants**	**tAt/tGt**	**Y139C**	**73.50**	**0.00**	**100.00**
mutL	DNA mismatch repair protein	Upstream gene variants	T/G		30.00	0.00	100.00
mutL	DNA mismatch repair protein	Upstream gene variants	A/C		79.60	0.00	100.00
nagA	N-acetylglucosamine-6-phosphate deacetylase	Frameshift gene variants	aaa/aa	K346	11.50	25.00	31.51
napA3	**Na(+)/H(+) antiporter**	**Missense gene variants**	**gTc/gCc**	**V316A**	**38.50**	**0.00**	**100.00**
Nizo1839_0932	DedA protein	Upstream gene variants	T/C		1007.50	0.00	100.00
Nizo2802_0180	Integral membrane protein	Synonymous gene variants	ggC/ggT	G230	132.50	0.00	100.00
Nizo2802_0857	**Uncharacterised protein**	**Missense gene variants**	**Gta/Ata**	**V48I**	**86.40**	**41.80**	**67.39**
Nizo2802_2173	**Uncharacterised protein**	**Missense gene variants**	**gCc/gTc**	**A25V**	**157.00**	**156.67**	**50.05**
Nizo2802_2173	Uncharacterised protein	Frameshift gene variants	ctt/cTtt	L28L	7.00	8.00	46.67
Nizo2802_2946	DedA protein	Synonymous gene variants	ttC/ttT	F96	104.00	35.50	74.55
pck	Phosphoenolpyruvate carboxykinase (ATP)	Frameshift gene variants	tta/tt1	L69	7.00	21.00	25.00
pepF1	**Oligoendopeptidase F**	**Missense gene variants**	**gGt/gTt**	**G358V**	**107.00**	**0.00**	**100.00**
pglH	**GalNAc-alpha-(1->4)-GalNAc-alpha-(1->3)-diNAcBac-PP-undecaprenol alpha-1,4-N-acetyl-D-galactosaminyl transferase**	**Missense gene variants**	**cGg/cAg**	**R44Q**	**34.00**	**0.33**	**99.03**
pox5	Pyruvate oxidase	Synonymous gene variants	gcG/gcA	A361	169.50	2.00	98.83
pts21A	**PTS system, EIIA component**	**Missense gene variants**	**gTg/gCg**	**V3A**	**32.00**	**0.00**	**100.00**
pts21A	PTS system, EIIA component	Upstream gene variants	TAA/TA		51.00	19.00	72.86
ptsG_2	PTS system glucose-specific EIICBA component	Synonymous gene variants	ttA/ttG	L137	58.00	0.00	100.00
purF	**Amidophosphoribosyltransferase**	**Missense gene variants**	**Gca/Aca**	**A151T**	**145.00**	**0.00**	**100.00**
purK1	N5-carboxyaminoimidazole ribonucleotide synthase	Synonymous gene variants	tcG/tcA	S137	26.25	0.00	100.00
rhaR	**Transcription regulator, AraC family, GlcNAc-like induced**	**Missense gene variants**	**gAc/gGc**	**D113G**	**15.50**	**0.00**	**100.00**
rnhA	Ribonuclease HI	Frameshift gene variants	tta/tt2	L131	8.00	22.00	26.67
rodA2	Rod-shape determining protein	Synonymous	ggA/ggG	G147	79.00	0.00	100.00

Table S4.2 continued

Gene name	Product	Mutation type	Mutation (base change) [REF/ALT]	Mutation (amino acid change)	Alternate allele observation count (AO)	Reference allele observation count (RO)	Frequency (%)
smc	Chromosome partition protein	Upstream gene variants	G/C		20.33	1.00	95.31
tarL	**Techoic acid poly(ribitol-phosphate) polymerase**	**Frameshift gene variants**	**tta/ttTa**	**L422F**	**35.00**	**22.00**	**61.40**
topA	DNA topoisomerase 1	Upstream gene variants	CAAAAAAAG/CAAAAAAAAG		22.00	13.00	62.86
truA	**tRNA pseudouridine synthase A**	**Frameshift gene variants**	**ttg/tt**	**L70**	**11.00**	**11.00**	**50.00**
trxA1	**Thioredoxin**	**Missense gene variants**	**aAg/aCg**	**K108T**	**27.00**	**0.00**	**100.00**
uvrA	UvrABC system protein A	Synonymous gene variants	ttA/ttG	L6	84.00	0.00	100.00
uvrC	UvrABC system protein C	Upstream gene variants	C/A		21.25	1.25	94.44
WP50_37325	**Uncharacterized protein**	**Missense gene variants**	**gAc/gGc**	**D43G**	**72.00**	**0.00**	**100.00**
xerC	**Tyrosine recombinase *xerC***	**Missense gene variants**	**Acg/Gcg**	**T336A**	**61.00**	**0.00**	**100.00**
yfkN	**Trifunctional nucleotide phosphoesterase protein YfkN**	**Missense gene variants**	**Cgc/Tgc**	**R184C**	**144.67**	**0.00**	**100.00**
yjaB	**Putative N-acetyltransferase YjaB**	**Missense gene variants**	**Agt/Ggt**	**S132G**	**122.00**	**0.00**	**100.00**

Table S4.3 All affected genes annotated as hypothetical proteins by Prokka. These were confirmed using NCBI BLAST, and the 'new' annotation is shown.

Gene name	Original annotation	BLAST annotation
A0A165S4L8	Hypothetical Protein	Uncharacterized protein
A0U96_14875	Hypothetical Protein	Uncharacterized protein
AVR82_04290	Hypothetical Protein	Aquaporin
AVR82_04325	Hypothetical Protein	Cell surface protein
AVR82_05345	Hypothetical Protein	Integral membrane protein
AVR82_06480	Hypothetical Protein	NADP oxidoreductase
AVR82_07550	Hypothetical Protein	Cytoskeleton protein RodZ
AVR82_08035	Hypothetical Protein	Glucose uptake protein GlcU
AVR82_11000	Hypothetical Protein	MurR/RpiR family transcriptional regulator
AVR82_12850	Hypothetical Protein	Putative phosphotransferase YtmP
AVR82_13025	Hypothetical Protein	Hypothetical protein
AVR82_13160	Hypothetical Protein	Cation efflux protein
AVR82_16315	Hypothetical Protein	cell surface SD repeat protein precursor, membrane-anchored
L_03143	Hypothetical Protein	Hypothetical protein
AYO51_01645	Hypothetical Protein	Uncharacterized protein
AYO51_02915	Hypothetical Protein	Uncharacterized protein
AYO51_04665	Hypothetical Protein	Putative transcriptional regulator
AYO51_04795	Hypothetical Protein	Uncharacterized protein
AYO51_08145	Hypothetical Protein	conjugal transfer protein TraG
AYO51_14370	Hypothetical Protein	Nickel transporter
BIZ32_01220	Hypothetical Protein	ATP-dependent helicase/deoxyribonuclease subunit B
BIZ32_05665	Hypothetical Protein	L-lysine permease
L_03120	Hypothetical Protein	Membrane protein
BIZ32_11210	Hypothetical Protein	PTS lactose/cellobiose transporter subunit IIA
BIZ32_12025	Hypothetical Protein	Uncharacterized protein
BV241_15255	Hypothetical Protein	Hypothetical protein
BV299_01050	Hypothetical Protein	VOC family virulence protein
C0682_00850	Hypothetical Protein	Uncharacterized protein
C0682_05590	Hypothetical Protein	Glycosyl transferase family 1
C0682_05720	Hypothetical Protein	HTH-type transcriptional regulator AdhR
C0682_11360	Hypothetical Protein	HlyC/CorC family transporter
C0682_12350	Hypothetical Protein	Serine/threonine protein kinase
C0682_12600	Hypothetical Protein	DUF2712 domain-containing protein
C0682_13070	Hypothetical Protein	DUF916 and DUF3324 domain-containing protein
C0682_14530	Hypothetical Protein	GNAT family N-acetyltransferase
C5Z26_10595	Hypothetical Protein	Uncharacterized protein

Table S4.3 continued

Gene name	Original annotation	BLAST annotation
C6Y09_01835	Hypothetical Protein	Uncharacterized protein
C6Y10_13250	Hypothetical Protein	Hypothetical protein
CFI62_03005	Hypothetical Protein	Hypothetical protein
CFI62_10080	Hypothetical Protein	TetR/AcrR family transcriptional regulator
L_00873	Hypothetical Protein	Membrane protein PlnW
CFM86_15580	Hypothetical Protein	IS30 family transposase
L_03243	Hypothetical Protein	Hypothetical protein
CUR48_01495	Hypothetical Protein	Uncharacterized protein
CUR48_02310	Hypothetical Protein	Choice-of-anchor A family protein
CUR48_02430	Hypothetical Protein	MFS transporter
CUR48_05135	Hypothetical Protein	DUF806 domain-containing protein
CUR48_05825	Hypothetical Protein	PAP2 family protein
CUR48_07430	Hypothetical Protein	DNA translocase FtsK
CUR48_07570	Hypothetical Protein	Uncharacterized protein
CUR48_10370	Hypothetical Protein	Gfo/Idh/MocA family oxidoreductase
CUR48_14360	Hypothetical Protein	DUF536 domain-containing protein
EQG58_08925	Hypothetical Protein	Uncharacterized protein
EQG58_13400	Hypothetical Protein	Helix-turn-helix domain-containing protein
EQG58_16085	Hypothetical Protein	Carbohydrate ABC transporter permease
EQJ86_02090	Hypothetical Protein	Mutator family transposase
EQJ86_05285	Hypothetical Protein	Uncharacterized protein
EQJ92_04985	Hypothetical Protein	Glucose-starvation inducible protein B
FC81_GL002113	Hypothetical Protein	Uncharacterized protein
FC92_GL001975	Hypothetical Protein	helix-turn-helix domain-containing protein
IV39_GL001561	Hypothetical Protein	Uncharacterized protein
IV39_GL002161	Hypothetical Protein	Integral membrane protein
IV39_GL002393	Hypothetical Protein	Lipoate--protein ligase
IV39_GL002658	Hypothetical Protein	Tryptophanyl-tRNA synthetase II
IV39_GL002953	Hypothetical Protein	Acetyltransferase, GNAT family
LAP9434_02799	Hypothetical Protein	Uncharacterized protein
LAP9492_01753	Hypothetical Protein	2-haloacrylate reductase
lp_0083	Hypothetical Protein	Transcription regulator, ArsR family
lp_0118	Hypothetical Protein	DNA-binding protein with HIRAN domain
lp_0209	Hypothetical Protein	RNA-binding protein, DUF814 family
lp_0287	Hypothetical Protein	Hypothetical protein
L_00288	Hypothetical Protein	Extracellular transglycosylase
lp_0348	Hypothetical Protein	Multidrug-efflux transporter, major facilitator superfamily (MFS), EmrB family

