ANALYSIS OF MICROBIAL POPULATIONS IN WINES THROUGH NGS METHODOLOGIES

Dimitrios Kioroglou

UNIVERSITAT ROVIRA i VIRGILI

Analysis of microbial populations in wines through NGS methodologies

DIMITRIOS KIOROGLOU

DOCTORAL THESIS
2020

UNIVERSITAT ROVIRA I VIRGILI
ANALYSIS OF MICROBIAL POPULATIONS IN WINES THROUGH NGS METHODOLOGIES
Dimitrios Kioroglou

UNIVERSITAT ROVIRA I VIRGILI
ANALYSIS OF MICROBIAL POPULATIONS IN WINES THROUGH NGS METHODOLOGIES
Dimitrios Kioroglou

Dimitrios Kioroglou

Analysis of microbial populations in wines through NGS methodologies

Doctoral Thesis

supervised by Dr. Albert Mas and Dra. Maria del Carmen Portillo
Department Biochemistry and Biotechnology
BIOTENOL research group

Tarragona 2020

UNIVERSITAT ROVIRA I VIRGILI
ANALYSIS OF MICROBIAL POPULATIONS IN WINES THROUGH NGS METHODOLOGIES
Dimitrios Kioroglou

Statement of supervision

UNIVERSITAT
ROVIRA i VIRGILI

Departament de Bioquímica i Biotecnologia
Facultat de Enologia de Tarragona
Marcel·lí Domingo 1
43007 Tarragona

We STATE that the present study, entitled "**Analysis of microbial populations in wines through NGS methodologies**", presented by **Dimitrios Kioroglou** for the award of the degree of Doctor, has been carried out under our supervision at the Department **Biochemistry and Biotechnology** of this university.

Tarragona, September 2020

Doctoral Thesis Supervisors:

Dr. Albert Mas **Dra. Maria del Carmen Portillo**

UNIVERSITAT ROVIRA I VIRGILI
ANALYSIS OF MICROBIAL POPULATIONS IN WINES THROUGH NGS METHODOLOGIES
Dimitrios Kioroglou

Acknowledgements

Developing my PhD in Rovira i Virgili the past three years has been a wonderful journey, and I would like to thank numerous people for their help and support throughout this process.

First and foremost, I would like to express my deepest appreciation and gratitude to my supervisors Dr. Albert Mas and Dra. Maria del Carmen Portillo for giving me the opportunity to conduct my PhD under their supervision and guidance. It has been a great pleasure working with you, providing me with valuable insights, mentorship and support and at the same time giving me the right flexibility to develop my own academic voice. In addition, I would like to thank the university Rovira i Virgili, where with the Martí Franquès COFUND doctoral fellowship programme provided me with a very generous financial support that was extremely valuable for the realization of this PhD.

Special thanks should be given to Dra. Maria Jesús Torija and Dra. Paloma Pontón Merino for making sure that my incorporation to the University and the department of Biochemistry and Biotechnology was as smooth as possible, and for helping me with any query on administrative procedures I might have had. Moreover, I would like to thank Dra. Jessica Lleixà for her support in the lab and the great collaboration during my first year.

I am particularly grateful to Dr. Jose Maria Gonzalez Izarzugaza for hosting me in his research group at the Technical University of Denmark and for mentoring me during my research stay, and to Dra. Motlhalamme from Stellenbosch university of South Africa for her kindness to provide me with a transcriptomic dataset from one her experiments. Furthermore, special thanks should be given to Dra. Elena Kraeva-Deloire and Dr. Leigh M. Schmidtke from Charles Sturt University of Australia for the great collaboration we had during my second year.

Finally, at a personal level I would like to express my gratitude to my wife Maitane for all the motivation and positive energy that she instilled in me during these years, and for all the support I received from her so as to cope with all sorts of adversities.

Thank you all.

UNIVERSITAT ROVIRA I VIRGILI
ANALYSIS OF MICROBIAL POPULATIONS IN WINES THROUGH NGS METHODOLOGIES
Dimitrios Kioroglou

Dedicated to the soul that has left and the one that is coming.

UNIVERSITAT ROVIRA I VIRGILI
ANALYSIS OF MICROBIAL POPULATIONS IN WINES THROUGH NGS METHODOLOGIES
Dimitrios Kioroglou

CONTENTS

CONTENTS

ABSTRACT

Vinification and winematuration consist the two main phases of the winemaking process. In each phase the microbial communities exhibit a dynamic behaviour, derived mainly from their endeavour to cope with the limited source of nutrients. This microbial activity is crucial for the final wine product, since it results in the production of metabolites which in turn modulate the organoleptic characteristics of the wine and, in combination with environmental factors, it may have beneficial or catastrophic consequences for the wine quality. The past years, the main focus has merely been the detection of certain microorganisms, notorious for their wine spoiling capacities, along with the adoption of empirical strategies for the prevention and control of undesirable microbial growth. Nevertheless, the known limitations of the conventional methodologies used for such detection and monitoring, as well as the realization of the multifactorial basis of the wine spoilage have urged for the development of innovative strategies that could overcome these drawbacks and provide an in-depth view of the microbial diversity. Next-generation-sequencing has emerged as the technology that promises to fulfill such expectations, and in this dissertation various aspects related to its implementation on the wine metataxonomic analysis have been evaluated. These aspects concern its performance in comparison to conventional methodologies, and the establishment of a NGS-based bioinformatic and statistical framework, suitable for the analysis of metataxonomic data. Furthermore, various factors closely related to the winemaking process have been studied in various chapters, in order to assess the alignment of the derived results to previous studies, and the ability of NGS to provide with new insights. Overall, in this work NGS managed to corroborate previous findings and suggest subsequent studies, while the developed analytical framework was proven to be a solid foundation to support further meta-analysis.

UNIVERSITAT ROVIRA I VIRGILI
ANALYSIS OF MICROBIAL POPULATIONS IN WINES THROUGH NGS METHODOLOGIES
Dimitrios Kioroglou

PROJECT OUTLINE

Project justification

The current thesis has been conducted at the Faculty of Oenology of the University Rovira i Virgili as part of the Oenological Biotechnology research group (BIOTENOL) coordinated by Dr. Albert Mas. As a research group, BIOTENOL has been for many years analyzing the field and industrial ecology of the wine microorganisms, having numerous published articles dedicated to the winemaking process. Moreover, the thesis was part of the project "**Aplicación de Metagenomica y Metatranscriptomica al Control Microbiologico del Vino de Crianza**" (abbreviated as METACONVIN) whose funding was provided by the Spanish government and granted to Dra. Maria del Carmen Portillo (project code AGL2015-73272). A series of objectives were attached to the project, however they could be summarized as the process of applying novel omic techniques for the study, evaluation and control of the microbial dynamics in ageing wine. Finally, the current work has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skodowska-Curie grant agreement No. 713679 and from the University Rovira i Virgili (Marti Franquès - COFUND). In order to be awarded the Doctorate with international mention, I conducted a 3-month research stay at the Technical University of Denmark from June 1st, 2018 to August 31st, 2018. The results from the conducted research during that stay are summarized in the appendix A. Moreover, chapter 3 has been the outcome of a research collaboration with the Charles Sturt University of Australia, where a bioinformatic study was performed on the provided metataxonomic raw data.

Hypothesis and objectives

The main focus of this work is to demonstrate the benefits of applying NGS technology to the wine analysis, since its implementation has been advocated and praised by numerous studies covering various biological fields. Therefore, the main hypothesis of the dissertation can be formulated as following:

Hypothesis:
The development of NGS-based bioinformatic and statistical framework can provide with novel insights into the microbial dynamics during winematuration.

In order to test this hypothesis, the structure of the thesis has been developed in the form of several objectives, presented in several chapters. During the introduction, various details are given regarding the wine making process, wine diversity, key microorganisms that affect wine quality, conventional methodologies that have been applied in wine analysis, as well as the necessity of developing and implementing NGS-based methodologies into the wine analysis.

Chapter 1 focuses on the control of alcoholic fermentation, since it has been for a long time under research due to its importance in winemaking. Its motivation arose from the fact that several studies have shown that non-*Saccharomyces* yeasts, previously considered as wine spoilers, not always lead to wine spoilage. In fact, various beneficial organoleptic characteristics have been attributed to their use. Furthermore, wine spoilage has been described as a consequence of intricate microbial interactions combined with exogenous environmental factors [1]. The advent of NGS technologies has opened new possibilities to studying these interactions, promising at the same time the ability of overpassing the limitations of conventional methodologies such as PCR-DGGE, qPCR and plate culturing. Therefore, chapter 1 is based on the following objectives:

Objective 1:
Determine the state-of-art of NGS implementation in wine metataxonomic analysis and its usefulness for the control of alcoholic fermentation based on literature review.

Since the literature review yielded positive results regarding the use of NGS technologies, the next research question concerned the comparison of the performance between NGS and conventional methodologies in a study of spontaneous fermentations of sound, sour rot and *Botrytis*-infected grapes (Chapter 2):

Objective 2:
Comparison between NGS and conventional microbiological methodologies on their ability to monitor alcoholic fermentation of grapes based on the factors grape health and fermentation stage.

The insights provided from chapter 2 corroborated the increasing utility of the NGS technologies, however raised certain concerns regarding its subsequent implementation. First and foremost, combination of conventional and NGS methodologies is not possible in every study due to the cost ineffectiveness and labour intensity of the whole process. Furthermore, NGS implementation is related with laboratory and bioinformatic workflows, each one inherently associated with a series of factors and parameters that potentially could influence the analytical outcome. All these considerations led to the following objective dealt in chapter 3:

Objective 3:
Evaluation of the parameters applied to the bioinformatic pipelines destined to be implemented to subsequent NGS studies.

As soon as the bioinformatic framework was established, the focus was turned towards the improvement of the statistical methodology addressing certain limitations of current implementations. Thus, the main objective of chapter 4 was the following:

Objective 4:
Establishment of a non-parametric statistical framework for an unbiased inference of the microbial communities by NGS technologies.

The completion of the last two objectives set the basis for studying the winematuration process. During this process the dynamics of the microbial communities were examined from wine samples that were aged in oak barrels. Additionally, the impact of various factors on these dynamics was assessed leading to the following objective (Chapter 5):

Objective 5:
Factor-based NGS analysis of microbial dynamics in wine samples aged in oak barrels.

Along with the study of the microbial dynamics, the analysis of the wine metabolome during the maturation process is equally important. Therefore, the same experimental

design as in chapter 5 was used in order to study the metabolome of oak-aged wine samples. Moreover, targeted and untargeted metabolomic analytical methodologies were compared using the same factors as in chapter 5 formulating the following objective:

Objective 6:
Study of wine metabolome of wine samples that aged in oak barrels and factor-based comparison of metabolomic analytical methodologies.

All the aforementioned objectives incorporate various aspects related to the laboratory and bioinformatic procedures involved in the winemaking and winematuration process that are elaborated in the introduction.

GENERAL INTRODUCTION

Winemaking consists one of the oldest crafts in the realm of alcoholic beverages, having for centuries great socioeconomic impact. As a process, it encompasses a series of stages that can be grouped into two phases. The first phase includes the stages grape harvesting, grape crushing and fermentation that collectively are referred as "***vinification***" process. The second phase is referred as "***winematuration***" process and includes the period of wine ageing into barrels till the wine bottling [2]. Nevertheless, not all wines undergo this phase, since wines, characterized as "***young wines***", are bottled and commercialized after the vinification process and without being aged, or after being stored in oak-barrels for a very short period of time.

All the aforementioned stages are inextricably connected to the wine quality, since intrinsic factors of each stage could potentially lead to wine spoilage. Although, this work has been restricted to focusing on the analysis of the fermentation and winematuration process in oak barrels, factors concering the grape harvesting stage have also been examined.

Wine microbial diversity

The wine microbiota is a mixture of fungal and bacterial communities. The fungal communities include filamentous fungi and yeasts, though the latter play the most significant role during the winemaking and are separated into the *Saccharomyces* and non-*Saccharomyces* yeasts. They are ubiquitous in nature and are transferred onto the grape surface via wind and insects. Nevertheless, a very narrowed number of yeasts can be isolated from the grape surface, with some of the most common genera belonging to the non-*Saccharomyces* yeasts, such as *Hanseniaspora*, *Debaryomyces*, *Metschnikowia*, *Issatchenkia* and *Pichia* [3]. Overall, the main purpose of the yeasts concerns the production of ethanol and carbon dioxide during the primary fermentation, a process known

as "***alcoholic fermentation***" with the *Saccharomyces* genera having higher fermentation capacity than non-*Saccharomyces* [3]. The most notorious *Saccharomyces* species that exhibits such high fermenting capability is *Saccharomyces cerevisiae*, which although is rarely isolated on the grape surface, it reaches high abundance during the alcoholic fermentation [3]. On the other hand, non-*Saccharomyces* are characterized by a less competitive nature and for many years they were considered undesirable and wine spoilers as there were species, with the most dreadful belonging to the genus *Brettanomyces*, that were connected to the production of foul odors and off-flavours [3]. Nevertheless, this notion has been contradicted by results showing beneficial organoleptic characteristics attributed to their use, since they increase the wine complexity [4, 5], and it has been shown that wine spoilage is a consequence of intricate microbial interactions combined with exogenous environmental factors [1, 6]. Moreover, due to these intrinsic attributes of the non-*Saccharomyces*, a typical description of the fungal dynamics during the alcoholic fermentation has been the initial prevalence of the non-*Saccharomyces* genera, such as *Hanseniaspora*, *Metschnikowia* and *Starmerella*, followed by the rise and dominance of *Saccharomyces* genera [7–12].

Regarding the bacterial communities, they are mainly grouped into the acetic acid (AAB) and lactic acid bacteria (LAB), whose growth usually follows the fungal growth. AAB are Gram-negative and strictly aerobic microorganisms that are found on the ripe grape, with the most common described genera being *Acetobacter*, *Gluconobacter* and *Gluconoacetobacter* [2]. Their main role is the oxidation of alcohols into the corresponding acids, with the main ones being glucose to gluconic acid and ethanol to acetic acid. Although, these compounds are usually found in the wine metabolome, their low perception threshold may lead to undesirable wine characteristics when produced in high amounts, classifying this way the AAB as potential wine spoilers [13].

As far as the LAB are concerned, they are Gram-positive bacteria having as main role the conversion of the malic acid into lactic acid during the secondary fermentation, a process known as "***malolactic fermentation***". Its implementation depends on the goals of the winemaker, since it is not performed by each cellar, and mainly aims at wine deacidification [2]. Among the most commonly described lactic acid bacteria in wine are *Lactobacillus*, *Leuconostoc*, *Pediococcus* and *Oenococcus*, though the latter represents the genus that has better been adapted to the malolactic fermentation and to the winemaking process overall [2].

The growth of the bacterial communities are greatly affected by the fungal activity, due to the production of high levels of ethanol and CO_2, causing it to follow fungal growth. However, bacterial activity may lead to stuck fermentations in cases where fungal

growth has been delayed [2, 14], showing their opportunistic nature. Moreover, it has been demonstrated that the production of acetic acid and lactic acid, in combination with low pH, may inhibit the growth of *S. cerevisiae* [15]. Although, the production of the acetic acid has been attributed to AAB activity, its production has also been reported from LAB [3]. Contrary to the fungal communities, still little is known regarding the growth competition withing the bacterial communities, with results coming from the studies of Joyeux *et al.* [16] and Guillamón *et al.* [17] suggesting the lack of rival relationship.

In general, the study of the microbial dynamics during the fermentation process is an active research field, since there is limited knowledge on the underlying microbial interactions. There could be numerous reasons behind these interactions, but they are mainly driven by the limited amount of nutrients. Thus, each microorganism has developed its own strategy to cope with such a competitive environment, producing compounds which in turn modulate the wine characteristics. These interactions become even more intricate during winematuration, since the wine microbiota comes in contact with the microbiota developed inside the barrel [18]. Under ideal conditions, during the winematuration the wine should be characterized by the lack of living microbial communities [3]. However, this seems impossible since events such pH alteration, oxidation and cell autolysis may consist the right ingredients for microbial growth [13]. Therefore, during wine ageing the final wine product is a result of compounds produced by the wine's intrinsic microbiota as well as those released by the microbiota and the wood of the barrel. Nevertheless, there are also additional factors that may influence the final product such as grape variety, storage conditions and barrel age.

Limitations of conventional methodologies

Since the microbial interactions represent a factor that is difficult to control, in the past years most endeavors have been restricted to the identification and monitor of wine spoilers with classical methodologies, such as plate culturing, and molecular techniques, such as quantitative polymerase chain reaction (qPCR) and denaturing gradient gel electrophoresis (PCR-DGGE), collectively referred as "***conventional***" methodologies in this dissertation.

Plate culturing is aiming at the isolation of microorganisms or microbial communities on culture media, that favour their growth by taking advantage their physiological features and peculiarities. Nevertheless, the main drawback of this methodology consists the fact that wine microbiota includes many unculturable microorganisms that enter the viable but non-culturable state (VBNC) due to the change of their environment. One

microbial group well known for this behaviour are the AAB [13].

Due to these shortcomings of the plate culturing, the molecular techniques appeared more promising approaches since they are culture-independent. The main principle behind qPCR is the microbial quantification based on the density changes of fluorescent probes, and has been used for the detection of specific microbial groups, genera or species [19, 20]. Nevertheless, its limitations becomes obvious in cases where a more broad exploratory study of the wine microbiota is necessary, due to the scalability constraints [21, 22].

On the other hand, PCR-DGGE aims at the separation of PCR products of rDNA amplicons and is more appropriate for microbial profiling. It has been utilized for the study of bacterial [23] and fungal communities [24] and has the advantage that the individual bands can be used for subsequent sequencing and species identification. However, its main drawback comes from the fact that less abundant microorganisms cannot be detected [21, 22].

In addition to these methodologies, the wine industry has adopted empirical approaches such as addition of microbial growth inhibitors (such as SO_2) and starter cultures. The growth inhibitors are utilized in order to suppress and maintain at low levels undesirable microorganisms that belong to non-*Saccharomyces* yeasts, AAB or LAB. Nonetheless, their efficacy may be reduced due to their binding by wine metabolites, such as gluconic acid and acetaldehyde [13], and adverse health effects have been ascribed to their usage [13]. Furthermore, the role of started cultures has been the enhancement of certain wine features, however their implementation embodies the risk of causing a fermentation that cannot be controlled and potentially lead to spoiled wine.

In addition to all these limitations, the aforementioned approaches do not ensure the final wine quality and most importantly they do not represent adequate predictors of wine spoilage, since the distinction of the microbial activity as beneficial or harmful is much more complex and poorly defined [25]. Therefore, the understanding of the underlying mechanisms that drive the microbial dynamics is of paramount importance since on one hand will reduce the loses that the wine industry suffers due to spoiled wine, and on the other hand will provide a better mean of controlling the quality and characteristics of the final wine product.

NGS aspects of main objectives

In order to control the microbial dynamics during the fermentation process, or to monitor the wine quality during the winematuration, we have to understand the way that

microbial interactions affect the dynamics of the microbial communities. To do so, we need to be able to study and quantify, at any given moment, the wine diversity in its entirety. Thus, a methodology is needed that overcomes the aforementioned limitations of the conventional approaches. As a solution, next-generation-sequencing (NGS) represents a new and innovative strategy promising the ability to tackle these challenges, with various studies corroborating this claim and mentioned in chapter 1. Additionally, this claim also was experimentally evaluated in chapter 2. Nevertheless, as a methodology, it incorporates various aspects each one associated with a series of pitfalls that this dissertation is trying to address.

Laboratory aspects

In Figure 1 a brief schema is given regarding the stages a researcher has to follow when applying NGS-based methodologies in wine analysis. The process starts with a wine sample where the microbial DNA has been extracted using a DNA extraction protocol. There is a variety of protocols, however the one implemented in this work follows the steps defined in chapter 2. The extracted microbial DNA could be analyzed via shotgun sequencing, an approach known as "***metagenomic analysis***", or by targeting a short amplicon of marker genes used for taxonomic classification, a strategy known as "***metataxonomic analysis***" [26]. The latter is the method commonly implemented in wine analysis, due to the cost-effectiveness and low computational resources needed, and in chapter 1 various studies are cited regarding the different regions of the rRNA gene that researchers have been targeting in yeasts and bacteria. Despite the difference between the two methods, as terms, they are frequently used interchangeably due to the fact that the metagenomic term can be generalized and include any type of analysis associated with the entire genetic content of a biological sample, irrespectively of the fact whether the genetic content concerns the whole genome or not.

Prior to sequencing, the chosen amplicon should be amplified through the PCR process. During this process, one of the parameters that should be considered is the number of PCR cycles, since it has been demonstrated that high number of PCR cycles produce higher number of chimeric sequences [27]. These chimeric sequences represent artifacts of the sequencing process where an artificial sequence is produced from two different biological sequences [28]. Regarding the amplicons per se, although for reconstructing the bacterial communities the use of the 16S amplicon is well established, for yeasts there was research interest regarding the performance of two common targets; the 18S and ITS amplicons. Furthermore, although Illumina represents the most popular NGS-platform, Ion Torrent is another platform still in use whose performance consisted an additional

aspect to investigate. These laboratory aspects have been incorporated into the objectives 1 and 3 and could be summarized as following:

Laboratory aspects of Objective 1:

- NGS-platforms used for metataxonomic analysis.
- rRNA gene regions commonly targeted in metataxonomic analysis.

Laboratory aspects of Objective 3:

- Performance evaluation of Ion Torrent and Illumina NGS-platforms.
- Amount of chimeras produced by 30 and 45 PCR cycles.
- Comparison of 18S and ITS amplicons for yeast classification.

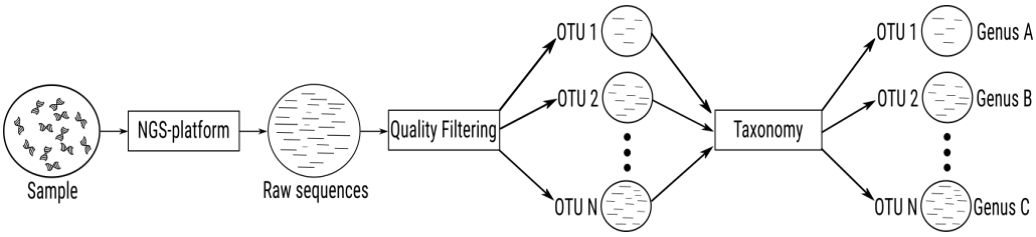


Figure 1: Brief representation of the typical stages of a metataxonomic analysis, from the biological sample until the taxonomic assignment of the amplicon-based sequences.

Bioinformatic aspects

During the sequencing process, the NGS-platform reads every amplicon and writes its sequence to a dedicated "***fastq***" file, a process known as "***base calling***". Depending on the NGS-platform, the reads in the resulting fastq file may include the barcodes attached to the amplicons during the PCR process. These barcodes serve to separate the resulted reads based on the biological sample they originated from, and are helpful when many samples are sequenced together. If barcode removal is necessary, it is performed through a process that is called "***demultiplexing***". Regarding the base calling, it is a probabilistic approach expressed in a form of a score, known as "***Phred quality score***", that refers to the base call error [29]; the higher the Phred quality score, the lower the base call error. This Phred quality score is also included in the fastq file for each read. Afterwards, the fastq file becomes the input of a subsequent bioinformatic analytical pipeline that aims to provide a taxonomy to each of the raw sequences, and thus reconstruct the microbial communities.

A known limitation of the NGS-platforms is the fact that Phred quality scores tend to decrease towards the end of the sequences due to the signal decay overtime. In order

to solve this issue, Illumina platforms produce, for each amplicon, a "***forward***" and a "***reverse***" read that are usually called "***paired-ends***". By merging the paired-ends into one read, the resulted sequence has higher Phred quality scores towards the end reducing this way the need for excessive trimming. Nevertheless, prior to taxonomic assignment, low Phred quality sequences should be filtered-out based on a defined quality threshold (Figure 1). The lower this threshold is set, the higher number of sequences are left to pass to the following steps. Additionally, as part of the quality filtering process of reads produced by Illumina platforms, the removal of optical duplicates is also included, a process known as "***reads dereplication***".

After the quality filtering step, taxonomies could be assigned to the filtered sequences. However, in order to reduce the demand in computational resources, the filtered sequences are grouped into the so-called "***Operational Taxonomic Units***" (OTUs), where each OTU represents a group of sequences that are similar based on a predefined similarity threshold (Figure 1). In the next step, OTUs that contain a single read, known as ***singletons***, are removed, and from each of the remaining OTUs, the sequence with the highest length is chosen as the representative sequence of that cluster. On these representative sequences, taxonomic assignment is performed by a taxonomic algorithm using dedicated and curated taxonomic databases based on the sequenced amplicon (Figure 1). Thus, each OTU is characterized by three pieces of information; the number of sequences it encompasses, known as "***OTU counts***", the representative sequence and the assigned taxonomy. All this information is incorporated into what is known as "***OTU table***". Before the OTU table is being used in the downstream analysis, taxonomies that do not pose research interest, such as those related to chloroplasts or mitochondria, are removed, along with taxonomies that have relative abundance below a certain threshold across all samples. Moreover, in case the assigned taxonomy has reached the species level, in addition to the OTU table the construction of a "***phylogenetic tree***" is a common practice. This phylogenetic tree is built by aligning the representative sequences to predefined templates provided by the curated databases, and holds the information regarding the phylogenetic relation of the sequences at species level.

Chapter 1 served as a starting point providing valuable insights into the state-of-art of the metataxonomic analysis, with the following bioinformatic aspects being incorporated into objective 1:

Bioinformatic aspects of Objective 1:

- Determination of available bioinformatic frameworks for metataxonomic analysis.
- Determination of available taxonomic databases for metataxonomic classification.

From the various bioinformatic frameworks that are available, QIIME was chosen in this work due to its popularity and its configuration based on the versatile Python programming language. By utilizing QIIME framework and applying parameters derived from the literature, the first bioinformatic pipeline was constructed, given in Table 1, in order to satisfy the objective 2.

Bioinformatic aspects of Objective 2:

- Implementation of QIIME-based bioinformatic pipeline with parameters derived from literature.

Steps		Parameters
1. Demultiplexing		
2. Phred quality filtering	→	Q10 Phred quality threshold.
3. Chimeras filtering		
4. OTU clustering	→	99% similarity threshold.
5. Singletons filtering	→	Remove OTUs with 1 read.
6. Representative sequences picking		
7. Taxonomic assignment	→	UCLUST taxonomic algorithm (species level classification), databases Greengenes (16S) and SILVA (18S).
8. Phylogenetic tree building		
9. Taxonomic filtering	→	Remove chloroplasts, mitochondria and taxonomies with relative abundance below 0.1% across all samples.

Table 1: Ion Torrent bioinformatic pipeline applied in chapter 2. Main implemented parameters are given.

Upon this initial bioinformatic implementation, further improvements were necessary in order to ensure, as much as possible, the robustness of its performance. One of the main parameters that were initially targeted was the Phred quality threshold, since very low thresholds may allow sequences with high base call error to pass, which in turn could lead to higher probability of erroneous taxonomic assignments. As a result, the dilemma

of higher number of sequences or higher overall quality was one of the aspects dealt in chapter 3.

Regarding the OTU clustering, it has been a classical approach implemented in numerous studies, with 97% been a popular similarity threshold. However, its efficiency and reliability has been questioned by Callahan *et al.* [30] suggesting an alternative strategy called "***Amplicon Sequence Variants***" (ASV). Nevertheless, a comparison of the two clustering strategies was conducted in the study of Van Der Pol *et al.* [31] where similar outcome was observed when the similarity threshold of the OTU clustering was set to 99%. Due to the popularity of the OTU clustering method, a confirmation on these findings was sought in chapter 3.

As far the taxonomic assignment is concerned, there are various classification algorithms currently available, however the main interest was confined to the incorporation of either the naive Bayes classification algorithm (SKLEARN) [32], or the BLAST+ algorithm [33] since their performance has been benchmarked by Bokulich *et al.* [34] against other algorithms that are included in QIIME. The former represents a classification approach based on machine learning, whereas the latter is based on the local alignment strategy. The assessment of their performance included the evaluation of their ability to reconstruct bacterial and fungal microbial communities of known composition, their established QIIME default parameters, as well as the possibility of improving taxonomic misclassification. All these bioinformatic aspects were included in the objective 3 that could be summarized as following:

Bioinformatic aspects of Objective 3:

- Performance evaluation of Q10 and Q20 Phred quality filtering thresholds.
- Comparison of OTU (99% similarity) and ASV clustering methods.
- Comparison between SKLEARN and BLAST+ classification algorithms.
- Improvement of taxonomic misclassification.
- Evaluation of QIIME default parameters.
- Establishment of a bioinformatic pipeline.

The investigation of the laboratory and bioinformatic aspects of the objective 3, led to the establishment of the bioinformatic pipeline that is given in Table 2 and to the decision of retaining the taxonomic assignment at genus level. This pipeline was destined to be implemented in chapter 5 formulating this way the bioinformatic aspect of the objective 5:

Bioinformatic aspects of Objective 5:

- Implementation of bioinformatic pipeline established in chapter 3.

Steps		Parameters
1. Paired ends merging		
2. Phred quality filtering	→	Q20 Phred quality threshold.
3. Reads dereplication		
4. OTU clustering	→	99% similarity threshold.
5. Chimeras filtering		
6. Singletons filtering	→	Remove OTUs with <10 reads.
7. Representative sequences picking		
8. Taxonomic assignment	→	BLAST+ taxonomic algorithm (genus level classification), databases SILVA (16S) and UNITE (ITS).
9. Taxonomic filtering	→	Remove chloroplasts, mitochondria and taxonomies with relative abundance below 1% across all samples.

Table 2: Illumina bioinformatic pipeline derived from chapter 3. Main implemented parameters are given.

Statistical aspects

After the construction and filtering of the OTU table, the next step concerns the comparison of the observed microbial diversity between the samples. Nevertheless, due to the sequencing and quality filtering process, each biological sample included in the OTU table contains different amount of sequences. Thus, in order to make comparisons between them this unevenness should be corrected, with "***rarefaction***" being the most common strategy applied to the OTU table. Rarefaction is a stochastic normalization process where, based on a threshold, assigns a fixed amount of sequences to each sample. This threshold is selected either from the minimum of the maximum amount of sequences observed between the samples, or by plotting "***rarefaction curves***" which , based on a diversity metric, show the maximum number of sequences after which the chosen metric does not change.

The rarefied OTU table is then utilized for the evaluation of the "***alpha***" and "***beta***" diversity. The alpha diversity refers to the observed biodiversity within each biological sample, and there are various metrics that can express it. However, the most commonly used are the phylogenetic diversity [35], and Shannon or Simpson indices [36]. The

phylogenetic diversity is derived from the previously constructed phylogenetic tree, and hence is applied in cases where the taxonomic assignment has reached species level. On the other hand, the Shannon and Simpson indices are based on the OTU counts.

The beta diversity refers to the observed biodiversity between the biological samples. As with the alpha diversity, it can be expressed from various metrics with weighted UniFrac distance [37] being implemented in cases of species level taxonomic classification. The weighted UniFrac distance is a quantitative method that shows differences between samples taking into consideration the information provided by the phylogenetic tree [38]. In contrast, in cases where the assigned taxonomies were constrained at genus level, the Bray-Curtis distance [39] is the common metric of choice as it based on the OTU counts.

These statistical aspects have been incorporated into the objectives 2 and 3 as following:

Statistical aspects of Objective 2:

- Evaluation of rarefaction as normalization methodology.
- Evaluation of species level taxonomic classification.
- Evaluation of phylogenetic diversity as alpha diversity metric.
- Evaluation of weighted UniFrac distance as beta diversity metric.

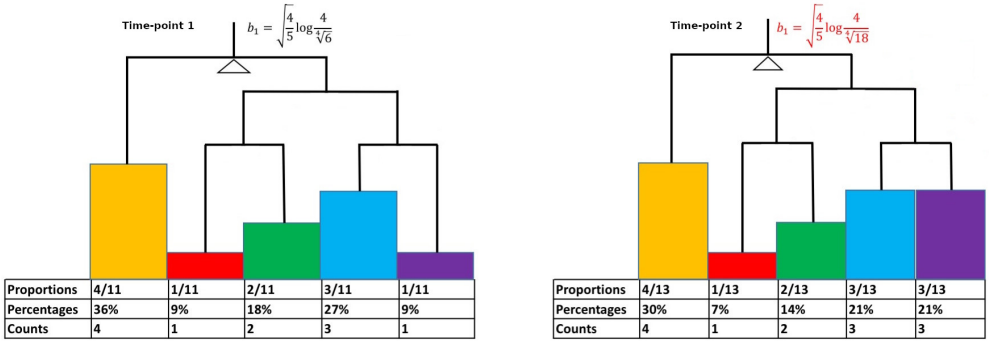
Statistical aspects of Objective 3:

- Evaluation of genus level taxonomic classification.
- Evaluation of Shannon index as alpha diversity metric.
- Evaluation of Bray-Curtis distance as beta diversity metric.

Although rarefaction is a popular normalization process, it has the limitation that certain amount of sequences have to be discarded to correct the sample size unevenness. Furthermore, one usual strategy of statistical inference on the microbial dynamics is the transformation of the OTU counts to relative abundances. This compositionality transformation process, adds the constrain of the abundances having to sum to 1 and could lead to misinterpretations especially when Pearson's correlation is applied. Moreover, implementation of parametric statistics, such as ANOVA, on the rarefied OTU counts could lead to false positives since they assume that the OTU counts follow a certain distribution (usually the normal distribution). These type of assumptions usually do not hold due to the sparse nature of this kind of data. Therefore, a non-parametric statistical framework was necessary to infer changes in the microbial communities.

The statistical framework chosen was the compositional analysis toolbox GNEISS [40] that introduces the concept of "***balances***" that could be explained with the following example. Figure 2 represents a hypothetical microbial community of a sample that contains 5 taxa (yellow, red, green, blue, purple) and has been analyzed in two different time-points to infer the changes among the microbial communities. From the 5 taxa, we know a priori that only the purple taxa has increased its abundance during the second time-point, whereas the rest have remained the same. Initially, the limitation of converting the OTU counts into relative abundances can be observed, since the change of one taxa causes the rest of the relative abundances to change as the constraint of having to sum to 1 must be applied. This gives the false impression that all the taxa have changed during the second time-point. GNEISS uses the unrarefied OTU counts, which initially imputes them by adding a pseudocount, and by applying Ward hierarchical clustering, clusters the taxa into two groups; one called "***numerator***" and the other "***denominator***". In this example, the numerator is consisted by the yellow taxa, whereas the denominator by the red, green, blue and purple taxa. It then calculates the number of taxa in the numerator (***r***) and the denominator (***s***), as well as the geometric mean of the OTU counts of the taxa in the numerator (***g(X$_r$)***) and the denominator (***g(X$_s$)***). Finally, for each time-point it calculates the balance using the following equation:

$$b = \sqrt{\frac{rs}{r+s}} log \frac{g(X_r)}{g(X_s)} \tag{1}$$



Proportions	4/11	1/11	2/11	3/11	1/11
Percentages	36%	9%	18%	27%	9%
Counts	4	1	2	3	1

Proportions	4/13	1/13	2/13	3/13	3/13
Percentages	30%	7%	14%	21%	21%
Counts	4	1	2	3	3

Figure 2: Example of the balances calculated by GNEISS for a sample at two different time-points. Each color represents a specific taxa. Counts refer to OTU counts and percentages to relative abundances. Image has been taken from Morton *et al.* [40] and modified for the purpose of the example.

Using the equation 1, the balance at time-point 1 is 0.84 and 0.59 at time-point 2. Since the balances are mainly affected by the log-ratio of the geometric means, this decrease reflects the increase of the purple taxa. However, in reality it is not possible to

know which taxa are changing. Therefore, an increase or decrease in the balances could indicate a change in the taxa of the numerator, the denominator or both. Therefore, the balances provided by GNEISS can be used to apply statistics such as PERMANOVA or regression in order to perform differential abundance or factor analysis. In order to infer which taxa drive the changes between the samples, additional visual analysis is required that could lead to a narrowed range of candidate microorganisms that could be studied in a subsequent more controlled experiment.

Biological aspects of main objectives

Apart from the NGS aspects, the main objectives included biological extensions that on one hand raised research interest due to their relevance in the winemaking process, and on the other hand they served as a mean to evaluate certain of the previously mentioned bioinformatic and statistical aspects.

In chapter 1, the view inside the alcoholic fermentation concerning the state-of-art of the metataxonomic analysis, was complemented with insights from the literature regarding the microbial diversity revealed by NGS approaches, as well as with strategies regarding the control of the alcoholic fermentation by NGS-based implementations.

Biological aspects of Objective 1:

- Literature review of the described microbial diversity of the alcoholic fermentation using NGS methodologies.
- Literature review of NGS-based implementations for the control of the alcoholic fermentation.

In chapter 2, the initial bioinformatic pipeline given in Table 1 along with the statistical aspects of the objective 2, were implemented for the performance comparison between NGS and conventional methodologies. However, this comparison was based on a spontaneous alcoholic fermentation study of grape must, with the aim to investigate the impact of the grape health on the microbial dynamics. Spontaneous fermentations are characterized by their unpredictability regarding the fermentation outcome, since the whole process is not controlled [3]. Therefore, the microbial dynamics can be observed in the absence of any human intervention.

The surface of the grape berry houses numerous microorganisms, and any grape damage could alter the microbial diversity in a way that it can influence negatively the wine quality [41]. The most common grape infections are sour rot and grey mold that cause significant crop losses. Thus, the factor grape health was divided into the groups of grapes

with rotten, *Botrytis*-affected, and healthy characteristics. Additionally, the alcoholic fermentation was divided into three stages (initial, mid and final fermentation), where the evolution of the microbial communities was followed during each stage leading to the following biological aspects:

Biological aspects of Objective 2:

- Impact of the factor grape health on the microbial dynamics of grape must spontaneous alcoholic fermentation.
- Impact of the factor fermentation stage on the microbial dynamics of grape must spontaneous alcoholic fermentation.

Apart from grape diseases, the wine quality can be influenced by other factors such as the grape variety, the grape maturation state and the regional characteristics of the vineyard, since they constitute factors that can influence the grape microbiota. The soil of the vineyard represents an important aspect for the winemaking, since it provides the necessary nutrients for the development of the plant and hosts fungi and bacteria that influence the plant's health and growth [42]. These characteristics of the soil are regional specific and, along with climatic conditions and human practices in the vineyard (such as irrigation), consist of what is termed as "***terroir***". Nowadays, there are numerous studies showing strong association between the terroir and the grape microbiome [42–49].

Furthermore, it is well understood that an additional factor that influences the grape microbiota is the berry development process [50, 51]. For instance, it has been demonstrated that the abundance of the fungal communities increases when the grape reaches maturation state, exhibiting the importance of the harvest time [22, 44, 50, 52–54]. Moreover, grape variety was another factor to consider, since the physiochemical changes that occur during the development of the grape varies between grapes [55]. Although, the influence of these factors on the grape microbiome is known, the amount of the variability concerning the observed microbial diversity that is accounted by each factor consisted the main focus. This research interest was pursued with a study involving grape samples from two grape varieties, originated from 4 different vineyards and harvested in two different harvest periods determined by the grape maturation state.

Biological aspects of Objective 4:

- Impact factor determination for the factors geographical origin, grape variety and maturation state on the microbial diversity of the grapes.

After establishing the bioinformatic and statistical framework in chapters 3 and 4, the research focus was directed towards the winematuration process. Wine ageing in

wooden barrels has become a common practice for the production of red wines of higher quality, due to the organoleptic characteristics added by this type of maturation process. However, still little is known regarding the microbial dynamics during the wine ageing and the underlying factors associated to these changes. There are many factors that could be incorporated to the studying of the wine ageing, however the ones considered in this work are time of wine ageing in oak barrels and amount of barrel's prior usage expressed in time. These two factors are shortly referred as ***time*** and ***barrel-type*** correspondingly. The samples originated from DOQ Priorat and DOCa Rioja regions, where at the samples of the latter an additional factor was studied. This factor concerned the comparison of wine ageing in oak barrels against glass bottles, a factor shortly referred as ***bottled-wine***. Thus, objective 5 embedded the following biological aspects:

Biological aspects of Objective 5:

- Determination of microbial evolution of wine samples that mature in oak barrels.
- Comparison of microbial dynamics between barrels based on the factor barrel-type and time.
- Comparison of 1-year microbial evolution between wine samples based on the factor bottled-wine.

In addition to studying the dynamics of the microbial communities, the study of the evolution of the metabolomic profile during the winematuration is equally important as the chemical composition represents the conditions during ageing and storage [56]. Although, targeted analysis, such as nuclear magnetic resonance (NMR), provide useful insights into a wide range of analytes, untargeted analyses have been proposed of being necessary for a more comprehensive analysis [56, 57]. Therefore, the wine samples from the experimental design of chapter 5 were used as a source for the study of the wine metabolome during winematuration. The same factors of chapter 5 were utilized in chapter 6 as the basis for comparing targeted and untargeted metabolomic analytical methodologies on their capacity to reveal informative patterns regarding the wine metabolome. Apart from the NMR method, the additional implemented methodologies were targeted and untargeted gas chromatography-mass spectrometry (GC-MS) and high-performance liquid chromatography-mass spectrometry (HPLC-MS). This is an exploratory approach due to the lack of established bioinformatic methodologies and dedicated databases regarding the organoleptic characteristics added to the wine by the detected analytes.

Biological aspects of Objective 6:

- Determination of metabolome of wine samples that mature in oak barrels.
- Comparison of metabolomic analytical methodologies based on their ability too reveal patterns within the wine metabolome that could explain differences between the groups of the factors time, barrel-type and bottled-wine.

Research value

Wine industry holds a significant share of the global market with revenues estimated of reaching 377 billion Euros by 2023 [58]. However, these projections cannot be easily attained if the losses due to wine spoilage cannot be controlled. Therefore, any research trying to shed light towards that direction would benefit the wine industry overall.

Furthermore, the study of microbial interactions extend also to other fields of biology with studies such as Yan *et al.* [59] showing growth inhibition of *S. aureus* by a colony of *C. pseudodiphtheriticum* in an agar plate. Similar studies have spurred the study of the human gut microbiome where, as in wine studies, it has been demonstrated that the pathogenicity of certain microorganisms, or the onset of certain diseases, has a multifactorial basis and is strongly associated with microbial interactions [60]. Although the context of the analysis of the microbial dynamics is different and field dependant, the need for the development of laboratory and bioinformatic methodologies for the analysis of microbial DNA and RNA in a biological sample appears universal.

