Copyright Statement

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

This thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- Any use you make of these documents or images must be for research or private study purposes only, and you may not make them available to any other person.
- Authors control the copyright of their thesis. You will recognise the author's right to be identified as the author of this thesis, and due acknowledgement will be made to the author where appropriate.
- You will obtain the author's permission before publishing any material from their thesis.

To request permissions please use the Feedback form on our webpage. [http://researchspace.auckland.ac.nz/feedback](http://researchspace.auckland.ac.nz/feedback)

General copyright and disclaimer

In addition to the above conditions, authors give their consent for the digital copy of their work to be used subject to the conditions specified on the Library Thesis Consent Form and Deposit Licence.
ABSTRACT

Microbes perform essential processes vital to the functioning of the biosphere and have wide reaching impacts on global economies for the roles they play in producing quality agricultural commodities. They are widely used as model systems to test fundamental hypotheses from a range of scientific disciplines, uncovering invaluable knowledge about biological processes. Despite their undisputed importance to both fundamental science and commercial enterprises, we have a poor understanding of microbial population biology and ecology and how these patterns and processes affect the production of quality agricultural goods. This thesis focuses on the model research eukaryote *Saccharomyces cerevisiae* and takes advantage of its and other fungal species', commercial applications in wine making.

I begin by quantifying the population processes occurring in the New Zealand (NZ) metapopulation of *S. cerevisiae*. I sampled over ten thousand isolates from a variety of native and vineyard associated niches from six different winemaking regions, spanning over 1 000 km across NZ. From these, hundreds of genotypes were obtained and used in a suite of comprehensive quantitative analyses of population structure and gene-flow. Within geographic regions, these reveal no differentiation between native or vineyard associate samples or between populations residing in different niches. Between regions (on scales larger than ~100 km), a complex picture of varying degrees of population differentiation and migration was revealed. These patterns are in line with the movement of fruit by the NZ wine industry and suggest human associated gene-flow may affect microbial population patterns and diversity.

From here I investigate the ecology of *S. cerevisiae* and target my research away from the well understood fruit and ferment niches. As fruit is ephemeral, *S. cerevisiae* requires a strategy to survive when this energy rich resource is not available. While it has been isolated from soil and bark samples in previous studies,
including the above population genetic analysis, what *S. cerevisiae* is doing in this ‘woodland’ niche is unknown. I hypothesised that *S. cerevisiae* employs a life history strategy targeted at self-preservation rather than growth outside of the fruit niche and thus resides in these alternate niches in a sporulated state. Using soil agar as a proxy for the soil niche, I provide evidence that *S. cerevisiae* is able to sporulate in the presence of soil nutrients and does so in a way that maximises its potential reproductive success upon germination. While there are many other aspects of this hypothesis that require experimental verification, this is the first step in understanding the ecology of *S. cerevisiae* outside of the fruit niche.

I then move on to investigate the potential consequences these observed population patterns have on a commercially important agricultural commodity: wine. Agricultural products derived from the same genotype display differential geographic phenotypes in their physical and sensorial signatures, adding economic value and distinctiveness to products. Historically this has been attributed to complex interactions between local soils, climate and agricultural practices and is collectively known as *terroir*, or sense of place. The potential for microbes to contribute to this regional distinction has been ignored until recently; however there is growing evidence to suggest that microbial communities and populations vary with geography. Here I perform the first general test for a microbial aspect to *terroir* using wine as a model system and take advantage of the genetically well characterised *S. cerevisiae* population described in this thesis. I experimentally demonstrate significant differentiation of wine phenotypes by yeasts derived from different geographic regions, providing the first evidence that microbes contribute to the regional distinctiveness of wine and potentially agricultural products generally. This reveals the importance of microbial populations on the regional identity of agricultural commodities and suggests that long-term implementation of agricultural practices that maintain differential microbial diversity could have direct economic implications as well as being desirable in terms of employing agricultural practices that increase responsible environmental stewardship and maintain microbial biodiversity.
Finally I investigated whether fungal species diversity in the grape juice and during fermentation is correlated with the final concentration of three volatile thiols important to wine aroma and flavour. The species of *Saccharomyces* driving the ferment was found to significantly correlate with thiol concentration, particularly 4MMP, with higher proportions of *S. uvarum* affording higher concentrations of 4MMP. Additionally, the fungal communities in the initial juice were found to correlate with the thiol concentrations in the wine. Genera identified as being the main drivers of this effect are known to influence vine and fruit health rather than contribute to fermentation itself, suggesting the effects of microbial populations on wine thiol concentration begins in the vineyard. This reiterates the need to have a better understanding of the interactions between microbial populations and agricultural products and has implications for the management of fungal diversity and disease in these systems.

Overall, this thesis provides a significant body of knowledge to both fundamentally and commercially important fields. It highlights the need to better understand the ecology of microbial populations not only in a fundamental sense but also for commercial imperatives.
ACKNOWLEDGEMENTS

I have had the great fortune to work with some fantastic friends and colleagues throughout my PhD and I feel privileged to be part of such a supportive and knowledgeable team. I would like to thank my supervisor Mat Goddard for all your advice, guidance and patience throughout the last four years. I have learnt many valuable lessons and skills that will remain with me throughout my career and I am grateful you gave me this opportunity. Additionally, thank you to my co-supervisor Shane Lavery for your support, insights and comments on the analyses and text.

This work would not have been possible if not for the support and effort of the “Epic Experiment” sampling team. Thank you to Soon Lee not only for your hard work and impeccable sampling techniques, but for all the behind the scenes work organising and managing the project. The samples would never have made it to Auckland without you! Also thank you for your patience in teaching me countless techniques and the many times you answered my (often absurd) questions. Thank you to my partner in crime, Peter Morrison-Whittle; for all your efforts during the madness that was sampling and for providing comic relief on many occasions when we’d both just had enough. Thank you to Velimir Gayevskiy, firstly for your help with our sampling back in 2011 but more so for the support and ranting conversations about writing in the more recent months.

For help with analyses; firstly thank you to Catrin Guenther and Mandy Herbst-Johnstone for all your support and help with the chemical analyses. I would have been way out of my depth without you both. Thank you to Claudia Buser for your patience and guidance in helping me with my statistical analyses and for getting me started with R. Miguel Roncoroni, thank you for your help with R and my growth curves; you saved me many torturous hours of trying to figure it out. Many thanks to Steffen Klaere not only for all of your guidance and advice, but for putting up with the many pestering questions and emails.

I would like to thank our industry partners whose passion and love of wine is inspirational. Thank you to Pernod Ricard for supplying the experimental juice and allowing me access to their wine scan facilities for analyses. Thank you to Michael Brajkovich and Kumeu River Estate for letting us practice our sampling techniques in Mates Vineyard. The team at Matua Valley for allowing us access to their Sauvignon Blanc vineyard for months on end, despite tearing it up and ending the experiment short. To all of the companies and personal involved in the “Epic Experiment” sampling effort of 2011, thank you for providing us with
expert guides, access to your land and samples; Amisfield, Ata Rangi, Churton, Coal Pit, Constellation, Delegats, Domain Road, Frey Vineyard, Huia, Misha’s Vineyard, Mt Difficulty, Mt Riley, Neudorf, Palliser, Pernod Ricard, Rippon, Seifried, Seresin, Te Kairanga, Te Whare Ra, Tohu, Trinity Hill, Villa Maria and Vita Brevis.

Thank you to the New Zealand Ministry of Business, Innovation and Employment, New Zealand Winegrowers and Plant and Food Research for funding this research and to the University of Auckland for providing my stipend.

I’d like to acknowledge the contribution of the NeSI high-performance computing facilities at the University of Auckland and the staff, particularly Jordi Blasco, at NeSI and Centre for eResearch for advice and access to the computing resources required for Chapter 3.

I would like to thank my lab mates, past and present; for all your guidance, company and friendship. Thank you Soon Lee, Claudia Buser, Catrin Guenther, Peter Morrison-Whittle, Velimir Gayevskiy, Ollie Karon, Emma Carroll, Paulina Giraldo, Nicole Anfang, Zachary Ardern and Beth Anderson for all the fun and ice cream breaks.

Last, but by no means least, thank you to my family and friends for all your support and for always being the first to volunteer to be my wine tasters, despite the lack of wine for you to taste. I love you all! To my partner Joe D’Ambrosio; thank you for your undying support and for keeping me grounded through all of the highs and lows of this epic project. I couldn’t have done it without you.
This form is to accompany the submission of any PhD that contains research reported in published or unpublished co-authored work. Please include one copy of this form for each co-authored work. Completed forms should be included in all copies of your thesis submitted for examination and library deposit (including digital deposit), following your thesis Acknowledgements.

**Chapter 3: Quantifying separation and similarity in a Saccharomyces cerevisiae metapopulation**

<table>
<thead>
<tr>
<th>Nature of contribution by PhD candidate</th>
<th>Nature of contribution</th>
<th>Extent of contribution by PhD candidate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sarah Knight</td>
<td>Performed the experiments, analysed the data and wrote the manuscript</td>
<td>75</td>
</tr>
</tbody>
</table>

**CO-AUTHORS**

<table>
<thead>
<tr>
<th>Name</th>
<th>Nature of Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matthew R. Goddard</td>
<td>Designed the experiment, assisted in data analyses and wrote the paper</td>
</tr>
</tbody>
</table>

**Certification by Co-Authors**

The undersigned hereby certify that:
- the above statement correctly reflects the nature and extent of the PhD candidate’s contribution to this work, and the nature of the contribution of each of the co-authors; and
- in cases where the PhD candidate was the lead author of the work that the candidate wrote the text.

<table>
<thead>
<tr>
<th>Name</th>
<th>Signature</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matthew R. Goddard</td>
<td></td>
<td>24/10/2014</td>
</tr>
</tbody>
</table>

Click here
This form is to accompany the submission of any PhD that contains research reported in published or unpublished co-authored work. **Please include one copy of this form for each co-authored work.** Completed forms should be included in all copies of your thesis submitted for examination and library deposit (including digital deposit), following your thesis Acknowledgements.

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

**Chapter 4: Sporulation in soil as an over-winter survival strategy in Saccharomyces cerevisiae**

<table>
<thead>
<tr>
<th>Nature of contribution by PhD candidate</th>
<th>Sarah Knight: Designed and performed the experiments, analysed the data and wrote the manuscript</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extent of contribution by PhD candidate (%)</td>
<td>90</td>
</tr>
</tbody>
</table>

**CO-AUTHORS**

<table>
<thead>
<tr>
<th>Name</th>
<th>Nature of Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matthew R. Goddard</td>
<td>Designed the research, assisted in data analyses and provided comments on the text</td>
</tr>
</tbody>
</table>

**Certification by Co-Authors**

The undersigned hereby certify that:

- the above statement correctly reflects the nature and extent of the PhD candidate’s contribution to this work, and the nature of the contribution of each of the co-authors; and
- in cases where the PhD candidate was the lead author of the work that the candidate wrote the text.

<table>
<thead>
<tr>
<th>Name</th>
<th>Signature</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matthew R. Goddard</td>
<td></td>
<td>24/10/2014</td>
</tr>
</tbody>
</table>
Co-Authorship Form

This form is to accompany the submission of any PhD that contains research reported in published or unpublished co-authored work. **Please include one copy of this form for each co-authored work.** Completed forms should be included in all copies of your thesis submitted for examination and library deposit (including digital deposit), following your thesis Acknowledgements.

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

Chapter 5: Regional microbial signatures translate to differential wine phenotypes providing evidence for a microbial aspect to terroir

<table>
<thead>
<tr>
<th>Nature of contribution by PhD candidate</th>
<th>Sarah Knight: Designed and performed the experiments, analysed the data and wrote the manuscript</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extent of contribution by PhD candidate (%)</td>
<td>75</td>
</tr>
</tbody>
</table>

**CO-AUTHORS**

<table>
<thead>
<tr>
<th>Name</th>
<th>Nature of Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Steffen Klaere</td>
<td>Analysed the data</td>
</tr>
<tr>
<td>Bruno Fedrizzi</td>
<td>Contributed reagents and analytical tools</td>
</tr>
<tr>
<td>Matthew Goddard</td>
<td>Designed the research, assisted in cata analyses and wrote the manuscript</td>
</tr>
</tbody>
</table>

**Certification by Co-Authors**

The undersigned hereby certify that:
- the above statement correctly reflects the nature and extent of the PhD candidate’s contribution to this work, and the nature of the contribution of each of the co-authors; and
- in cases where the PhD candidate was the lead author of the work that the candidate wrote the text.

<table>
<thead>
<tr>
<th>Name</th>
<th>Signature</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Steffen Klaere</td>
<td></td>
<td>24/10/2014</td>
</tr>
<tr>
<td>Bruno Fedrizzi</td>
<td></td>
<td>24/10/2014</td>
</tr>
<tr>
<td>Matthew R Goddard</td>
<td></td>
<td>24/10/2014</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Click here</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Click here</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Click here</td>
</tr>
</tbody>
</table>

Last updated: 25 March 2013
This form is to accompany the submission of any PhD that contains research reported in published or unpublished co-authored work. Please include one copy of this form for each co-authored work. Completed forms should be included in all copies of your thesis submitted for examination and library deposit (including digital deposit), following your thesis Acknowledgements.

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

Chapter 6: Fungal diversity during fermentation is indicative of thiol concentrations in wine with indications that microbes may modulate these throughout fruit development

| Nature of contribution by PhD candidate | Sarah Knight: Collected the Saccharomyces diversity data and wine samples, analysed the data and wrote the manuscript |
| Extent of contribution by PhD candidate (%) | 80 |

**CO-AUTHORS**

<table>
<thead>
<tr>
<th>Name</th>
<th>Nature of Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Steffen Klaere</td>
<td>Analysed the data</td>
</tr>
<tr>
<td>Peter Morrison-Whittle</td>
<td>Collected the 454 fungal community data</td>
</tr>
<tr>
<td>Matthew Goddard</td>
<td>Conceived the analysis, provided comments on the manuscript</td>
</tr>
</tbody>
</table>

**Certification by Co-Authors**

The undersigned hereby certify that:
- the above statement correctly reflects the nature and extent of the PhD candidate’s contribution to this work, and the nature of the contribution of each of the co-authors; and
- in cases where the PhD candidate was the lead author of the work that the candidate wrote the text.

<table>
<thead>
<tr>
<th>Name</th>
<th>Signature</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Steffen Klaere</td>
<td></td>
<td>24/10/2014</td>
</tr>
<tr>
<td>Peter Morrison Whittle</td>
<td></td>
<td>24/10/2014</td>
</tr>
<tr>
<td>Matthew R. Goddard</td>
<td></td>
<td>24/10/2014</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Click here</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Click here</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Click here</td>
</tr>
</tbody>
</table>

Last updated: 25 March 2013
# TABLE OF CONTENTS

Abstract ii
Acknowledgements v
Co-authorship forms vii
List of Figures xvii
List of Tables xix

1 GENERAL INTRODUCTION ................................................................. 1

1.1 POPULATION GENETICS ..................................................................... 2
  1.1.1 Hardy-Weinberg principle ........................................................... 4
  1.1.2 Coalescent theory ...................................................................... 5
  1.1.3 Applying the theory to infer population patterns .......................... 6
  1.1.4 Microbial population genetics ....................................................... 8

1.2 SACCHAROMYCES CEREVISIAE ......................................................... 10
  1.2.1 Classification ........................................................................ 10
  1.2.2 Global distribution and potential origin ........................................ 11
  1.2.3 Domestication ......................................................................... 12
  1.2.4 Life cycle .............................................................................. 12
  1.2.5 Ecology ................................................................................ 13
  1.2.6 The crab tree effect and ecosystem engineering ......................... 14
  1.2.7 Dispersal .............................................................................. 15

1.3 WINEMAKING ............................................................................. 16
  1.3.1 Yeast and alcoholic fermentation for winemaking ....................... 16
  1.3.2 Yeast-derived aroma and flavours ............................................. 17
  1.3.3 Vitis vinifera var. Sauvignon Blanc ........................................... 18
  1.3.4 History of wine making in New Zealand ..................................... 19
1.4 OBJECTIVES ...................................................................................................................... 20

1.5 THESIS OVERVIEW ......................................................................................................................... 21

1.5.1 Chapter 2: “Methodology” ........................................................................................................... 21

1.5.2 Chapter 3: “Quantifying separation and similarity in a *Saccharomyces cerevisiae* metapopulation” ......................................................................................................................... 21

1.5.3 Chapter 4: “Sporulation in soil as an over-winter survival strategy in *Saccharomyces cerevisiae*” ............................................................................................................................................ 22

1.5.4 Chapter 5: “Regional microbial signatures translate to differential wine phenotypes providing evidence for a microbial aspect to terroir” ........................................................................................................... 22

1.5.5 Chapter 6: “Fungal diversity during fermentation is indicative of thiol concentrations in wine with indications that microbes may modulate these throughout fruit development” ............................................................................................................ 23

1.5.6 Chapter 7: “General discussion and future directions” ........................................................................... 23

2 METHODOLOGY ................................................................................................................................ 24

2.1 OBJECTIVE 1: Quantify the degree to which New Zealand’s *S. cerevisiae* metapopulation is genetically structured by both environment and geography and go on to estimate the extent to which sub-populations are connected by gene-flow (Chapter 3) ................................................................................................................................. 24

2.1.1 Sampling ........................................................................................................................................ 25

2.1.2 Genotyping .................................................................................................................................... 25

2.1.3 Analysis .......................................................................................................................................... 30

  *i. Summary statistics* ................................................................................................................................. 31

  *ii. Bayesian approaches* ........................................................................................................................... 32

  *iii. ObStruct* .......................................................................................................................................... 34

2.2 OBJECTIVE 2: Investigate the role of the woodland niche in the life cycle of *S. cerevisiae* (Chapter 4) ........................................................................................................................................ 37

2.2.1 Attempts at observing sporulation in soil itself ................................................................................... 38

2.2.2 Soil ‘tea’.......................................................................................................................................... 39

2.2.3 Statistical analysis ............................................................................................................................... 39

2.3 OBJECTIVE 3: Quantify the contribution of regional populations of *S. cerevisiae* to terroir or regional distinctiveness of wine (Chapter 5) ................................................................................. 39

2.3.1 Microferments .................................................................................................................................. 40

2.3.2 Chemical profiles ............................................................................................................................... 40
5. REGIONAL MICROBIAL SIGNATURES TRANSLATE TO DIFFERENTIAL WINE PHENOTYPES PROVIDING EVIDENCE FOR A MICROBIAL ASPECT TO TERROIR ........................................ 80
   5.1 INTRODUCTION ............................................................................. 82
   5.2 METHODS .................................................................................. 84
      5.2.1 Genotype selection ................................................................. 84
      5.2.2 Micro-fermentation ............................................................... 85
      5.2.3 Blends ................................................................................... 85
      5.2.4 Chemical analyses ................................................................. 86
      5.2.5 Statistical analyses ............................................................... 86
   5.3 RESULTS .................................................................................. 88
      5.3.1 Selection of S. cerevisiae genotypes ........................................ 88
      5.3.2 Ferment performance ............................................................ 88
      5.3.3 Chemical profiles produced by single genotype ferments correlate with region of microbe origin ............................................................................................................. 89
      5.3.4 Chemical drivers of regional differentiation in single ferment samples .......................................................................................................................... 95
      5.3.5 The genetic basis for differences in chemical profiles .......... 96
      5.3.6 The effect of regionally co-fermented genotypes and blended wines on volatile profiles ........................................................................................................... 98
   5.4 DISCUSSION ............................................................................. 99
6. FUNGAL DIVERSITY DURING FERMENTATION IS INDICATIVE OF THIOL CONCENTRATIONS IN WINE WITH INDICATIONS THAT MICROBES MAY MODULATE THESE THROUGHOUT FRUIT DEVELOPMENT ........................................... 103
   6.1 INTRODUCTION ............................................................................. 105
   6.2 METHODS .................................................................................. 107
      6.2.1 Sample collection and fermentation ....................................... 107
      6.2.2 Fungal community analysis of the juice ................................ 108
6.2.3  *Saccharomyces* species analysis at the end of ferment                        109
6.2.4  Thiol analysis                                                                 109
6.2.5  Statistical analysis                                                            110

6.3  RESULTS                                                                                   111

6.3.1  Dataset                                                                                   111
   i. *Community diversity in the starting juice*                                               111
   ii. *Saccharomyces species diversity at the end of ferment*                                 112
   iii. Thiol concentrations                                                                  112

6.3.2  Data reduction of species diversity in the starting juice                               112
6.3.3  The effects of species diversity on the concentration of 4MMP                          114
6.3.4  The effects of species diversity on the molar sum of 3MH and 3MHA                     115
6.3.5  A closer look at *Saccharomyces* species diversity                                     116

6.4  DISCUSSION                                                                                 117

7.  GENERAL DISCUSSION AND FUTURE DIRECTIONS                                                  120

7.1  RESEARCH OUTCOMES                                                                          121

7.1.1  Objective 1: Quantify the degree to which New Zealand’s *S. cerevisiae* metapopulation is genetically structured by both environment and geography and go on to estimate the extent to which sub-populations are connected by gene-flow (Chapter 3) 121

7.1.2  Objective 2: Investigate the role of the ‘woodland’ niche in the life cycle of *S. cerevisiae* (Chapter 4) 123

7.1.3  Objective 3: Quantify the contribution of regional populations of *S. cerevisiae* to terroir or regional distinctiveness of wine (Chapter 5) 124

7.1.4  Objective 4: Investigate whether the fungal community in grape juice and the species of *Saccharomyces* driving the ferment affect final thiol concentration in wine (Chapter 6) 125

7.2  SYNTHESIS                                                                                 126

7.3  FUTURE DIRECTIONS                                                                         127

7.3.1  Population patterns                                                                     127

7.3.2  Ecology                                                                                130
LIST OF FIGURES

CHAPTER 1
Figure 1.1: An example of how coalescent theory can be used to infer migration ...................................................... 6

CHAPTER 2
Figure 2.1: Samples used to test the sources of error obtained during the genotyping of S. cerevisiae ... 29

CHAPTER 3
Figure 3.1: The location of New Zealand (NZ) sampling regions and analyses of population structure and connectivity ......................................................................................................................... 60

CHAPTER 4
Figure 4.1: The average proportion of sporulated cells on each of the three plate types in the pilot study ....................................................................................................................................................... 75
Figure 4.2: The average proportion of the total number of cells sporulated as well as the proportion of tetrads and dyads within that total across the eight days of the time course experiment for both the (a) plain agar and (b) soil agar plates ......................................................................................................................... 76

CHAPTER 5
Figure 5.1: A map of the regions the tested genotypes of S. cerevisiae were isolated from and the pairwise similarity in the chemical profiles produced by each .......................................................................................................................... 92
Figure 5.2: CCA analysis of the 105 single genotype ferments analysed. (a) All sample points colored by region. (b) Regional averages and 50 % ellipses ........................................................................................................................................ 94
Figure 5.3: Visualisation of the chemicals that individually explain more than 25 % of the regional variation as calculated by ANOVA analyses .............................................................................................................. 97

CHAPTER 6
Figure 6.1: A representation of the OTUs (points) comprising each component from the PLSR analysis. The large circles provide a scale for the magnitude of the loading vectors. The positions of the points represent the magnitude of the effect each OTU has on that component. OTUs with a magnitude larger than 0.2 are classified in Table 6.1 ........................................................................................................................................... 113

APPENDIX I
Figure AI.1: Map of New Zealand giving an indication of the scale between the geographic regions of sampling .................................................................................................................................................. 158
Figure AI.2: A map of the sampling locations in Hawke’s Bay with sample codes .......................................................................................................................... 159
Figure AI.3: A map of the sampling locations in Martinborough with sample codes ................................................................. 160
Figure AI.4: A map of the sampling locations in Nelson with sample codes ................................................................................................. 161
Figure AI.5: A map of the sampling locations in the Wairau Valley with sample codes ................................................................. 162
Figure AI.6: A map of the sampling locations in the Awatere Valley with sample codes ................................................................. 163
Figure AI.7: A map of the sampling locations in Central Otago with sample codes ................................................................................................. 164
APPENDIX II

Figure AII.1: Principal Coordinates Analysis investigating the genetic distance between individuals isolated from spontaneous ferment samples (blue) and environmental samples (red) encompassing vineyard soil and soil and fruit from native trees.

Figure AII.2: DISTRUCT plots. Vertical bars represent individuals and the colours denote inferred populations. Individuals to the left of the black line were isolated from spontaneous ferments and individuals to the right were isolated from vineyard soil and soil and fruit from native trees species. $R^2$ values and $P$-values were calculated using ObStruct.

Figure AII.3: Plots from the canonical discriminant analysis performed in ObStruct. (a) Mapping the median and 66% ellipse for each sampled population. Populations with higher degrees of overlap are more genetically similar. (b) Mapping the individual data from all 369 isolates analysed. The distance between points is indicative of the similarity in ancestry profiles.

APPENDIX III

Figure AIII.1: CCA analysis visualising the differences between co-ferments and wines blended from the single genotype ferments.

Figure AIII.2: Plots from the linear transformations used to visualise the co-ferments and blends in respect to the single genotype ferments for each region.

Figure AIII.3: The effect imposed by a single genotype with mixed ancestry derived from the Nelson region.
LIST OF TABLES

CHAPTER 3
Table 3.1: Summary of the populations isolated from each region .......................................................... 58

CHAPTER 5
Table 5.1: Summary of all PERMANOVA analyses .................................................................................. 91

CHAPTER 6
Table 6.1: The OTUs with loading vectors above 0.2 in the PLSR analysis, the components they relate to and their classification ...................................................................................... 114
Table 6.2: The results from the linear model against 4MMP concentration with the factors of Saccharomyces diversity, region and the six principle components representing the OTU diversity in the initial juice samples .................................................................................... 115
Table 6.3: The results from the linear model against the molar sum of 3MH and 3MHA with the factors of Saccharomyces diversity, region and the six principle components representing the OTU diversity in the initial juice samples .................................................................................... 116

APPENDIX II
Table AII.1: Further details about the samples obtained from Sauvignon Blanc vineyard sites across New Zealand .......................................................................................................................... 168
Table AII.2: Further details about the samples obtained from non-managed or native sites across New Zealand including the species of tree they were isolated from .............................................................................. 171
Table AII.3: Pairwise R² values from Obstruct ............................................................................................ 174
Table AII.4: Mean directional estimates of migration rate (Nm) as calculated by MIGRATE with 95% confidence intervals ........................................................................................................ 174

APPENDIX III
Table AIII.1: Regional pairwise PERMANOVA results showing the R² value for “region” and associated P-values. P-values corrected using a Benjamini-Hochberg correction for multiple hypothesis tests are also reported .......................................................................................... 178
Table AIII.2: Results from individual ANOVA analyses for each chemical using the single genotype ferments. R² values and the associated P-value for the factor “region” are reported ........................................................................ 179
Chapter 1: General Introduction

GENERAL INTRODUCTION

Microbes are a vital part of any ecosystem and play key roles in processes essential to the functioning of the biosphere (Kirchman, 2012). While many of these processes are vital to the health and productivity of agricultural products, microbes have also been adopted for more direct commercial purposes such as the conversion of plant materials to economically and socially important commodities including wine, bread, beer, spirits, sake, coffee, and chocolate and thus have wide reaching impacts on global economies (Fleet, 2006). Additionally, microbes are widely used as model research organisms to understand complex biological processes in a simple yet comparable system. While there is a vast and ever increasing knowledge of the cell biology, genetics and evolutionary mechanisms of many of these organisms, their ecology and population processes and subsequent impacts on agricultural industries is largely unknown.

Microbial ecology is a rapidly growing field, more recently driven by advances in genetic sequencing technology. While the ecologies of both macro- and microorganisms are jointly shaped by evolution and differential reproductive success (Atlas and Bartha, 1998) the application of theory to microbial ecology is largely lacking and there is sparse knowledge on whether ecological theory developed for macro-organisms generalise to microorganisms (Martiny et al., 2006; Prosser et al., 2007). In this thesis I investigate the population patterns and ecology of the commercially and fundamentally important yeast species *Saccharomyces cerevisiae* and test whether these patterns, as well as patterns of yeast community
Chapter 1: General Introduction

dissimilarity, affect wine styles in New Zealand. In doing so I aim to contribute to the growing knowledge of
the ecology of microbial species and to thus broaden our understanding of the evolutionary history and
future trajectories of these ecologically important species. Using population genetic analyses, the extent of
population differentiation and connectivity between native and cultivated niches as well as between
different geographic regions is quantified, providing the most comprehensive analysis of the population
patterns of this species to date. From here, the ecology of *S. cerevisiae* in the absence of fruit and how this
phase of the year affects its life cycle is investigated by assessing its propensity to sporulate in the
presence of soil. Whether the observed geographic differences in the *S. cerevisiae* population translate into
regionally distinct chemical profiles in wine is experimentally tested in *Vitis vinifera* var. Sauvignon Blanc,
providing the first objective evidence of a microbial aspect to the commercially important concept of *terroir*
(or sense of place). Finally, whether the species composition in the grape juice and the *Saccharomyces*
species that drive the ferment can predict the final concentration of volatile thiols in wine is modelled,
revealing the importance of microbial species on the chemical composition of a wine not only during the
fermentation process but throughout the fruit growing season.

This thesis contains chapters published or written in collaboration with co-authors. As per statute
regulation, co-authorship forms are provided following the acknowledgements at the start of this thesis.

This introductory chapter is divided into four main sections that detail 1) basic population genetic theory
and how it can be applied to microbes, 2) *S. cerevisiae* as a focal species for this thesis, 3) winemaking and
its history in New Zealand and 4) an outline of the objectives addressed by this thesis.

## 1.1 POPULATION GENETICS

Population patterns and processes occurring in natural populations have implications for the evolutionary
trajectory of species. If we are to better understand the ecology and evolution of microbial species, it is
crucial that we understand how their populations are structured and the frequency at which they interact
and exchange genetic material. While observational methods of inferring population patterns such as subdivision and connectivity are important for larger organisms, they cannot be used for microbial species due to their inherent small size; however, population genetics provides a powerful tool for inferring these processes for all forms of life.

Population genetics is the study of alleles and allele frequencies in populations. It can be used both predictively to infer future changes in a population and retrospectively to understand the processes that have resulted in the observed genetic composition of a population (Hartl and Clark, 1997; Nielsen and Slatkin, 2013). Understanding how and why these allele frequencies change has implications for understanding the processes driving evolution (Nielsen and Slatkin, 2013). Mathematical models both predicting patterns of allele frequencies in populations and inferring genealogical histories of individuals have been developed. These can be used to infer patterns of population subdivision and connectivity and ultimately to objectively test hypotheses regarding the evolutionary forces operating on a population. These mathematical models set their foundations on the Wright-Fisher model, named after its founders Sewall Wright and Ronald A. Fisher. In its simplest form, a Wright-Fisher population describes a population of constant size $2N$ (to represent a diploid population with $N$ individuals) with non-overlapping generations and no selection, mutation or migration and thus provides a mathematical basis on which to build potential hypotheses (Nielsen and Slatkin, 2013).

Selection and genetic drift are the two main processes driving changes in allele frequencies over time. While selection occurs when an allele in a population confers either an advantage or disadvantage to the individual possessing it, ultimately leading to differential reproductive success, genetic drift is a neutral process resulting from random fluctuations in allele frequencies due to chance differences in the number of offspring left by individuals of different genotypes (Griffiths et al., 2005; Nielsen and Slatkin, 2013). The rate at which genetic drift affects allele frequencies is dependent on population size, with large changes more
likely in small populations. Selection and genetic drift are not mutually exclusive and the random effects of genetic drift can counterintuitively result in advantageous alleles being lost and deleterious alleles becoming fixed in a population, particularly in small populations where the effects of genetic drift are stronger (Nielsen and Slatkin, 2013). When populations become subdivided, these processes of natural selection and genetic drift can result in the genetic differentiation of these subpopulations. Here I outline two major models in population genetics that are used in the inference of population differentiation (or population structure) and population connectivity of a species. The first is the Hardy-Weinberg Principle which acts a null hypothesis for a randomly mating population, and the second is coalescent theory which provides a basis for inferring the genealogical history of alleles and individuals.

1.1.1 Hardy-Weinberg principle

The Hardy-Weinberg principle is an important tool in population genetics and states that allele frequencies within a population will remain constant in the next generation unless one of the assumptions of the principle is violated (Chen, 2010; Nielsen and Slatkin, 2013). These assumptions include that mating is random within a population, the population is infinitely large, there is no gene-flow from other populations, there is no mutation, and natural selection is not acting on the population so that all individuals have equal probabilities of survival and reproduction (Hartl and Clark, 1997; Futuyma, 1998; Chen, 2010; Nielsen and Slatkin, 2013). In a sexual population with non-overlapping generations and equal proportions of males and females (analogous to a Wright-Fisher population explained above), Hardy-Weinberg Equilibrium (HWE) can be reached after only one generation of random mating (Hartl and Clark, 1997; Nielsen and Slatkin, 2013). This allows the Hardy-Weinberg principle to essentially act as a null hypothesis in which to test deviations of allele frequencies against. There are a number of explanations that may be plausible if a sample from a population deviates from what is expected under HWE, including 1) assortative mating where the same or similar genotypes are more likely to mate, 2) inbreeding where closely related individuals which inevitably have similar genotypes are more likely to mate, 3) population structure where
individuals within each sub-population are more likely to mate with each other than individuals from
different sub-populations, and 4) selection where there is differential survival and reproductive success of
individuals with certain genotypes (Nielsen and Slatkin, 2013). All of these scenarios are essentially a
deviation from random mating within the entirety of the population.

1.1.2 Coalescent theory

Rather than being predictive in terms of what allele frequencies we would expect to observe under a set of
assumptions, coalescent theory makes inferences about the genealogical history of a sample (Kingman,
1982; Nielsen and Slatkin, 2013). Again based on a Wright-Fisher model population, coalescent theory
estimates the time to the most recent common ancestor (or coalescent time) between all gene copies using
models of genetic drift. To better understand how coalescence works, consider Figure 1.1a. Starting in the
past and moving forward in time, there are a number of alleles in each generation with equal probabilities
of being transmitted to the next generation as expected under a model of random genetic drift. Some of
these alleles are passed forward to the next generation one or more times while others are not and are
consequently lost from the population. In the absence of new alleles arising either by mutation or
migration, eventually one allele in the population becomes fixed (Hartl and Clark, 1997). Therefore, using
coalesscent theory and models of genetic drift, the genealogy of all of the present day sampled gene copies
can be estimated by using either Bayesian or maximum likelihood inferences of the time since the most
recent common ancestor for all pairwise comparisons between samples. Since mutations naturally occur
throughout a species’ history, a number of alleles that may be descended from a single common ancestral
allele may no longer be identical; thus models of mutation are often built in to coalescent models.
Chapter 1: General Introduction

1.1.3 Applying the theory to infer population patterns

By knowing how we would expect a population to behave under defined conditions, we can test the genetic patterns we observe to make inferences about the processes operating within that population. Populations are naturally structured due to the distribution of suitable habitats and many organisms typically form aggregations such as herds, flocks, schools and colonies (Hartl and Clark, 1997). Through processes such as natural selection, genetic drift and founder effects, these subpopulations become genetically differentiated over time, resulting in population subdivision (Hartl and Clark, 1997). In contrast to these processes resulting in population subdivision, migration or gene-flow allows genetic variation to be shared among subpopulations, inhibiting population subdivision and thus having direct consequences for the evolutionary
trajectory of a species. It is therefore important to also quantify the level and direction of migration to understand the evolutionary dynamics of a population (Yamamichi and Innan, 2012).

The extent of subdivision, and the patterns of migration between subpopulations, can be examined using both gene frequency and coalescent approaches. When allele frequencies differ between subpopulations it becomes less likely that different alleles prevalent in different subpopulations will come together to form a heterozygote; thus the average total heterozygosity is less than expected under HWE (Hartl and Clark, 1997). Known as the Wahlund Effect (Wahlund, 1928), this can be used to detect population subdivision in a population that otherwise appears homogeneous. Migration or gene flow between populations can be estimated using allele frequencies through a simple model of population structure called the island model (Wright, 1931) that suggests a relationship between estimates of population differentiation such as $F_{ST}$ (discussed in more detail in Section 2.1.3) and the number of migrants per generation. The island model, however, makes a number of assumptions which are often not true in natural systems and this can provide misleading estimates of migration which should be interpreted with caution (Whitlock and McCauley, 1999).