Table S4.3 continued

Gene name	Original annotation	BLAST annotation
lp_0475	Hypothetical Protein	Hydrolase, HAD superfamily, Cof family
lp_0533	Hypothetical Protein	Transport protein, QueT family
lp_0691	Hypothetical Protein	Cytokinin riboside 5'-monophosphate phosphoribohydrolase
lp_0723	Hypothetical Protein	Putative ABC transporter ATP-binding protein
lp_0780	Hypothetical Protein	Putative gluconeogenesis factor
lp_0862	Hypothetical Protein	Uncharacterized protein
lp_0960	Hypothetical Protein	Hypothetical protein
lp_0988	Hypothetical Protein	Extracellular lipoprotein, Asp-rich
lp_1336	Hypothetical Protein	Bifunctional protein: ABC transporter, ATP-binding protein transcription regulator, LytR family
L_02894	Hypothetical Protein	Metal-dependent phosphohydrolase, HD family
lp_1915	Hypothetical Protein	Lipoprotein
lp_2488g	Hypothetical Protein	Uncharacterized protein
lp_2534	Hypothetical Protein	Protein DegV
lp_2745	Hypothetical Protein	Nucleotide-binding protein, universal stress protein UspA family
lp_2768	Hypothetical Protein	Transport protein, major facilitator superfamily (MFS)
lp_2797	Hypothetical Protein	Uncharacterized protein
lp_2799	Hypothetical Protein	Amino acid transport protein
lp_2817	Hypothetical Protein	Transcription regulator of multidrug-efflux transporter
lp_2847	Hypothetical Protein	Extracellular transglycosylase, with LysM peptidoglycan binding domain
lp_2925	Hypothetical Protein	Cell surface protein, LPXTG-motif cell wall anchor
lp_2926	Hypothetical Protein	Uncharacterized protein
lp_2929	Hypothetical Protein	Diguanylate cyclase/phosphodiesterase, GGDEF domain
lp_3042	Hypothetical Protein	Multidrug ABC transporter, ATP-binding and permease protein
lp_3098	Hypothetical Protein	NAD-dependent epimerase/dehydratase family protein
lp_3259	Hypothetical Protein	Zinc-dependent proteinase
lp_3323	Hypothetical Protein	Lipase/esterase, SGNH or GDSL hydrolase family
lp_3411	Hypothetical Protein	Extracellular protein, DUF1002 family
lp_3425	Hypothetical Protein	Uncharacterized protein
lp_3491	Hypothetical Protein	FMN-binding protein
Lp19_0430	Hypothetical Protein	Nickase
Lp19_1149	Hypothetical Protein	Uncharacterized protein
Lp19_2642	Hypothetical Protein	DL-alanine permease SerP2
L_03214	Hypothetical Protein	Uncharacterized protein

Table S4.3 continued

Gene name	Original annotation	BLAST annotation
LPJSA22_00247	Hypothetical Protein	PTS beta-glucoside transporter subunit IIABC
LPJSA22_01163	Hypothetical Protein	Uncharacterized protein
LPJSA22_01565	Hypothetical Protein	N-acetylmuramoyl-L-alanine amidase
LPJSA22_02223	Hypothetical Protein	Uncharacterized protein
LPJSA22_02902	Hypothetical Protein	Uncharacterized protein
LpLQ80_06975	Hypothetical Protein	Serine/threonine phosphatase stp
L_03213L_03242	Hypothetical Protein	Uncharacterized protein
LpLQ80_13370	Hypothetical Protein	Cell surface protein
LpLQ80_14210	Hypothetical Protein	Hypothetical protein
Nizo1839_0932	Hypothetical Protein	DedA protein
Nizo1839_1169	Hypothetical Protein	ATP-dependent Clp protease ATP-binding subunit
Nizo2802_0180	Hypothetical Protein	Integral membrane protein
Nizo2802_0857	Hypothetical Protein	Uncharacterised protein
L_03211	Hypothetical Protein	Holin
Nizo2802_2946	Hypothetical Protein	DedA protein
Q8KT06	Hypothetical Protein	Transposase
WP50_37325	Hypothetical Protein	Uncharacterized protein

Supplementary figures

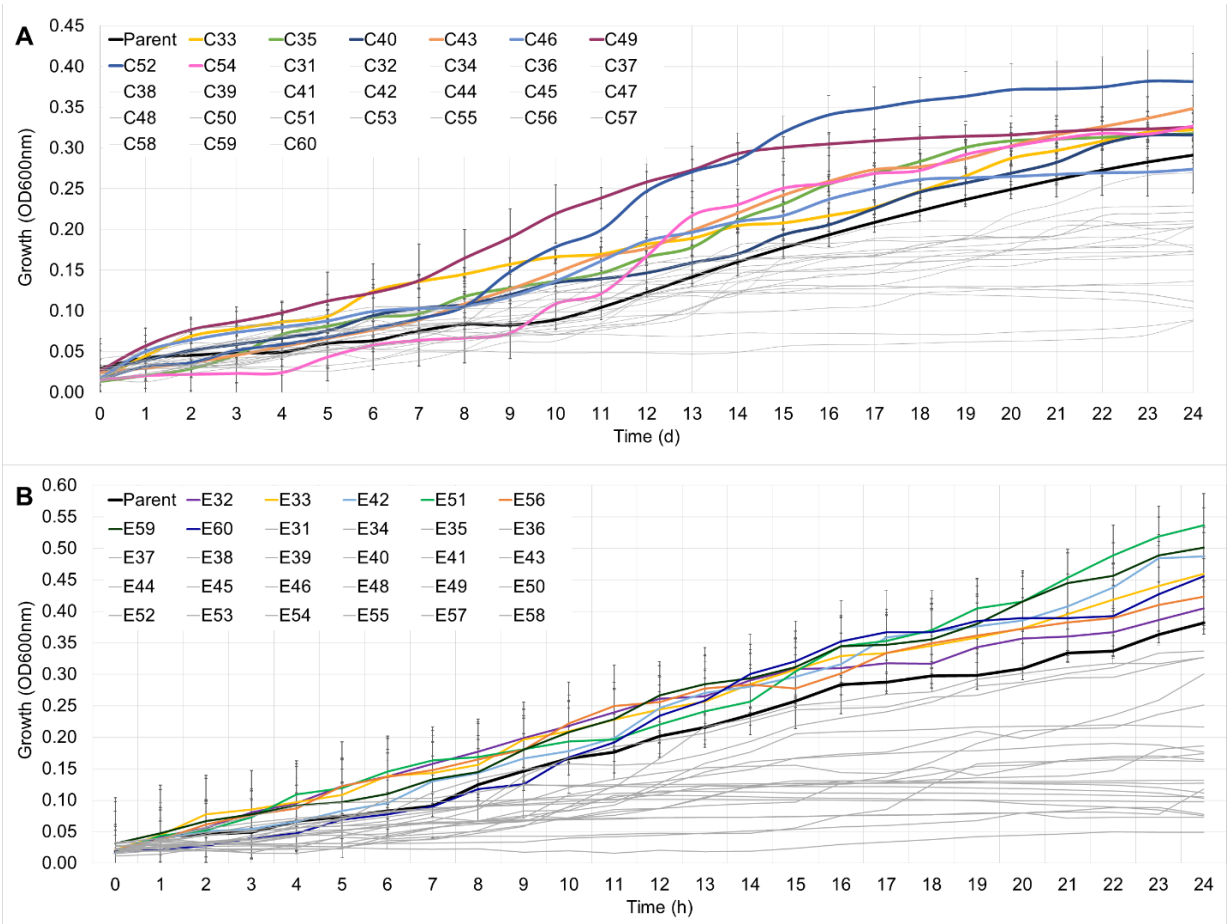

Figure S4.1 Microtitre plate screening of evolved isolates after 100 generations in SGM fermented with Cross Evolution (A) and EC1118 (B). The isolates were selected from the population that was evolved in the presence of the respective yeast strains. All isolates with a higher biomass compared to the parent strain (black line) are shown in colour. All other isolates are shown as grey lines. The error bars represent standard deviation. All data is the average of 3 biological repeats and 6 technical repeats.

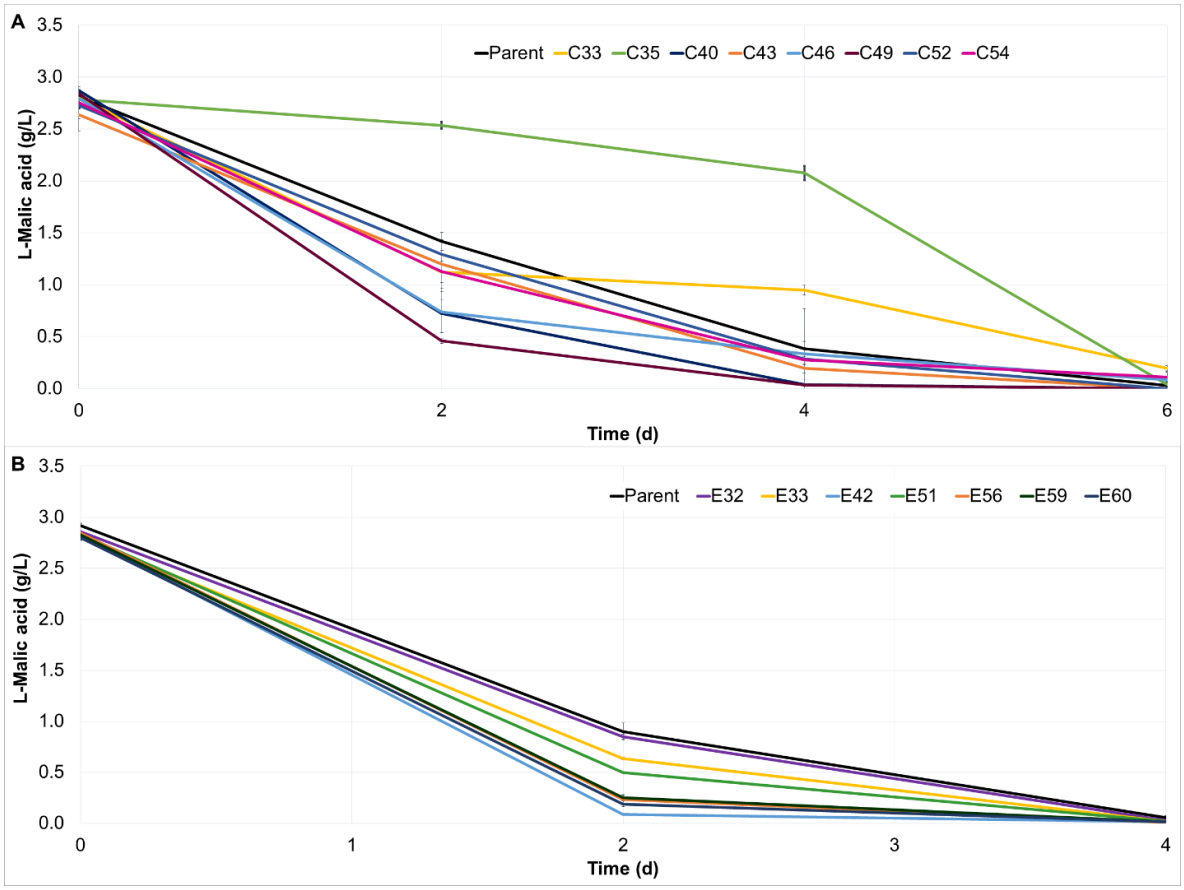

Figure S4.2 Residual L-malic acid in 15 ml SGM co-inoculated with each of the selected isolates from the microtitre plate experiments and the driver yeast Cross Evolution (A) and EC1118 (B). The parent strain is shown in black and each of the evolved isolates are coloured. All data shows the average of 3 biological repeats and error bars indicate standard deviation.