Finally, implementation of NGS methodologies into the wine analysis is a promising practice. However, its high degrees of freedom regarding the potential choices renders the whole process quite overwhelming. Thus, the systematic development and evaluation of various analytical and statistical frameworks may serve as a way of simplifying the whole process for the subsequent studies.

CHAPTER 1

Massive sequencing: A new tool for the control of alcoholic fermentation in wine?

Dimitrios Kioroglou[1], Jessica Lleixà[1], Albert Mas[1] and Maria del Carmen Portillo[1†]

Fermentation, 4(1), 7, 2018
https://doi.org/10.3390/fermentation4010007

[1] Departament de Bioquímica i Biotecnologia, Universitat Rovira i Virgili, Marcel·lí Domingo s/n, 43007, Tarragona, Spain
[†] Corresponding author

UNIVERSITAT ROVIRA I VIRGILI
ANALYSIS OF MICROBIAL POPULATIONS IN WINES THROUGH NGS METHODOLOGIES
Dimitrios Kioroglou

Abstract

In wine industry, there is a prevalent use of starter cultures to promote a controlled and efficient alcoholic fermentation preventing the growth of spoilage microbes. However, current trends in enology aim to combine the guaranteed success of monitored process and the complexity of fermentations either by inoculating autochthonous starters or by performing spontaneously to produce distinctive wines. To understand the complex roles of microorganisms on wine fermentation, we must understand their population dynamics and their relationships with wine quality and metabolome. Current metagenomics techniques based on massive sequencing are gaining relevance to study the diversity and evolution of microbial population on every stage of the wine making process. This new tool and technique increases the throughput and sensitivity to study microbial communities. This review focuses on the current knowledge about wine alcoholic fermentation, the contribution of massive sequencing techniques and the possibility of using this tool for microbial control.

Introduction

Wine is an alcoholic beverage with great cultural and economic importance, which results from the alcoholic fermentation process. During this process, yeasts derive energy by consuming sugars that occur naturally in the grapes and at the same time produce ethanol and carbon dioxide as byproducts [2]. From the yeasts genera the most widely used, due to its fermentation capacities, is the yeast *Saccharomyces* [2], whereas non-*Saccharomyces* yeasts contribute to wine flavor, although they can also spoil wines [61–63].

In addition to the different fermentation techniques used currently by the industry, the characteristics of the wine also depend upon other factors such as climate, soil and grape variety where variation of these factors attribute to the distinctiveness of the wine [64]. Moreover, the interplay between the wine microbiota and the microbiota of the fermentation facilities has been verified but not completely understood [65]. Due to this complexity of interactions between microorganisms in the wine itself, during fermentation, but also between wine microbiota and environment, the wine industry has adopted the use of starter cultures as a mean of control and quality improvement [66]. Nevertheless, more in-depth knowledge in needed in order to understand how microbial interactions may affect the wine quality.

The diversity of the vineyard and grape microbiota has been long ago investigated via traditional microbiological techniques involving agar plate cultivation, microscopy and biochemical characterization, focusing primarily on identifying pathogenic microorganisms or microorganisms that have been associated with wine spoilage. Nevertheless, these techniques fail to identify unculturable microorganisms that comprise a considerable fraction of the wine microbiota [67, 68].

Nowadays, molecular techniques such as qPCR (quantitative polymerase chain reaction) and PCR-DGEE (polymerase chain reaction denaturing gradient gel electrophoresis) are widely used for detection and monitoring of microbial communities in wine. The former technique is more appropriate for detection and monitoring of a desired microorganism, whereas the latter for microbial community profiling. Although both techniques are supplemented with culture-dependent methods, however PCR-DGEE fails to detect species in low abudance, and qPCR suffers from scalability problems when many strains should be targeted [21, 22].

These drawbacks of the aforementioned culture-dependent and molecular techniques come to solve recent novel techniques that are based on massive sequencing, and which in recent years have been regarded as the tool of choice for studying microbial communities

during the various stages of alcoholic fermentation. Although there have been encouraging findings demonstrating the superiority of the massive sequencing over the classical methods concerning speed, sensitivity and accuracy, however most of the research has been confined to describing the constituents microorganisms and their abundance fluctuation over time. Therefore, the aim of this review, apart from exhibiting the contribution of massive sequencing to monitoring alcoholic fermentation, is to demonstrate the possibility of using this method as a tool for microbial control.

Sequencing methods

The metagenomic analysis of wine samples is mainly performed via amplicon-based sequencing which through marker-genes amplification facilitates the taxonomic and phylogenetic profiling of the microbiome [69]. After the Sanger sequencing, that was developed back in 1977 by Sanger *et al.* [70] and was considered the first generation sequencing, second (SGS) and third (TGS) generation sequencing, collectivelly referred as next-generation sequencing (NGS), have been introduced in research as fast and cost-effective solutions.

Despite the fact that TGS solves many of the disadvantages of the SGS, still is under development and not widely applied in research. From the area of SGS, which is based on sequencing by synthesis method, the most popular platforms will be presented.

Ion Torrent

Introduced back in 2010 [71], Ion Torrent sequences the template DNA strand by detecting hydrogen ions that are released during the polymerization process. As a technology, with an error rate of 1.71% [72], it does not require modified nucleotides and it generates reads of around 200 bp in length allowing for multiple runs and more data generation [73].

Pyrosequencing

The most recent variant of pyrosequencing, 454 pyrosequencing, was introduced back in 2005 [74], and was the first affordable platform allowing whole genome sequencing. As a technology, it relies on the light signal detection that is emitted after the release of phosphate during the incorporation of a nucleotide by the DNA polymerase. With an error rate below 1% [75] 454 pyrosequencing is capable of generating reads of over 400 bp in length [76].

Illumina

With the first Illumina sequencer being available back in 2006, Illumina technology is based on the usage of fluorescently labeled dNTP terminators and the detection of light signal upon incorporation. Recent Illumina machines, HiSeq and MiSeq, have decreased the error rate below 1% and are capable of generating reads of around 300 bp in length [77].

From the aforementioned platforms, Illumina is the most widely used, with 52% of the published research citing it, followed by pyrosequencing that holds 48% of the total citations [78]. However, pyrosequencing technology has been discontinued, and currently Illumina is being considered as the largest contributor to SGS.

Amplified genomic regions

Apart from choosing the most appropriate sequencing platform, researchers have to decide the genomic region that is going to be used for the taxonomic classification of the metagenomic wine sample. As far as bacteria are concerned, the 16S ribosomal RNA (rRNA) gene is the common target that is used in research for taxonomic assignment.

The 16S rRNA gene contains nine hypervariable regions (V1-V9), which all have been used as potential classification targets generating different results. For instance, Bokulich *et al.* [79] used the V4 and V5 domain so as to ascertain which one is the most taxonomically informative for profiling bacterial communities. Based on the results, the V4 domain was regarded as more suitable for profiling lactic acid bacteria (LAB), as it gave more taxonomic depth comparing to the V5 domain.

Campisano *et al.* [80] used a 700 bp region that includes the domains from V5 up to V9 in order to assess the impact of pest management on bacterial endophytic communities of Merlot and Chardonnay grapevines, with the results indicating abundance differences of operational taxonomic units (OTUs) between organic and intergrated pest management (IPM) grapevines. The same genomic region was also targeted by Perazzolli *et al.* [81] in a study of leaf microbiota, that resulted in identification of beneficial microbial communities that could be used as a tool for crop protection. In the same manner, in the past years other researchers have been focusing on other domains for classification purposes. For instance, Sundquist *et al.* [82] favored the domains V1,V2 and V4, Liu *et al.* [83] the domains V2, V3 and V4, and Chakravorty *et al.* [84] the domains V2 and V3.

Regarding fungal classification, researchers have also displayed variability concerning genomic region preference. For instance, David *et al.* [85] sequenced the 18S rRNA

gene to show that 454 pyrosequencing is much more reliable than classical techniques for studying yeast communities in alcoholic fermentation. Holland *et al.* [86] pyrosequenced the D1-D2 regions of the 26S rRNA, demonstrating that changes in arbuscular mycorrhizal fungal communities do not depend on irrigation frequency. Bokulich and Mills [87] targeted the IT1, ITS2 and the whole ITS in order to compare their classification efficiency by utilizing a mock community. Although they favored the IT1 region, nevertheless they urged for caution as none of these regions reconstructed reliably the whole mock community. Encouraging results targeting the ITS region have been yielded also from the researches of Pinto *et al.* [43] and Stefanini *et al.* [88] indicating this region as a suitable target for yeast classification.

Bioinformatic Tools

Regardless the NGS platform a researcher decides to utilize, the sequencing of wine metagenomic samples generates a significant amount of data that necessitate the use of bioinformatic pipelines. Despite the plethora of bioinformatic tools available, the most widely used will be presented.

QIIME

QIIME, which stands for Quantitative Insights Into Microbial Ecology, is a bioinformatic package, offering a variety of microbial community analyses and visualizations, that wraps other software packages with python code [89]. Some of the most frequent wrapped applications include mothur [90], blast [91], PyNAST (Python Nearest Alignment Space Termination) [92], RDP (Ribosomal Database Project) Classifier [93], FastTree [94] and USEARCH (unique word count search) [95].

MOTHUR

Mothur is a bioinformatic package that re-implements in C and C^{++} code other software packages removing that way any external dependecies during installation. Some of the re-implemented algorithms include DOTUR (Distance-Based OTU and Richness), SONS (Shared OTUs and Similarity), TreeClimber, LIBSHUFF, and UniFrac, and additioanlly the mothur team has incorporated its own analytical features to the platform [90].

MG-RAST

MG-RAST, which stands for Metagenomics Rapid Annotation using Subsystem Technology, is a server based platform with initial aim the annotation of complete or draft microbial genomes [96]. Currently, MG-RAST offers an automated solution for phylogenetic classification and functional classification of metagenomic samples.

A comparison of these three bioinformatic pipelines has been conducted by Plummer *et al.* [97] using 16S rRNA gut microbial data. The study concluded that all of the three pipelines were able to generate similar and reliable results with common limitation the ability to classify at the species level due to the type of data. The main differences between the pipelines concerned the usability and duration of analysis. MG-RAST is a more user friendly pipeline compared to the command-line based QIIME and MOTHUR, whereas QIIME required approximately 1 h to complete the analysis with MOTHUR and MG-RAST 10 h and 2 days respectively.

Databases

One of the most crucial steps of metagenomic analysis is the taxonomic classification of the microbial community. Apart from other factors, such as the sequence length, the parameters used for quality filtering and the implemented algorithm, this step can be greatly influenced by the chosen database. Currently, there are a number of highly curated databases available, such as Greengenes for 16S rRNA [98], SILVA for small (16S/18S, SSU) and large (23S/28S, LSU) subunit rRNA [99], UNITE for ITS region [100] and RDP for 16S and 28S rRNA classification [101]. However, classification based on these databases should be regarded as a rough estimation of the microbial composition as genera abundances or even taxonomic assignments can be greatly influenced by the chosen percentage of homology.

Analysis of alcoholic fermentation

There are numerous studies dedicated to the microbial analysis of wine alcoholic fermentation, but until now great focus has been given on describing microbial abundance succession during the various stages of alcoholic fermentation. These studies have attested the superiority of NGS over classical methods [85] and offered novel insights into the microbial communities.

Although bacteria are not directly connected to wine quality, acetic acid bacteria (AAB) and lactic acid bacteria (LAB) play a significant role to the final wine product.

Portillo and Mas [102], in a Grenache variety wine fermentation study, showed that AAB and LAB are more abundant than previously thought, with a dominance of *Gluconobacter* during the mid fermentation. The latter finding contradicts the previous notion that *Gluconobacter*, being alcohol sensitive, usually declines during the alcoholic fermentation [16, 103, 104]. Similar results have also been yielded in other studies of low-sulfited or unsulfited wine fermentations [105].

Additionally, NGS analysis has created the notion that apart from AAB, other bacteria, not previously described, may be present during the process. Support to this hypothesis came from Godálová *et al.* [106] in a study of Blaufränkisch and Grüner Veltliner vines, where in addition to genera already found in other studies, such as *Sphingomonas*, *Variovorax*, *Pantoea*, *Enterobacter* and *Tatumella*, new genera were detected, namely *Amycolatopsis*, *Hydrogenophilus*, *Snodgrassella*, *Telluria*, *Gilliamella*, *Lelliottia*, and *Lonsdale quercina*. However, the possible impact of these newly described genera is still to be demonstrated.

Other studies come to supplement existing ones. For instance, Bokulich *et al.* [79] showed that *Acetobacter*, *Gluconobacter*, and *Gluconoacetobacter* are dominant in winemaking processes, whereas Campanaro *et al.* [107] in a grape marc study added that *Gluconobacter* and *Gluconoacetobacter* do not survive a prolonged grape marc storage period.

Moreover, the empirically based concept of the coined term *terroir*, that is distinction of wine quality due to regional features, has been put under the prism of NGS analysis and verified recently. Results from Zarraonaindia *et al.* [108] suggested that the soil serves as a bacterial reservoir for the vines and subsequently Bokulich *et al.* [109], in a 200 commercial wine fermentations study, demonstrated the correlation of wine microbiota, wine performance and wine metabolome. These authors even predicted the metabolome of the wine from the microbial composition by using machine learning techniques [109]. Similar results have also been generated from other studies [110, 111].

Besides bacteria, NGS analysis has also given significant insights into the yeast population during fermentation. The most frequent fungi described by NGS analysis are *Saccharomyces*, *Hanseniaspora*, *Issatchenkia*, *Rhodotorula*, *Penicillium*, *Cladosporium*, *Botrytis*, *Sporobolomyces*, *Aspergillus*, *Cryptococcus* and *Pichia* [42, 111, 112], with most studies reporting high abundance of *Hanseniaspora* and *Saccharomyces* during the mid and end of the fermentation respectively. Stefanini *et al.* [88] in a Vino Santo study, found that fungal species composition undergoes a dynamic change with a declining tendency overtime, and that small changes in fermentation procedures may result in significant differences in microbial communities. As advocates to these findings come

older studies that have demonstrated that aerobic yeasts are the first to decrease in abundance, and that the mid fermentation yeast genera, such as *Hanseniaspora*, *Candida*, *Metschnikowia* and *Torulaspora*, cannot be recovered on plates at high ethanol concentration in presence of *Saccharomyces* [8, 9]. Interestingly, *S. cerevisiae*, found in very low abundance at the beginning of the fermentation, manages to rise in dominance at the end of the it. In accordance to this, Lleixà *et al.* [113] drew a comparison between the dynamics of *Saccharomyces cerevisiae* and *Hanseniaspora vineae* after inoculation in Macabeo and Merlot grape varieties. The results indicated that fermentation of *S. cerevisiae* inoculated must was faster than the one with *H. vineae* inoculation, and that inoculation with *S. cerevisiae* is necessary as *H. vineae* alone leads to incomplete alcoholic fermentation. However *H. vineae* was able to dominate the microbiota in Macabeo must but not the Merlot perhaps due to high exhibited yeast diversity of Merlot must.

Another important question that NGS analysis has been called to answer, is whether grapes are the source of spoilage microorganisms [50], or the wine-making equipment [114]. Even though there is no clear answer to this debate, studies from Suárez *et al.* [115] and Pinto *et al.* [111] seem to support the latter hypothesis.

Control of alcoholic fermentation

Controlling the alcoholic fermentation of wine-making is a very complex process. Unlike fed-batch alcoholic fermentation in bioreactors, where algorithms have been developed for the estimation of parameters that may lead to higher biomass concentrations and yield of a specific compound [116], wine alcoholic fermentation incorporates higher order of complexity, as it concerns (i) the determination of all the microbial composition throughout the fermentation process; (ii) the comprehension of the interplay between different microbial communities; (iii) the definition of a series of metabolites that contribute to the wine quality, and (iv) the integration of all these information into a predictive machine-learning model.

In the past, a series of studies have set the ground for controlling alcoholic fermentation by monitoring or modifying certain fermentation parameters, but most of the results were empirical and their interpretation was not an easy task. Various studies have shown that yeasts increase their production of volatile compounds at low fermentation temperatures [117–119]. Therefore, wine-makers that aim at enhancing wine aroma could take advantage of this factor. Another popular method, is choosing a specific yeast strain for improving specific aspects of the wine, with studies having used this technique so as to improve wine characteristics of Sauvignon [120] and Chardonnay [121]. Furthermore,

addition of certain nutrients that will prevent the fermentation from stucking, is a common practice. For instance, Cramer *et al.* [122] developed a fermentation kinetic model which showed that fermentation rate can be increased upon addition of ammonium salts, whereas Birch *et al.* [123] supported that yeast growth rate and sugar degradation could be influenced by magnesium concentrations. On the other hand, adaptive evolution approaches are aiming towards the creation of non-recombinant yeast strains that could modify wine characteristics, as for instance in the study of McBryde *et al.* [124]. Additionally non-*S. cerevisiae* yeasts are known of adding distinct flavors to the wine but due to the fact that they can easily become replaced by *S. cerevisiae*, authors such as Soden *et al.* [125] have suggested the use of mixed cultures controlled by sequential inoculation.

Although all the above practices are means of manipulating specific aspects of wine fermentation towards a specific outcome, they treat alcoholic fermentation as a black-box without controlling the microbial composition of the wine and consequently the wine quality consistency they are aiming to provide may not be certain. NGS analysis is aiming to tackle these obstacles, but as a relatively new approach so far has yielded descriptive results on the bacteria and yeast genera abundances that have been encountered during the various fermentation stages. Until now, studies from the food industry have already evinced this type of analysis as a promising strategy for the detection of previously undescribed spoiler bacteria [126, 127], underlying its suitability for controlling alcoholic fermentation. Nevertheless, NGS analysis has as an intrinsic difficulty the overwhelming amount of metagenomic analysis tools, machine-learning algorithms, databases and parameters that the researcher has to choose from. Because small changes of parameters may result in singificantly altered taxonomic assignment results [83], a possible solution may come from the use of mock communities datasets with known species compositions [128]. This strategy has already been implemented in studies such as the one by Bokulich *et al.* [129] in order to compare the performance of different classifiers. Even though mock communities datasets cannot lead to the development of a standardized NGS analysis with fixed parameters, as metagenomic samples are bound to laboratory protocols, NGS platforms, environmental and grape variety differences, they may nonetheless serve as a way to validate the robustness of a bioinformatic pipeline or as a starting point for the subsequent metagenomic analysis.

Setting a solid ground for metagenomic analysis is of paramount importance, so additional analyses such as metatranscriptomics and metabolomics can function as determinant factors for the development of system-biology networks aiming for the understanding of microbial communities interaction, and machine-learning prediction models focusing on the quality of the final wine product. With encouraging results coming from

studies such as the one by Bokulich *et al.* [109] where it has been demonstrated that microbial composition of grape must can predict wine metabolome, the future of controlling alcoholic fermentation via NGS analysis seems nothing but promising.

Conclusions

The aim of this review is to cite contemporary contributions of massive sequencing techniques to wine alcoholic fermentation, and the possibility of being used as a tool for microbial control. Wine alcoholic fermentation is a complex process that encompasses an intricate and dynamic interaction between microbial populations that leads towards the composition of a wine metabolome that defines the final wine quality and characteristics.

As a way of controlling alcoholic fermentation, the industry has adopted various techniques, such as starter cultures and process monitoring and modification, but these approaches rely on empirical results as little is known about the relationships within the wine microbiome and its correlation to the final wine product.

High-throughput sequencing, based on NGS platforms, has been presented as a metagenomic analysis tool that offers higher speed, accuracy and taxonomic resolution compared to classical culture-dependent and molecular techniques. Till now, the implementation of this technology has yielded significant yet descriptive research results on microbial dynamics in connection to the fermentation stages. Although, NGS metagenomic analysis comprises a vast amount of bioinformatic tools, databases and machine-learning algorithms, however publicly available mock communities datasets may serve as ways of algorithm benchmarking, robustness check of bioinformatic pipelines, and parameters initialization.

These mock communities and highly curated taxonomic databases could set a solid foundation for the metagenomic analysis, upon which metatranscriptomics and metabolomics will be based and provide all the necessary knowledge for the development of system-biology networks and prediction models for deciphering microbial population dynamics and prediction of final wine product, correspondingly. Regarding the latter, research has provided encouraging results highlighting the potential and benefits of massive sequencing as a tool for controlling alcoholic fermentation.

CHAPTER 2

Microbiome dynamics during spontaneous fermentations of sound grapes in comparison with sour rot and Botrytis infected grapes.

Jessica Lleixà[1], Dimitrios Kioroglou[1], Albert Mas[1] and Maria del Carmen Portillo[1†]
International journal of food microbiology, 281, 36-46, 2018
https://doi.org/10.1016/j.ijfoodmicro.2018.05.016

[1] Departament de Bioquímica i Biotecnologia, Universitat Rovira i Virgili, Marcel·lí Domingo s/n, 43007, Tarragona, Spain
[†] Corresponding author

UNIVERSITAT ROVIRA I VIRGILI
ANALYSIS OF MICROBIAL POPULATIONS IN WINES THROUGH NGS METHODOLOGIES
Dimitrios Kioroglou

Abstract

The main losses in viticulture around the world are normally associated with rotten grapes affecting both the chemical composition and the grape microbiota that later might affect the alcoholic fermentation. We analyzed the population in musts obtained from sour rotten, botrytized and healthy Macabeo grapes and the population dynamics during the spontaneous alcoholic fermentation by culture dependent and various culture independent methods including, for the first time, qPCR and massive sequencing. Grape health state affected the fermentation kinetics and also the microbial diversity and composition. Unexpectedly, the fermentation proceeded the fastest in the rotten must followed by the healthy and the botrytized grapes. As in previous studies, plate cell counts and qPCR results confirmed the increase in the number of both bacteria and fungi in the musts from damaged grapes. Massive sequencing detected higher biodiversity than the other techniques at each stage, with *Saccharomyces* and *Oenococcus* found already in the grape must. *Hanseniaspora osmophila* replaced to *Hanseniaspora uvarum* as the predominant yeast during the mid-fermentation stage for both damaged grapes. Furthermore, musts and beginning of fermentation from rotten and botrytized grapes consistently had a higher presence of the fungi *Zygosaccharomyces*, *Penicillium* and *Aspergillus* while high abundance of *Botrytis* were observed just for botrytized grapes. As expected, the acetic acid bacteria number increased in musts from rotten and botrytized grapes, mostly due to changes in proportion of the genus *Gluconoacetobacter* which remained more abundant during damaged grapes fermentation than during healthy ones. Interestingly, the presence of *Oenococcus oeni* at the end of the alcoholic fermentation was strongly affected by the health status of the grapes.

Introduction

The grape berry surface hosts a microbiota of filamentous fungi, yeast, and bacteria that can have an impact on grape and wine quality [2, 130]. When the grape surface is altered (e.g. by damaged skin of the berry, highly compact bunches, excess of humidity, phytopathogen infections) the diversity and the population sizes of the microbiota are affected and can lead to the spoilage of the berry. Grape damage of the harvested bunches and the alteration of the grape ecological balance may compromise the vinification process and the final wine quality typically adding off-flavors [41]. Thus, it is important to further investigate the microbiota diversity changes in damaged grapes and its influence on the alcoholic fermentation.

Sour rot and *Botrytis* infections are the most common causes of heavy grape berry crop losses. The sour rot affects mostly dense bunches close to harvesting and is typically characterized by vinegar odour and brown berries [53]. Disease aetiology is related with the skin rupture of the berry caused by physical factors (e.g. rain, hail, berry abrasion) or biological factors (e.g. insects, birds, moulds). The injuries on grape skin contribute to the development of yeasts and bacteria considered as the main responsible agents of this rot [131]. Moreover, insects are an important source of microorganisms that can colonize grapes and proliferate once the injury in the skin is done [132]. *Botrytis* infection (also known as grey mold) is frequent in vineyards exposed to cold and wet conditions during the ripening period [133]. In the case of sweet wines, where the presence of *Botrytis cinerea* is desired, the grapes are subjected to an extended ripening before harvesting and to a prolonged period of drying before crushing to enhance the abundance of *B. cinerea* [88].

Previous studies have documented the microbiota in sound and damaged grapes, including sour rotten and *Botrytis*-affected grapes [132, 134–136].

The results described how grape spoilage affects the grape microbiota, with damaged grapes harboring the highest yeast and acetic acid bacteria (AAB) population [134, 135, 137]. However, most of these studies use culture based techniques probably leading to underestimation of the microbial species involved. Currently, it is accepted that culture-isolated microorganisms are not necessarily representative of the microbial diversity [67, 138]. Thus, the reported species selected during grape damaged by sour rot or *Botrytis* might be biased by the composition of culture media and the capacity of the microbes to grow on them [23, 139].

Recently, several culture-independent methods based on the genetic background have been used to analyze the microbial diversity from grapes to wine (reviewed in Cocolin

et al. [140]). Generally, the use of molecular biology methods has not only endorsed the traditional results but has also been able to identify higher microbial diversity than previously expected [141]. Despite the potential of molecular techniques, we have just found one work where these were applied to study the microbial diversity of *Botrytis*-affected grapes [142]. Specifically, these authors used PCR-DGGE to monitor the yeast population changes during spontaneous fermentations of sound and *Botrytis*-affected grapes. The results included the detection of some bacterial genera not detected before in sour rot or botrytized musts like *Enterobacter*, *Bacillus* and *Staphylococcus*, some of them capable to survive in fermenting musts [142].

Among molecular methods, massive sequencing (MS) technologies are becoming a widely used methodology to characterize more precisely the microbial community of complex environmental ecosystems, including food samples [143]. For example, MS technologies have allowed metagenomic analysis of vineyard and wine microbiome deciphering which microorganisms are present with higher sensitivity than previous techniques and how their communities are affected by several magnitude factors (reviewed in Morgan *et al.* [78]).

In this study, we aim to establish the relationship between the sour rot and *Botrytis* infection affecting Macabeo grapes with specific changes on the grape microbiota. In order to achieve this objective, sound and damaged grapes were harvested and their microbial diversity monitored during subsequent spontaneous alcoholic fermentations by both culture dependent and independent methods including PCR-DGGE, qPCR and MS to weigh the biases introduced by the techniques in an effort to estimate the community changes introduced by sour rot and *Botrytis* infection.

Materials and methods

Grape samples and experimental wines

During 2016 vintage, grape clusters from the experimental vineyard of the Faculty of Oenology (Mas dels Frares, Tarragona Spain) were collected. The sampled vineyard plot produced Macabeo cultivars. Samples were collected using gloves, ethanol, sterilized scissors and sterile plastic bags. Between 10 and 12 replicate grape clusters from different plants within the plot were collected from each grape state in order to capture the heterogeneity present in the sampled lot. Samples without damaged grapes or infection signals were denominated healthy or "H", grape clusters presenting brown, damaged grapes (typical from sour rot) were denominated rotten or "R" and grape clusters with gray mold in the surface (typical from the *Botrytis* affected grapes) were denominated

botrytized or "B". H and R grapes were collected just before normal harvest, at the beginning of September and the B ones were collected two weeks later. Samples were immediately transported to the experimental cellar located 100 m away from the sampled plot and were crushed by a manual press, skins and seeds were removed by using a sieve resulting in approximately 3 L of each grape health state. 50 mL of grape juice was directly sampled corresponding to Must samples. Afterwards, as a normal procedure in the cellar, 80 mg L-1 potassium metabisulphite (40 ppm SO_2) was added to the rest of the juice. Must samples and the rest of the juice were transported refrigerated to the laboratory within the next hour. Part of the must samples was directly used for microbiological culture and the rest of the must was stored at -80°C until DNA extraction. The sulfited juice was incubated during 24 h at 4°C to allow clarification. From each health status juice, triplicates of 400 mL clarified juice were incubated at 23°C under agitation of 120 rpm in 500 mL flask and allowed to ferment spontaneously without inoculation.

Sampling and monitoring during spontaneous fermentations

The fermentation kinetics was followed considering the time needed to consume the 50% (t_{50}) and the 90% (t_{90}) of sugars. In order to easily monitor the fermentations, the density was measured daily with Densito 30PX Portable Density Meter (Mettler Toledo, Spain).

Glucose and fructose concentration was daily measured by Miura One Multianalyzer (TDI, Barcelona, Spain) using the enzymatic kit from Biosystems S. A. (Barcelona, Spain). Acetic acid and ethanol were just evaluated during late fermentation, in the last juice sampling point (when the juice density was below 1000 g/L and stable for two consecutive days). Acetic acid content was analyzed by Miura One Multianalyzer (TDI, Barcelona, Spain) using the enzymatic kit from Biosystems S. A. (Barcelona, Spain). In the case of ethanol, due to volume limitation, it was measured on the last sampling point by enzymatic method using Ethanol Boehringer Mannheim kit (R-biopharm).

Samples for plating, qPCR, PCR-DGGE and massive sequencing were taken from the must, the beginning of the fermentation (24 h after the incubation), middle fermentation (juice density between 1050 and 1040 g/L) and, finally, late fermentation when the juice density was below 1000 g/L and stable for two consecutive days.

Plate culturing

Samples for plating were serially diluted in sterile MilliQ water (Millipore Q-PODTM Advantage A10), plated on (i) YPD medium (Glucose 2%, Peptone 2%, Yeast Extract 1%, Agar 1.7%) and (ii) lysine agar medium (Oxoid, England) plates incubated at 28°C

for 48h; (iii) MRS Agar medium [144] supplemented with 4 g/L L-malic acid, 5 g/L fructose, 0.5 g/L L-cysteine, 100 mg/L nystatin and 25 mg/L sodium azide adjusted to pH 5.0 and incubated at 28°C in a 10% CO_2 atmosphere and (iv) GYC Agar (glucose 5%, yeast extract 1%, $CaCO_3$ and agar 2%, pH 6.3) supplemented with 100 mg/L natamycin and incubated at 28°C for 35 days under aerobic conditions. Appropriate dilution plates were counted. The YPD medium provided the total yeast counts, whereas the lysine agar medium is considered to provide the non-*Saccharomyces* cell counts since most *S. cerevisiae* strains have limited growth using lysine as a unique nitrogen source [145, 146]. However, it has to be considered that probably not all the non-*Saccharomyces* yeast related to wine environment are able to use lysine as nitrogen source [147]. MRS medium and GYC-Ca provided LAB and AAB counts, respectively.

DNA extraction, qPCR and PCR-DGGE

Genomic DNA was extracted from grape must and spontaneous fermentation stages using the recommended procedure for the DNeasy Plant Mini kit (Qiagen, Hilden, Germany), including three bead-beating steps for 3 min in a FastPrep-24 bead beater (MP Bio, Solon, OH) to homogenize the samples. Extracted DNA concentration was measured by nanodrop, adjusted with molecular grade water to a concentration of 50 ng/µl and stored at -20°C until further processing.

Quantitative PCR (qPCR) was performed in an Applied Biosystems 7300 Fast Real-Time PCR System (Applied Biosystems). SYBR Premix Ex Taq (Tli RNase H Plus) was used according to the manufacturer's instructions (Takara). An ABI PRISM96 well optical plate was used for the reaction. This instrument automatically determined the Ct. Yeast quantification was performed using the primers YEASTF/YEASTR for total yeast [148], CESPF/SCERR for *Saccharomyces* genus, generic CESPF/HUVR for *Hanseniaspora* genus [149], AF/ 200R for *Starmerella bacillaris* [150], TodsL2/TodsR2 for *Torulaspora delbrueckii* [151], Mp5-fw/Mp3bw for *Metschnikowia spp.* [152] and Bc3F/Bc3R [153] for *B. cinerea*. Bacterial quantification was performed using AQ1F/AQ2R primers for general AAB [154] and WLAB1/WLAB2 for general LAB [155]. All the primers anneal the ribosomal gene region. Standard curves were calculated for each type of microorganism in triplicate samples using serial dilutions of purified DNA (Supplementary Table 1 [1]).

For the PCR-DGGE analysis, the primer pairs U1GC/U2 and 341fGC/518r were used to amplify the specific U1/U2 of the 28S ribosomal region of yeast [156] and the 16S ribosomal region of bacteria [157], respectively. The DGGE procedure followed the

[1] Supplementary materials are available at https://doi.org/10.1016/j.ijfoodmicro.2018.05.016.

description in Lleixà *et al.* [113]. DNA from excised bands was re-amplified with the same primer pair without the GC-clamp and sequenced by Macrogen Company (South Korea). The BLASTN algorithm was applied to the GenBank database to identify the closest relative at species level. However, the accuracy of the taxonomic identification at species level is not accurate due to the length of the sequences.

Sequencing library construction

The library construction was done with the amplification of 1 sample for each of the musts and 2 samples for each of the fermenting points in the case of bacterial library. In the case of fungal library, 1 sample for each of the must and 1 sample of each of the fermented points were taken. The universal primer pairs 515F/806R [158] and FR1/FF390 [159] with adapters for the sequencing by the equipment PMG from Ion Torrent with chips 318 were used to amplify a region of the 16S and 18S ribosomal gene of bacteria and fungi, respectively. The use of 18S as taxonomic marker for eukaryotic genera is considered limited because many yeast species have no 18S sequence available in the databases, thus we used SILVA (v119) database as described later on because it is more updated and includes more eukaryotic genera than other databases. The universal forward primers included a 10-bp barcode unique to each amplified sample. PCR reactions contained 5-100 ng DNA template, 1x GoTaq Green Master Mix (Promega), 1 mM MgCl 2 , and 2 pmol of each primer. Reaction conditions consisted of an initial 94°C for 3 min followed by 35 cycles of 94°C for 45 s, 50°C (for Bacteria) or 52°C (Fungi) for 60 s, and 72°C for 90 s, and a final extension of 72°C for 10 min. PCR reactions were performed in triplicate for each sample replicate, pooled by sample and cleaned using a GeneRead Size Selection kit (Qiagen, Hilden, Germany). Cleaned PCR products were submitted to Centre for Omic Sciences (Reus, Spain) where their quality was checked by a Bionalyzer and their quantity adjusted for sequencing.

Data analysis

Raw sequences were demultiplexed and quality filtered using QIIME v1.9.1 [89]. Reads were discarded if the length of the read was < 200 or > 1000 and if any read contained one or more ambiguous base calls. Additionally, reads were truncated at any site containing 3 or more consecutive bases receiving quality score below 10 and reevaluating the remaining length with the aforementioned length rule. After quality filtering, 3,672,972 sequences remained with an average of 306,081 sequences per sample (Supplementary Table 2). Operational taxonomic units (OTUs) were picked by using QIIME's open-reference pipeline, where Greengenes (13_8) and SILVA (v119) were used as ref-

erence databases for 16S and 18S rRNA sequences correspondingly, at a 99% similarity threshold. The same databases and threshold have also been used for sequence alignment using PYNAST [92] and OTU taxonomy assignment [99]. The taxonomic assignment up to level species is not accurate for such a small fragment of DNA so the genera level was indicated except when the species was confirmed by qPCR and PCR-DGGE analysis. A final OTU table was created, excluding singletons (sequences observed just once), sequences detected by less of 0.001 abundance and sequences matching plant mitochondria or chloroplast. To avoid biases generated by differences in sequencing depth, bacterial and eukaryotic reads were rarefied to an even depth of 790 and 84,000 sequences per sample, respectively.

Alpha diversity (within-sample species richness) estimates were calculated by analyzing the observed OTUs, the phylogenetic diversity by the PD whole tree index, and Shannon and Simpson diversity indexes.

Statistical analysis

Beta-diversity (between-sample microbial community dissimilarity) estimates were calculated within QIIME using weighted UniFrac distances [160] between samples for bacterial sequences and eukaryotic sequences. Principal coordinate analysis (PCoA) was used to summarize and visualize patterns in species composition. ANOSIM (an analogue of univariate ANOVA which tests for differences between groups of samples) was performed in QIIME to determine significant differences in phylogenetic or species diversity among experimental factors (grape health state and fermentation stage). Kruskal-Wallis test was used to determine which taxa differed between sample groups. Taxonomic groups were considered to present significant differences in abundances across samples when False discovery rate (FDR)-corrected P values were lower than 0.05 for bacteria and $P < 0.05$ for fungi with no FDR correction due to the lack of replicated samples.

Results

Effect of grape health state on fermentation kinetics

Clarified musts from each health state, healthy (H), rotten (R) and botrytized (B), were divided into three biological replicates and allowed to ferment spontaneously (with no yeast or bacteria inoculation, Fig. 3). Sugar concentration in R and B initial musts was higher than in the H one (Table 3). Despite the higher sugar content, R fermentations were the faster to consume the 50% (t_{50}) and the 90% (t_{90}) (Table 3). Both fermentations

from H and B grapes consumed the 50% of the sugars in 5 days. However, H fermentations reached t_{90} in 11 days, 5 days earlier than the B ones that resulted in the slowest fermentations.

We also evaluated the sugar, ethanol and acetic acid concentration of the last sampling point, when density reached 1000 mg/L. In the case of H grapes, the fermentation was not complete on this time point considering the high sugar and low ethanol concentration. On the other hand, the low sugar and amount of ethanol suggested that R and B fermentations were almost finished at the last measured point. Interestingly, H and damaged grape juices presented similar acetic acid content (Table 3).

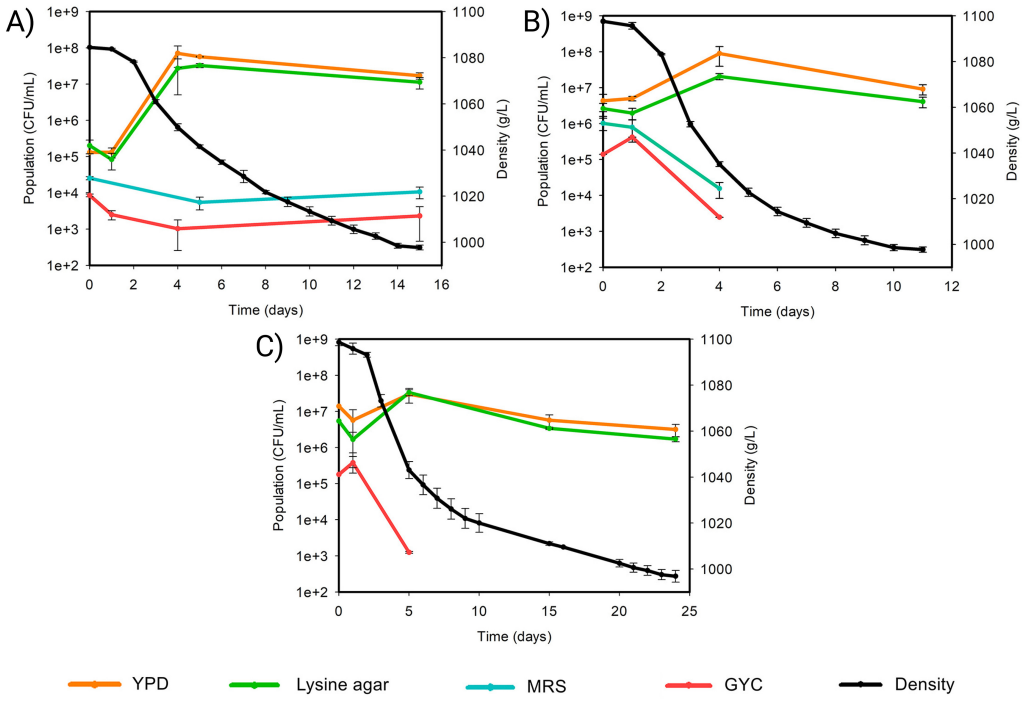


Figure 3: Fermentation density and population dynamics in YPD, Lysine Agar, MRS and GYC medium of (A) healthy, (B) rotten and (C) botrytized grapes fermentations.

Fermentation	Initial sugar content (g/L)	t_{50} (days)	t_{90} (days)	Residual sugars (g/L)	Ethanol (% v/v)	Acetic acid (g/L)
Healthy	205,26 ± 0,59	5	11	11,40 ± 1,36	11,30 ± 0,09	0,79 ± 0,07
Rotten	225,45 ± 4,12	4	7	2,97 ± 1,51	12,83 ± 0,65	0,52 ± 0,03
Botrytized	226,21 ± 1,12	5	16	5,21 ± 2,63	12,74 ± 0,54	0,79 ± 0,04

Table 3: Fermentation kinetics of healthy, rotten and botrytized grapes. The values indicate initial sugar content, sugar (residual sugars), ethanol and acetic acid concentration of the last sampling point of healthy, rotten and botrytized grapes. t_{50} and t_{90} are the time used to consume the 50% and 90% of initial sugars, respectively.

Fungal and bacterial taxonomic composition of healthy, rotten and botrytized musts and fermentations

Changes in microbial population were monitored along the alcoholic fermentation (must, beginning, middle fermentation and, finally, late fermentation (density was below 1000 g/L for two consecutive days) of H, R and B Macabeo grapes.

As we have previously mentioned, culture dependent (plate culturing in specific media) and independent techniques (qPCR, PCR-DGGE and MS) were applied and compared.

Plate culturing

The yeast population quantification was based on the colony growth in YPD (total yeast population) and lysine agar medium (most non-*Saccharomyces* yeasts) while LAB and AAB populations were quantified using MRS and GYC media, respectively. Total yeast, non-*Saccharomyces* yeast, and AAB counts were higher in the musts and the beginning of the fermentation from R and B grapes compared with the same stages from H grapes (Table 4). However, during the mid and late fermentation, yeast populations were comparable for both damaged and healthy grapes (Table 4).

LAB population was also higher in R must than in the H one. Nevertheless, LAB colonies count increased slightly through the end of H must fermentation while decreased to undetectable levels in R samples. In *Botrytis*-affected samples no LAB colonies were detected at any stage of the fermentation (Table 4).

Even if the AAB populations were higher in R and B musts samples, the number of colonies decreased through the fermentation to undetectable levels while it remained low but constant in H samples (Table 4).