Coalescent theory also provides the tools necessary to detect and quantify gene-flow between populations by assessing the genealogical histories of individuals sampled from each population. Migration can be detected if an individual that was sampled from one population (i.e. a particular geographic location) is considered to be genetically more closely related to individuals sampled from a second population (i.e. a different location; Figure 1.1). Therefore it can be assumed that the individual sampled from the first population originated from the second population (based on genetic relatedness) and had migrated to the first. Different loci may show slightly different patterns so by analysing a suite of unlinked loci, migration patterns through the genealogical history of the sampled individuals can be inferred to calculate the rate of migration.
Chapter 1: General Introduction

1.1.4 Microbial population genetics

Population processes occurring in microbial species remain relatively unknown despite their importance to ecosystem function (Martiny et al., 2006; Gerstein and Moore, 2011; Kirchman, 2012). With advances in genetic sequencing technologies, it has only become possible to make inferences about microbial population patterns in more recent scientific history since they cannot be directly observed. Population genetic theory however, is largely derived from diploid, out-crossed species while microbes are typically asexual or those that have a sexual phase are highly inbred. This violates the assumption of random mating imposed by many population genetic analyses as inbreeding is a type of non-random mating. Non-random mating can affect the frequency of alleles and total heterozygosity observed in a population (Hartl and Clark, 1997; Nielsen and Slatkin, 2013), and can thus affect estimates of population structure and migration. In fact, the effects of inbreeding in microbial species do not even follow predictions made for ‘macrobes’ or higher eukaryotes due to the nature of their sexual reproduction. For example, in ascomycetes yeast, meiosis results in the formation of four haploid spores enclosed in an ascus. These spores are highly likely to fuse or mate with others of the opposite mating type within the same ascus upon germination (Johnson et al., 2004). Therefore they are not mating with a close relative as would be the case in an inbred higher eukaryote population, but they are mating with a gamete from the same meiotic division. This has consequences for the rate at which heterozygosity is lost (Johnson et al., 2004). In other systems such as plants, selfing results in the fusion of gametes from independent meiotic events resulting in a reduction in heterozygosity by one-half every generation; however in these microbial populations selfing events are more likely to occur between spores from the same meiotic division, resulting in a reduction of heterozygosity of only one-third as each haploid spore from a meiotic division of a heterozygous cell shares an allele with only one other spore in the ascus and is different from the other two (Johnson et al., 2004).
Chapter 1: General Introduction

It is often hypothesised that microbial populations do not show the same biogeographic patterns and population structure seen in larger organisms due to their large population sizes, low extinction rates and high dispersal rates (Fenchel and Finlay, 2004; Gerstein and Moore, 2011). Additionally, population differentiation is predicted to decrease in increasingly asexual organisms such as microbes as clonal reproduction has theoretically been shown to increase heterozygosity at a single locus but decrease overall genotypic diversity in a population compared to sexual reproduction (Balloux et al., 2003). Despite these predictions, it has been argued that microbial species do show endemism and vicariant distributions (Lachance, 2004) and with the use of more informative genetic markers (Arnaud-Haond et al., 2005) there is increasing evidence for population subdivision in microbial populations reminiscent of that observed in larger species (Gerstein and Moore, 2011).

The scale at which non-pathogenic microbial populations are structured is largely unknown, as too are the forces driving differentiation. Some studies suggest that ecological function is more important in defining microbial populations than geography (Fay and Benavides, 2005; Legras et al., 2007; Liti et al., 2009); however, there is increasing evidence of differing degrees of population subdivision by geography (Buschbom, 2007; Gayevskiy and Goddard, 2012; Lowe et al., 2012). The nature of microbial reproduction can also have consequences on the scale at which populations are structured and ultimately the evolution of a species. While it seems logical that microbes are easily dispersed due to their small size, the extent and frequency of this dispersal is unknown. Additionally, since microbes have a high propensity to self-fertilise, dispersed individuals may not outcross with other individuals in the population and thus could create very small isolated populations entirely derived from one cell, resulting in very small scales of population structure.

Despite the current constraints of inferring accurate population patterns in microbial species, we can build and refine methods for analysis as we advance our understanding of these processes. Revealing these
population patterns and accumulating knowledge of the ecology of microbial species has vast consequences for our understanding of microbial evolution. Therefore by focusing on these population patterns in the model research eukaryote S. cerevisiae, we can draw on the comprehensive knowledge of this species’ biology to help understand the patterns we observe.

1.2 SACCHAROMYCES CEREVISIAE

Saccharomyces cerevisiae, a budding yeast, has had a close association with humans for over 7000 years due to its fermentative capabilities and has thus come to be of great commercial importance in industries involved in the production of breads, wine, beer and other alcoholic beverages (McGovern et al., 1996; Cavalieri et al., 2003; This et al., 2006; Chambers and Pretorius, 2010; Goddard, 2010a). Since the discovery that fermentation was a biological process conducted by microorganisms such as S. cerevisiae (reviewed in: Barnett, 1998 and Pretorius, 2000), this yeast has become a model eukaryote for scientific research and is widely used in fields such as cell biology, genetics, and increasingly ecology and evolution (Jessup et al., 2004; Chambers and Pretorius, 2010; Dujon, 2010; Gray and Goddard, 2012; Hittinger, 2013; Hyma and Fay, 2013) and was the first eukaryote genome to be fully sequenced (Goffeau et al., 1996). S. cerevisiae is easily stored, has a short generation time, a relatively small genome (~12 Mb) and is capable of both asexual and sexual reproduction making it ideal to investigate intricate evolutionary hypotheses using a simple eukaryote system. While much is known about S. cerevisiae’s biology, comparatively little is known about the population patterns and ecology of this species.

1.2.1 Classification

Species identification of S. cerevisiae can readily be performed using genetic approaches (de Melo Pereira et al., 2010). Before DNA sequencing technologies were readily available yeast taxonomy was based on morphological, reproductive and physiological characteristics (Replansky et al., 2008; Hittinger, 2013).
Chapter 1: General Introduction

Unfortunately many of these characteristics did not provide sufficient information and were often subject to homoplasy, resulting in the frequent misclassification of species to the genus *Saccharomyces* (Hittinger, 2013). Advances in molecular identification technologies have begun to clarify many of these issues and it is currently accepted through interbreeding and genetic analyses that the *Saccharomyces sensu stricto* complex consists of seven naturally occurring species including *S. cerevisiae*, *S. paradoxus*, *S. uvarum*, *S. eubayanus*, *S. arboricola*, *S. mikatae* and *S. kudriavzevii* (Liti et al., 2006; Naumov et al., 2010; Libkind et al., 2011; Naumov et al., 2011; Hittinger, 2013). Natural hybrids also exist within this group between *S. cerevisiae* and *S. kudriavzevii*, *S. cerevisiae* and *S. eubayanus* (known as *S. pastorianus*), and between *S. cerevisiae*, *S. eubayanus* and *S. uvarum* (known as *S. bayanus*) and are largely used industrially for beer brewing and cold wine fermentation (Hittinger, 2013).

### 1.2.2 Global distribution and potential origin

Evidence suggests that *S. cerevisiae* has a global distribution and inhabits a variety of ecological niches, both natural and agricultural or cultivated (Wang et al., 2012; Hittinger, 2013). *S. cerevisiae* is differentiated from *S. paradoxus*, its next closest relative, by approximately 15 % sequence divergence (Cliften et al., 2001; Liti et al., 2006) and it is estimated using currently accepted mutation rates and generation times that these two species diverged between 0.4 and 3.4 MYA (Liti et al., 2006). *S. cerevisiae* has naturally been found in sympatry with *S. paradoxus* at a woodland site in North America (Sniegowski et al., 2002), and with *S. paradoxus* and *S. kudriavzevii* on Oak (*Quercus sp.*.) bark in Portugal (Sampaio and Gonçalves, 2008).

While the origin of the *Saccharomyces* genus itself is debated (Libkind et al., 2011; Wang et al., 2012; Bing et al., 2014), evidence suggests that *S. cerevisiae* originated in Far East Asia as great genetic diversity has been observed in this region including basal lineages only isolated in China (Wang et al., 2012).
Chapter 1: General Introduction

1.2.3 Domestication

For many years now researchers have debated over the extent to which \textit{S. cerevisiae} is domesticated, defined as a species that is genetically distinct from its ancestors due to artificial or human-mediated breeding targeted to increase its usefulness to humans, and is therefore removed from the taming of wild-born individuals (Diamond, 2002). Typically the argument for domestication is that the diversity of strains found in vineyard and wine associated areas is low (Fay and Benavides, 2005). Humans have been making wine and cultivating vines for over 7000 years (McGovern et al., 1996; Cavalieri et al., 2003; This et al., 2006). This has provided a long evolutionary time period for \textit{S. cerevisiae} to be influenced by human farming and wine making practices and thus it would be expected that this has undoubtedly had some effect on its evolutionary trajectory. While it is not thought that \textit{S. cerevisiae} is domesticated in terms of the above definition, evidence of the effects of human activity and human associated gene-flow can been seen in previous comprehensive genetic analyses (Legras et al., 2007; Liti et al., 2009; Wang et al., 2012).

1.2.4 Life cycle

\textit{S. cerevisiae} is usually found in a diploid state and is capable of reproducing both asexually and sexually. When conditions are nutritionally and environmentally favourable the cells undergo asexual reproduction or mitosis via budding. Budding produces a daughter cell made from new cell surface material that eventually separates and is not to be confused with fission where a cell enlarges and subsequently pinches into two daughter cells (Herskowitz, 1988). The daughter cell is smaller than the parent and is required to increase in size before it can undergo budding itself. With sufficient nutrients, this cycle takes approximately 100 minutes to complete (Herskowitz, 1988).

The sexual cycle or sporulation requires the cell to be diploid and is a starvation response triggered by the absence of a nitrogen source combined with the presence of a non-fermentable carbon source (Esposito...
Chapter 1: General Introduction

and Klapholz, 1981; Honigberg and Purnapatre, 2003; Neiman, 2005; Taxis et al., 2005; Neiman, 2011). This
meiotic division typically results in four haploid spores being produced which are subsequently enclosed in
an ascus (Esposito and Klapholz, 1981; Neiman, 2011). Spores, like stationary phase cells, are much more
resistant to environmental stresses such as repeated freeze-thaw cycles, increased osmolarity and
desiccation; however, spores alone have a higher tolerance to short periods of heat shock, high salt
concentrations and extremes of pH (Coluccio et al., 2008).

The haploid spores produced are either \textit{MATa} or \textit{MAT\textalpha} mating type and can also reproduce via budding.
Diploid cells are formed again from the fusion of two haploid cells of opposite mating types. These cells can
be from the same ascus (inbreeding) or from asci of different mother cells (outcrossing). Additionally, if the
cells are homothallic, the haploid cell can bud and produce a daughter cell and then switch its own mating
type through the production of a site-specific endonuclease (Kostriken and Heffron, 1984). The mother and
daughter cell immediately mate or fuse, creating an entirely homozygous genome. This mate-type
switching is only observed in the mother cell and none of its haploid daughter cells (Klar, 1987). Most cells
that are found in the wild are homothallic, although heterothallic cells that cannot mate-type switch also
exist (Landry et al., 2006).

1.2.5 Ecology

In the natural environment and early grape must fermentation \textit{S. cerevisiae} is very rare (Mortimer and
Polsinelli, 1999; Pretorius, 2000; Xufre et al., 2006; Goddard, 2008; Taylor et al., 2014). It has been
estimated that \textit{S. cerevisiae} is present on approximately one in one thousand grape berries, increasing to
one in four if the berries are damaged (Mortimer and Polsinelli, 1999). Fruit however, is ephemeral and
there is limited knowledge about where \textit{S. cerevisiae} may reside and what it’s doing when fruit is not
available. \textit{S. cerevisiae} has been isolated from native woodland environments (tree bark and soil) in the
northern hemisphere (Sniegowski et al., 2002; Wang et al., 2012; Hyma and Fay, 2013), exotic \textit{Quercus}
species in New Zealand (NZ) and from *Nothofagus* in Patagonia (Zhang et al., 2010; Libkind et al., 2011). These populations have been shown to be genetically homogenous with nearby vineyard populations (Goddard et al., 2010; Hyma and Fay, 2013), suggesting they are part of one connected population. Whether *S. cerevisiae* is actively growing or dormant in niches other than fruit is unknown.

### 1.2.6 The crab tree effect and ecosystem engineering

Biologically, *S. cerevisiae* and the *Saccharomyces sensu stricto* group is an oddity. Known as the Crabtree effect, these yeasts preferentially ferment in the presence of oxygen rather than respire when sugar is in excess, despite the latter being energetically more efficient (van Dijken et al., 1993; Thomson et al., 2005; Piskur et al., 2006; Merico et al., 2007). The energy generated by fermentation is in the form of adenosine triphosphate (ATP). Yeast metabolise sugar to pyruvate through the glycolytic pathway which is subsequently decarboxylated to acetaldehyde and carbon dioxide (Campbell and Farrell, 2003). Acetaldehyde is then reduced to ethanol, producing a net gain of two ATP molecules. This is extremely inefficient considering aerobic respiration yields a net gain of 32 ATP molecules (Campbell and Farrell, 2003). Since this imposes such a large cost on these species, there must be an advantage gained from this behaviour that has allowed this trait to persist through evolutionary time (Thomson et al., 2005). One explanation put forward to explain this behaviour is that *S. cerevisiae* and other species of the *sensu stricto* group uses fermentation as a niche engineering tool (Goddard, 2008). *S. cerevisiae* is a ripe fruit specialist and ethanol and heat production during fermentation may be a way of defending this resource (Goddard, 2008). Ethanol is both a general antimicrobial and a deterrent to many vertebrates (Janzen, 1977; Thomson et al., 2005). Non-*Saccharomyces* species are seen to decrease in frequency in the ferment as the levels of ethanol rise; however, some non-*Saccharomyces* species also show a high tolerance to ethanol so this alone may not be enough to exclude them (Gao and Fleet, 1988; Pina et al., 2004). Heat is also produced during fermentation and has been shown to have a greater impact on the fitness of *S. cerevisiae* than ethanol production alone, suggesting that this additional dimension of ecosystem engineering is utilised by
Chapter 1: General Introduction

*S. cerevisiae* to dominate the ferment environment (Goddard, 2008). This provides an evolutionary explanation to why *Saccharomyces* species exhibit the Crabtree effect and how it has persisted in this genus.

### 1.2.7 Dispersal

*S. cerevisiae* is sessile and unable to disperse without a vector; however, ecosystem engineering in *S. cerevisiae* has recently been reported to facilitate a mutualistic interaction with *Drosophila* (Buser et al., 2014). During fermentation *S. cerevisiae* produces a variety of energetically costly volatile compounds of which the biological function is unclear. While it has been suggested that some of these compounds are part of detoxification mechanisms, or serve as an intermediate to other important metabolites (Saerens et al., 2010), recent evidence suggests they also play a role in the attraction of potential insect vectors (Becher et al., 2012; Palanca et al., 2013; Buser et al., 2014). Buser et al. (2014) were able to show experimentally that there is variation in *Drosophila*’s attraction to the odorants produced by different *S. cerevisiae* strains in grape juice and the attractiveness of the strains is positively correlated with yeast dispersal and fly fecundity. Thus, this mutualistic interaction provides a mechanism allowing *S. cerevisiae* to disperse. There are many other reports of associations of *S. cerevisiae* with other insects such as bees and wasps (Mortimer and Polsinelli, 1999; Goddard et al., 2010; Stefanini et al., 2012) and it is possible that similar interactions exist here too. A study conducted by Stefanini et al. (2012) has shown that *S. cerevisiae* can survive over winter with queens of social wasps and that the queens transmit *S. cerevisiae* to their progeny the following spring, providing an additional mode of transport for these yeast.

Evidence has also been presented for dispersal of *S. cerevisiae* by human vectors. Over 50 isolates of *S. cerevisiae* representing 40 different genotypes were found in New Zealand in a new oak barrel imported from Chagny in the Burgundy region of France (Goddard et al., 2010). In fact strains associated with different technological uses are more closely related to one another than to other strains, even when
isolated from different geographic locations (Legras et al., 2007; Liti et al., 2009; Wang et al., 2012). It is therefore highly likely that \textit{S. cerevisiae} was and is transported by the activities of humans.

### 1.3 WINEMAKING

#### 1.3.1 Yeast and alcoholic fermentation for winemaking

\textit{Saccharomyces} species have been utilised by humans for their fermentative properties for thousands of years. The earliest evidence of wine production dates back to between 7400 – 7000 BP from a pottery jar found in a Neolithic village in Iran’s northern Zargros mountains (McGovern et al., 1996; This et al., 2006). Yeasts are essential for the transformation of grape must to wine, converting sugar in the must to ethanol and carbon dioxide while producing a range of metabolites important to the final aroma and flavour.

Traditionally wine is fermented naturally or spontaneously by a succession of microbes naturally present in the grape must with a diversity of non-\textit{Saccharomyces} yeasts present in the early stages of fermentation, which diminish and are replaced by \textit{Saccharomyces} yeasts, typically \textit{S. cerevisiae} (Pretorius, 2000; Xufre et al., 2006; Goddard, 2008). This eventual dominance of \textit{Saccharomyces} species is driven by their adaptive ability to engineer its environment through fermentation (explained in Section 1.2.6). While this method of winemaking can be unpredictable and risks the production of off-flavours or the onset of stuck and sluggish fermentations, it is often preferred for the greater complexity and unique sensory profiles potentially produced by the indigenous yeast (Pretorius, 2000; Mateo et al., 2001; Fleet, 2003).

Commercial yeast strains, typically \textit{S. cerevisiae}, are often used to inoculate grape must to initiate and complete fermentation. The use of these commercial strains in the wine making process significantly reduces the proportion of indigenous yeasts, with higher levels of inoculum resulting in a more rapid reduction in non-\textit{Saccharomyces} yeasts (Erten et al., 2006). While these types of ferments have improved reproducibility and predictability in terms of the quality of the wine produced, there is a common
perception that these wines are ordinary with generic aromas and flavours and do not possess the same complexity as spontaneously fermented wine (Rainieri and Pretorius, 2000).

### 1.3.2 Yeast-derived aroma and flavours

During fermentation yeasts produce a range of volatile compounds, both novel and from the conversion of odourless precursors in the juice, all of which are important to the wines aroma and flavour and thus final quality (Lambrechts and Pretorius, 2000; Swiegers and Pretorius, 2005; Swiegers et al., 2005a). The potency of fermentation volatiles primarily depends on their odour detection thresholds and therefore volatiles with low thresholds but produced in small concentrations, such as thiols, can significantly affect wine sensory characteristics. Yeast-derived compounds in wine include esters, higher alcohols, carbonyls, volatile fatty acids, sulfur compounds and more and are derived from sugar and amino acid metabolism (Swiegers et al., 2005a).

Different compound classes have different effects on the organoleptic properties of wines. Esters typically impart a fresh fruity aroma and flavour in wine and are derived from lipid and acetyl-CoA metabolism (Swiegers and Pretorius, 2005; Swiegers et al., 2005a). Higher alcohols or fusel alcohols can be detrimental to wine flavour and aroma at high concentrations, resulting in a strong, pungent smell and taste; however at concentrations below 300 mg/L they add to the complexity and fruity sensory properties (Swiegers and Pretorius, 2005; Swiegers et al., 2005a). They are also important precursors for the production of esters, aldehydes and acids (Swiegers and Pretorius, 2005). Some carbonyl compounds such as volatile aldehydes can contribute to sensory properties of apple, citrus and nuts (Swiegers and Pretorius, 2005). Sulfur compounds generally have a negative effect on wine sensory properties, contributing descriptors such as cabbage, rotten egg, garlic, onion and rubber (Mestres et al., 2000); however some sulfur compounds are pleasant at low concentrations including some volatile thiols that impart tropical characteristics (Tominaga et al., 1998; Swiegers et al., 2005a).
Chapter 1: General Introduction

The aroma and flavour compounds and their concentrations in the final wine is dependent on a multitude of interacting factors including, but not limited to, grape and must composition, precursor availability, fermentation temperature, oxygen level and the presence of different yeast species and strains (Mauricio et al., 1997; Murat et al., 2001; Bell and Henschke, 2005; Anfang et al., 2009; Swiegers et al., 2009; Dennis et al., 2012). These factors interact to produce distinctive wine styles important for the economic viability of a product.

1.3.3 *Vitis vinifera* var. Sauvignon Blanc

Sauvignon Blanc is a white variety of wine typically described as either ‘green’ referring to flavours and aromas of capsicum, tomato leaf, asparagus and grass, or ‘tropical’ referring to passion fruit, citrus, grapefruit, gooseberry (Swiegers et al., 2009; Coetzee and du Toit, 2012). While the ‘green’ characters are largely derived from methoxypyrazines produced by the plant itself, the ‘tropical’ characters are predominately derived from volatile thiols and esters produced by yeast during fermentation (Swiegers et al., 2009). For Sauvignon Blanc three of the most important thiols have been identified as 4-mercapto-4-methyl-pentan-2-one (4MMP) that contributes to the box tree, passion fruit, broom and black current bud aromas, as well as 3-mercapto-hexan-1-ol (3MH) and 3-mercapto-hexan-1-ol acetate (3MHA) that contribute to the passion fruit, grapefruit, guava and citrus flavours and aromas; however these perceptions are concentration dependent and in high concentrations have been described as sweat and cat urine (Dubourdieu et al., 2006; Swiegers et al., 2009; Coetzee and du Toit, 2012). Fermentative esters are also considered to be important in the flavour and aroma of Sauvignon Blanc, adding to the fruity, capsicum and fresh asparagus notes (Benkwitz et al., 2012). The release of these compounds and others by yeasts thus modulates the flavour and aroma of the final wine, recapitulating the importance of yeasts in the wine making process.
1.3.4 History of wine making in New Zealand

New Zealand (NZ) has a relatively young wine industry with the first grape vines planted in 1819 by Samuel Marsden in Kerikeri and the first record of wine being made from vines planted 1836 in Waitangi by James Busby (Halliday, 1991). While the first commercial winemaking attempts were located in Northland, throughout the 1800’s, further attempts to establish vineyards were made throughout the country, including those by French immigrants in 1840 in Akaroa and by German immigrants in 1843 in Nelson; however long term these were largely unsuccessful (Halliday, 1991). Mission Estate, the oldest existing vineyard and winery in NZ, was established by French missionaries in 1851 in Hawke’s Bay. Today Hawke’s Bay is the second largest wine producing region in New Zealand, pressing nearly 39 thousand tonnes of grapes in the 2013 harvest (New Zealand Winegrowers, 2013).

The major advancement of the New Zealand wine industry came with the success of Sauvignon Blanc, particularly from Marlborough. Matua Valley in Auckland was the first to trial Sauvignon Blanc in New Zealand, planting the first vines in 1974, followed by Montana in Marlborough (Halliday, 1991). It was not until 1980 that the first commercial release of Marlborough Sauvignon Blanc was made. By 1984 Ernie Hunter, an early exporter of New Zealand wine, had pushed Hunter’s Sauvignon Blanc in to the United States of America, Britain, Ireland and Australian markets (Stewart, 2010). In 1986 he obtained a number of high-profile trophies including a gold medal for the 1985 Sauvignon Blanc, boosting publicity for New Zealand wine in Britain (Stewart, 2010). This, in combination with a multitude of accolades for New Zealand wine, including winning the Marquis de Goulaine Trophy for champion Sauvignon Blanc at the British International Wine and Spirit competition from 1990-1993, solidified Sauvignon Blanc as New Zealand’s flagship wine (Stewart, 2010).

In its early days the New Zealand wine industry was purely domestic and demand was low with per capita consumption of wine in 1898 being just 720 mL (Stewart, 2010). Today wine is New Zealand’s largest
Chapter 1: General Introduction

horticultural export by value, exporting to over 80 countries at a value of $1.2 billion per annum (New Zealand Winegrowers, 2013). Vineyards are spread throughout New Zealand from Northland to Central Otago. Marlborough is by far the largest wine producing region pressing over 250 thousand tonnes of grapes during the 2013 harvest (New Zealand Winegrowers, 2013). By variety nothing compares to the magnitude of Sauvignon Blanc with nearly 230 thousand tonnes produced in 2013 and accounting for 84.5% of exports, followed by Pinot Noir with about 32 thousand tonnes (New Zealand Winegrowers, 2013).

1.4 OBJECTIVES

The main objective of this thesis is to provide a deeper understanding of the ecology and population patterns of S. cerevisiae, and investigate how these patterns, in combination with yeast community dissimilarity, contribute to wine styles in NZ.

More specifically, the objectives are to:

1. Quantify the degree to which New Zealand’s S. cerevisiae metapopulation is genetically structured by both environment and geography and go on to estimate the extent to which sub-populations are connected by gene-flow (Chapter 3)

2. Investigate the role of the ‘woodland’ niche in the life cycle of S. cerevisiae (Chapter 4)

3. Quantify the contribution of regional populations of S. cerevisiae to terroir or regional distinctiveness of wine (Chapter 5)

4. Investigate whether the fungal community in grape juice and the species of Saccharomyces driving the ferment affect thiol concentrations in wine (Chapter 6)
Chapter 1: General Introduction

1.5 THESIS OVERVIEW

This thesis comprises seven chapters in total. In this first chapter I have provided general background pertaining to the research performed in this thesis and outlined the main objectives. The title and aims of subsequent chapters are detailed below.

1.5.1 Chapter 2: “Methodology”

The methods pertaining to experimental design, data generation and analysis were chosen to allow a rigorous investigation of each of the objectives outlined above. Here I detail available approaches for each objective and justify why the final methods used were chosen. As part of the first objective I also outline a statistical method for analysing the driving forces of population structure that I was part of developing in collaboration with Velimir Gayevskiy, Dr Steffen Klaere and Dr Matthew R. Goddard during my PhD. This method has been published in PLoS One (Gayevskiy et al., 2014).

1.5.2 Chapter 3: “Quantifying separation and similarity in a *Saccharomyces cerevisiae* metapopulation”

This chapter addresses the first objective and quantifies population patterns in the NZ *S. cerevisiae* metapopulation, sampling both vineyard and native environments from six major wine growing regions.

The major aims of this chapter are to:

1. Test the hypothesis that populations of *S. cerevisiae* residing in native and vineyard environments are ubiquitous within geographic regions
2. Test the hypothesis that populations of *S. cerevisiae* isolated from environmental niches and spontaneous ferments from the same geographic region are ubiquitous
3. Test the hypothesis that the New Zealand population of *S. cerevisiae* is structured by geography
4. Quantify the degree to which these geographic subpopulations are subdivided and subsequently connected by gene-flow or migration

**1.5.3 Chapter 4: “Sporulation in soil as an over-winter survival strategy in *Saccharomyces cerevisiae*”**

This chapter addresses the second objective and investigates the hypothesis that *S. cerevisiae* resides in soil in a sporulated state. Since fruit is ephemeral, *S. cerevisiae* requires a life history strategy to survive over winter and its potential to do so in a sporulated state in the ‘woodland’ niche is addressed for the first time.

The main aims of this chapter are to:

1. Articulate the hypothesis that *S. cerevisiae* resides outside of the fruit niche in a sporulated state as a self-preservation strategy when fruit is not available
2. Experimentally test a crucial aspect of this hypothesis by investigating if soil can initiate a sporulation response in *S. cerevisiae*

**1.5.4 Chapter 5: “Regional microbial signatures translate to differential wine phenotypes providing evidence for a microbial aspect to *terroir*”**

This chapter addresses objective three and experimentally tests and quantifies the contribution of *S. cerevisiae* to regionally distinctive wine flavours and aromas with implications for microbial species and agricultural products generally.

The main aims of this chapter are to:

1. Test the hypothesis that *S. cerevisiae* isolated from different geographic regions produce distinct chemical profiles in the wine
Chapter 1: General Introduction

2. Quantify the amount of variation in the chemical profiles that can be attributed to the region of isolation of the *S. cerevisiae* genotypes

3. Examine if regional genotypes interact during fermentation to produce distinct chemical profiles in the wine by testing for differences between regional co-ferments and blends

4. Test if there is regional differentiation in these regional co-ferments

1.5.5 Chapter 6: “Fungal diversity during fermentation is indicative of thiol concentrations in wine with indications that microbes may modulate these throughout fruit development”

Chapter 6 addresses objective four and investigates what effects fungal species have on the flavour and aroma of wine by correlating the fungal community present in grape juice pre-fermentation and the species of *Saccharomyces* that drives the ferment with three important varietal thiols in Sauvignon Blanc (3MH, 3MHA and 4MMP).

The main aims of this chapter are to:

1. Investigate whether the fungal community in the grape juice correlates with the final concentration of 3MH, 3MHA and 4MMP

2. Examine whether different species of *Saccharomyces* impart different concentrations of volatile thiols during fermentation

1.5.6 Chapter 7: “General discussion and future directions”

Here I summarise the findings of this thesis and relate them back to the aims and objectives outlined above. I discuss how these findings have contributed to the field of microbial ecology and winemaking and propose future directions for further research.
Chapter 2

METHODOLOGY

As with any scientific investigation, the methods used must ultimately provide robust and appropriate information to address the objectives. Here I discuss the different methods used to address each objective and clarify why each method was chosen. Specific details of the methods used can be found in each chapter.

2.1 OBJECTIVE 1: Quantify the degree to which New Zealand’s S. cerevisiae metapopulation is genetically structured by both environment and geography and go on to estimate the extent to which sub-populations are connected by gene-flow (Chapter 3)

Since microbes cannot be observed directly in nature, genetic methods are relied upon for accurate identification and subsequent analyses to infer population patterns and processes. To investigate the population process occurring in New Zealand’s S. cerevisiae I use a microsatellite genotyping protocol developed at the University of Auckland (Richards et al., 2009) followed by a suite of objective statistical analyses described below.
Chapter 2: Methodology

2.1.1 Sampling

Samples were collected from *Vitis vinifera* var. Sauvignon Blanc vineyards and surrounding native bush in six major wine growing regions in NZ (refer to Appendix I for locations). In the vineyard a variety of niches were targeted including soil beneath the vines, bark and fruit. Samples of each were collected from three dispersed locations within each vineyard and homogenised to create the final sample from each niche for that site. Additionally, juice pressed from fruit from the same vineyard blocks was collected from winery settling tanks and sent to the University of Auckland. Samples of the juice were taken before letting it ferment spontaneously, and further samples were taken after 21 days. The native samples were taken from nearby unmanaged reserves, targeting native species fruiting around the same time as the grapevines (refer to Table AII.2 in Appendix II for tree species sampled). Six native trees in each region were targeted taking samples of soil and fruit from each.

Since *S. cerevisiae* is rare in niches other than actively fermenting fruits, a means to extract it from the environmental samples (soil, bark, fruit, juice from the vineyards and soil and fruit from native samples) was required (Mortimer and Polsinelli, 1999; Pretorius, 2000; Xufre et al., 2006; Goddard, 2008; Taylor et al., 2014). Previous studies have had success with an enrichment media containing high sugar and moderate ethanol concentrations (Serjeant et al., 2008; Goddard et al., 2010) and this was also used here. This initiates the clonal expansion of any *S. cerevisiae* cells in the sample, increasing their population size and allowing individuals to be isolated. This is analogous to the spontaneous ferment samples which naturally enrich for *S. cerevisiae*.

2.1.2 Genotyping

In this genomic era, it is becoming increasingly common to differentiate individuals using whole genome sequencing or genotyping-by-sequencing approaches (Narum et al., 2013). While this is becoming
increasingly more cost effective (Wetterstrand, 2014), traditional methods of genotyping remain faster and cheaper while still providing sufficient information to address the questions posed. This is particularly important for studies such as this where large numbers of individuals need to be genotyped to provide the statistical power needed to make informed inferences of the population processes occurring.

Microsatellites are short tandem repeats of nuclear DNA consisting of repeat units of 1-6 base pairs in length (Tautz and Schlötterer, 1994; Schlotterer, 1998). They are highly polymorphic markers making them useful for determining paternity and kinship and are a valuable tool for investigating population structure (Schlotterer, 1998). Richards et al. (2009) developed a genotyping protocol that allows for the discrimination of different strains of *S. cerevisiae* based on a multiplex reaction targeting 10 microsatellite loci. Genetic distance data from the microsatellite profiles and from base pair differences from whole genome sequencing are significantly correlated, suggesting that the data obtained from the microsatellite profiling system is a useful shorthand of the absolute genetic distance from whole genome sequencing (Richards et al., 2009). As expected given the high diversity of the strains tested and the more phylogenetically informative data obtained from whole genome sequencing, when networks were constructed from the distance matrices it was found that the microsatellite data was useful at discriminating close genetic relationships and was able to cluster genetically similar strains but was not as effective at discriminating more distant evolutionary relationships (Richards et al., 2009). Here I am interested in the clustering of genetically similar groups to infer population structure and migration within the NZ metapopulation, not estimates of historical evolutionary relationships between strains. This, in combination with the reduced cost of obtaining microsatellite data in comparison to genome sequencing, denotes that genotyping via microsatellites is sufficient for my purposes.

Errors are inherent in any genotyping technology and can be introduced at multiple steps within the process. A genotyping error occurs when the genotype determined does not correspond to the true genotype of the individual (Bonin et al., 2004). For capillary electrophoresis systems such as the one used here, the main sources of error include allelic dropout where there is preferential amplification of smaller
alleles, null alleles where an allele fails to amplify due to a mutation in the primer binding site, plus-A artefacts where an additional nucleotide is added in error during replication, PCR slippage and misprints where PCR errors alter the length of the fragment, and human error in identifying an allele (Bonin et al., 2004; Pompanon et al., 2005; Kelly et al., 2011). By being aware of these potential errors, steps can be taken to minimise them.

Allelic drop out and null alleles can bias estimates of population genetic parameters as a portion of the genetic diversity in a population remains undetected (Chapuis and Estoup, 2007). There are many methods available for detecting the presence of these alleles in a dataset and these do so by testing for deficiencies in heterozygotes compared to the expected frequencies under Hardy-Weinberg Equilibrium (Dąbrowski et al., 2014). Essentially a deficiency of heterozygotes is considered indicative of allelic drop out or a null allele as individuals that are truly heterozygous would be analysed as homozygous if the second allele is not detected. While this seems like a reasonable assumption, a reduction in heterozygosity can result from other factors such as inbreeding, assortative mating and population structure (Nielsen and Slatkin, 2013), which results in false positives when attempting to identify null alleles (Dąbrowski et al., 2014). To further complicate the situation, microbes are typically inbred and do not conform to Hardy-Weinberg Equilibrium. Therefore these methods are not suitable for this analysis as a large proportion of informative data would be wrongfully excluded, reducing the power of the results and thus no controls for allelic drop out and null alleles were used in this study.

To investigate the sources of error in the microsatellite genotyping protocol described in Richards et al. (2009), a simple test was devised. Three strains of \textit{S. cerevisiae} were grown up independently three times. DNA from each of these colonies were extracted and each extract was amplified three separate times using the multiplex PCR protocol developed by Richards et al. (2009). Each PCR product was genotyped three separate times resulting in 81 genotypes, 27 of each strain (Figure 2.1). The peak sizes were called to the
Chapter 2: Methodology

nearest base pair and no binning was used. By comparing these genotypes I could identify what steps of the process error is occurring and what actions I could take to minimise them. The genotypes across different DNA extracts, PCR runs and genotyping runs were largely consistent; however some alleles varied by one base pair for some loci. This was not consistent with a particular DNA extraction or PCR run and has likely occurred during the calculation of the peak size. This is easily overcome by defining bins that alleles are forced in to. While this may mask any real single base pair differences in allele size, this effect is too common to ignore and by compressing the alleles into bins I took a conservative approach by not creating new alleles and artificially enhancing the variation observed in the samples. These results provided me with the confidence that within a plate submitted for genotyping, the samples were comparable.