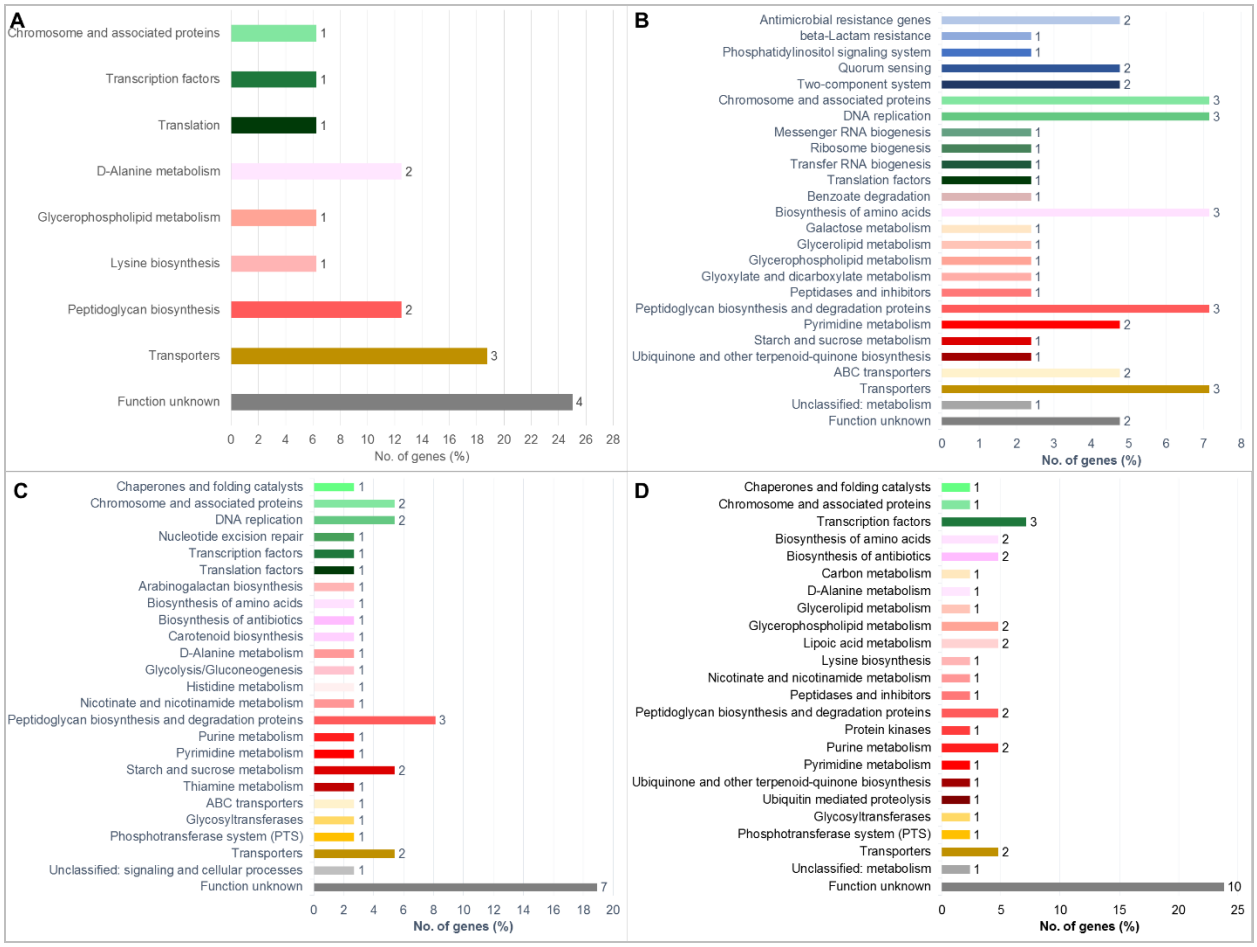

Figure S4.4 Distribution of affected genes into KEGG categories. A) G50C, B) G50E, C) G100C, and D) G100E. The x-axis shows the percentage number of genes in the particular group that fit the assigned category, while the data labels shows the actual gene count. Categories coloured according to similarity.

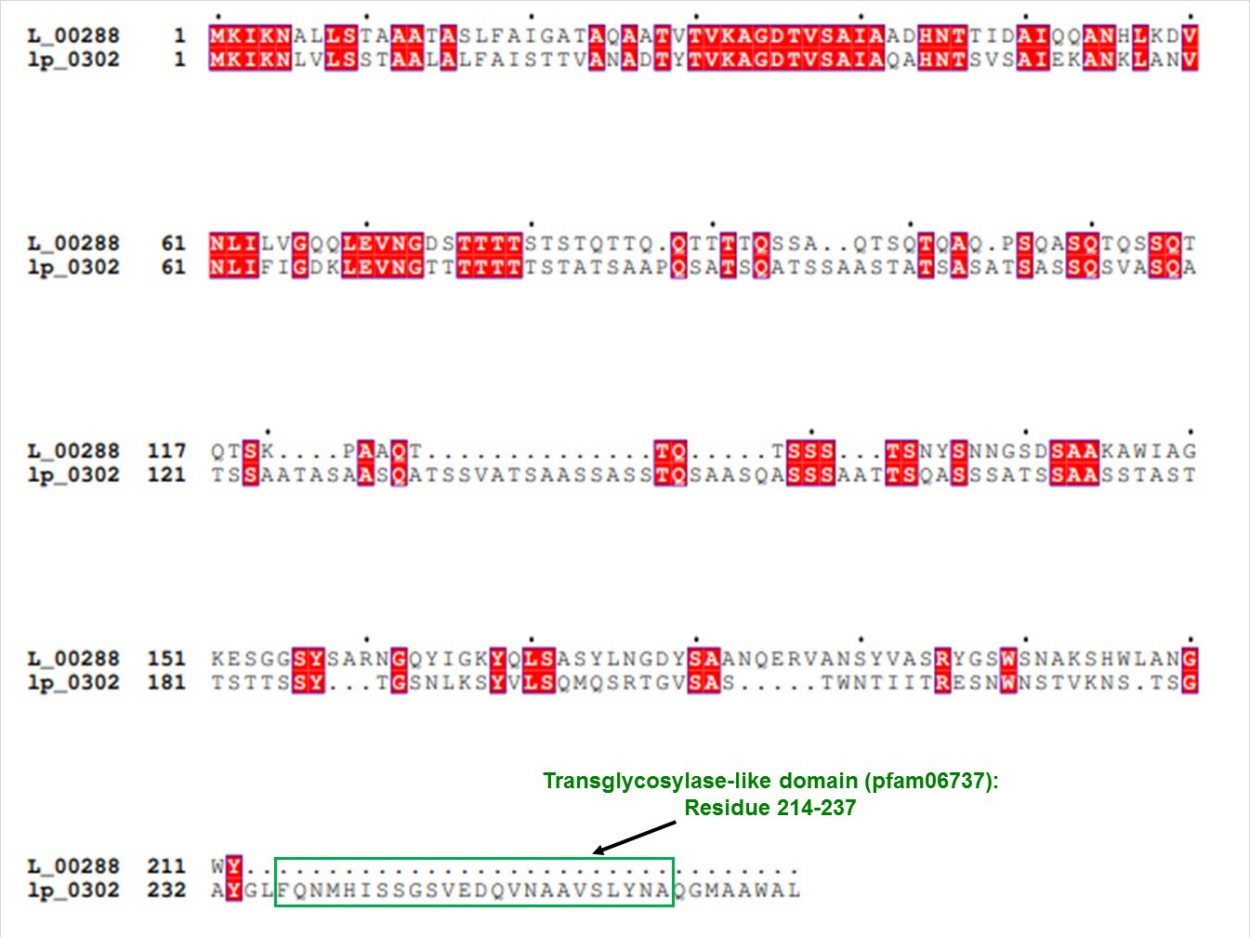

Figure S4.5 Sequence alignment of the L_00288 protein sequence of the parent strain used in this study and the lp_0302 extracellular transglycosylase of *L. plantarum* WCFS1. When L_00288 DNA sequence was searched using BLAST, the sequence identity was 58.49% (E-value: 3e-68). The green box shows the residues that make up the transglycosylase-like domain (pfam06737), which do not appear in the parent strain.

Chapter 5

Going too deep? Downstream sequencing assembly errors in whole genome sequencing using the Ion Proton System

Chapter 5: Going too deep? Downstream sequencing assembly errors in whole genome sequencing using the Ion Proton System

5.1. Abstract

Next generation sequencing (NGS) has taken the world by storm. Rapid declines in the cost of sequencing has made this technology accessible to many commercial as well as research laboratories. Additionally, fast turnaround times and thus high throughput has made NGS the technology of choice in most laboratories. It is a decade since the launch of NGS technology, yet many of its platforms are still plagued by high sequencing errors (0.1-1%). Therefore, several techniques have been employed to counteract this, such as deep and ultra-deep sequencing, which is sequencing at >100x and >1000x, respectively. This technique was shown to not be ideal for small genomes in spite of its popular use. Data show that the deeper one goes; the larger errors are discovered during the assembly and variant call analyses. Conversely, other works have shown that deep sequencing results in high accuracy base calls. Therefore, I applied deep and ultra-deep whole genome sequencing on *Lactobacillus plantarum* IWBT B063 using the Ion Proton sequencer to compare the genome assembly and variant analysis results at different depths. I sequenced two samples at 300x and 4000x depth: the 300x dataset was subsampled to 50x and 100x; and the 4000x dataset was subsampled to 800x to generate a total of 5 datasets. The data obtained showed that at a depth of 300x the number of contigs and number of indels (insertion/deletion) per 100 kbp are reduced, while the N50/NGA50 are higher. Moreover, the number of variants increased with increasing coverage, except there were significantly less indels at 4000x coverage. This data is significant because it suggests that a lower read depth (<300x) is adequate to generate a 'good' assembly and accurately call variants, especially for small genomes such as *L. plantarum*.

5.2. Introduction

With the advent of Sanger sequencing came the need to develop cost-effective, high throughput and time-efficient sequencers (Lander *et al.*, 2001). For almost two decades, work has been done to develop new high throughput sequencing technology that is fast and cheap, referred to as Next Generation Sequencing (NGS) (Quail *et al.*, 2012). About a decade ago, a few new sequencing platforms were made available for commercial use: Illumina MiSeq, PacBio RS, Ion Torrent Personal Genome Machine (PGM) and Ion Proton (Quail *et al.*, 2012; Rhoads & Au, 2015; Damiati *et al.*, 2016). These platforms produce short reads (100 bp up to 60 kbp) and high sequencing error (0.1-1%) (Quail *et al.*, 2012; Rhoads & Au, 2015; Besser *et al.*, 2018). Moreover, it was found that NGS technologies suffer sequence coverage bias, which is the uneven depth of sequencing coverage of reads across the genome (Ross *et al.*, 2013). Taken together, these shortcomings

affect genome assembly and downstream analyses, resulting in false-positive results (Ross *et al.*, 2013; Abnizova *et al.*, 2017).

To reduce or correct sequencing errors, high depth sequencing has been applied to both large (60x - ≥100x) and smaller (≥100x) genomes (Desai *et al.*, 2013; Kishikawa *et al.*, 2019). Deep (>100x) and ultra-deep (≥1000x) sequencing has been used to identify rare mutations in the genomes of small disease-causing viruses and bacteria (see review by McElroy *et al.*, 2014). Deep sequencing can be used on available NGS technology like the Ion Proton system (ThermoFisher Scientific, Waltham, MA) which uses a semiconductor technology, whereby hydrogen ions (H^+) are released when a nucleotide base is incorporated into a strand of DNA. The change in pH resulting from the released H^+, is converted to voltage by an ion sensor, indicating the incorporated nucleotide (Merriman *et al.*, 2012). The Ion Proton sequencer's main attraction is its cost-effectiveness, fast turnaround time, and accurate single nucleotide variant (SNV) calls (Damiati *et al.*, 2016; Gampawar *et al.*, 2019).

Semiconductor technology is limited in that the quality around GC content extremes and high polymer regions is lower due to the ineffective detection of the measured voltage in the ion sensor when same-nucleotide repeats are incorporated in the DNA strand, resulting in lower coverage at these sites (Lahens *et al.*, 2017; Besser *et al.*, 2018). Therefore, Ion Torrent –based sequencing is known to overestimate indel mutations (Churchill *et al.*, 2016; Seo *et al.*, 2017). This limitation in Ion Torrent sequencing remains a challenge for downstream analysis, especially for the identification of variants impacting gene/protein function (Ross *et al.*, 2013; Laehnemann *et al.*, 2016). Therefore, Ion Torrent released the Hi-Q™ Sequencing Chemistry (ThermoFisher Scientific) which was shown to improve accuracy at homopolymeric sites (Churchill *et al.*, 2016).

Some authors have suggested using ultra-deep and deep sequencing to counteract the sequencing error of sequencers, yet others have shown that ultra-deep sequencing increases errors and results in 'bad' genome assemblies (Fang *et al.*, 2014; Lonardi *et al.*, 2015; Mirebrahim *et al.*, 2015; Fujita *et al.*, 2017). In fact, in previous work it was shown that sequencing at >1000x depth increases genome assembly error regardless of the genome assembler used (Lonardi *et al.*, 2015). Interestingly, it was also shown that the SPAdes genome assembler was less sensitive to high sequencing error and high coverage (Mirebrahim *et al.*, 2015). In both studies bacterial artificial chromosomes (BACs) and simulated datasets were used for all analyses. There is only one study (at the time of writing) which investigates the impact of deep sequencing (410x) of whole genome sequencing (WGS) data on downstream variant calling (Kishikawa *et al.*, 2019), however this study focused on the large human genome (3 234.83 Mb). Therefore, I aim to compare genome assembly and variant calling results of 5 whole genome sequencing (WGS) datasets of

Lactobacillus plantarum IWBT B063 (3.37 Mb) obtained by Ion Proton deep (>300x) and ultra-deep (>4000x) sequencing. Each of these samples were subsampled to generate 50x and 100x (300x), and 800x depth (4000x) reads. The data presented here shows that in as much as SPAdes genome assembly is less affected by high coverage and high sequencing error, compared to other genome assemblers, downstream variant calling presents with a large number of false positive indel calls and a large number of potentially erroneous SNP (single nucleotide polymorphism) calls at ultra-deep sequencing, suggesting that ultra-deep sequencing is not forcibly a good option for whole genome sequencing of small genomes.