Quantitative PCR (qPCR)

The population levels of total yeast, total LAB, total AAB, *Saccharomyces spp.*, *Hanseniaspora spp.*, *Torulaspora delbrueckii*, *Metschnikowia spp.*, *Starmerella bacillaris* and *Botrytis cinerea* were separately quantified by qPCR with specific primers (Table 4). The total yeast population determined by qPCR was higher in R and B musts than in H one. Apparently, the increase of total yeast in R and B musts was due to an increase in the genera *Hanseniaspora* and *S. bacillaris* while *Saccharomyces* remained at the same level than in H must. Moreover, *Saccharomyces spp.* population in R and B did not increase as much as in H during the fermentation (Table 4). The quantification of *B. cinerea* was only positive for damaged grape samples and it was considerably higher

in the B ones that were obtained from grapes visibly affected by this filamentous fungus. However, *B. cinerea* population gradually decreased through the end of fermentation (Table 4). The anaerobic conditions during alcoholic fermentation would explain the sharp decrease of *B. cinerea* in the last fermentation stages and, probably, the quantification could correspond to DNA from dead cells as no viable fungi was recovered on YPD from the mid fermentation stage onwards.

T. delbrueckii was detected in low proportion in the three musts just increasing through the mid and end of alcoholic fermentation of H samples. The last yeast species quantified, *Metschnikowia spp.*, was only detected in low proportion in B samples (Table 4).

In the case of bacteria, the quantification of AAB was at least three orders of magnitude higher in musts from damaged grapes than in the healthy ones. Finally, it was remarkable the increase of LAB population on the late fermentation of H samples (Table 4).

	Healthy				Rotten				Botrytized			
	Must	IF	MF	FF	Must	IF	MF	FF	Must	IF	MF	FF
Plate culture (CFU/mL)												
YPD	1.30E+05	1.30E+05	5.70E+07	1.70E+07	4.40E+06	5.00E+06	9.00E+07	9.20E+06	1.40E+07	5.70E+06	3.00E+07	3.20E+06
Lysine Agar	2.00E+05	8.30E+04	3.30E+07	1.10E+07	2.60E+06	2.00E+06	2.10E+07	4.10E+06	1.30E+07	1.70E+06	3.30E+07	1.70E+06
MRS	4.10E+03	2.50E+03	5.50E+03	1.10E+04	1.00E+06	7.90E+05	1.60E+04	Nd	Nd	Nd	Nd	Nd
GYC-Ca	7.00E+03	2.50E+03	1.00E+03	4.80E+03	1.40E+05	4.20E+05	2.50E+03	Nd	1.80E+05	2.60E+05	1.30E+03	Nd
qPCR (cells/mL)												
Total yeast	1.30E+06	1.90E+07	4.80E+08	4.80E+08	2.60E+08	6.40E+07	4.80E+08	1.00E+08	3.70E+07	6.20E+07	2.20E+08	2.00E+07
Hanseniaspora spp.	4.00E+05	1.80E+05	1.50E+08	5.60E+07	2.10E+07	1.10E+07	1.30E+08	1.20E+07	6.60E+06	8.80E+06	3.30E+07	1.40E+06
Sacharomyces spp.	1.60E+04	8.40E+03	2.30E+06	3.50E+07	2.70E+04	1.30E+04	7.30E+05	6.00E+06	2.60E+03	4.60E+03	1.10E+05	2.20E+06
Starmerella bacillaris	8.40E+05	7.20E+05	2.20E+07	1.00E+08	1.80E+07	9.30E+06	8.70E+07	2.50E+07	5.70E+06	6.60E+06	4.50E+07	4.30E+06
Torulaspora delbrueckii	1.20E+03	7.30E+02	5.20E+04	1.30E+05	4.70E+03	2.60E+03	3.70E+04	9.90E+03	1.50E+03	2.80E+03	2.00E+04	7.10E+02
Metschnikovia spp.	Nd	Nd	Nd	Nd	Nd	Nd	Nd	Nd	5.10E+03	8.00E+03	3.20E+03	2.20E+02
Botrytis cinerea	Nd	Nd	Nd	Nd	7.50E+04	2.10E+04	4.90E+03	1.40E+04	1.60E+07	1.80E+07	5.30E+07	1.10E+06
Total LAB	3.90E+03	1.50E+03	7.70E+01	7.50E+05	1.10E+04	2.90E+03	3.50E+02	7.60E+01	6.60E+03	1.10E+04	9.60E+02	1.50E+02
Total AAB	4.70E+04	1.30E+04	2.00E+04	4.20E+03	1.50E+07	5.90E+06	1.50E+06	8.30E+04	1.50E+08	5.30E+07	8.20E+07	1.90E+07
PCR-DGGE (-/+/++)												
Kazachstania africana [a] (4) [b]	–	–	–	–	+	+	–	–	+	+	–	–
Rizhopus stolonifer (0)	++	+	–	–	++	++	+	+	+	+	+	–
Botrytis cinerea (2)	–	–	–	–	+	+	–	–	+	+	+	+
Zygosaccharomyces bisporus (5)	–	–	–	–	+	+	+	+	+	+	+	+
Aerobasidium pullulans (0)	–	–	–	–	–	–	–	–	+	+	–	–
Issatchenkia hanoiensis (2)	+	+	–	–	+	+	–	–	+	+	–	–
Candida californica (1)	+	+	–	–	+	+	–	–	+	+	–	–
Starmerella bacillaris (0)	+	+	+	++	+	+	++	++	+	+	++	++
Hanseniaspora uvarum (0)	+	+	++	++	+	–	++	+	+	+	+	+
Saccharomyces cerevisiae (0)	–	–	+	++	–	–	+	++	–	–	–	++

Acetobacter pastorianus (4)	–	–	–	–	+	+	–	–	–	–	–	–
Gluconoacetobacter xylinus (0)	–	–	–	–	+	+	–	–	–	–	–	–
NGS (> 1% on average)												
Botrytis cinerea	0	0.3	0	0	0	0	0	0	40.6	36.4	2.3	0
Cladosporium	4.4	3.3	0	0	0.5	0.8	0	0	3	2.2	0.2	0
Hanseniaspora osmophila	3.8	3.8	1.2	2.5	2.3	5.1	47	25.7	6.4	7.9	53.1	3.1
Hanseniaspora uvarum	10	36.9	95.4	25.8	6.9	12.1	36.7	8.5	14.3	16.4	30.4	1.9
Aspergillus	14.9	3.9	0	0	23	22.4	1.6	0.8	10.3	9.8	1.4	0
Rhizopus	52.3	21.6	0	0	51.1	39.2	2.1	0.6	4.9	4.5	0.6	0.1
Saccharomyces cerevisiae	1.1	4.7	2.1	59.9	0.5	0.8	1.5	50.2	0.5	0.4	0.6	86.6
Starmerella spp.	1.5	14.2	1	6.2	1.6	2.1	5.4	4.7	0.7	0.8	3	0.5
Penicillium	0	0	0	0	3.3	3.1	0.2	0.1	5.2	4.4	0.5	0
Zygosaccharomyces	0.2	0.7	0.1	2	6	9.2	5	8.9	6.5	9.8	6.9	7.5
Acetobacter	1.5	2.9	0.3	3.4	3.6	4	4.2	4.7	1.5	1.2	2	2.6
Ameyamaea	0.4	0.3	0	0	4.5	4.8	5.4	5.1	1.3	1.6	1.8	0.5
Gluconacetobacter	9.5	11.9	2.2	0.2	45.1	44.7	48.7	42.3	34.4	36	28.9	14.6
Gluconobacter	67.2	52.7	88.6	5.3	42.9	41.8	34.6	41.4	52.3	49.8	44.5	52.3
Tanticharoenia	0.4	0.8	0	0	3.1	3.3	5.5	0.6	8.1	8.3	8.6	0.2
Oenococcus	15.1	19.2	4.1	90.9	0	0.1	0.2	2.3	0.1	0.1	9.6	17.8

[a] Closest relative according to BLAST analysis.
[b] Sequence difference (number of substitutions/indels) from type strain.

Table 4: Abundance and detection of the fungal (upper) and bacterial (lower) groups determined by the different compared techniques. Results are the mean of three different biological replicates evaluated from the must (Must), the beginning of the fermentation (IF), the middle of the fermentation (MF) and the final sampling point (FF) taken when juice density was below 1000 g/L for two consecutive days. Nd for not detected.

PCR-DGGE

Microbial communities from H, R and B grapes were analyzed at different fermentation stages by PCR-DGGE for Eukarya. The excised DGGE bands were re-amplified and identified by sequencing. Occasionally, some bands migrating differently were identified as the same taxon. Though the identification to species level from the short sequences obtained by PCR-DGGE analysis was not reliable, we were able to detect eight different yeast species (closest relatives: *S. cerevisiae*, *Hanseniaspora uvarum*, *S. bacillaris*, *Candida spp.*, *Issatchenkia spp.*, *Kazachstania spp.*, *Zygosaccharomyces spp.* and *Aureobasidium pullulans*) and two filamentous fungi (*Rhizopus spp.* and *B. cinerea*). *Saccharomyces cerevisiae* was not detected with this technique in any grape must (Table 4). However, it was detected during mid and late fermentation in H, R and B fermentations. Moreover, the highest intensity of *S. cerevisiae* was reached at late fermentations regardless of grape health status. *Hanseniaspora uvarum* and *S. bacillaris* exhibited a similar behavior being present along all grape fermentations and showing higher band intensity from mid to late fermentation (Table 4).

Kazachstania spp., *Zygosaccharomyces spp.*, *B. cinerea* and *A. pullulans* were just identified in damaged grape samples. Concretely, *Kazachstania* was detected just in the must and the beginning of the fermentation while *Zygosaccharomyces* was present in all stages. Besides, *B. cinerea* was observed during all B fermentation phases while it was just detected in the must and the beginning of R fermentations. As previously mentioned, the detection of *B. cinerea* in the last fermentation phases could correspond to DNA from dead cells. In the case of *A. pullulans*, this yeast like fungus was only identified in the first part of B grape fermentation (Table 4).

Apart from *B. cinerea*, we observed another filamentous fungus identified as *Rhizopus spp.* This fungus was present in all grape musts and it was detected until the middle of the R fermentation and late fermentation of B (Table 4).

The PCR for the DGGE analysis with bacterial specific primers did not result in strong amplifications indicating less proportion of bacteria in comparison with yeast population (results not shown). The different DGGE bands from bacterial profiles were excised and amplified for their identification, but most of the resulting sequences did not have a match on the NCBI database probably due to co-migration of bands from similar species and thus, cloning of the excised bands should have been done in order to have single sequences from co-migrating bands. Just two bands recovered from must and the beginning of the fermentation of R grapes were identified as *Gluconoacetobacter* and *Gluconobacter* (Table 4).

Massive sequencing

Barcode amplicon sequencing was used to analyze the bacterial and fungal communities of the different grapes through their fermentation. A total of 382,990 bacterial sequences and 1,954,049 eukaryotic sequences were used to build the OTU tables with an average of 31,916 and 162,837 sequences per sample, respectively (Supplementary Table 2). The massive sequencing analysis detected a higher diversity of fungal and bacterial genera than the other techniques (Table 4 and Supplementary Fig. S1). However, considering those genera more abundant than 1% on average, 9 fungal and 6 bacterial genera were detected (Table 4).

The most abundant yeast on average across all samples was *Hanseniaspora* (38.2%), detected mainly in the beginning and mid fermentation (Fig. 4). Interestingly, two different abundant OTUs within *Hanseniaspora* were identified and the closest relatives were *H. uvarum* (23.1%) and *H. osmophila* (15.1%). *Hanseniaspora uvarum* was more abundant in H than in R or B samples, while *H. osmophila* was more abundant in mid fermentations of R and B (Fig. 4). Other non-*Saccharomyces* yeast were detected in less proportion on average, for example, *Starmerella* (3.3%), and *Zygosaccharomyces* (5.3%) (Fig. 4). *Saccharomyces* (19.8% on average) was detected in all musts and every stage of fermentations, being the predominant yeast (between 50.2 and 59.9% of sequences) during late fermentations samples. Yeast like *Hanseniaspora* or *Saccharomyces* quickly replaced to filamentous fungi or molds detected by this technique in the first stages of the fermentation. Within these molds, *Rhizopus* was abundant (13.6% on average) in H and R, while *B. cinerea*, (6.1% on average) predominated in must and beginning of the fermentation of B grapes (ranging between 36.4 and 40.6%), *Aspergillus* (6.9% on average) was more abundant in R must at the beginning of the fermentation (23 and 22.4%, respectively) than in the rest of the samples, *Penicillium* was just detected in damaged samples (ranging between 3.1 and 5.2% and *Cladosporium* (1.1%) slightly more abundant on H and B musts and the beginning of fermentation (ranging between 2.2 and 4.4%) than in the respective R samples (0.50.8%). Other fungal genera detected in lower proportion than 1% but higher than 0.1% on average across all samples are indicated on the heatmap (Supplementary Fig. S1). Some of these low abundance genera were present just in samples from damaged grapes, like *Saccharomycopsis*. On the other hand, *Fusarium* was detected just in H samples. All these taxa, except the fermentative yeast *Saccharomycodes*, disappeared at the late fermentation, indicating a low implication during wine fermentation.

In the case of bacteria detected by MS, the 6 most abundant genera were the AAB genera *Gluconobacter*, *Gluconoacetobacter*, *Acetobacter*, *Tantiocharoenia*, and *Ameya-*

maea (accounting for 82.4% on average across all samples) and the LAB genus Oenococcus. The abundance of these bacterial genera varied among the samples with different health states (Fig. 4B). *Oenococcus* was predominant during late fermentation of H (90.9%) and also represented an important proportion of the sequences during the rest of H fermentation stages while it was scarcely detected in damaged grapes samples (Fig. 4). R and B samples harbored higher proportion of *Gluconoacetobacter* than H samples and the genus *Gluconobacter* was clearly the most abundant from the must to mid fermentation of H grapes (52.7 - 88.6%). In addition, sequences related to *Tantiocharoenia* were more abundant in damaged samples than in H ones. Finally, R samples harbored higher proportions of the genera *Acetobacter* and *Ameyamaea* than H or B. Other bacterial genera detected in lower abundance than 1% but higher than 0.1% are listed on Supplementary Fig. S1. Within these genera, some LAB like *Aerococcus*, *Lactococcus* or *Streptococcus* were also identified. All these genera disappeared during late fermentation of H grapes while some of them remained in damaged grapes (Supplementary Fig. S1). In addition, some of the genera detected just at late fermentation of R and B grapes increased their abundance with respect to the must and beginning of fermentation samples, for example, *Acinetobacter*, *Bacillus*, *Staphylococcus* and *Tatumella*.

Bacterial and fungal alpha diversity

The highest microbial diversity as determined by the number of different genera identified by the PCR-DGGE analysis was observed in the must and at the beginning of the

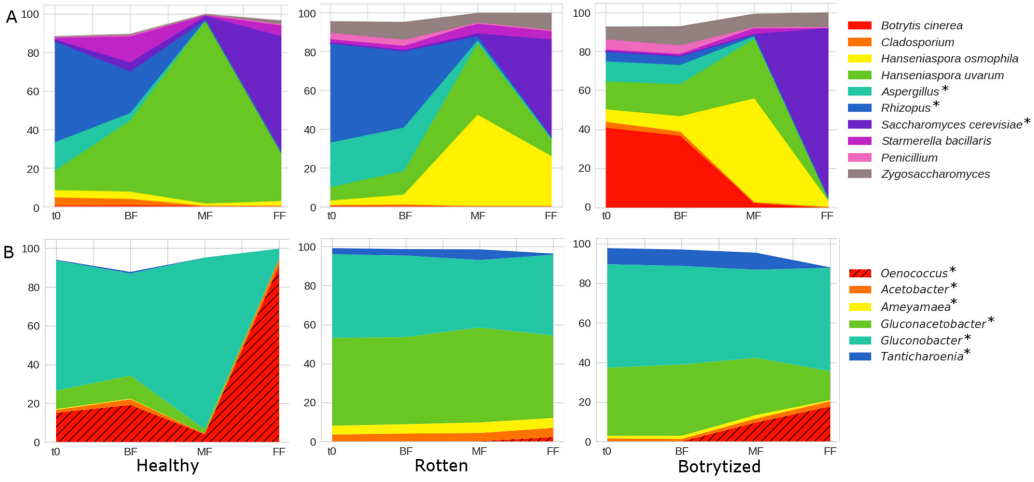


Figure 4: Relative abundance of fungal (A) and bacterial (B) taxa detected at $> 1\%$ by MS. Taxa that differed significantly (P value < 0.05) by fermentation stage (for fungi) or by health status (for bacteria) are indicated by an asterisk.

fermentation of each health type grape, with higher diversity in the must samples and also more diversity in R and B samples than in H ones (Table 4). The diversity was lower through the end of H fermentation that for the damaged ones (Table 4).

According to MS analysis, fungal diversity ranged from 1 to 1.6 for the PD whole tree index and from 42 to 68 observed OTUs (Fig. 5A, Supplementary Table 3). Higher diversities were reached for all samples in the musts and during the first stages of fermentation with similar values for damaged and H samples in those stages. However, diversity decreased sharply for H samples during the second half of the fermentation while, in the case R samples, diversity remained high and relatively constant along the fermentation and in the case of B samples, diversity decreased just a little during late fermentation stage (Fig. 5A). The lowest fungal diversity belonged to late fermentation of H grapes. Simpson and Shannon indexes pointed to H samples during the mid-fermentation as the ones with the lowest diversity values (Supplementary Table 3).

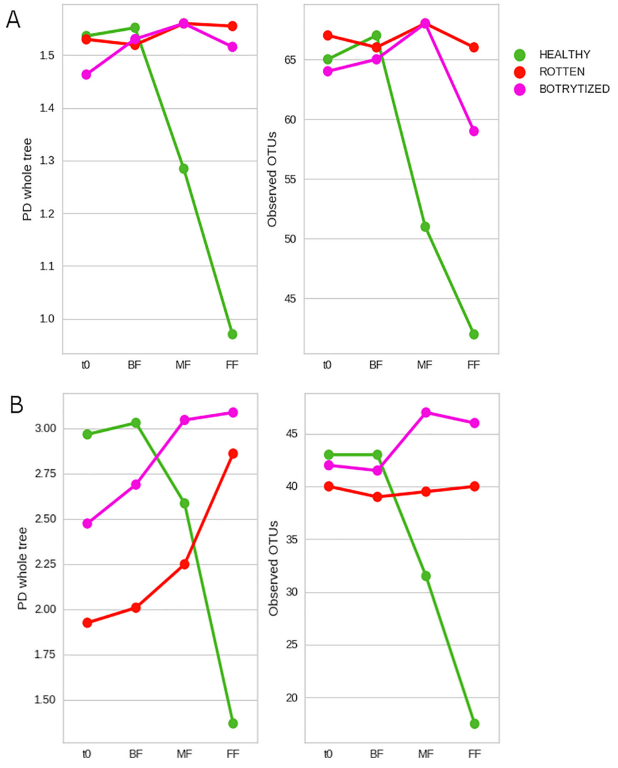


Figure 5: Alpha diversity graphs showing the PD whole tree index (left) and number of different OTUs (right) for the fungal (A) and bacterial (B) communities determined by MS.

This scenario was slightly different for bacterial diversity. The PD whole tree index for bacterial sequences ranged from 1 and 3.1 while observed OTUs ranged from 15 to 49 (Fig. 5B, Supplementary Table 4). The bacterial taxonomic diversity increased

through the end of fermentations for damaged samples while decreased sharply from the first 24 h to late fermentation for H samples. The same tendency was observed for the number of OTUs with the exception of R samples harbored a relatively constant number of OTUs through the fermentation (Fig. 5B). Other indexes like Simpson or Shannon also revealed that the lower bacterial diversity was observed for Healthy samples from mid to late fermentation while the values for the other samples remained quite constant (Supplementary Table 4).

Health status of the grapes influences must and fermentation communities

Fungal and bacterial communities changed across the different fermentation stages and between the different health statuses of the grapes used for the alcoholic fermentation (Fig. 6). Unifrac distance matrices [160] were calculated with the taxonomic composition and abundance data from samples analyzed by MS in order to be used for the analysis of similarities (ANOSIM) of the microbial communities from the different samples. Fungal communities resulted significantly different and clustered by the different fermentation stages (Table 5, Fig. 6A). According to statistical analysis, *Aspergillus*, *Rhizopus* and *Saccharomyces* were the fungal genera that varied significantly across all fermentation stages, thought other additional genera showed variation in their proportions across the fermentation stages and different health statuses (Fig. 4). Bacterial populations from the different samples of H, R and B were significantly different and clustered by health status (Table 5 and Fig. 6B). The bacterial genera that varied significantly in abundance between the H, R and B samples were *Acetobacter*, *Aeyamaea*, *Gluconoacetobacter*, *Gluconobacter*, *Oenococcus* and *Tanticharoenia* (Fig. 4).

Factor	ANOSIM bacterial		ANOSIM fungal	
	R	P	R	P
Health	0.355	0.001	0.013	0.356
Ferm. stage	0.005	0.455	0.598	0.003

Table 5: ANOSIM results showing the analysis of similarities of the different fungal and bacterial communities calculated from the weighted Unifrac distances matrices for the factors health status of the grape and the fermentation stage.

Discussion

Grape health status is a primordial fact during winemaking and it can negatively impact on the fermentation process and the composition and quality of wine [2]. In the present work, we described the ecological changes along the fermentation of Macabeo grapes with different health status, H, R and B, using various techniques.

Analyzing the influence of grape health state on fermentation kinetics, *Botrytis* infection had the strongest effect on the delay of the fermentation evolution since fermentations affected by this fungus were the slowest to consume 90% of the sugars. Nevertheless, undamaged grape fermentation presented the highest amount of sugar and ethanol in the last sampling point (density below 1000 g/L for two consecutive days). Previous studies have reported higher residual sugar when non-*Saccharomyces* yeasts were abundant during alcoholic fermentation [161, 162], which can occur in spontaneous fermentations [163, 164].

In the present study, the techniques of plate culturing, qPCR, PCR-DGGE and MS have been used to monitor the changes of microbial community on grapes with three different health statuses. All these techniques allowed for the differentiation of the microbial communities in musts and fermentations of the three types of grapes, but differences in the results were observed depending on the technique.

Most of the studies on sour rot and *Botrytis*-affected grape ecology have been based on plate counts [53, 134–136, 142]. However, the inability of some microorganisms to grow in some media and/or under certain conditions [67] can give a biased result of

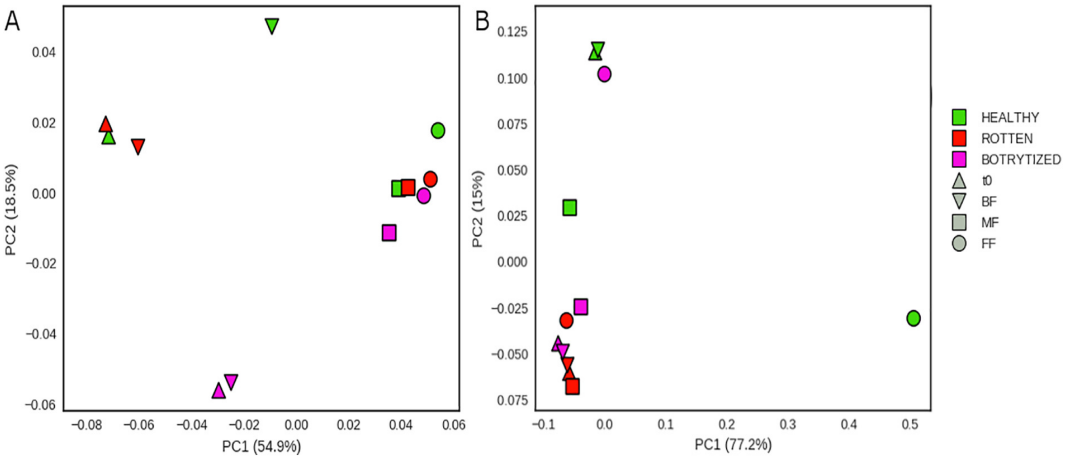


Figure 6: Weighted Unifrac distance PCoA plots for fungal (A) and bacterial (B) communities from Macabeo must and fermentations.

the microbial diversity [138]. Considering these facts, we additionally used molecular methods since they have shown to be more informative about environmental microbial diversity.

One of the most used molecular techniques to quantify microbial populations is the qPCR. Nevertheless, the specific primer design limits the quantification to the targeted groups or species. The PCR-DGGE using general primers is a good molecular technique to obtain a fingerprint of the microbial community in a sample, but hardly detects populations with lower density than 10^3 CFU/mL or two orders of magnitude lower than the most abundant members [22, 163, 165]. Recently, high-throughput sequencing or MS techniques can be used to obtain a more detailed image of the microbial communities of various ecosystems, including food processing [143]. To our knowledge, this is the first study analyzing microbial populations in sour rot or *Botrytis*-affected grape musts and fermentations by qPCR and MS. It is important to consider that the used molecular methods detect both viable and non-viable cells. Thus, it is possible that DNA of dead or non viable cells lead at some point to the overestimation of a taxonomic group. However, an increase in the proportion of DNA probably will correspond to an increase of the population.

In general, our plate counts agreed with similar studies analyzing sound and damaged grapes with a higher fungal and bacterial population in the affected grapes [130, 134, 166].

Total yeast and AAB bacteria quantification by qPCR was higher than the counts detected by plating probably due to the quantification of viable but non culturable and dead cells by qPCR [19, 149]. In addition, the primers used to quantify total yeast have been described to also detect many filamentous fungi apart from yeast [148].

Fungal communities varied significantly across the different fermentation stages as shown in this study by the used techniques. As in previous studies, the yeast population number and diversity resulted higher in damaged grape musts than in H one [134, 137, 142, 167]. The higher yeast number might have been induced by physically damaged grapes [134, 167] together with the release of nutrients from the berry that encourage their growth [130].

The high proportion of non-*Saccharomyces* in damaged musts, determined by plate counts, qPCR and MS, could interfere with *Saccharomyces* imposition along the fermentation as a consequence of interactions between both populations. Among others, these interactions involve the competition for substrate, yeast-yeast cell contact or the release of antimicrobial compounds [168, 169]. However, R must presented higher difference between total yeast and non-*Saccharomyces*, indicating a higher initial concentration of

Saccharomyces, which could explain why the R microbial population was the fastest to consume the 90% of the sugars.

Higher populations of *Hanseniaspora* and *Candida* (or *Starmerella*) observed by qPCR in damaged grapes coincided with previous ecological studies on damaged grape berries [22, 25, 170]. Nevertheless, these species are also predominant worldwide in healthy grapes and during the first stages of fermentation [4, 7, 25, 171]. In our study, independently on the grape status, *H. uvarum* and *S. bacillaris* were detected in high proportions across the alcoholic fermentations by qPCR and PCR-DGGE. A previous study using PCR-DGGE to monitor yeast populations during sound and *Botrytis*-affected fermentations [136] found a similar behavior of *H. uvarum* to what we describe by PCR-DGGE but they did not found *Saccharomyces*. In our study, *Saccharomyces* was not detected in any must sample by PCR-DGGE but was detected from mid fermentation of damaged and H grapes fermentation. MS together with qPCR allowed *Saccharomyces* detection and quantification, respectively, in all samples from the must onwards.

MS also enabled us to identify abundantly *H. uvarum* sequences in all fermentations, but above all, during the middle of H fermentations. Despite the low proportion of *S. bacillaris* identified by MS, its quantification by qPCR was proportional to the values obtained for *H. uvarum*. Other sequences related to *H. osmophila* were more abundant than *H. uvarum* in damaged grapes but there were not detected or differentiated by PCR-DGGE or qPCR techniques. A previous study on Dolce wine fermentation was able to differentiate *H. osmophila* from *H. uvarum* by PCR-DGGE analysis [170], indicating that their bands migrated differently. However, differentiation of different species of *Hanseniaspora* is not possible by qPCR with the used primers.

Some key yeast species previously associated with damaged grapes as *Botrytis*, *Kazachstania* and *Zygosaccharomyces* [136, 137] were observed by PCR-DGGE just in R and B samples. Barata *et al.* [53] proposed as biomarkers for sour rot the presence of the yeast *Zygoascus hellenicus* and *Issatchenkia*. However, in our case, *Zygoascus* was not detected by PCR-DGGE in any sample while *Issatchenkia* was detected in both damaged and H samples by both PCR-DGGE and MS techniques.

B. cinerea was detected by qPCR and PCR-DGGE in damaged samples, although its quantification in R samples was low and constant. In contrast, MS analysis revealed a very small proportion of *Botrytis* in H must and even lower proportion in R grapes. In B samples though, *Botrytis* represented $> 30\%$ of the sequences analyzed by MS in the must and beginning of the fermentation. This proportion lowered close to the end of the fermentation evidencing the sensibility of this fungus to the semi anaerobic conditions and the increasing concentration of ethanol along the fermentation [41].

Rhizopus and *A. pullulans* were detected also by both DGGE and MS techniques. *Rhizopus* was present in all must and beginning stages. This fungus has been described as a saprophytic organism that can be a secondary bunch rot invader infecting grape berries [41] and lead to organoleptic defects in grapes and wines when is associated with *B. cinerea* [172]. Furthermore, *A. pullulans* was identified in must and initial fermentations of H and damaged grapes by MS analysis. This coincides with previous studies where it was isolated from both sound and damaged grapes [22, 137, 173].

In general, MS analysis revealed a higher number of yeast genera than PCR-DGGE. The higher proportion of some yeasts could inhibit or impede the detection of other less abundant yeasts or microorganisms by PCR-DGGE [22]. However, some of the genera that PCR-DGGE failed to detect were more abundant than 1% on average as determined by MS (for example, *Aspergillus*, *Penicillium* and *Cladosporium*). On the other hand, PCR-DGGE analysis detected additional genera that were not abundant or even not detected by MS, like *Kazachstania*, *Issatchenkia* or *Candida*. These differences in the detection of genera by both techniques might be due to PCR amplification preferences as the primers used for DGGE and MS were not the same. Recent studies using MS technique to analyze the wine fermentation process of different grape varieties have detected the fungal genera *Hanseniaspora*, *Issatchenkia*, *Rhodotorula*, *Penicillium*, *Cladosporium*, *Botrytis*, *Sporobolomyces*, *Aspergillus*, *Cryptococcus* and *Pichia* [42, 44, 111], all of them also detected in the present study and making our fungal community results solid (Table 2, Supplementary Fig. S1).

LAB and AAB are the most relevant bacterial groups related to grapes and wine fermentation. In this study, GYC and MRS media were employed to count AAB and LAB populations, respectively. As in previous studies, our plate counts revealed an evident increase of AAB population in musts and beginning of the fermentations of R and B grapes [134, 136, 137, 142, 167] . As explained above, the reason could be the release of nutrients from the berry that encourages AAB and yeast growth [130]. The evaluation of AAB population by plate culture is usually complicated [19, 174] mainly for its ability to enter in VBNC (viable but non-culturable state) [139] or because they die under inappropriate conditions. Thus, the use of specific primers to quantify AAB by qPCR [154] allowed us to detect higher populations of AAB in all fermentations than the plate culturing, indicating the capacity of qPCR to detect VBNC and dead bacteria (Table 2). In order to identify the AAB genera, PCR-DGGE and MS techniques were applied. Unfortunately, PCR-DGGE allowed the identification of just *Gluconoacetobacter* and *Acetobacter* in the must and beginning of R fermentation. This might be due to the limitation of PCR-DGGE to detect populations two orders of magnitude lower than

the most abundant members [22], and, as noticed by qPCR results, yeast population was mostly two or even three orders above the bacterial one. On the other hand, MS technique allowed the identification of up to 21 bacterial genera, most of them related to AAB genera. Clear differences in bacterial composition were detected between H and damaged grapes. *Gluconobacter* followed by *Gluconoacetobacter* were the most abundant until the mid-fermentation of H grapes. In R grapes though, *Gluconoacetobacter* represented the most abundant genus in all fermentation stages and it was also more abundant in Botrytized samples than in H ones. Thus, the abundance ratio between *Gluconobacter* and *Gluconoacetobacter* was higher in H samples than in R and B ones. This fact is really aligned with previous observations where the health status of the grapes indicated that *Gluconobacter* is more abundant in healthy grapes, whereas *Gluconoacetobacter* (or even *Acetobacter*) are more abundant in damaged grapes [13].

In a wine fermentation study in a Grenache variety using MS [102], we showed that AAB and LAB were more abundant during fermentation than previously thought, with a dominance of *Gluconobacter* during the mid-fermentation. The latter finding contradicts the previous notion that *Gluconobacter*, being alcohol sensitive, usually declines during the alcoholic fermentation [103, 104, 175]. Similar results have also been reported in other studies using MS analysis on low-sulfited or unsulfited wine fermentations [176]. The same authors found *Acetobacter*, *Gluconobacter*, and *Gluconoacetobacter* as dominant bacteria during winemaking processes [79].

Plate culturing also allowed the quantification of considerable LAB populations in H and R musts, contrasting with previous studies where LAB populations were not detected or detected in low concentrations [137, 142]. In fact, our MRS counts overestimated LAB population respect to those of qPCR analysis using LAB specific primers, probably due to non-LAB species that may grow in MRS media [137]. However, both techniques detected a LAB population increase at the H late fermentation.

MS analysis also supported qPCR results, with low percentages of LAB taxa in comparison with those of AAB populations, except during the last sampled point of H fermentation, making solid this tendency. This LAB population increase at late H fermentation suggests that the spontaneous evolution of malolactic fermentation might not occur spontaneously in damaged grapes fermentations. No LAB genus was identified by the PCR-DGGE technique but MS analysis deciphered the LAB community composition in our samples and the main player was *Oenococcus oeni*. Thus, to our knowledge, this is the first study relating the presence of *Oenococcus oeni* to the grape health status.

The diversity of fungal and bacterial communities as determined by the quantification of the identified genera by DGGE gels resulted in higher diversity in the musts and

beginning of every type of fermentations and, in general, higher diversity for damaged grapes. However, according to MS results, similar diversity indexes were observed for the different types of grapes in the musts and the health status of the grapes influenced on fungal and bacterial diversities in a different way. For the fungal communities, H fermentation decreased the diversity through the last fermentation point while R and B samples remain almost constant. In the case of bacterial communities, the diversity declined sharply along H fermentation while increased (PD whole index) or remained relatively constant (number of OTUs) for damaged grapes fermentations. Both DGGE and MS analysis suggest that the microbial diversity of must obtained from H grapes decreased along the alcoholic fermentation while musts from damaged grapes maintain or increase their diversity. The higher diversity during the mid and late damaged fermentations may result from the additional metabolisms present in the infected grapes and musts, making possible the survival of non-conventional yeast for longer time respect to H fermentations but also, compromising the success of the alcoholic fermentation or including off-flavours to the final wine.

Conclusions

The present study is the first to include the molecular techniques qPCR and MS to evaluate the population evolution along spontaneous fermentation of sour rot and *Botrytis* affected grapes in comparison with healthy grapes. Both culture and molecular based analyses showed differences in fungal and bacterial communities of Macabeo grapes depending on its health status. However, MS analysis provided higher diversity at each stage than the other compared techniques and detected *Saccharomyces* and *Oenococcus* even in the initial must samples. The main differences in the fermentations revealed by MS were that *H. osmophila* was predominant during mid-fermentation of damaged samples instead of *H. uvarum*. Besides, *Oenococcus oeni* and *Gluconobacter* were more abundant in healthy samples than in damaged ones, while the later had higher proportion of *Gluconoacetobacter* with respect to the healthy samples. The microbial diversity of healthy fermentations decreased from the middle to the end. Similarly to other studies that used MS to describe the microbial population, in this work MS was the technique that contributed the most in the deciphering of the community microbiome and for the first time, the health status of the grape was related to the relative abundance of *Oenococcus oeni* during the alcoholic fermentation.

UNIVERSITAT ROVIRA I VIRGILI
ANALYSIS OF MICROBIAL POPULATIONS IN WINES THROUGH NGS METHODOLOGIES
Dimitrios Kioroglou

CHAPTER 3

Evaluating the effect of QIIME balanced default parameters on metataxonomic analysis workflows with a mock community.

Dimitrios Kioroglou[1], Albert Mas[1] and Maria del Carmen Portillo[1†]

Frontiers in microbiology, 10, 1084, 2019
https://doi.org/10.3389/fmicb.2019.01084

[1] Departament de Bioquímica i Biotecnologia, Universitat Rovira i Virgili, Marcel·lí Domingo s/n, 43007, Tarragona, Spain

[†] Corresponding author

UNIVERSITAT ROVIRA I VIRGILI
ANALYSIS OF MICROBIAL POPULATIONS IN WINES THROUGH NGS METHODOLOGIES
Dimitrios Kioroglou

Abstract

Metataxonomic analysis represents a fast and cost-effective approach for acquiring informative insight into the composition of the microbiome of samples with variable diversity, such as wine samples. Nevertheless, it comprises a vast amount of laboratory procedures and bioinformatic frameworks each one associated with an inherent variability of protocols and algorithms, respectively. As a solution to the bioinformatic maze, QIIME bioinformatic framework has incorporated benchmarked, and balanced parameters as default parameters. In the current study, metataxonomic analysis of two types of mock community standards with the same microbial composition has been performed for evaluating the effectivess of QIIME balanced default parameters on a variety of aspects related to different laboratory and bioinformatic workflows. These aspects concern NGS platforms, PCR protocols, bioinformatic pipelines, and taxonomic classification algorithms. Several qualitative performance expectations have been the outcome of the analysis, rendering the mock community a useful evaluation tool.

Introduction

During the past years significant improvements in Next Generation Sequencing (NGS) platforms and computational performance have given a considerable momentum to the research of microbial communities. Primarily there are two sequencing-based methods for the classification analysis of a microbiome, the metagenomic approach which concerns the shotgun sequencing of microbial DNA, and the metataxonomic approach which refers to the sequencing of a marker gene, having as a usual target the ribosomal RNA gene [26]. Due to the cost-effectiveness and decreased demands on computational resources of the latter, it has been used quite broadly in research and consists the focus of the current study.

A typical metataxonomic analysis includes a process that combines laboratory and bioinformatic workflows. The steps involved in the laboratory process concern the collection of a microbiome sample, the DNA extraction, the library preparation based on the preferred rRNA gene marker and the massive sequencing with the NGS platform of choice. The bioinformatic workflow concerns the quality filtering of the resulted data, the clustering of sequences based on a specific clustering strategy and the taxonomic assignment to the representative sequence of each cluster.

There are a plethora of bioinformatic frameworks for the analysis of the microbiome data with Quantitative Insights Into Microbial Ecology (QIIME) being one of the most popular and thus, implemented in the current study [89, 177]. As a bioinformatic framework, it contains a significant amount of algorithms and parameters to select and tweak, respectively, but studies such as Bokulich *et al.* [34, 178] have provided informative and useful benchmarks with the resulted balanced parameters being incorporated into QIIME as default parameters. Nevertheless, microbiome samples are subjects to different laboratory procedures and protocols and as such implementation of parameters must be evaluated. For that reason, a mock community, which represents a microbiome sample of known composition [128], consists a valuable tool in assessing both laboratory and bioinformatic workflows prior to establishment of parameters. There are many studies dedicated to mock communities, such as Yuan *et al.* [179] where a mock community was used for the comparison of six common DNA extraction protocols, or Yeh *et al.* [180] where mock communities were the tool for the establishment of a methodology that could verify similar performance between sequencing runs. However, the way that the current study differs from the rest is based on the fact that the main focus is given on assessing the effectiveness of QIIME balanced default parameters on our laboratory and bioinformatic workflows destined to the metataxonomic analysis of wine samples.

Wine samples are characterized by extremely dynamic microbial populations. During wine ageing, these populations tend to be quite sparse with most of the microorganisms being difficult to detect as they enter the viable but non-culturable state (VBNC) [139], and thus making NGS technology the most appropriate detection tool. Therefore, sparse microbial communities are quite important since wine spoilage microorganisms may go undetected due to their low abundance and significantly alter the wine quality later on. For that reason, the mock community in the current study was chosen to be simple. Additionally to the main focus, the mock community will serve a double qualitative role on a series of aspects related to our workflows. Regarding the laboratory procedure, to evaluate 16S metataxonomic analysis on data produced by Ion Torrent and Illumina platforms, the impact of 18S and ITS amplicons on the metataxonomic classification and the effect of the PCR cycles during the library preparation on the downstream bioinformatic analysis of the Ion Torrent data. As far as the bioinformatic analysis is concerned, the mock community will assist in ascertaining the impact on classification of different quality filtering thresholds, the performance of different sequence clustering methods and the classification performance of two different algorithms. Moreover, we are examining the possibility of utilizing the confidence of the assigned taxonomy, as reported by the classification algorithms, as a tool for eliminating false positives.

Methods

Laboratory workflow

Two microbial community standards from ZymoBIOMICS^{TM} with the same microbial composition of 8 prokaryotes and 2 eukaryotes and impurity level $< 0.01\%$ have been used. The first standard contained DNA extracted from pure cultures (DNA standard D6305 200 ng), whereas the second standard was constructed by pooling pure cultures (Microbial Community standard D6300). The microbial species along with the 16S theoretical relative abundance, as provided by the standards specifications, are given in Table 6. The theoretical relative abundances have been calculated by the standards provider taking into consideration differences in the number of copies each amplicon has among the species. However, such correction is rendered impossible when estimating relative abundances in real wine samples. Therefore, the estimated relative abundances have not been corrected in order to examine the amount of deviation between estimated and ideal relative abundance. The aim of using the DNA standard (DS) was to assess the performance of different PCR primers and amplicons used with the NGS platforms, the impact of PCR cycles on the number of chimeric sequences in the Ion Torrent platform, as well

as the performance of the bioinformatic pipelines at reconstructing the 16S theoretical relative abundance as well as assigning correct taxonomy to the eukaryotic DNA. The additional goal of using the culture standard (CS) was to ascertain the effectiveness of the in-house DNA extraction protocol that follows the recommended procedure of the DNeasy Plant Mini kit (Qiagen, Hilden, Germany), including three bead-beating steps for 3 minutes in a FastPrep-24 bead beater (MP Bio, Solon, OH) [181].

Species	**NRRL Accession NO.**	Theoretical Composition of 16S rRNA(%)	
		Culture standard	**DNA standard**
Pseudomonas aeruginosa	B-3509	4.2	4.6
Escherichia coli	B-1109	10.1	10.0
Salmonella enterica	B-4212	10.4	11.3
Lactobacillus fermentum	B-1840	18.4	18.8
Enterococcus faecalis	B-537	9.9	10.4
Staphylococcus aureus	B-41012	15.5	13.3
Listeria monocytogenes	B-33116	14.1	15.9
Bacillus subtilis	B-354	17.4	15.7
Saccharomyces cerevisiae	Y-567	-	-
Cryptococcus neoformans	Y-2534	-	-

Table 6: Culture and DNA standard microbial composition of the mock communities used during the current study and 16S theoretical relative abundance. Based on ZymoBIOMICS^{TM}, the strain information was extracted from the website of the Agricultural Research Service Culture Collection and can be accessed with the NRRL accession number (NRRL, https://nrrl.ncaur.usda.gov/).