Multiple plates of my unknown samples were run during the analysis. Each plate contained three positive controls of DNA from the same S. cerevisiae strain and a negative control to detect contamination during the procedure. An automated allele calling program called GeneMapper was used to make the initial allele calls which were then confirmed by eye, saving a large amount of time and reducing the potential error that can be introduced by manual scoring (Pompanon et al., 2005). Upon examination of the positive controls combined across plates, I was able to identify problematic loci that were subsequently removed from the analysis. These loci included YBR240C that reported inconsistent allele sizes between plates, even with the binning protocol, and YOR267C which frequently failed to amplify any alleles. Finally, all genotypes obtained were scrutinised for near identical individuals and where two individuals were identical with the exception of one locus, the raw electropherograms were examined to ensure peaks were not missed or misinterpreted.
Figure 2.1: Samples used to test the sources of error obtained during the genotyping of *S. cerevisiae*. This process was repeated for three separate strains of *S. cerevisiae* but only one is represented for simplicity. (a) DNA from three individual colonies was extracted. (b) From each DNA extract, three PCR reactions were performed. (c) Each PCR product was then genotyped three independent times.
2.1.3 Analysis

As explained in Section 1.1.4, clonal reproduction has theoretically been shown to increase heterozygosity at a single locus but decrease overall genotypic diversity in a population (Balloux et al., 2003). This increase in heterozygosity can have implications for inferring patterns of population structure and connectivity. As described in Section 2.1.1, an enrichment process was used to isolate *S. cerevisiae*. *S. cerevisiae* preferentially replicates via budding (clonal or asexual reproduction) under these enrichment conditions so it is likely that identical individuals arose from the same original mother cell. In terms of the original sample, this is analogous to sampling the same individual multiple times and including that replicated sample in the dataset. It was thus assumed that duplicate genotypes of *S. cerevisiae* within a sample were a result of clonal expansion and were therefore collapsed to a single isolate for further population genetic analyses. I believe this is a conservative approach as including all individuals of the same genotype from a sample will alter the allele frequencies in the population, increasing the similarity within a sample and decreasing the variability between samples, leading to an inflated signal for population structure. If I can still detect patterns of population subdivision upon removing these samples, I can have higher confidence that I have observed a real effect rather than that of a sampling bias.

There are multiple methods and programs that can be utilised for analysing population structure and migration or gene-flow ranging from classic summary statistics such as $F_{ST}$ to more complex Bayesian estimation approaches. Each method makes a series of assumptions and it is important to understand what these mean and how they may affect the results. Overall I utilise multiple methods of analyses to infer patterns of population structure and migration and gain confidence from my results as they all converge on the same answer. Here I discuss the methods used to examine population differentiation and connectivity in *S. cerevisiae* in New Zealand.
Chapter 2: Methodology

i. Summary statistics

Wright’s F-statistics are the most widely used descriptive statistics in population and evolutionary genetics and are used to describe the genetic structure of populations, providing insights into the evolutionary processes driving these patterns (Wright, 1965; Holsinger and Weir, 2009). $F_{ST}$ calculates the mean reduction in heterozygosity of a subpopulation relative to the total population (reporting a value between 0 and 1) as an indication of the extent of genetic differentiation between populations (Holsinger and Weir, 2009). This reduction in heterozygosity is indicative of population structure in sexual, outcrossed populations due to the Wahlund effect described in Section 1.1.3 (Wahlund, 1928; Hartl and Clark, 1997).

As well as being a measure for genetic differentiation, Wright’s $F_{ST}$ can be used to estimate migration rates between predefined populations under the assumptions of the island model of population structure (Wright, 1931) with the relationship:

$$F_{ST} = \frac{1}{4Nm + 1}$$

where $N$ is the effective population size and $m$ is the proportion of the population that migrates per generation (Whitlock and McCauley, 1999; Holsinger and Weir, 2009). Therefore $Nm$ together denotes the number of migrants per generation. While this provides a quick first estimate of migration between populations, the underlying mathematical assumptions used to infer these values are often not biologically relevant and the estimates may be incorrect by an order of magnitude; thus the use of alternative methods are also recommended to verify the results (Whitlock and McCauley, 1999; Holsinger and Weir, 2009).

F-statistics all require prior knowledge of the subpopulations and where the samples were taken from; however, population substructure is not always evident and can lead to a lower than expected value for the observed heterozygosity (Wahlund, 1928). Since microbes cannot be directly observed, the scale at which
Chapter 2: Methodology

their subpopulations are defined is unknown. Therefore these statistics are not ideal for microbial populations; however due to their universal use and understanding they are still reported in this thesis and are backed up with more computationally expensive and appropriate methods.

\[\textit{Bayesian approaches}\]

Bayesian approaches to estimating both population structure and migration are now widely available. Bayesian methods have the advantage of allowing prior knowledge about a dataset to be built into the model which is then adjusted with the additional knowledge provided by the data itself (O’Hara et al., 2008). This results in a posterior distribution of the parameters of interest which describes the probability that the parameters take a certain value given the data while also accounting for prior information (Holsinger and Weir, 2009). In contrast to this approach, Maximum Likelihood approaches attempt to find estimates of the parameters that best predict the data; however these methods can be biased by local optima resulting in non-optimal estimates of the parameters of interest (Templeton, 2006; Holsinger and Weir, 2009).

Bayesian clustering programs such as STRUCTURE (Pritchard et al., 2000) and BAPS (Corander et al., 2003) use advanced algorithms that analyse the allele frequencies of multiple loci and cluster the individuals into groups that are in HWE. It is thus assumed that these clusters are randomly mating sub-populations within the larger sampled population. These methods are problematic when investigating the population patterns of microbes. Microbes have limited dispersal capabilities and thus create sub-populations much more readily than macro-organisms. This, combined with their propensity to inbreed and even self, inevitably creates deviations from HWE (Nielsen and Slatkin, 2013). When applying these analyses to highly inbred and even partially selfing populations it is not viable to assume that deviations from random mating other than that driven by population structure are occurring and thus it can lead to spurious results (Falush et al., 2003). These methods are therefore not suitable for microbial populations.
A modification to the Bayesian approach used in STRUCTURE has been developed by Gao et al. (2007) that accounts for clonal reproduction and inbreeding. The algorithm implemented in the program InStruct, assumes a model of K populations that are each characterised by a set of allele frequencies for each locus and an inbreeding rate. The number of generations of selfing or inbreeding for each individual is modelled back to the most recent outcrossing event and they are probabilistically assigned to one or more of these K populations. By simultaneously estimating inbreeding rates and population structure InStruct is not reliant on the HWE and simulations show it is able to successfully classify individuals into appropriate subpopulations (Gao et al., 2007). This is much more appropriate for microbial populations and this method was utilised in the analyses.

Estimates of migration rates based on coalescent theory outperform those based on $F_{ST}$. Of those based on coalescent theory, Bayesian inference typically outperforms Maximum Likelihood inference (Beerli, 1999; Beerli, 2006). To infer patterns of migration I chose to use a Bayesian coalescence approach implemented in the program MIGRATE that is able to calculate bidirectional migration rates between populations and thus removes the assumption of symmetry in migration rates imposed by some other methods (Beerli, 1999; Beerli, 2009). This method estimates mutation-scaled population sizes as well as mutation-scaled migration rates. From these, estimates of the number of migrants per generation ($Nm$) can be calculated using the equation below:

$$4Nm_{j\rightarrow i} = \theta_i M_{j\rightarrow i}$$

where $\theta$ is the mutation-scaled population size, $M$ is the mutation-scaled immigration rate and $i$ and $j$ denote the populations of interest. As with other inferences of migration, this approach assumes constant population sizes, random mating, a constant mutation rate, and that populations are connected only through migration, not population divergence (Beerli and Felsenstein, 2001; Beerli, 2006; Beerli, 2009; Beerli and Palczewski, 2010). These assumptions can be relaxed but for interpreting the data it is important to understand what violations of these assumptions mean for the results. For example, simulations have
shown estimates of population size by MIGRATE are affected by different population histories (Beerli, 2009). This is particularly evident for recent changes in population size, such as those experienced during a population bottleneck and subsequent recovery, where MIGRATE underestimates the true population size due to the reduced genetic diversity (Beerli, 2009). It would therefore be assumed that reduced genetic diversity due to high levels of inbreeding, such as those observed in microbial species, would also result in an underestimation of the true population size. When using the equation above, underestimating the population size will also underestimate of the number of migrants per generation ($Nm$).

While MIGRATE can simultaneously estimate the mutation-scaled population sizes and directional mutation-scaled migration rates, due to the large number of populations and potential migration routes in this data, the program struggled to report informative estimates of all of the parameters. I therefore calculated the mutation-scaled population sizes using the number of sampled alleles (Haasl and Payseur, 2010) and fixed these values in the model. This reduced the number of parameters the program was required to estimate, providing more power for the estimation of the directional migration rates and thus resulted in more informative and consistent estimates. The migration routes between Nelson and Central Otago were also omitted after successive MIGRATE analyses reported negligible migration between these regions with confidence intervals encompassing zero, again reducing the number of parameters estimated by the program.

### iii. ObStruct

While InStruct robustly infers genetic subgroups in the data, the interpretation of the results is subjective and there is no objective test for the potential drivers of the inferred patterns. In collaboration with Velimir Gayevskiy (PhD student), Dr Steffen Klaere (statistician) and Dr Matthew R. Goddard we developed a program that can do just that (Gayevskiy et al., 2014). Now published in PLoS One, ObStruct directly takes the resulting ancestry profiles from the outputs of STRUCTURE (Pritchard et al., 2000), InStruct (Gao et al.,...
Chapter 2: Methodology

2007) and BAPS (Corander et al., 2004; Corander et al., 2008) and determines the extent to which it reflects the predetermined populations or groupings that the user defines (Gayevskiy et al., 2014). These predetermined populations can be any factor of interest to the user, often spatial factors such as geographic location or ecological factors such as the niche of isolation. The method calculates an $R^2$ statistic and tests the null hypothesis that the inferred ancestry profiles do not correlate with the predefined populations using a permutation approach. The $R^2$ value is calculated by evaluating the variance within and between the predefined populations and is essentially a measure of how much of the total variation in the data is explained by the predefined populations. The proportions in the ancestry profile undergo a logit transformation to account for issues of heterogeneity in variance typical of proportion data.

A permutation approach is used to evaluate the statistical significance of the $R^2$ value by randomly shuffling the ancestry profiles and recalculating the $R^2$ value 10,000 times. The $P$-value is calculated as the proportion of $R^2$ values from the 10,000 permuted datasets that exceed the original $R^2$ value calculated from the predefined populations.

With the total level of structure in the data now quantified, ObStruct can investigate the relative contribution of each predefined population and each inferred population to this structure. Each predefined or inferred population is individually excluded and the $R^2$ value is recalculated. An increase in the $R^2$ value from the original value indicates that the population excluded was more homogeneous than the average and thus did not contribute to the overall structure in the dataset, while a decrease in the $R^2$ value indicates that the population excluded was contributing to the signal for structure in the total dataset. This allows the user to identify which populations are driving the observed structure in the dataset and which are more homogeneous.
Differences between different predefined populations are also calculated using a pairwise approach. The $R^2$ statistic calculation is applied to all pairwise combinations of predefined populations, with each calculation statistically evaluated using the same permutation approach used for the full dataset. A Bonferroni correction is used to correct for multiple sampling. This allows the user to investigate differences between predefined populations.

Finally, ObStruct outputs a R-script, executable in the statistical software R (http://www.r-project.org), that uses the R specific packages of “candisc” and “heplots” (Friendly, 2007) to perform canonical discriminant analysis (CDA; e.g. Gittins, 1985) to visualise the data. Using the logit transformed ancestry profiles, the CDA fits a linear model between the predefined populations (factor) and inferred populations (response) and assesses the correlation between the variables. The divergences in the data are then visualised using a suggested set of transformed, orthogonal variables. This differs from the popular visualisation tool known as Principle Component Analysis (PCA) by accounting for variations in the explanatory variable in the calculation and providing a more informative visualisation of the variation of interest. Three plots are constructed using the script. The first figure visualises the two axes explaining the highest proportion of the variation in the data and places all of the individuals coloured by their predefined population within it. Two ellipsoids centred at the hypothetical average of all individuals are shown, one encapsulating approximately 50% of all individuals sampled and one encapsulating approximately 95%. The second figure plots ellipsoids for each predefined population. These are centred at the population mean and contain approximately 66% of the individuals in that population. This visualises the position of each of the populations given the transformed variables and larger ellipsoids represent populations with larger variances. The third plot is known as a hypothesis error (HE) plot. Using the same axes, just the centres of each predefined population are shown, and the positions of the inferred populations are represented as arrows. The ellipsoid labelled group indicates the range of individuals in the analysis, while the red ellipsoid represents the error. If the predefined populations and inferred populations do not resemble each other, the red error ellipsoid will be large and could potentially exceed the group ellipsoid.
Chapter 2: Methodology

This method was validated using both simulated and real datasets with varying degrees of population structure and admixture which was reflected in the magnitude of the $R^2$ value. It allows the user to objectively test for drivers of structure in their datasets and provides an alternative approach to the subjective interpretation of plots of ancestry profiles. This analysis is particularly useful for this objective as it provides a means to quantify the effects of both geography and the niche of isolation on the genetic structure in the dataset, revealing insights into the forces potentially driving population structure in the NZ S. cerevisiae population.

2.2 OBJECTIVE 2: Investigate the role of the ‘woodland’ niche in the life cycle of S. cerevisiae (Chapter 4)

No other studies that I am aware of have attempted to investigate sporulation in S. cerevisiae in the natural environment. Here I hypothesise that S. cerevisiae resides in the ‘woodland’ niche in a sporulated state as a survival strategy when fruit is unavailable (see Chapter 4 for a more detailed explanation of this hypothesis). I test one aspect of this hypothesis and investigate whether soil nutrients can induce a sporulation response in S. cerevisiae. Previous studies suggest strains isolated from different ecological niches vary in their ability and efficiency to sporulate (Gerke et al., 2006). Therefore six genetically diverse strains originally isolated from both soil and spontaneous wine ferments were selected for testing. Through a series of observations detailed in Chapter 4, soil was considered to be a candidate niche for sporulation to occur in nature. Since it has been observed that S. cerevisiae behaves differently in solid and liquid environments when undergoing meiosis (Vopálenská et al., 2005), experiments testing this hypothesis also needed to be performed on solid media. Soil for experiments was collected from Mates Vineyard, a Chardonnay vineyard at Kumeu River in West Auckland, NZ. I dried the soil to standardise the mass of soil that would be used in each trial. I monitored the change in weight during drying so I could estimate the amount of moisture that was removed.
Chapter 2: Methodology

2.2.1 Attempts at observing sporulation in soil itself

Ideally trials would be conducted in soil itself as this most accurately reflects natural environmental conditions; however that presented many challenges. I would not only need to be able to recover the yeast cells from the soil but also be able to observe any signs of sporulation. Known quantities of vegetative and sporulated cells were inoculated into soil, restoring the soil to its original moisture content (as calculated from the weight lost during drying). Attempts to wash cells from the soil and observe them under a microscope proved unsuccessful due to the small nature of the soil particles and the inability to separate these particles and the cells.

A series of subsequent trials trying to wash cells from the soil with water and plate on to YPD agar (1 % yeast extract, 2 % peptone, 2 % glucose) were performed. Known concentrations of both vegetative and sporulated cells were inoculated into soil and the sample was subsequently washed with sterile water. Firstly an aliquot of the wash was plated onto YPD in serial dilutions. Since I knew the concentration of both vegetative and sporulated cells I had inoculated into the samples, I could estimate number of colonies I would expect to see if I had a 100 % recovery rate. A second portion of the wash underwent a random spore isolation procedure using zymolyase to kill off vegetative cells and separate spores. This was also plated on solid YPD in serial dilutions. The idea behind this protocol was that I would be able to determine the number of sporulated cells versus vegetative cells by comparing the number of colonies on each plate. The untreated portion of the wash plated first would contain both vegetative cells (unsporulated cells) and tetrads (sporulated cells), while the treated wash would only contain spores that were no longer bound as a tetrad. By dividing the number of colonies on the treated plate by four (there are four spores in a tetrad), I would theoretically be able to calculate the number of tetrads, and compare it to the total number of cells in the sample, providing a proportion of sporulated versus unsporulated cells. Unfortunately the recovery rate of all cells and spores was highly variable, ranging from between 0 – 86 % of what I was expecting to observe and this protocol was deemed unviable.
2.2.2 Soil ‘tea’

Upon deciding that using soil itself was not a viable option, I attempted to make a solid soil agar proxy. A more detailed description can be found in Section 4.2.2 but briefly, a soil ‘tea’ was firstly made by soaking soil from the vineyard in water overnight. The large particles of soil were strained off and the solution was sterilised with DMDC. The soil tea was solidified with agar to create solid soil agar plates. *S. cerevisiae* cell suspensions were plated on the soil agar and signs of sporulation were observed after two days. It was thus considered a success and the soil agar was used for experiments.

2.2.3 Statistical analysis

The statistical analysis of this data was relatively straight forward using a repeated measures linear mixed effects model, allowing the consideration of the random effects resulting from the time point of sampling, the strain and the replicate. Since the number of sporulated cells per sample was reported as a proportion of the cells counted, all data underwent an arcsine transformation prior to analyses to account for heterogeneity of variance (Sokal and Rohlf, 1995).

2.3 OBJECTIVE 3: Quantify the contribution of regional populations of *S. cerevisiae* to terroir or regional distinctiveness of wine (Chapter 5)

This objective used an experimental approach to quantify the contribution of *S. cerevisiae* to the terroir or regionality of a wine. I utilised *S. cerevisiae* genotypes isolated from different geographic regions in the first objective to perform micro-ferments for which I quantified a suite of 39 yeast-derived chemical compounds and properties. Ferments were performed with both single genotypes and regional co-inoculations of genotypes. Multivariate analyses were performed on the chemical profiles to quantify the amount of variation that could be attributed to the region the genotypes were isolated from. This was combined with univariate analyses to infer the individual chemicals responsible for the observed regional signal. Additional
comparisons were made between the single genotype ferments, the regional co-ferments and regional blends (constructed from blends of wine from the single genotype ferments) to investigate the interaction between strains during fermentation.

2.3.1 Micro-ferments

By using homogenised Sauvignon Blanc juice, I was able to remove external influences on wine flavour and aroma and solely focus on the yeasts contribution. Since Sauvignon Blanc in NZ is typically fermented at cooler temperatures, micro-ferments were performed at 15 °C with 150 rpm shaking to simulate the mixing that would occur in larger tanks. Due to constraints on the number of flasks that could fit in the incubator at any one time, three batches were run. Each batch contained one replicate of each treatment which allowed me to control for any variations in the results due to batch by including it as a factor in the statistical analyses.

2.3.2 Chemical profiles

The parameters and compounds targeted in the chemical analyses of the resulting wines included wine quality indicators (ethanol, pH, residual sugar, volatile acid and total acid) and a suite of yeast-derived chemical compounds including the volatile thiols 3MH and 3MHA known to be important to Sauvignon Blanc aroma and flavour (Swiegers et al., 2009; Coetzee and du Toit, 2012). I also attempted to quantify the volatile thiol 4MMP but its concentration in all samples was too low for the method to reliably quantify and it was removed from the analysis. As courtesy of Pernod Ricard, their wine scan facilities were used to quantify the wine quality indicators while the methods for the quantification of the remaining compounds had previously been developed and implemented at the University of Auckland, providing me with the tools I required.
Chapter 2: Methodology

2.3.3 Statistical analysis

A permutational multivariate analysis of variance PERMANOVA was employed to test and quantify the effect of the region the genotypes were isolated from on the overall chemical profiles. PERMANOVA is a nonparametric multivariate analysis of variance (MANOVA) that calculates a test statistic analogous to Fisher’s $F$-ratio directly from a distance or dissimilarity matrix (Anderson, 2001). Unlike other nonparametric options, PERMANOVA can incorporate both single and multifactorial designs and calculates $P$-values using a permutation approach (Anderson, 2001). It also allows constraints on the permutations performed for hypothesis testing. Since we know that different strains of *S. cerevisiae* produce different chemical profiles in the wine (Howell et al., 2004; Anfang et al., 2009), it was important that the replicates from each strain were shuffled together and not separated during the permutations to avoid biasing the calculation of the $P$-values. Nonparametric analyses are typically more conservative and I can be assured that my analyses were not biased by any violations of the assumptions enforced by parametric approaches, providing me with robust results.

Constrained Correspondence Analysis (CCA) was chosen to visualise the variation in the chemical profiles as it allows the model used in the PERMANOVA analyses to be built into the visualisation (Ter Braak, 1986). This is analogous to a Principle Component Analysis (PCA) in that transformations of the data are performed to provide components that allow the data to be visualised in 2-D plots; however, the CCA additionally partitions these components into a part that is explained by the specified linear model and a part that is residual to that model. The components specified thus provide a visualisation that has rotated the data to the best orientation to observe the variation explained by the factors of interest that were specified in the model.

The CCA additionally provides vectors indicating the direction and magnitude of the influence that each chemical property has on the distribution of data points in the plot. While vectors of a larger size have a
larger influence on the positioning of the data points within the plot, they do not necessarily have a great effect on the factors of interest (i.e. they are important to the variation in the data but not necessarily the variation between data points explained by the factors being investigated). Therefore, to determine which chemicals were important to the regional chemical signal detected in the data, individual ANOVA analyses for the factor “region” were performed on each chemical independently and the $P$-values were adjusted for multiple comparisons. From these results, the chemicals that were significantly influencing the chemical profiles of different regions could be identified and the chemical loadings (vectors) in the CCA analysis provided information on the direction of these effects.

**2.4 OBJECTIVE 4: Investigate whether the fungal community in grape juice and the species of *Saccharomyces* driving the ferment affect thiol concentrations in wine (Chapter 6)**

This objective investigates the correlation between the final thiol concentration of a wine and the fungal communities during fermentation, ultimately investigating the role microbes may play on the final chemical properties of wine. Next generation sequencing technology was used to gain unprecedented insight into the fungal community in the starting juice and species specific primers identified the ratio of *Saccharomyces* species towards the end of fermentation. The volatile thiols 4-mercapto-4methylpentan-2-one (4MMP), 3-mercapto-hexan-1-ol (3MH) and 3-mercapto-hexan-1-ol acetate (3MHA) are known to be important in Sauvignon Blanc aroma and flavour (Swiegers et al., 2009; Coetzee and du Toit, 2012) and were targeted for this analysis.

**2.4.1 Sample collection and fermentation**

The samples used for this objective are the same spontaneous ferment samples used in Objective 1 (Chapter 3) for the population genetic analyses. Therefore the juice samples for fermentation originated
Chapter 2: Methodology

from 37 Sauvignon Blanc vineyards from six different geographic regions in NZ. Since these samples were intended to isolate populations of *S. cerevisiae* for the population genetic analyses, some factors potentially important to this analysis could not be controlled for. Firstly, volatile thiols in wine are converted from odourless precursors in the grape juice by yeast metabolism (Darriet et al., 1995; Tominaga et al., 1998a; Swiegers and Pretorius, 2005; Dubourdieu et al., 2006; Coetzee and du Toit, 2012). Many factors affect the concentration of the precursors and the final concentration of volatile thiols in wine for example soil quality, grape ripeness, harvesting techniques, skin contact and pressing pressure (des Gachons et al., 2002; Swiegers and Pretorius, 2005; Maggu et al., 2007; Allen et al., 2011; Martins et al., 2014). Therefore it would be assumed that each juice obtained for fermentation would have a different thiol potential which ideally would be controlled for in the analysis. As part of a related study, a sample of each of these juices was taken, sterilised and fermented under the same conditions with the commercial yeast strain EC1118. Using the same yeast strain for fermentation provided a control for the ability of different species and strains to liberate volatile thiols and thus provided a measure of juice thiol potential; however, due to the nature of this related study, diammonium phosphate (DAP) was added only to juices with low nitrogen levels. Since the addition of nitrogen via DAP has been shown to induce a decrease of 3MH production (Subileau et al., 2008), these ferments were not suitable to act as a control in this study and thus no control for initial juice composition could be used.

Secondly, there was no control for any fungicide treatments applied to the fruit during its development. These treatments would alter the fungal communities on the fruit and thus in the juice where the samples were taken from. However, limits are imposed on the concentration of residues allowed in the final wine and certain sprays cannot be used within certain time periods before harvest (New Zealand Winegrowers, 2013a). Thus it is assumed this effect is minimal by the time the fruit is harvested and pressed. Regardless, as this analysis is correlative and not causative, the lack of controls for these factors does not impose large restrictions on the interpretation of the data.
Chapter 2: Methodology

Some factors could be controlled for. Region has been shown to correlate with the final concentration of volatile thiols in NZ Sauvignon Blanc (Lund et al., 2009; Benkwitz et al., 2012). Since the location the juice was sampled from was known, this could be accounted for in the analyses. Additionally, all ferments were temperature controlled, removing any effects temperature may have had on the production of volatile thiols during fermentation (Masneuf-Pomarède et al., 2006).

2.4.2 Community analyses

Next generation sequencing technology allows massive parallel sequencing of entire microbial communities obtained directly from environmental samples in a fast and cost effective manner (Shokralla et al., 2012) and was used here to analyse the fungal community in the starting juice. This contrasts with traditional Sanger sequencing (Sanger et al., 1977) that is only able to sequence individuals independently and is therefore not appropriate for capturing the vast diversity of the thousands of microbes present in environmental samples (Shokralla et al., 2012). At the time of sequencing, Roche 454 technology provided the longest read length of the sequencing platforms available to us (Shokralla et al., 2012) which was important to ensure the DNA fragments sequenced were long enough to contain relevant phylogenetic information for species identification. Briefly, the Roche 454 platform determines the DNA sequence of a sample using pyrosequencing. Here the attachment of a nucleotide by DNA polymerase ultimately results in the production of light with the amount of light produced being proportional to the number of nucleotides incorporated (Margulies et al., 2005; Mardis, 2008). Firstly single template molecules are attached to beads which are amplified using emulsion PCR. The beads are loaded onto a picotiter plate containing millions of wells that each holds a single bead, providing a fixed location for the sequencing of each fragment to be monitored. Single nucleotide solutions are sequentially washed over the plate and the fluorescence at each bead location on the plate after each wash is monitored. The occurrence of the light emission signifies the incorporation of that particular nucleotide to the sequence and the intensity indicates how many nucleotides were added sequentially (Margulies et al., 2005; Mardis, 2008). The beads are therefore
sequenced in parallel with the patterns and intensities of light emitted from each bead representing the DNA sequence of the original fragments attached to each bead.

Many precautions were taking in the post-processing of the sequencing data to minimise potential errors inherent of any sequencing technology. A particular issue with pyrosequencing technologies such as 454 is its inability to distinguish the length of long stretches of the same nucleotide (homopolymer sequence regions). As the length of the homopolymer region increases, the ability of this technology to interpret the number of bases incorporated at one time, represented by the intensity of light emitted, diminishes (Hudson, 2008; Mardis, 2008). Another potential error to be aware of when sequencing amplified DNA fragments is the presence of chimeras. Chimeric sequences are artefacts of the PCR process often caused when amplification of a fragment terminates early in one cycle and is added to in subsequent cycles, resulting in unique sequences which could be identified as a novel species in downstream analyses (Edgar et al., 2011). To account for these errors, post-processing of the 454 data was performed using the software Mothur version 1.30 (Schloss et al., 2009). All primers, low quality and short reads less than 200 bp were removed as these sequences do not contain enough information for accurate operational taxonomic unit (OTU) and species classification. OTUs are clusters of read that share more than 98 % sequence identity. Multiple species of Ascomycota and Basidiomycota (Fungi) have empirically been shown to differ by less than 2 % at the 26S rDNA gene (Kurtzman and Robnett, 2003; Romanelli et al., 2010) and thus these groups are considered to approximate species. Homopolymer errors identified using the PyroNoise algorithm (Quince et al., 2009) and PCR chimeras identified using the UCHIME algorithm (Edgar et al., 2011) were also removed. Of these good quality reads, all unique reads were compared to a fungal database and all sequences not deemed to be of fungal origin were removed.
Chapter 2: Methodology

2.4.3 Statistical analyses

The aim of this analysis was to examine whether any of the OTUs from the initial juice samples, as well the species of *Saccharomyces* found at the end of ferment, correlate with any of the variation in the final thiol concentrations. The analysis of the fungal community in the starting juice using next-generation sequencing technology revealed many more OTUs than samples taken (Appendix IV Dataset AIV.5). Due to constraints on parameter estimation and the number of degrees of freedom, a model cannot be built using more variables than samples taken. Therefore a data reduction technique was required. Partial Least Squares Regression analysis (PLSR) is a valuable statistical tool typically used to model a linear relationship between a set of variables and a set of responses (Wold, 1966; Hubert and Branden, 2003). It compresses the numerous variables into components or latent variables while accounting for potential collinearity or multicollinearity between variables and thus is useful as a method for data reduction (Hubert and Branden, 2003; Mevik and Wehrens, 2007). PLSR is a similar approach to a Principle Component Analysis (PCA) in that it captures the variation in the X-variable (in this case the OTU diversity); however PLSR is considered to be a supervised approach as it also captures the information in the Y-variables (for this dataset, the thiol concentrations) and considers how the X-variables may be related to these (Mevik and Wehrens, 2007).

More generally, PLSR compresses all of the OTUs into a manageable number of components that still describe the maximum amount of variation in the OTU proportions. Therefore instead of building a linear model using all of the OTUs against the thiol concentrations, a model is built using the components from the PLSR model. The region the juice was isolated from was also included in the linear model as previous research has indicated that wines from different geographic regions have different thiol concentrations (Lund et al., 2009; Benkwitz et al., 2012). Components that are considered to significantly affect the concentration of thiols in the final wine can then be investigated further to identify which OTUs are driving these differences. This is done by examining the loadings of each OTU in relation to each component (i.e. the magnitude of effect each OTU has on each component). While components are typically made up of many OTUs with loadings of varying degrees of magnitude, those with larger loadings are considered to have a large effect on the variation of interest. Due to the vast number of OTUs obtained, a loading value of
Chapter 2: Methodology

0.2 and above was considered to represent the main OTUs driving the observed correlations while keeping the number identified to a manageable and interpretable number.
Chapter 3: QUANTIFYING SEPARATION AND SIMILARITY IN A SACCHAROMYCES CEREVISIAE METAPOPULATION

Status of chapter:

This chapter has been published in the International Society of Microbial Ecology (ISME) Journal. The co-authorship form is presented at the start of this thesis, after the acknowledgements.

Abstract

Eukaryotic microbes are key ecosystem drivers, but we have little theory and few data elucidating the processes influencing their observed population patterns. Here we provide an in-depth quantitative analysis of population separation and similarity in the yeast *Saccharomyces cerevisiae* with the aim of providing a more detailed account of the population processes occurring in microbes. Over ten thousand individual isolates were collected from native plants, vineyards, and spontaneous ferments of fruit from six major regions spanning 1 000 km across New Zealand. From these, hundreds of *S. cerevisiae* genotypes were obtained, and using a suite of analytical methods we provide comprehensive quantitative estimates for both population structure and rates of gene-flow or migration. No genetic differentiation was detected within geographic regions, even between populations inhabiting native forests and vineyards. We do, however, reveal a picture of national population structure at scales above ~100 km with distinctive populations in the more remote Nelson and Central Otago regions primarily contributing to this. In addition, differential degrees of connectivity between regional populations are observed and correlate with the movement of fruit by the New Zealand wine industry. This suggests some anthropogenic influence on these observed population patterns.
3.1 INTRODUCTION

Research to date attempting to elucidate the patterns and processes involved in shaping natural populations has largely focused on readily observed macro-organisms but comparatively little work has been conducted on microbial species (Anderson and Kohn, 1998; Martiny et al., 2006; Prosser et al., 2007). Because of their large population sizes, and ease of transfer, one might expect microbial populations to be well mixed (Finlay, 2002), but there is increasing evidence showing that many are not homogeneous but structured (Taylor et al., 2006; Whitaker and Banfield, 2006; Hanson et al., 2012). Most microbial ecology studies have focused on bacteria, but eukaryotic microbes, which undergo sex (with recombination), also play key ecosystem roles (Green et al., 2008; Van Der Heijden et al., 2008). It is not clear whether the population patterns estimated for eukaryotic ‘macrobes’ generally hold for eukaryotic microbes (Hartl and Clark, 1997; Anderson and Kohn, 1998; Halkett et al., 2005; Taylor et al., 2006; Prosser et al., 2007; Tsai et al., 2008).

A metapopulation comprises a number of spatially separated populations of the same species that interact to some extent. To date all studies examining microbial populations have simply examined whether population structure is evident or not (Aa et al., 2006; Achtman, 2008; Liti et al., 2009; Goddard et al., 2010; Anderson and Shearer, 2011; Härnström et al., 2011; Gayevskiy and Goddard, 2012; Wang et al., 2012). Merely defining microbial populations as either structured or homogeneous is highly unlikely to reflect the true biological situation. A more accurate approach is not only to assess the degree to which populations are structured, but also connected by gene-flow, and crucially go onto quantify these processes; however, to the best of our knowledge, there are no previous studies that have used a unified framework to do this. Here we analyse the natural population of *Saccharomyces cerevisiae* in New Zealand, and in doing so take the first steps towards quantifying microbial population structure and similarity.

*S. cerevisiae*, a budding yeast, has been closely associated with humans since the dawn of civilisation due to its fermentative capabilities, and has come to be of significant commercial importance in the production of
bread, wine, beer and other alcoholic beverages (McGovern et al., 1996; Pretorius, 2000; Cavalieri et al., 2003; Chambers and Pretorius, 2010). S. cerevisiae is also a classic model organism for research into cell biology, genetics, and increasingly ecology and evolution (Chambers and Pretorius, 2010; Dujon, 2010; Gray and Goddard, 2012; Hittinger, 2013; Hyma and Fay, 2013). Recent studies have revealed a large genetic diversity within S. cerevisiae, and there is good evidence for population structure at inter-continental scales (Fay and Benavides, 2005; Schuller et al., 2005; Aa et al., 2006; Lopandic et al., 2008; Liti et al., 2009; Goddard et al., 2010; Mercado et al., 2011; Di Maio et al., 2012; Wang et al., 2012). Similar inferences have been made at finer scales with reports that some genotypes were unique to different geographic locations in Austria, although many were also ubiquitous across regions (Lopandic et al., 2008). Additionally, Bayesian inference shows genetic differentiation between populations spanning hundreds of kilometres in New Zealand (NZ) (Gayevskiy and Goddard, 2012). While the scales of these studies differ, they all commonly report the presence of hybrid or mosaic strains indicative of some level of connectivity between populations via gene-flow.

Global scale analyses have suggested that ecological function may define population structure to a greater extent than geographic origin (Fay and Benavides, 2005; Legras et al., 2007). Strains associated with wine appear somewhat distinct from those isolated from distilling, bread making, fermented milk, rice wine, ale and lager, with geographic origin only explaining 28 % of variability (Legras et al., 2007). Furthermore, whole genome analyses of a limited number of strains suggest specific S. cerevisiae populations associated with vineyards, sake and related ferments, although some of these clusters are confounded with geographic origin (Liti et al., 2009; Schacherer et al., 2009). In contrast, a recent rigorous genome wide population study has provided evidence of gene-flow across small distances (<17 km) between distinct populations inhabiting vineyards and oak trees, showing connectivity between ecological niches at small scales (Hyma and Fay, 2013).
Despite these excellent efforts to date, most studies have drawn from widely dispersed isolates often from different ecological niches, with relatively small sample sizes from any one discrete population (Fay and Benavides, 2005; Aa et al., 2006; Legras et al., 2007; Liti et al., 2009; Schacherer et al., 2009; Wang et al., 2012), and have thus not afforded adequate power to quantify ‘ecological scale’ population processes such as gene-flow. Previous work using both microsatellite and RAD-seq analyses show a distinct S. cerevisiae population resides in NZ, suggesting this population is not subject to rampant inward international gene-flow (Goddard et al., 2010; Cromie et al., 2013). Therefore in addition to its geographic isolation, the NZ S. cerevisiae population appears relatively self-contained, and thus provides a good population to study the processes we are interested in. Here we analyse close to a thousand S. cerevisiae isolates from four niches across six regions spanning over 1 000 km. We quantify both the degree to which this population is structured, and go on to quantify the extent to which the various regional populations are connected by gene-flow in one of the most comprehensive studies of a microbial metapopulation to date. Lastly, this study sheds light on the connection between farmed (managed) and native ecosystems by examining the relationship between microbial populations residing in vineyards and native NZ forest.