5.3. Materials and Methods

5.3.1. Bacterial samples and DNA extraction

The *L. plantarum* IWBT B063 strain used in this study was obtained from the IWBT collection (Institute for Wine Biotechnology, Stellenbosch University, South Africa). Genomic DNA extraction was conducted using the method by (Lewington *et al.*, 1987).

5.3.2. Ion Torrent sequencing

Ion Torrent sequencing was conducted at the Central Analytical Facilities: Next Generation Sequencing Unit (Stellenbosch University, South Africa). Two samples were sequenced at >300x and >4000x depth. An adapter-ligated library was prepared from 100 ng gDNA using the Ion Plus Fragment Library Kit according to the protocol (ThermoFischer Scientific). The adapter-ligated, barcoded library was purified with Agencourt™ AMPure™ XP reagent (Beckmann Coulter). Following purification, the size-selected library fragments were quantified using the Ion Universal Library Quantitation Kit (ThermoFischer Scientific) and the StepOnePlus™ Real-time PCR system (ThermoFisher Scientific). Libraries were diluted to a target concentration of 100 pM. The diluted, barcoded, libraries were combined in equimolar amounts for template preparation using the Ion PI™ Hi-Q™ OT2 200 Kit (ThermoFisher Scientific). Pooled library samples were prepared for emulsion amplification and template enrichment using the Ion OneTouch™ 2 System and MyOne Streptavidin C1 beads (both from ThermoFischer Scientific) according to the manufacturer's protocol. Ion sphere particles matching positively with the template were enriched and sequenced on the Ion Proton sequencer using Ion PI™ Chip (v3).

5.3.3. Genome assembly and variant calling

All data from the Ion Proton run were analyzed on the Galaxy software framework (Afgan *et al.*, 2018). To begin with reads with 300x coverage were subsampled to ~30% and ~15%; and reads with 4000x coverage were subsampled to ~20% of the total reads. This was done to 'artificially'

remove excessive coverage, post sequencing, resulting in 5 datasets of coverage 50x, 100x, 300x, 800x, and 4000x. To remove adapter sequences and poor-quality reads (Q<20) a sliding window trimming was performed on the reads using default setting in Trimmomatic (Version 0.36.0). Trimmed reads were aligned to *L. plantarum* WCFS1 (GenBank accession AL935263) with BWA-MEM (Galaxy Version 0.36.6). FastQC Read Quality Reports was used to determine sequence quality.

De *novo* assemblies of reads from the 5 datasets were performed with SPAdes genome assembler (Galaxy version 1.0) to assemble genomes. The default settings were used for the assembly. A QUAST analysis was performed to generate summary statistics for the genome assemblies (Gurevich *et al.*, 2013). The N50 (the length to which adding contigs of equal or longer lengths sum up to half the length of the assembly), NGA50 (minimum contig length required that using equal or longer sized contigs sum up to 50% of the *L. plantarum* WCFS1 genome length), number of contigs, the number of misassemblies, and the percent indels per 100 kbp were used to determine the quality of the assembly. Variant calls were made with FreeBayes (Galaxy version 1.1.046-0) using default settings, except the ploidy was set to 1 and left alignment of indels was turned on (Garrison & Marth, 2012). The variants were filtered based on QUAL>20 and DP>15 using VCFilter (Galaxy version 1.0.0_rc1+galaxy2; (Song *et al.*, 2016)). Moreover, all variants that occurred on sequences mapping to the plasmid sequences of the reference genome were filtered manually and comparisons were made for variant calling.

5.3.4. Using 300x and 4000x datasets to generate draft genomes to be used as reference genomes for evolved isolates

I assembled the reads from the 300x and 4000x datasets as described above. These were then annotated using *L. plantarum* WCFS1 as the reference genome on Prokka (Version 1.11.0) with default settings. Then, previously evolved isolates (Chapter 3 of this dissertation) of the *L. plantarum* IWBT B063 strain were sequenced and analysed (Chapter 4 of this dissertation). The reads for each of the evolved isolates were mapped to the 300x and 4000x genomes (obtained from Prokka) using Bowtie2 (version 2.3.4.2). Thereafter, the aligned reads were then assembled using SPAdes (with default settings). Genome assembly quality was assessed using QUAST. Each of the assembled genomes were annotated using Prokka and the 300x and 4000x datasets were used as reference genomes.

5.4. Results

I applied WGS to a strain of *L. plantarum* IWBT B063 which was sequenced at two sequencing depths (>300x and >4000x). Previous work suggests that deep sequencing and ultra-deep

sequencing may be useful in reducing the number of sequencing errors that normally occur with next generation sequencing (McElroy *et al.*, 2014; Mirebrahim *et al.*, 2015; Kishikawa *et al.*, 2019). In this study I compare the post-sequencing results of a genome sequenced at a deep and ultra-deep coverage using the Ion Proton sequencer to see if there are major differences in data. I subsampled the 300x reads to a coverage of 50x and 100x; and the 4000x reads to a coverage of 800x, to have 5 sequencing depths for analysis: 50x, 100x, 300x, 800x, and 4000x.

5.4.1. Sequence read quality

Reads were trimmed and the quality of the reads was measured. Figure 5.1 shows the Phred scores of all bases in each dataset. The Phred score distribution of 50x-300x reads is skewed towards base calls of high quality, with 82.93-82.99% of them having a score ≥25 (Figure 5.1). There was no difference between the Phred score distribution at 800x and 4000x depth, with 47.41-47.42% of reads being of high quality (Figure 5.1). This data shows that >50% of the base calls at 800x and 4000x depth will be discarded with downstream analysis which shows the amount of wasted resources incurred with ultra-deep sequencing (Daley & Smith, 2013).

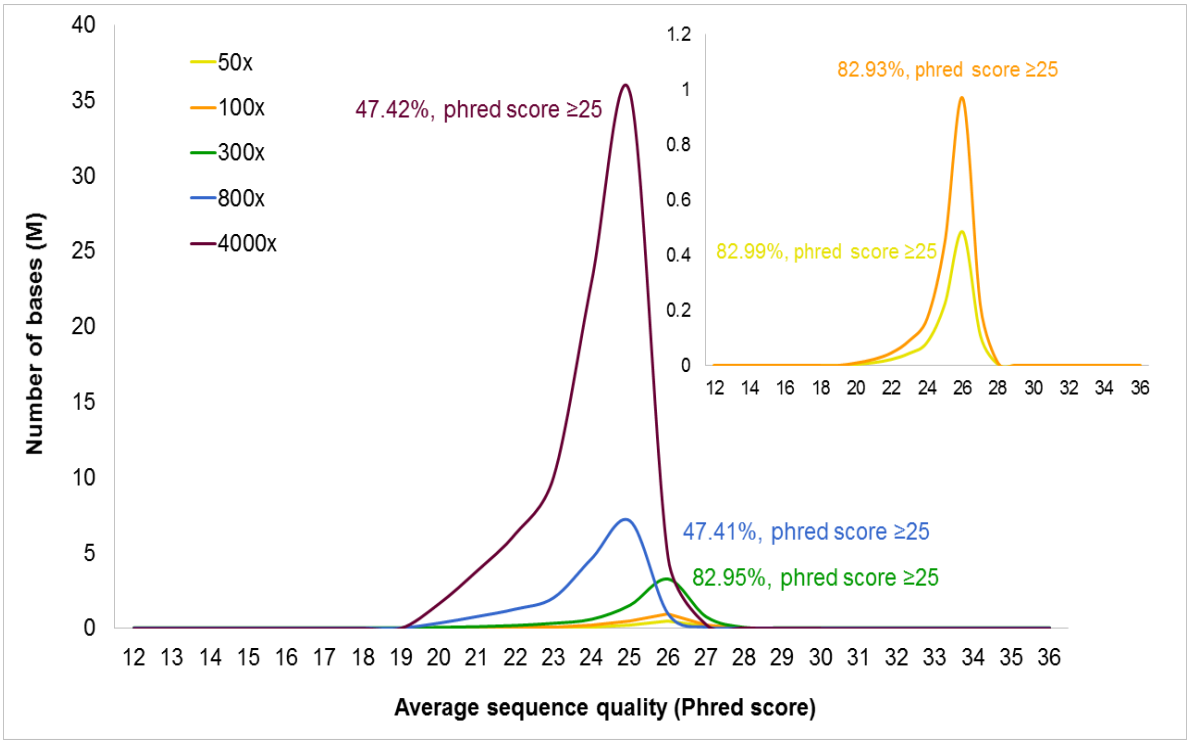

Figure 5.1 Per sequence quality score plot for 300x (green) depth and 4000x (plum) depth downsampled to 50x (yellow), 100x (orange), and 800x (blue). The percentage total number of bases with a Phred score ≥25 (base call accuracy 99.996%) is shown for each curve.

5.4.2. De novo assembly

To compare the impact of read depth on genome assembly, I performed de novo assemblies at the different read depths (50x-4000x). The data show that the best assembly was obtained at 300x read depth (Table 5.1) as indicated by the duplication ratio (1.003), number of contigs (147), largest contig (258 434), N50 (60 488), and NGA50 (27 790). Moreover, at 300x depth the proportion of predicted indels per 100 kbp is lower (25.92%) than at all other read depths (Table 5.1). However, significant differences were observed for the total length (especially between 50x and all other depths), number of contigs and the proportion of indels (chi square p-value < 0.0001). This data suggests that at 300x depth, the number of contigs is reduced significantly ($p < 0.0001$) meaning that the majority of contigs are of longer lengths. The need for fewer contigs is important to reduce gaps in the genome and reduce possible assembly errors (Kuśmirek *et al.*, 2019). The number of genes covered were not significantly different (chi square p-value = 0.058) for depths ≥100x to 4000x (2590-2597), although at 4000x there were slightly more (2597), which suggests that at ultra-deep sequencing some genomic features are detected which would otherwise be missed.

Table 5.1 Genome assembly statistics for different sequencing depths (for contigs ≥500 bp). Genomes were mapped against the *L. plantarum* WCFS1 reference genome. The numbers shown in bold represent the best assembly statistic per row.

Assembly features	50x	100x	300x	800x	4000x
Total length	3 338 431	3 347 124	3 356 590	**3 390 887**	3 390 079
Genome fraction (%)	84.793	84.868	85.036	**85.086**	85.062
Duplication ratio	1.006	1.005	**1.003**	1.004	1.004
Number of contigs	272	199	**147**	200	212
Largest contig	83 990	234 100	**258 434**	258 394	258 394
N50	25 758	47 762	**60 488**	52 342	56 794
NGA50	16 533	24 536	**27 790**	26 768	26 769
Misassemblies	72	**69**	75	75	71
Indels per 100 kbp	31.31	29.14	**25.92**	42.92	41.95
Genomic features	2 566 + 92 partial	2 590 + 72 partial	2 595 + 65 partial	2 595 + 72 partial	**2 597 + 70 partial**

5.4.3. Variant calling

The impact of deep and ultra-deep sequencing on variant calling was determined by aligning reads of *L. plantarum* IWBT B063 to the closely related reference genome of *L. plantarum* WCFS1 (Kleerebezem *et al.*, 2003; Siezen *et al.*, 2012). I then compared the variants obtained at all selected read depths (Figure 5.2). There is an increase in the number of variants with an increase in sequence depth, however at 4000x the number of indels (Figure 5.2B) and MNPs (multiple nucleotide polymorphism; Figure 5.2C) is significantly lower than at all other depths ($p = 0.012$).