Amplicon based sequences were generated by two different platforms, Ion Torrent (Centre for Omics Sciences, Reus, Spain) and Illumina (Centre for Genomic regulation, Barcelona, Spain). In the case of Ion Torrent, the sequencing libraries were prepared in the in-house laboratory of the University Rovira i Virgili using both the DNA and culture standard. For the libraries creation, the 16S rRNA region was amplified by PCR with the primers 515F and 806R [158] whereas the 18S rRNA region was amplified using the primers FR1 and FF390 [159]. Since a positive correlation between PCR cycles and amount of chimeric sequences has been reported [27], 30 and 45 PCR cycles were used for the libraries creation. The PCR products were purified using GeneRed Size selection Kit (Qiagen, Hilden, Germany) and sent to COS for sequencing with the 530 chip using the Gene Studio S5 System of the Ion Torrent platform. On the other side, the DNA standard and extracted DNA from the culture standard were sent directly to CRG to be sequenced by Illumina MiSeq 2x300 yielding paired end sequences for the v3 region of the 16S (primers 341F and 785R, [182]) and for the ITS region (primers ITS1F/ITS2R, [183]). Schematic representation of the experimental design is given in Figure 7.

The Ion Torrent platform generated in average 300 bp reads for the 16S amplicon and 350 bp reads for the 18S amplicon, with an average Phred33 quality score of 29 and 27, respectively. On the other hand, Illumina generated in average 300 bp reads for both amplicons with an average Phred33 quality score of 36 for both 16S and ITS forward reads and 34 and 35 for the 16S and ITS reverse reads, respectively. Due to the fact that the Phred33 quality of the Ion Torrent reads dropped below 10 in positions located in the middle of the read, two filtering strategies were applied. One applying a quality threshold at 10 (Q10) and one at 20 (Q20). The motivation behind these two strategies was to examine whether higher number of sequences or higher overall quality will produce better results. Contrarily, for the Illumina reads, only the Q20 threshold was applied.

Bioinformatic workflow

Bokulich *et al.* [178] benchmarked different quality filtering strategies with QIIME 1 and Bokulich *et al.* [34] benchmarked the performance of difference classification algorithms between QIIME 1 and QIIME 2. Therefore, the bioinformatic pipelines were based on two versions of QIIME, QIIME 1 (version 1.9.1) and QIIME 2 (version 2018.2),

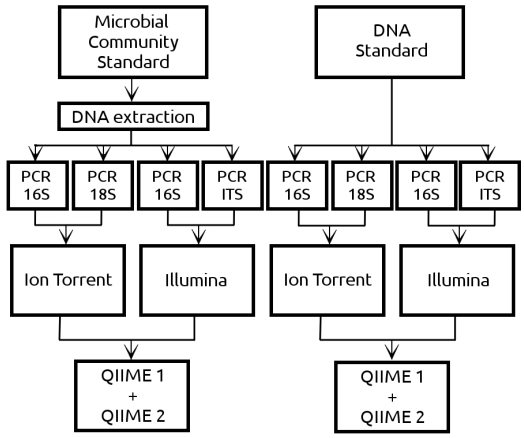


Figure 7: Two commercial mock community standards from ZymoBIOMICS^{TM} with exactly the same microbial composition of 8 prokaryotes and 2 eukaryotes have been used in the current study. The Microbial Community standard (referred as CS) consisted of microbial cells from which DNA was extracted using an in-house DNA extraction protocol. The DNA standard (referred as DS) contained DNA from the same 10 microbial cells as the CS but extracted by ZymoBIOMICS^{TM}. Both standards were sequenced using Ion Torrent and Illumina platforms. Regarding the DNA from the prokaryotic cells, both platforms sequenced the 16S amplicon. Regarding the DNA from the eukaryotic cells, Ion Torrent sequenced the 18S amplicon whereas Illumina the ITS amplicon. In the case of Ion Torrent 30 and 45 PCR cycles have been implemented in both amplicons, whereas in Illumina only 30 PCR cycles were implemented. Sequencing data derived from both NGS platforms have been analyzed using QIIME 1 and QIIME 2.

with the processing and taxonomic assignment steps mentioned in Table 7. Along with QIIME, bioinformatic tools such as FastQC [184], Trimmomatic [185] and FLASH [186] were executed externally.

Ion Torrent OTU	**Illumina OTU**	**Illumina ASV**
Barcode extraction[1]	Paired ends merging[3]	Paired ends merging[3]
Quality filtering (Q10 or Q20)[1]	Quality filtering (Q20)[4]	DADA2 quality filtering (Q20)[2]
Reads dereplication[2]	Reads dereplication[2]	DADA2 reads dereplication[2]
Open reference OTU[2]	Open reference OTU[2]	DADA2 Chimeras filtering (only ITS)[2]
Chimeras filtering[2]	Chimeras filtering[2]	DADA2 ASV[2]
SKLEARN classifier training[2]	SKLEARN classifier training[2]	SKLEARN classifier training[2]
SKLEARN taxonomy assignment[2]	SKLEARN taxonomy assignment[2]	SKLEARN taxonomy assignment[2]
BLAST+ taxonomy assignment[2]	BLAST+ taxonomy assignment[2]	BLAST+ taxonomy assignment[2]

[1] QIIME 1 (version 1.9.1)
[2] QIIME 2 (version 2018.2)
[3] FLASH
[4] Trimmomatic

Table 7: Bioinformatic pipelines based on NGS platform and method of clustering used during this study for comparisson of their performance over the mock community standards.

From the default parameters of QIIME 1 for the quality filtering of raws reads, only the Phred33 quality threshold was altered. Generally, the quality filtering concerned discarding reads with consecutive bases above a given Phred33 threshold but occupying $<$ 75% of the total read length, truncating reads at positions with more than 3 consecutive bases with Phred33 quality less than the desired and reassessing the discarding rule after truncation. Due to the fact that QIIME 1 quality filtering steps require the sequences to be multiplexed, for the demultiplexed Illumina sequences the quality filtering steps of QIIME 1 were replicated in Trimmomatic. Moreover, the DADA2 algorithm [187], as incorporated into QIIME, truncated reads at the first base instance of undesired quality and discarded reads with $>$2 expected errors. An additional filtering step was implemented by removing chimeric sequences with VSEARCH UCHIME de novo [188] or DADA2.

Regarding the Illumina reads two clustering methods were applied. One that creates clusters of sequences, called operational taxonomic units (OTU) based on a similarity threshold [189] and one that defines sequence variants called amplicon sequence variants (ASV) [30]. The OTU method produces an OTU-table where, for each sample, the number of sequences in each OTU has been recorded [188], whereas the ASV method is related with an ASV-table of the frequency that each ASV has been observed in each sample [187]. OTUs containing $<$ 10 sequences across all samples were filtered-out as noise [190], and the similarity threshold for the OTU clustering was set to 99% as this threshold returns more comparable results between OTU and ASV [31].

For the metataxonomic classification the database SILVA (version 132) has been the

source of taxonomy for the 16S and 18S amplicons [99] as it is the most recent and updated database, whereas the ITS taxonomy relied on the UNITE database (version 7.2) [191]. The taxonomic assignment was carried out by two algorithms, the k-mer based multinomial naive Bayes algorithm integrated in the Python Scikit-learn library (SKLEARN) [32] and the Basic Local Alignment Search Tool+ (BLAST+) algorithm which represents an enhanced version of the very popular BLAST algorithm available from 1997 [33]. Both algorithms report a confidence percentage, with the SKLEARN algorithm referring to the amount of confidence for the taxonomy assigned at a specific taxonomic level and BLAST+ referring to the fraction of top hits that matched the consensus taxonomy at a given level. As SKLEARN represents a machine learning approach, the additional flexibility provided was to assign taxonomy after training the algorithm with extracted reference sequences from the SILVA and UNITE databases using the aforementioned PCR primers and trimmed to a length equal to the maximum length of the reads after quality filtering. The training process of SKLEARN is based on k-mers where the value 7 was used as it is the default balanced QIIME 2 parameter. In relaxed terms, during the training process SKLEARN splits each reference sequence into a series of overlapping heptamers and assigns a level of taxonomy to a given collection of heptamers. Later on, during the classification process SKLEARN splits each sequence once again into a collection of overlapping heptamers, and tries to assign a level of taxonomy by taking into consideration the collections of heptamers from the reference sequences. The balanced default parameters of BLAST+ remained unaltered whereas the performance of SKLEARN improved after reducing the confidence parameter from the default 0.7 value down to 0.5.

Results

Figure 8 shows the number of sequences for each sample after applying Phred33 quality filtering and removing chimeras. For the Ion Torrent a mild filtering was applied after setting the quality threshold at Q10 with an average of 8.6% of the sequences filtered, across all samples, for the 16S amplicon and 14.1% for the 18S whereas at Q20 an average of 62 and 72.4% was removed, respectively. An additional average of 13.5% of the sequences were identified as chimeras for the 16S amplicon and 1.2% for the 18S at Q10, while at Q20 the identified chimeras were 5.9 and 1.3%, respectively. Considering the PCR cycles, their impact on the production of chimeras was not clear for the 16S amplicon as at Q10, 45 cycles generated 3.5% more chimeras than 30 cycles for the CS but for the DS they produced 4.2% less. The same pattern repeated for the 16S amplicon at

Q20 with 45 cycles of the CS producing 1.6% more chimeras but for the DS 3.5% more chimeras produced from 30 cycles. On the other hand, the difference was more apparent for the 18S amplicon producing more chimeras at 45 than 30 cycles, but the difference was marginal representing only 1.6% of the sequences in average (Figure 8A).

For the Illumina platform, the merging of the paired ends caused a $\approx 2\%$ loss of reads for the 16S amplicon in both standards, whereas for the ITS amplicon of the DS the loss was 38%. Due to the fact that the sequencing of the ITS amplicon for the CS generated very low amount of sequences which had very low Phred33 quality, this sample was excluded from the study. This was the additional reason for not reporting the theoretical abundance of 18S and ITS amplicons, along with the fact that from the two standards only the CS reports 18S theoretical abundance in the specifications. However, research interest still remained on examining whether the classification algorithms could assign correct taxonomy to the eukaryotic DNA and which amplicon of the two improves classification performance. For the 16S amplicon of the CS, the Illumina OTU pipeline removed 1.2% of sequences during the quality filtering step and an additional 23.7% was identified as chimeras. The pipeline performed quite similar for the DS removing 1 and 17.9%, respectively. On the contrary, for the 16S amplicon of the two standards the Illumina ASV pipeline identified $\approx 80\%$ of the sequences as chimeric. This high percentage could be justified in cases where non-biological nucleotides, such as primers or adapters, have not been removed prior to analysis [2], but since this rationale did not hold for the given dataset, the chimera filtering step was omitted for both standards. Therefore, the only loss was during the quality filtering with both standards losing$\approx 5\%$ of sequences. Regarding the ITS amplicon of the DS, the Illumina OTU pipeline filtered 0.8% of sequences based on quality but did not identify any chimeras, and the Illumina ASV pipeline removed 1.9% during quality filtering and a further 5% during chimera filtering (Figure 8B).

The metataxonomic classification was performed at genus level since accurate classification at species level is a known limitation of rRNA amplicon sequencing due to the fact that it is a highly conserved region [192]. This limitation became apparent also in the current study as the only bacterium identified consistently and accurately at species level was *Listeria monocytogenes* whereas *Salmonella* was the only one whose classification never reached species level. From the rest, *Bacillus* demonstrated the highest variability with overall 7 different species being identified, 5 species for *Staphylococcus* and *Pseudomonas*, and ≤ 3 for *Escherichia*, *Lactobacillus*, and *Enterococcus*. Although this broad variability concerned the OTU clustering method, the variability in the ASV method was more constrained including only the cases of either correct species identifi-

[2] https://benjjneb.github.io/dada2/tutorial.html

cation, no species identification or species identification as *uncultured* bacterium.

Figures 9 - 12 depict 16S estimated relative abundance (orange color) being juxtaposed against theoretical relative abundance (blue color) for both standards and NGS platforms. Overlapping between the two abundances is being represented with dark gray color and estimated abundance below 1% or undefined (0%) is being represented numerically. Excess of orange color at the bar edges denotes abundance overestimation whereas excess of blue color abundance underestimation. Next to each figure the taxonomic assignment confidence is being displayed as it has been reported by the classification algorithm at genus level (All). An additional step has been performed where the assigned taxonomies have been filtered by setting a confidence threshold which is displayed next to the unfiltered confidence. This threshold was initially set to 90% (>0.90) and gradually decreased until an optimal balance between amount of false positives and theoretical abundance reconstruction is achieved. Apart from Figures 11B, 12B,D this confidence threshold matches the minimum unfiltered confidence reported by the classification algorithm giving an identical estimated relative abundance before and after confidence filtering as well as the same amount of false positives (FP).

For the Ion Torrent platform, SKLEARN failed to identify *Salmonella* regardless

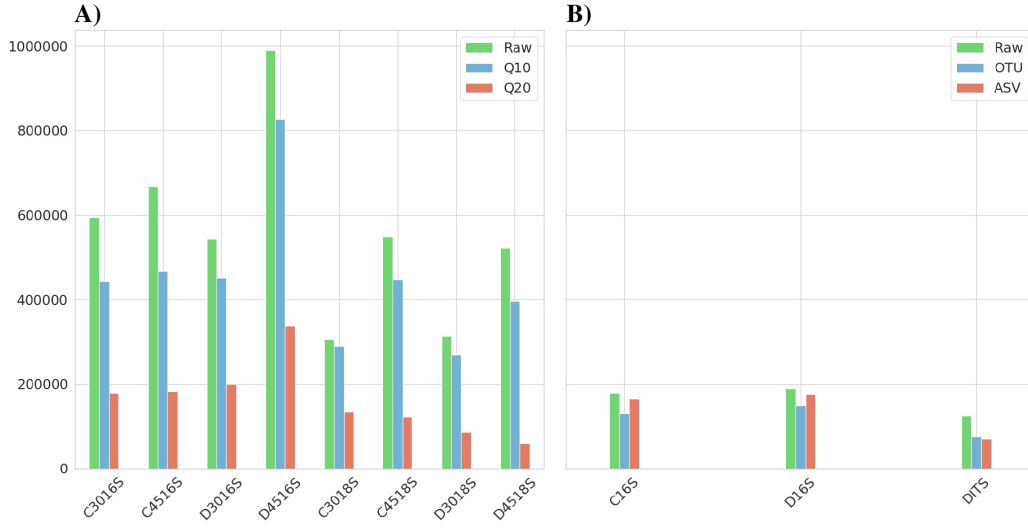


Figure 8: Number of sequences resulted after applying quality and chimeras filtering. (A) Ion Torrent. First letter of the sample names (C or D) represents type of mock community standard (Culture or DNA). What follows is the number of PCR cycles (30 or 45) with the amplicon (16S or 18S) at the end. The raw number of sequences are represented in green and in red and blue the two Phred33 quality filtering strategies Q20 and Q10, respectively. (B) Illumina. First letter of the sample names (C or D) represents type of mock community standard (Culture or DNA) with the amplicon (16S or ITS) at the end. The raw number of sequences are represented in green and in red and blue the sequences resulted from the filtering steps of the Illumina ASV and Illumina OTU pipeline, respectively.

quality filtering threshold, PCR cycles or standard type, while achieved best performance with the DS, 45 PCR cycles, Q20 and confidence threshold 80% (Figure 10G). Overall, the maximum number of false positives was 2 with the genera *Carnobacterium*, *Citrobacter*, *Oenococcus*, and *Pediococcus* consisting the pool of false positives. At the same time, BLAST+ seems to have exhibited a better performance than SKLEARN with optimal performance also with the DS, 45 cycles and Q20 (Figure 10H), but generating higher amounts of false positives and requiring a lower confidence threshold for optimal performance. In general, BLAST+ proved to be more sensitive than SKLEARN with 5 as the maximum number of false positives and a persistent confidence threshold of 60%. The false positives identified by BLAST+ were the genera *Cedecea*, *Citrobacter*, *Enterobacter*, *Klebsiella*, *Oenococcus*, and *Pediococcus*.

With Illumina generated data, the landscape was more clear. Both pipelines, Illumina OTU and ASV, yielded similar results with both classification algorithms performing better with the DS (Figure 12). Once again BLAST+ held the best performance managing to approximate quite accurately the theoretical composition (Figures 12B,D). However, it demonstrated overall higher sensitivity producing more false positives with their number being affected by even a slight increase of the confidence threshold by just 1% from the minimum reported confidence of 69% (Figures 11B, 12B,D). The pool of false positives for SKLEARN was comprising the genera *Acetobacter*, *Enterobacter*, and *Oenococcus*, whereas for BLAST+ the genera *Citrobacter*, *Acetobacter*, *Cronobacter*, *Enterobacter*, and *Oenococcus*. In general, although the relative abundance of the false positives remained below 0.01%, the only exemption was with the CS and the Illumina ASV pipeline where *Cronobacter* reached 0.3%. Moreover, even if the confidence level of the classification assignment was quite low for the false positives in both algorithms (60%), the genera that defied this trend were *Acetobacter*, *Enterobacter* and *Oenococcus* reaching as high as 90% confidence.

With respect to fungi, none of the algorithms detected *Cryptococcus* regardless NGS platform or standard type, contrary to *Saccharomyces* which was detected though not always at species level. In both Illumina OTU and ASV pipelines, both algorithms exhibited similar performance by identifying only *Saccharomyces* with 100% confidence without yielding any false positives. On the other hand, BLAST+ in Ion Torrent managed to identify *Saccharomyces* with 99.9% confidence in both standards regardless quality threshold and PCR cycles, but produced *Zygosaccharomyces* as a false positive with CS at Q10 and 30 cycles and *Kazachstania* with DS at Q20 and 45 cycles having a 60% confidence in both cases. On the side of SKLEARN, *Saccharomyces* occupied $\approx 61\%$ of the relative abundance in average across the different PCR cycles in both standards

at Q10 with the rest of the abundance being occupied by a taxonomy assigned as *uncultured* fungus. At Q20, *Saccharomyces* occupied 99% of the relative abundance with the DS at 45 cycles and 50% in the rest of the samples, with the remaining abundance once again assigned as *uncultured* fungus. Although in the case of BLAST+ the false positives could be removed by raising the confidence threshold, in the case of SKLEARN confidence filtering did not improve the result as the confidence level was in average 90% for *Sacchraromyces* and 85% for the false positives.

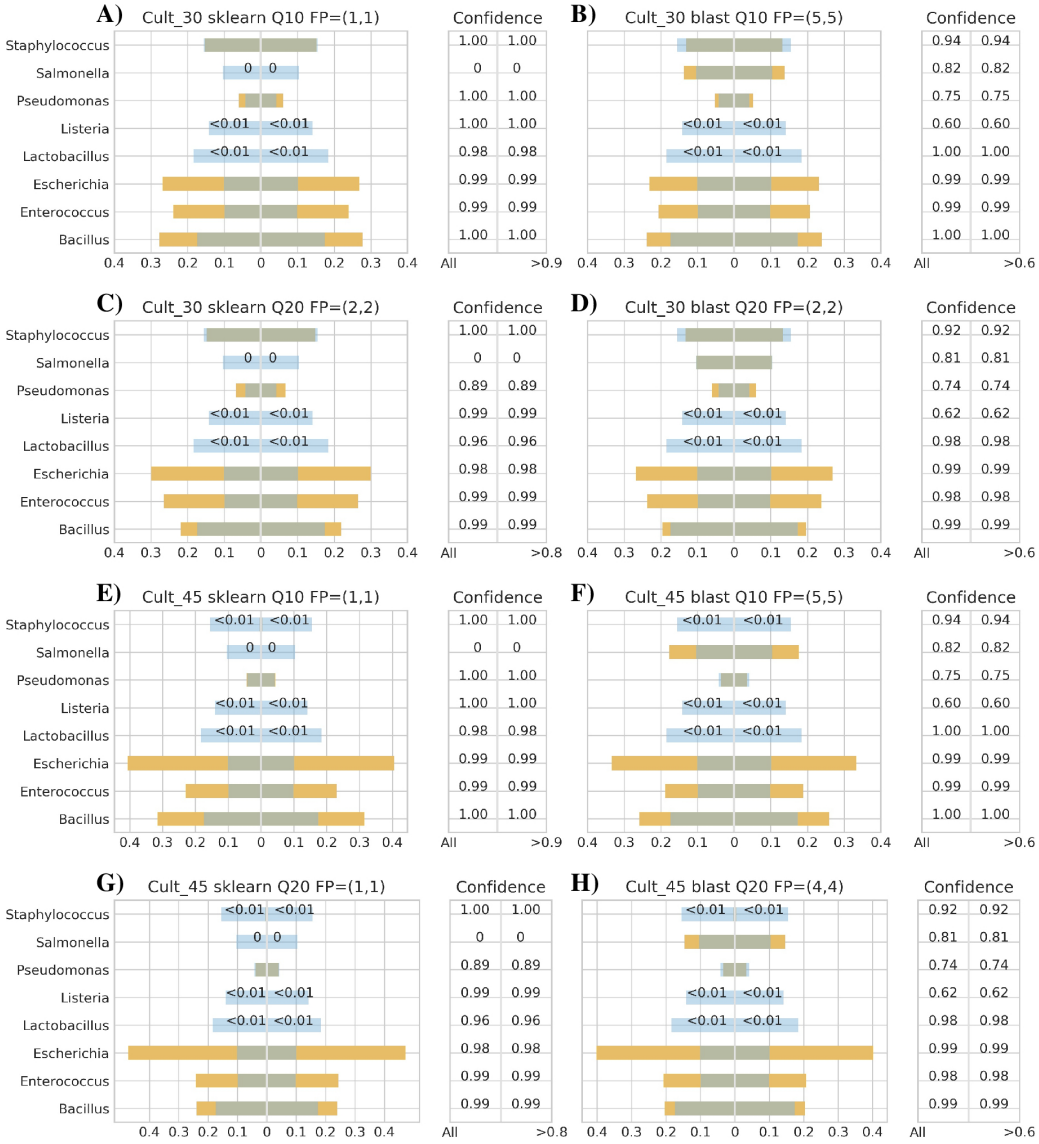


Figure 9: 16S theoretical (blue color) and estimated (orange color) relative abundance for culture standard using Ion Torrent. Overlapping between the two abundances is being represented with dark gray color. Cult_30 and Cult_45 represent 30 and 45 PCR cycles, Q10, and Q20 Phred33 quality filtering threshold and FP false positives without (first number) and with confidence filtering (second number). Figures to the left (A,C,E,G) represent estimated abundance based on SKLEARN algorithm and to the right (B,D,F,H) based on BLAST+. Estimated relative abundance to the left side of 0 is based on unfiltered confidence (All) and to the right on filtered ($> \%$).

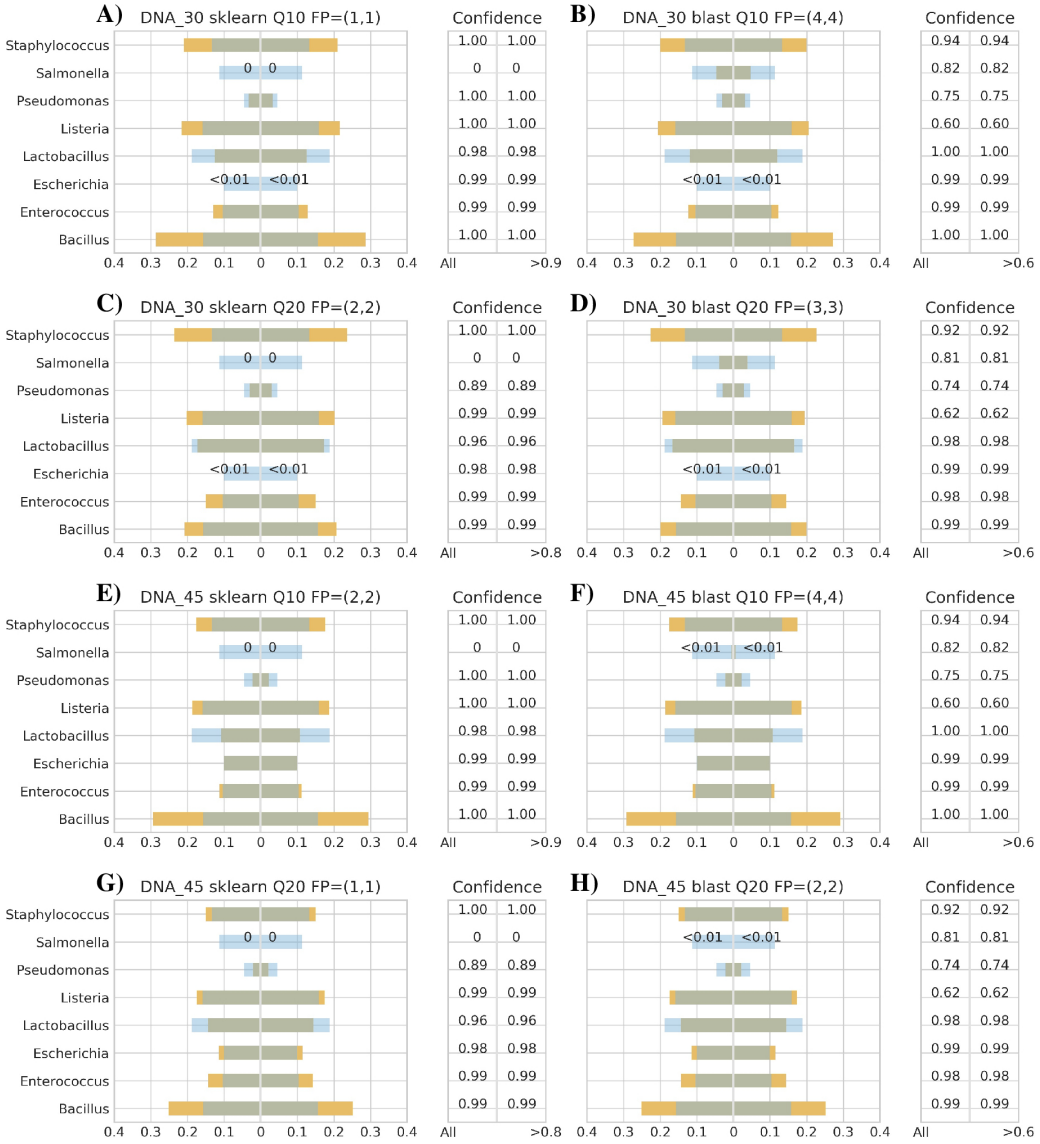


Figure 10: 16S theoretical (blue color) and estimated (orange color) relative abundance for DNA standard using Ion Torrent. Overlapping between the two abundances is being represented with dark gray color. DNA_30 and DNA_45 represent 30 and 45 PCR cycles, Q10, and Q20 Phred33 quality filtering threshold and FP false positives without (first number) and with confidence filtering (second number). Figures to the left (A,C,E,G) represent estimated abundance based on SKLEARN algorithm and to the right (B,D,F,H) based on BLAST+. Estimated relative abundance to the left side of 0 is based on unfiltered confidence (All) and to the right on filtered ($> \%$).

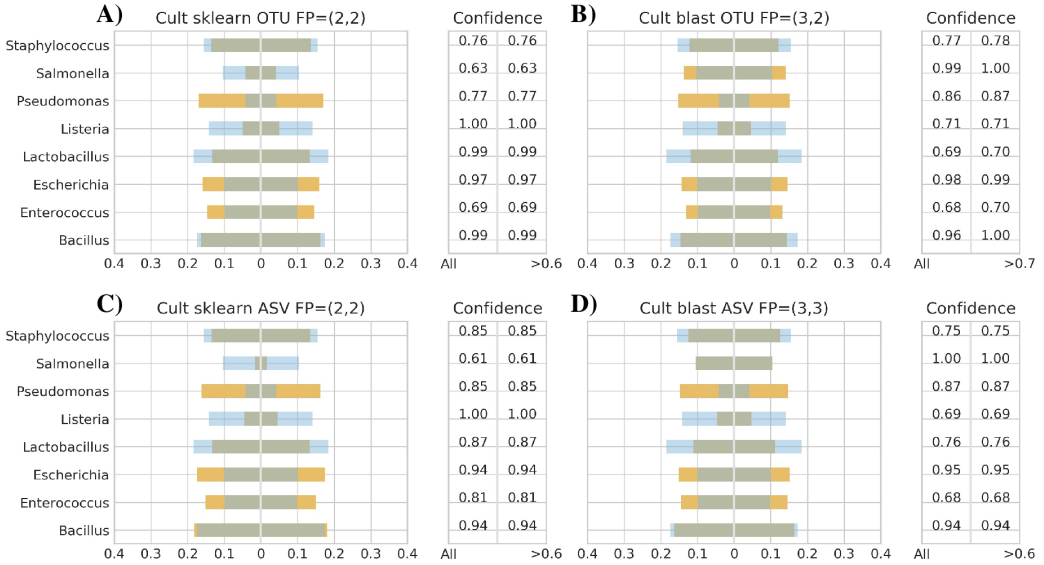


Figure 11: 16S theoretical (blue color) and estimated (orange color) relative abundance for culture standard using Illumina. Overlapping between the two abundances is being represented with dark gray color. OTU and ASV represent Illumina OTU and Illumina ASV pipelines and FP false positives without (first number) and with confidence filtering (second number). Figures to the left (A,C) represent estimated abundance based on SKLEARN algorithm and to the right (B,D) based on BLAST+. Estimated relative abundance to the left side of 0 is based on unfiltered confidence (All) and to the right on filtered ($> \%$).

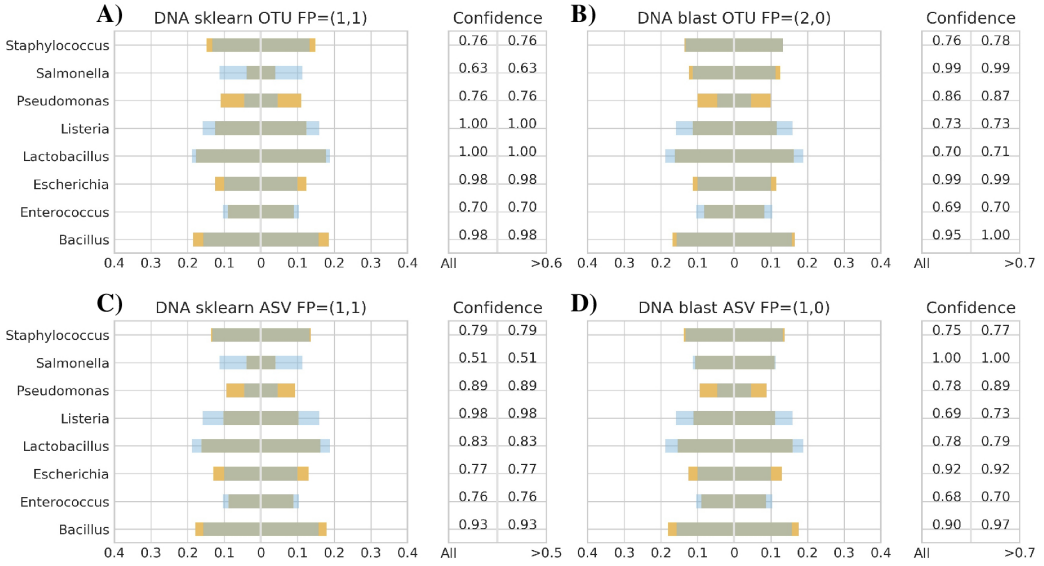


Figure 12: 16S theoretical (blue color) and estimated (orange color) relative abundance for DNA standard using Illumina. Overlapping between the two abundances is being represented with dark gray color. OTU and ASV represent Illumina OTU and Illumina ASV pipelines and FP false positives without (first number) and with confidence filtering (second number). Figures to the left (A,C) represent estimated abundance based on SKLEARN algorithm and to the right (B,D) based on BLAST+. Estimated relative abundance to the left side of 0 is based on unfiltered confidence (All) and to the right on filtered ($> \%$).

Discussion

A mock community represents a microbiome sample of known microbial composition and in the current study two types of mock community standards with the same species composition have become the tool for evaluating the effectiveness of QIIME balanced default parameters on metataxonomic analysis workflows destined to the analysis of wine aging samples. The evaluation was performed with QIIME framework and two classification algorithms, one representing a popular local alignment algorithm (BLAST+) and the other one a popular machine learning approach (SKLEARN). These two algorithms have been introduced for the first time in QIIME 2 and their performance compared to the classification algorithms of QIIME 1 have been benchmarked by Bokulich *et al.* [34] where they exhibited similar as well as enhanced performance on different performance metrics. Moreover, Bokulich *et al.* [178] in QIIME 1 benchmarked different quality-filtering strategies so as to provide guidelines for processing Illumina amplicon-based sequencing data. Although the suggested parameters of these studies have been incorporated as balanced default parameters in QIIME, microbiome samples undergo different laboratory procedures and protocols and thus these parameters should be evaluated prior to implementation. Therefore, the aim of the present study was to examine the effect of these parameters on a series of aspects related to our laboratory and bioinformatic workflows using a mock community and focusing on reconstructing the theoretical 16S relative abundance or yeast composition based on 18S and ITS amplicon sequencing. Furthermore, the mock community facilitated the qualitative assessment of other aspects such as the performance of the classification algorithms, the possibility of utilizing the reported taxonomic assignment confidence from the classification algorithms as a tool for eliminating false positives, the performance of Ion Torrent and Illumina NGS platforms with the 16S amplicon, the effect of PCR cycles on the analysis of Ion Torrent data, as well as the outcome of the in-house DNA extraction protocol by using a culture based standard (CS).

The 16S metataxonomic analysis of the CS approximated quite closely the outcome of the DS analysis in the Illumina platform, while it demonstrated an apparent variability in the case of the Ion Torrent platform. On the other hand, the Ion Torrent 18S analysis produced similar results in both standards. This denotes that pinpointing a performance culprit among the NGS platforms, PCR protocols or bioinformatic pipelines is rendered difficult as a further variability is being added by the DNA extraction protocol. Regarding the discard of the ITS amplicon based sample of the CS due to low quality, it has been attributed to the poor performance of the DNA extraction protocol since good quality

Illumina sequences were generated with the corresponding sample of the DS.

With Ion Torrent, both classification algorithms performed better with the DS linked to 45 PCR cycles and Q20 as a quality threshold signifying that optimal performance is more related to better overall sequence quality rather than higher amount of sequences as produced by the Q10 threshold. This could be associated with the fact that Q20 is related to 1% base call error rate while Q10 to 10% [29], indicating that low Phred33 quality threshold might lead to higher possibility of misclassification. Nevertheless, this result could not be easily attributed to the PCR cycles as 45 cycles in DS produced the highest amount of sequences among all samples and on the other hand in CS both algorithms favored 30 cycles. Moreover, the impact of PCR cycles on the amount of chimeric sequences was either marginal or unclear, however a negative correlation between quality threshold and amount of chimeras became apparent with the 16S amplicon, with fewer chimeras being identified at Q20 threshold. This indicates that a small increase of the PCR cycles does not influence greatly the production of chimeras and many of those chimeric sequences had overall low quality as they represent PCR artifacts. Similarly, slight difference on the production of chimeric sequences was also observed by a small increase of PCR cycles in the study of Ahn *et al.* [27] when 25 PCR cycles were compared to 30 cycles, however great disparity on the amount of chimeras was observed between 15 and 30 cycles with the authors suggesting the lowest PCR cycles possible.

As Van Der Pol *et al.* [31] suggested, setting the similarity threshold to 99% for the OTU clustering method produced similar results as the ASV method in Illumina, however the latter demonstrated a narrower variability of taxonomic assignment at species level. Furthermore, the omitted chimera filtering step in Illumina ASV pipeline for the 16S amplicon highlighted its importance as false positives above the impurity level of 0.01% were emerged. Additionally, the two NGS platforms presented different filtering behaviors at Q20 with Ion Torrent removing more sequences during the Phred33 quality filtering and less during chimera filtering, whereas Illumina performed the opposite. That could indicate that more chimeric sequences with high Phred33 quality score were generated with Illumina.

As a whole, BLAST+ exhibited better and more balanced performance in both NGS platforms than SKLEARN, however it demonstrated higher sensitivity producing more false positives and overall lower confidence regarding taxonomic assignment. The low amount of false positives generated by SKLEARN with the 16S amplicon could be associated with its training process as higher amount of reference sequences were extracted from the database with the PCR primers of this amplicon compared to 18S and ITS. Nonetheless, its enhanced performance with the Illumina data could be connected to

the fact that its default parameters were linked with this NGS platform in the study of Bokulich *et al.* [34]. Moreover, the lack of false positives from both algorithms with the ITS amplicon could be explained by its higher specificity compared to 18S [193], and overall the reported taxonomic assignment confidence from the algorithms could not lead to an effective filtering tool of false positives as some of the false taxonomies have been assigned with high confidence level.

Conclusions

Overall, the mock community standards have been proven a useful tool demonstrating good performance of QIIME balanced default parameters on our workflows especially with the Illumina platform. Nevertheless, the performance of the NGS platforms or the classification algorithms should not be considered deterministic since an exhaustive benchmarking process is needed for that purpose. As underlined by Bokulich *et al.* [34], further fine-tuning of the QIIME default parameters with limited number of mock communities could lead closer to an overfitted rather than generalized performance. Moreover, a series of qualitative performance expectations could be proposed that could be summarized as better metataxonomic outcome when setting the Phred33 quality filtering threshold as high as possible, marginal difference in chimeras production between 30 and 45 PCR cycles, less false positives with ITS amplicon sequencing compared to 18S, similar performance between ASV and OTU clustering method when the clustering similarity threshold of the latter is set to 99% and more comparable results between Ion Torrent and Illumina platforms using the BLAST+ classification algorithm.

UNIVERSITAT ROVIRA I VIRGILI
ANALYSIS OF MICROBIAL POPULATIONS IN WINES THROUGH NGS METHODOLOGIES
Dimitrios Kioroglou

CHAPTER 4

Geographical origin has a greater impact on grape berry fungal community than grape variety and maturation state.

Dimitrios Kioroglou[1], Elena Kraeva-Deloire[2], Leigh M. Schmidtke[2,3], Albert Mas[1] and Maria C. Portillo[1†]

Microorganisms, 7(12), 669, 2019
https://doi.org/10.3390/microorganisms7120669

[1] Departament de Bioquímica i Biotecnologia, Universitat Rovira i Virgili, Marcel·lí Domingo s/n, 43007, Tarragona, Spain
[2] National Wine and Grape Industry Centre, Charles Sturt University, Wagga Wagga, NSW 2678, Australia
[3] School of Agricultural and Wine Sciences, Charles Sturt University, Wagga Wagga, NSW 2650, Australia
[†] Corresponding author

UNIVERSITAT ROVIRA I VIRGILI
ANALYSIS OF MICROBIAL POPULATIONS IN WINES THROUGH NGS METHODOLOGIES
Dimitrios Kioroglou

Abstract

We used barcoded sequencing to analyze the eukaryotic population in the grape berries at different ripening states in four Australian vineyards. Furthermore, we used an innovative compositional data analysis for assessing the diversity of microbiome communities. The novelty was the introduction of log-ratio balances between the detected genera. Altogether, our results suggest that fungal communities were more impacted by the geographical origin of the Australian vineyards than grape variety and harvest time. Even if the most abundant genera were *Aureobasidium* and *Mycosphaerella*, they were ubiquitous to all samples and were not discriminative. In fact, the balances and the fungal community structure seemed to be greatly affected by changes of the genera *Penicillium*, *Colletotrichum*, *Aspergillus*, *Rhodotorula*, and *Botrytis*. These results were not evident from the comparison of relative abundance based on OTU counts alone, remarking the importance of the balance analysis for microbiome studies.

Introduction

The concept of terroir in oenology refers to a geographic area characterized mainly by its climate, soil, and human factors that contribute to producing typical wines. However, the term microbial terroir is recently gaining interest in viticultural studies to indicate the importance of the vineyard microbiome composition over the regional wine typicity. The grapevine microbiome is the complex community of microorganisms, including fungi and bacteria that interact with the whole plant and play a key role in plant health, growth, and nutrient uptake [194].

Recently, due to advances in metagenomics and the development of high-throughput sequencing (HTS) techniques, the grapevine microbiome is receiving increasing attention. Metagenomic analyses suggest that the microbial communities associated with grapes and grape must resemble the ones present on leaves [42, 43] and have their source mainly in the soil and surrounding fields [108]. Furthermore, grape microorganisms can be transferred to the winery where, ultimately, they may affect wine chemical composition and influence its quality, even at the regional scale [78, 109, 110, 195].

Several factors and vineyard characteristics have recently shown that grape microbiome is influenced by vineyard characteristics like climate, region, site, and grape cultivar, suggesting that there is a nonrandom microbial fingerprint associated with the terroir [42–49]. One important factor that has been proven to induce changes in grape microbiome composition and structure is the berry development process [50, 51]. The grape ripening stages are defined mainly by physiochemical changes, such as increase in levels of phenolic compounds and accumulation of sugars [196]. Large numbers of yeast species have been identified on grape berries with population densities ranging from 101 to 103 CFU/g (Colony Forming Units) on immature grapes but increasing to 103106 CFU/g at harvest time [22, 42, 50, 52–54]. However, previous studies analyzing microbiome changes through berry maturation have been mostly based on culture-dependent techniques that have been proven to be insufficient to reveal the environmental microbial diversity and ecology [67, 138]. As a consequence, currently little is known about the real influence of the grape berry maturation state on its microbiome.

One of the major drawbacks of the studies analyzing microbiome diversity by means of HTS techniques is the derivation of statistical inferences after converting the OTU (operational taxonomic unit) counts of the identified genera to relative abundance. The transformation of the OTU counts to compositional data, on one hand, adds the constraint of the abundances having to sum to a constant (i.e., 1), and on the other hand, may lead to misinterpretations when multivariate statistics are applied [197]. Moreover, the nature

of compositional data is known to hinder proper differential abundance analysis since various normalization methods and statistical assumptions could potentially not be appropriate for this type of data [198]. Therefore, in the current study inferences on the differences of the microbial communities have been derived by using the compositional analysis toolbox GNEISS [40], as incorporated into Quantitative Insights Into Microbial Ecology framework (QIIME version 2019.1) [199]. GNEIIS introduces the concept of balances which refer to the log-ratio between specific microbial subsets of the community, eliminating the need of using relative abundances and statistical assumptions.