### 3.2 METHODS

#### 3.2.1 Sample collection and processing

Six to seven Vitis vinifera var. Sauvignon Blanc vineyards were selected from each of Hawke’s Bay, Martinborough, Nelson, Awatere Valley, Wairau Valley, and Central Otago in NZ (Figure 3.1). Approximately 5 g of soil were aseptically taken from each of these 37 vineyards between one and four weeks before harvest in mid-March 2011. 10 L of juice deriving from the same vineyards was collected from commercial settling tanks (one vineyard provided juice samples from two pressing tiers, resulting in a total of 38 juice samples). Soil and fruit samples were taken from six native NZ plants located in non-managed native bush reserves within each region (Appendix II Table AII.2), ranging from 0.1 to 50 km from the vineyard sites, totalling 72 native samples (36 soil and 36 fruit). S. cerevisiae is rare in niches other than in actively
fermenting fruit so equivalent selective culturing methods were employed for all samples to control for the effects of high sugar and ethanol (Mortimer and Polsinelli, 1999; Pretorius, 2000; Xufre et al., 2006; Goddard, 2008; Taylor et al., 2014). An enrichment method emulating fermenting selection pressures was employed for all 147 environmental samples (Mortimer and Polsinelli, 1999; Serjeant et al., 2008). Samples were submerged in 10 mL SelMed media (1 % yeast extract, 2 % peptone, 10 % glucose and 5 % ethanol) for six days; 500 µL was then transferred to 10 mL fresh SelMed for four additional days, and then dilutions plated onto YPD (1 % yeast extract, 2 % peptone, 2 % glucose and 1.5 % agar) with 50 µg/mL chloramphenicol to retard bacterial growth. All incubation was at 28 °C. Up to 94 colonies were taken from each sample and stored in 15 % glycerol at -80 °C. A total of 7 144 individuals were isolated from environmental samples. A natural enrichment of the juice samples was performed by allowing them to ferment spontaneously at 15 °C. 100 mL was concentrated by centrifugation after 21 days, and plated onto YPD with 50 µg/mL chloramphenicol. Again 94 colonies were isolated from each ferment sample totalling 3 572 individuals. All niches were thus evenly sampled and in total 10 716 individuals were collected.

3.2.2 Molecular methods

Genomic DNA was extracted from colonies with 15 µL of 1.25 mg/mL Zymolyase solution dissolved in 1.2 M sorbitol and 0.1 M KH2PO4 at pH 7.2 and treated with ethidium monoazide (EMA) to bind unwanted DNA fragments (Rueckert and Morgan, 2007). We employed a multiplex PCR reaction to distinguish S. cerevisiae, and this also identifies S. uvarum (de Melo Pereira et al., 2010). DNA from eight S. cerevisiae colonies from each sample were initially amplified and scored at 10 unlinked loci as described by Richards et al (2009) using capillary electrophoresis on an ABI3130XL (Applied Biosystems). If all eight initial isolates were genotypically identical then no further genotyping was performed for that sample, but if more than one genotype was recovered, another eight were genotyped until either no new genotypes were seen, or all isolates from the sample had been genotyped. A number of control samples were submitted for the calculation of error rates per allele and per locus as described by Pompanon et al. (2005). To further ascertain the reliability of
microsatellite loci amplification and scoring we analysed an additional 96 well control plate replicating the same strain for DNA extraction, PCR amplification and genotyping.

### 3.2.3 Data analyses

A ±1 bp error in size calling from run to run variation and plus-A effects was observed and loci were binned accordingly using Genemapper (Version 4). F-statistics, migration estimates (Nm values) and Mantel tests were performed with GenAlEx (Genetic Analyses in Excel) version 6.5 (Peakall and Smouse, 2006; Peakall and Smouse, 2012). Estimates of population diversity were calculated by rarefaction (which controls for unequal sample sizes) using EstimateS (Colwell, 2006). The maximum likelihood outcrossing rates were estimated in Mathematica 7 following the method used by Johnson et al. (2004) which estimates the proportion of matings between spores from the same meiotic event (i.e. that are ascimates), and those from independent meiotic events (code available at http://goddardlab.auckland.ac.nz/data-and-code/). Allelic richness was estimated using rarefaction with HP-rare, again controlling for unequal sample sizes, based on the lowest number of 94 observed alleles among sampled populations (Kalinowski, 2005).

Population structure was evaluated using the Bayesian clustering method implemented in InStruct, which does not assume Hardy-Weinberg Equilibrium, accounts for inbreeding, and makes no a priori assumptions about the sampling location of the genotypes (Gao et al., 2007). This method estimates the most likely number of populations and assigns genotypes to these probabilistically. Admixture was allowed and the proportion of each genotype’s ancestry in each inferred population was estimated. Three chains of one million MCMC iterations with a burn-in of 10 000 were run for K = 1 to 25. Convergence of the MCMC chain was confirmed using the Gelman-Rubin statistic (Gelman and Rubin, 1992). Analyses of the resulting ancestry profiles evaluating and quantifying the contribution of niche and geographic region to population structure was conducted with ObStruct (Gayevskiy et al., 2014).
Directional migration rates were quantified using the Bayesian coalescent approach implemented in MIGRATE which assumes constant population sizes, random mating, a constant mutation rate, and that populations are connected only through migration, not population divergence (Beerli and Felsenstein, 2001; Beerli, 2006; Beerli, 2009; Beerli and Palczewski, 2010). Mutation-scaled population sizes ($\theta$) were calculated using the number of sampled alleles (Haasl and Payseur, 2010). We employed a Brownian motion allele mutation model with starting estimates of the mutation-scaled migration rate derived from $F_{ST}$ calculations to estimate all possible migration routes. Chains of one million steps with a burn-in of 50 000 were run with 10 replicates, sampling every 100 steps (Beerli, 2009). The analysis was run in parallel on the NeSI pan cluster at the University of Auckland.

3.3 RESULTS

3.3.1 *S. cerevisiae* presence, abundance and genetic diversity

PCR analyses revealed that 3 900 (36 %) of the 10 716 isolates were *S. cerevisiae*. Of the 3 780 isolates from spontaneous ferments, 2 210 (56 %) were *S. cerevisiae* and 1 570 (40 %) *S. uvarum*, revealing the co-existence of a sister *Saccharomyces* species in this niche. Here we do not pursue the population genetics of *S. uvarum*. *S. cerevisiae* was detected in 13 of the 37 vineyard soils, and four and one of the 36 native soil and fruit samples respectively. The breakdown of samples that yielded *S. cerevisiae* is shown in Appendix II (Tables AII.1 and AII.2).

From controls samples, two loci (YOR267C and YBR240C) amplified unreliably and were removed from all analyses. Overall the mean error rates per allele and locus were ± 4.08 % and 4.35 % respectively. In total 850 individuals were genotyped, with 681 isolates from spontaneous ferments, 130 from vineyard soil, 31 from native soil and 8 from native fruits. Identical genotypes within the same sample were collapsed to conservatively account for clonal expansion during enrichment and fermentation meaning the data set was
Chapter 3: *S. cerevisiae* Population Structure and Migration

compressed to 380 genotype profiles. Just eleven genotypes matched commercially available wine strains commonly used in NZ (Richards et al., 2009), and were removed from further analyses. This resulted in a final dataset comprising 369 microsatellite profiles (Appendix IV Dataset AIV.1). Interestingly, no genotypes matched a genetically and ecologically diverse set of international strains (Liti et al., 2009) genotyped using the same method (Richards et al., 2009; Goddard et al., 2010).

For the entire dataset, a large allelic diversity was detected at all loci. YFR028C and YML091C had the greatest diversity with 25 and 30 alleles respectively, and all other loci had between 11 and 16 alleles. Overall 295 different genotypes were recovered and only 38 of these were identified in more than one sample. On average samples that yielded *S. cerevisiae* contained 4.7 unique genotypes although most alleles were shared between populations (Appendix IV Dataset AIV.1). Rarefaction analyses (Chao, 1987; Colwell, 2006) estimates these genotypes were sampled from an underlying NZ population containing approximately 1 700 different genotypes (with 95 % confidence limits of 1 159 - 2 486).

### 3.3.2 Testing for ecological drivers of population structure

Only four different genotypes were derived from native fruits and soil, and 21 from vineyard soil (Appendix II Table AII.1 AII.2). This translates to low statistical power to test how the niche of isolation affects population structure. Despite this, observations of identical genotypes between niches within, but not between, regions are striking. For example: genotypes recovered from native soil and fruit in the Martinborough region (Waiohine Gorge) were identical to spontaneous ferment isolates recovered 20 km away in Martinborough vineyards, but were not found in other NZ regions; three of the genotypes isolated from vineyard soil in the Wairau Valley were identical to isolates from spontaneous ferments sourced in the same region, with one of these vineyard soil genotypes being identical to an isolate from the spontaneous ferment from the same vineyard. *F*<sub>ST</sub> values between environmental samples (from native soil and fruit and vineyard soil) and spontaneous ferment samples within regions are extremely low (< 0.005) and insignificant (*P* > 0.33) with the
exception of Martinborough ($F_{ST} = 0.046, \ P = 0.003$); however, this $F_{ST}$ value is classed as only representing ‘low’ differentiation (Wright, 1978). There is complete overlap of isolates deriving from all niches in a Principal Component Analysis (PCA) of genetic distances between genotypes (Appendix II Figure AII.1), and no significant population differentiation between niches within regions using InStruct and subsequent ObStruct analyses ($P > 0.119$, Appendix II Figure AII.2). Together this provides no substantial evidence of an effect of niche on population structure within regions. Some of the *S. cerevisiae* genotypes contributing to spontaneous ferments may have derived from wineries, as opposed to the ‘environment’ (Bokulich et al., 2013). Since all of these wineries reside within the geographic regions the fruit was collected from, these potentially winery derived genotypes form part of the local population we wish to study. Thus individuals from various niches within regions comprise homogenous populations, and so we combined all genotypes from different niches within regions to form regional populations for further analyses.

### 3.3.3 Testing for geographic drivers of population structure

There was significant genetic differentiation, as estimated by pairwise $F_{ST}$ values, between populations deriving from all six regions ($P < 0.01$), with the exception of those between the Wairau and Awatere Valley’s ($F_{ST} = 0.001, \ P = 0.310$). These two valleys comprise the wider Marlborough region and were thus combined to represent one population residing in Marlborough. The subsequent pairwise $F_{ST}$ values between regions are shown in Figure 3.1. A low albeit significant correlation was observed between genetic and geographic distance (Mantel Test: $R^2 = 0.181, \ P < 0.001$). Population diversity, as estimated by rarefaction analyses to control for uneven numbers of genotypes, differs by as much as three-fold between regions (Table 3.1). Hawke’s Bay and Marlborough harbour the greatest diversity, while Nelson and Central Otago the least. Allelic richness across regions is comparable with estimates within one standard deviation of each other (Table 3.1). All eight loci in all regions are significantly out of Hardy-Weinberg equilibrium ($P < 0.001$), and show strong signals for inbreeding; however, outcrossing rates are significantly above zero within each region (Table 3.1).
Table 3.1: Summary of the populations isolated from each region.

<table>
<thead>
<tr>
<th>Region</th>
<th>Number of individuals (^a)</th>
<th>Number of different genotypes</th>
<th>Estimated total number of genotypes (^b)</th>
<th>Mean allelic richness (^c)</th>
<th>Inbreeding Coefficient F(_{IS})</th>
<th>Outcrossing rate (^d)</th>
<th>Region-specific genotypes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hawke’s Bay</td>
<td>94</td>
<td>87</td>
<td>482: 272-932</td>
<td>10.72 ± 4.86</td>
<td>0.6483</td>
<td>0.12: 0.09-0.16</td>
<td>79/87: 91%</td>
</tr>
<tr>
<td>Martinborough</td>
<td>68</td>
<td>56</td>
<td>291: 150-647</td>
<td>11.02 ± 4.29</td>
<td>0.6067</td>
<td>0.11: 0.08-0.16</td>
<td>52/56: 93%</td>
</tr>
<tr>
<td>Nelson</td>
<td>66</td>
<td>48</td>
<td>131: 82-252</td>
<td>7.29 ± 1.83</td>
<td>0.2620</td>
<td>0.42: 0.32-0.54</td>
<td>44/48: 92%</td>
</tr>
<tr>
<td>Marlborough</td>
<td>91</td>
<td>76</td>
<td>415: 138-1610</td>
<td>10.47 ± 3.64</td>
<td>0.5331</td>
<td>0.18: 0.14-0.24</td>
<td>70/76: 92%</td>
</tr>
<tr>
<td>Central Otago</td>
<td>50</td>
<td>44</td>
<td>144: 84-293</td>
<td>8.29 ± 2.84</td>
<td>0.5869</td>
<td>0.14: 0.09-0.20</td>
<td>39/44: 89%</td>
</tr>
</tbody>
</table>

\(^a\) The number of individuals included in the final analysis after the collapsing of identical individuals due to clonal expansion.

\(^b\) The estimated number of genotypes in the population by rarefaction with 95% upper and lower rarefaction limit calculated using EstimateS (Colwell, 2006).

\(^c\) Calculated for each locus independently using HP-Rare and based on 94 genes. The mean ± 1 standard deviation are reported.

\(^d\) The average proportion of spores mating from independent tetrads per meiotic generation surrounded by 2 log likelihood support limits (equivalent to a 95% confidence interval), calculated following Johnson et al. (2006).
3.3.4 Quantifying geographic population structure

InStruct analyses (Gao et al., 2007) indicate the optimal number of populations given the data is 16. Examination of the ancestry profile plots (Rosenberg, 2004) resulting from this analysis are indicative of population structure by region to some degree (e.g. the blocks of green, red and yellow in Nelson, Central Otago and Martinborough respectively) and are in agreement with the magnitude of the pairwise $F_{ST}$ estimates (Figure 3.1). Subsequent ObStruct analyses revealed the inferred population structure is significantly correlated with geographic location ($R^2 = 0.16, P < 0.0001$), and this explains about one sixth of the genetic variability observed. Individuals from the Nelson and Central Otago regions contributed the greatest signal to overall population structure with significant decreases in the $R^2$ values observed when these are removed ($R^2 = \Delta-0.05$ and $\Delta-0.02$ respectively). The $R^2$ value remained constant when data from Martinborough was removed, but increased when Hawke’s Bay and Marlborough data were independently removed ($R^2 = \Delta+0.03$ for both). Increases in $R^2$ suggest individuals from these regions add noise to any signal for structure (i.e. have homogenised not localised populations). Further, canonical discriminant analysis show approximately 80 % of the variation in ancestry profiles can be represented with the first and second axes, suggesting most of the variation can be visualised in these graphical representations of the data (Appendix II Figure AII.3). Ancestry profiles from Central Otago and Nelson cluster the most discretely in these plots, recapitulating that populations from these regions provide the strongest signals for differentiation. Pairwise comparisons between regions all significantly differ ($P < 0.001$ or $P = 0.06$ between Hawke’s Bay and Marlborough), but the $R^2$ values vary from 0.02 to 0.23 (Appendix II Table AII.3).
Figure 3.1: The location of New Zealand (NZ) sampling regions and analyses of population structure and connectivity. Plots of the ancestry profiles are shown beside each region: each vertical line represents an individual with the different colours showing the proportion of ancestry of each individual to each of the 16 inferred populations. Arrows connecting different regions show directional migration rates as calculated in MIGRATE with the width of the arrows representing the number of migrants per generation as indicated in the scale. Absolute numbers can be found in Appendix II Table AII.4. The table reports pairwise $F_{ST}$ values below the diagonal and the number of migrants per generation ($Nm$) as calculated from $F_{ST}$ above the diagonal. All $F_{ST}$ values are significant ($P < 0.01$).
3.3.5 Quantifying population connectivity and migration

Pairwise estimates of migration between the regions (Nm values) using classic methods derived from FST values (Hartl and Clark, 1997) suggest Hawke’s Bay and Marlborough are the most connected, closely followed by Marlborough and Martinborough, and Hawke’s Bay and Martinborough (Figure 3.1). Nelson and Central Otago share the lowest number of migrants with an estimate of just one per generation (Figure 3.1). MIGRATE analyses showed an acceptance ratio for each parameter ranging from 0.38 to 0.65, and an effective sample size of approximately two million suggesting the chain length was sufficient. The autocorrelation between parameters and the prior was high and estimated to be around 0.96 overall, indicating a lack of information in the data. This is reflected in the high confidence intervals surrounding the estimates (Appendix II Table AII.4). However, consistent patterns between multiple runs were evident, allowing meaningful estimates of gene-flow between regions to be made. Inferred mean rates of movement between regions span an order of magnitude ranging from 6 to 63 migrants per generation (Figure 3.1 and Appendix II Table AII.4), and show differential inward and outward movement for some regions. Correlating with the classic Nm estimates, and the analyses of population structure, Nelson and Central Otago show greatest isolation with 2-fold greater rates of outward than inward migration, and show an average of just 51 inward migrants per generation, 3.2-fold less than the overall average inward migration rate of 164 migrants per generation for all other regions (Figure 3.1 and Appendix II Table AII.4). Conversely Marlborough and the Hawke’s Bay, which harbours some of the least distinctive and most diverse populations, experience some of the greatest inward migration rates at an average of 171 migrants per generation, 1.4-fold more than the average inward migration rate (Appendix II Table AII.4). In line with the low FST estimates, a high degree of individuals with shared ancestry from InStruct, and a large proportion of admixed individuals (Figure 3.1), Marlborough and the Hawke’s Bay are the most connected regions, and experience an average of 2-fold more migration between these regions than the average overall migration rate. The extent of migration between regions does not correlate with geographic location (P > 0.21) showing the difference in the extent of gene-flow is not simply a function of distance.
3.4 DISCUSSION

We have very few models attempting to generally describe the population biology of microbes. Accurate quantification of short-term population level processes is necessary to understand the likely longer-term evolutionary trajectories of populations (Smadja and Butlin, 2011; Gray and Goddard, 2012), as well as how microbes may interact with other members of the community (Ruxton et al., 2014). We have attempted to make a significant step forward: rather than simply describing this \textit{S. cerevisiae} population as either structured or not, here we paint a more biologically realistic picture by quantifying the role that geography plays in defining structure, and go on to provide quantitative estimates of gene-flow between populations residing in different regions.

\textit{S. cerevisiae} has clearly been isolated many times from managed vineyard ecosystems and ferments of fruit (Lopandic et al., 2008; Liti et al., 2009; Schacherer et al., 2009; Goddard et al., 2010; Gayevskiy and Goddard, 2012; Bokulich et al., 2014). This species is also well reported from native niches in the northern hemisphere (Sniegowski et al., 2002; Wang et al., 2012; Hyma and Fay, 2013), and in the Southern hemisphere has been isolated from exotic \textit{Quercus} species in NZ, and from \textit{Nothofagus} in Patagonia (Zhang et al., 2010; Libkind et al., 2011). Here we provide the first report of \textit{S. cerevisiae} from multiple native tree species in the South Pacific region. Overall this NZ \textit{S. cerevisiae} metapopulation displays large genetic variance, compounding evidence that NZ harbours a large and diverse population of this species (Goddard et al., 2010; Gayevskiy and Goddard, 2012; Cromie et al., 2013). Within regions, which typically encompass a radius of under 100 km, there is no compelling evidence for genetic differentiation between niches within managed ecosystems, nor more strikingly between managed and native ecosystems. The lack of genetic differentiation between managed and native ecosystems seen here does not permit us to determine if vineyards or native forests are the sources or sinks of these populations, just that they are connected. Thus, there appears to be a free flow of individuals between these various niches at sub-regional scales, supporting previous reports from NZ and the USA (Goddard et al., 2010; Hyma and Fay, 2013). The
inference of little differentiation between niches at regional scales is in contrast to previous reports showing differentiation between isolates from various ecological niches at global scales (Fay and Benavides, 2005; Legras et al., 2007). One explanation for this is the extent of sample effort within any one population. The studies, including this one, reporting a minor effect of niche examined a large number of individuals from specific more localised populations, and in some sense evaluate ‘ecological scale’ processes: it may be that some strains are less well adapted to various niches and that selection will eventually result in their removal. Studies evaluating strains from different geographic and ecological sources only include a handful of strains from any one specific population and unfortunately tend to confound geographic location with niche, but conclude niche plays a stronger role; in some sense these studies might examine populations where selection has possibly had more time to operate. Perhaps the drivers of population structure differ at different scales? Lastly, it might be that NZ has relatively recently been colonised by only one of the inferred lineages of S. cerevisiae, and this has radiated to all niches. This would also provide a signal for the lack of differentiation between niches. Estimates of the rates of global flux for S. cerevisiae would help disentangle these possibilities.

While populations appear homogeneous within regions, analyses provide compelling evidence for various degrees of genetic differentiation between populations inhabiting major NZ regions. This differentiation is not absolute and there is also a degree of connectivity between regions. This is in line with a previous smaller scale study with this species which reported both differentiation and connection between regions in the North Island of NZ (Gayevskiy and Goddard, 2012), and this is also echoed at global scales (Liti et al., 2009; Wang et al., 2012). Here all analyses, both classic and more sophisticated Bayesian approaches, converge on the same conclusion. $F_{ST}$, Bayesian, ancestry profile and migration analyses show the populations residing in Nelson and Central Otago are the most distinct and experience the least inward migration. Conversely, Marlborough and Hawke’s Bay have smaller pairwise $F_{ST}$ values and Bayesian, ancestry profile and migration analyses show these regions are the most mixed and connected. Marlborough and Hawke’s Bay experience the most inward migration at approximately three times that
Chapter 3: *S. cerevisiae* Population Structure and Migration

into Nelson and Central Otago. This is consistent with the higher genetic diversity observed in these regions and implies they accumulate genetic diversity from around the country.

*S. cerevisiae* cells and spores are sessile, but there are a variety of possible vectors that may move this unicellular eukaryote around. *S. cerevisiae* has been shown to be associated with both wasps and bees, and has long been known to be associated with fruit flies (Reuter et al., 2007; Goddard et al., 2010; Stefanini et al., 2012). Recent work provides evidence that certain volatiles released by *S. cerevisiae* attract *Drosophila*, and this enhances the likelihood of movement, and potentially facilitates a mutualism between these species (Palanca et al., 2013). These insect species easily move over regional scales, and so presumably play some part in the homogenisation of *S. cerevisiae* within regions. Insects less likely move *S. cerevisiae* over hundreds of kilometres between regions, although *S. cerevisiae* may also be associated with birds which can easily cover these distances (Francesca et al., 2012). Humans are also obvious vectors. Indeed, the patterns of separation, and rates of migration in and out the various regions shown here are nicely in line with the flow of fruit and equipment due to the actions of the NZ wine industry. Marlborough and the Hawke’s Bay are the two largest viticultural and winemaking regions in the country, and fruit from other regions is often transferred to them, mirroring the inferred migration of *S. cerevisiae* into these regions. This national ‘ecological’ scale picture complements and mirrors the global ‘evolutionary’ scale picture revealed for this species: that this is a genetically diverse species that shows some degree of structure and connectivity, and these patterns are consistent with human influenced dispersal (Fay and Benavides, 2005; Legras et al., 2007).

While the above interpretation fits nicely with the population patterns observed here, it is important to consider alternate explanations. The connections between populations could instead be indicative of recent divergence events. The NZ wine industry is very young in evolutionary terms and it is possible that *S. cerevisiae* was introduced to these regions via the introduction of vines and winery equipment such as
barrels (Goddard et al., 2010). The patterns observed in this analysis could be explained by the large wine producing regions of Hawke’s Bay and Marlborough being the source of variation and the outlying regions resulting from founder events with subsequent population expansion and divergence (Hartl and Clark, 1997). The method of migration analysis employed here assumes that population divergence has not occurred, and only invokes migration to explain any similarity in genetic diversity between populations (Beerli, 2009). One issue with a divergence (as opposed to migration) explanation is that source populations must exist prior to the populations they are proposed to have founded. While Hawke’s Bay is one of the oldest wine producing regions in NZ, Marlborough is one of the youngest having only been established around 1970. Thus, the divergence hypothesis fits less well given the vast diversity and admixture observed in the recently established Marlborough region. In addition, whilst it appears that the NZ S. cerevisiae population is reasonably internationally distinct, these patterns of differentiation may also be explained by the inward migration of genotypes from offshore. These explanations are not mutually exclusive however, and it is likely that population divergence from founding populations is occurring alongside inevitable national and international migration of strains due to the vast movement of fruit, equipment and people by the wine industry.

The demonstration that certain regions have ‘signature’ microbial populations is of relevance to the wine industry. It is often suggested that certain wines reflect their geographic origin, and this is encapsulated in the concept of terroir (Bokulich et al., 2014). Classically this was thought to largely result from the interaction between specific Vitis vinifera varieties and the local soils, geography and climate; however, there is limited but increasing evidence showing that the microbes which influence vine growth, fermentation and wine style (as S. cerevisiae does) also exhibit regional differentiation (Gayevskiy and Goddard, 2012; Bokulich et al., 2014; Taylor et al., 2014), as we again demonstrate here. Thus, these data further support the concept that there could be a microbial aspect to terroir. Metabolic profiling of regionally defined genotypes is necessary to determine whether the genetic differentiation demonstrated
Chapter 3: *S. cerevisiae* Population Structure and Migration

here translates to phenotypes that are relevant to wine, and thus whether microbes contribute to terroir in a predictable and consistent way.

Here we provide a more advanced insight into the population biology of a well-established model microbial eukaryote that has also been biotechnologically harnessed by humans since the dawn of civilisation. We take a significant step toward quantifying these processes by providing the first estimates for metapopulation separation and similarity. We reveal *S. cerevisiae* population differentiation in NZ at scales over 100 km, with the most signal provided by the more remote regions, but no differentiation within regions, even between populations inhabiting native forests and vineyards. We also show differential migration of this species between regions, and postulate that this may be due, at least in part, to human influence. By quantifying the magnitude of these forces in microbes we begin to provide one crucial aspect of an inclusive framework attempting to more fully integrate ecological and evolutionary processes.
Chapter 4: Ecology of Sporulation in \textit{S. cerevisiae}

Chapter 4:

SPORULATION IN SOIL AS AN OVER-WINTER SURVIVAL STRATEGY IN \textit{SACCHAROMYCES CEREVISIAE}

Status of chapter:

This chapter has not yet been submitted for publication. I am the lead author in collaboration with Dr Matthew R. Goddard. The co-authorship form is presented at the start of this thesis, after the acknowledgements.
Chapter 4: Ecology of Sporulation in *S. cerevisiae*

**Abstract**

Model microbial systems afford a valuable tool for investigating intricate ecological and evolutionary hypotheses. Here we examine the ecology of *Saccharomyces cerevisiae*, a widely used model microbial eukaryote for which a large body of knowledge exists about its cell biology and genetics. While there is extensive knowledge about how this species behaves in the fruit niche, fruit is ephemeral and thus *S. cerevisiae* requires a strategy to survive when this preferred niche is not available. We hypothesise that *S. cerevisiae* employs a life history strategy targeted at self-preservation rather than growth outside of the fruit niche and resides in the ‘woodland’ niche in a sporulated state, returning to fruit when it’s available via insect vectors. One crucial aspect of this hypothesis is that *S. cerevisiae* must be able to sporulate in this environment. Here we provide evidence that *S. cerevisiae* sporulates in the presence of soil nutrients and does so in a way that maximises its potential reproductive success upon germination. A larger proportion of four-spored asci (tetrads) compared to two-spored asci (dyads) were observed on soil agar compared to controls, resulting in a larger number of potentially viable spores. While there are many more aspects of this hypothesis that require experimental verification, this is the first step towards an inclusive understanding of how this model research organism behaves in its natural environment.
4.1 INTRODUCTION

Single-celled, microbial eukaryotes are essential tools to investigate intricate ecological and evolutionary hypotheses that cannot otherwise be studied in more complex, longer living animals (Jessup et al., 2004). One of the more widely used model research eukaryotes is the budding yeast *Saccharomyces cerevisiae* which is easily stored and cultured, has a short generation time, and is capable of sexual reproduction. These attributes have allowed it to be utilised as a model research organism in various fields including cell biology, physiology, genetics, evolutionary biology and other aspects of molecular biology (Chambers and Pretorius, 2010; Dujon, 2010; Gray and Goddard, 2012; Hittinger, 2013; Hyma and Fay, 2013). Despite this vast knowledge about its biology, comparatively little is known about its ecology.

To begin to understand the ecology of this budding yeast, it is important to have a good understanding of its life-cycle. Using enrichment protocols (Mortimer and Polsinelli, 1999; Serjeant et al., 2008), *S. cerevisiae* is typically isolated from the environment in a diploid state. These diploid cells can undergo mitosis and replicate vegetatively via budding, or meiosis and form up to four haploid spores enclosed in an ascus of either MATa or MATα mating type (Esposito and Klapholz, 1981). Diploid cells are formed again from the fusion of two haploid cells of opposite mating types originating from either the same ascus or tetrad (inbreeding) or from a tetrad of another mother cell (outcrossing). Additionally, if the cells are homothallic, the haploid cell can divide mitotically producing an identical daughter cell, and subsequently switch its own mating type through the production of a site-specific endonuclease (Kostriken and Heffron, 1984). The mother and daughter cell immediately mate or fuse, resulting in a diploid cell with an entirely homozygous genome. This mate-type switching is only observed in the mother cell and none of its haploid daughter cells (Klar, 1987). Most cells that are found in nature are homothallic, although heterothallic cells that cannot mate-type switch also exist (Landry et al., 2006).
Chapter 4: Ecology of Sporulation in *S. cerevisiae*

By considering the potential selective pressures that may have shaped *S. cerevisiae*’s life history strategies that we see today, it is possible to develop hypotheses as to how its life cycle may fit into its environment. Firstly, mitosis or vegetative growth of diploid cells occurs when essential nutrients are abundant such as in fruits or grape juice ready for fermentation (Pringle and Hartwell, 1981). *S. cerevisiae* is considered a fruit specialist and a high genetic diversity is often observed in this niche (Diezmann and Dietrich, 2009; Goddard et al., 2010; Murphy and Zeyl, 2010). Due to its essential role in the wine industry through its capacity to ferment in the presence of oxygen, there is vast knowledge about how it behaves under these conditions ranging from its ability to competitively exclude other yeast species through the production of heat and alcohol (Goddard, 2008) to the variation in the metabolites produced by individual genotypes (Pretorius, 2000; Lopandic et al., 2007; King et al., 2008; Swiegers et al., 2009; Barbosa et al., 2014); however little is known about its life outside the ephemeral fruit niche. Since the ability of *Saccharomyces* yeasts to ferment is considerably more ancient than man-made fermenting technology (Hagman et al., 2013), an alternate life-history strategy must be employed when the sugar rich fruit is not available; however what it is doing during this period and how it survives until the next season of fruit is largely unknown.

Sporulation is a starvation response and is induced in diploid cells containing both *MAT*a and *MAT*a mating types when nitrogen and fermentable carbon sources such as glucose are absent and a non-fermentable carbon source such as acetate is present (Esposito and Klapholz, 1981; Honigberg and Purnapatre, 2003; Neiman, 2005; Piccirillo and Honigberg, 2010). In *S. cerevisiae*’s environment, conditions where sporulation is more likely to be induced are when the abundant nutrients in the fruit are exhausted and winter approaches. Spores and stationary phase cells have been shown to be more resistant than log phase cells to environmental conditions such as repeated freeze-thaw cycles, increased osmolarity and desiccation; however, spores alone have a higher tolerance to short periods of heat shock, high salt concentrations and extremes of pH (Coluccio et al., 2008). Thus a transition into a sporulated state when the fruit drops and winter sets in would provide the cell with increased protection against harsh conditions potentially experienced over-winter. *S. cerevisiae* is known to reside in the ‘woodland’ niche including soil and bark of
Chapter 4: Ecology of Sporulation in *S. cerevisiae*

and around fruiting tree species and oak (Goddard et al., 2010; Wang et al., 2012; Hyma and Fay, 2013; Knight and Goddard, 2015). Isolates from these niches are typically homozygous (Diezmann and Dietrich, 2009; Magwene et al., 2011; Knight and Goddard, 2015), suggesting they have recently been in a sporulated state and have returned to the preferred diploid state via a mate-type switching event and the subsequent fusing of mother and daughter cells; however this is purely speculative and requires experimental validation.

Sporulation is inherently linked to dispersal as it is triggered by starvation and thus the cell needs to move to survive. *S. cerevisiae* is not able to disperse itself thus insects have been inferred to be a potential vector (Mortimer and Polsinelli, 1999; Reuter et al., 2007; Goddard et al., 2010; Stefanini et al., 2012; Palanca et al., 2013; Buser et al., 2014). Past research indicates that spores can survive passage through the gut of flies and cells that have done so exhibit higher rates of outcrossing (Reuter et al., 2007; Coluccio et al., 2008). Temperate species of fruit flies typically overwinter as diapausing pupae and enter the soil after leaving the fruit as winter approaches, emerging as adults the following summer (Bateman, 1972). Therefore, if *S. cerevisiae* proceeds into a sporulated state when the fruit drops to the soil or onto the bark of the tree and flies are potentially ingesting *S. cerevisiae* spores in this ‘woodland’ niche, the cells are more likely to survive passage through the fly gut and be dispersed to the fruit niche by adult flies the next summer. Higher rates of outcrossing exhibited by cells that have passed through the fly gut additionally fits with the higher genetic diversity typically observed in fermenting environments (Diezmann and Dietrich, 2009; Goddard et al., 2010; Murphy and Zeyl, 2010). Thus not only do the cells have increased protection overwinter by being in a sporulated state, but they are also more likely to survive transport back to the fruit niche when it becomes available.

These ideas are very speculative and require experimental investigation. Crucially it relies on the ability of *S. cerevisiae* to sporulate in the ‘woodland’ niche. Here we investigate just that and hypothesise that
sporulation in *S. cerevisiae* can be induced by the presence of soil nutrients. We test how the presence of soil nutrients affects sporulation efficiency in twelve genetically diverse genotypes of *S. cerevisiae* isolated as part of a previous study from both vineyard soil and the ferment of fruits (Knight and Goddard, 2015) and take significant steps forward in our understanding of *S. cerevisiae*’s ecology.

4.2 METHODS

4.2.1 Genotype selection and preparing cultures

Six genotypes isolated from vineyard soil and six genotypes isolated from spontaneous *Vitis vinifera var. Sauvignon Blanc* ferments were selected for analysis from a genetically diverse collection at the University of Auckland (Knight and Goddard, 2015). These genotypes were selected on the basis of being genetically different and diverse representatives of each of these niches as ascertained by microsatellite genotyping at eight loci (Knight and Goddard, 2015). None of the genotypes are genetically similar to a diverse set of international isolates (Liti et al., 2009) or to commonly used commercial strains in New Zealand and are therefore considered to be derived from the natural New Zealand population. All isolates were stored in glycerol (15 %) at -80 °C and were revived at 25 °C in liquid YPD (1 % yeast extract, 2 % peptone and 2 % glucose). Once each culture reached an optical density of 0.6 at a wavelength of 600 nm (about the point where the cells reach lag phase growth) it was centrifuged at 3000 rpm for three minutes and washed twice with 10 mL of sterile water, centrifuging to pellet the cells between each wash. The cells were re-suspended in 1 mL of sterile water, ready for plating.

4.2.2 Preparing soil agar

Since we cannot observe cells directly in a soil environment, a proxy was required. *S. cerevisiae* behaves differently in liquid and solid environments (Vopálenská et al., 2005) so solid soil agar was made to emulate the natural conditions while still being able to observe the cells. Firstly a liquid soil solution or soil ‘tea’ was
made. 200 g of dry soil from Mates Vineyard at Kumeu Wine Estate (West Auckland NZ) was put in 1 L of distilled water and was rocked at room temperature for six hours and settled over night at 4 °C. The supernatant was poured off to separate it from the larger soil particles and then filtered with a 40 µm cell strainer. To sterilise the ‘soil tea’ dimethyl dicarbonate (DMDC) was used in two doses. Firstly it was added at a concentration of 200 µL/L and the solution was stirred for six hours. Another addition was then added at 400 µL/L and the solution was stirred overnight. The sterilised soil ‘tea’ was subsequently mixed with an autoclaved agar solution to a final concentration of 1.5 % agar to create solid soil agar plates. For the initial pilot study just 50 g of dry soil was used to make the soil ‘tea’ but this was scaled up to 200 g of soil for the larger test to try and elicit a larger response.

4.2.3 Pilot study

A pilot study was performed with 100 µl of the cell solution (described above) of each genotype being plated in triplicate on sporulation agar (1 % potassium acetate, 0.1 % yeast extract, 0.05 % glucose and 1.5 % agar) to control for optimal sporulation conditions, plain agar (1.5 % agar) to control for the addition of agar to the soil medium and finally soil agar (described above) as the media of interest, and incubated at 25 °C. At two days and two weeks after inoculation, the proportion of sporulated cells was calculated by scoring over 100 cells (binary result yes/no; ambiguously sporulated cells were not counted; Appendix IV Dataset AIV.2).

4.2.4 Time course study

Again 100 µl of the cell solution (described above) of each genotype was plated in triplicate on to soil agar and plain agar as a control and incubated at 25 °C. Due to the observations of two-spored asci in the pilot study, the number of unsporulated cells, four-spored asci (tetrads) and two-spored asci (dyads) were counted each day for eight days by scoring over 150 cells (ambiguously sporulated cells were not counted).
Chapter 4: Ecology of Sporulation in *S. cerevisiae*

The proportion of sporulated versus unsporulated cells was thus calculated for each sample on each day as well as the proportion of tetrads with respect to the total number of sporulated cells (Appendix IV Dataset AIV.3).