Therefore, 99.9% of all called variants are SNPs at 4000x, which was significantly higher (chi square p-value = 0.000) than at 800x (89.9%) (Figure 5.2A). There was a significant difference in the number of MNPs from 50x to 800x (chi square p-value = 0.0.024), and a significant increase in the number of indels from 100x to 800x (chi square p-value = 0.0009; Figure 5.2B).

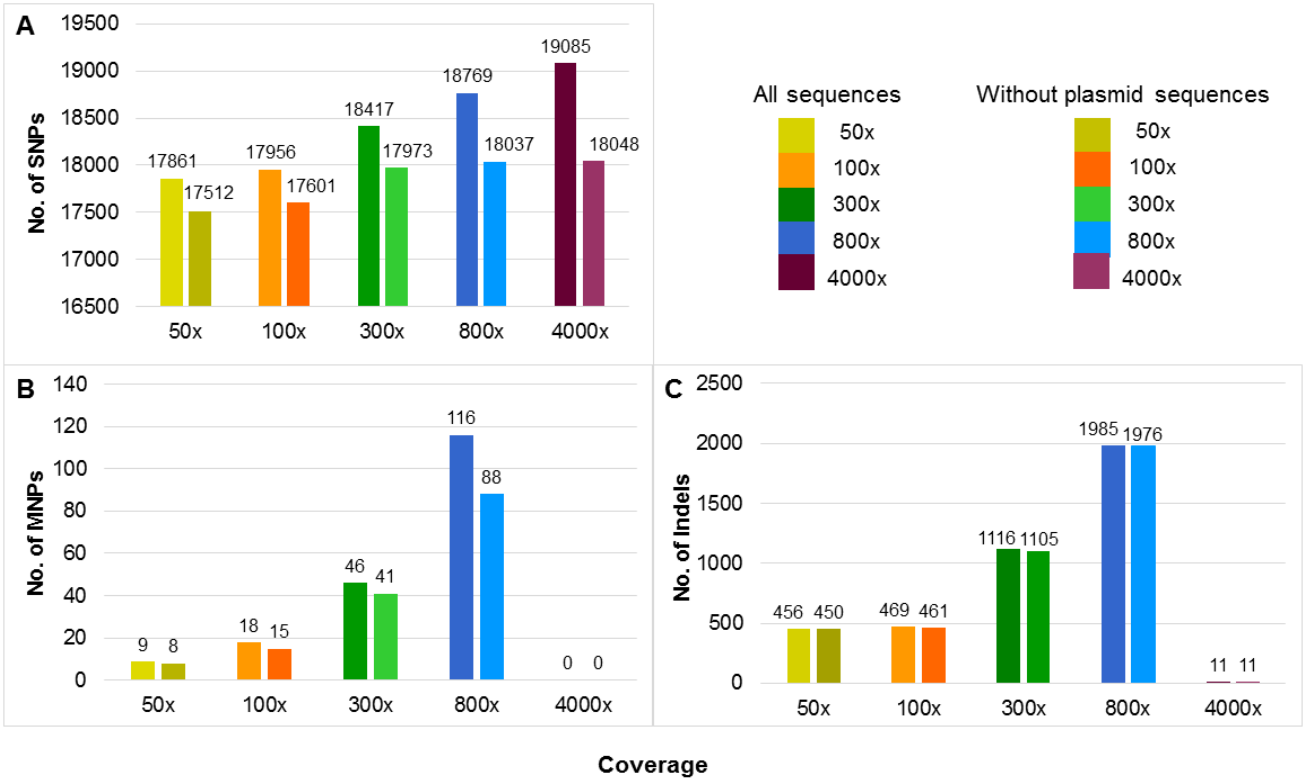

Figure 5.2 Total number of predicted variants overall and after the removal of plasmidsequences at each read depth. A) Number of SNPs. B) Number of Indels. C) Number of MNPs. The actual values are shown as data labels above each bar. Sequence depths: 50x (yellow/gold), 100x (light/dark orange), 300x (dark/light green); 800x (dark/light blue); 4000x (dark plum/plum).

From this dataset variants appearing on sequences that mapped to plasmid sequences of the reference genome were manually removed to reduce variant call errors (Quail *et al.*, 2012). I then compared the variant calls at the different read depths without the plasmid sequences (Figure 5.2). The data show a significant decrease in the number of SNPs at all read depths (p-value = 0.025) (Figure 5.2A). There were no significant differences in the number of MNPs at 800x (chi square p-value = 0.754; Figure 5.2B). There were no significant differences in the number of indels (chi square p-value = 0.987) after the removal of plasmid sequences, which may suggest that these indels are true variants between *L. plantarum* WCFS1 and *L. plantarum* IWBT B063.

5.4.4. Comparison of downstream analysis of isolates evolved from *L. plantarum* IWBT B063 for 50 generations at sequencing depth 300x and 4000x

Genome assembly data suggest that reads with sequence depth of 300x are assembled better with SPAdes compared to 4000x reads (Table 5.1). Using these datasets, I mapped isolates generated by the directed evolution of the *L. plantarum* IWBT B063 strain (see Chapter 3 of this dissertation), i.e. used the 300x and 4000x datasets as reference datasets to assemble and annotate the evolved isolates. Table 5.2 shows global genome statistics for each of the evolved isolates when the 4000x and 300x datasets are used as reference sequences, respectively. The data show that using a reference genome sequenced at 4000x depth results in false data as there are large differences between the evolved isolates and their parent (reference) in terms of the genome size, number of contigs, predicted protein coding genes, hypothetical proteins and tRNAs (Table 5.2). Whereas when using a reference with 300x sequencing depth the data between the evolved isolates and their parent is more coherent and makes more biological sense (Table 5.2).

Table 5.2 Genome stats based on the annotation by Prokka. The *L. plantarum* IWBT B063 strain was sequenced at 4000x coverage and used as reference genome for the evolved isolates. The bold, red values vary the most between the reference and evolved isolates.

Genome	Genome size (bp)	No. of contigs	GC content (%)	Predicted protein coding genes	rRNA	tRNA	No. of hypothetical proteins
4000x coverage							
Parent	**3 538 359**	**314**	44.37	3 384	5	**58**	1 458
C5	3 218 466	268	44.39	3 106	6	66	1 358
C26	3 217 267	293	44.41	3 072	5	60	1 339
E5	3 361 487	262	44.37	3 199	5	61	1 393
E22	3 368 244	277	44.37	3 239	6	64	1 430
E26	3 366 003	267	44.36	3 230	4	62	1 416
300x coverage							
Parent	**3 371 245**	**192**	44.33	3 243	4	**68**	1 430
C5	3 366 394	143	44.33	3 237	4	65	1 429
C26	3 367 885	145	44.33	3 238	4	64	1 431
E5	3 364 361	153	44.33	3 237	5	64	1 427
E22	3 365 592	154	44.33	3 245	5	65	1 434
E26	3 365 984	150	44.33	3 237	5	66	1 439

5.5. Discussion

It is over two decades since the development of genome sequencing technologies, and newer, faster platforms have been developed recently (Kchouk *et al.*, 2017), but these technologies are still plagued with errors (Quail *et al.*, 2012; Besser *et al.*, 2018). These errors can be introduced at different stages of the sequencing project i.e. at the PCR amplification step, in regions of the DNA with high GC content and polymer sequences, during the genome assembly step and when calling variants (Fang *et al.*, 2014; see review by Abnizova *et al.*, 2017). Many authors have come up with techniques that can be applied to reduce errors detected with downstream analysis, which include

PCR-free sequencing, use of error correction tools such as Genome Analysis Toolkit (GATK; McKenna *et al.*, 2010), 'slicing' genome sequences, improving sequencing chemistry (Churchill *et al.*, 2016), and finally selecting optimum sequencing coverage (Desai *et al.*, 2013; Fang *et al.*, 2014; Lonardi *et al.*, 2015; Laehnemann *et al.*, 2016).

The aim of this study was to compare genome assembly and variant call datasets obtained from 50x, 100x, 300x, 800x and 4000x read coverage datasets, to highlight the potential errors that may arise from ultra-deep sequencing. I obtained reads generated using the Ion Proton sequencer and applied SPAdes to assemble the genome. To start with, the number of reads with high quality was reduced at 800x and 4000x depth which is evidence that ultra-deep sequencing is not the best option for whole genome sequencing of small genomes, especially if one plans to call variants downstream (Desai *et al.*, 2013; Lonardi *et al.*, 2015). This also shows that ultra-deep sequencing results in a lot of data that is unusable, therefore wasting resources (Daley & Smith, 2013).

The genome assembly data shows that the data obtained was variable, although at 300x sequencing depth the number of contigs and the number of indels per 100 kbp was reduced, the duplication ratio was also lower at this depth, but there was a large number of misassemblies. As was shown previously, a read depth <300x resulted in less misassemblies (Desai *et al.*, 2013; Mirebrahim *et al.*, 2015; Gampawar *et al.*, 2019). Moreover, employing a reference-guided assembly instead of a *de novo* assembly could also improve the genome assembly (Lischer & Shimizu, 2017). A step up can be added, where multiple references are used in the assembly, especially for short Ion Torrent reads and for strains that have a high genome diversity compared to the reference genome (Schneeberger *et al.*, 2009; Lischer & Shimizu, 2017) as is the case between *L. plantarum* WCFS1 and *L. plantarum* IWBT B063. Additionally, the use of longer reads may significantly improve genome assemblies, reducing gaps and missassemblies. Ion Torrent can now make use of 600 bp reads (Kchouk *et al.*, 2017).

As expected, more indels were predicted by the genome assembly method, than by variant calling for ultra-deep sequenced reads (4000x). This shows that assembly based variant analysis method is more likely to call false positive indels, because unlike variant callers, assemblers build contigs based on reads, and the differences in reads making up contigs may be identified as an indel when it is not. Whereas, with variant caller-specific indel detection specific parameters are set when calling variants (Garrison & Marth, 2012). On that note, the large number of SNPs observed here is due to the genome diversity between *L. plantarum* IWBT B063 and *L. plantarum* WCFS1 which was used as reference. Additionally, it was observed that at ultra-deep coverage (4000x) the number of SNPs increases, and indels decrease which is in line with what was shown for Ion Torrent datasets that this technology accurately detects SNPs, but is poor at detecting indels

(Merriman *et al.*, 2012; Salipante *et al.*, 2014; Song *et al.*, 2017). This is observed especially, after the removal of plasmid sequences where 1.9-5.4% (50x to 4000x, respectively) of SNPs are removed. Interestingly, there was no significant difference in the number of SNPs called after the removal of plasmid sequences, suggesting that at sequencing depths >300x the variants were correctly called (Quail *et al.*, 2012). This also highlights the accuracy of Ion Torrent technology when detecting SNPs (Vasudevan *et al.*, 2019). In all genome datasets, regardless of depth (with the exception of 4000x) 6-11 indels were removed with the removal of plasmid sequences, suggesting that Ion Proton has a threshold at which it detects indels (probably <1000x) (Quail *et al.*, 2012; Lonardi *et al.*, 2015) and that plasmid sequences are less likely to contain long homopolymer sequences and therefore will not map to these regions, thus less likely to influence indel calling.

Additionally, it was clearly evident that ultra-deep sequencing may significantly affect downstream analysis, especially in the case of unknown genomes. It is observed here that when using a genome sequenced at 4000x depth as a reference there will be a high genome diversity between strains that are closely related, which will affect data interpretation. The data in Table 5.2 suggests that the reference and each of the evolved isolates are completely different strains, which is impossible after only 50 generations of evolution (McDonald, 2019). Moreover, the data in Table 5.2 show that Prokka struggled to annotate the genome of *L. plantarum* IWBT B063 as indicated by the variability in the number of tRNA sequences. *L. plantarum* genomes have between 62-75 tRNAs (Siezen *et al.*, 2012; Golneshin *et al.*, 2015; Li *et al.*, 2016; El Halfawy *et al.*, 2017; Inglin *et al.*, 2017; Suryavanshi *et al.*, 2017; Amoranto *et al.*, 2018), here I observed 58 tRNAs in the parent strain (4000x) whereas the evolved isolates have >60 tRNAs which suggests that these isolates gained tRNAs. This is not possible, and an explanation will again lie with genome annotation. Large errors in sequencing may introduce a lot of 'artifact' variants (González-Escalona *et al.*, 2019) that may have resulted in the assembly recognizing reads that should be grouped together into contigs as different, increasing the number of contigs and altering genome assembly results (Olson *et al.*, 2015), which ultimately impact genome annotation and data interpretation.