In this work, we use HTS to investigate the fungal biota (mycobiome) composition at two maturation stages of grape berries from two grapevine cultivars, Cabernet Sauvignon and Syrah, growing under two different geographical and environmental conditions. We also compare the usefulness of balances obtained by the GNEISS toolbox for the microbial diversity analysis on our data set.

Materials and Methods

Experimental Vineyard and Harvesting

Grapes were sourced from two Australian wine regions sampled during 2015 and denominated Griffith (G) (Riverina, New South Wales, Australia) and Orange (O) (Coordinates and elevation in Table S1 [3]). These wine regions represent two distinctively different grape growing regions. The Griffith region is classified as a warm to very warm grape growing area with temperate nights [200] and is characterized by a flat terrain (around 130 m above sea level, a.s.l.) and secure water supply, enabling it to maintain a 15% share in the total Australian grape production. In contrast, the Orange (O) region has an undulating to mountainous terrain with vineyard elevations spanning from 600 up to 1000 m a.s.l. The Orange region is classified as temperate to temperate/warm with cool to very cool nights. Two commercial vineyards were selected in both regions (designated G1, G2, O1, and O2), for Shiraz vines (S) whereas Cabernet Sauvignon (C) vines were sampled only at vineyard G1 and O2.

Between G1 and G2 there is less than 5 m altitude difference whereas the O1 site is at 607 m a.s.l. and the O2 site at 876 m a.s.l. thus having an approximate difference of 270 m (Table S1). Both S and C vines were own rooted, grown under drip irrigation, and trellised to a sprawling system in G. In O, vines were trellised to vertical shoot positioning. The nitrogen management throughout the season was similar for both cultivars and

[3] Supplementary materials are available at https://www.mdpi.com/2076-2607/7/12/669/s1

the average crop yields of both plots were approximately 15-20 tons per hectare. During the season, mesoclimatic temperatures, stem water potential, and soil moisture were monitored in an attempt to characterize experimental plots. Harvest dates for vineyards (H1 and H3) were determined at the point where sugar accumulation per berry and berry fresh mass in conjunction with °Brix. The first harvest for both cultivars occurred at approximately 21°Brix and was designated H1. The second harvest, designated as H3, occurred at 23°Brix for both cultivars. At each harvest date, 60 kg of grapes was randomly harvested across the vineyards for each variety with an addition of 40 mg/kg of potassium metabisulfite prior to transport to the Charles Sturt University (CSU)/ National Wine and Grape Industry Centre (NWGIC) experimental winery. On arrival, a 100-berry subsample from each replicate was collected and immediately frozen at -20°C for further analyses with a total of 50 subsamples.

DNA Extraction

Must samples were defrosted and centrifuged at 3500 g for 15 min, washed three times with ice cold phosphate buffered saline and the pellet resuspended in 200 L of DNeasy lysis buffer (Qiagen, Valencia, CA, USA) supplemented with 40 mg/mL lysozyme and incubated at 37°C for 30min. After this point, DNA extraction continued following the protocol of the QIAmp Fast DNA Stool Mini Kit (Qiagen,), with the addition of a bead beater cell lysis step for 2 min using a FastPrep-24 (MP Bio) and 100 µL of DNA eluted using AE buffer (Qiagen). DNA concentration and quality were assessed using a Quantus Fluorometer (Promega, Madison, WA, USA) followed by gel electrophoresis of 5 µL of eluant in 1.5% agarose submerged in 1X TAE buffer. Gels were stained with GelRed^{TM} (Biotium, Fremont, CA, USA) nucleic acid gel stain and viewed under UV light using Gel Doc XR+ Imaging system (Bio-Rad, Hercules, CA, USA) DNA samples (approx. 70 ng) were subject to PCR amplification and sequencing performed by the Australian Genome Research Facility. PCR amplicons were generated using as forward primer (ITS1: CTTGGTCATTTAGAGGAAGTAA or 341 F: CCTAYGGGRBGCASCAG) and reverse primer (ITS2: GCTGCGTTCTTCATCGATGC or 806 R: GGACTACNNGGGTATCTAAT), with 35 cycles, an initiation temperature of 95°C for 7min, disassociate conditions of 94°C for 30sec, annealing conditions of 55°C for 45sec for ITS or 50°C for 60sec for 16S RNA, extension at 72°C for 60sec and a final temperature of 72°C for 7min. Thermocycling was completed with an Applied Biosystem 384 Veriti and using AmpliTaq Gold 360 mastermix (Applied Biosystems, Foster City, CA, USA) for the primary PCR. The first stage PCR was cleaned using magnetic beads, and samples were visualized on 2% Sybr Egel (Thermo-Fisher, Carlsbad, CA, USA). A secondary

PCR to index the amplicons was performed with TaKaRa Taq DNA Polymerase (Takara Shuzo, Otsu, Japan). The resulting amplicons were cleaned again using magnetic beads, quantified by fluorometry by the Promega Quantifluor ST fluorometer (Promega), and normalized. The eqimolar pool was cleaned a final time using magnetic beads to concentrate the pool and then measured using a High-Sensitivity D1000 Tape on an Agilent 2200 TapeStation. The pool was diluted to 5nM and molarity was confirmed again using a High-Sensitivity D1000 Tape. This was followed by sequencing on a MiSeq platform (Illumina, San Diego, CA, USA) with a V3, 600 cycle kit (2 x 300 base pairs paired-end).

Paired-ends reads were assembled by aligning the forward and reverse reads using PEAR (version 0.9.5) [201]. Primers were identified and trimmed. Trimmed sequences were processed using Quantitative Insights into Microbial Ecology (QIIME 1.8) [89] USEARCH [95] (version 8.0.1623) and UPARSE software (version 8.1.1861) [202].

Using USEARCH tools sequences were quality filtered, full length duplicate sequences were removed and sorted by abundance. Singletons or unique reads in the data set were discarded.

The 16S rRNA sequences were clustered followed by chimera filtering using rdp_gold database as the reference. To obtain the number of reads in each OTU, reads were clustered with a minimum identity of 97%. Using Qiime taxonomy was assigned using Greengenes database (version 13_8, Aug 2013).

ITS sequences were clustered followed by chimera filtering using "Unite" database as reference. To obtain number of reads in each OTU, reads were clustered with a minimum identity of 97%. Using Qiime, taxonomy was assigned based on Unite database [203] (Unite Version7.1 Dated: 22.08.2016).

Data Analysis

The data processing and part of the statistical analysis has been performed with QIIME (version 2019.1). The OTU table has undergone a series of filtering steps including removing OTUs with < 10 counts across all samples, removing OTUs whose assigned taxonomy did not reach genus level and removing genera whose relative abundance was $< 1\%$ across all samples. After collapsing the OTU table at genus level, in order to compensate for the uneven sequencing depth across the samples, the OTU table was rarefied at a value equal to the maximum amount of sequences observed across all samples so as each sample to include 11.579 sequences

Statistical Analysis

The factors considered for the statistical analysis were region (G1, G2, O1, and O2), variety (C and S) and harvest period (H1 and H3). Statistical analysis has been performed in QIIME with the ADONIS permutation-based statistical test [204] and GNEISS, as well as externally using the Python (version 3.7) libraries STATSMODELS [205], SCIPY [206], and PANDAS [207]. Using the rarefied OTU table, alpha diversity was calculated based on the Shannon index, whereas beta diversity was based on the Bray-Curtis index since taxonomy was constrained at genus level. Using the Shannon index, the replicates were examined for outliers resulting in the removal of two samples. The distribution of the Shannon index proved of being bimodal, with one mode concerning only the region O1. The two modes were separated and two-way ANOVA was performed on each mode. Prior to ANOVA, the assumptions of heteroskedasticity and normality on each mode were examined and satisfied using the Levene and Shapiro-Wilk test, respectively. Similarities between regions, varieties, and harvest periods were examined with Principal Coordinate Analysis (PCoA) using the Bray-Curtis distance metric, whereas ADONIS multivariate analysis of variance (MANOVA) with 999 permutations helped to identify significance. The unrarefied OTU table became the input source for GNEISS since it applies its own normalization method. The analytical pipeline for GNEISS included the initial steps of imputing zero OTU counts with a pseudocount equal to 1 and partitioning of genera into two groups using Ward hierarchical clustering. Each group contains genera that are highly correlated based on their co-occurrence and therefore the two groups are anti-correlated. Subsequently, GNEISS applied isometric log-ratio transformation which calculates, for each sample, the log-ratio between these two groups. That means one group represents the numerator of the ratio and the other group the denominator, whereas the log-ratio is referred as balance. This balance may have positive or negative value signifying that for a given sample the abundances of some genera have changed and these genera are either from the numerator, the denominator, or both the numerator and denominator. Finally, based on these balances GNEISS performed ordinary least squares regression (OLS) using the regressors region, variety, and harvest period, where 10-fold cross validation of 10 partitions showed that over fitting did not occur. Since OLS regression is more appropriate for continuous than categorical independent variables, from the reported results the only statistical measure considered was the explained variance (R^2 adjusted).

Results

Sequences Analysis Results

DNA of 50 samples of grapes representative of the 4 vineyards (G1, G2, O1, and O2) were massively sequenced by the Illumina platform resulting in a minimum of 30K reads for both the ITS and 16S rRNA gene regions, respectively. After quality filtering and exclusion of sequences matching to chloroplast or mitochondria, 1.187.046 and 24.610 reads remained for fungal and bacterial community analysis, respectively. In the case of ITS sequences the median number of sequences per sample was 23.028 whereas for the 16S rRNA sequences it was 366. Given the low number of 16S rRNA sequences per sample combined with the fact that the majority of these sequences have been identified of belonging to the genus *Sphingomonas*, we considered that the obtained sequences for this amplicon were not enough for a robust analysis of the bacterial community.

Fungal Diversity Was Mainly Impacted by the Wine Region

Figure 13 shows alpha diversity of the fungal community based on Shannon index. The Shiraz samples from the Griffith region (G1 and G2) exhibited higher diversity compared to the rest of the samples, whereas the samples with the lowest diversity have been the ones from the O1 region. Additionally, the harvest period (H1 and H3) seemed to have affected the observed diversity of the Shiraz samples from O1. Two-way ANOVA on the Shannon index for the region O1 revealed significant differences between the groups of the factors variety and harvest which had relatively equal amount of impact on the total variance explained by these two factors (54% in total) (Table 8). For the rest of the samples significance was observed only between the groups of the factors region and variety which combined explained 73% of the total variance with the factor region having the greatest impact (53%) (Table 9).

Factor	**DF**	**R^2**	**F**	**Pr(>F)**
Variety	1	0.25	5.218	0.048
Harvest	1	0.29	5.522	0.043

Table 8: Results from Shannon index two-way ANOVA on region O1.

Fungal Community Clustered Distinctly According to Wine Region, Variety and Harvest

The clustering of the samples in the PCoA based on Bray-Curtis distance metric (Figure 14) suggested that the factor region has the greatest effect on the distinction of the

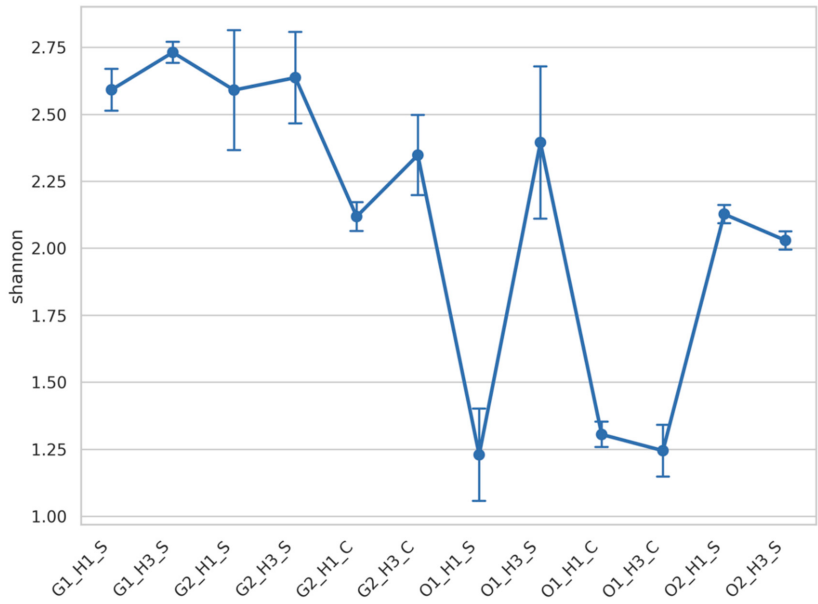


Figure 13: Samples alpha diversity based on Shannon index. Each value corresponds to the average of samples replicates and bars correspond to standard deviation. Samples abbreviation includes information of the region, Griffith (G1 and G2) or Orange (O1 and O2), the harvest time points (H1 or H3), and the grape varietals, Shiraz (S) or Cabernet (C).

Factor	**DF**	**R^2**	**F**	**Pr(>F)**
Region	2	0.53	33.575	1.74E-08
Variety	1	0.2	25.71	1.75E-05
Harvest	1	0.02	2.807	0.103

Table 9: Results from Shannon index three-way ANOVA on all the regions apart region O1.

samples with O2 samples being the most different. Moreover, higher order of influence was observed on the G2 samples by the factor variety, whereas on the O1 samples the factor harvest was the most influent, as suggested also by the Shannon index. After performing MANOVA with ADONIS on the BrayCurtis distance metric, the results showed that there are significant differences between the groups of the factors region, variety and harvest period (Table 10). From the three factors, region accounts for the highest amount of variance (53%) followed by variety (7%) and harvest period (2%).

Factor	**DF**	**R^2**	**F**	**Pr(>F)**
Region	3	0.532	20.694	0.001
Variety	1	0.079	9.267	0.001
Harvest	1	0.027	3.229	0.029

Table 10: Results of the MANOVA analysis performed with ADONIS on BrayCurtis distance metric.

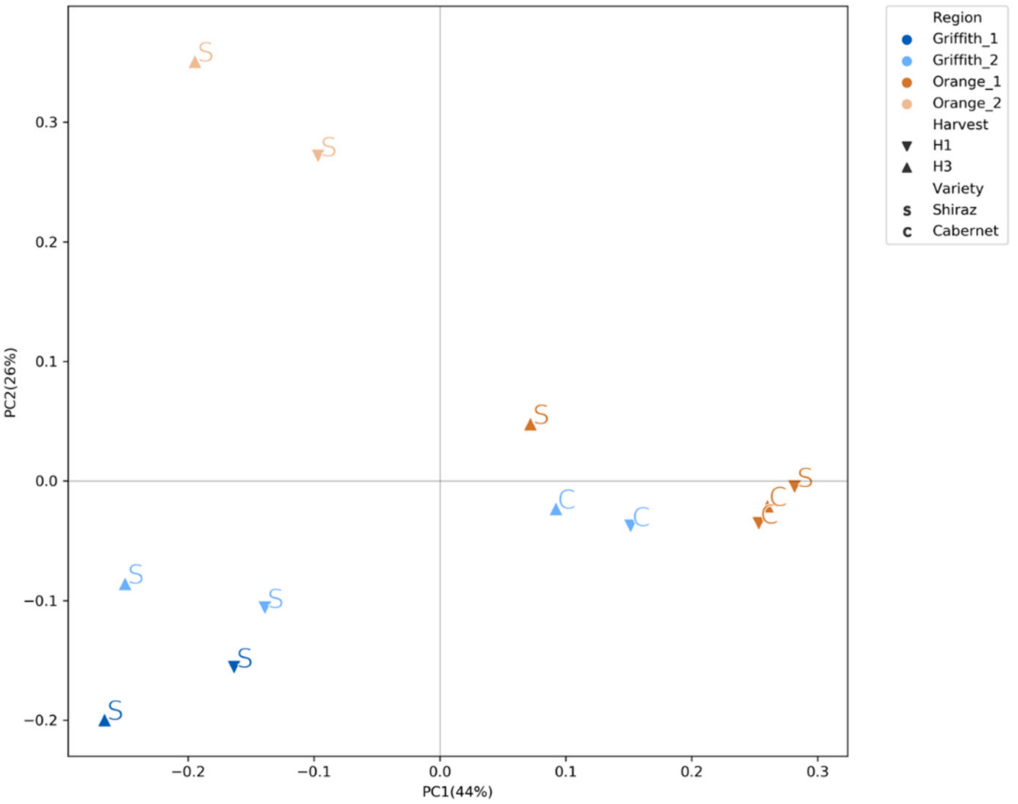


Figure 14: PCoA based on Bray-Curtis distance.

Genera Balances Affected the Fungal Community Structure

The high-throughput sequencing analysis allowed the detection of 18 different genera represented by more than 1% of the sequences for each sample. Table 11 shows all the genera that have been identified as well as the range of their OTU counts based on the rarefied OTU table. The most abundant genera in general were *Aureobasidium* and *Mycosphaerella* with a range of OTUs representing 5099 and 1991 on average, respectively. Other abundant genera were *Botrytis*, *Aspergillus*, *Colletotrichum*, *Rhodotorulla*, and *Penicillium*.

Using the tool GNEISS, 9 genera were included by GNEISS in the numerator and 9 in the denominator in order for the balances to be calculated (Figure 15). Furthermore, the range of the OTU counts for each genus is being depicted along with the fold-change between the minimum and maximum observed OTU count. Additionally, for each genus, the samples were grouped based on the factors (region, variety and harvest) and KruskalWallis H-test was performed in order to identify non-significant genera. After applying Bonferroni correction on the resulted p-values and setting the significance

threshold at 0.01, the genera *Aureobasidium*, *Phoma*, and *Diplodia* were identified as non-significant. Figure 16 shows the balances calculated by GNEISS for each sample showing again that the factor region seems to have the greatest impact on the separation of the samples.

Taxonomy	G1_H1_S	G1_H3_S	G2_H1_S	G2_H3_S	G2_H1_C	G2_H3_C	O1_H1_S	O1_H3_S	O1_H1_C	O1_H3_C	O2_H1_S	O2_H3_S
Diplodia	25	52	21	45	14	11	9	119	2	3	1	4
	146	0128	12 33	2680	618	1011	312	66163	16	23	01	25
Cladosporium	2	1	2	2	3	2	54	98	53	21	6	7
	14	03	04	17	36	17	2154	91135	4158	1925	67	77
Mycosphaerella	709	630	2046	1554	3876	2754	2536	3406	1750	1780	1784	1068
	5781041	467796	15683104	8522780	35114283	21052764	18452783	30733512	16951840	15192026	17822349	10111500
Aureobasidium	4190	3492	3783	2390	5384	4333	8178	5257	8545	8712	3393	2939
	33735254	28494175	19964907	15025085	52375559	41224625	79639262	32935927	85138696	83319090	32894590	28683802
Alternaria	354	284	508	304	517	399	53	536	64	61	47	62
	286440	206366	268804	227580	440531	287428	5355	415833	5364	4779	4480	4272
Stemphylium	196	162	376	282	346	300	86	188	49	58	27	18
	107263	126232	226479	207409	303404	211337	86139	176214	4265	3561	2348	1332
Phoma	198	147	138	67	279	304	150	118	88	92	341	14
	183327	118180	60319	41150	266331	226316	142186	89144	7892	76116	70349	1423
Aspergillus	1567	2263	1634	3734	44	32	6	14	1	2	3	0
	4652626	18263485	2345548	8116248	3867	636	593	635	06	112	24	01
Penicillium	520	1320	685	460	34	17	35	164	25	2	2	1
	360943	3242727	3041414	268933	2154	519	16129	106307	10406	144	04	04
Erysiphe	4	9	17	30	0	9	6	9	1	8	44	260
	28	315	722	1483	01	59	512	213	012	29	4189	251310
Botrytis	540	1121	50	1235	32	21	18	202	8	6	4302	5314
	4503317	3181631	10659	1301723	2254	925	1221	132508	39	27	23154794	43835398
Metschnikowia	6	51	2	0	4	1	4	103	5	4	0	0
	058	0148	05	02	25	14	15	16151	26	317	02	01
Acremonium	12	16	122	232	44	32	0	2	0	0	1	1
	135	132	47662	63316	4367	1633	00	14	02	00	04	11
Colletotrichum	112	250	1	14	331	2567	0	0	1	1	0	0
	7575	42509	02	517	187428	25543623	01	00	02	01	00	00
Rhodotorula	1638	1153	8	2	10	8	138	131	763	622	47	60
	304445	812663	020	110	812	612	60252	104179	512780	562716	3162	4577
Sporobolomyces	139	90	236	148	184	249	11	288	24	22	12	7
	84264	35175	89361	108311	182228	127287	314	159434	1832	1241	631	512
Bullera	19	13	84	86	118	125	5	29	9	10	8	4
	1226	1017	32132	51115	115128	74160	015	1735	515	425	79	5
Cryptococcus	159	104	902	538	327	410	78	1303	97	136	1559	1770
	81256	58170	271563	3261425	277353	234468	3488	6911739	87101	110199	11951926	14411872

Table 11: OTU counts of rarefied OTU table collapsed at genus level. Values represent median and minimum-maximum range of OTU counts of sample replicates.

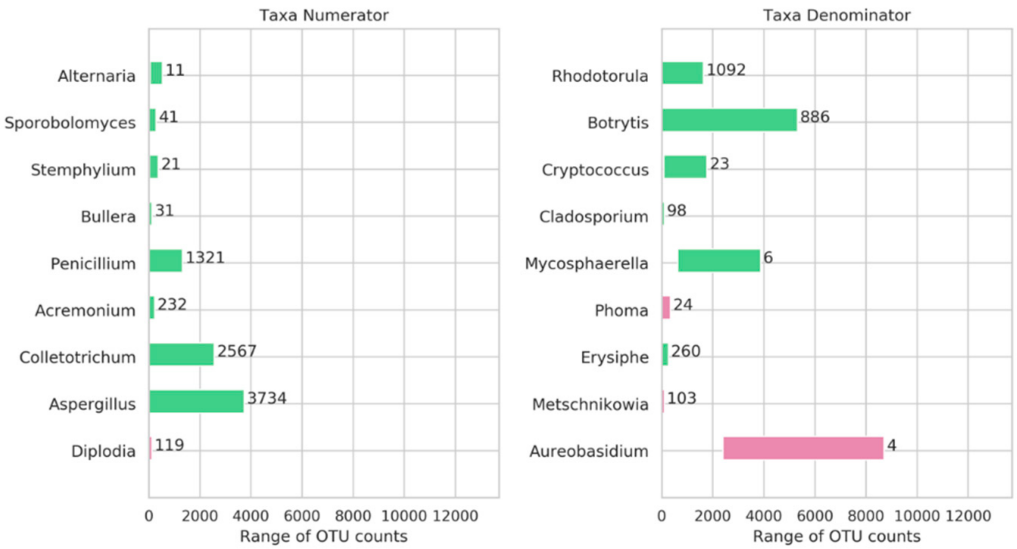


Figure 15: Range of collapsed OTU counts for genera in the numerator and denominator of the balances. OTU counts concern the rarefied OTU table. To the right of each bar, the fold-change between the minimum and maximum observed OTU count is shown. Purple color represents non-significant genera whereas green color significant.

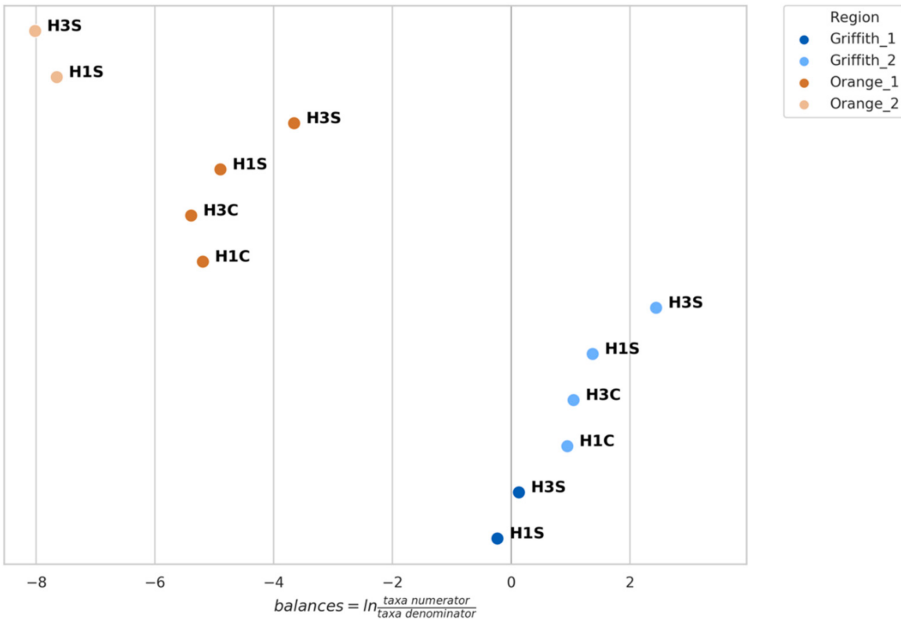


Figure 16: Balances calculated for each sample. Values represent balances median value of replicates.

Since GNEISS calculates log-ratio for each sample, it is difficult to conclude whether the calculated log-ratio has been the outcome of changes in the genera of either the nu-

merator, the denominator or both. However, based on the Figure 15 we may assume that the genera that could greatly influence the resulted log-ratio are genera with quite high fold-change between the minimum and maximum observed OTU count. These genera are *Penicillium*, *Colletotrichum*, and *Aspergillus* for the numerator, and *Rhodotorula* and *Botrytis* for the denominator. Figure 17 shows the abundances of the identified genera based on their observed OTU counts. The OTU counts have been converted to relative abundance in order to be compared to their log2 values.

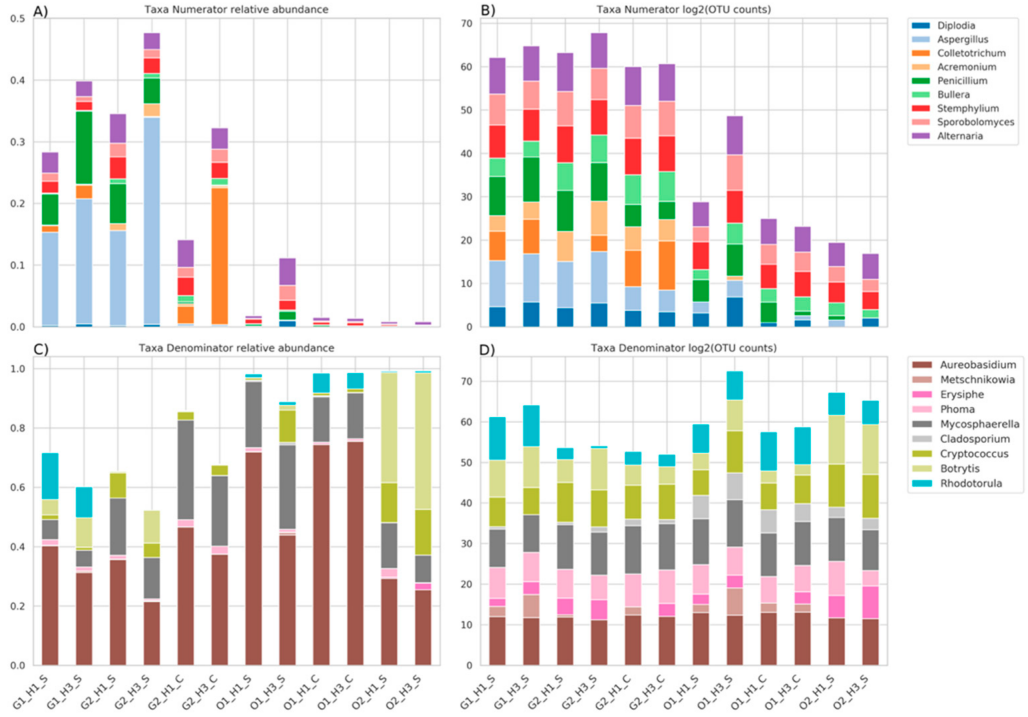


Figure 17: Plots AC represent relative abundances of the OTU counts for the genera of the rarefied OTU table. Plots BD represent log2 transformation of the OTU counts for the genera of the rarefied OTU table. Genera have been split into the groups Numerator (**A** and **B**) and Denominator (**C** and **D**) as defined by GNEISS.

The log2 transformation of the OTU counts in Figure 17 will be used as a means of explaining the calculated balances depicted in Figure 16, taking into consideration the aforementioned genera of Figure 15 that could greatly influence the resulted balance. Starting with the Shiraz samples from G1, the calculated balances are close to zero signifying that the abundances of the genera in the numerator are counter-balanced by the abundances of the genera in the denominator (Figure 16). For instance, the abundances of *Aspergillus*, *Colletotrichum*, and *Penicillium* are relatively equal to those of *Rhodotorula* and *Botrytis*. Therefore, the samples from G1 could conveniently be used as a reference

for explaining the change of the balances for the rest of the samples. Consequently, although the abundances of *Aspergillus* and *Penicillium* of the Cabernet samples from G2 are lower than the samples of G1, the higher positive balance of these samples could be attributed to the lower abundance of *Rhodotorula* and *Botrytis* as well as the higher abundance of *Colletotrichum* resulting in an overall more positive balance for G2 samples (Figure 16). Within the G2 Shiraz samples, *Aspergillus* and *Penicillium* had almost the same abundance as the G1 samples, whereas great difference was observed with *Rhodotorula* and *Colletotrichum* (Figure 17). Moreover, the H3 harvest period seems to have influenced positively the abundance of *Aspergillus* and *Colletotrichum*, comparing to H1, and quite negatively the abundance of *Rhodotorula* making these samples the ones with the highest positive balance. All the genera considered for the numerator in the balances had very low relative abundance and log2 of the OTU counts for O1 and O2 samples (Figure 17), resulting in negative balances (Figure 16). As far as the O1 region is concerned, the negative balance was the result of the very low abundances of *Aspergillus*, *Colletotrichum*, and *Penicillium* comparing to G1. Also, comparing the Shiraz and Cabernet samples from O1, the lower abundance of *Aspergillus* and *Colletotrichum* resulted in higher negative balance. Considering the harvest period, the genus that seems to be greatly affected for the Cabernet samples is *Penicillium*, whereas for the Shiraz samples it is *Botrytis*. The Shiraz samples from O2 followed the same pattern as the corresponding samples from O1 (Figure 16), however the higher abundance of *Botrytis* (Figure 17) compared to the rest of the samples was the reason for having the highest negative balance.

The OLS regression performed by GNEISS on the calculated balances using the regressors region, variety and harvest revealed a 75% of total variance explained. From the three factors, region is responsible for the 63% of the total observed variance, whereas variety for 11% and harvest period for 1%.

Discussion

Microbial communities on grape surfaces have been previously studied due to their perceived importance for contribution to wine characteristics, style, and quality [49, 108, 109]. Bacterial and fungal populations on the grape surface and the vine plant are affected by various biotic and abiotic factors, such as insects, interactions between resident populations, geography, climate, and viticultural practices [42, 48, 108]. Generally, many of these variables are not independent and may be clustered into broad groups of effects. Particular attention needs to be paid to the population dynamics of fungi

during grape berry development which may be related to the increased surface area of each berry, and to the availability of nutrients such as carbohydrates and organics acids [50, 54, 208]. Most of the studies analyzing microbial changes during the maturation of the grape have been based on culture-dependent analysis [22, 44, 50, 52–54]. Those studies found changes in structure and dynamics of the bacterial and fungal communities during grape maturation. However, the limitation of culture-dependent techniques to assess the real microbial diversity in natural environments is well recognized [67, 138]. Recent investigations have characterized significant and consistent changes in grape and wine composition and wine sensory profiles, associated with grape maturities at harvest and vineyard site [209, 210]. Besides, higher levels of carotenoids are present in grapes from hot or dry climates, or exposed vineyards to solar radiation [211].

In the present study we used barcoded sequencing to analyze the mycobiome of Cabernet Sauvignon and Shiraz grapes varieties sampled at two ripening times in four vineyards situated at two different Australian wine regions.

Our results show that fungal diversity of the grapes was mainly influenced by region while the varietal and harvest time had a slighter weight. As the climate at O is colder and drier than at G, the pattern of lower fungal species richness in the highest altitude regions hints that selection might have a role in determining these patterns. Within the two O vineyards, greater differentiation between fungal communities was observed than for the G vineyards probably due to a higher heterogeneity of the terrain and differences in altitude (about 270 m) between sampling points O1 and O2. In fact, previous studies have observed changes in microbial diversity and composition due to altitude and geographical orientation [49, 212]. In addition, within the O1 samples, the factors variety and harvest had significant influence on the fungal diversity while the weight of harvest or maturation time for the rest of samples was not significant. Except for the absence of *Acremonium* or *Colletotrichium* in most O samples, the list of genera is the same in the rest of samples, thus, the observed changes in alpha diversity should be due to changes in species within each genus. In fact, both genera are usually related to humid or moist climates, which would justify their low abundance in the area O that is qualified as dry.

Our results also show that the fungal community composition varied significantly across the different vineyards. The BrayCurtis distance metric was used for the clustering of the samples and indicated that the samples clustered significantly different by region, variety and harvest time, being again the region the factor that had the strongest effect on sample differentiation by taxa composition.

The most abundant genera across all samples were *Aureobasidium* and *Mycosphaerella*. While *Aureobasidium* has been frequently isolated and also detected by culture-independent

techniques in previous studies analyzing wine grape berries surfaces around the world [42, 43, 46, 51, 53, 166, 212], the genera *Mycosphaerella* has been seldom reported. Because several species of *Mycosphaerella* are considered plant pathogens, the presence of this genus may be directly related to the vine health. However, as different species of the same genus may behave totally different and *Mycosphaerella* was highly abundant in all the analyzed samples, it could represent a characteristic genus of the Australian grapes. In fact, previous studies using HTS analysis to analyze the grape microbiome have also suggested the presence of specific genera or species in different wine regions [42, 46, 112, 213]. Recently, Dissanayake *et al.* [214] identified both *Aureobasidium* and *Mycosphaerella* within the endophytic community in stems grapevine. Other epiphytic filamentous fungi usually associated with plant diseases and frequently found by HTS in the grape mycobiome were *Aspergillus*, *Botrytis*, *Colletotrichium*, *Rhodotorula*, and *Penicillium*. The results obtained by QIIME corroborated these findings and additionally GNEISS identified these five genera as the ones with the strongest effect on the balances driving the differentiation in fungal composition across samples. Most analyses of microbiome based on HTS relies on multivariate analysis. From the comparison between relative abundance and log2 OTU counts it becomes apparent that the constraint applied when the OTU counts are converted to compositional data could lead to misinterpretations. For instance, the fluctuation of the relative abundance of *Aureobasidium* across all samples seems significant. However, the log2 transformation of the OTU counts of *Aureobasidium* reveals a relatively stable abundance across all samples corroborating its identification as non-significant genus by the current analysis. Thus, the balances analysis applied during the present study lead to more realistic results than those of transforming OTU counts of genera to relative abundance.

Limitations

Metataxonomic analysis is notorious of incorporating various laboratory and bioinformatic procedures that assign a degree of inaccuracy to the overall analysis by introducing variability among the samples. The numerous factors that are associated with these procedures, along with the fact that the taxonomy and microbial abundance of the samples is unknown render difficult the estimation of the produced variability. Moreover, the nature of the microbial data necessitates the use on non-parametric statistics such as Kruskal-Wallis that do not make assumptions regarding the distribution of the OTU counts. Nevertheless, such non-parametric statistics may lose statistical power in small datasets. Finally, GNEISS does not provide an explicit information regarding differential

abundance of taxa. As the results demonstrate, it rather servers as an useful exploratory analysis tool aiming at producing comparative insights between the samples utilizing the concept of balances. This way it sets the basis for a subsequent controlled experimental design that focuses to analysing specific microbial dynamics.

Conclusions

This study used barcoded massive sequencing to analyze the effect of the grape ripening state, the vineyard region, and grape variety on the grape mycobiome. The results revealed that both fungal composition and diversity were mainly influenced by the vineyard region while the grape variety or the ripening state had less impact. However, within each region the fungal communities were affected differentially by the ripening state apparently due to the climatology. Even if the most abundant genera across samples were *Aureobasidium* and *Mycosphaerella*, the results obtained by GNEISS identified five genera with the strongest effect on the balances driving the changes in fungal composition. This result manifests that the analysis of the microbiome changes based on transformed OTU counts to relative abundance could lead to misinterpretations.

UNIVERSITAT ROVIRA I VIRGILI
ANALYSIS OF MICROBIAL POPULATIONS IN WINES THROUGH NGS METHODOLOGIES
Dimitrios Kioroglou

CHAPTER 5

High throughput sequencing approach to analyze the effect of ageing time and barrel usage on the microbial communities composition of red wines.

Dimitrios Kioroglou[1], Albert Mas[1] and Maria del Carmen Portillo[1†]

Frontiers in microbiology
under final edition

[1] Departament de Bioquímica i Biotecnologia, Universitat Rovira i Virgili, Marcel·lí Domingo s/n, 43007, Tarragona, Spain

[†] Corresponding author

UNIVERSITAT ROVIRA I VIRGILI
ANALYSIS OF MICROBIAL POPULATIONS IN WINES THROUGH NGS METHODOLOGIES
Dimitrios Kioroglou

Abstract

Wine aged in barrels or bottles is susceptible to alteration by microorganisms that affect the final product quality. However, our knowledge of the microbiota during ageing and the factors modulating the microbial communities is still quite limited. The present work uses high throughput (HTS) techniques to deal with the meta-taxonomic characterization of microbial consortia present in red wines along 12 months ageing. The wines obtained from two different grape varieties were aged at two different cellars and compared based on time of wine ageing in the barrels, previous usage of the barrels and differences between wine ageing in oak barrels or glass bottles. The ageing in barrels did not affect significantly the microbial diversity but changed the structure and composition of fungal and bacterial populations. The main microorganisms driving these changes were the bacterial genera *Acetobacter*, *Oenococcus*, *Lactobacillus*, *Gluconobacter*, *Lactococcus* and *Komagataeibacter* and the fungal genera *Malassezia*, *Hanseniaspora* and *Torulaspora*. Our results showed that the oak barrels increased effect on the microbial diversity in comparison with the glass bottles, in which the microbial community was very similar to that of the wine introduced in the barrels at the beginning of the ageing. Furthermore, wine in the bottles harbored higher proportion of *Lactobacillus* but lower of *Acetobacter*. Finally, it seems that one year of previous usage of the barrels was not enough to induce significant changes in the diversity or composition of microbiota through ageing compared with new barrels. This is the first meta-taxonomic study on microbial communities during wine ageing and shows that the microorganism composition of barrel-aged wines was similar at both cellars. These results hint the possibility of a common and stable microbiota after ageing in the absence of exogenous alterations. Further corroborations on the current outcome would be valuable for the comparison and detection of microbial alterations during ageing that could potentially prevent economic losses in the wine industry.

Introduction

Winemaking is a process in which *Saccharomyces cerevisiae* is the main yeast responsible of the alcoholic fermentation of grape must to produce wine. However, a wide diversity of yeast and bacterial species from the grape surfaces, the field or the cellar facilities and equipment might contribute to the final wine quality [2, 109]. These microorganisms can exert a positive or negative influence through all the winemaking process including wine maturation and ageing [50, 215]. The maturation and ageing process starts with the introduction of wine in wooden barrels and it continues after bottling until its consumption.

Nowadays, the wood barrel ageing is a common practice in winemaking for wine maturation of higher red quality wine [216]. The main reason is that barrel ageing improves many red wines from a visual, olfactory and gustatory point of view because the extractable compounds of the casks induce positive changes in the composition and flavour of the aged wine [217–220]. However, during ageing, microorganisms surviving the winemaking process, remain in the pores of the wood or occasional contaminants might produce metabolic compounds that can cause deviations from the olfactory optimum and spoilage of the wine. Bottle ageing is also susceptible to undesirable microbial growth even though fining or racking (filtration and clarification) are applied with the aim of microbiological stabilization [50, 221].

Several authors have manifested that controlling the growth of spoiler microorganisms is one of the most important challenges of the current winemaking process [222, 223]. This issue is critical for aged wines because of their added value. Multiple methods detect spoilage wine bacteria and yeast, but most are based on culture-dependent techniques [224–228]. This resistance phenomenon may be reversed when the environmental parameters change (SO_2, pH, O_2) and trigger additional fermentation start during barrel maturation or bottle-ageing. The metabolic activities of microorganisms at these stages might be detrimental to the wine flavour [50]. The introduction of molecular methods based on DNA has improved the detection of present cells even at low concentrations. However, most of the studies dealing with microbiological spoilage of wine are focused on the detection of a few specific species that had been previously associated with such deterioration [148, 229, 230]. Thus, the use of high throughput sequencing (HTS) techniques could provide a more realistic view of the complex microbial community present during wine ageing. These techniques have been recently used in wine samples mostly focusing in grapes, grape must or fermentation stages [42, 105, 109, 111, 181, 212, 231–235]. However, little attention has been paid to changes of microbial communities during

wine ageing process or factors driving its evolution. It is well recognized that factors affecting wine composition are grape variety, ageing time, wood origin along with its toasting level during barrel ageing [217–220] and SO_2 addition or the stopper composition during bottle aging [236, 237]. Nevertheless, the influence of these factors over the present microbial communities during ageing is not known.

In the present study, red wines were analysed during ageing to monitor the taxonomic composition of prokaryotic and eukaryotic communities by HTS of short amplicons of hypervariable domains of 16S rDNA gene and ITS1-ITS2. The ageing process of wines was performed in the only two Spanish Qualified Appellations of Origin; DOQ Priorat (Catalonia) and DOCa Rioja (Spain) regions. The factors considered for the comparison included time of wine ageing in the barrels, prior usage of the barrels and, in the case of Rioja wines, differences between wine ageing in oak barrels or glass bottles.