### 4.2.5 Statistical analyses

Statistical analyses were performed in JMP (version 10.0.0). Differences in the proportion of cells sporulated as calculated in both the pilot study and time course study, as well as the proportion of tetrads in relation to the total number of sporulated cells in the time course study were analysed. As proportion data is known to have heterogeneous variance, all data underwent an arcsine transformation prior to analyses (Sokal and Rohlf, 1995). For all analyses, a repeated measures linear mixed effects model was used with the niche of isolation and plate type as fixed effects in a full factorial model with the time point, genotype and replicate included as random factors.

### 4.3 RESULTS

The pilot study revealed a significant effect of plate type on the proportion of sporulated cells ($F_{2,197} = 51.8, P < 0.0001$). The original niche of isolation had no significant effect ($F_{1,10} = 0.83, P = 0.38$), nor did the interaction between the two factors ($F_{2,197} = 1.06, P = 0.35$). The average proportion of cells sporulated for each time point and each plate type can be seen in Figure 4.1. Subsequent Tukey HSD ($\alpha = 0.05$) analysis suggests that all plate types are significantly different from each other with the sporulation media having the highest least square mean of the transformed values being 0.78, followed by soil agar with 0.64 and lastly plain agar with 0.52.
Chapter 4: Ecology of Sporulation in *S. cerevisiae*

Figure 4.1: The average proportion of sporulated cells on each of the three plate types in the pilot study. The size of the error bars represents the standard error around each average.

For the time course study, all genotypes on all plates exhibited some degree of sporulation after eight days of incubation on the soil agar and agar plates (Appendix IV Dataset AIV.3). Significant differences in the proportion of cells sporulated between the plain agar control and the soil agar ($F_{1, 553} = 97.9$, $P < 0.0001$) were observed with a higher proportion of sporulated cells on the soil agar (Figure 4.2). Additionally, as seen in Figure 4.2, of the sporulated cells, the soil agar exhibited a larger proportion of tetrads than the plain agar plates ($F_{1, 553} = 104.8$, $P < 0.0001$). The original niche that the genotype was isolated from had no significant effect on the proportion of sporulated cells or the number of tetrads versus dyads formed ($F_{1, 553} = 0.017$, $P = 0.90$ and $F_{1, 533} = 0.021$, $P = 0.89$ respectively).
Chapter 4: Ecology of Sporulation in *S. cerevisiae*

**Figure 4.2**: The average proportion of the total number of cells sporulated as well as the proportion of tetrads and dyads within that total across the eight days of the time course experiment for both the (a) plain agar and (b) soil agar plates. The error bars represent the standard error around each average.

### 4.4 DISCUSSION

Understanding the basic biology of microbes in their natural environment is essential for understanding the evolutionary trajectories that have shaped them. A vast amount is known about the model microbial eukaryote, *S. cerevisiae*, in fermenting fruit niches but there is scare information regarding its ecology when fruit is not available. Here we provide evidence that *S. cerevisiae* may reside in the soil niche in a sporulated state and hypothesise this response is one of self-preservation when its preferred fruit niche is unavailable.

If *S. cerevisiae* was to cycle between the fruit and soil niches, the populations occupying these niches must be connected. Population genetic studies investigating *S. cerevisiae* report no compelling evidence for population differentiation between fruit associated and ‘woodland’ niches on small geographic scales suggesting these populations are homogeneous and thus connected (Goddard et al., 2010; Hyma and Fay, 2013; Knight and Goddard, 2015). The phenotypic data here support this with the original niche of isolation of the tested genotypes having no effect on the sporulation efficiency or, of the sporulated cells, the proportion of tetrads formed. Therefore the genotypes isolated from the different niches are not
differentially adapted to these conditions in terms of their sporulation efficiency and are phenotypically homogenous for these traits. This is in contrast to previous findings that suggest genotypes isolated from oak trees were more efficient at sporulating and formed asci with predominantly four-spores (tetrad) compared to genotypes isolated from wine fermentations that formed large numbers of two- (dyads) and three-spored (triad) asci (Gerke et al., 2006). These genotypes were not isolated in close geographic proximity with the oak isolates originating solely from North America and the vineyard isolates mostly from wider Europe but also Australia, South Africa and California, thus these findings are potentially confounded by geographic origin.

While the initial niche of isolation does not affect sporulation efficiency or tetrad versus dyad formation in this study, the media does (Figures 4.1 and 4.2). Higher sporulation efficiency was observed on the soil agar when compared to the plain agar control in both the pilot and the time course studies. While the sporulation efficiency on the soil agar was not as high as on laboratory defined sporulation agar as seen in the pilot study, these results do suggest that the presence of soil ‘tea’ in the media stimulates sporulation to a greater extent than the agar control. This is the first direct evidence that the presence of soil nutrients can induce sporulation in *S. cerevisiae*. The capability for *S. cerevisiae* to exist in soil and potentially other ‘woodland’ niches in a sporulated form fits with previous observations of the genetic diversity in these niches. Typically isolates found outside of the fruit niche are largely homozygous (Diezmann and Dietrich, 2009; Magwene et al., 2011; Knight and Goddard, 2015) suggesting recent inbreeding either between sister spores or self-fertilisation of a haploid spore from a mate-type switching event (genome renewal) possibly during enrichment processes required to isolate *S. cerevisiae* from these niches (Mortimer and Polsinelli, 1999; Serjeant et al., 2008).

Of the sporulated cells, a higher proportion of tetrads were observed on the soil agar compared to plain agar. Dyads, or an ascus containing two spores, can result from mutation (Neiman, 2005); however this
Chapter 4: Ecology of Sporulation in *S. cerevisiae*

would result in all cells with that genotype being unable to form tetrads. Since all of the genotypes here were able to form tetrads to some extent, we suggest that the dyads observed here are non-sister dyads (NSDs) formed by a metabolic response. Depletion of the carbon source (such as acetate) after commitment to sporulation does not arrest meiosis due to a lack of nutrients, but instead the cell utilises what external energy it does have and initiates a switch from forming tetrads to forming the less energy expensive NSDs (Davidow et al., 1980; Neiman, 2005; Taxis et al., 2005). This occurs by meiosis II outer plaques only being formed by two of the four spindle pole bodies, one from each spindle, resulting in two spores with homologous chromosomes rather than sister chromatids (Davidow et al., 1980; Neiman, 2005; Taxis et al., 2005; Neiman, 2011) and thus encapsulates the maximum genetic diversity of the mother cell and ensures the two spores formed are of opposite mating type (Taxis et al., 2005; Neiman, 2011). This is an advantage to the cell as maintaining heterozygosity has been associated with an immediate fitness benefit upon germination (Taxis et al., 2005); however, if given the available resources to do so, completing meiosis and forming a four-spored ascus or tetrad should be preferred as it would contain the same amount of genetic diversity as NSDs, but with twice the number of potentially viable cells, increasing the cells fitness through an increased likelihood that one or more of those cells would survive to the next generation.

It is not surprising that a higher proportion of tetrads were observed on the soil agar as it would be assumed this media contains more nutrients than plain agar alone; however, even on the soil agar many more dyads were observed than tetrads. Soil itself would contain a higher concentration of nutrients compared to the soil agar used here. Here the nutrients contained in soil have firstly been extracted into water (the soil ‘tea’) and then further diluted with agar to solidify it; therefore it is possible that soil itself may elicit the formation of a higher proportion of tetrads. It could be argued that any media without a fermentable carbon source but more nutrients than agar could elicit a larger sporulation response than plain agar alone. Chemical analyses of the soils and bark that *S. cerevisiae* is found in would allow us to examine whether the necessary nutrients required for efficient sporulation are present in these niches. If
Chapter 4: Ecology of Sporulation in *S. cerevisiae*

so, this would add to the evidence provided here in support of *S. cerevisiae* being to exist in the ‘woodland’ niche in a sporulated form.

Based on examination of past research about the biology of *S. cerevisiae*, we propose this species employs a life-history strategy favouring self-preservation in unfavourable environments and enters a state of dormancy by initiating sporulation. While there are more aspects of this hypothesis that remain to be tested, we provide evidence that *S. cerevisiae* is able to sporulate in the presence of soil nutrients and does so in a way that maximises its potential genetic diversity upon germination. As this behaviour would affect the frequency and nature of sex experienced by *S. cerevisiae* in its natural environment, the levels of homozygosity and frequency of recombination in the population would be altered and thus it has implications for estimates of population structure and connectivity, as well as inferences of its evolutionary history and potential future trajectories (Magwene et al., 2011).
Chapter 5: Evidence for a Microbial Aspect to Terroir

Chapter 5:

REGIONAL MICROBIAL SIGNATURES TRANSLATE TO DIFFERENTIAL WINE PHENOTYPES PROVIDING EVIDENCE FOR A MICROBIAL ASPECT TO TERROIR

Status of chapter:

A shortened version of this chapter is currently under review with the journal Scientific Reports. This journal formats its articles with the methods at the end; however in keeping with the rest of this thesis, this has been rearranged here and like other chapters, the methods are placed after the introduction. I am the lead author of this publication in collaboration with Dr Steffen Klaere, Dr Bruno Fedrizzi and Dr Matthew R. Goddard. The co-authorship form is presented at the start of this thesis, after the acknowledgements.
Chapter 5: Evidence for a Microbial Aspect to Terroir

Abstract

Many crops with the same genotype display differential geographic phenotypes in terms of physical and sensorial signatures, but the factors that drive these differences remain elusive. The potential contribution of, and link between, microbes and differential agricultural phenotypes has been ignored until recently. Microbial communities and populations are increasingly shown to significantly vary between geographic regions, and here we conduct the first general test to evaluate whether this translates into differential crop phenotypes. Wines from different geographic regions around the world are known for their unique aroma and flavour characteristics, collectively encapsulated in the concept of terroir. Here we utilise the genetically well described and genetically differentiated population of *Saccharomyces cerevisiae* in New Zealand and quantify the contribution of this variance to the distinctiveness of wine. We perform small scale ferments using single genotypes and co-inoculations of six representative genotypes from each of six major regions and analyse the resulting wine phenotype for regional signal: a suite of 39 volatile compounds and properties derived by the yeast during fermentation. We reveal that genetically distinct yeast populations from different regions significantly affect wine phenotypes, and show that a complex mix of chemicals drives this. These findings reveal the importance of microbial populations on the regional identity of wine, and may potentially extend to other important agricultural commodities. Moreover, this suggests that the long-term implementation of methods that maintain differential biodiversity may have tangible economic imperatives as well as being desirable in terms of employing agricultural practices that increase responsible environmental stewardship.
Chapter 5: Evidence for a Microbial Aspect to Terroir

5.1 INTRODUCTION

Microbes play key roles in the production of quality agricultural commodities for reasons ranging from their effect on crop nutrient availability via rhizosphere interactions with roots, through to their role in crop disease pressure: ultimately microbes influence plant and fruit health (Whipps, 2001; Peiffer et al., 2013; Philippot et al., 2013). Additionally microbes transform plant products to economically and socially important commodities such as coffee, chocolate, bread, wine, beer and other fermented beverages (Fleet, 2006). Many crops that comprise the same or very similar genotypes display differential geographic phenotypes in terms of the physical and sensorial signatures of their produce: this is generally encapsulated by the concept of terroir (Bokulich et al., 2014). Often the chemical descriptors of these differential geographic phenotypes are well documented (Costa Freitas and Mosca, 1999; Lund et al., 2009; Son et al., 2009; Obuchowicz et al., 2011; Benkwitz et al., 2012; López-Rituerto et al., 2012; Robinson et al., 2012; Torres-Moreno et al., 2015); however, the environmental factors that drive these differences remain elusive (Van Leeuwen and Seguin, 2006). These differential agricultural geographic phenotypes are classically thought to result from complex interactions between specific crop species and local soils, climate and agricultural practices, and are commercially important as they add distinctiveness and thus value to products (Van Leeuwen and Seguin, 2006). However, the potential contribution of, and link between, microbes and differential agricultural geographic phenotypes have been ignored until recently (Gayevskiy and Goddard, 2012; Bokulich et al., 2014; Taylor et al., 2014; Knight and Goddard, 2015).

Logically there are two necessary requirements if microbes are to contribute to differential geographic agricultural phenotype signatures: first, microbial communities and populations must vary geographically; and, second, these differential microbial populations must measurably influence crop phenotypes. Wine has been made by humans since the dawn of civilisation, is still an important social and economic commodity, arguably displays the strongest geographic signatures of all agricultural products, and thus is a superb model to evaluate the degree to which there might be a microbial aspect to terroir. Microbes,
predominantly fungi, potentially significantly affect the ‘phenotype’ of wine firstly by affecting grapevine and fruit health and development, and thus quality (Barata et al., 2012), and secondly by manipulating wine flavour, aroma and style due to their fermentative actions (Dubourdieu et al., 2006; Swiegers et al., 2009). During alcoholic fermentation fungi not only convert sugars into ethanol but also produce an array of secondary metabolites including volatile compounds that are important to wine aroma and flavour (Lambrechts and Pretorius, 2000; Swiegers et al., 2005a). While grape-derived compounds may provide varietal distinctions, yeast-derived acids, alcohols, carbonyl compounds, phenols, esters, sulfur compounds and monoterpenoids all significantly contribute to wine quality and aroma (Swiegers et al., 2005a; Sumby et al., 2010).

Here we focus on the potential for microbes to influence wine phenotypes via fermentation. Wine may be made by either attempting to remove the array of microbes that are naturally associated with grapes and deliberately inoculating with a commercial strain of yeast, or allowing the microbes naturally associated with grapes to conduct the ferment (Goddard, 2008). The former inoculated option reduces any potential for microbes to contribute to terroir, during fermentation at least, and has only been available commercially to winemakers since 1965 (Pretorius, 2000). The latter has been employed by humans since the dawn of civilisation and is known as spontaneous or wild fermentation, and may comprise at least tens of species and hundreds of strains of *S. cerevisiae* (Goddard et al., 2010; Bokulich et al., 2014). There is recent evidence for regional delineations of microbial communities and populations associated with vines, and so the spontaneous ferment option may maximise the potential for microbes to impart regional signatures to wine (Bokulich et al., 2014; Taylor et al., 2014; Knight and Goddard, 2015). In addition to the possibility that individual yeast genotypes may impart different properties to wine, experiments show that different yeast species and strains of *Saccharomyces* may synergistically interact during fermentation to create different flavour and aroma profiles (Howell et al., 2006; Anfang et al., 2009). Since spontaneously fermented wine comprises a diversity of yeast species and strains of *S. cerevisiae*, these metabolic
interactions may potentially be the key to any microbial signature contributing to terroir. As far as we are aware neither of these ideas have been previously objectively tested for wine or any agricultural product.

As well as recent robust experimental evidence revealing that microbial communities associated with grape vines significantly vary with respect to geographic region (Bokulich et al., 2014; Taylor et al., 2014), we have also recently provided compelling evidence for genetically differentiated sub-populations of S. cerevisiae contributing to spontaneous ferments in major regions in New Zealand (NZ) (Knight and Goddard, 2015). Using population genetic analyses here we select appropriate genetic representatives from these regional S. cerevisiae sub-populations, and analyse their fermentative effects on a suite of chemicals known to significantly affect the phenotype of wine. Here we test whether single genotypes from geographically distinct sub-populations, or mixes of genotypes representing these sub-populations, impart any differential volatile chemical properties to wine, and in doing so conduct the first general test for whether there is a microbial aspect to terroir.

5.2 METHODS

5.2.1 Genotype selection

Six genotypes from six major wine growing regions in NZ were selected to represent the genetic diversity in each region (See Figure 5.1 for geographic locations). Here we specifically employed genotypes isolated from spontaneous ferments by Knight and Goddard (2014). We selected one genotype from each region that harbour ed at least one allele that was unique to that region while the remaining genotypes were selected to cover the diversity of ancestry profiles reported in each regional population as reported in Knight and Goddard (2014).
Chapter 5: Evidence for a Microbial Aspect to Terroir

5.2.2 Micro-fermentation

The 36 single genotype ferments were conducted in triplicate across three batches due to space constraints. For statistical purposes, each batch contained one replicate of every treatment. Each ferment contained 230 mL of Marlborough (NZ) Sauvignon Blanc juice from the 2012 vintage (pH = 3.1, 22.1 °brix) supplied by Pernod Ricard NZ, sterilized with 200 µL/L dimethyl dicarbonate (DMDC) and with the SO2 levels adjusted to 10 mg/L. Each *S. cerevisiae* genotype was grown up independently in liquid YPD (1 % yeast extract, 2 % peptone, 2 % glucose) prior to inoculation. The live cell concentration of each culture was determined using a haemocytometer with methylene blue staining, and cells were inoculated to give a final concentration of $2.5 \times 10^6$ cells/mL. Regional co-ferments were performed by inoculating all six genotypes isolated from each region in equal proportions in to the same ferment to the same final concentration of $2.5 \times 10^6$ cells/mL and again fermented in triplicate along with the single genotype ferments. Replicate uninoculated controls were included in each batch to monitor and control for weight loss via evaporation and identify potential contamination issues. This totalled 126 experimental ferments and 9 uninoculated controls. Ferments were conducted at 15 °C with 150 rpm shaking in 250 mL Erlenmeyer flasks with airlocks. Fermentation progress was monitored by weighing the flasks daily (El Haloui et al., 1988) and were considered finished when the rate of weight loss was below 0.001 g/hr (after controlling for evaporation as calculated from the controls) or when they reached 30 days. Once finished, ferments were centrifuged at $6000 \times g$ for 10 minutes to pellet cells. The supernatant was decanted and frozen at -20 °C until chemical analyses were performed.

5.2.3 Blends

After fermentation, regional blends were constructed from the single genotype ferments. Equal proportions of wine from ferments of each of the six genotypes from each region were homogenised, creating triplicate regional blends for each of the six regions. This resulted in a total of 144 wine samples for chemical analyses.
5.2.4 Chemical analyses

In total, 40 chemical compounds and quality parameters were quantified for analysis. Final ethanol concentration, pH, residual sugar, volatile acidity (VA) and titratable acidity (TA) were quantified using FTIR (Fourier Transform Infrared Spectroscopy) with a FOSS WineScan™ FT120. The varietal thiols 3MH, 3MHA and 4MMP were quantified using an ethyl propiolate derivatisation and analysed on an Agilent 6890N gas chromatograph (Santa Clara, CA, USA) equipped with a 7683B automatic liquid sampler, a G2614A autosampler and a 593 mass selective detector as outlined in Herbst-Johnstone et al. (2013). 32 esters, higher alcohols, terpenes, C6-alcohols and fatty acids were quantified simultaneously using a HS-SPME/GC-MS method outlined in Herbst-Johnstone et al. (2013a). Raw data was transformed with GCMSD Translator and peak integration was performed using MS Quantitative Analysis, both part of the Agilent MassHunter Workstation Software (Version B.04.00, Agilent Technologies). Unfortunately 4MMP was consistently below the detection threshold of the quantification method used and could not be quantified accurately. It was therefore removed from the analysis, leaving 39 chemical compounds and quality parameters for analysis.

5.2.5 Statistical analyses

The sigmoid or altered Gompertz decay function described by Tronchoni et al. (2009) was used to build a model of fermentation kinetics for each ferment from the weight loss data to infer the lag phase. The data was fitted using the non-linear least squares method implemented in the R package “nlstools” (Bates and Watts, 1988). Differences in the lag phase between batches were tested using a mixed linear model in JMP (version 10.0.0) accounting for genotype and stuck ferments as random factors.

Statistical tests for regional signal were performed on the chemical profiles for all datasets separately using a PERMANOVA approach as implemented in the R package “vegan” (Anderson, 2001). Jaccard distances
were used to calculate pairwise distances in the model and 10,000 permutations of the raw data constrained at the genotype level to account for the dependency between genotypes and their replicates, were performed for the hypothesis tests ($F$-tests). Full factorial models were implemented and subsequently reduced upon analysis of the results to obtain the model of best fit. Pairwise PERMANOVA analyses were performed between all combinations of regions for the single genotype ferments and the $P$-values were subsequently corrected for multiple comparisons using the Benjamini-Hochberg (BH) method (Benjamini and Hochberg, 1995). Constrained Correspondence Analysis (CCA), implemented in the package "vegan," was used to visualise the data. This method allows the model to be built into the visualisation, providing the most appropriate transformation and orientation of the data to visualise differences between the factors of interest.

The chemical drivers of the observed regional differentiation were determined using both individual ANOVA analyses testing the factorial model of "region" and "batch" for each of the chemical properties, and the chemical vectors explaining the distribution of data points in the CCA plot. Chemicals with significant $R^2$ values that exceeded 0.25 in the ANOVA analyses were visualised as vectors in the CCA plot to determine which regions they were mostly influencing.

A Mantel test was performed between genetic distance, calculated from microsatellite profiles (Knight and Goddard, 2015) in GenAlEx (Genetic Analyses in Excel) version 6.5 (Peakall and Smouse, 2006; Peakall and Smouse, 2012) and a chemical distance matrix calculated using the Jaccard similarity coefficient. The inferred ancestry profiles of each genotype was previously calculated by Knight and Goddard (2014) and clustered the individual genotypes by genetic similarity. These clusters are indicated in Appendix IV Dataset AIV.4 and largely correspond to the region of isolation with a few exceptions where the genotypes are considered to be of mixed ancestry. PERMANOVA analysis replacing the factor of "region" with "inferred
population” allowed us to investigate whether genetic similarity is a better predictor of chemical differences than region.

The CCA analysis did not provide the transformations used in the original visualisation of the single genotype ferments so to incorporate the co-ferments and regional blends in to this visualisation a linear model was fitted to the first two principle components. This was subsequently used to predict the position of the blend and co-ferment data points within the same 2-D space. These linear models explained 100 % of the variation in the CCA components and thus accurately reflect the transformation of the data in the original visualisation.

5.3 RESULTS

5.3.1 Selection of *S. cerevisiae* genotypes

Recently Knight and Goddard (Knight and Goddard, 2015) isolated 3 900 *S. cerevisiae* from native forests, vineyards, and the spontaneous ferments of *Vitis vinifera* var. Sauvignon Blanc from six major wine growing regions in NZ (Hawke’s Bay, Martinborough, Nelson, Wairau Valley, Awatere Valley and Central Otago). Microsatellite genotype profiling of these isolates revealed the presence of 295 different genotypes. Bayesian population structure analyses revealed evidence for distinct sub-populations, and statistical analyses of the resulting ancestry profiles (Gayevskiy et al., 2014) showed these sub-populations were significantly correlated with geographic location (Knight and Goddard, 2015). Here we select six *S. cerevisiae* genotypes from each of these six regional populations that span and represent the genetic diversity within each area.
Chapter 5: Evidence for a Microbial Aspect to Terroir

5.3.2 Ferment performance

All ferments were conducted using the same batch of homogenised and sterilised Sauvignon Blanc juice from Marlborough in New Zealand. Individual *S. cerevisiae* genotypes, and co-inoculations of genotypes representing regional populations, were fermented in triplicate across three separate batches. The extent to which sugars were fermented was analysed by weight loss (El Haloui et al., 1988), and most lost approximately 25 g indicating complete fermentation given the 220 g of sugar in the juice initially. One genotype from the Wairau Valley failed to ferment at all and was removed from all analyses. Eleven single genotype ferments, all in the third batch, displayed significantly less weight loss than the remaining ferments ($F_{1,108} = 905.9, P < 0.0001$), indicating incomplete fermentation. Incomplete fermentation is known to affect the volatile profiles of wines (Malherbe et al., 2009), and so we conservatively removed these ferments from all further analyses. Lag phase, the time taken for fermentation to initiate, differed significantly between batches ($F_{2,89} = 7.73, P = 0.0008$), and since each batch contained one replicate of each sample, this was controlled for in subsequent statistical analyses by introducing a “batch” factor.

5.3.3 Chemical profiles produced by single genotype ferments correlate with region of microbe origin

We quantified the concentrations of 39 volatile compounds and wine quality parameters produced in each of the ferments using targeted GC-MS and FTIR analyses. First we analysed the volatile profiles deriving from ferments conducted by single yeast genotypes only. PERMANOVA (permutational multivariate analysis of variance) employing a full factorial model with “region” and “batch” as main effects, where permutations kept replicates of each genotype together, revealed that both factors significantly affected volatile profiles (both $P = 0.001$), but provided no evidence of an interaction between these main effects (Table 5.1a). The $R^2$ value for the region effect was greatest reporting the geographic origin of the *S. cerevisiae* genotypes explained approximately 10% of the total variation in the chemical profiles (Table 5.1a). The lack of a significant interaction term means this result is not confounded by the batch effect. In
addition, we analysed these differential chemical profiles by also accounting for human perception thresholds of compounds. Where available, we used empirically determined odour activity values (OAVs) to standardise the various chemical concentrations in these ferments (Francis and Newton, 2005; Swiegers and Pretorius, 2005). The results of PERMANOVA agreed with the initial analyses and again revealed a highly significant effect of the region of *S. cerevisiae* isolation on these wine phenotypes (Region: $R^2 = 0.127, P = 0.002$; Table 5.1b).

Pairwise PERMANOVA analyses between the chemical profiles produced by *S. cerevisiae* genotypes originating from different regions revealed that just three of the 15 pairwise region comparisons were significant (Appendix III Table AIII.1), and when multiple correction weightings are applied none are deemed significant. Recalling that overall the region of *S. cerevisiae* genotype origin significantly affects volatile profiles, it has been argued that when dissecting differences more emphasis should be placed on the magnitude of the effect rather than the *P*-value (Nuzzo, 2014). Here the magnitude of effect represents the distinction between chemical profiles of genotypes originating from different regions, and are visualised in Figure 5.1. The profiles of yeasts originating from Nelson and Central Otago are the least similar to other regions (Figure 5.1) and report the highest overall pairwise $R^2$ values ranging from 0.032 – 0.121 with means of 0.077 and 0.078 respectively (Appendix III Table AIII.1). Nelson’s similarity to all regions is relatively low with the exception of the Awatere Valley (Figure 5.1). The Awatere and Wairau Valleys are the most similar to other regions (Figure 5.1) and report the least distinct chemical profiles compared to other regions with $R^2$ values ranging from 0.026 – 0.077 with means of 0.042 and 0.052 respectively (Appendix III Table AIII.1). Martinborough and Hawke’s Bay are intermediate with a mix of both highly similar and more distinct relationships with other regions (Figure 5.1; Appendix III Table AIII.1).
Chapter 5: Evidence for a Microbial Aspect to Terroir

Table 5.1: Summary of all PERMANOVA analyses.

<table>
<thead>
<tr>
<th>Factors</th>
<th>Df</th>
<th>F Model</th>
<th>$R^2$</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Single strain ferments only</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Region</td>
<td>5</td>
<td>2.056</td>
<td>0.100</td>
<td>0.001 ***</td>
</tr>
<tr>
<td>Batch</td>
<td>2</td>
<td>3.687</td>
<td>0.072</td>
<td>0.001 ***</td>
</tr>
<tr>
<td>Region*Batch</td>
<td>9</td>
<td>0.860</td>
<td>0.076</td>
<td>0.093</td>
</tr>
<tr>
<td>Residuals</td>
<td>77</td>
<td>0.752</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>93</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(b) Single strain ferments only, with chemicals standardised by OAV</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Region</td>
<td>5</td>
<td>2.758</td>
<td>0.127</td>
<td>0.002 **</td>
</tr>
<tr>
<td>Batch</td>
<td>2</td>
<td>4.302</td>
<td>0.079</td>
<td>0.001 ***</td>
</tr>
<tr>
<td>Region*Batch</td>
<td>9</td>
<td>0.987</td>
<td>0.082</td>
<td>0.166</td>
</tr>
<tr>
<td>Residuals</td>
<td>77</td>
<td>0.711</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>93</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(c) Single strain ferments with strains with mixed ancestry removed</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Region</td>
<td>5</td>
<td>3.176</td>
<td>0.198</td>
<td>0.006 **</td>
</tr>
<tr>
<td>Batch</td>
<td>2</td>
<td>3.092</td>
<td>0.077</td>
<td>0.005 **</td>
</tr>
<tr>
<td>Region*Batch</td>
<td>9</td>
<td>1.482</td>
<td>0.166</td>
<td>0.056</td>
</tr>
<tr>
<td>Residuals</td>
<td>45</td>
<td>0.560</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>61</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(d) Co-ferments and blends only, testing for effect of the type of ferment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type</td>
<td>1</td>
<td>2.425</td>
<td>0.061</td>
<td>0.014 *</td>
</tr>
<tr>
<td>Batch</td>
<td>2</td>
<td>3.698</td>
<td>0.186</td>
<td>0.014 *</td>
</tr>
<tr>
<td>Residuals</td>
<td>30</td>
<td>0.753</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>33</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(e) All co-ferment samples only</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Region</td>
<td>5</td>
<td>1.555</td>
<td>0.346</td>
<td>0.073</td>
</tr>
<tr>
<td>Batch</td>
<td>2</td>
<td>2.364</td>
<td>0.210</td>
<td>0.073</td>
</tr>
<tr>
<td>Residuals</td>
<td>10</td>
<td>0.444</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>17</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(f) All blend samples only</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Region</td>
<td>5</td>
<td>1.375</td>
<td>0.339</td>
<td>0.196</td>
</tr>
<tr>
<td>Batch</td>
<td>2</td>
<td>1.704</td>
<td>0.168</td>
<td>0.196</td>
</tr>
<tr>
<td>Residuals</td>
<td>10</td>
<td>0.493</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>17</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(g) All single strain ferments but with odd Nelson samples removed</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Region</td>
<td>5</td>
<td>2.412</td>
<td>0.117</td>
<td>0.003 ***</td>
</tr>
<tr>
<td>Batch</td>
<td>2</td>
<td>4.259</td>
<td>0.083</td>
<td>0.001 ***</td>
</tr>
<tr>
<td>Region*Batch</td>
<td>9</td>
<td>0.941</td>
<td>0.082</td>
<td>0.095</td>
</tr>
<tr>
<td>Residuals</td>
<td>74</td>
<td>0.718</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>90</td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 5.1: A map of the regions the tested genotypes of *S. cerevisiae* were isolated from and the pairwise similarity in the chemical profiles produced by each. Values are calculated as $1-R^2$ from the pairwise PERMANOVA analyses (Appendix III Table AIII.1). The inset indicates the portion of NZ highlighted in the larger map.
Chapter 5: Evidence for a Microbial Aspect to Terroir

To effectively visualise the differences in chemical profiles of each wine produced, the data were transformed and plotted using Constrained Correspondence Analysis (CCA) (Ter Braak, 1986). This is analogous to a principle component analysis in that transformations of the data are performed to provide components that allow the data to be visualised in 2-D plots. The CCA additionally partitions these components into a part that is explained by the specified linear model (in this case “region + batch”) and a part that is residual to that model. The plot that is produced rotates the data to the best orientation to observe the variation explained by the model. Overall a large overlap is observed between chemical profiles derived from genotypes from different regions (Figure 5.2); however, the chemical profiles of Central Otago genotypes cluster in the upper half and those from Nelson mostly toward the lower left quadrant with the exception of the three replicate samples from one genotype that is located in the upper right quadrant (Figure 5.2a). The genotypes from Wairau and Awatere Valleys have the largest ellipses indicating a larger variability in the chemical profiles of these samples (Figure 5.2b).
Figure 5.2: CCA analysis of the 105 single genotype ferments analysed. (a) All sample points coloured by region. (b) Regional averages and 50 % ellipses.
Chapter 5: Evidence for a Microbial Aspect to Terroir

5.3.4 Chemical drivers of regional differentiation in single ferment samples

Next we evaluated which components of the volatile profiles might be driving these differences in wine phenotype. After adjusting the $P$-values for multiple comparisons, individual ANOVA analyses revealed that concentrations of 25 of the 39 compounds significantly vary with respect to the region of origin of the yeast genotype (Appendix III Table AIII.2). $R^2$ values range from zero to 38% of the variation being explained by the $S.\ cerevisiae$ genotype region of isolation, and no one class of chemical compound is exclusively responsible for the regional signal for wine phenotypes (Appendix III Table AIII.2).

The CCA additionally provides vectors indicating the direction and magnitude of the influence that each chemical property has on the distribution of data points in the plot, providing a mechanism to infer which chemicals are prominent in regions. These vectors show four compounds (three esters and one fatty acid) that have the greatest impact on the differentiation of wine phenotypes with a magnitude larger than 0.25 (Figure 5.3a), although this does not necessarily imply that these chemicals are solely driving the regional signal. By focusing only on the vectors for the chemical properties that reported a high and significant $R^2$ value (above 0.25; Appendix III Table AIII.2), we can decipher and visualise the properties most important to the differences in chemical profiles between ferments conducted by yeasts derived from different regions (Figure 5.3b and c). This reveals that concentrations of ethyl isobutyrate and ethyl-2-methylbutanoate, which have apple and sweet fruit sensory descriptors, are on average both greatest in the ferments conducted by the genotypes deriving from Nelson and least in those from Central Otago and Martinborough. In addition, concentrations of ethyl butanoate (sensory descriptors of peach, apple and sweet) are on average greatest in ferments conducted by genotypes derived from Martinborough, and least in ferments conducted by genotypes derived from Nelson (Figure 5.3). β-damascenone (sensory descriptors of apple, honey and floral) concentrations are on average greater in the ferments conducted by yeast genotypes derived from the Awatere and Wairau Valleys comprising the larger Marlborough region, and least from the fermentations conducted by genotypes deriving from the Hawke’s Bay. Together this paints
an intuitively sensible picture and reveals that the differential wine phenotype signatures driven by yeasts derived from different regions are not one-dimensional but multi-faceted.

5.3.5 The genetic basis for differences in chemical profiles

While not exclusively genetically determined, the types and concentrations of metabolites produced by S. cerevisiae are significantly influenced by yeast genotype. It is thus not surprising that a Mantel test evaluating the correlation between S. cerevisiae genotype genetic distance (using microsatellite profiles) (Knight and Goddard, 2015) and volatile chemical profile similarities (calculated using Jaccard dissimilarity) reveal they are significantly correlated ($R^2 = 0.189; P < 0.0001$). Additionally, a PERMANOVA analysis using the assignment of genotypes to inferred genetic clusters calculated using InStruct (Knight and Goddard, 2015) as a factor, as opposed to region of origin, increased the $R^2$ value by 0.051 to 0.151 or 15% ($P = 0.007$). Some of the genotypes do not have a high proportion of ancestry to any one inferred population, and thus have mixed ancestry to different regions (Appendix IV Dataset AIV.4). If these hybrid genotypes are removed and only those genotypes with a ‘clean’ geographic signal are analysed, the PERMANOVA analysis with region as a factor reveals an increase in $R^2$ to 0.198 ($P = 0.006$), double that of the original analysis (Table 5.1c).
Chapter 5: Evidence for a Microbial Aspect to Terroir

Figure 5.3: Visualisation of the chemicals that individually explain more than 25% of the regional variation as calculated by ANOVA analyses. (a) The direction and magnitude of all chemical loading vectors, with labels for the chemicals that reported a magnitude above 0.25. The blue circles represent the position of 0.1 and 0.25. (b) The chemical loading vectors in the CCA analysis for those that reported an $R^2$ value for region larger than 0.25 in the ANOVA analyses (Appendix III Table AIII.2). The blue circles represent the position of 0.1 and 0.25. (c) The same chemical loading vectors reported in b with respect to the regional centres of the chemical profiles.
5.3.6 The effect of regionally co-fermented genotypes and blended wines on volatile profiles

There is evidence to show that the presence of other yeasts during fermentation, be they conspecifics or other species, may affect the subsequent volatile profiles of wine compared to the profiles produced when genotypes ferment in isolation (Howell et al., 2006; Anfang et al., 2009). Having quantified the regional signal demonstrated by genotypes fermented in isolation, we moved to evaluate whether there are interactions between genotypes from each region that may affect and potentially alter regional signals for wine phenotypes. We compared the volatile profile of regional co-ferments, produced by inoculating all six genotypes from a region together in equal proportions, to regional blends, created by mixing the final wine produced by single genotypes from each region in equal proportions. PERMANOVA reveals that the type of ferment (co-ferment or blend) has a significant effect on chemical profiles ($R^2 = 0.061$, $P = 0.014$; Table 5.1d). Again CCA was used to visualise the differences between the chemical profiles, and while overlap between the blends and co-ferments is evident, the blended ferments show less variability than the co-ferments, and are typically placed in the lower right of the plot (Appendix III Figure AIII.1a). Individual chemical ANOVA and the resulting CCA analysis show the main differences between the co-ferments and blends are driven by ethyl decanoate, ethyl dodecanoate, methyl octanoate and ethyl octanoate (Appendix III Figure AIII.1b).