5.6. Conclusion

Small genomes such as those of bacteria are normally sequenced at high sequencing depths (>100x) (Song *et al.*, 2018). However, it has been shown in recent years that low depth can achieve high quality variants, at least for Illumina data (Desai *et al.*, 2013; Lonardi *et al.*, 2015; Mirebrahim *et al.*, 2015). For the first time it was shown that deep and ultra-deep sequencing of whole genome bacterial datasets is not forcibly a good choice for genome assembly and downstream variant call analysis. Sequence read depth of 300x resulted in a 'better' assembly and SNP analysis compared to read depths ≤100x or ≥800x. However, consistent with literature, our

data shows the limitations of Ion Proton to accurately detect indels. The data show that indel detection is limited at 4000x depth. Moreover, I show that sequencing at 4000x depth results in less coherent data when using the higher coverage data as reference genome for closely related strains, resulting in possible misinterpretation of data.

References

Abnizova, I., Boekhorst, R. te, et al., 2017. Computational errors and biases in short read next generation sequencing J. Proteomics Bioinform. 10, 1, 1–17.

Afgan, E., Baker, D., et al., 2018. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update Nucleic Acids Res. 46, W1, W537–W544.

Amoranto, M.C., Oh, J.K., et al., 2018. Complete genome sequence of *Lactobacillus plantarum* SK151 isolated from kimchi Korean J. Microbiol. 54, 3, 295–298.

Besser, J., Carleton, H.A., et al., 2018. Next-generation sequencing technologies and their application to the study and control of bacterial infections Clin. Microbiol. Infect. 24, 4, 335–341.

Churchill, J.D., King, J.L., et al., 2016. Effects of the Ion PGMTM Hi-QTM sequencing chemistry on sequence data quality Int. J. Legal Med. 130, 5, 1169–1180.

Daley, T. & Smith, A.D., 2013. Predicting the molecular complexity of sequencing libraries Nat. Methods 10, 4, 325–327.

Damiati, E., Borsani, G., et al., 2016. Amplicon-based semiconductor sequencing of human exomes: performance evaluation and optimization strategies Hum. Genet. 135, 5, 499–511.

Desai, A., Marwah, V.S., et al., 2013. Identification of optimum sequencing depth especially for de novo genome assembly of small genomes using next generation sequencing data PLoS One 8, 4, 60204.

El Halfawy, N.M., El-Naggar, M.Y., et al., 2017. Complete genome sequence of *Lactobacillus plantarum* 10CH, a potential probiotic lactic acid bacterium with potent antimicrobial activity Genome Announc. 5, 48, 1–2.

Fang, H., Wu, Y., et al., 2014. Reducing INDEL calling errors in whole genome and exome sequencing data Genome Med. 6, 89, 1–17.

Fujita, S., Masago, K., et al., 2017. Single nucleotide variant sequencing errors in whole exome sequencing using the Ion Proton System Biomed. Reports 7, 17–20.

Gampawar, P., Saba, Y., et al., 2019. Evaluation of the performance of AmpliSeq and SureSelect exome sequencing libraries for Ion Proton Front. Genet. 10, 856.

Garrison, E. & Marth, G., 2012. Haplotype-based variant detection from short-read sequencing 1–9.

Golneshin, A., Adetutu, E., et al., 2015. Complete genome sequence of *Lactobacillus plantarum* Strain B21, a bacteriocin-producing strain isolated from Vietnamese fermented sausage Nem Chua Genome Announc. 3, 2, 1–2.

González-Escalona, N., Allard, M.A., et al., 2019. Nanopore sequencing for fast determination of plasmids, phages, virulence markers, and antimicrobial resistance genes in Shiga toxin-producing Escherichia coli PLoS One 14, 7, e0220494.

Gurevich, A., Saveliev, V., et al., 2013. QUAST: Quality assessment tool for genome assemblies Bioinformatics 29, 8, 1072–1075.

Inglin, R.C., Meile, L., et al., 2017. Complete and assembled genome sequence of *Lactobacillus plantarum* RI-113 isolated from salami Genome 5, 16, 1.

Kchouk, M., Gibrat, J.F., et al., 2017. Generations of sequencing technologies: from first to next generation Biol. Med. 09, 03, 1–9.

Kishikawa, T., Momozawa, Y., et al., 2019. Empirical evaluation of variant calling accuracy using ultra-deep whole-genome sequencing data Sci. Rep. 9, 1784, 1–10.

Kleerebezem, M., Boekhorst, J., et al., 2003. Complete genome sequence of *Lactobacillus plantarum* WCFS1 Proc. Natl. Acad. Sci. U. S. A. 100, 4, 1990–1995.

Kuśmirek, W., Franus, W., et al., 2019. Linking de novo assembly results with long DNA reads using the dnaasm-link application Biomed Res. Int. 2019, 7847064.

Laehnemann, D., Borkhardt, A., et al., 2016. Denoising DNA deep sequencing data-high-throughput sequencing errors and their correction Brief. Bioinform. 17, 1, 154–179.

Lahens, N.F., Ricciotti, E., et al., 2017. A comparison of Illumina and Ion Torrent sequencing platforms in the context of differential gene expression BMC Genomics 18, 1, 602.

Lander, E.S., Linton, L.M., et al., 2001. Initial sequencing and analysis of the human genome Nature 409, 6822, 860–921.

Lewington, J., Greenaway, S.D., et al., 1987. Rapid small-scale preparation of bacterial genomic DNA, suitable for cloning and hybridization analysis Lett. Appl. Microbiol. 5, 3, 51–53.

Li, P., Gu, Q., et al., 2016. Complete genome sequence of *Lactobacillus plantarum* LZ206, a potential probiotic strain with antimicrobial activity against food-borne pathogenic microorganisms J. Biotechnol. 238, 52–55.

Lischer, H.E.L. & Shimizu, K.K., 2017. Reference-guided de novo assembly approach improves genome reconstruction for related species BMC Bioinformatics 18, 1, 1–12.

Lonardi, S., Mirebrahim, H., et al., 2015. When less is more: "Slicing" sequencing data improves read decoding accuracy and de novo assembly quality Bioinformatics 31, 18, 2972–2980.

McDonald, M.J., 2019. Microbial Experimental Evolution – a proving ground for evolutionary theory and a tool for discovery EMBO Rep. 20, e46992.

McElroy, K., Thomas, T., et al., 2014. Deep sequencing of evolving pathogen populations: applications, errors, and bioinformatic solutions Microb. Inform. Exp. 4, 1, 1.

McKenna, A., Hanna, M., et al., 2010. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data Genome Res. 20, 9, 1297–1303.

Merriman, B., Rothberg, J.M., et al., 2012. Progress in Ion Torrent semiconductor chip based sequencing Electrophoresis 33, 23, 3397–3417.

Mirebrahim, H., Close, T.J., et al., 2015. De novo meta-assembly of ultra-deep sequencing data Bioinformatics 31, 12, i9–i16.

Olson, N.D., Lund, S.P., et al., 2015. Best practices for evaluating single nucleotide variant calling methods for microbial genomics Front. Genet. 6, 235.

Quail, M.A., Smith, M., et al., 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers BMC Genomics 13, 341.

Rhoads, A. & Au, K.F., 2015. PacBio sequencing and its applications Genomics, Proteomics Bioinforma. 13, 5, 278–289.

Ross, M.G., Russ, C., et al., 2013. Characterizing and measuring bias in sequence data Genome Biol. 14, 5, R51.

Salipante, S.J., Kawashima, T., et al., 2014. Performance comparison of Illumina and Ion Torrent next-generation sequencing platforms for 16S rRNA-based bacterial community profiling Appl. Environ. Microbiol. 80, 24, 7583–7591.

Schneeberger, K., Hagmann, J., et al., 2009. Simultaneous alignment of short reads against multiple genomes Genome Biol. 10, 98, 1–12.

Seo, H., Park, Y., et al., 2017. Evaluation of exome variants using the Ion Proton Platform to sequence error-prone regions PLoS One 12, 1–12.

Siezen, R.J., Francke, C., et al., 2012. Complete resequencing and reannotation of the *Lactobacillus plantarum* WCFS1 genome J. Bacteriol. 194, 1, 195–196.

Song, K., Li, L., et al., 2016. Coverage recommendation for genotyping analysis of highly heterologous species using next-generation sequencing technology Sci. Rep. 6, 35736, 1–7.

Song, L., Huang, W., et al., 2017. Comparison of error correction algorithms for Ion Torrent PGM data: Application to hepatitis B virus Sci. Rep. 7, 8106, 1–11.

Song, Y., He, Q., et al., 2018. Genomic variations in probiotic *Lactobacillus plantarum* P-8 in the human and rat gut Front. Microbiol. 9, 893, 1–11.

Suryavanshi, M. V, Paul, D., et al., 2017. Draft genome sequence of *Lactobacillus plantarum* strains E2C2 and E2C5 isolated from human stool culture Stand. Genomic Sci. 12, 15, 1–9.

Vasudevan, K., Devanga Ragupathi, N.K., et al., 2019. Highly accurate-single chromosomal complete genomes using IonTorrent and MinION sequencing of clinical pathogens Genomics (April), 1–7

.

Chapter 6

General discussion and conclusions

Chapter 6: General discussion and conclusions

6.1. Introduction

The wine environment provides a good opportunity to model interspecies interactions and population dynamics, as the raw material is not sterilised and therefore the natural microflora is part of the fermentation whether spontaneous or inoculated. It is a relatively well-described and characterised anthropogenic environment, and the fermentation microbiome has been studied (Boynton & Greig, 2016; Bagheri *et al.*, 2017, 2018). The data show that this environment is highly selective and a limited number of robust species dominate the system (Albergaria & Arneborg, 2016; Boynton & Greig, 2016) due to harsh physicochemical parameters, including acidic pH, lack of oxygen, high sugar and high ethanol concentrations (Sumby *et al.*, 2019). Data show that the yeast species *Saccharomyces cerevisiae* consistently dominates over other microbial species, in spite of this species being generally present in very low numbers in freshly pressed grape must (Ciani & Comitini, 2015; Albergaria & Arneborg, 2016). The reason for this is that *S. cerevisiae* has evolved to be highly competitive through the acquisition of genes by horizontal gene transfer (HGT) (Novo *et al.*, 2009) and by continuous propagation in the wine environment by humans (Steensels *et al.*, 2019). This has allowed this species to be metabolically more efficient, and to be considered a “domesticated” microorganism (Ciani *et al.*, 2016).

Lactic acid bacteria (LAB) interact with the yeast throughout the course of the fermentation of grape must to wine in both spontaneous and inoculated fermentations (Liu *et al.*, 2017). The interactions that exist between yeast and LAB are complex and broadly either negative (competition for nutrients, inhibition by inhibitory metabolites) or positive (stimulation by release of essential nutrients) (Zapparoli *et al.*, 2003; Bayrock & Ingledew, 2004; Rizk *et al.*, 2016; Ponomarova *et al.*, 2017). These interactions have been investigated mainly from a metabolic response perspective (Muñoz *et al.*, 2014; Tristezza *et al.*, 2016; du Toit, 2018; Lucio *et al.*, 2018). The impact of interspecies interactions within a particular environment on the evolution of co-inhabitants is however not well characterised.

As part of ongoing research at the South African Grape and Wine Research Institute (SAGWRI) that focuses on understanding mixed species interactions and their influence on shaping evolutionary selection, I applied a different and novel approach to better characterise these interactions on a molecular level. I implemented a directed evolution (DE)-based strategy to evolve a strain of LAB by subjecting the strain to biotic selection pressures exerted by fermenting wine yeast strains. In this approach I used two driver *S. cerevisiae* strains to evaluate whether the bacteria would evolve yeast strain-specific properties or primarily evolve to respond to general fermentation stresses (i.e. rising ethanol concentrations, etc.). The major outcomes of this study are presented below and the potential avenues for further research are proposed.