Methods

Samples

The barrels source of red wine samples were the traditional bordelaise barrel of 225L and made of French oak, mid-toasted. Two of them located in a winery of the DOQ Priorat (cellar Ferrer Bobet, FB) and the other two in the DOCa Rioja (Bodega institucional, Instituto de Ciencias de la Vid y el Vino, ICVV). Besides the high price of the HTS analysis, we have to consider that the sampled barrels had an increase probability of contamination or oxygenation due to sampling and that is the main reason to keep the number of barrels and the sampled volume low. The procedure of sampling and experimental set up is represented Figure 18. In each region the two barrels differed in time of usage, with one barrel being new, without any prior usage (BAN), while the other had been used for one year and is referred as old (BAO). Cleaning of the used barrels were done with the standard cellar practices (washing with pressurized hot water and rinsing). The main parameters of the wine before being introduced in the barrels were similar: 13.8 and 14,1% ethanol; pH 3,3 and 3,4; 0,29 and 0.34 g/L acetic acid; 4,4 and 4,3 g/L tartaric acid (total acidity); 80 and 90 ppm total SO_2 ; 1,5 and 1,2 g/L residual sugar; 0,88 and 0,94 g/L malic acid at the end of malolactic fementation for FB and ICVV, respectively. The barrels followed the habitual cellar management and were maintained with the rest of the barrels from the same vintage. In FB, grape variety was Carignan, which is the main and characteristic variety in DOQ Priorat, and the wine samples were collected at the end of malolactic fermentation inoculated with an autochthonous strain of *Oenococcus oeni*, completed inside BAO and BAN and denoted as 0 time-point, at the

time-points of 3, 6 and 9-months of barrel ageing from both barrels, and at the 12-month time-point from BAN only, as BAO was accidentally used to refill other barrels due to common practices in the cellar. On the other hand, the grape variety at ICVV winery was Tempranillo, which is the main variety in DOCa Rioja, and the wine samples were collected at the end of spontaneous malolactic fermentation, completed inside the steel tank and denoted as FML or 0 time-point, and after 3, 9 and 12-months of barrel ageing from both barrels. Additionally, the same day that the wine finished the FML at ICVV winery and transferred into BAO and BAN, a sample of 750mL from each barrel was taken and bottled into a dark glass bottled as the cellar uses for its wine commercialization. These bottle-aged wine samples, from the old (BTO) and new (BTN) barrel, were stored in the same cellar as the barrels and analyzed after 12-months of bottle ageing. At each sampling point, we sampled 3 bottles of 50mL of aged wines with a sterilized pipette of 100mL introduced into the barrel by a top overture and used for stirring the wine in the barrel and sampling. At the laboratory, we used one of the sample bottles and keep the others at -80°C . From the 50mL of one bottle, we used 3-10mL for plating on the different culture media (described at section 2.5) and 40mL were filtered through a 0.2 µm polycarbonate filter. The filter was frozen at -80°C and used to extract the DNA once all the samples had been collected in order to avoid differences due to DNA extraction. All the acronyms used for the samples and their descriptions are in Table 12.

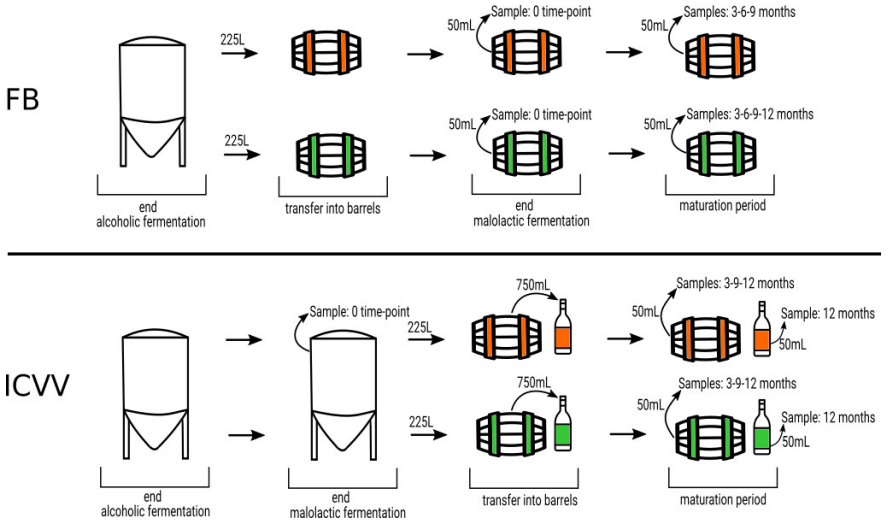


Figure 18: Schematic representation of the experimental setup. The green color represents new barrels and the orange color the used ones, at FB and ICVV cellars. Barrels of 225L were sampled at different time-points, expressed in months, taking 50 mL of wine. In FB, the malolactic fermentation was performed inside the steel tank, whereas in ICVV it was performed inside the barrels. In addition, in ICVV the moment the wine was introduced in the barrels, a sample of 750mL was taken and placed into glass bottles. The bottled wines were sampled after 12 months of maturation at cellar conditions.

Abbreviation	Description
BAO	barrel with 1-year prior usage.
BAN	barrel with no prior usage.
BTO	1-year wine sample that was sampled from BAO barrel and aged in bottles.
BTN	1-year wine sample that was sampled from BAN barrel and aged in bottles.
FML_0	malolactic fermentation completed inside steel tank.
BAO_0	malolactic fermentation completed inside BAO.
BAO_3	3-months wine sample aged in BAO.
BAO_6	6-months wine sample aged in BAO.
BAO_9	9-months wine sample aged in BAO.
BAO_12	12-months wine sample aged in BAO.
BAN_0	malolactic fermentation completed inside BAN.
BAN_3	3-months wine sample aged in BAN.
BAN_6	6-months wine sample aged in BAN.
BAN_9	9-months wine sample aged in BAN.
BAN_12	12-months wine sample aged in BAN.

Table 12: Description of the acronyms used for the samples of this study.

Bioinformatic and statistical analysis

The processing of the raw amplicon sequences has been performed using Quantitative Insights into Microbial Ecology (QIIME versions 1.9.1 and 2018.2) implementing the Illumina OTU pipeline steps previously described [238] with Phred33 quality filtering threshold of <20, 99% similarity threshold during OTU clustering and BLAST+ as taxonomic classification algorithm [33]. After quality filtering and taxonomic classification, exclusion of sequences matching to chloroplast or mitochondria was performed. The sequences obtained during this study have been included in the SRA database of the NCBI under the BioProject accession number PRJNA635684. Moreover, due to the nature of the OTU counts data, such as sparsity and lack of normality, as well as the compositionality constraint applied after converting the OTU counts to relative abundancies, nonparametric methodologies are necessary for the statistical analysis of the resulted OTU counts that do not depend on relative abundancies and assumptions [197, 198]. Therefore, in the current study we have implemented the compositional analysis toolbox GNEISS [40], as incorporated into Quantitative Insights Into Microbial Ecology framework (QIIME version 2019.1) [199].

Statistical analysis has been based on the factors barrel-type and time. For FB, the

factor barrel-type included the 0, 3, 6 and 9-month barrel-aged wine samples separated in the groups of old and new barrel resulting in 4-samples per group, whereas the factor time concerned the barrel-aged wine from old and new barrel grouped by the attributes 0, 3, 6 and 9-month timepoints leading to 2-samples per group. Similarly, for ICVV the factor barrel-type concerned the 3, 9 and 12-month barrel-aged wine samples divided into the groups of old and new barrel, and the factor time comprised the 4-groups of 3, 9, 12-month barrel-aged and 12-month bottle-aged 1 wine samples. Moreover, ICVV included the additional factor bottled-wine which included the 12-month barrel-aged and 12-month bottle-aged wine samples. Summarizing, for the statistical analysis the samples were distributed as 4 of old barrel compared to 4 of new barrels (at each cellar), 2 samples for each of the four time points (at each cellar) and 2 samples for glass bottle compared with barrel samples (just for ICVV cellar) (Table 12). Using the rarefied OTU table, alpha diversity was calculated based on the Shannon index and statistical significance at alpha-level 0.05 was evaluated using Students t-test for the factor barrel-type and ANOVA for the factor time. The rarefied OTU table became also the source for assessing the beta diversity that was based on the Bray-Curtis index since taxonomy was constrained at genus level. The resulted Bray-Curtis distance matrix became the input for principal coordinate analysis (PCoA) as well as permutational multivariate analysis of variance (PERMANOVA) using alpha-level 0.05 and the factors barrel-type and time. On the other hand, GNEISS utilized the unrarefied OTU table since it applies its own normalization. In a nutshell, GNEISS performed zero OTU counts imputation and clustering of the genera into two groups via Ward-hierarchical clustering. Upon these two groups, which are considered anti-correlated, GNEISS applied isometric log-ratio transformation with one group being the numerator and the other the denominator. Therefore, this transformation provides a log-ratio, which is referred as balance, that may have positive or negative value and reflects for a given sample the changes that might have occurred in the OTU counts of the genera from the numerator, the denominator or both in relation to another sample. Finally, the average impact of each one of the identified genera on the balances has been calculated by the following equation:

$$impact_i = \frac{1}{n}\sum_{j=1}^{n} 2^{|B_j - b_j|} \tag{2}$$

where $\boldsymbol{n}$ refers to the total number of samples, $\boldsymbol{B}_j$ to the log2-ratio of all OTU counts of the genera included in the numerator and denominator, and $\boldsymbol{b}_j$ to the log2-ratio for the j-th sample after subtracting the OTU counts of the i-th genus belonging to either the numerator or the denominator. Therefore, this impact represents the average fold-change

caused on the log2-ratio from a given genus. The greater the impact, the greater the influence of this genus on the balances calculated by GNEISS.

qPCR analysis

Quantitative PCR (qPCR) was performed on the extracted DNA to quantify the main microorganisms previously detected in wine according to Lleixà *et al.* [181]. The used primers anneal the ribosomal gene region and allowed the quantification of total yeasts, *Saccharomyces* genus, *Hanseniaspora* genus, *Starmerella bacillaris*, *Botrytis cinerea*, acetic acid bacteria (AAB) and lactic acid bacteria [181] and DBRUXF/DBRUXR for *Brettanomyces bruxellensis* [239]. Standard curves were created by plotting the Ct (Cycle threshold) values of the qPCR performed on dilution series of cells against the log input cells/mL. Samples and cultures for standard curves were analysed in triplicate.

Plate culture

Samples were serially diluted in sterile water and plated on (i) YPD medium (Glucose 2% Peptone 2%, Yeast Extract 1%, Agar 1.7%) and incubated at 28°C for 49h; (ii) modified WLN medium (Difco^{TM} WL Nutrient Medium, BD) with the addition of cycloheximide to suppress yeast growth (100 mg/L) and incubated from 7 to 10 days at 28°C ; (iii) MRS Agar medium [144] supplemented with 4 g/L L-malic acid, 5 g/L fructose, 0.5 g/L L-cysteine, 100 mg/L nystatin and 25 mg/l sodium azide adjusted to pH 5.0 and incubated at 28°C in a 10% CO_2 atmosphere; (iv) GYC-Ca Agar medium (glucose 5%, yeast extract 1%, $CaCO_3$ and agar 2%, pH 6.3) supplemented with 100 mg/L natamycin to suppress yeast growth and incubated at 28°C for 3-5 days under aerobic conditions. Appropriate dilution plates were counted. The YPD medium provided the total yeast counts, modified WLN medium is selective for Brettanomyces genus, whereas MRS medium and GYC-Ca provided LAB and AAB counts, respectively.

Results

Sequence analysis and alpha diversity

From the initial 1,066,085 16S amplicon raw sequences for FB and 1,520,976 for ICVV, a corresponding sequence filtering of 11% and 5% resulted after applying all the filtering steps, leading to a rarefication threshold of 30,815 sequences per sample for FB and 10,1243 sequences per sample for ICVV. The rarefaction curves of 16S and ITS sequences are included in the supplementary Figure S25. Shannon alpha diversity did

not reveal statistical significance for FB (Figure 19A) for any of the factors, whereas statistical significance (p-value 0.02) for the factor time was observed for ICVVs samples (Figure 19B). Overall, the barrel-aged wine from old and new barrel exhibited similar trends in both cellars, whereas the 12th month bottle-aged wine resulted in lower diversity than the 12th month barrel-aged wine in the case of ICVV.

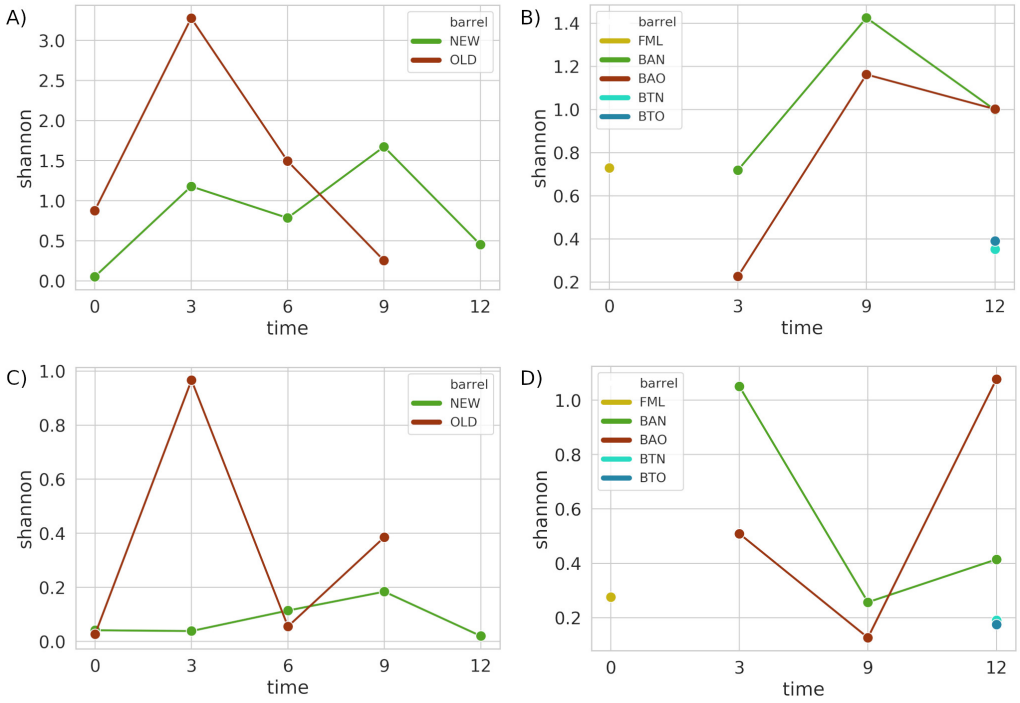


Figure 19: Alpha diversity based on Shannon index for FB 16S (A) and ITS (C) amplicon samples, as well as for ICVV 16S (B) and ITS (D) amplicon samples. Acronyms BAN or NEW and BAO or OLD refer to barreled wine from new and old barrel respectively, FML to final malolactic fermentation stage and BTN and BTO to bottled wine from new and old barrel respectively.

For the ITS amplicon, the initial 1,252,877 raw sequences of FB were filtered by 38% and by 29% the initial 1,366,522 raw sequences of ICVV, resulting in a rarefication threshold of 2,832 sequences per sample for FB and 2,381 sequences per sample for ICVV. Shannon alpha diversity was found of having statistically non-significant differences between the groups of the factors barrel-type and time for FB (Figure 19C) and ICVV (Figure 19D) with the bottle-aged wine having once again lower diversity than the barrel-aged wine.

Principal Coordinate Analysis of samples

After performing PCoA based on Bray-Curtis distance metric, bacterial communities presented a separation between early (<6) and late (>9) maturation wine samples across the first principal component for FB (Figure 21A) accompanied by statistical significance (p-value 0.05) for the factor time. Similar clustering of bacterial communities became apparent also across the first principal component between the early (<9) and late (>9) maturation wine samples of ICVV (Figure 21B) with a reported statistical significance (p-value 0.03) for the factor time. Along the same component, a clear separation between the barrel and bottle-aged wine bacterial communities could also be observed.

For FB, PCoA analysis based on Bray-Curtis of the fungal communities (Figure 21C) presented separation of the 3-month and 9-month BAO samples from the rest, however PERMANOVA reported non-significant differences between the groups of factors barrel-type and time. On the other hand, for the fungal communities of ICVV samples (Figure 21D) the reported statistical significance based on the factor time (p-value 0.03) seems to refer to the 3-month BAN and 12-month BAO samples due to their greater distance to the rest of the samples. Regarding the factor bottled-wine, a higher degree of separation could be observed between the fungal communities of BTO and BAO samples than between BTN and BAN samples.

Influence of studied factors on bacterial communities

The identified bacterial genera for FB along with their rarefied OTU counts are given in Table 13, and the calculated balances by GNEISS based on these bacterial genera are provided in Figure 22A. Overall, the balances did not show statistically significant differences between the groups of the factors barrel-type and time and the genera that seem to have greatly influenced the balances are *Acinetobacter*, *Cutibacterium*, *Lactobacillus*, *Pelomonas*, *Acetobacter* and *Oenococcus*. In Figure 20A the relative abundances of these genera are given, and in Figure 23A,B their log2-transformed OTU counts are shown.

The most abundant bacterial genus at FB was Oenococcus that had an average of over 80% of the bacterial sequences the first 6 months of maturation, followed by Acetobacter, which increased in abundance at the end of barrel maturation representing avobe 85% of the sequences at the 9th month samples (Figure 20A).

Regarding ICVV samples, Table 15 holds the rarefied OTU counts of the identified bacterial genera, Figure 22B provides their calculated balances, and their relative abundance is displayed in Figure 20C. Similarly to FB, the most abundant bacterial genus was *Oenococcus* that had an average of relative abundance above 60% the first 9 months of

maturation. However, even though Acetobacter increased at the end of barrel maturation, its abundance was slightly lower than in FB, representing 30 and 50% of the bacterial sequences at the 9th and 12th months of barrel maturation (Figure 20C). Interestingly, the relative abundances of the detected bacterial genera did not change after 12 month of glass bottle maturation (Figure 20C). The balances revealed statistically significant differences between the groups of the factor time with *Acetobacter*, *Oenococcus*, *Lactobacillus*, *Gluconobacter*, *Lactococcus* and *Komagataeibacter* being the main genera that drove these differences (Figure 24A,B). In both cellars *Acetobacter* and *Oenococcus* have been included in the same group by GNEISS as they have been identified of being correlated. *Acetobacter* exhibited increasing trend through time in both cellars, whereas the abundance of *Oenococcus* was relatively stable in ICVV but decreased over time in the case of FB. That could be explained by the fact that the initial wine samples of FB were taken at the end of malolactic fermentation where the abundance of *Oenococcus* was at higher levels. On the other hand *Lactobacillus* displayed different behaviour between the two cellars, with an increasing tendency in the case of FB and a decreasing one in the case of ICVV. Overall, the barrel-aged wine did not present great differences between the old and new barrels. The observed differences between the 12th month bottle and barrel-aged wine, could be attributed to *Acetobacter* and *Lactobacillus* whose abundances in the bottle-aged wine were similar to those of the early maturation period ($<$9).

The analysis of qPCR showed a deep decrease with time of LAB after the introduction of the wine in barrels but the number of AAB remained constant with time in the barrels and also in the glass bottles (Table 17). The cfu of LAB on plates were just one order bellow the cells detected by qPCR. However, no cells were recovered on the medium for AAB (Table 17).

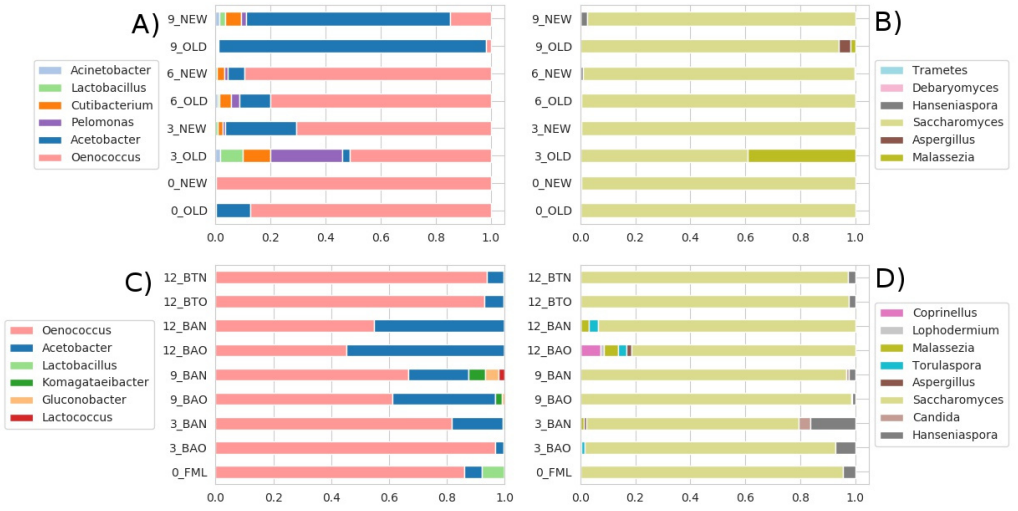


Figure 20: Relative abundances for FB 16S (A) and ITS (B) amplicon samples, as well as for ICVV 16S (C) and ITS (D) amplicon samples. Acronyms BAN or NEW and BAO or OLD refer to barreled wine from new and old barrel respectively, FML to final malolactic fermentation stage and BTN and BTO to bottled wine from new and old barrel respectively.

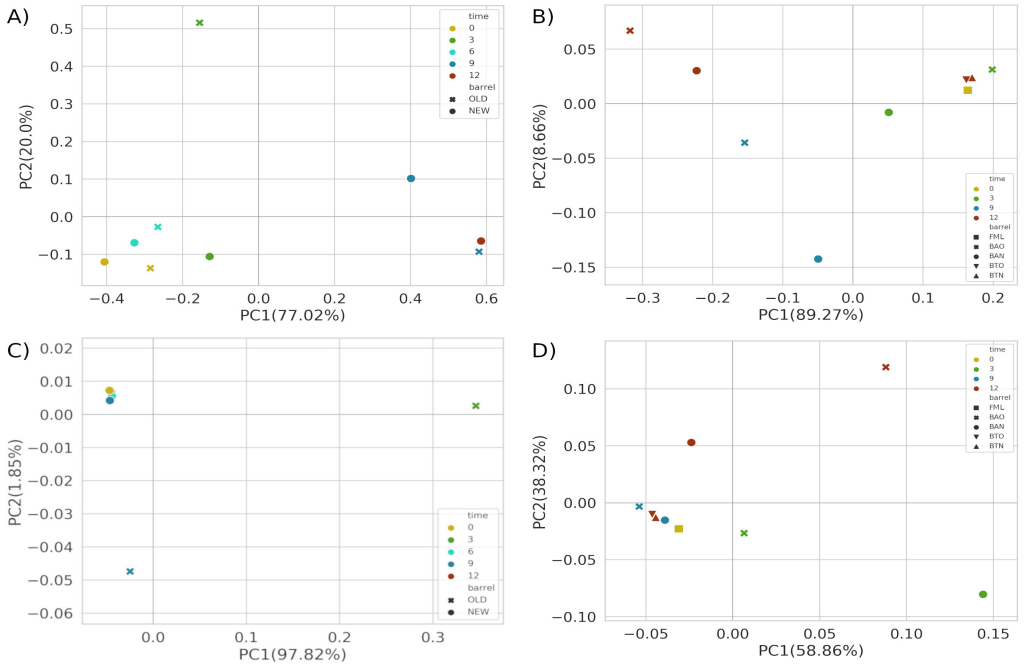


Figure 21: PCoA based on Bray-Curtis distance metric for FB 16S (A) and ITS (C) amplicon samples, as well as for ICVV 16S (B) and ITS (D) amplicon samples. Acronyms BAN or NEW and BAO or OLD refer to barreled wine from new and old barrel respectively, FML to final malolactic fermentation stage and BTN and BTO to bottled wine from new and old barrel respectively.

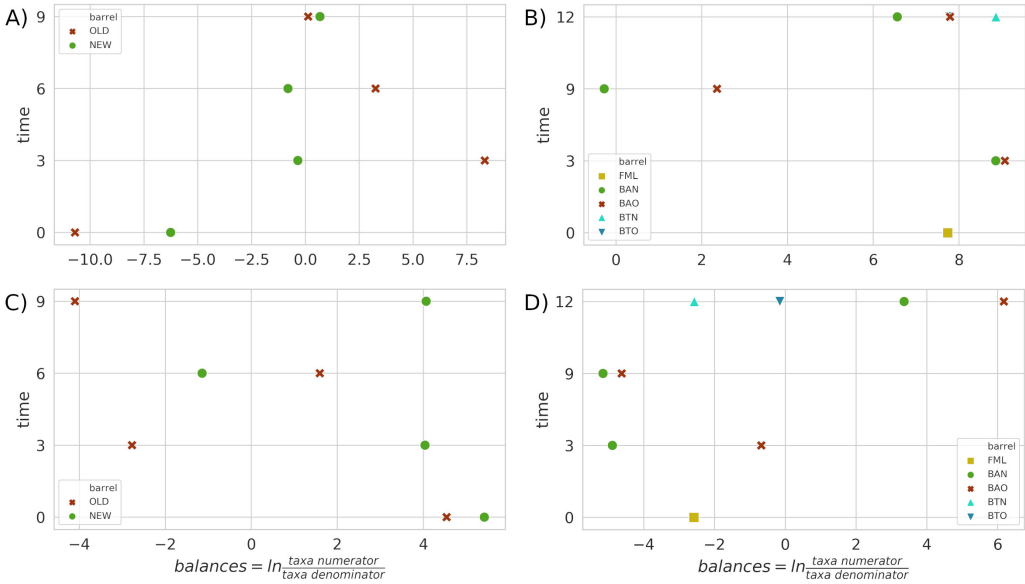


Figure 22: Calculated balances by GNEISS for FB 16S (A) and ITS (C) amplicon samples, as well as for ICVV 16S (B) and ITS (D) amplicon samples. Acronyms BAN or NEW and BAO or OLD refer to barreled wine from new and old barrel respectively, FML to final malolactic fermentation stage and BTN and BTO to bottled wine from new and old barrel respectively.

taxonomy	FB_BAO_0	FB_BAO_3	FB_BAO_6	FB_BAO_9	FB_BAN_0	FB_BAN_3	FB_BAN_6	FB_BAN_9	FB_BAN_12
Leuconostoc [a]	0	1	0	0	0	27	0	0	0
Bradyrhizobium [a]	0	604	0	1	1	0	52	0	0
Lactobacillus [a]	0	1570	221	134	4	187	169	633	276
Corynebacterium [a]	0	926	0	0	0	1	0	0	0
Delftia [a]	3	1576	142	0	0	25	219	741	0
Staphylococcus [a]	4	622	146	28	1	58	0	257	109
Streptococcus [a]	0	0	134	0	0	100	233	501	145
Deinococcus [a]	0	0	1518	1	0	0	0	0	0
Stenotrophomonas [a]	2	1547	155	2	0	0	0	72	0
Pelomonas [a]	21	4929	917	106	24	313	387	510	274
Acinetobacter [a]	13	295	184	18	1	0	0	359	477
Thermicanus [a]	0	863	0	0	0	0	0	0	0
Rothia [a]	0	398	0	0	0	56	0	1	0
Cloacibacterium [a]	0	3253	0	0	0	0	0	0	0
Vulcaniibacterium [a]	0	0	197	19	0	0	0	0	0
Pseudomonas [a]	4	1667	0	25	0	32	0	229	57
Bacteroides [a]	0	13	0	0	0	0	0	0	0
Cutibacterium [a]	0	1919	1183	75	3	594	824	1679	230
Glutamicibacter [a]	0	400	3	0	0	0	0	3	0
Oenococcus [b]	25553	9697	22808	493	30679	21541	27101	4305	20
Acidovorax [b]	1032	3	0	0	0	0	0	1	1
Aquabacterium [b]	474	1	0	0	0	0	0	138	0
Komagataeibacter [b]	0	0	0	0	9	0	0	0	0
Lonsdalea [b]	2	0	0	0	0	0	0	0	0
Acetobacter [b]	3669	531	3207	29913	55	7870	1830	21385	29158
Candidatus Finniella [b]	3	0	0	0	0	0	0	0	0

Gluconobacter [b]	0	0	0	0	21	11	0	1	0
Flavobacterium [b]	35	0	0	0	17	0	0	0	68

Table 13: Rarefied OTU counts for FB (Priorat) cellar 16S taxonomy. Genera included in the numerator of the calculated balances by GNEISS are denoted with (a) whereas those included in the denominator are denoted with (b). BAN, Barrel New; BAO, Barrel Old.

taxonomy	FB_BAO_0	FB_BAO_3	FB_BAO_6	FB_BAO_9	FB_BAN_0	FB_BAN_3	FB_BAN_6	FB_BAN_9	FB_BAN_12
unidentified [a]	1	0	1	1	1	1	0	2	0
Trametes [a]	0	0	0	0	0	0	0	6	0
Debaryomyces [a]	0	0	0	0	0	0	0	0	1
Hanseniaspora [a]	6	0	13	1	11	10	32	63	0
Saccharomyces [a]	2825	1720	2816	2663	2820	2821	2793	2761	2827
Aspergillus [b]	0	0	0	114	0	0	0	0	0
Malassezia [b]	0	1112	2	53	0	0	7	0	4

Table 14: Rarefied OTU counts for FB (Priorat) cellar ITS taxonomy. Genera included in the numerator of the calculated balances by GNEISS are denoted with (a) whereas those included in the denominator are denoted with (b). BAN, Barrel New; BAO, Barrel Old.

taxonomy	Rioja_FML_0	Rioja_BAN_3	Rioja_BAN_9	Rioja_BAN_12	Rioja_BAO_3	Rioja_BAO_9	Rioja_BAO_12	Rioja_BTN_12	Rioja_BTO_12
Oenococcus [a]	87265	82877	67457	55563	97918	61883	45720	95026	94129
Acetobacter [a]	5996	17848	21020	45646	2935	35985	55457	5893	6715
Lactobacillus [a]	7796	494	20	4	380	26	30	295	364
Komagataeibacter [b]	18	6	6000	10	8	2443	10	0	5
Gluconobacter [b]	159	11	4653	13	1	821	25	27	27
Lactococcus [b]	9	7	2093	7	1	85	1	2	3

Table 15: Rarefied OTU counts for ICVV (Rioja) cellar 16S taxonomy. Genera included in the numerator of the calculated balances by GNEISS are denoted with (a) whereas those included in the denominator are denoted with (b). FML, Final Malolactic Fermentation; BAN, Barrel New; BAO, Barrel Old; BTN, Bottle New; BTO, Bottle Old.

taxonomy	Rioja_FML_0	Rioja_BAN_3	Rioja_BAN_9	Rioja_BAN_12	Rioja_BAO_3	Rioja_BAO_9	Rioja_BAO_12	Rioja_BTN_12	Rioja_BTO_12
Coprinellus [a]	0	0	0	0	0	0	178	0	0
Lophodermium [a]	0	0	0	0	0	0	28	0	0
Malassezia [a]	0	32	3	77	11	0	124	0	0
Torulaspora [a]	2	0	1	79	27	0	71	1	0
Aspergillus [a]	1	21	0	0	0	0	44	0	2
Saccharomyces [a]	2273	1838	2298	2225	2168	2344	1936	2313	2321
Candida [b]	0	99	21	0	0	7	0	0	0
Hanseniaspora [b]	105	391	58	0	175	30	0	67	58

Table 16: Rarefied OTU counts for ICVV (Rioja) cellar ITS taxonomy. Genera included in the numerator of the calculated balances by GNEISS are denoted with (a) whereas those included in the denominator are denoted with (b). FML, Final Malolactic Fermentation; BAN, Barrel New; BAO, Barrel Old; BTN, Bottle New; BTO, Bottle Old.

A)

		FB								
		BAO 0	**BAO 3**	**BAO 6**	**BAO 9**	**BAN 0**	**BAN 3**	**BAN 6**	**BAN 9**	**BAN 12**
qPCR	Total yeast	2,95E+06	5,46E+03	4,12E+05	1,98E+04	2,14E+06	6,71E+05	2,81E+05	2,05E+04	4,06E+04
	S. cerevisiae	1,22E+06	1,67E+03	1,51E+05	1,95E+03	6,77E+05	3,16E+05	1,95E+03	5,41E+03	2,29E+04
	S. bacillaris	2,28E+04	-	-	-	2,44E+04	-	-	-	-
	H. uvarum	2,71E+04	-	-	-	2,66E+04	1,23E+04	-	-	-
	B. bruxellensis	-	-	-	6,80E+01	-	-	2,70E+01	-	-
	LAB	1,27E+06	-	-	-	2,08E+06	1,59E+04	-	-	-
	AAB	5,31E+06	-	-	2,06E+06	2,06E+06	8,76E+05	2,06E+06	1,85E+05	8,62E+05
Plate count	YPD (Yeasts)	1,01E+03	2,00E+02	1,24E+03	5,00E+01	4,25E+02	9,00E+01	1,70E+02	7,35E+02	5,70E+02
	WLN (Brettanomyces)	-	-	-	9,50E+01	-	-	-	-	-
	MRS (LAB)	1,67E+05	1,30E+03	-	-	1,13E05	-	-	-	-
	GYC-Ca (AAB)	-	-	-	-	-	-	-	-	-

B)

		ICVV								
		FML 0	**BAO 3**	**BAO 9**	**BAO 12**	**BAN 3**	**BAN 9**	**BAN 12**	**BTO 12**	**BTN 12**
qPCR	Total yeast	2,34E+06	4,76E+03	6,34E+04	3,39E+03	6,80E+05	1,97E+04	1,16E04	2,14E+05	9,28E+05
	S. cerevisiae	5,97E+05	1,59E+03	2,30E+04	6,58E+02	1,87E+04	5,23E+04	3,74E+03	8,26E+04	2,87E+05
	S. bacillaris	4,12E+05	-	-	-	-	-	-	5,39E+04	3,72E+04
	H. uvarum	2,14E+04	-	-	-	3,92E+03	1,33E+04	-	2,56E+04	8,10E+04
	B. bruxellensis	-	-	2,14E+03	3,06E+02	2,74E+02	5,34E+02	-	2,12E+02	-
	LAB	3,58E+06	1,59E+06	2,08E+05	1,06E+05	4,38E+06	2,41E+05	6,48E+04	6,41E+05	1,69E+06
	AAB	2,98E+06	1,40E+06	4,06E+06	2,80E+06	2,13E+06	2,26E+06	1,85E+06	3,09E+06	3,41E+06
Plate count	YPD (Yeasts)	1,30E+03	4,50E+02	3,00E+02	2,50E+02	2,38E+02	7,00E+01	9,50E+01	5,20E+02	3,00E+02
	WLN (Brettanomyces)	-	-	-	-	-	-	-	-	-
	MRS (LAB)	6,72E+05	1,20E+04	-	-	3,15E+05	-	-	-	-
	GYC-Ca (AAB)	-	-	-	-	-	-	-	-	-

Table 17: Mean values of qPCR (cells/mL) and plate culture count (cfu) analysis on the studied samples form FB (A) and ICVV (B)

Influence of studied factors on fungal communities

The identified fungal genera are reported in Table 14 for FB and in Table 16 for ICVV. The calculated balances of the fungal detected genera are displayed in Figures 22C and 22D and their corresponding relative abundances are displayed in Figures 20B and 20D, respectively. According to the relative abundances of the detected fungal genera (Figure 20B,D), *Saccharomyces* was the predominant yeast at both cellars with an average of 90% of the sequences. Secondly, *Malassezia* was detected at the FB old barrels just at the 3rd and 9th month samples, whereas *Hanseniaspora* was detected at ICCV the firsts 9 months of the new and old barrel maturation (6%) and at the 12-month bottle-aged samples (3%). Statistical analysis based on the balances of fungal genera verified the statistically non-significant differences between the groups of barrel-type and time for FB and showed statistically significant differences for the factor time in ICVV (p-value 0.03). In both cellars the most abundant genus was *Saccharomyces* with relatively stable abundance over time, and the genera mainly responsible for the observed differences between the samples were *Malassezia* and *Hanseniaspora* for FB (Figure 23C,D) and *Malassezia*, *Hanseniaspora* and *Torulaspora* for ICVV (Figure 24C,D). GNEIIS analysis of fungal genera identified *Hanseniaspora* and *Malassezia* as anti-correlated in both cellars. The statistical significance for the factor time in ICVV appear to concern the higher number of fungal genera of the 3-month and 12-month old barrel samples in comparison to the rest (Figure 24C,D). Finally, once again the fungal communities of the 12-month bottle-aged wine were different than that of the barrel-aged wine with the BTN and BTO samples having similar composition to FML samples (Figure 24C).

Taking into account qPCR results, the number of yeast represented mostly by *S. cerevisiae*, trended to decrease with ageing 2 or 3 orders of magnitude. However, yeast number in the bottles remained constant or decreased just one order (Table 17). The detection by qPCR of yeasts species other than *S. cerevisiae* was non important except in the case of *S. bacillaris* at ICCV wine at the beginning of ageing. The number of yeast cells recovered in plate media was three order of magnitude lower than the number detected by qPCR but also trend to decrease with ageing time (Table 17).

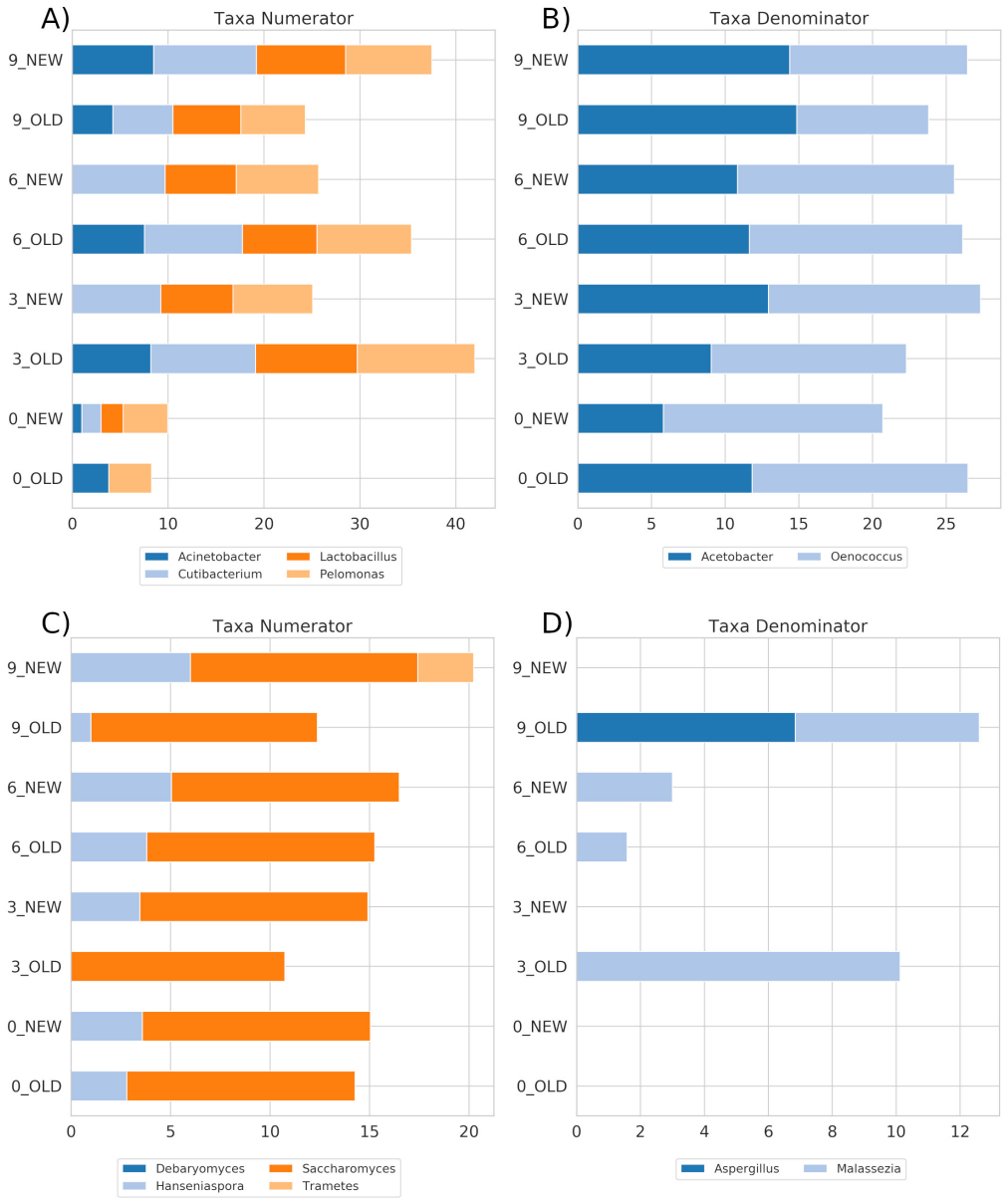


Figure 23: FB log2-transformed OTU counts of the genera with the highest impact on the balances for the amplicons 16S (A,B) and ITS (C,D) that correspond to the numerator or the denominator of the balances. The indicators NEW and OLD refer to the new and old barrel respectively and numbers at the beginning of the indicators refer to the sampling period.

Figure 24: ICVV log2-transformed OTU counts of the genera with the highest impact on the balances for the amplicons 16S (A,B) and ITS (C,D) that correspond to the numerator or the denominator of the balances. Acronyms BAN and BAO refer to barreled wine from new and old barrel respectively, FML to final malolactic fermentation stage and BTN and BTO to bottled wine from new and old barrel respectively. Numbers at the beginning of the acronyms refer to the sampling period.

Discussion

Although deterioration of ageing wines in barrels or bottles caused by microorganisms consists a real threat for product quality and marketability, our knowledge on the microbial status of those wines is still quite limited. The present work is one of the first studies dealing with the meta-taxonomic characterization of microbial consortia during ageing of wines using a HTS approach. To our knowledge, just one previous meta-taxonomic study included samples from aged wine prior to bottling but it was mainly focused on the vineyard microbiota and its correlation with the chemical composition of the finished wines [240]. That study, used HTS at aged wines and detected high number of bacterial and fungal genera allowing to differentiate wines from different vineyards [240]. In addition, for the first time, our work used a meta-taxonomic approach to monitor the effect of time of ageing in the barrels, prior usage of the barrels and ageing in glass bottles over the bacterial and eukaryotic communities.

Diverse studies found that species diversity trend to decrease over time during the winemaking process, being the two most significant decrease during alcoholic fermentation and after SO_2 addition once fermentation is finished [228]. The lower species diversity could be explained by the stressing conditions (high ethanol concentration, low pH and scarcity of nutrients) that characterise the process resulting in a strong selection.