Where the profiles of co-ferments and blends sit within their respective regional population of single genotype ferment profiles is of interest for investigating potential interactions between genotypes during fermentation. Due to the significant difference between co-ferments and blends, we hypothesised that CCA analyses would reveal the profile of blends to reside toward the centre of the population of single genotype ferments profiles from that same region, while the co-ferments would not show this pattern. As hypothesised, the chemical profiles from the blended ferments tended towards the centre of the single genotype ferments for most regions (Appendix III Figure AIII.2), but the co-ferments typically do not, and
reinforce the concept that blends and co-ferments provide different volatile signatures (Howell et al., 2006; Anfang et al., 2009).

While the co-fermentation of multiple genotypes significantly affects the phenotype of wine compared to blending, it appears to erode signal for wine phenotype regionality, as PERMANOVA analysis reveals no strong regional co-ferment effect on volatile profiles ($R^2 = 0.346, P = 0.073$; Table 5.1e). However, this may be an issue of statistical power – only three replicates of regional co-ferments and blends were implemented compared to the six volatile profiles from each of six genotypes from each region in the initial analysis. It is worth noting that the $P$-value for the effect of region reported by the co-ferments is marginal ($P = 0.073$), but the value for blends is not ($P = 0.196$; Table 5.1 e and f), and might suggest that blending more significantly erodes any signal for regional wine phenotype than co-fermentation does.

5.4 DISCUSSION

Microbes play essential roles in the production of agricultural commodities, and influence crop quality and sensory attributes (Fleet, 2006; Barata et al., 2012). Using wine as a model system, we experimentally tested and quantified the extent to which genetically distinct regional populations of *S. cerevisiae* affect wine phenotype in terms of volatile composition. We provide the first direct evidence showing that regionally distinct microbial populations translate into distinctions in wine chemical properties, and thus show that microbes significantly contribute to the regional identity or *terroir* of wine. These findings may potentially extend to the differential effects of microbes on other important agricultural crops and produce generally.

The ability of microbes to affect differential crop phenotypes is potentially greater than we estimate here. First, we have not evaluated microbes’ effect on crop development and how this might vary between
Chapter 5: Evidence for a Microbial Aspect to Terroir

differential geographic communities and populations. This is apparent in some sense, as different crops
tend to suffer different levels of disease in different areas. The subtler effects of microbes on crop
development and quality are mostly not understood. Moreover, many other species of fungi and bacteria
contribute to the natural conversion of juice to wine and many of these also significantly affect wine
phenotype, and there is good evidence to show these may synergistically interact (Howell et al., 2006;
Anfang et al., 2009). Thus, the presence of regionally differentiated communities of yeast and bacteria
associated with ripe fruit, as has been demonstrated (Gayevskiy and Goddard, 2012; Bokulich et al., 2014;
Taylor et al., 2014; Knight and Goddard, 2015), may further affect differences in wine phenotype over what
we have revealed here. Here we conservatively remove both these effects as we use the same
homogenised batch of grape juice and only examine the ability of differential populations of just one
species to manipulate crop produce. Even so, we provide evidence that different natural populations of \textit{S.}
\textit{cerevisiae} residing in different regions may significantly affect wine phenotype.

This regional microbial effect was most apparent when the volatile profiles of individual members from
populations residing in different areas were compared. Strains of \textit{S. cerevisiae} are known to metabolically
interact during fermentation and the final profile of a co-fermented wine cannot be mimicked by blending
wines fermented by monocultures of the same strains (Howell et al., 2006), and we see this again here.
However, the regional effect appears diminished when the members from different populations were
fermented together, though this may have been due to lower statistical power.

The chemicals responsible for the differences between regions are not consistently from any particular
class (Appendix III Table AIII.2), and thus the signals for difference in wine phenotype by region are
complex, which makes intuitive sense. We attempted to evaluate the impact of how humans might
perceive these differences in wine phenotypes by standardising chemical concentrations with published
OAVs (Francis and Newton, 2005; Swiegers and Pretorius, 2005). This analysis again reported a significant
effect of regionally differentiated microbes on wine phenotypes; however, OAVs are subjective to an extent, and interactions between chemicals that may lead to enhancement or masking of aromas are not accounted for here (Francis and Newton, 2005). Ultimately the inclusion of sensory trials in these kinds of studies would add an extra layer to evaluate the extent that microbes play in the geographic differentiation of wine phenotypes.

It is well characterised that the types and amounts of volatiles produced by *S. cerevisiae* have a strong genetic component (Murat et al., 2001; Lopandic et al., 2007; Swiegers et al., 2009), and it is thus not surprising that the varying degrees of regional differentiation between chemical profiles largely mirrors the genetic differences between populations (Knight and Goddard, 2015). Genotypes from Central Otago are the most genetically distinct and produce the most distinct wine phenotypes. In contrast, genotypes from the Awatere and Wairau Valleys, that together comprise the Marlborough region, are the least genetically differentiated and in turn do not produce wines that have a strong regional chemical signal. The genotypes isolated from Nelson show strong genetic differentiation (Knight and Goddard, 2015) but the chemical profiles produced are distinct from most regions apart from the Awatere. This appears largely influenced by three replicates of one genotype that clusters away from the other Nelson samples (Figure 5.2a). This genotype was designated as having mixed ancestry in Bayesian analyses (Appendix IV Dataset AIV.5; Knight and Goddard, 2015) and is more aligned with the genetically homogenous populations residing in the Awatere and Wairau Valleys (Marlborough). If this genotype is removed from the analysis the $R^2$ value for “region” increases ($R^2 = 0.117, P = 0.003$; Table 5.1g), and correspondingly the ellipse around the remaining Nelson samples is substantially reduced in size (Appendix III Figure AIII.3), providing evidence that this genotype is masking the regional signal provided by the remaining Nelson genotypes. The original population genetic analyses of these *S. cerevisiae* populations infers reasonable levels of movement of individuals between regions in NZ, which is in line with the movement of fruit and equipment by the wine industry (Knight and Goddard, 2015). It may be that this genotype has recently been moved from the Marlborough region to Nelson, likely by humans’ activities.
Recently a handful of studies have shown that the communities and populations of microbes associated with vines and wines vary by region (Gayevskiy and Goddard, 2012; Bokulich et al., 2014; Taylor et al., 2014; Knight and Goddard, 2015), and these are the first demonstrations of geographic variance in microbes associated with agriculture generally. Here we conduct a crucial follow-on to these observations: to test whether the variance in microbes translates into altered crop phenotypes. Geographic variance in crop physical and sensorial signatures are well described, and have important economic and consumer preference consequences (Van Leeuwen and Seguin, 2006), but until now the drivers behind these differences have not been evaluated and quantified. These data show there is a quantifiable microbial aspect to terroir thus revealing the importance of microbial populations on the regional identity of wine, and may potentially extend to other important agricultural commodities. With a better understanding of the forces driving microbial population and community differentiation, food and agricultural sectors can develop systems to better control and manage these communities to help conserve the regional identity of products. More generally this finding indicates the importance of characterising and understanding biodiversity and the services it may provide. Together this suggests that the long-term implementation of methods that maintain biodiversity may have tangible economic imperatives as well as being driven by a desire to employ agricultural practices that increase responsible environmental stewardship.
Chapter 6

Fungal diversity during fermentation is indicative of thiol concentrations in wine with indications that microbes may modulate these throughout fruit development.

Status of chapter:

This chapter has not yet been submitted for publication. I am the lead author in collaboration with Dr Steffen Klaere, Peter Morrison-Whittle and Dr Matthew R. Goddard. The co-authorship form is presented at the start of this thesis, after the acknowledgements.
Abstract

Agricultural products produced by the same genotypic clone often have different physical and sensorial properties, influencing their overall quality and economic value. Microbes play key roles throughout the production of these goods, affecting plant and fruit health and modifying plant materials to produce socially and economically important commodities. Using wine as a model system, we investigate whether the fungal diversity in the initial grape juice and during fermentation is correlated with the final concentration of three volatile thiols important to wine aroma and flavour. The species of *Saccharomyces* yeast, the main driver of fermentation, was found to have a significant effect on the concentration of volatile thiols in the final wine, particularly 4MMP. We additionally used next-generation sequencing technologies to gain unprecedented insight into the fungal diversity in the initial grape juice revealing a complexity of OTUs important to the final concentration of thiols. Identification of main genera influencing this effect suggests that fungi not only modulate volatile thiol production during fermentation both directly by contributing metabolites and indirectly by altering the microbial community via antimicrobials, but may also be influencing the accumulation of odourless precursors in the grape itself via pathogenic effects during fruit ripening. This recapitulates the need for a better understanding of the interactions between microbial populations and agricultural products and has implications for the management of fungal diversity and disease in these systems.
Chapter 6: Fungal Diversity and Wine Thiol Concentration

6.1 INTRODUCTION

Microbes play key roles in the production of quality agricultural goods, affecting plant and fruit health and converting plant materials into socially and economically important commodities (Whipps, 2001; Fleet, 2006; Peiffer et al., 2013; Philippot et al., 2013). During the processing of plant materials such as grapes for winemaking, different species and strains of yeast are known to produce different concentrations of volatile compounds and thus contribute different sensorial properties to the final products (Howell et al., 2004; Viana et al., 2008; Anfang et al., 2009). In wine, many of these yeast-derived aromas and flavours produced during fermentation result from the conversion of odourless precursors in the grape must (Darriet et al., 1995; Tominaga et al., 1998a; Swiegers and Pretorius, 2005; Dubourdieu et al., 2006). Pathogenic fungi present in the vineyard on vines and grapes potentially alter concentrations of odourless precursors in the grapes of infected vines and bunches (Thibon et al., 2009; Barata et al., 2012) and thus can potentially affect the final flavour and aroma of a wine. Here we investigate the relationship between volatile thiol concentrations in wine and fungal species diversity in the grape juice and during fermentation to reveal the effects of microbial species diversity on this important agricultural commodity.

Spontaneous ferments of grape juice to wine are completed by a succession of indigenous yeasts that naturally occur on grapes and on winery surfaces and are transferred to the grape must (Pretorius, 2000; Xufre et al., 2006). During the early stages of a spontaneous ferment, a diversity of yeast species is observed with *Saccharomyces* species being very rare (Pretorius, 2000; Xufre et al., 2006; Goddard, 2008). As the ferment progresses, *Saccharomyces* species outcompete other hemiascomycete (non-*Saccharomyces*) yeasts by engineering its environment through the preferential fermentation of sugars in the must, even in the presence of oxygen, creating toxic ethanol and heat (Goddard, 2008). During this fermentative process, *Saccharomyces* species also produce a wide range of metabolites that have a positive influence on wine sensory attributes including volatile thiols (Lambrechts and Pretorius, 2000; Swiegers et al., 2006; Swiegers et al., 2009). Typically fermentation is performed by *S. cerevisiae*, although other
Chapter 6: Fungal Diversity and Wine Thiol Concentration

Saccharomyces species such as S. uvarum have been reported in large quantities in finished wine (Torriani et al., 1999; Naumov et al., 2000; Demuyter et al., 2004). Different species of Saccharomyces have been shown to produce different levels of volatile compounds during fermentation (Murat et al., 2001; Masneuf et al., 2002; Dubourdieu et al., 2006; Masneuf-Pomarède et al., 2010).

Non-Saccharomyces species present in the early stages of fermentation also contribute desirable sensory properties and complexity to the final wine (Romano et al., 2003; Clemente-Jimenez et al., 2005; Ciani et al., 2006; Hernández-Orte et al., 2008; Anfang et al., 2009; Comitini et al., 2011; Gobbi et al., 2013) although negative effects have also been reported (Comitini et al., 2011). Much work has been done to evaluate the contribution of particular non-Saccharomyces yeasts to wine composition, often using co-inoculation trials with S. cerevisiae, and different species have been shown to contribute different attributes (Ciani and Maccarelli, 1998; Romano et al., 2003a; Clemente-Jimenez et al., 2005; Hernández-Orte et al., 2008; Ciani et al., 2010; Comitini et al., 2011). While non-Saccharomyces yeasts appear to play an important role in the final flavour and aroma of a wine, Saccharomyces species are required to complete the ferment with different species and strains interacting to produce unique flavour and aroma profiles in the finished wine (Howell et al., 2006; Anfang et al., 2009; Sadoudi et al., 2012). Both non-Saccharomyces communities and S. cerevisiae populations have been shown to vary with geographic region (Bokulich et al., 2014; Taylor et al., 2014; Knight and Goddard, 2015) affording the potential for microbes to contribute to regionally distinct wine phenotypes. While tested for S. cerevisiae (Chapter 5), the differential effects of non-Saccharomyces communities on wine phenotype are uncharacterised.

Sauvignon Blanc is a major contributor to the New Zealand (NZ) wine industry comprising 85% of wine exports (New Zealand Winegrowers, 2013). The volatile thiols 4-mercapto-4-methylpentan-2-one (4MMP), 3-mercapto-hexan-1-ol (3MH) and 3-mercapto-hexan-1-ol acetate (3MHA) are important in Sauvignon Blanc aroma and flavour and are typically described as having box tree, passion fruit, black current bud,
Chapter 6: Fungal Diversity and Wine Thiol Concentration

broom, grapefruit and guava characteristics (Dubourdieu et al., 2006; Swiegers et al., 2009; Coetzee and du Toit, 2012). These compounds are highly potent with low sensory detection thresholds and are made by yeast during fermentation from odourless precursors in the grape must (Darriet et al., 1995; Tominaga et al., 1998a; Dubourdieu et al., 2006; Coetzee and du Toit, 2012).

Here we spontaneously ferment *Vitis vinifera* var. Sauvignon Blanc grape juice sourced from different geographic regions in NZ and correlate the resulting thiol concentrations with fungal diversity. Fungi present at the start and end of fermentation were quantified as well as the final concentration of 3MH, 3MHA and 4MMP. Results suggest the species of *Saccharomyces* driving fermentation has a significant effect on the production of 4MMP, while the genera of non-*Saccharomyces* that correlate with thiol concentration may alter the microbial community present by potentially excreting antimicrobials or are largely associated with vine and fruit health. This highlights that microbes not only affect the quality and sensory properties of the final wine during fermentation but may also modulate their potential development by affecting the surrounding microbial communities and vine and fruit health, having implications for the management of fungal biodiversity and disease.

6.2 METHODS

6.2.1 Sample collection and fermentation

Sauvignon Blanc juice was collected from 37 vineyards across NZ. Juices from six vineyards were collected from Hawke’s Bay, Martinborough, Nelson, the Awatere Valley and Central Otago and seven were collected from the Wairau Valley. A 50 mL sample of each juice was centrifuged at 3 000 rpm for 5 minutes to pellet the yeast cells. The supernatant was discarded and the yeast cells were frozen at -20 °C for further community sequencing analyses detailed below. Spontaneous ferments of each juice were performed at 15 °C in 10 L volumes. After 21 days, 50 mL samples were taken and again centrifuged at 3 000 rpm for
5 minutes. The supernatant was poured off and frozen for later chemical analyses and the yeast pellet was analysed for *Saccharomyces* species diversity explained below.

### 6.2.2 Fungal community analysis of the juice

The fungal community composition in the initial juice samples was quantified using Roche 454 next generation sequencing technology (Margulies et al., 2005). Total DNA was extracted from the juice samples using the Zymo Research Soil Microbe DNA MiniPrep™ kit. A 600 bp fragment of the D1/D2 26S ribosomal RNA locus, known to provide good signal for fungal community differentiation (Taylor et al., 2014), was amplified using the fungal specific primers NL1 and NL4 (Kurtzman and Robnett, 2003). Distinct multiplex identifiers were added to the primers to bioinformatically distinguish between samples. AmpureXP beads were used to clean the PCR products and remove primer dimers, with the final quality confirmed using Agilent DNA1000 chips. The samples were multiplexed and uni-directionally sequenced using a Roche 454 GS Junior sequencer.

Post-processing of the DNA sequencing data was performed using Mothur version 1.30 (Schloss et al., 2009). Low quality and erroneous sequences were removed starting with primers, low quality reads and reads smaller than 200 bp. Subsequently homopolymer errors as identified using the PyroNoise algorithm (Quince et al., 2009) and finally PCR chimeras identified using the UCHIME algorithm (Edgar et al., 2011) were removed. Individual sample identifiers were assigned to the remaining reads and the data was merged for further analyses. To further account for potential error in the dataset, unique sequences were compared to a fungal reference database and those not assigned to Fungi were removed. The remaining sequences were clustered into groups or operational taxonomic units (OTUs) sharing more than 98 % identity. Multiple species of Ascomycota and Basidiomycota (Fungi) have empirically been shown to differ by less than 2 % at the 26S rDNA gene (Kurtzman and Robnett, 2003; Romanelli et al., 2010) thus these
groups are considered to approximate species. Any OTUs that only contained one read were conservatively removed from further analyses as they are potentially errors.

6.2.3 *Saccharomyces* species analysis at the end of ferment

Yeast from the 21 day samples of the spontaneous ferments were plated in serial dilutions on to YPD agar (1 % yeast extract, 2 % peptone, 2 % glucose, 1.5 % agar with 50 µg/mL chloramphenicol to retard bacterial growth) and incubated at 28 °C for two days. 94 individual colonies from these plates were isolated and genomic DNA extracted with a 1.25 mg/mL zymolyase solution dissolved in 1.2 M Sorbitol and 0.1 M KH$_2$PO$_4$ at pH 7.2 and treated with EMA to bind unwanted DNA fragments (Rueckert and Morgan, 2007). DNA extractions were incubated at 37 °C for 30 minutes followed by 95 °C for 10 minutes. Using primers designed by de Melo Pereira et al. (2010) a multiplex PCR to distinguish between *S. cerevisiae* and *S. uvarum* was performed. Isolates were then scored as being *S. cerevisiae*, *S. uvarum* or unknown. Of the identified isolates, the proportion of *S. uvarum* in each sample was calculated and arcsine transformed to account for heterogeneous variance known to be associated with proportion data (Sokal and Rohlf, 1995).

It is important to note that since only two *Saccharomyces* species were identified in the ferment samples, the inverse of the proportion of *S. uvarum* is the proportion of *S. cerevisiae*.

6.2.4 Thiol analysis

Wine samples from the end of ferment (50 mL) were sent to Hill Laboratories Limited, Hamilton, NZ for chemical analysis. Quantitative analyses were performed for the volatile thiols 3-mercaptohexanol (3MH), 3-mercaptohexyl acetate (3MHA) and 4-mercapto-4-methylpentan-2-one (4MMP). Chemical compounds were extracted from the wine samples using Solid Phase Micro Extraction (SPME) and quantified in extracts using Gas Chromatography coupled with Mass Spectrometry (GCMS). 3MHA in wine is formed via the esterification of 3MH by an alcohol acetyltransferase produced by yeast and can be revert back to 3MH by
an esterase enzyme also produced by yeast (Swiegers et al., 2006). These two volatile thiols thus exist in wine in equilibrium and should be correlated with one another. Pearson’s product-moment correlation coefficient verifies this for this dataset with 3MH and 3MHA being significantly positively correlated ($r = 0.72$, $t_{34} = 6.10$, $P = 6.37 \times 10^{-7}$). Since these concentrations are not independent, the molar sum of 3MH and 3MHA is used for further analyses.

### 6.2.5 Statistical analysis

Due to the large number of OTUs identified in the starting juice a data reduction technique was required (i.e. there were more explanatory variables than there were samples). We used a partial least squares regression (PLSR) analysis using a matrix of the scaled OTU proportions as the explanatory variable and a matrix of the concentration of 4MMP and molar sum of 3MH and 3MHA as the response variable. This is similar to a Principle Component Analysis (PCA) that captures the variation in the X-variables; however PLSR also captures the information in the Y-variables and considers how the X-variables may be related to these (Mevik and Wehrens, 2007). More generally, the PLSR approach compresses the OTU’s into a manageable number of components that still describe the maximum amount of variation in the OTU proportions. The first six components from this analysis were used to avoid over fitting the model. These components along with the geographic region the juice was collected from (since previous research has suggested that wines from different regions have different concentrations of volatile thiols (Lund et al., 2009; Benkwitz et al., 2012)) and the scaled proportions of *S. uvarum* in the final ferment were used to build a linear model against both 4MMP concentration and the molar sum of 3MH and 3MHA independently. The components that significantly contributed to the variation in each volatile thiol were examined and OTUs that had a loading vector above 0.2 within these components were deemed to have a driving force on the thiol concentration in the wine. OTUs were identified to the genus level using a Bayesian approach implemented in Mothur version 1.30 (Schloss et al., 2009). Using the ‘classify.seqs’ command, a representative sequence from each OTU (i.e. a sequence that had a minimum distance to other sequences in the same cluster) was
compared to a fungal taxonomic database and classified to all levels including and above the genus level. Consensus sequences with less than a 70 % match at any taxonomic level were listed as unclassified.

To further investigate the effect of *Saccharomyces* species driving the ferment, Pearson’s product-moment correlations were performed between the scaled proportions of *S. uvarum* and the concentration of 4MMP and the molar sum of 3MH and 3MHA independently.

### 6.3 RESULTS

#### 6.3.1 Dataset

*Community diversity in the starting juice*

The 454-sequencing of the initial juice samples resulted in a total of 29 253 quality reads after processing. For each sample the number of quality reads range from 151 – 1372 and a total of 173 OTUs were defined (Appendix IV Dataset AIV.5). The raw counts of reads assigned to each OTU were converted into proportions for each sample to standardise for the variation in reads per sample and scaled to account for heterogeneity in variance, typical of proportion data. It was observed that in a number of samples the proportion of OTU002 was high (Appendix IV Dataset AIV.5). This OTU was identified as the genus *Saccharomyces* which is known to be rare in the early stages of fermentation we were attempting to sample (Goddard, 2008); therefore their high frequency indicates fermentation has begun and thus these samples represent a different time point. Sampling the same time point is important as *Saccharomyces* species are known to engineer their environment to exclude competing yeast species (Goddard, 2008) and their presence has likely altered the community composition in the sample. Therefore samples that had more than 10 % *Saccharomyces* (OTU002) were removed as their presence may have masked the initial starting fungal population and thus our ability to detect any contribution that population might have had on thiol concentrations. This reduced the dataset to 26 samples and 164 OTUs.
Chapter 6: Fungal Diversity and Wine Thiol Concentration

\[ ii. \quad \text{Saccharomyces species diversity at the end of ferment} \]

At the end of fermentation, two species of \textit{Saccharomyces} were identified; \textit{Saccharomyces cerevisiae} and \textit{S. uvarum}. These species were isolated in varying proportions in the final wine of each sample (Appendix IV Dataset AIV.5).

\[ iii. \quad \text{Thiol concentrations} \]

A wide range of concentrations were seen for all three volatile thiols quantified with 3MH reporting the highest concentrations up to 21 910 ng/L, followed by 3MHA then 4MMP (Appendix IV Dataset AIV.5). 4MMP was unable to be quantified in six samples as it was below the detection threshold of the analysis (<10 ng/L). For analytical purposes the midpoint of 5 ng/L was used for these samples. One sample from Martinborough (MARLI) reported concentrations of 3MH and 3MHA below the detection threshold of the analysis and 4MMP was very low at just 16 ng/L. As two of the three response variables were unable to be quantified, this sample contains a large portion of missing data and was subsequently removed from further analyses. The removal of this sample reduced the dataset to 25 samples and 162 OTUs. As reported above in Section 6.2.4, 3MH and 3MHA are correlated and therefore the molar sum of these two compounds is used for further analyses.

\[ 6.3.2 \quad \text{Data reduction of species diversity in the starting juice} \]

The first six components of the PLSR analysis explained 99.2 % and 88.9 % of the variation in 4MMP and the molar sum of 3MH and 3MHA respectively; however just 43.0 % of the variation in the OTUs were explained. As we are interested in explaining variation in the thiol concentrations, these six components are sufficient for our purposes. Each component consists of multiple OTUs with different magnitudes of effect. These are visually represented in Figure 6.1. The classifications of the OTUs with loading vectors larger than 0.2 are reported in Table 6.1.
Figure 6.1: A representation of the OTUs (points) comprising each component from the PLSR analysis. The large circles provide a scale for the magnitude of the loading vectors. The positions of the points represent the magnitude of the effect each OTU has on that component. OTUs with a magnitude larger than 0.2 are classified in Table 6.1.
Chapter 6: Fungal Diversity and Wine Thiol Concentration

**Table 6.1:** The OTUs with loading vectors above 0.2 in the PLSR analysis, the components they relate to and their classification

<table>
<thead>
<tr>
<th>OTU</th>
<th>Component</th>
<th>Classification to genus (if possible)</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>1</td>
<td><em>Alternaria</em></td>
</tr>
<tr>
<td>12</td>
<td>6</td>
<td>Unclassified genus; Phylum Basidiomycota</td>
</tr>
<tr>
<td>53</td>
<td>1</td>
<td>Unclassified genus; Phylum Ascomycota; Family Didymellaceae</td>
</tr>
<tr>
<td>76</td>
<td>3</td>
<td>Unclassified genus; Phylum Ascomycota; Order Dothideales</td>
</tr>
<tr>
<td>122</td>
<td>3</td>
<td>Unclassified genus; Phylum Ascomycota; Family Psathyrellaceae</td>
</tr>
<tr>
<td>128</td>
<td>3</td>
<td><em>Preussia</em></td>
</tr>
<tr>
<td>231</td>
<td>3</td>
<td>Unclassified genus; Phylum Ascomycota; Family Didymellaceae</td>
</tr>
<tr>
<td>850</td>
<td>3</td>
<td><em>Columnosphaeria</em>; (Order Dothideales)</td>
</tr>
<tr>
<td>2043</td>
<td>4</td>
<td>Unclassified fungi</td>
</tr>
<tr>
<td>2135</td>
<td>3</td>
<td><em>Preussia</em></td>
</tr>
<tr>
<td>3511</td>
<td>3</td>
<td>Unclassified genus; Phylum Ascomycota</td>
</tr>
</tbody>
</table>

### 6.3.3 The effects of species diversity on the concentration of 4MMP

The first six components from the PLSR analysis, the proportion of *S. uvarum* after fermentation, and the region the juice was sourced from, were used as factors to construct a linear model explaining the variation in 4MMP concentration. Overall the model explained 99.5% of the variation in 4MMP concentration ($R^2 = 0.995$, $F_{8,16} = 412.8$, $P < 2.2e-16$). The proportion of *S. uvarum* present at the end of fermentation significantly affects the concentration of 4MMP ($t_{16} = 3.16$, $P = 0.006$) as do all components except component five (Table 6.2). Of all the OTUs that comprise these components, eleven were observed to have a loading over 0.2 suggesting they are the main drivers behind the observed variation (Table 6.1). Each of these OTUs correspond to the genera *Alternaria, Preussia* and *Columnosphaeria* and seven remain unclassified to the genus level, although two were classified to the family Didymellaceae, one to the family Psathyrellaceae and one to the order Dothideales (Table 6.1). The region the samples were sourced from did not significantly affect the concentration of 4MMP (Table 6.2).
Table 6.2: The results from the linear model against 4MMP concentration with the factors of *Saccharomyces* diversity, region and the six principle components representing the OTU diversity in the initial juice samples. The model reported a residual error of 6.748 on 16 degrees of freedom and $F_{8,16} = 412.8$, $P < 2.2e-16$.

<table>
<thead>
<tr>
<th>Factor</th>
<th>t-value</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(intercept)</td>
<td>11.91</td>
<td>2.3e-9 ***</td>
</tr>
<tr>
<td>Region</td>
<td>-0.49</td>
<td>0.628</td>
</tr>
<tr>
<td><em>S. uvarum</em> proportion</td>
<td>3.16</td>
<td>0.0061 **</td>
</tr>
<tr>
<td>Component 1</td>
<td>-35.10</td>
<td>&lt;2e-16 ***</td>
</tr>
<tr>
<td>Component 2</td>
<td>-14.60</td>
<td>1.15e-10 ***</td>
</tr>
<tr>
<td>Component 3</td>
<td>-7.30</td>
<td>1.79e-6 ***</td>
</tr>
<tr>
<td>Component 4</td>
<td>-7.52</td>
<td>1.23e-6 ***</td>
</tr>
<tr>
<td>Component 5</td>
<td>-1.03</td>
<td>0.318</td>
</tr>
<tr>
<td>Component 6</td>
<td>-5.25</td>
<td>8.00e-5 ***</td>
</tr>
</tbody>
</table>

6.3.4 The effects of species diversity on the molar sum of 3MH and 3MHA

The same linear model was used to evaluate the factors affecting the molar sum of 3MH and 3MHA (Table 6.3). Overall this model explained 92 % of the variation in the molar sum of 3MH and 3MHA ($R^2 = 0.921$, $F_{8,16} = 23.33$, $P = 2.01e-7$). The proportion of *S. uvarum* present at the end of fermentation significantly affects the molar sum of 3MH and 3MHA as do components one, three, four and five (Table 6.3). These components comprise ten OTUs with loadings over 0.2 suggesting they are the main drivers of the differences observed. These correspond to the same genera as above for 4MMP, with the exception of one unidentified Basidiomycete (Table 6.1). The region the samples were sourced from does not significantly affect the molar sum of 3MH and 3MHA (Table 6.3).
Chapter 6: Fungal Diversity and Wine Thiol Concentration

Table 6.3: The results from the linear model against the molar sum of 3MH and 3MHA with the factors of *Saccharomyces* diversity, region and the six principle components representing the OTU diversity in the initial juice samples. The model reported a residual error of 9.24 on 16 degrees of freedom and $F_{8,16} = 23.33, P = 2.70e-7$.

<table>
<thead>
<tr>
<th>Factor</th>
<th>t-value</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(intercept)</td>
<td>2.84</td>
<td>0.0117 *</td>
</tr>
<tr>
<td>Region</td>
<td>0.13</td>
<td>0.901</td>
</tr>
<tr>
<td><em>S. uvarum</em> proportion</td>
<td>2.57</td>
<td>0.0207 *</td>
</tr>
<tr>
<td>Component 1</td>
<td>-3.25</td>
<td>0.0051 **</td>
</tr>
<tr>
<td>Component 2</td>
<td>0.06</td>
<td>0.952</td>
</tr>
<tr>
<td>Component 3</td>
<td>8.93</td>
<td>1.3e-7 ***</td>
</tr>
<tr>
<td>Component 4</td>
<td>-5.04</td>
<td>0.00012 ***</td>
</tr>
<tr>
<td>Component 5</td>
<td>4.16</td>
<td>0.00072 ***</td>
</tr>
<tr>
<td>Component 6</td>
<td>-0.33</td>
<td>0.750</td>
</tr>
</tbody>
</table>

6.3.5 A closer look at *Saccharomyces* species diversity

Analyses above indicate that the proportion of *S. uvarum* at the end of the ferment has an effect on final thiol concentrations. These analyses excluded samples that had begun to ferment before juice samples were taken as the presence of *Saccharomyces* species could have altered the community composition (Goddard, 2008). Since this does not affect the species of *Saccharomyces* sampled at the end of ferment, we performed additional analyses with all 36 samples testing the effect of *Saccharomyces* species composition on thiol concentration. The scaled proportions of *S. uvarum* are significantly positively correlated with the concentration of 4MMP ($r = 0.607$, $t_{34} = 4.46$, $P = 8.55 \times 10^{-5}$) suggesting that the higher the proportion of *S. uvarum*, the higher the concentration of 4MMP in the final wine. The molar sum of 3MH and 3MHA however is not significantly correlated with the proportion of *S. uvarum* at the end ferment ($r = 0.285$, $t_{34} = 1.74$, $P = 0.092$).
6.4 DISCUSSION

Microbes are vital for the production of quality agricultural commodities, affecting product quality throughout the development process (Whipps, 2001; Fleet, 2006; Peiffer et al., 2013; Philippot et al., 2013). Using wine as a model system, here we attempt to elucidate which yeast species potentially modulate the production of three important volatile thiols in Sauvignon Blanc. We reveal that different proportions of Saccharomyces species driving fermentation and differences in the fungal community in the starting juice significantly correlate with thiol concentration in the wine, highlighting the potential importance of microbes in the production of quality agricultural commodities.

Saccharomyces species are responsible for completing fermentation and different species and strains have been shown to produce different metabolites important to the final aroma and flavour of a wine (Masneuf et al., 2002; Howell et al., 2004; Dubourdieu et al., 2006). Here we provide evidence that higher proportions of S. uvarum at the end of fermentation correlate with higher concentrations of 4MMP. This is consistent with previous studies that report S. uvarum produces higher concentrations of 4MMP compared to S. cerevisiae and may be attributed to differences in the IRC7 gene (Masneuf et al., 2002; Dubourdieu et al., 2006; Roncoroni et al., 2011). IRC7 is necessary for 4MMP production by yeast and while it is largely functional in S. uvarum, it often is not in S. cerevisiae due to a 38 bp deletion and therefore may be responsible for higher levels of 4MMP observed when higher proportions of S. uvarum were detected (Masneuf et al., 2002; Dubourdieu et al., 2006; Roncoroni et al., 2011).

While some apparently benign non-Saccharomyces species have been shown to interact with various S. cerevisiae strains during fermentation to effect aroma production (Ciani et al., 2006; Viana et al., 2008; Anfang et al., 2009; Comitini et al., 2011), the species of fungi identified in this analysis as having a more pronounced effect on the final thiol concentration in wine are not known to actively grow in grape juice and therefore would not contribute metabolites during fermentation. Some fungi identified in this analysis
may affect or be indicative of plant and fruit health. *Alternaria* spp. on grapes are typically associated with grape bunch rot (Lorenzini and Zapparoli, 2014) and the family Didymellaceae consists of a range of plant pathogens including many representatives of the genus *Phoma*, a known cause of leaf and stem spots (Aveskamp et al., 2008; Zhang et al., 2009). Conversely, *Aureobasidium pullulans*, a member of the order Dothideales, is associated with healthy, unripe grapes (Fleet, 2003) and may therefore be an indication of a disease free vine. The three volatile thiols examined in these analyses are converted by yeast from odourless precursors in the juice. As pathogenic fungi present on vines and grapes may potentially alter precursor accumulation in grapes and thus the thiol potential of the must (Thibon et al., 2009; Barata et al., 2012), these results suggest that the modulation of thiol concentrations by these fungi is perhaps linked to their effects on vine and fruit health and thus grape and must composition rather than fermentative effects via metabolism in the grape must itself.

Alternatively, fungi may be affecting the microbial community via antimicrobials. The genus *Preussia* are predominantly coprophilous with some endophytes reported (Mapperson et al., 2014); however, several species of *Preussia* secrete antimicrobial preussomerins and thus have the potential to modulate the surrounding microbial community in the vineyard that may interact with, and affect, the vine and fruit itself, as well as microbes actively growing in the grape must. Many members of *Psathyrella*, a genus within the family Psathyrellaceae, also display antimicrobial activity (Suay et al., 2000) and therefore also have the potential to alter the microbial community both in the vineyard and the grape must. These interactions may indirectly have an effect on the final thiol concentration of wine by altering the microbial community affecting vine and fruit health and thus thiol precursor accumulation, as well as the microbial community actively metabolising and producing these compounds during fermentation. These inferences regarding the effects of non-*Saccharomyces* genera present in grape juice are crude and require further experimental validation as there is a degree uncertainty around the identification and ecological function of the genera reported here. Not only is information about the identified taxonomic classes sparse, but the taxonomic
As an alternate explanation, the fungal community in the starting juice may have been correlated with the
ripeness of the grapes when harvested. Aroma compounds are accumulated by grapes throughout the
ripening process (González-Barreiro et al., 2015), and if fungal species diversity also changes during fruit
ripening, the patterns observed in this analysis may reflect the ripeness of the grapes at harvest, as
represented by the fungal community, rather than any affects imparted by the fungal community itself.
Grape berries have been shown to harbour different yeast communities at different stages of their
development with unripe berries reported to have a predominance of *Rhodotorula*, *Cryptococcus* and
*Candida* species, along with the yeast-like fungus *Aureobasidium pullulans*, whereas ripe berries
additionally harbour *Hanseniaspora* and *Metschnikowia* species (Fleet, 2003). Statistical analyses suggest
differences in the fungal communities is apparent between grapes at the start of berry ripening and those
that are over ripe; however at stages potentially experienced around harvest (berries not quite ripe
through to over-ripe), no consistent differences in fungal diversity was detected (Martins et al., 2014). As
these are the extremes of the grape ripening stages that our juice was obtained from, this effect should not
have confounded our results.