6.2. Major outcomes and future work

To date, there are only two studies that have applied directed evolution to wine LAB, however both studies used abiotic selective pressures (Betteridge *et al.*, 2018; Jiang *et al.*, 2018). It is only recently that the evolution of one microbe was evaluated in the presence of other microbes (biotic stresses), however, this study focused on yeast evolution (Zhou, Bottagisi, *et al.*, 2017; Zhou, Swamy, *et al.*, 2017). Therefore, this is the first study to our knowledge that evolved wine LAB using a biotic selective driver.

6.2.1. LAB evolutionary response altered by conflicting selection pressures

The research presented in this PhD is the first to use an untargeted directed evolution approach to evolve a wine *L. plantarum* strain, which means that the evolving populations (evolved with either Cross Evolution or EC1118) were exposed to fluctuating environments. The different stresses that the evolving bacterial population presumably encountered ranged from SGM parameters (high sugar, low pH) to the production of possibly inhibitory metabolites by the yeast strains to actual physical contact with the yeast strains. Although, none of these (apart from SGM parameters) were measured in this study, each may have exerted strong selection pressure on the evolving populations over time.

This approach, unlike in targeted directed evolution studies (Betteridge et al., 2018; López-González et al., 2018), should allow us to have a broader view of what occurs in natural ecosystems as it mimics dynamic natural environments which continuously fluctuate (Brooks et al., 2011). Indeed, in natural environments it is frequently difficult to determine what specific selective pressures are acting on microbial populations. Our approach reduces the complexity of these natural environments by only allowing interactions between two specific strains, but nevertheless provides us with a broader view of microbial phenotypic and molecular responses to fluctuating environments.

It is worthwhile noting that such fluctuating environments may result in conflicting selection pressures at varying stages of the DE experiments (Harrison et al., 2017). This may be one reason that I observed widely diverging phenotypes of individual strains such as highly variable L-malic acid degradation capabilities (from very limited to better than parent) and variable growth (same as for MLF), while the phenotype of the evolved population was improved for both growth and MLF. Identifying yeast-specific genetic changes proved quite challenging, as there were no mutations or gene targets unique to isolates evolved with either EC1118 or Cross Evolution. Nonetheless, some of the identified genes in the KEGG pathway analysis were unique to isolates evolved for either 50

or 100 generations and with either yeast strain. Overall this suggests that yeast-driven DE impacts similar pathways in *L. plantarum* isolates (at least for the strains used here).

The data show an increase of growth over time in populations evolved in the presence of both EC1118 and Cross Evolution, but individually selected isolates from these populations did not all perform better than the parent strain, which may allude to the indirect selection of a population phenotype. It is expected in microbial communities that individuals have varying roles, where some individuals are 'co-operators' while others are cheaters or rather benefit from the more efficient individuals in terms of metabolic processing (Harrington & Sanchez, 2014). Moreover, it has been suggested that throughout the evolutionary process and different generations phenotypic switching plays an important role in the survival of individuals by allowing them to fill the different niches that might occur due to the dynamics of the environment and the contribution of other microbes to the same environment (Van Boxtel *et al.*, 2017). Therefore, the few selected isolates in this current study show a range of L-malic acid degradation, in spite of the efficiency with which the larger population metabolised this acid. This is not entirely surprising as individuals in a population generally play different roles and fill various niches which may be beneficial to the population as a whole or to the individual only (West *et al.*, 2006; Xavier, 2011). It was suggested that traits that delay growth tend to have a higher probability of being selected, especially in the case of populations which undergo continuous bottlenecks (Wahl *et al.*, 2002) as was the case here with the serial transfers. This would then suggest that the selected isolates potentially carry mutations that will confer a fitness benefit in the long-term as they have a longer time to establish in the population as opposed to mutations that appear early and promote growth only to be lost in the bottleneck (Wahl & Zhu, 2015).

Furthermore, and although malic acid and growth were used to select evolved isolates (Chapter 3), it is not clear what other traits might have been selected for, especially since it is evident that selection was done at the population level. Therefore, future work should include the application of a phenotypic microarray (PM) strategy which may help elucidate some of the non-obvious phenotypes (Borglin et al., 2012). For example, PMs were used to evaluate the nutrient utilisation of *Lactobacillus sakei* after exposure to increasing levels of ultrasound. The data revealed that the metabolic pathways involved with carbon, nitrogen, sulphur, and phosphorous utilisation varied between different cultures (Ojha et al., 2018). The application of PMs allows for a large screen of microbial isolates and populations to elucidate multiple phenotypes (Borglin et al., 2012) which might have been missed in this current study.

6.2.2. Evolved strains show yeast-specific phenotypes

The variable selective pressures resulted in significant phenotypic changes that were observed after only 50 generations (Chapter 3 of this dissertation) compared to ≥330 generations in other studies (Betteridge *et al.*, 2018; Jiang *et al.*, 2018). Therefore, future work will fully characterise the phenotypes of isolates evolved for 100 generations to do a full comparative analysis between the two generations. The majority of the isolates that completely degraded L-malic acid were those that were evolved with EC1118, while the majority of the isolates that were poorer at L-malic acid degradation were evolved with Cross Evolution. The latter suggests that either the selection pressure imposed by Cross Evolution was not sufficient to have isolates with increased competitive advantage or that the larger population evolved with this yeast managed to degrade the L-malic acid substrate faster because there were more individuals to consume the substrate i.e. individuals competed for the same substrate and so it was used up faster (Dai *et al.*, 2012). Alternatively, the high growth observed during the DE experiments may be due to the high initial inoculation of cells (log 8.3 CFU/ml) and not due to increased fitness. Cross Evolution produces inhibitory metabolites which directly impact the degradation of L-malic acid and the growth of individual isolates (Scott Laboratories, 2018). Therefore, had the starting population been smaller the increase in growth would not have been evident suggesting that this population was indeed supressed by Cross Evolution.

On the other hand, the competitive behaviour between EC1118 and the evolved isolates was strong, as suggested by the consistently superior phenotypes of isolate E22 compared to the parent (Hall *et al.*, 2018). Inasmuch as the majority of isolates showed distinct phenotypes, a full spectrum of overlapping phenotypes existed in both cases. For example, some isolates evolved with EC1118 were poorer at degrading malic acid compared to the parent strain, showing that indeed random isolates were selected but highlighting the limitation of selecting only 10 isolates per evolutionary line. Moreover, the evolving population was subjected to various stresses other than those of the experimental evolution i.e. filtration of the fermentation through a 0.22 µm pore disc filter. This strategy may have introduced a selection biased in favour of cells that can fit through the pores. In the event that cells formed large clumps or chains, these cells were most likely selected against. These limitations can be overcome by selecting more colony isolates per evolutionary line and instead of filtering, adding a fungicide as a means of removing yeasts before transferring the bacteria to fresh synthetic grape must (SGM).

6.2.3. Use of yeast-free fermentations as standards

The interest in this work was primarily evaluating yeast strain-specific versus generic evolution targets. The use of a monoculture evolutionary line as a further ‘control’ would not have contributed to this evaluation, because yeast drive the metabolic fluctuation in grape must fermentation

(Albergaria & Arneborg, 2016) and thus provide a unique environment in which the LAB was evolved. In monoculture fermentation, the environment would be completely different and comparing the two would not provide the intended insights. However, the use of a monoculture control may have helped to compare the general features and the speed of evolutionary change in the presence or absence of yeast. Moreover, the wide and seemingly random spread of mutations between the different isolates, suggests that the results would have probably been as random in monocultures. Here the major finding was that there are obvious yeast-strain specific phenotypes.

6.2.4. Peptidoglycan biosynthesis and degradation is suggested as a major target for interspecies interactions

Chapter 4 reveals the potential evolutionary targets of the directed evolution experiments. Different genes were affected in the different isolates evolved for 50 and 100 generations. However, KEGG pathway analysis revealed that most of the genes that were affected belonged to the *Peptidoglycan Biosynthesis and Biodegradation* pathway. This is a major finding because it may suggest that interspecies interactions are affected by physical contact (Chapot-Chartier & Kulakauskas, 2014) and that the presence of yeast in the fermentation medium results in LAB modulating their cell wall as a means of protection against the biotic stress (Chapot-Chartier & Kulakauskas, 2014; Papadimitriou *et al.*, 2016). Further investigations are needed to elucidate whether the mutations on cell-wall related genes have resulted in phenotypic changes which either negatively (in the case of isolates evolved with Cross Evolution) or positively (EC1118 evolved isolates) impact the growth and malic acid degradation capabilities of the evolved isolates. This can be done first through electron microscopy to visualise cell surface changes (if any). Secondly, isolates with cell-wall related mutations can be tested for Gram variability compared to the parent strain (Beveridge, 1990). Lastly, the impact of physical interactions can be investigated by evolving the LAB in a system where they are not in direct contact with the yeast, but share metabolites. A similar system was used previously to investigate yeast-yeast interactions in a bioreactor where the yeast were physically separated by a porous membrane, allowing metabolic exchanges (Lopez *et al.*, 2014; Taillandier *et al.*, 2014; Luyt, 2015). The point of this study was to mimic natural environments as far as possible, and this system can be applied to batch cultures with no control of other parameters. On the other hand, the lack of controlled environmental parameters presented a challenge and I could not conclusively identify which genetic targets are a direct consequence of the interactions of the LAB with the yeast and which are due to general adaptation to the wine environment or which are simply neutral mutations that have no bearing on the fitness of selected isolates. Therefore, using a system where the yeast is separated from the bacteria can be implemented in future under both controlled and uncontrolled parameters and compared to elucidate the extent of this effect.

6.2.5. Biotic factor-driven evolution may prefer diversity within populations

In this dissertation I present, for the first time, a comparative genomics study for short-term evolution in wine LAB. Most research has focused on the comparative genomics of LAB that have evolved over millennia (Makarova *et al.*, 2006; Kant *et al.*, 2011; Sun *et al.*, 2015). This study focused on nonsynonymous mutations as these can easily be followed up with *in silico* analyses (Chapter 4). However, there was a large number of upstream gene mutations, which are possibly in regulatory regions of genes. Mutations on regulatory regions may affect gene regulation and expression (Scholten & Tommassen, 1994; Eckdahl & Eckdahl, 2016) making them important for further investigations. Additionally, there was a large number of synonymous mutations which have been shown recently to affect the transcription of genes and subsequently increase growth of *Escherichia coli* which shows that synonymous mutations are also relevant as they can result in phenotypic changes (Kershner *et al.*, 2016). Therefore, gene expression analysis by quantitative polymerase chain reaction (qPCR) should be conducted to gain a full understanding of the genome-wide impact of interspecies interactions on *L. plantarum* in wine.

Although, isolates evolved with EC1118 mainly display the 'good MLF/growth' phenotypes and the isolates evolved with Cross Evolution have 'poor MLF/growth' after 50 generations, the results in Chapter 4 do not show a clear picture of which mutations or gene targets are responsible for these phenotypes, therefore transcriptomic analysis may help resolve this. Moreover, in hindsight, comparing isolates with different phenotypes could not have resulted in similarities of genotype, therefore it is important to analyse isolates from the 100th generation and above. As shown in the preliminary screens of these isolates, the majority completely degraded L-malic acid, and further characterisation will allow for the selection of isolates with similar phenotypes, which can then be evaluated for genomic differences or similarities.

Furthermore, selection of isolates based only on growth and malic acid disregards other potentially interesting phenotypes that may not be directly linked to growth or malic acid (Cosetta & Wolfe, 2019). As mentioned earlier, phenotypic microarrays may help elucidate underlying differences between the evolved isolates which can be linked to the observed differences in genotypes. These diverse mutations can possibly explain the observed strain-dependent phenotypes in LAB (Thierry *et al.*, 2015). Alternatively, differences between isolates can be evaluated based on the mutations that are unique to specific isolates, which may be easier to follow up with qPCR since these are fewer than mutations that occur in more than one isolate.