However, through the ageing period the microbial diversity has been shown to be relatively constant although the number of cells tend to decrease [150]. Overall, our results showed non-significant changes in microbial diversity during ageing for the factor ageing-time and barrel-type. The exception was found for ICVV samples that showed an inflexion point in bacterial and fungal diversity over time at 9 months of ageing, harbouring the 3 and 12 month samples higher diversity. Meanwhile, our qPCR results pointed to a decrease in yeast and LAB numbers through time whereas AAB population were kept constant, as previously observed [229]. However, no AAB cell were recovered by culturing in each stage manifesting the difficulty to grow them in laboratory conditions or their VBNC state [19, 241]. These results point out that either the death of the cells or their entrance into VBNC state in the barrels, lead to a decreased cell number but nonetheless a scarce fluctuation of their diversity based on the factors time or barrel-type. However, the 12 month bottle-aged wine resulted in significantly lower diversity of bacterial and fungal communities compared to the 12 month barrel-aged wine while the number of cells were similar or even one order higher than in the bottles, suggesting that the barrel have a positive influence on the microbial diversity. In fact, the diversity and number of cells of bottle samples were similar to those of the wine samples just introduced in the barrels

after FML. Wood is more or less porous depending on being the origin of the wood. For example, American and French oak have different porosity and their absorbent structure allows progressive microbial penetration, especially during the first time it is used [50]. These microorganisms can develop when they come into contact with wine, increasing the diversity and with a possibly harmful effect on wine quality.

The relative abundances alone of the detected bacterial and fungal genera give little information about the microbial communities changes according to the studied factors. Thus, in this study, we analysed the calculated balances by GNEISS based on these genera. Even if no profound changes were observed for microbial diversity across ageing-time in barrels, bacterial communities of early and late maturation differed significantly at both ICVV and FB wines. The bacterial changes between early and late ageing are in agreement with the chemical evolution previously observed for the wines during ageing [217–220]. On the other hand, fungal communities behaved differently at the two cellars but they have in common that the final communities at 12th months of barrel ageing were similar to the samples of the initial ageing, harbouring the intermediate ageing samples different fungal communities. The HTS and qPCR techniques based on DNA do not allow the differentiation between lives, VBNC or death cells. But the combination of both techniques allowed to known that number of yeasts was decreasing with time while their structure and diversity was not changing deeply. Also, the cells detected by plate culture were the those culturable while the rest detected by qPCR but not recovered on plates would be either dead or in VBNC state. The VBNC state may be reversed when the environmental parameters are adequate and the metabolic activities of recovered microorganisms might be detrimental to the wine flavour [50].

The number of times the barrels are used determines the oak composition and the rate of chemical compounds extracted from the wood [218]. Similarly, it is well known that aroma of wines aged in oak barrels differ significantly from that aged in glass bottles [242]. Thus, changes in the concentration of different compounds during oak ageing due to those factors could potentially affect the population of microorganisms in the samples. In fact, in our study 12 month barrel and bottle-aged wines harboured different bacterial communities. This was also the case for fungal communities specially for wine in older barrels. Changes in microbial composition together with the higher diversity observed in the barrel respect the glass bottle indicated at possible positive effect of the former on the development of new species even if the total number of yeasts and LAB was decreasing with time. However, the factor barrel-type did not influence significantly the bacterial or fungal communities composition, probably because just one year of barrel usage was not enough to infer deep changes.

Nisiotou and Gibson [227] were the first to study culturable yeast on bottled wine and the yeast isolates were mainly *Brettanomyces bruxellensis*, *Saccharomyces cerevisiae* and *Rhodotorula pinicola*. Other microorganisms like the bacteria *Oenococcus oeni* or *Pediococcus parvulus* or the yeasts *Pichia anomala* or *Zygosacharomyces bailii* have been previously isolated and detected during wine ageing [228]. Furthermore, Rubio *et al.* [243] found that *Brettanomyces* presence (cfu/ml and strains) and ethylphenol production during ageing was affected more by the ageing conditions (aerobic/anaerobic and sulfiting) than by the origin of the oak. However, most of the microbiological studies of ageing wines have been focused on specific spoilage microorganisms and their effects over wine quality. In our study, the HTS allowed us the detection of high diversity of bacterial and fungal genera in the absence of any sign of wine spoilage or off-odors according to cellar monitoring. Overall, the observed changes in bacterial communities across the different studied factors resulted from changes in the balances of the genera *Acetobacter*, *Oenococcus*, *Lactobacillus*, *Lactococcus* and *Komagataeibacter*, with the two firsts being the most abundant at both cellars. In the case of fungal genera, the most abundant was *Saccharomyces* and together with *Malassezia* and *Hanseniaspora* (and *Torulaspora* for ICVV) determined the differentiation of the samples at intermediate ageing time. The fact that *Hanseniaspora* and *Malassezia* have been identified as anti-correlated in both cellars. could suggest an underlying competition between these two genera. *Saccharomyces*, *Hanseniaspora* and *Torulaspora* are frequently reported genera in wine using either culture or molecular based techniques [1, 111]. *Malassezia* though, is found on the skin surface thus is possible that this yeast-like fungus contaminated the samples or the DNA during extraction. Nevertheless, this genus has been recently reported in studies of must and wine samples using HTS [244, 245]. This methodology has detected minor and rare species that are sometimes overlooked with culture-dependent methods and can detect non-culturable cells at the end of fermentation. However, it is still unknown if these microorganisms have wine environments as their natural niche and have a specific role during winemaking or if they are simply contaminants.

Our results showed that two wines from two cellars obtained from different grape varieties and aged under different conditions resulted in a common number of genera indicating that the microbial community detected could be the normal in the absence of wine deterioration. However, Bokulich *et al.* [109] used also HTS on wine samples after several months of barrel ageing and detected more than 95% of bacterial sequences belonging to *Leuconostoc* (same family than *Oenococcus*) and fungal sequences related to *Cladosporium*, *Botrytis* and *S. cerevisiae*, in that order of abundance and accounting over 80% of the eukaryotic sequences. Herein, further studies using the newest sequencing

technologies would be necessary to elucidate the regular microbial communities during wine ageing.

Conclusions

Barrel ageing of wine improves its organoleptic characteristics by the physicochemical reactions occurring between wine and wood compounds. The microorganism are supposed not to interfere or have a relevant role during wine ageing unless uncontrolled growth occurs, affecting thus wine quality. A plethora of studies about the presence of contaminant microorganisms and their by-products both during winemaking and barrel ageing are available. However, it is still missing an holistic view of the normal microbiota of aged wine and how the different factors and management affect that microbiota. In the present study we have used HTS of short amplicons to characterize the bacterial and fungal communities of wines aged for 12 months. The ageing in barrels did not affect significantly the microbial diversity with time but changed the structure and composition of fungal and bacterial population. Also, the barrels exert a positive effect on the microbial diversity in comparison with the glass bottles, in which the microbial communities were very similar to those of the samples at the beginning of the ageing. Finally, one year difference in the usage of the barrels was not enough to induce significant changes in the diversity or composition of wine microbiota through ageing. Our results showed that wines from different grape varieties and from different cellars, aged under different conditions resulted in a similar microbial composition. Nevertheless, more studies would be necessary to know if that microbiota is the standard after barrel ageing or if other factors not considered here could influence it.

UNIVERSITAT ROVIRA I VIRGILI
ANALYSIS OF MICROBIAL POPULATIONS IN WINES THROUGH NGS METHODOLOGIES
Dimitrios Kioroglou

CHAPTER 5: SUPPLEMENTARY MATERIALS

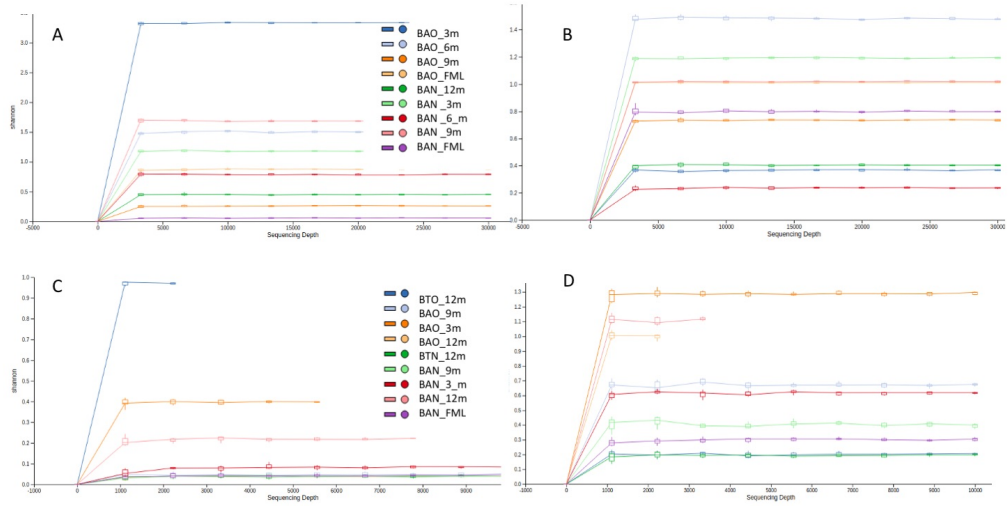


Figure S25: Rarefaction curves based on Shannon index of 16S (A, B) and ITS (C, D) amplicons obtained for FB (A, C) and ICVV (B, D) samples.

UNIVERSITAT ROVIRA I VIRGILI
ANALYSIS OF MICROBIAL POPULATIONS IN WINES THROUGH NGS METHODOLOGIES
Dimitrios Kioroglou

CHAPTER 6

Qualitative factor-based comparison of NMR, targeted and untargeted GC-MS and LC-MS on the metabolomic profiles of Rioja and Priorat red wines.

Dimitrios Kioroglou[1], Albert Mas[1] and Maria del Carmen Portillo[1†]
submitted

[1] Departament de Bioquímica i Biotecnologia, Universitat Rovira i Virgili, Marcel·lí Domingo s/n, 43007, Tarragona, Spain
[†] Corresponding author

UNIVERSITAT ROVIRA I VIRGILI
ANALYSIS OF MICROBIAL POPULATIONS IN WINES THROUGH NGS METHODOLOGIES
Dimitrios Kioroglou

Abstract

Wine origin and ageing are two factors related to wine quality which in turn is associated to wine metabolome. In the current study, we implement NMR, targeted and untargeted GC-MS and LC-MS metabolomic analytical techniques so as to gain insights into the volatile and non-volatile wine metabolome composition of red wines from two cellars located in the only two Spanish Qualified Appellations of Origin; DOQ Priorat and DOCa Rioja regions. The main goal was to qualitatively evaluate these metabolomic methodologies on their ability to provide informative patterns regarding wine metabolome based on a set of factors, such as ageing of barrel-aged wine (factor time), prior usage of the barrels (factor barrel-type) and differences between wine ageing in barrels or glass bottles (factor bottled-wine). Overall, 95 differentially significant metabolites have been identified facilitating the evaluation of the analytical methodologies performance. The results did not favor NMR as an effective technique on the current dataset whereas suggested LC-MS as an adequate technique for revealing differences based on the factor time, targeted GC-MS on the factor barrel-type and untargeted GC-MS on the factor bottled-wine. Overall, a combination of different metabolomic techniques is necessary for a complete overview of the metabolome changes.

Introduction

Wine is a complex hydroalcoholic solution including hundreds to thousands of different molecules (e.g. sugars, amino acids, organic acids, lipids, phenolics, alkaloids, sterols, lignans, terpenes, fatty acids). These compounds account for the metabolomic profile of the wine which confers and modulates the quality and sensory properties of the final wine [56, 246]. Several factors are involved in the wine metabolome and quality, such as the grape variety, the yeast and bacteria performing the alcoholic and malolactic fermentations, the winemaking practices (e.g. SO_2 addition, fining agents) and ultimately, the ageing process [56, 246–248].

Wine ageing process is the period that starts at the end of winemaking with its introduction in wooden barrels and continues after bottling until its consumption. Barrel ageing improves wine stability, colour, aroma and flavour. It is well recognized that the main factors related to the quality of barrel aged wines are the wine composition, ageing time and wood composition along with its toasting level [218]. In addition, after barrel ageing, factors such as storage conditions, SO_2 addition and stopper composition may also influence wine chemical composition during bottle aging [236, 237, 249].

Overall, the chemical composition of ageing wine reflects the history and conditions during ageing and storage [56]. For instance, it is well recognized that compounds such as 5-methylfurfural are formed during the toasting process and later transfered to the wine during the ageing process, whereas other such as 4-ethylpenol and 4-ethylguaiacol have microbial origin [250, 251]. Some authors have investigated the effect of selected factors on the metabolomic profile of wines during ageing using mainly targeted metabolomic or single analytical techniques [218, 249, 252]. The most widely analytical techniques used in wine metabolomics are gas chromatography-mass spectrometry (GC-MS), high-performance liquid chromatography-mass spectrometry (HPLC-MS) and nuclear magnetic resonance (NMR). GC-MS has preferably been used to profile wine volatile metabolites whereas HPLC-MS has been the most widely used for non-volatiles. NMR spectroscopy gives a reproducible direct identification and quantification of a broad range of analytes without sample pre-treatment [253]. Though, NMR is limited as it is unable to detect metabolites that are present in very small concentrations. Nevertheless, it has been generally accepted that untargeted analysis is needed for a more comprehensive and holistic analysis [56, 57]. The use of the kinetic correlations in time-dependent processes as wine ages can further contribute to metabolomic monitoring, discovery of new biomarkers and metabolic network investigations. Frequently, data are analyzed by multivariate statistical methods but the choice of the proper statistical treatment plays an

important role in drawing conclusions. The most frequently implemented methodologies include Principal Component Analysis (PCA), correlation analysis, ANOVA, t-tests and hierarchical clustering analysis. However, the reported statistical significance alone does not provide enough evidence for the importance of the findings without estimating the magnitude of the effect [254]. Thus, the determination of the practical significance (effect size) of different conditions or treatments is also of outstanding importance so as to discern the most relevant changes in metabolites.

The aim of this study was to combine different metabolomic analytic techniques (NMR, targeted and untargeted GC-MS and untargeted LC-MS) to monitor and compare the kinetic evolutions of the detected metabolites between red wines aged in oak barrels from two cellars located in the only two Spanish Qualified Appellations of Origin; DOQ Priorat (Catalonia) and DOCa Rioja (Spain) regions. The factors considered for the comparison included time of wine ageing in the barrels, prior usage of the barrels and, in the case of Rioja wines, differences between wine ageing in oak barrels or glass bottles. These factors are shortly referred as time, barrel-type and bottled-wine respectively in the study. Moreover, it needs to be underlined that among the aims of the current study is not the direct comparison of the wine metabolome between the two cellars, as other factors such as grape variety and climatic conditions would render such comparison incoherent. Instead, the study focuses on the comparison between the different metabolomic analytical techniques in relation to their ability to reveal informative patterns regarding the wine metabolome.

Methods

Samples

French oak mid-toasted barrels were the source of red wine samples. Two of them located in a winery of the DOQ Priotat (cellar Ferrer Bobet, FB) and the other two in the DOCa Rioja (Bodega institucional, Instituto de Ciencias de la Vid y el Vino, ICVV). In each region the two barrels differed in time of usage, with one barrel being new, without any prior usage (BAN), while the other had been used for one year and is referred as old (BAO). The barrels followed the habitual cellar management and were maintained with the rest of the barrels from the same vintage. In FB, grape variety was Carinyena, which is the main and characteristic variety in DOQ Priorat, and the wine samples were collected at the end of malolactic fermentation, completed inside BAO and BAN and denoted as 0 time-point, at the time-points of 3, 6 and 9-months of barrel ageing from both barrels, and at the 12-month time-point from BAN only as BAO was accidentally

used to refill other barrels due to common practices in the cellar. On the other hand, the grape variety at ICVV winery was Tempranillo, which is the main variety in DOCa Rioja, and the wine samples were collected at the end of malolactic fermentation, completed inside the steel tank and denoted as FML or 0 time-point, and after 3, 9 and 12-months of barrel ageing from both barrels. Additionally, the same day that the wine finished the FML at ICVV winery and transferred into BAO and BAN, a sample from each barrel was taken and bottled. These bottle-aged wine samples, from the old (BTO) and new (BTN) barrel, were stored in the same cellar as the barrels and analyzed after 12-months of bottle ageing. All collected wine samples were immediately frozen and preserved at -80°C prior to analysis.

1H-NMR

^{1}H NMR spectra were recorded at 300K on an Avance III 600 spectrometer (Bruker®, Germany) operating at a proton frequency of 600.20MHz using a 5mm PBBO gradient probe. Wine aqueous samples were measured and recorded in procno 11 using a One-dimensional ^{1}H pulse experiments were carried out using the nuclear Overhauser effect spectroscopy (NOESY) pre-saturation sequence (RD–90°–t1–90°–tm–90° ACQ) to suppress the residual water peak, and the mixing time was set at 100ms. Solvent presaturation with irradiation power of 75Hz was applied during relaxation delay (RD = 5s) and mixing time. (noesypr1d pulse program in Bruker®) to eliminate the residual water moisture of deuterated water. The acquisition time (ACQ) was 3.42s for a total recycling delay (RD+ACQ) of 8.42s. The 90° pulse length was calibrated for each sample and varied from 10.12 to 11.68ms. The spectral width was 10kHz (20ppm), and a total of 256 transients were collected into 64k data points for each ^{1}H spectrum. The exponential line broadening applied before Fourier transformation was of 0.5Hz. The frequency domain spectra were manually phased and baseline-corrected using TopSpin software (version 3.2, Bruker).

After pre-processing and visually assessing the NMR dataset, specific ^{1}H regions of compounds were identified in the spectra using a comparison into AMIX 3.9 software and Chenomx 7.2 software. Curated identified regions across the spectra were integrated using the same AMIX 3.9 software package and exported to excel spreadsheet in order to give relative concentrations.

GC-MS

GC-MS analysis were performed on a GCxGC-TOF Pegasus 4D from Leco Instruments equipped with a MPS autosampler from Guerstel. Chromatographic column was

a CP-Sil 24 CB (30m x 0.25mm i.d., 0.25µm film) from Agilent Technologies. The injection volume was 1µl and it was performed in pulsed splitless mode in a split/splitless injector at 250°C. He (99.999%) was used as mobile phase at a constant flow of 1.2 mL/min. For the elution of compounds, the following temperature program was used: 50°C for 2min, 50-150°C at 5°C/min, 150-240°C at 10°C/min. The transfer line temperature was 250°C and ionization was made by electron impact at 70eV with a source temperature of 250°C. The MS acquisition was in full scan after a solvent delay of 5min between 35-600m/z at 20scan/seg.

Data analysis for both target and untargeted experiments were performed in Chromatof 4.50.8 software from LECO. For untargeted analysis, the chromatograms were deconvoluted by fixing a baseline offset of 1, a peak width of 1 and signal/noise ratio of 100. For targeted analysis, the method was validated by evaluating the Limits of detection (LOD), limits of quantification (LOQ), linearity (R2), recovery, accuracy and repeatability using standard solutions and standard additions to a representative pool of samples. Quantification of target compounds was performed by an internal standard calibration method.

LC-MS

LC-MS analysis were performed on a UHPLC-qTOF 6550 from Agilent Technologies. Chromatographic column was an Aquity BEH-C18 (100mm x 2.1mm., 1.7µm) from Waters. For the elution of compounds, two different chromatographic methods were used with a 5mM ammonium formate (pH=3.8) for positive ionization mode and 5mM ammonium acetate (pH=4.5) for negative ionization mode as aqueous mobile phase and pure methanol as organic mobile phase component in both methods. The elution gradient was the same for both mobile phases, consisting on (0%-0%, 1 min; 0-65% 7 min; 65-100% 8 min; and 100%-100%, 11min). The injection volume was 1µl, the flow rate was 0.6 mL/min and column temperature was 40°C. The ionization was performed both in positive and negative electrospray in two separate runs and mass spectra was recorded between 100-1100m/z at 3spec/seg.

Data analysis was performed with Mass Profinder Software from Agilent. This software deconvolutes the chromatograms to find the molecular features present in the samples and align their mass and retention times resulting in a matrix containing the neutral mass of the feature their retention time and the area of the chromatographic peak.

Chemical classes

The assignment of chemical classes to the identified metabolites has been based on the food database (fooddb.ca), the yeast metabolome database [255] and the human metabolome database [256].

Statical analysis

Statistical analysis has been based on the factors barrel-type and time. For FB, the factor barrel-type included the 0, 3, 6 and 9-month barrel-aged wine samples separated in the groups of old and new barrel resulting in 4-samples per group, whereas the factor time concerned the barrel-aged wine from old and new barrel grouped by the attributes 0, 3, 6 and 9-month time-points leading to 2-samples per group. Similarly, for ICVV the factor barrel-type concerned the 3, 9 and 12-month barrel-aged wine samples divided into the groups of old and new barrel, and the factor time comprised the 4-groups of 3, 9, 12-month barrel-aged and 12-month bottle-aged wine samples. Moreover, ICVV included the additional factor bottled-wine which included the 12-month barrel-aged and 12-month bottle-aged wine samples.

For the analytical methods NMR and targeted GC-MS, statistical significance for each metabolite was derived from Student's T-Test based on the factor barrel-type and ANOVA on the factor time using the Python module STATSMODELS [205]. The resulted p-values were FDR-corrected (q-values) and the statistical significance (q-value $\leq$0.05) was coupled with practical significance which is defined as a minimum 2-fold change between minimum and maximum value observed among the samples (FCMM $\geq$2).

For the methods LC-MS and GC-MS, differential metabolomic analysis has been performed using the R package MetaboDiff [257]. The analytical steps followed by MetaboDiff included the imputation of missing values using k-nearest neighbor imputation, k-means clustering outlier detection, variance stabilizing normalization and differential analysis based on Student's T-Test or ANOVA using the factors barrel-type and time respectively. The metabolites that were chosen for further analysis were those that presented statistical significance with FDR-corrected p-value $\leq$0.05 and practical significance with FCMM $\geq$2.

Principal coordinate analysis (PCoA) for the methods NMR, LC-MS and GC-MS was based on the euclidean distance and permutational multivariate analysis of variance (PERMANOVA) on the distance matrix was performed using the Python module SCIKIT-BIO [258] and the factors barrel-type and time. Finally, hierarchical clustering

was performed using the Python module SCIPY [206] after calculating the growth rates of the barrel-aged wine samples for the periods 0-3, 3-6, 6-9 and 9-12-month time-points for FB, and 0-3, 3-9 and 9-12-month time-points for ICVV. For comparison between ICVV's 12th-month bottle and barrel-aged wine samples, the hierarchical clustering was based on their growth rates for the period 0-12-month.

Results

NMR

The 39 identified metabolites with NMR are reported in Table 18 for FB and ICVV. For both cellars the concentration for the majority of the metabolites was very low ranging between 0-90 mmols with exception ethanol that ranged between 1000-1600 mmols. After performing differential analysis, none of the metabolites of FB was found to be statistically significant for the factors barrel-type and time, whereas for the cellar ICVV the metabolite formate was found to be statistically significant for the factor barrel-type and methanol for the factor time. However, methanol had FCMM <2, leaving formate as the only metabolite with both statical and practical significance (FCMM=5.6).

After performing PCoA for the cellar FB (supplementary Figure S27A), the metabolites with the highest loadings across the principal components were saccharopine, 2,3 butanediol, tartaric acid and histidine, however without demonstrating practical significance. The only metabolites with practical significance were formate (FCMM=2, supplementary Figure S28A) and ethanal (FCMM=3, supplementary Figure S28B), with the former having an impact on separating the early maturation samples ($\leq$3 months) from those of late maturation ($\geq$6 months), and the latter showing differences between old and new barrel for the samples of 3, 6 and 9 months. Nevertheless, their effect on the samples clustering was minimal since PERMANOVA reported non-significant differences for the factors barrel-type (p-value=0.46) and time (p-value=0.09), and overall the hierarchical clustering based on the growth rates of the NMR metabolites did not reveal any informative clustering structure (Figure 26A, supplementary Figure S37).

Similarly to FB, most of the metabolites of the ICVV's samples with the highest loadings across the principal components (supplementary Figure S27B) were practically non-significant apart sorbate (FCMM=2 , supplementary Figure S28D) that showed differences between early ($\leq$3 months) and late ($\geq$9 months) maturation samples . The rest of the metabolites with practical significance were acetoin (FCMM=2, supplementary Figure S28C) with similar PCoA loadings as gluconate that showed differences between early ($\leq$3 months) and late ($\geq$9 months) maturation samples, and formate (FCMM=5.6,

supplementary Figure S28E) and ethanal (FCMM=3.5, supplementary Figure S28H) where both showed differences between old and new barrel and had similar PCoA loadings as sorbate. Practical significance was also shown by uracil (FCMM=2.4, supplementary Figure S28F) and shikimic acid (FCMM=2.0, supplementary Figure S28G) which had similar PCoA loadings as choline, however their capacity on demonstrating differences between the factors was minimal and mainly concerned the discrepancy between the 9th month samples with the rest. After performing PERMANOVA, statistical significance was found only for the factor time (p-value=0.02) which upon the calculation of the growth factors was attributed to the higher growth rates of formate and ethanal for the period of 0-3 months of the new barrel and the fact that the majority of the metabolites in both barrels had negative growth rates for the period 9-12 month (Figure 26C, supplementary Figure S39).

Overall, the metabolites did not present differences between the 12th month samples of bottle and barrel-aged wine with exception the case of formate that demonstrated practical significance between the 12th month BTN and BAN samples, an exception that also influenced the hierarchical clustering of the samples due to its high growth rate for the BAN sample (Figure 26B, supplementary Figure S38).

targeted GC-MS

The results for targeted GC-MS are given in Table 18 for FB and ICVV. For FB the metabolites 4-ethylphenol (4-EP) and 4-ethylguaiacol (4-EG) demonstrated similar trends between old and new barrel, with the former having almost identical values between the barrel-types across the time-points and the latter showing a converging tendency of the barrel-types after the 3rd month (supplementary Figure S29A). After performing differential analysis, the metabolite 5-methylfurfural (5-MF) showed statistical significance for the factor barrel-type (q-value=0.009) and the metabolite 4-EG for the factor time (q-value=0.002). However, only 5-MF was considered of presenting practical significance having a median fold-change of 12.4 between the barrel-types.

Regarding ICVV, the metabolite 5-MF displayed practical significance of 3.9 fold-change difference between the 12th month barrel and bottle-aged wine from new barrel without being accompanied by statistical significance for the factor bottled-wine (supplementary Figure S29B). However, it presented statistical significance for the factor barrel-type (p-value=0.033) and practical significance of median fold-change of 96.4 between the barrel-types (supplementary Figure S29C). The rest of the metabolites were statistically non-significant and ranged in low concentrations ($<23\mu$g/L).

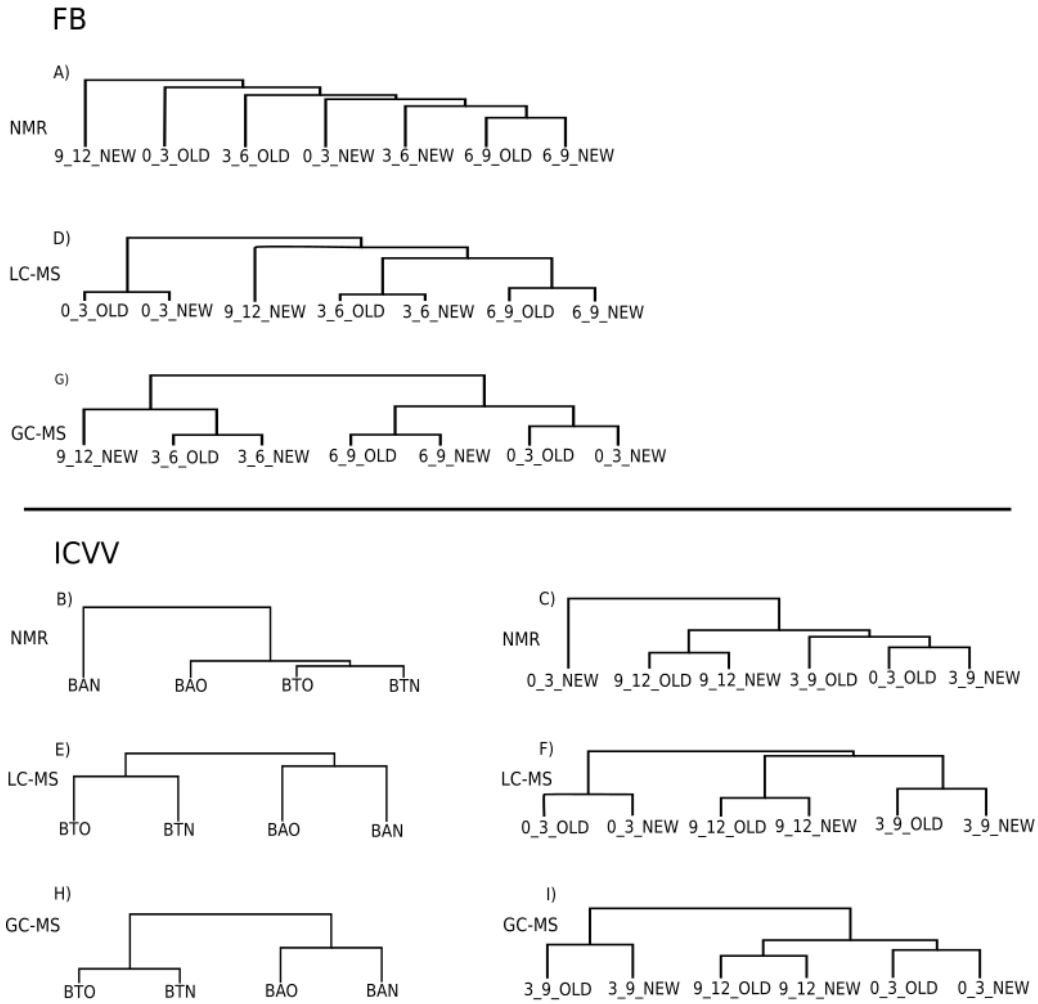


Figure 26: Growth rates hierarchical clustering of samples for region FB (A,D,G) and ICVV (B,C,E,F,H,I) based on different analytical methods. Acronyms BAN and BAO refer to 12th month barrel-aged wine from new and old barrel respectively and BTN and BTO to bottle-aged wine from new and old barrel respectively. Samples containing the labels OLD and NEW refer to barrel-aged wine from old and new barrel respectively, whereas numbers at the beginning of the labels represent growth rate periods.

		FB									ICVV								
	Compound	BAO_0	BAO_3	BAO_6	BAO_9	BAN_0	BAN_3	BAN_6	BAN_9	BAN_12	FML_0	BAO_3	BAO_9	BAO_12	BAN_3	BAN_9	BAN_12	BTO_12	BTN_12
	ethanol	1153.660	1403.470	1274.930	1446.270	1203.490	1413.400	1470.880	1457.710	1137.700	1034.400	1357.820	1486.880	1184.620	1352.390	1459.150	1067.110	1212.080	1186.600
	glycerol	67.800	82.850	77.200	84.140	71.460	82.480	85.420	85.260	65.650	68.310	78.660	89.960	73.810	86.260	88.320	67.710	75.575	73.030
	1-propanol	0.470	0.590	0.540	0.560	0.470	0.570	0.640	0.490	0.540	0.470	0.700	0.600	0.530	0.680	0.740	0.510	0.595	0.600
	isobutanol+isopentanol	1.850	2.230	2.060	2.250	1.910	2.240	2.300	2.290	1.790	1.360	1.740	1.910	1.500	1.750	1.840	1.340	1.555	1.500
	2,3 butanediol	4.890	5.970	5.610	6.100	5.260	5.960	6.210	6.180	4.750	4.490	5.440	6.150	4.900	5.960	6.070	4.500	5.150	5.020
	methanol	2.350	2.790	2.660	2.880	2.430	2.820	3.000	3.030	2.510	3.340	4.290	4.350	3.610	4.220	4.410	3.670	3.775	3.720
	lactic acid	4.050	4.920	4.590	5.070	4.260	4.870	5.080	5.100	3.840	12.280	13.660	15.890	13.200	15.320	15.410	12.280	13.405	13.080
	acetic acid	4.900	6.120	5.890	6.680	5.350	6.330	6.870	7.040	5.550	5.740	7.110	8.160	6.820	7.840	8.830	6.970	6.485	6.390
	succinic acid	4.950	5.880	5.460	5.910	5.240	5.870	6.070	6.020	4.580	4.620	5.100	5.880	4.800	5.710	5.780	4.450	4.890	4.800
	acetoin	0.110	0.170	0.160	0.130	0.130	0.160	0.140	0.120	0.100	0.110	0.150	0.220	0.210	0.160	0.210	0.170	0.180	0.160
	2-hydroxybutiric acid	1.340	1.580	1.550	1.640	1.410	1.600	1.640	1.680	1.300	1.000	1.230	1.340	1.060	1.190	1.240	0.930	1.090	1.020
	proline	1.320	1.520	1.450	1.520	1.340	1.510	1.550	1.550	1.260	4.950	5.260	5.860	5.200	5.830	5.860	4.810	5.290	5.040
	butanone	0.040	0.040	0.040	0.040	0.030	0.040	0.050	0.050	0.040	0.060	0.070	0.070	0.070	0.080	0.090	0.070	0.070	0.060
	tartaric acid	1.500	1.290	1.250	1.350	1.290	1.260	1.040	1.110	1.420	2.100	3.010	2.290	3.030	2.600	2.330	3.110	2.250	2.250
	galacturonic acid	0.340	0.520	0.570	0.630	0.400	0.530	0.590	0.640	0.570	1.790	2.140	2.210	1.890	2.240	2.170	1.810	1.925	1.750
	levulinic acid	0.380	0.470	0.420	0.460	0.400	0.460	0.470	0.470	0.340	0.230	0.260	0.300	0.240	0.300	0.300	0.220	0.250	0.240
	malic acid	1.020	1.200	1.180	1.250	1.060	1.210	1.210	1.200	1.040	1.100	1.340	1.480	1.360	1.370	1.440	1.230	1.385	1.290
	lysine	0.140	0.130	0.140	0.140	0.120	0.140	0.140	0.150	0.140	0.500	0.480	0.480	0.490	0.500	0.500	0.440	0.505	0.470
	sorbate	0.050	0.050	0.050	0.040	0.060	0.060	0.050	0.050	0.040	0.080	0.060	0.050	0.040	0.080	0.070	0.040	0.040	0.040
	2-hydroxy-3-methylvalerate	0.170	0.210	0.200	0.220	0.190	0.210	0.220	0.220	0.160	0.190	0.230	0.250	0.210	0.250	0.260	0.200	0.215	0.200
	ornithine	0.190	0.200	0.180	0.190	0.190	0.190	0.200	0.200	0.160	0.570	0.540	0.590	0.550	0.600	0.560	0.510	0.560	0.540
	choline	0.120	0.140	0.130	0.150	0.120	0.140	0.150	0.150	0.120	0.170	0.190	0.220	0.180	0.210	0.220	0.170	0.190	0.180
	o-phosphocholine	0.030	0.040	0.040	0.030	0.030	0.040	0.030	0.030	0.030	0.070	0.070	0.090	0.080	0.080	0.080	0.070	0.075	0.070
	trigonelline	0.120	0.120	0.120	0.130	0.140	0.110	0.150	0.140	0.120	0.140	0.130	0.150	0.160	0.140	0.140	0.130	0.135	0.120
	histidine	0.050	0.040	0.060	0.060	0.050	0.050	0.050	0.060	0.060	0.110	0.090	0.100	0.120	0.110	0.110	0.100	0.105	0.100
	myo-inositol	1.650	1.880	1.810	1.920	1.670	1.870	1.940	1.920	1.730	1.410	1.400	1.620	1.490	1.560	1.710	1.450	1.495	1.400
	tyrosine	0.150	0.170	0.160	0.180	0.150	0.160	0.180	0.180	0.140	0.140	0.140	0.180	0.150	0.160	0.150	0.120	0.145	0.130
	gallic acid	0.110	0.140	0.140	0.160	0.120	0.150	0.170	0.180	0.130	0.180	0.220	0.300	0.250	0.270	0.310	0.230	0.235	0.210
	shikimic acid	0.070	0.080	0.090	0.090	0.100	0.090	0.120	0.100	0.070	0.090	0.080	0.110	0.080	0.130	0.160	0.100	0.095	0.080
	caffeic acid	0.170	0.200	0.190	0.200	0.180	0.190	0.200	0.190	0.130	0.260	0.290	0.300	0.250	0.310	0.270	0.210	0.255	0.230
	formate	0.050	0.050	0.080	0.060	0.050	0.070	0.100	0.090	0.090	0.050	0.070	0.070	0.070	0.210	0.280	0.230	0.060	0.060
	phenylalanine	0.060	0.090	0.090	0.100	0.070	0.080	0.080	0.090	0.090	0.090	0.090	0.110	0.090	0.090	0.100	0.090	0.100	0.100
	3-phenylpropionate	0.520	0.600	0.550	0.580	0.520	0.590	0.590	0.600	0.450	0.360	0.390	0.450	0.350	0.440	0.450	0.320	0.375	0.340
	uracil	0.050	0.060	0.070	0.070	0.050	0.040	0.060	0.050	0.050	0.080	0.060	0.120	0.070	0.080	0.090	0.090	0.075	0.050
	guanosine	0.050	0.060	0.070	0.050	0.060	0.070	0.060	0.070	0.050	0.080	0.070	0.100	0.060	0.070	0.070	0.070	0.075	0.070
	gluconate	1.570	1.770	1.750	1.830	1.670	1.780	1.870	1.870	1.620	1.370	1.380	1.490	1.540	1.360	1.440	1.440	1.530	1.420
	alanine	0.260	0.240	0.310	0.310	0.210	0.280	0.330	0.370	0.290	0.340	0.360	0.480	0.430	0.400	0.440	0.380	0.380	0.350
	saccharopine	0.340	0.370	0.320	0.340	0.370	0.350	0.350	0.360	0.280	0.740	0.650	0.670	0.590	0.730	0.680	0.550	0.605	0.570
	ethanal	0.020	0.010	0.020	0.020	0.020	0.030	0.030	0.030	0.020	0.030	0.020	0.030	0.020	0.070	0.070	0.030	0.020	0.020
B	4-ethylphenol	28	38	31	98	2	2	17	66	46	5	11	11	11	2	2	23	3.5	3
B	4-ethylguaiacol	0	1	2	6	0	1	3	7	4	0	0	0	0	0	2	0	0	0
B	5-methylfurfural	0	7	9	6	36	106	63	58	55	0	7	5	5	675	1134	358	0	91

Table 18: NMR (A) and targeted GC-MS (B) metabolites for cellars FB and ICVV. Numbers at the end of the acronyms represent sampling month. NMR values correspond to concentrations which are expressed in arbitrary relative units, but in a perfect experimental could be taken as absolute molar concentration into the wine in mmols in comparison with the ERETIC signal of 0.6616 mM inserted in each spectra. Targeted GC-MS values correspond to concentrations expressed in μg/L.

LC-MS

From the 502 metabolites initially identified with LC-MS, only 14 metabolites were found to be statistically and practically significant for FB (Table 19) with average FCMM of 6.3 and 17 for ICVV (Table 20) with average FCMM of 6.

After performing PERMANOVA on the distance matrix for the cellar FB, statistically significant differences were found only for the factor time (p-value=0.01). The impact of the metabolites on the samples clustering (supplementary Figure S30A) could be distinguished in two groups with the metabolites of each group having similar loadings across the first principal component. The first group contained the metabolites whose concentrations had an increasing trend and had similar PCoA loadings as eriodictyol and the second group those metabolites with decreasing trend and PCoA loadings similar to isorhamnetin-3-o-glucoside (supplementary Figure S31A). From the comparison of the groups becomes apparent that the calculated FCMM has been derived from the differences between the samples of 0 and 9th or 12th month. This is also being depicted in the hierarchical clustering based on the growth factors where in both barrels the first trimester appears to be the most distant to the rest mainly due to the high growth factors of caffeic acid, jaceosidin, naringin and luteolin 7-glucoside during that period (Figure 26D, supplementary Figure S40).

In the same manner, the LC-MS metabolites for the ICVV samples were divided in those with increasing and those with decreasing tendency (supplementary Figure S32), with the metabolites of each group having similar PCoA loadings across the first principal component (supplementary Figure S30B). Once more, PERMANOVA reported statical significance only for the factor time (p-value=0.04), and in combination with the hierarchical clustering based on the growth factors it seems that these differences concern the early ($\leq$3 month) and late ($\geq$9 month) maturation periods (Figure 26F, supplementary Figure S42). Regarding the bottle and barrel-aged 12th month wine samples, although they demonstrated practically non-significant differences with exception the metabolite scopoletin with FCMM of 2 between BAO and BTO and 2.5 between BAN and BTN (supplementary Figure S32), minor differences in their growth factors created a cumulative effect able to differentiate the two sample groups (Figure 26E, supplementary Figure S41). However, these differences could only be observed across the second principal component of the LC-MS PCoA (supplementary Figure S30B), that accounted only for 9.5% of the overall observed variation in ICVV's samples, suggesting weak differences.

untargeted GC-MS

From the 394 metabolites initially detected with GC-MS, 16 metabolites were found of having statistical and practical significance for FB with mean FCMM of 8.2 (Table 21) and 48 for ICVV with mean FCMM of 16 (Table 22).

As with the previous methods, PERMANOVA on the distance matrix showed statistical differences only for the factor time in both cellars (p-value=0.006 for FB and p-value=0.01 for ICVV). For FB three groups of metabolites could be observed influencing the sample clustering (supplementary Figure S33A). The first group included the metabolites with increasing trend having similar PCoA loadings to diethyl succinate (supplementary Figure S33A), the second group metabolites with decreasing trend and PCoA loadings similar to indole-3-methyl acetate and the third group metabolites whose calculated FCMM derived from the differences between the 6th month sample against the rest and had similar PCoA loadings to 4,6,8-trimethylon-1-ene (supplementary Figure S34). The latter group appears to be the reason for the clustering of the periods 9-12 and 3-6 months after performing hierarchical clustering based on the growth rates (Figure 26G, supplementary Figure S43).