These findings highlight that the final quality and aroma profile of wine may potentially be influenced by
fungi throughout the growing season and not just during the fermentation process itself. This recapitulates
the importance of an integrated approach to the study of agricultural phenotypes and quality
characteristics and verifies our need for a better understanding of microbial ecology throughout the
entirety of these systems.
Our knowledge of the ecology and population patterns of microbes is beginning to expand with the application of ecological theory to these ecologically, scientifically and commercially important species (Martiny et al., 2006; Prosser et al., 2007; Hanson et al., 2012; Soininen, 2012; Barberan et al., 2014). This thesis significantly adds to this knowledge base and fulfils its main objective of providing a deeper understanding of the ecology and population patterns of *S. cerevisiae* and how these patterns, in combination with yeast community dissimilarity, contribute to wine styles in NZ.

To summarise, a comprehensive analysis and quantification of the population processes occurring in the New Zealand population of *S. cerevisiae* is provided revealing a complex picture of varying degrees of population subdivision and connection. The ecology of *S. cerevisiae* in the ‘woodland’ niche is investigated and evidence that this species resides in the soil niche in a sporulated state is provided. Microbes’ contribution to the regional distinctiveness of agricultural products, specifically wine, is experimentally tested and quantified, confirming these populations play important roles in differential agricultural geographic phenotypes. Finally the effect of different fungal species and communities on the final thiol concentration in wine is investigated, revealing a correlation between thiol concentration and both the species of *Saccharomyces* driving the ferment and complex communities of fungi in the starting juice that may be indicative of plant and fruit health at the time of harvest. This implies that fungi not only modulate
the final aroma and flavour of a wine during fermentation but also during fruit development by affecting
vine and fruit health.

Here I summarise the key findings of this thesis as they relate to each objective outlined in the General
Introduction (Section 1.4), and discuss potential avenues for future research.

7.1 RESEARCH OUTCOMES

7.1.1 Objective 1: Quantify the degree to which New Zealand’s *S. cerevisiae* metapopulation is
genetically structured by both environment and geography and go on to estimate the
extent to which sub-populations are connected by gene-flow (Chapter 3)

In Chapter 3 I reported results from genetic analyses detailing the population patterns of *S. cerevisiae* in NZ.
These results provide interesting insight into the population biology of this species and yield many key
findings. Firstly, while *S. cerevisiae* has been isolated from exotic *Quercus* species in New Zealand and
*Nothofagus* in Patagonia (Zhang et al., 2010; Libkind et al., 2011), this is the first report of *S. cerevisiae* from
multiple native tree species in the South Pacific region, providing evidence of a wider potential ecological
range for this species.

Within geographic regions, the populations of *S. cerevisiae* residing in native, vineyard and spontaneous
ferment environments were shown to be undifferentiated (i.e. they show no significant genetic
differentiation by the niche of isolation). This is in line with previous studies on the same geographic scale
from New Zealand and the United States of America (Goddard et al., 2010; Hyma and Fay, 2013), but in
conflict with global studies that suggest differentiation by ecological niche (Fay and Benavides, 2005; Legras
et al., 2007). As discussed in Section 3.4, these global studies evaluate fewer strains from a diverse range of
niches, which are often confounded with geography. Alternatively it may be that *S. cerevisiae* was
introduced to NZ relatively recently and has subsequently radiated to all niches but not yet had sufficient
time to genetically diverge. Despite this, evidence shows that populations of *S. cerevisiae* residing in different niches within geographic regions are connected and comprise one homogenous regional population.

At distances greater than ~100 km, the NZ *S. cerevisiae* population exhibits varying degrees of population structure or subdivision. Overall, analyses quantified that 16 % or one sixth ($P < 0.0001$) of the observed genetic variation in the data is explained by geographic region alone with Nelson and Central Otago more genetically distinct, Hawke’s Bay and Marlborough more homogeneous, and Martinborough intermediate. These results are consistent with previous studies investigating the population structure of *S. cerevisiae* across a range of scales from inter-continental to finer regional scales that all report varying degrees of genetic differentiation with the presence of hybrid or mosaic strains, indicative of some gene-flow or migration occurring between populations (Fay and Benavides, 2005; Schuller et al., 2005; Aa et al., 2006; Lopandic et al., 2008; Liti et al., 2009; Goddard et al., 2010; Mercado et al., 2011; Di Maio et al., 2012; Gayevskiy and Goddard, 2012; Wang et al., 2012).

This is complemented with estimates of bidirectional patterns of migration between different geographic regions in NZ. These mirror the transportation of fruit by the NZ wine industry, suggesting human associated dispersal over these distances. Hawke’s Bay and Marlborough, NZ’s two largest viticultural and winemaking regions, often import grapes from other regions for fermentation. This pattern aligns with the migration estimates for these two regions in that they experience a net gain of migrants with more individuals entering these regions than leaving. The opposite is observed in the smaller wine producing regions of Nelson and Central Otago that experience a net loss of migrants with more individuals leaving these regions than entering (Figure 3.1). This is the first attempt to quantify migration rates in *S. cerevisiae* and the alignment of these estimates with the movements of fruit by the wine industry suggests anthropogenic activities affect microbial population patterns and diversity.


Chapter 7: General Discussion and Future Directions

7.1.2 Objective 2: Investigate the role of the ‘woodland’ niche in the life cycle of *S. cerevisiae* (Chapter 4)

The ecology of *S. cerevisiae* outside of the ephemeral fruit niche is poorly understood. There are many reports of *S. cerevisiae* being isolated from soil and bark of and around fruiting tree species and oak (Goddard et al., 2010; Wang et al., 2012; Hyma and Fay, 2013); however, what *S. cerevisiae* is doing in these ‘woodland’ niches is not known. In Chapter 4 I hypothesised that *S. cerevisiae* resides in the soil niche in a sporulated state. Using a soil agar media, I tested the sporulation efficiency of *S. cerevisiae* genotypes originally isolated from soil and spontaneous fermenters and recorded the frequency of both four-spored (tetrad) and two-spored (dyad) asci. Firstly, further evidence for connectivity between the soil and ferment niches was established with no differences observed in sporulation efficiency or, of the sporulating cells, the proportion of tetrads formed between strains originally isolated from either niche. This suggests these strains are not differentially adapted to either of these niches, at least in terms of their ability to sporulate, and comprise one connected population.

For the first time, experimental evidence that *S. cerevisiae* sporulates in the presence of soil nutrients and does so at a significantly higher rate than controls is provided. Additionally, of the sporulated cells, a higher proportion of tetrads were observed on the soil agar compared to the plain control agar. Both of these results culminate in the samples on soil agar having a higher number of potentially viable spores and thus potentially greater fitness compared to the controls. Since all genotypes were able to form tetrads to some extent, the formation of dyads was not considered to result from a genetic mutation; however, the formation of nonsister dyads (NSDs) can be elicited by a metabolic response to a depletion of the carbon source after commitment to sporulation (Davidow et al., 1980; Neiman, 2005; Taxis et al., 2005). As explained in Chapter 4 (Section 4.4), these NSDs encapsulate the maximum genetic diversity of the mother cell, potentially providing an advantage as heterozygosity is associated with fitness benefits upon germination (Taxis et al., 2005). This however, should not be preferred over forming a 4-spored ascus (tetrad) as this would contain the same amount of genetic diversity but with twice the number of viable
cells as a NSD, increasing the mother cells fitness through an increased likelihood that one or more of those spores will survive to germination.

7.1.3 Objective 3: Quantify the contribution of regional populations of *S. cerevisiae* to terroir or regional distinctiveness of wine (Chapter 5)

In Chapter 5 I performed the first experimental test of whether there is a microbial component to the concept of terroir or sense of place of wine. Using a combination of multivariate and univariate statistics, isolates of *S. cerevisiae* sourced from different geographic regions were found to produce significantly different chemical profiles in the final wine. Of the total variation in the chemical profiles, 10 % could be attributed to the geographic region the isolates were sourced from. Central Otago isolates produced the most distinct chemical profiles while the isolates from Hawke’s Bay and the Marlborough sub-regions of the Wairau and Awatere Valley’s produced the least distinct chemical profiles. The chemical compounds mainly responsible for the observed differences between regions were not from any one chemical class and without sensory trials it is not possible to draw conclusions on the perceived differences in aroma and flavour between the wines.

The variation in the chemical profiles produced was found to significantly correlate with the genetic distance of the strains tested ($R^2 = 0.189; P < 0.0001$). Additionally, the patterns of chemical profile differentiation between regions largely mirror the patterns of genetic differentiation from all samples reported in the first objective (Chapter 3). Thus the strains selected from regions with more genetically differentiated populations produced more chemically differentiated profiles in the final wine (e.g. Central Otago) and vice versa (e.g. Hawke’s Bay and Marlborough). The one exception is Nelson which was found to be more genetically differentiated but failed to show strong chemical profile differentiation in the wines produced. Upon closer examination of the chemical profiles, it was noted that this was largely caused by one genotype that was considered genetically admixed in population genetic analyses and shared ancestry.
with strains isolated from other geographic regions. When this strain was removed from the full analysis, the total variation attributed to the geographic region of isolation increase by 2% to 12%. In fact when all tested genotypes with mixed ancestry were removed from the analysis the variation in the chemical profiles attributed to geographic region of isolation doubled to 20%. This finding is unsurprising since it is known that there is a genetic basis to volatile compound production in yeast (Murat et al., 2001; Lopandic et al., 2007; Swiegers et al., 2009).

Wine co-fermented with the six genotypes selected from each region and blends of wine from the same genotypes fermented in isolation were significantly different ($R^2 = 0.061$, $P = 0.014$), confirming a metabolic interaction between *S. cerevisiae* genotypes during fermentation (Howell et al., 2006). No significant regional distinctions between the co-fermented wines could be detected, although statistical power was low with only three replicates per region.

### 7.1.4 Objective 4: Investigate whether the fungal community in grape juice and the species of *Saccharomyces* driving the ferment affect thiol concentrations in wine (Chapter 6)

Chapter 6 revealed some interesting findings regarding the impact of microbes on thiol concentrations in wine. Firstly, a complex consortium of species was revealed to correlate with 4MMP and the molar sum of 3MH and 3MHA in Sauvignon Blanc wine. The fungal genera identified as being the main drivers of these patterns potentially affect the composition of the microbial community via antimicrobials and vine and fruit health rather than contributing metabolites to the fermentation process itself. Grape damage associated with microbes such as grape rots and mildews affect fruit composition in a multitude of ways, including the production of volatile thiol precursors (Barata et al., 2012). This would have a direct impact on the final aroma and flavour of a wine and highlights the importance of microbial species throughout the entire grape-growing season with potentially very real economic implications.
Chapter 7: General Discussion and Future Directions

The species of *Saccharomyces* driving the ferment had a significant effect on final thiol concentrations. While this effect was seen for both 4MMP and the molar sum of 3MH and 3MHA in the full linear models, subsequent correlations between *S. uvarum* proportion and thiol concentration revealed the effect to be stronger for 4MMP. A positive correlation was revealed with higher proportions of *S. uvarum* compared to *S. cerevisiae* being associated with higher concentrations of 4MMP. This is consistent with previous research suggesting *S. uvarum* isolates commonly harbour a functional copy of the IRC7 gene, known to contribute to 4MMP production, while *S. cerevisiae* isolates often carry a non-functional copy, affecting their ability to produce 4MMP (Masneuf et al., 2002; Roncoroni et al., 2011).

7.2 SYNTHESIS

While the ecology and population patterns of microbes in general are poorly understood, this thesis begins to reveal their intricacies in the model research eukaryote, *Saccharomyces cerevisiae*. The NZ population of *S. cerevisiae* exhibits a complex pattern of varying degrees of population differentiation and migration. No genetic differentiation between individuals isolated from different ecological niches is evident below scales of ~100 km showing homogenous populations within regions. Connectivity between the soil and ferment niches is also reflected in the phenotype data of sporulation efficiency. No detectable difference in sporulation efficiency was observed between individuals originally sampled from soil or spontaneous ferment niches. If these populations were physically isolated from each other, selection and genetic drift would result in phenotypic differentiation. Thus the finding of no phenotypic differentiation, at least in terms of their sporulation efficiency, also suggests these populations are homogenous. These, in combination with soil agar inducing a sporulation response, are important findings for the hypothesis that *S. cerevisiae* cycles between the fruit and ‘woodland’ niches, persisting in the ‘woodland’ niche in a sporulated state.
On larger scales between geographic regions, population genetic analyses report varying degrees of differentiation and migration that reflect the movement of fruit by the wine industry for fermentation and bottling. This suggests anthropogenic activities may have an effect on microbial population diversity and dispersal. These population patterns may also affect agricultural commodities as more genetically differentiated populations of *S. cerevisiae* were found to impart more distinctive chemical profiles in wine in the first objective experimental test for a microbial aspect to *terroir* or regional distinctiveness of an agricultural product. Furthermore, not only do the species of microbes involved in fermentation have an effect on the final chemical composition of a wine, but also those associated with vine and fruit health, suggesting that agricultural practices that maintain differential microbial diversity may have direct economic implications.

### 7.3 Future Directions

Several potential avenues for future research could be pursued upon the findings of this thesis to investigate the population patterns and ecology of microbial species and how they affect economic imperatives. These are broken down into three main categories: 1) Population patterns, 2) Ecology, and 3) Agriculture and winemaking.

#### 7.3.1 Population patterns

As discussed in Section 2.1.2, microsatellite loci were used in this thesis to genotype *S. cerevisiae* isolates. Genotyping using microsatellites is fast, cost-effective and useful for estimating patterns of population structure and gene-flow; however it provides limited information on a small set of variable loci and is unable to provide information on the forces driving the observed patterns. Genotyping-by-sequencing methods such as RADSeq (Davey and Blaxter, 2010) are alternate methods that utilise next-generation sequencing technologies. They sequence targeted areas across the genome, greatly increasing the number
of markers assayed and in turn improve the precision of population structure and gene-flow estimates (Narum et al., 2013). The vast numbers of polymorphisms that are screened are distributed across the genome and are subject to variations in evolutionary forces such as selection and drift. This affords the power to infer which regions of the genome are subject to selection and adaptive evolution (Davey and Blaxter, 2010; Narum et al., 2013). By identifying these regions, inferences about the forces and selective pressures driving the observed population patterns can be made. Genotyping-by-sequencing methods are becoming increasingly more cost-effective and are now feasible for a large number of individuals with small genomes such as *S. cerevisiae* (genome size of approximately 12 Mb (Goffeau et al., 1996)). Therefore it would be interesting to perform a similar study on the population patterns of *S. cerevisiae* using these technologies such as RADSeq (Davey and Blaxter, 2010) to not only obtain more accurate estimates of population differentiation and gene-flow between *S. cerevisiae* populations but also detailed information on regions of the genome under selection. This would yield a more comprehensive picture of the patterns of genetic diversity in this model microbial species as well as provide valuable information on the forces driving these differences, having consequences for our understanding of the evolutionary forces shaping microbial populations.

The patterns of population differentiation and migration for *S. cerevisiae* reported in this thesis were only elucidated for one season and their stability needs to be verified over time. Previous research investigating the community composition of grape-surface microbiota between two years, 2010 and 2012, found only a weak or insignificant effect of time on the fungal communities across regional scales, but differences between years at individual sites (Bokulich et al., 2014). This suggests there is a degree of stability in microbial communities over time, at least at larger regional scales; however the effect of time or season on the genetic diversity and population patterns in a single microbial species is not well understood. Ultimately it would be interesting to repeat the sampling scheme applied in Chapter 3 of this thesis over consecutive years to investigate whether the regional population patterns observed hold over time; however this would be costly and time consuming given the scale. During the course of this research an attempt was made to
track a *S. cerevisiae* population in a single vineyard across two successive seasons using the same sampling methods as the larger population genetic analyses in Chapter 3. Vineyard samples of soil, bark and fruit (when present) were taken at three fixed locations within a single vineyard every four weeks and enriched for *S. cerevisiae*. At harvest additional samples of juice from that vineyard were taken from the winery settling tank and spontaneously fermented. Unfortunately the vineyard was destroyed part way through this study and the full time course was not completed; however, consistent with *S. cerevisiae* being rare in environments other than fermenting fruit (Mortimer and Polsinelli, 1999; Pretorius, 2000; Xufre et al., 2006; Goddard, 2008; Taylor et al., 2014), very few samples yielded any *S. cerevisiae* isolates. Therefore the part of the time course that was completed had a large amount of missing data and may not have had enough power to truly address the question of whether populations of *S. cerevisiae* are genetically stable over time. Therefore, I suggest a different approach. Since the research reported in this thesis shows *S. cerevisiae* populations in the vineyard and surrounding environment are connected to those in spontaneous ferments (Knight and Goddard, 2015), sampling effort should be concentrated on spontaneous ferment samples as these harboured the highest frequency and largest genetic diversity of *S. cerevisiae* isolates. For statistical purposes replication is required so samples of juice for spontaneous fermentation should be taken from multiple vineyards sites. To control for population subdivision between different regions and potentially between different grape varieties these sites should be within ~100 km from each other and from the same grape variety (Gayevskiy and Goddard, 2012; Knight and Goddard, 2015). Sampling would be repeated across consecutive years and isolates of *S. cerevisiae* obtained from these samples would be genotyped and statistically analysed for differences between years. This would be repeated over a minimum of three years to ensure any differences (or lack of) in the patterns observed between consecutive years were not due to chance.

Previous research suggests the NZ population of *S. cerevisiae* is relatively isolated from other geographic locations (Goddard et al., 2010; Cromie et al., 2013); however if it is assumed this species was brought to
NZ with the introduction of vines and winemaking equipment it may not have been established long
(Halliday, 1991). Therefore the population patterns that were observed here may be more indicative of a
young, expanding population rather than an older, well established population. While other aspects of the
data do not offer support for this explanation (discussed in Section 3.4) it cannot be discounted entirely and
it is likely that population divergence is occurring to some extent. To further investigate the potential
effects human activity may have on the population patterns and genetic diversity of *S. cerevisiae*, I believe
it would be interesting to see if the same patterns of genetic differentiation with potential human
associated gene-flow hold for older, well established wine making areas of the world. It could then be
assumed that the populations are more stable and not undergoing rapid divergence, providing greater
assurances that the genetic signal observed is a result of migration. For migration to be correlated with
human activities the regions tested must also behave in a similar manner as the NZ wine industry and
transport fruit between regions for fermentation and bottling. If this is not occurring, it would be
interesting to see if the extent of migration across larger geographic regions diminishes. Additionally
whether these patterns hold on a larger, more global scale is of interest. Barrels have been reported to
contain live *S. cerevisiae* cells that could potentially become resident in the local population (Goddard et al.,
2010). Therefore, countries that actively import winery equipment and barrels from certain regions of the
world may harbour more genetically similar and connected populations. These studies would need to
ensure adequate levels of replication to allow for appropriate statistical analyses.

### 7.3.2 Ecology

Microbes play key ecosystem roles and are vital to the functioning of the biosphere (Kirchman, 2012).
Understanding the ecology of microbial species is not only fundamentally important but has implications
for the way we sustainably manage ecosystems while maintaining essential resources, particularly in the
face of climate change (Prosser et al., 2007). The application of theory is essential for advancements in
microbial ecology, providing predictions that can be experimentally tested. While ecological theory for
larger plants and animals is well established, it is largely unknown whether these extend to microbes (Prosser et al., 2007). Microbes provide a powerful tool to investigate the effect of spatial scale on ecological processes due to their small size but large distribution (Jessup et al., 2004). The inherent difficulties in observing microbes has largely inhibited the development of and testability of ecological theory with microbes; however, with the development of new cultivation-independent techniques, some of these barriers are now being overcome (Prosser et al., 2007; Barberan et al., 2014).

Understanding how an organism behaves in its natural environment is essential for inferring ecological processes and potential inter- and intra-species interactions. In a high sugar environment such as fruit, S. cerevisiae is known to engineer its environment via fermentation by producing ethanol and heat to eliminate competitors and gain exclusive access to this energy rich resource (Goddard, 2008). Fruit, however, is ephemeral and S. cerevisiae has been found to also reside in ‘woodland’ niches such as soil and bark (Goddard et al., 2010; Wang et al., 2012; Hyma and Fay, 2013; Knight and Goddard, 2015). How S. cerevisiae behaves and interacts with other organisms in these environments is largely unknown. It was proposed in Chapter 4 that S. cerevisiae resides in these niches in a sporulated state and that it is not actively growing or dividing. Preliminary evidence is put forward in support of this with sporulation observed in multiple strains of S. cerevisiae on a soil agar media; however this is just the first step in many towards understanding S. cerevisiae’s behaviour in the ‘woodland’ niche and much investigation is required to confirm this hypothesis. Firstly, to truly demonstrate that S. cerevisiae resides in a sporulated state in the ‘woodland’ niche, I would suggest attempting to observe cells in this environment using microscopy techniques. S. cerevisiae is rare in environments other than fermenting fruit (Mortimer and Polsinelli, 1999; Pretorius, 2000; Xufre et al., 2006; Goddard, 2008; Taylor et al., 2014) and thus the success rate of finding cells in the ‘woodland’ niche is extremely low, making it unfeasible to do this for environmental samples. I suggest inoculating particles of soil or fragments of bark with a sufficiently large number of S. cerevisiae cells and leaving them on this substrate for approximately five to ten days. This should allow sufficient time for S. cerevisiae to respond to the new environment. These samples should then be prepared and imaged.
with an instrument such as an electron microscope to directly visualise the state of the cells in this niche. If sporulated cells are observed, this would add to the evidence that *S. cerevisiae* resides in a sporulated state in the ‘woodland’ niche and is largely dormant overwinter.

A second potential approach to investigate what *S. cerevisiae* is doing in ‘woodland’ niches would utilise high-throughput next-generation sequencing technologies. RNA-Seq provides DNA sequence of the entire transcriptome and allows quantitative analysis of gene expression levels (Nagalakshmi et al., 2008). I propose using this method to determine what genes are actively expressed and to what extent in *S. cerevisiae* inoculated in soil or on bark as representative alternate ‘woodland’ niches and compare these to juice or fruit samples to identify any differences. This would clarify whether *S. cerevisiae* is actively growing and metabolising in these niches, in which case genes involved in metabolic activity would be active, or if it is undergoing meiosis and sporulation, in which case genes essential to the meiotic cycle would be active. Understanding how *S. cerevisiae* behaves in its natural environment is important for inferring potential intra- and inter-species interactions. These potential interactions have implications for ecological processes such as competition, predation, parasitism and mutualism which are important in structuring communities (Jessup et al., 2004). It also has implications for the number of generations and the modes of reproduction experienced by this species during different seasons, having consequences for estimates of population patterns and processes and thus our understanding of the evolutionary history and future trajectories of this species.

### 7.3.3 Agriculture and winemaking

In Chapter 5 a microbial aspect to *terroir* was experimentally confirmed using *S. cerevisiae* and wine fermentation. In a fermentative sense, *terroir* must originate from microbes involved in the spontaneous fermentation of wine as inoculating with a commercial strain would mask any affect from the regional yeast population. Spontaneous fermentations are typically performed by a succession of yeast species and a
multitude of different strains that metabolically interact, creating vastly different aroma and flavour profiles compared to inoculated wines (Howell et al., 2006; Goddard, 2008; Anfang et al., 2009). As a first step this thesis investigated the interaction between *S. cerevisiae* genotypes on the regional distinctiveness or *terroir* of a wine as part of Chapter 5. Wines were fermented by a combination of six genotypes from each geographic region and the resulting chemical profiles were tested for differentiation by region. No significant difference between the chemical profiles by region was detected, although this *P*-value was marginal (*P* = 0.073). As interactions between species will not happen in exactly the same fashion each time and would thus be expected to increase the variance between replicates, it is possible that the effect of region may have been relatively small and could not be detected by only three replicates. I would therefore recommend repeating this experiment with at least double the number of replicates to evaluate the confidence in the results obtained. In addition, further data could be collected from this experiment to begin to unravel the interactions between these strains in an ecological sense. The abundance of each genotype could be determined at the end of fermentation by genotyping, indicating whether one or a small number of genotypes were dominant in the ferment, or if all strains were able to exist and ferment in sympatry. This is not only interesting in terms of determining which strains contributed more to the final aroma and flavour of a wine, but also affords the opportunity to examine the ecological processes involved in the partitioning of resources and structuring of microbial communities.

As suggested in Chapter 6, microbes may alter the concentration of volatile thiols in a wine by affecting vine and fruit health during fruit ripening; however this analysis was performed on a diverse dataset of samples from different geographic regions and wine producers with few controls (discussed in Section 2.4.1). To better understand how microbes associated with the vine during fruit development affect the final aroma and flavour of wine, I propose tracking the microbial communities associated with a particular vine throughout the growing season. A comprehensive vine microbiome could be established for each vine by quantifying the microbial communities associated with the soil and roots as well as the fruit itself. Next-generation sequencing technologies allow these comprehensive studies to be performed, providing a
powerful tool to quantify the microbial community associated with a variety of different substrates. Controls for abiotic factors also thought to influence grape composition would need to be imposed during vine selection to avoid confounding the results. The simplest way to do this would be to select replicate vines from the same block as these should experience the same temperature, rainfall and sunshine hours and have similar soil types; although if this is not possible these factors should be accounted for in the analysis as random effects. The same viticultural practices should also be used on all vines to ensure all vines encounter the same growing conditions. When ripe, the fruit from each vine would need to be pressed and fermented independently with the same strain of \textit{S. cerevisiae} to control for differences produced during fermentation. In combination with the abiotic and vineyard management controls imposed during vine selection, this allows the chemical composition of the wine at the end of fermentation to be attributed to the vine microbiome during fruit development. The focus of the chemical analyses of the wine would not only target volatile compounds converted from precursors in the juice by yeast metabolism, by also grape-derived compounds such as methoxypyrazines (Coetzee and du Toit, 2012). Identifying microbes that impact the final flavour and aroma of wine via interactions with the vine itself could have implications for the management of microbial diversity or alternatively, disease in vineyards. This comprehensive analysis could also provide information on microbial interactions with vines and generate new, testable hypotheses investigating how microbial biodiversity impacts agricultural products generally.

7.4 CONCLUSIONS

The research presented in this thesis addresses both fundamentally and commercially important concepts and adds a significant body of knowledge to these fields. It firstly provides a significant step forward in our understanding of the ecology and population patterns in the fundamental research eukaryote \textit{Saccharomyces cerevisiae}. By better understanding and quantifying the forces operating in and between microbial populations, a more comprehensive and inclusive framework fully integrating ecological and
Chapter 7: General Discussion and Future Directions

evolutionary processes can start to be established. Secondly this thesis reveals the importance of microbial populations and communities on the regional identity, quality and sensorial properties of wine and potentially agricultural commodities in general. This suggests that long-term implementation of agricultural practices that maintain differential microbial diversity while managing disease could have direct economic implications as well as being desirable in terms of employing agricultural practices that increase responsible environmental stewardship.
References


de Melo Pereira, G. V., Ramos, C. L., Galvão, C., Souza Dias, E. and Schwan, R. F. (2010). Use of specific PCR primers to identify three important industrial species of *Saccharomyces* genus: *Saccharomyces cerevisiae*, *Saccharomyces bayanus* and *Saccharomyces pastorianus*. Letters in Applied Microbiology 51(2): 131-137.


140


Sampaio, J. P. and Gonçalves, P. (2008). Natural populations of *Saccharomyces kudriavzevii* in Portugal are associated with Oak bark and are sympatric with *S. cerevisiae* and *S. paradoxus*. Applied and Environmental Microbiology 74(7): 2144-2152.


Snigowski, P. D., Dombrowski, P. G. and Fingerman, E. (2002). *Saccharomyces cerevisiae* and *Saccharomyces paradoxus* coexist in a natural woodland site in North America and display different levels of reproductive isolation from European conspecifics. FEMS Yeast Research 1(4): 299-306.


eq 1/(4N_{m}+1)$. Heredity 82(2): 117-125.


APPENDICES

APPENDIX I: Sample codes and locations

All sample codes follow the same format and include information on the geographic location, vineyard company, specific vineyard site and sample type. Some sample codes also contain an isolate number which refers to that isolates position in that samples 96 well plate.

Here’s an example for a vineyard sample:

```
ACNCCsf_C3
```

Region  
Vineyard Company  
Vineyard Block  
Sample Type  
Isolate number

Here’s an example for a native sample:

```
UMBMfrR_A1
```

Specifies Native Sample  
Region  
Tree Species  
Sample Type  
Isolate number

A list of the sample codes identifying what company and block they belong to is provided in Appendix II Tables S1 and S2 and is indicated on the following maps of the sample sites.
Figure A1.1: Map of New Zealand giving an indication of the scale between the geographic regions of sampling.
Figure A1.2: A map of the sampling locations in Hawke’s Bay with sample codes. The boxes with a blue background indicate native sampling sites.
Figure AI.3: A map of the sampling locations in Martinborough with sample codes. The boxes with a blue background indicate native sampling sites.
Figure A1.4: A map of the sampling locations in Nelson with sample codes. The boxes with a blue background indicate native sampling sites.
Figure A1.5: A map of the sampling locations in the Wairau Valley with sample codes. The boxes with a blue background indicate native sampling sites.
Figure A1.6: A map of the sampling locations in the Awatere Valley with sample codes. The boxes with a blue background indicate native sampling sites.
Figure AI.7: A map of the sampling locations in Central Otago with sample codes. The boxes with a blue background indicate native sampling sites.
Figure AII.1: Principal Coordinates Analysis investigating the genetic distance between individuals isolated from spontaneous ferment samples (blue) and environmental samples (red) encompassing vineyard soil and soil and fruit from native trees.
Figure AII.2: DISTRACT plots. Vertical bars represent individuals and the colours denote inferred populations. Individuals to the left of the black line were isolated from spontaneous ferments and individuals to the right were isolated from vineyard soil and soil and fruit from native trees species. $R^2$ values and $P$-values were calculated using ObStruct.
Figure AII.3: Plots from the canonical discriminant analysis performed in ObStruct. (a) Mapping the median and 66% ellipse for each sampled population. Populations with higher degrees of overlap are more genetically similar. (b) Mapping the individual data from all 369 isolates analysed. The distance between points is indicative of the similarity in ancestry profiles.
Table AII.1: Further details about the samples obtained from Sauvignon Blanc vineyard sites across New Zealand. NG refers to no growth meaning the sample did not yield any yeast isolates.

<table>
<thead>
<tr>
<th>Region</th>
<th>Company</th>
<th>Vineyard</th>
<th>Niche</th>
<th>Sample Code</th>
<th>Number of S. cerevisiae ( /94)</th>
<th>Number successfully genotyped</th>
<th>Genotypes unique to NZ</th>
<th>Genotypes unique to that sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hawke’s Bay</td>
<td>Constellation NZ</td>
<td>Kikowhero</td>
<td>Soil</td>
<td>HCNKISO</td>
<td>94</td>
<td>9</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Hawke’s Bay</td>
<td>Constellation NZ</td>
<td>Kikowhero</td>
<td>Spontaneous Ferment</td>
<td>HCNKISF</td>
<td>93</td>
<td>26</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>Hawke’s Bay</td>
<td>Constellation NZ</td>
<td>Te Tua</td>
<td>Soil</td>
<td>HCNTHISO</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Hawke’s Bay</td>
<td>Constellation NZ</td>
<td>Te Tua</td>
<td>Spontaneous Ferment</td>
<td>HCNTHISF</td>
<td>64</td>
<td>21</td>
<td>17</td>
<td>12</td>
</tr>
<tr>
<td>Hawke’s Bay</td>
<td>Pernod Ricard</td>
<td>Matapiro</td>
<td>Soil</td>
<td>HPRMAISO</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Hawke’s Bay</td>
<td>Pernod Ricard</td>
<td>Matapiro</td>
<td>Spontaneous Ferment</td>
<td>HPRMAISF</td>
<td>93</td>
<td>22</td>
<td>20</td>
<td>18</td>
</tr>
<tr>
<td>Hawke’s Bay</td>
<td>Pernod Ricard</td>
<td>Montana Terraces</td>
<td>Soil</td>
<td>HPRMTISO</td>
<td>94</td>
<td>16</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Hawke’s Bay</td>
<td>Trinity Hill</td>
<td>Esprit</td>
<td>Spontaneous Ferment</td>
<td>HTHESSO</td>
<td>NG</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Hawke’s Bay</td>
<td>Trinity Hill</td>
<td>Esprit</td>
<td>Soil</td>
<td>HTHESSF</td>
<td>91</td>
<td>20</td>
<td>16</td>
<td>9</td>
</tr>
<tr>
<td>Hawke’s Bay</td>
<td>Trinity Hill</td>
<td>Mangaorapa</td>
<td>Soil</td>
<td>HTHMASO</td>
<td>94</td>
<td>7</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Hawke’s Bay</td>
<td>Trinity Hill</td>
<td>Mangaorapa</td>
<td>Spontaneous Ferment</td>
<td>HTHMASF</td>
<td>82</td>
<td>20</td>
<td>12</td>
<td>9</td>
</tr>
<tr>
<td>Martinborough</td>
<td>Ata Rangi</td>
<td>Lismore</td>
<td>Soil</td>
<td>MARLISO</td>
<td>NG</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Martinborough</td>
<td>Ata Rangi</td>
<td>Lismore</td>
<td>Spontaneous Ferment</td>
<td>MARLISF</td>
<td>17</td>
<td>16</td>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td>Martinborough</td>
<td>Ata Rangi</td>
<td>Ariki</td>
<td>Soil</td>
<td>MARARISO</td>
<td>94</td>
<td>6</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Martinborough</td>
<td>Ata Rangi</td>
<td>Ariki</td>
<td>Spontaneous Ferment</td>
<td>MARARISF</td>
<td>73</td>
<td>24</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>Martinborough</td>
<td>Te Kairanga</td>
<td>SBLTKYNG</td>
<td>Soil</td>
<td>MTKSISO</td>
<td>89</td>
<td>19</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>Martinborough</td>
<td>Te Kairanga</td>
<td>SBLTKYNG</td>
<td>Spontaneous Ferment</td>
<td>MTKSISF</td>
<td>94</td>
<td>20</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Martinborough</td>
<td>Te Kairanga</td>
<td>SBLB2</td>
<td>Soil</td>
<td>MTKSBISO</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Martinborough</td>
<td>Te Kairanga</td>
<td>SBLB2</td>
<td>Spontaneous Ferment</td>
<td>MTKSBISF</td>
<td>94</td>
<td>20</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Martinborough</td>
<td>Palliser</td>
<td>Clouston</td>
<td>Soil</td>
<td>MPACLISO</td>
<td>94</td>
<td>9</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Martinborough</td>
<td>Palliser</td>
<td>Clouston</td>
<td>Spontaneous Ferment</td>
<td>MPACLISF</td>
<td>91</td>
<td>24</td>
<td>16</td>
<td>13</td>
</tr>
<tr>
<td>Martinborough</td>
<td>Palliser</td>
<td>Pencarrow</td>
<td>Soil</td>
<td>MPAPEISO</td>
<td>94</td>
<td>12</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Martinborough</td>
<td>Palliser</td>
<td>Pencarrow</td>
<td>Spontaneous Ferment</td>
<td>MPAPEISF</td>
<td>23</td>
<td>23</td>
<td>17</td>
<td>13</td>
</tr>
<tr>
<td>Nelson</td>
<td>SelFried</td>
<td>Challies</td>
<td>Soil</td>
<td>NSECHISO</td>
<td>NG</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Nelson</td>
<td>SelFried</td>
<td>Challies</td>
<td>Spontaneous Ferment</td>
<td>NSECHISF</td>
<td>94</td>
<td>20</td>
<td>7</td>
<td>3</td>
</tr>
</tbody>
</table>