Ultra-deep sequencing results in significant downstream data analysis errors

Whole genome sequencing was conducted on the parent strains and each of the evolved isolates sampled at 50 and 100 generations. The parent genome was sequenced at ultra-deep coverage

(4000x) to ensure that a higher quality genome was obtained, because this strain served as the reference genome for the evolved isolates (Chapter 4). This proved to be a challenge because downstream data analysis revealed that at this high coverage a large number of errors occurred. For example, the number of contigs, tRNAs and the GC content differed significantly between the parent and the evolved isolates (Chapter 5). Therefore, the parent genome was re-sequenced at a lower coverage (300x). This resulted in more biologically sound data (Chapter 4). Therefore, it is important to consider the sequencing depth prior to sequencing whole genomes. Furthermore, our data revealed that sequencing at a depth between 100x and 300x reduced the number of errors in genome assembly and subsequent downstream analysis.

To further increase accuracy in whole genome sequence data, it is important to consider the reference genome used, especially for *de novo* sequencing and assembly. In this work *Lactobacillus plantarum* WCFS1 and LB1-2 were used as reference genomes for the assembly of the parent genome (Chapter 4). WCFS1 was further used to annotate the parental genome. It was observed that there is high genome variation between WCFS1 and the parent genome which resulted in a large number of genes being annotated as 'hypothetical proteins'. Future work should involve manual annotation of these hypothetical proteins using NCBI BLAST and Pfam to provide a full view of the genome (Johnson *et al.*, 2008; El-Gebali *et al.*, 2019). Moreover, gaps in the genome could be filled by using gap filling programs such as GapFiller and LR_Gapcloser (Nadalin *et al.*, 2012; Xu *et al.*, 2018) or by using long and short reads from different sequencing platforms (Vasudevan *et al.*, 2019). These can help generate a highly accurate, single contig genome for use as a reference for the *de novo* assembly of the evolved isolates. This may help identify genes of unknown function that may have also been affected by the DE experiments, revealing more genetic targets and increasing our understanding of yeast-bacteria interactions on a molecular level.

6.3. Conclusions

Microbial interactions are a topic of interest currently due to the prevalence of mixed communities in various environments including food and beverages. Wine fermentation provides a good model to study microbial interactions and evolution because it has dynamic physicochemical properties and has a high species richness. The work presented in this PhD is the first to investigate the impact of *S. cerevisiae* on the evolution of *L. plantarum* under wine-like environments. It is shown here that yeast-driven directed evolution results in yeast-specific phenotypes. Moreover, EC1118-driven evolution results in more nonsynonymous mutations possibly suggesting positive selection, compared to Cross Evolution-driven evolution which resulted in more synonymous mutations. There were no clear yeast-specific gene targets. Nonetheless, similar pathways were affected between the isolates evolved with either yeast i.e. cell wall-related genes. The data also highlight

the complexity of using a biotic selective driver, as opposed to a targeted abiotic selective driver where predictions can be made as to what genes may be affected (Bracher *et al.*, 2017; Betteridge *et al.*, 2018). Moreover, I show that small genomes should be sequenced at ≤300x coverage for accurate and comprehensive results.

In conclusion, this data reveals the potential use of directed evolution using a biotic selective pressure as a tool to improve strains for industrial use. Furthermore, the strategy employed here can be used to select for novel strains or for the investigation of interspecies interactions to enhance our understanding of microbial communities. The data suggest that variable and perhaps sometimes conflicting selective pressures at different fermentation stages being imposed on the evolving populations results in a wide-range of phenotypes. However, I also observed a strong selection for yeast-specific phenotypes, but no yeast-specific genetic targets could be identified. Furthermore, the use of both *de novo* and reference-guided assembly as was used here allows for accurate data analysis and interpretation which is essential in the development of new strains with potentially interesting new phenotypes.

References

Albergaria, H. & Arneborg, N., 2016. Dominance of *Saccharomyces cerevisiae* in alcoholic fermentation processes: role of physiological fitness and microbial interactions Appl. Microbiol. Biotechnol. 100, 5, 2035–2046.

Bagheri, B., Bauer, F.F., et al., 2017. The impact of *Saccharomyces cerevisiae* on a wine yeast consortium in natural and inoculated fermentations Front. Microbiol. 8, 1988.

Bagheri, B., Zambelli, P., et al., 2018. Investigating the effect of selected non-*Saccharomyces* species on wine ecosystem function and major volatiles Front. Bioeng. Biotechnol. 6, 169.

Bayrock, D.P. & Ingledew, W.M., 2004. Inhibition of yeast by lactic acid bacteria in continuous culture: Nutrient depletion and/or acid toxicity? J. Ind. Microbiol. Biotechnol. 31, 8, 362–368.

Betteridge, A.L., Sumby, K.M., et al., 2018. Application of directed evolution to develop ethanol tolerant *Oenococcus oeni* for more efficient malolactic fermentation Appl. Microbiol. Biotechnol. 102, 2, 921–932.

Beveridge, T.J., 1990. Mechanism of gram variability in select bacteria J. Bacteriol. 172, 3, 1609–1620.

Boynton, P.J. & Greig, D., 2016. Species richness influences wine ecosystem function through a dominant species Fungal Ecol. 22, 61–72.

Bracher, J.M., de Hulster, E., et al., 2017. Laboratory evolution of a biotin-requiring *Saccharomyces cerevisiaeSaccharomyces cerevisiae* strain for full biotin prototrophy and identification of causal mutations Appl. Environ. Microbiol. 83, 16, e00892-17.

Brooks, A.N., Turkarslan, S., et al., 2011. Adaptation of cells to new environments Wiley Interdiscip Rev Syst Biol Med 3, 5, 544–561.

Chapot-Chartier, M.-P. & Kulakauskas, S., 2014. Cell wall structure and function in lactic acid bacteria Microb. Cell Fact. 13, Suppl1, S9.

Ciani, M. & Comitini, F., 2015. Yeast interactions in multi-starter wine fermentation Curr. Opin. Food Sci. 1, 1–6.

Ciani, M., Capece, A., et al., 2016. Yeast interactions in inoculated wine fermentation Front. Microbiol. 7, 1–7.

du Toit, S.C., 2018. Coevolution of *Saccharomyces cerevisiae* and *Lactobacillus plantarum* : Engineering interspecies cooperation Stellenbosch University.

Eckdahl, A.J. & Eckdahl, T.T., 2016. Mutational Analysis of Transcriptional Initiation in Bacteria 31, 3, 1–8.

El-Gebali, S., Mistry, J., et al., 2019. The Pfam protein families database in 2019 Nucleic Acids Res. 47, D427–D432.

Jiang, J., Sumby, K.M., et al., 2018. Directed evolution of *Oenococcus oeni* strains for more efficient malolactic fermentation in a multi-stressor wine environment Food Microbiol. 73, 150–159.

Johnson, M., Zaretskaya, I., et al., 2008. NCBI BLAST: a better web interface. Nucleic Acids Res. 36, Suppl_2, W5–W9.

Kant, R., Blom, J., et al., 2011. Comparative genomics of *Lactobacillus* Microb. Biotechnol. 4, 3, 323–332.

Kershner, J.P., McLoughlin, S.Y., et al., 2016. A synonymous mutation upstream of the gene encoding a weak-link enzyme causes an ultrasensitive response in growth rate J. Bacteriol. 198, 20, 2853–2863.

Liu, Y., Rousseaux, S., et al., 2017. Wine microbiome: A dynamic world of microbial interactions Crit. Rev. Food Sci. Nutr. 57, 4, 856–873.

Lopez, C.L.F., Beaufort, S., et al., 2014. Interactions between *Kluyveromyces marxianus* and *Saccharomyces cerevisiae* in tequila must type medium fermentation World J. Microbiol. Biotechnol. 30, 8, 2223–2229.

López-González, M.J., Escobedo, S., et al., 2018. Adaptive evolution of Industrial *Lactococcus lactis* under cell envelope stress provides phenotypic diversity Front. Microbiol. 9, November, 1–17.

Lucio, O., Pardo, I., et al., 2018. Influence of yeast strains on managing wine acidity using *Lactobacillus plantarum* Food Control 92, 471–478.

Luyt, N.A., 2015. Interaction of multiple yeast species during fermentation by Stellenbosch University.

Makarova, K.S., Slesarev, A., et al., 2006. Comparative genomics of lactic acid bacteria Proc. Natl. Acad. Sci. U. S. A. 103, 42, 15611–15616.

Muñoz, V., Beccaria, B., et al., 2014. Simultaneous and successive inoculations of yeasts and lactic acid bacteria on the fermentation of an unsulfited Tannat grape must Brazilian J. Microbiol. 45, 1, 59–66.

Nadalin, F., Vezzi, F., et al., 2012. GapFiller: A de novo assembly approach to fill the gap within paired reads BMC Bioinformatics 13, Suppl 14, S8.

Novo, M., Bigey, F., et al., 2009. Eukaryote-to-eukaryote gene transfer events revealed by the genome sequence of the wine yeast *Saccharomyces cerevisiae* EC1118 Proc. Natl. Acad. Sci. U. S. A. 106, 38, 16333–16338.

Papadimitriou, K., Alegria, A., et al., 2016. Stress physiology of lactic acid bacteria Microbiol. Mol. Biol. Rev. 80, 3, 837–888.

Ponomarova, O., Gabrielli, N., et al., 2017. Yeast creates a niche for symbiotic lactic acid bacteria through nitrogen overflow Cell Syst. 5, 4, 345–357.e6.

Rizk, Z., El Rayess, Y., et al., 2016. Impact of inhibitory peptides released by *Saccharomyces cerevisiae* BDX on the malolactic fermentation performed by *Oenococcus oeni* Vitilactic F Int. J. Food Microbiol. 233, 90–96.

Scholten, M. & Tommassen, J., 1994. Effect of mutations in the -10 region of the *phoE* promoter in *Escherichia coli* on regulation of gene expression MGG Mol. Gen. Genet. 245, 218–223.

Steensels, J., Gallone, B., et al., 2019. Domestication of industrial microbes Curr. Biol. 29, R381–R393.

Sumby, K.M., Niimi, J., et al., 2019. Ethanol-tolerant lactic acid bacteria strains as a basis for efficient malolactic fermentation in wine: evaluation of experimentally evolved lactic acid bacteria and winery isolates Aust. J. Grape Wine Res. 25, 4, 404–413.

Sun, Z., Harris, H.M.B., et al., 2015. Expanding the biotechnology potential of lactobacilli through comparative genomics of 213 strains and associated genera Nat. Commun. 6, 8322, 1–13.

Taillandier, P., Lai, Q.P., et al., 2014. Interactions between *Torulaspora delbrueckii* and *Saccharomyces cerevisiae* in wine fermentation: Influence of inoculation and nitrogen content World J. Microbiol. Biotechnol. 30, 7, 1959–1967.

Thierry, A., Valence, F., et al., 2015. Strain-to-strain differences within lactic and propionic acid bacteria species strongly impact the properties of cheese–A review Dairy Sci. Technol. 95, 895–918.

Tristezza, M., di Feo, L., et al., 2016. Simultaneous inoculation of yeasts and lactic acid bacteria: Effects on fermentation dynamics and chemical composition of Negroamaro wine LWT - Food Sci. Technol.

Vasudevan, K., Devanga Ragupathi, N.K., et al., 2019. Highly accurate-single chromosomal complete genomes using IonTorrent and MinION sequencing of clinical pathogens Genomics (April), 1–7.

Xu, G.C., Xu, T.J., et al., 2018. LR-Gapcloser: A tiling path-based gap closer that uses long reads to complete genome assembly Gigascience 8, 1–14.

Zapparoli, G., Torriani, S., et al., 2003. Interactions between *Saccharomyces* and *Oenococcus oeni* strains from Amarone wine affect malolactic fermentation and wine composition Vitis 42, 2, 107–108.

Zhou, N., Bottagisi, S., et al., 2017. Yeast-bacteria competition induced new metabolic traits through large-

scale genomic rearrangements in *Lachancea kluyveri* FEMS Yeast Res. 17, 6, 1–14.

Zhou, N., Swamy, K.B.S., et al., 2017. Coevolution with bacteria drives the evolution of aerobic fermentation in *Lachancea kluyveri* PLoS One 12, 3, 1–19.