(Continued over page)
<table>
<thead>
<tr>
<th>Region</th>
<th>Company</th>
<th>Vineyard</th>
<th>Niche</th>
<th>Sample Code</th>
<th>Number of S. cerevisiae ( /94)</th>
<th>Number successfully genotyped</th>
<th>Genotypes unique to NZ</th>
<th>Genotypes unique to that sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nelson</td>
<td>Seifried</td>
<td>Winery</td>
<td>Soil</td>
<td>NSEWIso</td>
<td>94</td>
<td>20</td>
<td>11</td>
<td>5</td>
</tr>
<tr>
<td>Nelson</td>
<td>Seifried</td>
<td>Redwood Valley, Malling Rd</td>
<td>Soil</td>
<td>NSERVso</td>
<td>63</td>
<td>23</td>
<td>11</td>
<td>6</td>
</tr>
<tr>
<td>Nelson</td>
<td>Seifried</td>
<td>Edens Rd</td>
<td>Soil</td>
<td>NSEERso</td>
<td>74</td>
<td>24</td>
<td>18</td>
<td>10</td>
</tr>
<tr>
<td>Nelson</td>
<td>Seifried</td>
<td>Brightwater</td>
<td>Soil</td>
<td>NSEBRso</td>
<td>91</td>
<td>19</td>
<td>11</td>
<td>5</td>
</tr>
<tr>
<td>Nelson</td>
<td>Neudorf</td>
<td>Maggies</td>
<td>Soil</td>
<td>NNEMAso</td>
<td>82</td>
<td>19</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>Wairau</td>
<td>Pernod Ricard</td>
<td>Squire SBL D block</td>
<td>Soil</td>
<td>WPRSQso</td>
<td>94</td>
<td>16</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Wairau</td>
<td>Pernod Ricard</td>
<td>Squire SBL D block</td>
<td>Spontaneous Ferment</td>
<td>WPRSQsf</td>
<td>64</td>
<td>20</td>
<td>11</td>
<td>8</td>
</tr>
<tr>
<td>Wairau</td>
<td>Pernod Ricard</td>
<td>Booker SBL K block</td>
<td>Soil</td>
<td>WPRBOso</td>
<td>19</td>
<td>19</td>
<td>14</td>
<td>11</td>
</tr>
<tr>
<td>Wairau</td>
<td>Churton</td>
<td>Loin</td>
<td>Soil</td>
<td>WCHLOso</td>
<td>93</td>
<td>20</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>Wairau</td>
<td>Vita Brevis</td>
<td>Vita Brevis</td>
<td>Soil</td>
<td>WVBVBso</td>
<td>94</td>
<td>8</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Wairau</td>
<td>Delegat's</td>
<td>Airfields Oyster Bay vineyard Sav blanc F block</td>
<td>Soil</td>
<td>WDEOBso</td>
<td>94</td>
<td>14</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Wairau</td>
<td>Delegat's</td>
<td>Airfields Oyster Bay vineyard Sav blanc F block</td>
<td>Spontaneous Ferment</td>
<td>WDEOBsf</td>
<td>14</td>
<td>12</td>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td>Wairau</td>
<td>Delegat's</td>
<td>Airfields Oyster Bay vineyard Sav blanc F block</td>
<td>Spontaneous Ferment</td>
<td>WDEOBsfP</td>
<td>10</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Wairau</td>
<td>Villa Maria</td>
<td>Winery block</td>
<td>Soil</td>
<td>WVMWBso</td>
<td>94</td>
<td>6</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Wairau</td>
<td>Mount Riley</td>
<td>Mount Riley</td>
<td>Soil</td>
<td>WMRMRso</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Wairau</td>
<td>Mount Riley</td>
<td>Mount Riley</td>
<td>Spontaneous Ferment</td>
<td>APRFVsso</td>
<td>94</td>
<td>8</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Awatere</td>
<td>Pernod Ricard</td>
<td>Frey Vineyard Ltd</td>
<td>Soil</td>
<td>APRFVs</td>
<td>44</td>
<td>27</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Awatere</td>
<td>Pernod Ricard</td>
<td>Frey Vineyard Ltd</td>
<td>Spontaneous Ferment</td>
<td>APRSEso</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

(Continued over page)
<table>
<thead>
<tr>
<th>Region</th>
<th>Company</th>
<th>Vineyard</th>
<th>Niche</th>
<th>Sample Code</th>
<th>Number of <em>S. cerevisiae</em> ( /94)</th>
<th>Number successfully genotyped</th>
<th>Genotypes unique to NZ</th>
<th>Genotypes unique to that sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Awatere</td>
<td>Pernod Ricard</td>
<td>Seaview SBL F block</td>
<td>Spontaneous Ferment</td>
<td>APRSEsf</td>
<td>11 ( /94)</td>
<td>11</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>Awatere</td>
<td>Constellation NZ</td>
<td>Castle Cliffs</td>
<td>Soil</td>
<td>ACNCCso</td>
<td>94 ( /94)</td>
<td>12</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Awatere</td>
<td>Constellation NZ</td>
<td>Castle Cliffs</td>
<td>Spontaneous Ferment</td>
<td>ACNCCsf</td>
<td>8 ( /94)</td>
<td>7</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Awatere</td>
<td>Constellation NZ</td>
<td>The Favourite</td>
<td>Soil</td>
<td>ACNTFsso</td>
<td>0 ( /94)</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Awatere</td>
<td>Constellation NZ</td>
<td>The Favourite</td>
<td>Spontaneous Ferment</td>
<td>ACNTFsf</td>
<td>4 ( /94)</td>
<td>4</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Awatere</td>
<td>Constellation NZ</td>
<td>Casey’s Rd</td>
<td>Soil</td>
<td>ACNCRso</td>
<td>0 ( /94)</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Awatere</td>
<td>Constellation NZ</td>
<td>Casey’s Rd</td>
<td>Spontaneous Ferment</td>
<td>ACNCRsf</td>
<td>93 ( /94)</td>
<td>29</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Awatere</td>
<td>Tohu</td>
<td>Arapoto Conv</td>
<td>Soil</td>
<td>ATOACso</td>
<td>94 ( /94)</td>
<td>7</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Central Otago</td>
<td>Rippon</td>
<td>Rippon</td>
<td>Soil</td>
<td>CRIRIsso</td>
<td>NG ( /94)</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Central Otago</td>
<td>Misha’s</td>
<td>Misha’s Vineyard</td>
<td>Soil</td>
<td>CMVMVso</td>
<td>NG ( /94)</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Central Otago</td>
<td>Misha’s</td>
<td>Misha’s Vineyard</td>
<td>Spontaneous Ferment</td>
<td>CMVMVsf</td>
<td>3 ( /94)</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Central Otago</td>
<td>Mt. Difficulty</td>
<td>Mt. Difficulty</td>
<td>Soil</td>
<td>CMDMDso</td>
<td>0 ( /94)</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Central Otago</td>
<td>Mt. Difficulty</td>
<td>Mt. Difficulty</td>
<td>Spontaneous Ferment</td>
<td>CMDMDsf</td>
<td>1 ( /94)</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Central Otago</td>
<td>Domain Road</td>
<td>Domain Road</td>
<td>Soil</td>
<td>CDRDRso</td>
<td>0 ( /94)</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Central Otago</td>
<td>Domain Road</td>
<td>Domain Road</td>
<td>Spontaneous Ferment</td>
<td>CDRDRsf</td>
<td>68 ( /94)</td>
<td>24</td>
<td>13</td>
<td>10</td>
</tr>
<tr>
<td>Central Otago</td>
<td>Amisfield</td>
<td>Block 2</td>
<td>Soil</td>
<td>CAMBIsso</td>
<td>0 ( /94)</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Central Otago</td>
<td>Amisfield</td>
<td>Block 2</td>
<td>Spontaneous Ferment</td>
<td>CAMBIsf</td>
<td>91 ( /94)</td>
<td>22</td>
<td>16</td>
<td>11</td>
</tr>
<tr>
<td>Central Otago</td>
<td>Coal Pit</td>
<td>Coal Pit</td>
<td>Soil</td>
<td>CPEGIso</td>
<td>0 ( /94)</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Central Otago</td>
<td>Coal Pit</td>
<td>Coal Pit</td>
<td>Spontaneous Ferment</td>
<td>CPEGIsf</td>
<td>87 ( /94)</td>
<td>27</td>
<td>13</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Total</td>
<td>3432</td>
<td>811</td>
<td>365</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Total ferment</td>
<td>2210</td>
<td>681</td>
<td>344</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Total soil</td>
<td>1222</td>
<td>130</td>
<td>21</td>
</tr>
</tbody>
</table>
Table AII.2: Further details about the samples obtained from non-managed or native sites across New Zealand including the species of tree they were isolated from. NG refers to no growth meaning the sample did not yield any yeast isolates.

<table>
<thead>
<tr>
<th>Region</th>
<th>Site</th>
<th>Tree Species</th>
<th>Common Name</th>
<th>Niche</th>
<th>Sample Code</th>
<th>Number of S. cerevisiae ( /94)</th>
<th>Number successfully genotyped</th>
<th>Genotypes unique to NZ</th>
<th>Genotypes unique to that sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hawke’s Bay</td>
<td>Boundary Stream</td>
<td>Pseudopanax arboreus</td>
<td>Five Finger</td>
<td>Soil</td>
<td>UHTTso1</td>
<td>NG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hawke’s Bay</td>
<td>Boundary Stream</td>
<td>Pseudopanax arboreus</td>
<td>Five Finger</td>
<td>Fruit</td>
<td>UHTTfr1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hawke’s Bay</td>
<td>Boundary Stream</td>
<td>Alseuosmia pusilla</td>
<td>Soil</td>
<td>UHTTso2</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hawke’s Bay</td>
<td>Boundary Stream</td>
<td>Alseuosmia pusilla</td>
<td>Fruit</td>
<td>UHTTfr2</td>
<td>NG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hawke’s Bay</td>
<td>Boundary Stream</td>
<td>Coprosma sp.</td>
<td>Coprosma</td>
<td>Soil</td>
<td>UHCOso1</td>
<td>NG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hawke’s Bay</td>
<td>Boundary Stream</td>
<td>Coprosma sp.</td>
<td>Coprosma</td>
<td>Fruit</td>
<td>UHCOfr1</td>
<td>NG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hawke’s Bay</td>
<td>Boundary Stream</td>
<td>Coprosma sp.</td>
<td>Coprosma</td>
<td>Soil</td>
<td>UHCOso2</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hawke’s Bay</td>
<td>Boundary Stream</td>
<td>Coprosma sp.</td>
<td>Coprosma</td>
<td>Fruit</td>
<td>UHCOfr2</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hawke’s Bay</td>
<td>Boundary Stream</td>
<td>Gaultheria antipoda</td>
<td>Snowberry or False Beech</td>
<td>Soil</td>
<td>UHBMso</td>
<td>NG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hawke’s Bay</td>
<td>Boundary Stream</td>
<td>Gaultheria antipoda</td>
<td>Snowberry or False Beech</td>
<td>Fruit</td>
<td>UHBMfr</td>
<td>NG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hawke’s Bay</td>
<td>Boundary Stream</td>
<td>Melicytus ramiflorus</td>
<td>Whiteywood or Māhoe</td>
<td>Soil</td>
<td>UHMASo</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hawke’s Bay</td>
<td>Boundary Stream</td>
<td>Melicytus ramiflorus</td>
<td>Whiteywood or Māhoe</td>
<td>Fruit</td>
<td>UHMAfr</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Martinborough</td>
<td>Waiohine Gorge</td>
<td>Coriaria arborea</td>
<td>Tutu</td>
<td>Soil</td>
<td>UMTTso1</td>
<td>94</td>
<td>7</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Martinborough</td>
<td>Waiohine Gorge</td>
<td>Coriaria arborea</td>
<td>Tutu</td>
<td>Fruit</td>
<td>UMTTfr1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Martinborough</td>
<td>Waiohine Gorge</td>
<td>Rhipogonum scandens</td>
<td>Supplejack</td>
<td>Soil</td>
<td>UMTTso2</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Martinborough</td>
<td>Waiohine Gorge</td>
<td>Rhipogonum scandens</td>
<td>Supplejack</td>
<td>Fruit</td>
<td>UMTTfr2</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Martinborough</td>
<td>Waiohine Gorge</td>
<td>Coprosma sp.</td>
<td>Coprosma</td>
<td>Soil</td>
<td>UMCOso1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Martinborough</td>
<td>Waiohine Gorge</td>
<td>Coprosma sp.</td>
<td>Coprosma</td>
<td>Fruit</td>
<td>UMCOfr1</td>
<td>NG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Martinborough</td>
<td>Waiohine Gorge</td>
<td>Coprosma sp.</td>
<td>Coprosma</td>
<td>Soil</td>
<td>UMCOso2</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Martinborough</td>
<td>Waiohine Gorge</td>
<td>Coprosma sp.</td>
<td>Coprosma</td>
<td>Fruit</td>
<td>UMCOfr2</td>
<td>NG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Martinborough</td>
<td>Waiohine Gorge</td>
<td>Pseudopanax arboreus</td>
<td>Five Finger</td>
<td>Soil</td>
<td>UMBMso</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Martinborough</td>
<td>Waiohine Gorge</td>
<td>Pseudopanax arboreus</td>
<td>Five Finger</td>
<td>Fruit</td>
<td>UMBMfr</td>
<td>92</td>
<td>8</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Martinborough</td>
<td>Waiohine Gorge</td>
<td>Coprosma sp.</td>
<td>Coprosma</td>
<td>Soil</td>
<td>UMMAso</td>
<td>94</td>
<td>8</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

(Continued over page)
<table>
<thead>
<tr>
<th>Region</th>
<th>Site</th>
<th>Tree Species</th>
<th>Common Name</th>
<th>Niche</th>
<th>Sample Code</th>
<th>Number of <em>S. cerevisiae</em> ( /94)</th>
<th>Number successfully genotyped</th>
<th>Genotypes unique to NZ</th>
<th>Genotypes unique to that sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Martinborough</td>
<td>Waiohine Gorge</td>
<td>Coprosma sp.</td>
<td>Coprosma</td>
<td>Fruit</td>
<td>UNMAfr</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nelson</td>
<td>Brook Sanctuary</td>
<td>Coriaria arboarea</td>
<td>Tutu</td>
<td>Soil</td>
<td>UNTTs01</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nelson</td>
<td>Brook Sanctuary</td>
<td>Coriaria arboarea</td>
<td>Tutu</td>
<td>Fruit</td>
<td>UNTTfr1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nelson</td>
<td>Brook Sanctuary</td>
<td>Coriaria arboarea</td>
<td>Tutu</td>
<td>Soil</td>
<td>UNTTs02</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nelson</td>
<td>Snowden’s Bush</td>
<td>Coprosma sp.</td>
<td>Coprosma</td>
<td>Soil</td>
<td>UNCOs01</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nelson</td>
<td>Snowden’s Bush</td>
<td>Coprosma sp.</td>
<td>Coprosma</td>
<td>Fruit</td>
<td>UNCOfr1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nelson</td>
<td>Snowden’s Bush</td>
<td><em>Melicytus ramiflorus</em></td>
<td>Whiteywood or Māhoe</td>
<td>Soil</td>
<td>UNCOs02</td>
<td>NG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nelson</td>
<td>Snowden’s Bush</td>
<td><em>Melicytus ramiflorus</em></td>
<td>Whiteywood or Māhoe</td>
<td>Fruit</td>
<td>UNCOfr2</td>
<td>NG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nelson</td>
<td>Snowden’s Bush</td>
<td><em>Pittosporum tenuifolium</em></td>
<td>Black Matipo or Kōhūhū</td>
<td>Soil</td>
<td>UNBMso</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nelson</td>
<td>Snowden’s Bush</td>
<td><em>Pittosporum tenuifolium</em></td>
<td>Black Matipo or Kōhūhū</td>
<td>Fruit</td>
<td>UNBMfr</td>
<td>NG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nelson</td>
<td>Snowden’s Bush</td>
<td><em>Melicytus ramiflorus</em></td>
<td>Whiteywood or Māhoe</td>
<td>Soil</td>
<td>UNMAso</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nelson</td>
<td>Snowden’s Bush</td>
<td><em>Melicytus ramiflorus</em></td>
<td>Whiteywood or Māhoe</td>
<td>Fruit</td>
<td>UNMAfr</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wairau</td>
<td>Onalamatu Scenic Reserve</td>
<td>Coriaria arboarea</td>
<td>Tutu</td>
<td>Soil</td>
<td>UWTTs01</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wairau</td>
<td>Onalamatu Scenic Reserve</td>
<td>Coriaria arboarea</td>
<td>Tutu</td>
<td>Fruit</td>
<td>UWTTfr1</td>
<td>NG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wairau</td>
<td>Onalamatu Scenic Reserve</td>
<td>Coriaria arboarea</td>
<td>Tutu</td>
<td>Soil</td>
<td>UWTTs02</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wairau</td>
<td>Onalamatu Scenic Reserve</td>
<td>Coriaria arboarea</td>
<td>Tutu</td>
<td>Fruit</td>
<td>UWTTfr2</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wairau</td>
<td>Onalamatu Scenic Reserve</td>
<td>Coprosma sp.</td>
<td>Coprosma</td>
<td>Soil</td>
<td>UWCOs01</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wairau</td>
<td>Onalamatu Scenic Reserve</td>
<td>Coprosma sp.</td>
<td>Coprosma</td>
<td>Fruit</td>
<td>UWCOf1</td>
<td>NG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wairau</td>
<td>Onalamatu Scenic Reserve</td>
<td>Coprosma sp.</td>
<td>Coprosma</td>
<td>Soil</td>
<td>UWCOs02</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wairau</td>
<td>Onalamatu Scenic Reserve</td>
<td>Coprosma sp.</td>
<td>Coprosma</td>
<td>Fruit</td>
<td>UWCOf2</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wairau</td>
<td>Onalamatu Scenic Reserve</td>
<td><em>Pittosporum tenuifolium</em></td>
<td>Black Matipo or Kōhūhū</td>
<td>Soil</td>
<td>UWBMso</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wairau</td>
<td>Onalamatu Scenic Reserve</td>
<td><em>Pittosporum tenuifolium</em></td>
<td>Black Matipo or Kōhūhū</td>
<td>Fruit</td>
<td>UWBMfr</td>
<td>NG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wairau</td>
<td>Onalamatu Scenic Reserve</td>
<td><em>Melicytus ramiflorus</em></td>
<td>Whiteywood or Māhoe</td>
<td>Soil</td>
<td>UWMAs0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wairau</td>
<td>Onalamatu Scenic Reserve</td>
<td><em>Melicytus ramiflorus</em></td>
<td>Whiteywood or Māhoe</td>
<td>Fruit</td>
<td>UWMAfr</td>
<td>NG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Awatere</td>
<td>Tohu Sign</td>
<td><em>Cordyline australis</em></td>
<td>Cabbage Tree</td>
<td>Soil</td>
<td>UATTs01</td>
<td>94</td>
<td>8</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

(Continued over page)
<table>
<thead>
<tr>
<th>Region</th>
<th>Site</th>
<th>Tree Species</th>
<th>Common Name</th>
<th>Niche</th>
<th>Sample Code</th>
<th>Number of S. cerevisiae (/94)</th>
<th>Number successfully genotyped</th>
<th>Genotypes unique to NZ</th>
<th>Genotypes unique to that sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Awatere</td>
<td>Tohu Sign</td>
<td>Cordyline australis</td>
<td>Cabbage Tree</td>
<td>Fruit</td>
<td>UATTfr1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Awatere</td>
<td>Tohu Sign</td>
<td>Griselinia littoralis</td>
<td>Kapuka</td>
<td>Soil</td>
<td>UATTso2</td>
<td>94</td>
<td>8</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Awatere</td>
<td>Tohu Sign</td>
<td>Griselinia littoralis</td>
<td>Kapuka</td>
<td>Fruit</td>
<td>UATTfr2</td>
<td>NG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Awatere</td>
<td>Tohu Sign</td>
<td>Griselinia littoralis</td>
<td>Kapuka</td>
<td>Soil</td>
<td>UACOso1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Awatere</td>
<td>Tohu Sign</td>
<td>Griselinia littoralis</td>
<td>Kapuka</td>
<td>Fruit</td>
<td>UACOfr1</td>
<td>NG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Awatere</td>
<td>Road side leaving Tohu</td>
<td>Coprosma sp.</td>
<td>Coprosma</td>
<td>Soil</td>
<td>UACOso2</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Awatere</td>
<td>Road side leaving Tohu</td>
<td>Coprosma sp.</td>
<td>Coprosma</td>
<td>Fruit</td>
<td>UACOfr2</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Awatere</td>
<td>Upton Brook</td>
<td>Pittosporum tenuifolium</td>
<td>Black Matipo or Kōhūhū</td>
<td>Soil</td>
<td>UABMso</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Awatere</td>
<td>Upton Brook</td>
<td>Pittosporum tenuifolium</td>
<td>Black Matipo or Kōhūhū</td>
<td>Fruit</td>
<td>UABMfr</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Awatere</td>
<td>Tohu Sign</td>
<td>Cordyline australis</td>
<td>Cabbage Tree</td>
<td>Soil</td>
<td>UAMAso</td>
<td>NG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Awatere</td>
<td>Tohu Sign</td>
<td>Cordyline australis</td>
<td>Cabbage Tree</td>
<td>Fruit</td>
<td>UAMAfr</td>
<td>NG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central Otago</td>
<td>Queenstown-Glenorchy Rd</td>
<td>Cordyline australis</td>
<td>Cabbage Tree</td>
<td>Soil</td>
<td>UCTTso1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central Otago</td>
<td>Queenstown-Glenorchy Rd</td>
<td>Cordyline australis</td>
<td>Cabbage Tree</td>
<td>Fruit</td>
<td>UCTTfr1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central Otago</td>
<td>Queenstown-Glenorchy Rd</td>
<td>Coprosma sp.</td>
<td>Coprosma</td>
<td>Soil</td>
<td>UCTTso2</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central Otago</td>
<td>Queenstown-Glenorchy Rd</td>
<td>Coprosma sp.</td>
<td>Coprosma</td>
<td>Fruit</td>
<td>UCTTfr2</td>
<td>NG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central Otago</td>
<td>Queenstown-Glenorchy Rd</td>
<td>Griselinia littoralis</td>
<td>Kapuka</td>
<td>Soil</td>
<td>UCCOso1</td>
<td>NG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central Otago</td>
<td>Queenstown-Glenorchy Rd</td>
<td>Griselinia littoralis</td>
<td>Kapuka</td>
<td>Fruit</td>
<td>UCCOfr1</td>
<td>NG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central Otago</td>
<td>Queenstown-Glenorchy Rd</td>
<td>Coriaria arborea</td>
<td>Tutu</td>
<td>Soil</td>
<td>UCCOso2</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central Otago</td>
<td>Queenstown-Glenorchy Rd</td>
<td>Coriaria arborea</td>
<td>Tutu</td>
<td>Fruit</td>
<td>UCCOfr2</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central Otago</td>
<td>Queenstown-Glenorchy Rd</td>
<td>Coriaria arborea</td>
<td>Tutu</td>
<td>Soil</td>
<td>UCBMso</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central Otago</td>
<td>Queenstown-Glenorchy Rd</td>
<td>Coriaria arborea</td>
<td>Tutu</td>
<td>Fruit</td>
<td>UCBMfr</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central Otago</td>
<td>Queenstown-Glenorchy Rd</td>
<td>Cordyline australis</td>
<td>Cabbage Tree</td>
<td>Soil</td>
<td>UCMAso</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central Otago</td>
<td>Queenstown-Glenorchy Rd</td>
<td>Cordyline australis</td>
<td>Cabbage Tree</td>
<td>Fruit</td>
<td>UCMAfr</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Total**          |                          |                      |                      |       |             | 468                          | 39                          | 4                      | 1                              |
Table AII.3: Pairwise $R^2$ values from Obstruct. All values are significant with $P < 0.001$ for all comparisons except between Hawke’s Bay and Marlborough ($P = 0.06$).

<table>
<thead>
<tr>
<th></th>
<th>Hawke’s Bay</th>
<th>Martinborough</th>
<th>Nelson</th>
<th>Marlborough</th>
<th>Central Otago</th>
</tr>
</thead>
<tbody>
<tr>
<td>Martinborough</td>
<td>0.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nelson</td>
<td>0.12</td>
<td>0.23</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marlborough</td>
<td>0.02</td>
<td>0.05</td>
<td>0.16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central Otago</td>
<td>0.13</td>
<td>0.10</td>
<td>0.20</td>
<td>0.12</td>
<td></td>
</tr>
</tbody>
</table>

Table AII.4: Mean directional estimates of migration rate (Nm) as calculated by MIGRATE with 95% confidence intervals (mean estimate: lower 95% confidence interval – upper 95% confidence interval). All estimates go from the region on the left, towards the region along the top row.

<table>
<thead>
<tr>
<th></th>
<th>Hawke’s Bay</th>
<th>Martinborough</th>
<th>Nelson</th>
<th>Marlborough</th>
<th>Central Otago</th>
<th>Total Outward Migration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hawke’s Bay</td>
<td></td>
<td>45: 29-61</td>
<td>16: 8-23</td>
<td>63: 43-84</td>
<td>16: 7-27</td>
<td>140</td>
</tr>
</tbody>
</table>

**Total Inward Migration**

|              | 188 | 150 | 44  | 153 | 57  |
Figure AIII.1: CCA analysis visualising the differences between co-ferments and wines blended from the single genotype ferments. (a) All sample points and 50 % ellipses for co-ferments and blends. (b) The direction and magnitude of chemical loading vectors with labels provided for chemicals that reported significant $R^2$ values above 0.25 for the factor “type” in ANOVA analyses. Hexanol is the only chemical with a $R^2$ value above 0.25 that does not have a loading vector in the CCA analysis above 0.25 in magnitude.
Figure AIII.2: Plots from the linear transformations used to visualise the co-ferments and blends in respect to the single genotype ferments for each region. The 50% ellipse for the single genotype samples is shown for reference. (a) Awatere Valley. The ferments that were removed from the single genotype analysis for not finishing (i.e. becoming stuck) are also included for reference to the blends. (b) Central Otago. Again the stuck ferments from the single genotype analysis are also included for reference. (c) Hawke’s Bay. (d) Martinborough. (e) Nelson. The three sample points that correspond to the genotype of mixed ancestry are shown for reference but not included in the 50% ellipse. (f) Wairau Valley.
Figure AIII.3: The effect imposed by a single genotype with mixed ancestry derived from the Nelson region. (a) The original CCA analysis showing the regional centers and 50 % ellipses. (b) Recalculation of the regional centers and 50 % ellipses after the removal of the single mixed ancestry genotype derived from Nelson. The regional center for Nelson is shifted further to the lower left and the 50 % ellipse is substantially reduced in size, resulting in increased separation from other regions.
Table AIII.1: Regional pairwise PERMANOVA results showing the $R^2$ value for “region” and associated $P$-values. $P$-values corrected using a Benjamini-Hochberg correction for multiple hypothesis tests are also reported.

<table>
<thead>
<tr>
<th>Region 1</th>
<th>Region 2</th>
<th>$R^2$</th>
<th>$P$-Value</th>
<th>BH corrected $P$-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Awatere Valley</td>
<td>Central Otago</td>
<td>0.067</td>
<td>0.227</td>
<td>0.310</td>
</tr>
<tr>
<td>Awatere Valley</td>
<td>Hawke's Bay</td>
<td>0.057</td>
<td>0.071</td>
<td>0.159</td>
</tr>
<tr>
<td>Awatere Valley</td>
<td>Martinborough</td>
<td>0.030</td>
<td>0.037*</td>
<td>0.158</td>
</tr>
<tr>
<td></td>
<td>Nelson</td>
<td>0.032</td>
<td>0.343</td>
<td>0.429</td>
</tr>
<tr>
<td>Awatere Valley</td>
<td>Wairau Valley</td>
<td>0.026</td>
<td>0.637</td>
<td>0.637</td>
</tr>
<tr>
<td>Central Otago</td>
<td>Hawke's Bay</td>
<td>0.063</td>
<td>0.055</td>
<td>0.159</td>
</tr>
<tr>
<td>Central Otago</td>
<td>Martinborough</td>
<td>0.076</td>
<td>0.021*</td>
<td>0.158</td>
</tr>
<tr>
<td>Central Otago</td>
<td>Nelson</td>
<td>0.121</td>
<td>0.225</td>
<td>0.310</td>
</tr>
<tr>
<td>Central Otago</td>
<td>Wairau Valley</td>
<td>0.063</td>
<td>0.510</td>
<td>0.588</td>
</tr>
<tr>
<td>Hawke's Bay</td>
<td>Martinborough</td>
<td>0.072</td>
<td>0.011*</td>
<td>0.158</td>
</tr>
<tr>
<td>Hawke's Bay</td>
<td>Nelson</td>
<td>0.077</td>
<td>0.074</td>
<td>0.159</td>
</tr>
<tr>
<td>Hawke's Bay</td>
<td>Wairau Valley</td>
<td>0.037</td>
<td>0.161</td>
<td>0.302</td>
</tr>
<tr>
<td>Martinborough</td>
<td>Nelson</td>
<td>0.078</td>
<td>0.042</td>
<td>0.158</td>
</tr>
<tr>
<td>Martinborough</td>
<td>Wairau Valley</td>
<td>0.056</td>
<td>0.200</td>
<td>0.310</td>
</tr>
<tr>
<td>Nelson</td>
<td>Wairau Valley</td>
<td>0.077</td>
<td>0.559</td>
<td>0.599</td>
</tr>
</tbody>
</table>
Table AIII.2: Results from individual ANOVA analyses for each chemical using the single genotype ferments. R² values and the associated P-value for the factor “region” are reported. R² values of 0 and P-values reported as NA indicate these chemicals did not show enough regional distinction to be added to the model and could not be calculated.

<table>
<thead>
<tr>
<th>Chemical</th>
<th>Class</th>
<th>R²</th>
<th>P-value</th>
<th>BH corrected P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ethanol</td>
<td></td>
<td>0.167</td>
<td>7.66 x 10⁻⁵***</td>
<td>0.0002***</td>
</tr>
<tr>
<td>Total Acid</td>
<td></td>
<td>0.310</td>
<td>3.53 x 10⁻⁶***</td>
<td>1.41 x 10⁻⁵***</td>
</tr>
<tr>
<td>Volatile Acid</td>
<td></td>
<td>0.321</td>
<td>1.61 x 10⁻⁶***</td>
<td>7.51 x 10⁻⁶***</td>
</tr>
<tr>
<td>pH</td>
<td></td>
<td>0.375</td>
<td>5.70 x 10⁻⁸***</td>
<td>7.55 x 10⁻⁷***</td>
</tr>
<tr>
<td>Residual Sugar</td>
<td></td>
<td>0</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>3MHA</td>
<td>Thiol</td>
<td>0.161</td>
<td>0.004**</td>
<td>0.007**</td>
</tr>
<tr>
<td>3MH</td>
<td>Thiol</td>
<td>0.257</td>
<td>5.25 x 10⁻⁵***</td>
<td>0.0002***</td>
</tr>
<tr>
<td>Ethyl isobutyrate</td>
<td>Ester</td>
<td>0.342</td>
<td>8.09 x 10⁻⁸***</td>
<td>7.55 x 10⁻⁷***</td>
</tr>
<tr>
<td>Ethyl butanoate</td>
<td>Ester</td>
<td>0.254</td>
<td>6.23 x 10⁻⁵***</td>
<td>0.0002***</td>
</tr>
<tr>
<td>Ethyl 2-methyl butanoate</td>
<td>Ester</td>
<td>0.316</td>
<td>7.69 x 10⁻⁷***</td>
<td>5.38 x 10⁻⁶***</td>
</tr>
<tr>
<td>Ethyl isovalerate</td>
<td>Ester</td>
<td>0.120</td>
<td>0.032*</td>
<td>0.041*</td>
</tr>
<tr>
<td>Ethyl hexanoate</td>
<td>Ester</td>
<td>0.114</td>
<td>0.020*</td>
<td>0.028*</td>
</tr>
<tr>
<td>Ethyl octanoate</td>
<td>Ester</td>
<td>0.137</td>
<td>0.006**</td>
<td>0.009**</td>
</tr>
<tr>
<td>Ethyl decanoate</td>
<td>Ester</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Ethyl dodecanoate</td>
<td>Ester</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Ethyl acetate</td>
<td>Ester</td>
<td>0.114</td>
<td>0.055</td>
<td>0.055</td>
</tr>
<tr>
<td>Isobutyl acetate</td>
<td>Ester</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Isoamyl acetate</td>
<td>Ester</td>
<td>0.175</td>
<td>0.002**</td>
<td>0.004**</td>
</tr>
<tr>
<td>Hexylacetate</td>
<td>Ester</td>
<td>0.175</td>
<td>0.0002***</td>
<td>0.0005***</td>
</tr>
<tr>
<td>Cis-3-hexenyl acetate</td>
<td>Ester</td>
<td>0.098</td>
<td>0.041*</td>
<td>0.048*</td>
</tr>
<tr>
<td>Ethyl phenylacetate</td>
<td>Ester</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>β-phenylethyl acetate</td>
<td>Ester</td>
<td>0.147</td>
<td>0.011*</td>
<td>0.016*</td>
</tr>
<tr>
<td>Methyl octanoate</td>
<td>Ester</td>
<td>0.146</td>
<td>0.004**</td>
<td>0.007**</td>
</tr>
<tr>
<td>Isobutanol</td>
<td>Alcohol</td>
<td>0.116</td>
<td>0.029*</td>
<td>0.039*</td>
</tr>
<tr>
<td>Isoamylalcohol</td>
<td>Alcohol</td>
<td>0.108</td>
<td>0.045*</td>
<td>0.050*</td>
</tr>
<tr>
<td>Methionol</td>
<td>Alcohol</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Phenylethyl alcohol</td>
<td>Alcohol</td>
<td>0.114</td>
<td>0.054</td>
<td>0.055</td>
</tr>
<tr>
<td>β-damascenone</td>
<td>Norisoprenoids Terpene</td>
<td>0.318</td>
<td>1.59 x 10⁻⁶***</td>
<td>7.51 x 10⁻⁶***</td>
</tr>
<tr>
<td>β-ionone</td>
<td>Norisoprenoids Terpene</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Cis/trans-rose-oxide</td>
<td>Norisoprenoids Terpene</td>
<td>0.178</td>
<td>0.001***</td>
<td>0.002**</td>
</tr>
<tr>
<td>Linalool</td>
<td>Norisoprenoids Terpene</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>β-citronellol</td>
<td>Norisoprenoids Terpene</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Hexanol</td>
<td>C6 Compound</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Trans-3-hexen-1-ol</td>
<td>C6 Compound</td>
<td>0.122</td>
<td>0.039*</td>
<td>0.047*</td>
</tr>
<tr>
<td>Cis-3-hexen-1-ol</td>
<td>C6 Compound</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Isovaleric acid</td>
<td>Fatty Acid</td>
<td>0.370</td>
<td>5.89 x 10⁻⁸***</td>
<td>7.55 x 10⁻⁷***</td>
</tr>
<tr>
<td>Hexanoic acid</td>
<td>Fatty Acid</td>
<td>0.183</td>
<td>0.002**</td>
<td>0.004**</td>
</tr>
<tr>
<td>Octanoic acid</td>
<td>Fatty Acid</td>
<td>0.185</td>
<td>0.0009***</td>
<td>0.002**</td>
</tr>
<tr>
<td>Decanoic acid</td>
<td>Fatty Acid</td>
<td>0.104</td>
<td>0.052</td>
<td>0.055</td>
</tr>
</tbody>
</table>
APPENDIX IV: Chapter dataset (electronic)

The dataset pertaining to each chapter are provided in the electronic appendix that accompanies this thesis. Below are descriptions of each:

**Dataset AIV.1**: Dataset to accompany Chapter 3. This excel spreadsheet contains the allele lengths for all analysed isolates at all loci. Missing data is indicated by 0. This format is appropriate for input into GenAIEx.

**Dataset AIV.2**: Dataset to accompany Chapter 4. This excel spreadsheet contains all the raw cell counts from the pilot study experiment and their relevant proportions. “SM” under plate type refers to lab defined sporulation media.

**Dataset AIV.3**: Dataset to accompany Chapter 4. This excel spreadsheet contains all the raw cell counts from the time series experiment and their relevant proportions. The proportions of tetrads is in relation to the total number of sporulating cells.

**Dataset AIV.4**: Dataset to accompany Chapter 5. This excel spreadsheet contains the chemical concentrations and factors that were utilized in the analysis. Inferred pop refers to the inferred population the InStruct analysis group each isolate. Those labelled “M” were not strongly assigned to any one inferred population. Under the column stuck, “F” refers to normal fermentation and “S” refers to stuck fermentation.

**Dataset AIV.5**: Dataset from Chapter 6. The concentration of the three volatile thiols 3MH, 3MHA and 4MMP and the molar sum of 3MH and 3MHA along with the proportion of *S. uvarum* at the end of fermentation and the proportion of OTUs identified in the grape juice. The inverse of the proportion of *S. uvarum* at the end of ferment represents the proportion of *S. cerevisiae* as only two *Saccharomyces* species were identified. Thiol concentrations of 0 indicate the actual value was below the detection threshold of the analysis.