For ICVV, two main groups of metabolites seemed to be influencing the sample clustering (supplementary Figure S33B). As with LC-MS, the first group demonstrated an increasing trend and had similar PCoA loadings to 2-methyltetrahydrothiophen-3-one (supplementary Figure S33B) and the second group included metabolites with decreasing tendency and similar PCoA loadings to dibutyl phthalate (supplementary Figures S35, S36). These two groups could differentiate the early ($\leq$3) from the late ($\geq$9) maturation wine samples, however due to the occasional non-linearity of the trends caused mainly by the 9th month samples in cases such as 6-tridecene (supplementary Figure S35), palmic acid and stearic acid (supplementary Figure S36) the periods 0-3 and 9-12 months clustered together during the hierarchical clustering based on growth rates (Figure 26I, supplementary Figure S45).

With regards the 12th month bottle and barrel-aged wine, a number of metabolites displayed practical significance with a mean FCMM of 7.3 between BAO and BTO and mean FCMM of 6.9 between BAN and BTN that could be separated into two groups. The first group included the metabolites 4,6,8-trimethylnon-1-ene, palmitic acid, stearic acid, 1-dodecanol, and 6-tridecene whose values were higher for the bottle-aged wine samples (supplementary Figures S35, S36) and had similar PCoA loadings to 4,6,8-trimethylnon-1-ene (supplementary Figure S33B). The second group included the metabolites methyl 2-methoxy-2-phenylacetate, oxoglutaric acid, 2,10-dimethylundecane, 5-ethoxyoxolan-2-one, 1-tetradecene and ethyl-3-hydroxybutyrate that displayed higher values for the

barrel-aged wine samples (supplementary Figures S35, S36) and received similar PCoA loadings to phenol, 2-(2h-benzotriazol-2-yl)-4,6-bis(1,1-dimethylpropyl) (supplementary Figure S33B). These two groups appear to mainly have influenced the separation of the 12th month bottle and barrel-aged wine samples after performing growth rates hierarchical clustering (Figure 26H, supplementary Figure S44). Additionally, the separation of these two groups of the factor bottled-wine occurs towards the second principal component of GC-MS PCoA (supplementary Figure S33B) that accounts for 26.2% of the total observed variation suggesting better separation capacity of GC-MS than LC-MS for this factor.

Compound	BAO_0	BAO_3	BAO_6	BAO_9	BAN_0	BAN_3	BAN_6	BAN_9	BAN_12
4-vinylphenol^{1U}	2.401E+05	4.445E+05	5.751E+05	6.785E+05	1.976E+05	3.537E+05	5.062E+05	5.740E+05	5.335E+05
3-hydroxycinnamic acid^{2U}	4.681E+05	7.217E+05	8.987E+05	1.038E+06	4.253E+05	6.029E+05	8.160E+05	8.886E+05	8.553E+05
caffeic acid^{2U}	9.369E+04	5.058E+05	6.973E+05	8.137E+05	9.421E+04	4.157E+05	6.112E+05	6.945E+05	6.772E+05
(+/-)-eriodictyol^{3U}	5.958E+04	7.900E+04	1.420E+05	2.327E+05	4.878E+04	8.147E+04	1.447E+05	2.421E+05	3.093E+05
6-methoxyluteolin^{3D}	2.990E+05	1.479E+05	1.337E+05	4.167E+04	2.791E+05	1.308E+05	1.268E+05	5.896E+04	1.723E+04
dihydroquercetin^{3D}	3.194E+05	4.837E+05	4.758E+05	1.236E+05	3.049E+05	4.278E+05	4.495E+05	9.960E+04	9.015E+04
hyperoside^{3D}	5.994E+06	2.578E+06	1.620E+06	9.887E+05	5.775E+06	2.424E+06	1.513E+06	9.386E+05	7.660E+05
isorhamnetin-3-o-glucoside^{3D}	8.401E+05	4.768E+05	3.681E+05	2.859E+05	8.273E+05	5.055E+05	3.612E+05	2.567E+05	2.129E+05
jaceosidin^{3U}	4.673E+04	1.297E+05	1.901E+05	3.032E+05	3.983E+04	1.118E+05	1.844E+05	2.855E+05	1.641E+05
luteolin 7-glucoside^{3D}	3.374E+05	1.188E+05	8.155E+04	5.768E+04	3.229E+05	1.187E+05	7.258E+04	4.625E+04	5.087E+04
myricetin 3-galactoside^{3D}	1.127E+07	7.719E+06	6.981E+06	5.992E+06	1.085E+07	7.697E+06	6.774E+06	5.744E+06	4.433E+06
naringin^{3D}	1.818E+06	5.272E+05	2.757E+05	1.794E+05	1.791E+06	5.219E+05	2.692E+05	2.132E+05	1.973E+05
procyanidin B3^{3D}	2.860E+05	2.871E+05	2.628E+05	2.482E+05	2.881E+05	2.745E+05	2.582E+05	2.327E+05	7.758E+04
4-glucogallic acid^{4D}	2.246E+05	2.041E+05	1.749E+05	1.634E+05	2.345E+05	2.105E+05	1.710E+05	1.390E+05	1.106E+05

Table 19: LC-MS metabolites, with statistical and practical significance, for cellar FB. Numbers at the end of the acronyms represent sampling month. Values represend the area of the detected chromatographic peaks after applying variance stabilization. The metabolites can be grouped in the following chemical classes (1) benzene and substituted derivatives, (2) cinnamic acids and derivatives, (3) flavonoids and (4) organooxygen compounds. Metabolites that exhibited increasing tendency are indicated as (U) and those with decreasing tendency as (D).

Compound	FML_0	BAN_3	BAN_9	BAN_12	BAO_3	BAO_9	BAO_12	BTN_12	BTO_12
3,4-dihydroxybenzaldehyde^{1U}	1.420E+05	3.854E+05	6.214E+05	6.202E+05	2.213E+05	3.810E+05	3.685E+05	4.502E+05	4.223E+05
benzoic acid^{1U}	1.342E+05	3.549E+05	4.017E+05	4.196E+05	2.935E+05	3.050E+05	3.083E+05	2.725E+05	2.461E+05
ethyl gallate^{1U}	9.749E+04	1.937E+05	4.171E+05	4.528E+05	2.114E+05	3.925E+05	4.431E+05	4.526E+05	4.347E+05
gallic acid^{1D}	5.902E+05	4.846E+05	1.026E+05	1.899E+05	4.911E+05	1.911E+05	1.897E+05	3.634E+05	3.377E+05
vanillic acid^{1D}	7.071E+04	3.281E+04	2.825E+04	3.191E+04	3.943E+04	2.812E+04	3.096E+04	3.255E+04	3.325E+04
caffeic acid^{2U}	4.492E+04	3.065E+05	4.644E+05	4.421E+05	2.749E+05	5.147E+05	5.360E+05	5.399E+05	4.651E+05
scopoletin^{3U}	7.789E+04	2.285E+05	4.743E+05	4.419E+05	1.819E+05	3.303E+05	3.212E+05	1.757E+05	1.580E+05
6-methoxyluteolin^{4U}	1.014E+05	5.635E+04	3.068E+05	2.069E+05	6.785E+04	1.450E+05	1.490E+05	9.607E+04	1.909E+05
astilbin^{4D}	2.386E+05	2.096E+05	2.401E+04	1.655E+04	2.202E+05	2.373E+04	1.246E+04	1.279E+04	1.003E+04
dihydromyricetin 3-rhamnoside^{4D}	5.119E+04	3.240E+04	2.719E+04	2.328E+04	2.807E+04	2.399E+04	1.772E+04	2.485E+04	2.639E+04
dihydroquercetin^{4D}	5.209E+05	2.730E+05	2.378E+05	1.754E+05	5.209E+05	2.061E+05	2.201E+05	1.975E+05	2.701E+05
hyperoside^{4D}	6.571E+06	4.652E+06	3.087E+06	2.211E+06	4.426E+06	2.807E+06	2.269E+06	2.097E+06	2.242E+06
luteolin 7-glucoside^{4D}	9.870E+05	4.829E+05	2.121E+05	1.412E+05	4.527E+05	1.902E+05	1.418E+05	1.329E+05	1.365E+05
naringin^{4D}	3.984E+05	1.781E+05	1.400E+05	1.213E+05	1.877E+05	1.514E+05	1.337E+05	1.267E+05	1.183E+05
procyanidin C2^{4D}	2.273E+05	1.362E+05	4.477E+04	3.378E+04	1.692E+05	7.353E+04	5.864E+04	3.868E+04	4.560E+04
4-methoxyphenylacetic acid^{5U}	5.709E+04	7.191E+04	1.136E+05	1.390E+05	1.183E+05	1.784E+05	1.802E+05	1.590E+05	2.084E+05
syringaldehyde^{6U}	5.742E+05	9.709E+05	1.505E+06	1.574E+06	1.397E+06	2.269E+06	2.250E+06	1.780E+06	2.416E+06

Table 20: LC-MS metabolites, with statistical and practical significance, for cellar ICVV. Numbers at the end of the acronyms represent sampling month. Values represent the area of the detected chromatographic peaks after applying variance stabilization.The metabolites can be grouped in the following chemical classes (1) benzene and substituted derivatives, (2) cinnamic acids and derivatives, (3) coumarins and derivatives, (4) flavonoids, (5) phenol ethers and (6) phenols. Metabolites that exhibited increasing tendency are indicated as (U) and those with decreasing tendency as (D).

Compound	BAN_0	BAN_3	BAN_6	BAN_9	BAN_12	BAO_0	BAO_3	BAO_6	BAO_9
n-phenethylacetamide^{1S}	2.852E+05	1.940E+05	5.383E+05	8.346E+04	2.829E+05	1.512E+05	1.338E+05	5.773E+05	8.381E+04
2,4-di-tert-butylphenol^{2S}	3.717E+06	2.697E+06	6.253E+06	2.396E+06	1.413E+06	3.408E+06	3.440E+06	6.976E+06	2.546E+06
2-phenethyl acetate^{2D}	8.860E+06	7.812E+06	3.742E+06	5.750E+06	2.981E+06	9.366E+06	8.369E+06	4.480E+06	5.083E+06
3,5-bis(1,1-dimethylethyl)-4-hydroxy-benzoic acid ethyl ester^{2S}	1.619E+05	8.739E+04	5.094E+05	5.437E+04	2.829E+05	8.547E+04	1.313E+05	4.922E+05	6.937E+04
ethyl 5-oxo-dl-prolinate^{3U}	4.017E+06	9.371E+06	6.607E+06	1.640E+07	8.974E+06	4.770E+06	8.509E+06	7.601E+06	1.421E+07
2-butyl-1-octanol^{4S}	7.991E+05	3.638E+05	1.807E+06	1.243E+05	4.386E+06	6.227E+05	5.332E+05	2.046E+06	2.829E+05
diethyl succinate^{4U}	3.195E+07	1.130E+08	9.060E+07	2.611E+08	2.169E+08	4.494E+07	1.359E+08	1.238E+08	2.448E+08
diethyl malate^{5U}	1.840E+06	4.602E+06	3.518E+06	9.499E+06	5.617E+06	2.215E+06	4.972E+06	4.129E+06	8.456E+06
indole-3-methyl acetate^{6D}	7.433E+05	6.886E+05	3.444E+05	5.188E+05	2.810E+05	9.426E+05	7.030E+05	4.249E+05	4.657E+05
oxoglutaric acid^{7U}	3.762E+06	6.090E+06	5.393E+06	1.436E+07	1.182E+07	5.554E+06	5.322E+06	6.230E+06	1.155E+07
gamma-butyrolactone^{8U}	3.682E+07	7.239E+07	5.101E+07	1.379E+08	1.023E+08	3.503E+07	7.649E+07	6.008E+07	1.175E+08
2,3-dimethylphenol^{9U}	2.725E+07	2.241E+07	1.080E+07	2.427E+07	2.727E+07	2.926E+07	2.439E+07	1.298E+07	2.114E+07
acetaminophen^{9D}	7.490E+05	1.088E+06	7.082E+05	1.493E+06	9.043E+04	7.203E+05	8.893E+05	6.914E+05	1.142E+06
2-methylnonane^{10S}	3.791E+06	1.683E+06	6.895E+06	1.190E+06	4.384E+06	1.918E+06	1.510E+06	7.666E+06	1.461E+06
3-eicosyne^{11S}	7.315E+05	3.703E+05	1.902E+06	2.730E+05	6.101E+05	3.795E+05	5.738E+05	2.504E+06	2.907E+05
4,6,8-trimethylnon-1-ene^{11S}	2.062E+06	9.257E+05	4.637E+06	5.638E+05	3.014E+06	1.531E+06	1.509E+06	5.021E+06	6.459E+05

Table 21: GC-MS metabolites, with statistical and practical significance, for cellar FB. Numbers at the end of the acronyms represent sampling month. Values represend the area of the detected chromatographic peaks after applying variance stabilization. The metabolites can be grouped in the following chemical classes (1) amines, (2) benzene and substituted derivatives, (3) carboxylic acids, (4) fatty acyls, (5) hydroxy acids, (6) indoles, (7) keto acids, (8) lactones, (9) phenols, (10) saturated hydrocarbons, (11) unsaturated hydrocarbons. Metabolites that exhibited increasing tendency are indicated as (U) and those with decreasing tendency as (D), whereas metabolites with relatively stable concentrations but with a peak on the 6th month are denoted as (S).

Compound	FML_0	BAO_3	BAO_9	BAO_12	BAN_3	BAN_9	BAN_12	BTO_12	BTN_12
2,4-di-tert-butylphenol^{1D}	2.414E+06	3.082E+06	4.827E+05	1.640E+06	1.498E+06	4.557E+05	9.088E+05	5.615E+05	5.328E+05
2-hydroxy-3,4-dimethoxybenzoic acid^{1U}	7.967E+05	1.243E+06	1.724E+06	7.664E+05	2.255E+06	2.486E+06	1.443E+06	5.913E+05	8.285E+05
2-phenethyl acetate^{1D}	1.771E+06	1.065E+06	5.997E+05	3.329E+05	1.278E+06	5.619E+05	4.411E+05	3.625E+05	4.176E+05
4-nonylphenol^{1U}	4.401E+05	4.296E+05	9.836E+05	5.848E+05	4.518E+05	9.891E+05	8.520E+05	5.391E+05	5.566E+05
benzyl alcohol^{1U}	2.647E+05	1.491E+05	1.769E+06	9.843E+05	4.066E+05	1.594E+06	1.436E+06	1.103E+06	1.136E+06
dibutyl phthalate^{1D}	1.442E+06	1.329E+06	2.231E+05	8.112E+05	1.255E+06	2.825E+05	4.175E+05	4.088E+05	5.382E+05
diphenylketone^{1D}	1.217E+06	7.776E+05	2.536E+05	4.332E+05	7.676E+05	3.267E+05	3.006E+05	5.087E+05	6.440E+05
ethyl-4-hydroxybenzoate^{1U}	1.790E+05	5.168E+05	7.788E+05	4.233E+05	2.670E+05	4.014E+05	3.703E+05	5.368E+05	3.978E+05
methyl 2-methoxy-2-phenylacetate^{1D}	2.290E+05	9.653E+04	9.087E+04	1.174E+05	8.450E+04	1.910E+05	1.801E+05	2.569E+04	2.869E+04
2-isopropyl-5-methyl-2-hexenal^{2D}	5.162E+05	5.378E+05	1.339E+05	3.172E+05	4.449E+05	1.433E+05	2.122E+05	3.733E+05	3.721E+05
ethyl 5-oxo-dl-prolinate^{3U}	2.260E+06	4.942E+06	7.916E+06	5.939E+06	4.556E+06	6.388E+06	6.910E+06	5.252E+06	5.053E+06
hexadecyl trichloroacetate^{3U}	2.866E+05	1.244E+05	2.553E+05	5.535E+05	1.884E+05	2.886E+05	2.735E+06	7.558E+05	7.893E+05
l,l-cyclo(leucylprolyl)^{3U}	1.687E+05	2.657E+05	4.116E+05	2.409E+05	2.463E+05	4.332E+05	3.298E+05	2.775E+05	2.951E+05
methionol^{3U}	3.071E+05	2.197E+06	4.677E+06	2.789E+06	7.484E+05	1.828E+06	1.583E+06	2.471E+06	3.843E+06
oxalic acid dibutyl ester^{3U}	2.980E+05	1.697E+05	5.164E+05	5.782E+05	2.975E+05	7.468E+05	1.436E+06	1.185E+06	1.756E+06
1-dodecanol^{4U}	4.293E+06	3.660E+05	4.844E+06	4.440E+06	3.252E+05	4.995E+06	4.349E+06	5.263E+06	4.578E+06
ethyl octanoate^{4D}	4.344E+06	3.960E+06	4.470E+06	4.260E+06	3.577E+06	6.498E+05	4.309E+06	4.587E+06	4.326E+06
ethyl undecanoate^{4D}	5.462E+05	4.444E+05	4.840E+05	3.484E+05	4.449E+05	4.625E+05	2.422E+05	3.346E+05	3.244E+05
palmitic acid^{4U}	3.554E+06	3.967E+05	9.465E+06	3.065E+06	4.054E+06	4.079E+07	4.994E+06	4.778E+07	4.855E+07
stearic acid^{4U}	9.744E+05	6.436E+04	5.274E+06	3.468E+06	6.305E+04	7.064E+06	2.380E+06	4.047E+07	4.088E+07
valeric acid^{4D}	4.754E+06	3.826E+06	4.355E+06	9.926E+05	4.388E+06	2.063E+06	2.466E+06	7.722E+05	6.348E+05
diethyl malate^{5U}	5.358E+05	5.345E+06	5.573E+06	5.553E+06	5.528E+06	5.093E+06	5.366E+06	7.755E+05	7.567E+05
ethyl-3-hydroxybutyrate^{5D}	5.874E+06	3.257E+06	5.053E+06	5.677E+05	5.884E+06	9.959E+05	5.509E+06	5.558E+05	5.902E+05
d-tryptophan^{6U}	9.486E+04	8.392E+04	2.575E+05	5.559E+05	5.207E+05	2.077E+05	2.044E+05	5.540E+05	5.468E+05
indole-3-methyl acetate^{6D}	3.065E+05	2.214E+05	1.207E+05	1.340E+05	2.160E+05	1.308E+05	5.713E+04	1.091E+05	6.159E+04
tryptophol^{6U}	2.296E+07	2.835E+07	1.040E+08	5.032E+07	4.100E+07	7.894E+07	8.590E+07	5.759E+07	5.743E+07

oxoglutaric acid^{7U}	2.866E+05	2.326E+05	2.713E+05	1.421E+06	1.884E+05	2.165E+05	2.735E+06	7.169E+05	5.352E+05
5-ethoxyoxolan-2-one^{8D}	8.963E+06	1.060E+05	2.523E+06	5.880E+05	3.386E+06	3.106E+06	1.077E+06	7.264E+04	6.515E+04
gamma-butyrolactone^{8U}	1.441E+07	2.206E+07	3.811E+07	2.603E+07	2.107E+07	3.036E+07	3.486E+07	2.257E+07	2.148E+07
1-tetradecene^{9D}	5.955E+05	9.176E+05	2.731E+05	3.713E+05	4.860E+05	3.491E+04	1.984E+05	8.382E+04	4.462E+04
succinic anhydride^{10D}	2.049E+07	1.910E+06	3.480E+06	3.435E+06	7.023E+06	2.101E+06	9.742E+05	1.755E+06	1.053E+06
2-methoxy-4-vinylphenol^{11D}	6.205E+05	5.566E+05	4.360E+05	2.259E+05	5.516E+05	2.764E+05	2.926E+05	2.833E+05	2.837E+05
homovanillic acid^{11U}	1.167E+05	2.533E+05	4.914E+05	2.101E+05	2.075E+05	4.610E+05	3.557E+05	1.880E+05	1.923E+05
tyrosol^{11U}	7.442E+06	9.684E+06	3.490E+07	2.626E+07	1.438E+07	3.244E+07	1.994E+07	2.360E+07	2.290E+07
methyl 3-(4-hydroxyphenyl)propanoate^{12U}	3.741E+05	7.971E+05	1.672E+06	1.004E+06	8.380E+05	1.592E+06	1.603E+06	9.248E+05	8.229E+05
phenol, 2-(2h-benzotriazol-2-yl)-4,6- -bis(1,1-dimethylpropyl)^{12D}	1.292E+06	9.899E+05	1.466E+06	7.854E+05	1.322E+06	1.547E+06	7.180E+05	3.657E+05	6.467E+05
borneol^{13D}	5.848E+05	4.724E+05	4.555E+05	2.466E+05	4.963E+05	5.482E+05	4.351E+05	2.741E+05	2.769E+05
carvacrol^{13D}	4.005E+05	3.721E+05	1.327E+05	1.945E+05	2.153E+05	1.449E+05	1.874E+05	1.714E+05	2.248E+05
2,10-dimethylundecane^{14U}	8.454E+04	1.643E+05	7.687E+05	3.971E+05	4.740E+04	4.754E+05	3.988E+05	5.803E+04	5.652E+04
2,7,10-trimethyldodecane^{14D}	1.768E+06	9.446E+05	2.138E+05	7.174E+05	1.397E+06	6.136E+05	7.810E+05	1.501E+06	1.200E+06
3,8-dimethylundecane^{14U}	2.217E+05	2.894E+05	1.107E+05	3.496E+05	2.130E+05	1.018E+05	1.812E+05	3.457E+05	3.181E+05
isoheptane^{14D}	2.466E+06	2.102E+06	8.155E+05	1.685E+06	1.862E+06	6.706E+05	1.141E+06	2.187E+06	1.870E+06
1,2-diphenylcyclobutane^{15D}	3.344E+05	3.456E+05	2.727E+05	1.778E+05	1.901E+05	3.178E+05	1.034E+05	1.233E+05	8.945E+04
7,9-di-tert-butyl-1-oxaspiro[4.5]deca- -6,9-diene-2,8-dione^{16D}	2.023E+06	1.914E+06	5.922E+05	7.056E+05	1.537E+06	6.204E+05	1.013E+06	4.003E+05	3.849E+05
2-methyltetrahydrothiophen-3-one^{17U}	3.982E+06	3.889E+06	3.687E+07	1.726E+07	5.165E+06	2.733E+07	2.304E+07	1.782E+07	1.776E+07
4,6,8-trimethylnon-1-ene^{18D}	1.063E+06	7.240E+05	5.813E+05	7.267E+05	9.725E+05	4.534E+05	7.950E+05	1.480E+06	1.663E+06
5-eicosene^{18U}	2.744E+05	1.413E+05	3.084E+05	5.904E+05	1.741E+05	3.222E+05	4.296E+05	7.640E+05	7.920E+05
6-tridecene^{18U}	1.847E+05	5.702E+05	2.045E+06	2.598E+05	1.011E+05	2.129E+06	7.004E+05	8.030E+06	7.017E+06

Table 22: GC-MS metabolites, with statistical and practical significance, for cellar ICVV. Numbers at the end of the acronyms represent sampling month. Values represent the area of the detected chromatographic peaks after applying variance stabilization. The metabolites can be grouped in the following chemical classes (1) benzene and substituted derivatives, (2) carbonyl compounds, (3) carboxylic acids and derivatives, (4) fatty acyls, (5) hydroxy acids and derivatives, (6) indoles and derivatives, (7) keto acids and derivatives, (8) lactones, (9) olefins, (10) oxolanes, (11) phenols, (12) phenylpropanoic acids, (13) phenol lipids, (14) saturated hydrocarbons, (15) stilbenes, (16) tetrahydrofurans, (17) thiolanes, (18) unsaturated hydrocarbons.

Discussion

In the current study the metabolomic profile of barrel-aged wines from two cellars was surveyed by combining NMR, targeted GC-MS and untargeted GC-MS and LC-MS. The statistical significance (p-value or q-value $\leq$0.05) has been accompanied by practical significance (FCMM $\geq$2) for evaluating the differences between the groups of the factors barrel-type, bottled-wine and time, since statistical significance alone does not provide enough evidence for the importance of the findings [254].

From previous studies we know that NMR is an useful technique to differentiate vintage, geographical origin, climate and ageing effect on bottle-aged wine quality [252, 259, 260]. Consonni *et al.* [252] used NMR to analyse different vintages and ageing times of Amarone wines and found an increase of amino acids during ageing. These authors attributed the increase in amino acids to grape protein degradation ascribed to hydrolysis of yeast and bacteria proteins after their autolysis during ageing process. Nevertheless, in the current study NMR had the least effectiveness in providing informative differences between the groups of the studied factors. From all the metabolites detected by NMR, formate was the only metabolite with both statistical and practical significance for the factor barrel-type in wine samples coming only from ICVV. Although, formic acid has been detected in wine before using NMR [261], neither of the two metabolites has been linked to barrel ageing.

We designed a targeted GC-MS analysis focused on absolute quantification of 4-ethylphenol (4-EP), 4-ethylguaiacol (4-EG), 5-methylfurfural (5-MF), 2-methylisoborneol (2-MIB) and 5-hidroximethylfurfural (5-HMF) as previous studies have related those compounds to wine quality [249, 262]. Additionally, forced ageing conditions during bottle-aged wine storage resulted in a considerable influence on wine quality increasing the production of dioxane and dioxolane isomers, furfural and 5-HMF [249]. However, in ICVV and FB samples only 4-EP, 4-EG and 5-MF were detected by this methodology. In fact, both cellars showed significant differences for the concentration of 5-MF between new and old barrel, having lower concentration in the latter. This could be reasoned from the fact that this compound is connected with the barrel's toasting process justifying its high concentration in the new barrel. Additionally, the high discrepancy of the 5-MF levels between the new barrels in FB and ICVV could be attributed to the intensity of the toasting process, whereas the detection of 5-MF in BTN indicates that merely a few hours are enough to transfer this compound to the wine in the case of new barrel and that the compound is stable in the bottle-aged wine even after a 12-month period. Finally, in both cellars the compounds 4-EP and 4-EG were non-significant and ranged below their

perception thresholds, 620 µg/L and 140 µg/L respectively [263].

Regarding LC-MS, in both cellars the majority of the identified metabolites were flavonoids demonstrating a decreasing tendency with the exception of the metabolites eriodictyol and jaceosidin in FB and 6-methoxyluteolin in ICVV that increased over time. Total flavonoid content has been reported of decreasing after a 70-days storage period [264], however the temporal concentrations of these three metabolites in the ageing wine have not been monitored before despite the fact that their antioxidant properties have been reported [265–267]. Moreover, flavonoid compounds in the wine are represented by groups of flavonols, flavan-3-ols, and anthocyanins. Contact between wine and wood results in a continuous decrease in the anthocyanins content [268] that could be explained by oxidation reactions during ageing or from condensation reactions between anthocyanins and certain wood molecules, all of which generate large, insoluble and precipitable polymers. The second major group of metabolites identified was benzene derivatives that displayed an increasing trend in both cellars, except the metabolites gallic acid and vanillic acid in ICVV which despite their decreasing tendencies they exhibited overall stable levels. Although the lack of time-series studies of benzene derivatives in ageing wine, studies such as Kalua and Boss [269] have observed an increase of this chemical class during the late Cabernet Sauvignon grape ripening stage. From the chemical class of cinnamic acids, caffeic acid has been detected with increasing trend in both cellars although it reached higher concentrations in FB. Assuming that the intensity of the toasting process in FB was lower than ICVV, based on the levels of 5-methylfurfural, we could explain this discrepancy of caffeic acid concentrations since studies such as Alañón *et al.* [270] have shown that the content of caffeic acid was significantly lower in toasted French oak woods compared to non-toasted. The same study also reports scopoletin and syringaldehyde of having correspondingly similar and inverse relation to the toasting process compared to caffeic acid. Although these two metabolites have been detected only in ICVV's wine samples, the similar levels of scopoletin to caffeic acid and the higher concentrations of syringaldehyde in relation to caffeic acid could potentially corroborate these findings. Overall, the two groups of metabolites with temporal increasing and decreasing concentrations, showed statistical and practical significance between the early ($\leq$3 months) and late ($>$3 months) maturation periods, in both cellars, but revealed no differences for the factor barrel-type. Furthermore, although no statistical and practical significance was reported for the factors barrel-type and bottled-wine, for the latter factor subtle differences between the metabolites growth rates created a cumulative effect able to separate the 12-month bottle-aged wines from the 12-month barrel-aged wines.

As far GC-MS is concerned, a higher variety of chemical classes was identified in

both cellars, compared to the other methods, with carboxylic acids being one of the chemical classes that included solely metabolites with increasing trends in both cellars. Among these metabolites was methionol, in ICVV's samples, whose degradation has been suggested as a good indicator of oxidation in the wine and that its concentration could be maintaned depending on the levels of oxygen and the amount of oxygen scavengers [271]. Given that the level of methionol increased during the first trimester and thereafter remained relatively stable, suggests low initial oxidation levels that increased over time. Another chemical class comprised of metabolites with increasing concentrations in both cellars was keto acids with oxoglutaric acid being mutually identified and being described as a metabolite that binds free SO_2 [272]. The rest of the chemical classes included metabolites that presented both increasing and decreasing trends, that could be explained by small fluctuations of temperature and oxygen levels as well as lysing events. Overall, in both cellars GC-MS did not reveal differences based on the factor barrel-type and had better performance than NMR but worse than LC-MS in clustering the samples in a sensible manner based on the growth rates suggesting an underlying noise. Nevertheless, GC-MS was the only method that detected metabolites with practical significance for the factor bottled-wine, that mainly influenced the separation between the 12-month bottle and barrel-aged wines. Regarding this factor, studies such as Aiken and Noble [242] have shown significant differences in the aroma after comparing wine ageing in oak barrels and glass bottles with a trained tasting panel. However, in the current study, among these metabolites, the ones that have been connected to aromatic characteristics are palmitic acid, stearic acid and 1-dodecanol [273, 274] all of them having higher concentrations in the bottle-aged wines.

Conclusions

From the four metabolomic analytical techniques implemented in the current study, NMR was the least effective in providing informative insights based on the given dataset, targeted GC-MS was the only technique that presented significant differences based on the factor barrel-type, whereas LC-MS and GC-MS were the only methods displaying significant differences for the factor time in both regions. From the latter two methods, GC-MS was also the only one with sufficient separating capacity based on the factor bottled-wine. Moreover, the lack of dedicated open source metabolomic database on the organoleptic characteristics of metabolites, renders difficult the inference of the changes imposed on the ageing wine based on the identified metabolites from the untargeted analytical methods. However, in the current study the identified metabolites from these

methods appear to aggregate in two groups; one with increasing and the other with decreasing concentrations.

CHAPTER 6: SUPPLEMENTARY MATERIALS

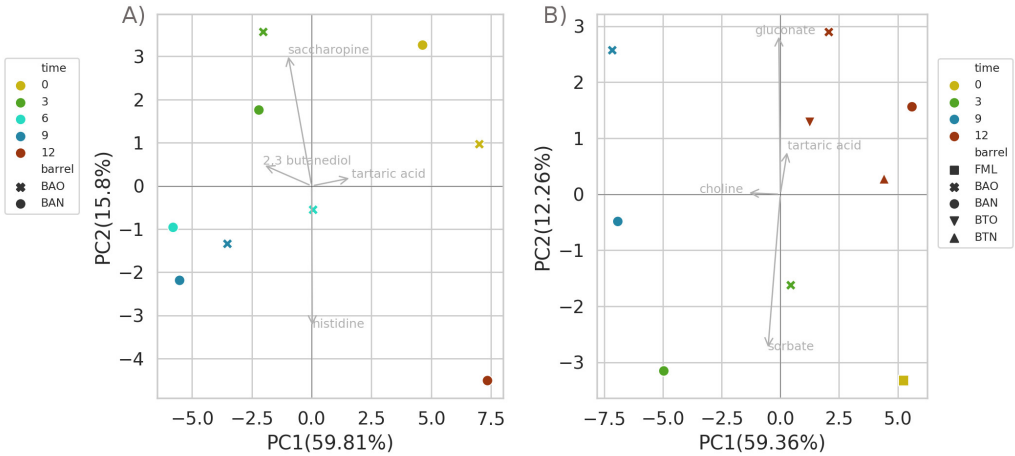


Figure S27: PCA for cellars FB (A) and ICVV (B) based on the NMR metabolites. Metabolites with highest loadings or practical significance are shown. Acronyms BAN and BAO refer to new and old barrel respectively, FML to final malolactic fermentation stage and BTN and BTO to bottle-aged wine from new and old barrel respectively.

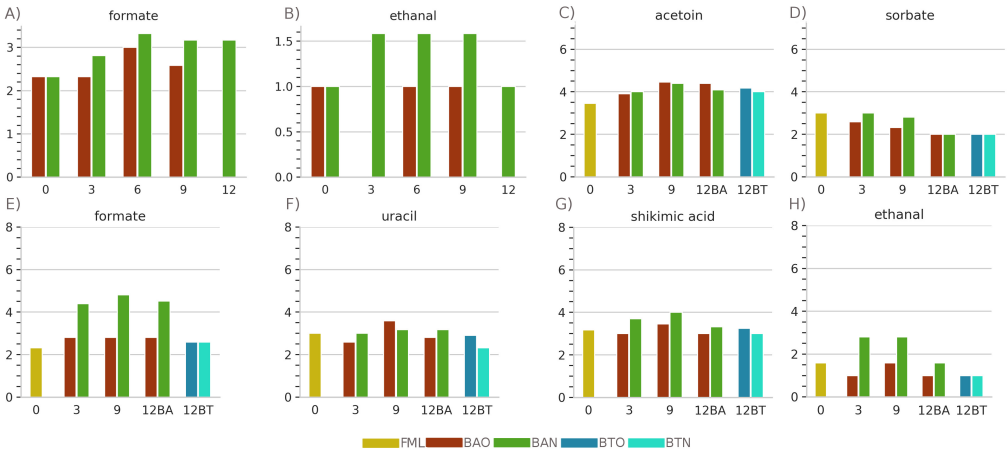


Figure S28: Log2 values of NMR metabolites with practical significance for the cellars FB (A-B) and ICVV (C-H). Acronyms BAN and BAO refer to barrel-aged wine from new and old barrel respectively, FML to final malolactic fermentation stage and BTN and BTO to bottle-aged wine from new and old barrel respectively.

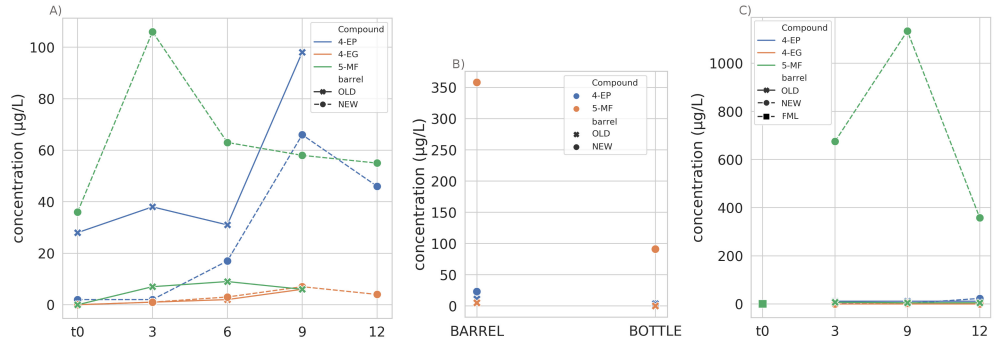


Figure S29: GC-MS targeted based on the factors barrel-type and time for cellar FB (A) and ICVV (C). Figure B concerns the comparison between the 12th month bottle and barrel-aged wine from ICVV cellar. For these samples the concentration of the 4-EG metabolite was zero.

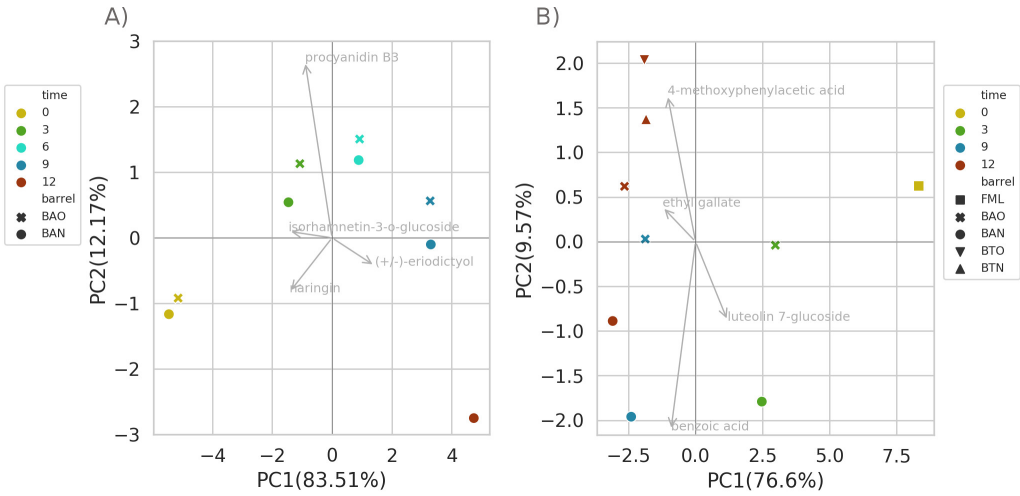


Figure S30: PCA for cellars FB (A) and ICVV (B) based on the LC-MS metabolites. Metabolites with highest loadings acrosss the principal components are shown. Acronyms BAN and BAO refer to new and old barrel respectively, FML to final malolactic fermentation stage and BTN and BTO to bottle-aged wine from new and old barrel respectively.

Figure S31: LC-MS log2 values of metabolites for cellar FB. Acronyms BAO and BAN refer to barrel-aged wine from old and new barrel respectively. Barplots of metabolites with blue background represent the group of metabolites whose concentration has an increasing trend whereas in orange background the group of metabolites with decreasing tendency.

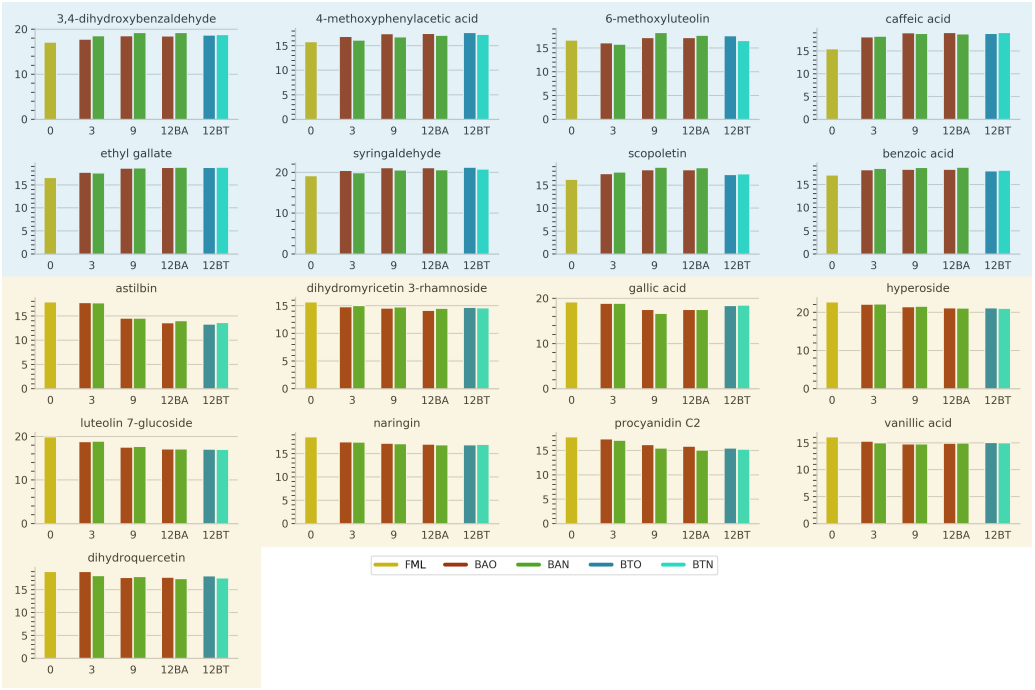


Figure S32: LC-MS log2 values of metabolites for cellar ICVV. Acronyms BAN and BAO refer to new and old barrel respectively, FML to final malolactic fermentation stage and BTN and BTO to bottle-aged wine from new and old barrel respectively. Barplots of metabolites with blue background represent the group of metabolites whose concentration has an increasing trend whereas in orange background the group of metabolites with decreasing tendency.

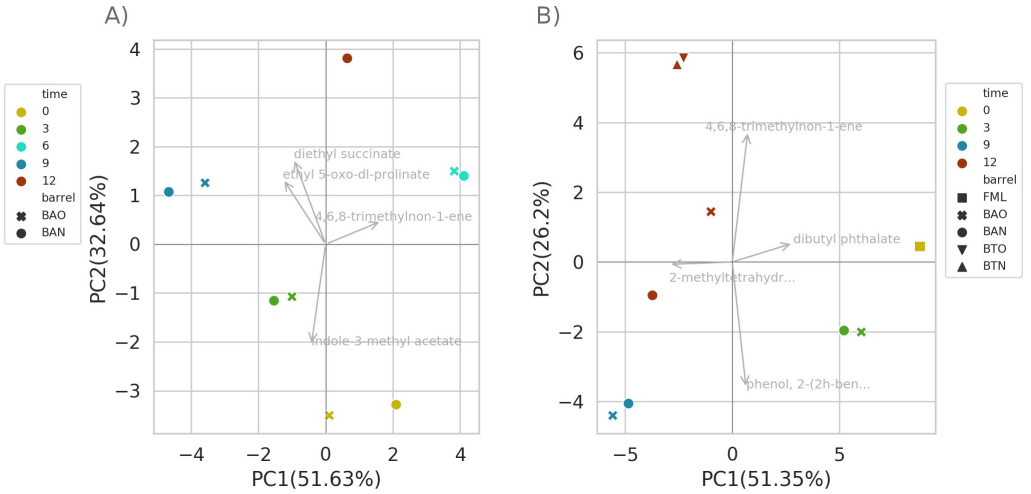


Figure S33: PCA for cellars FB (A) and ICVV (B) based on the GC-MS metabolites. Metabolites with highest loadings acrosss the principal components are shown. Acronyms BAN and BAO refer to new and old barrel respectively, FML to final malolactic fermentation stage and BTN and BTO to bottle-aged wine from new and old barrel respectively.

Figure S34: GC-MS log2 values of metabolites for cellar FB. Acronyms BAO and BAN refer to barrel-aged wine from old and new barrel respectively. Barplots of metabolites with blue background represent the group of metabolites whose concentration has an increasing trend, in orange background the group of metabolites with decreasing tendency and in red background the group of metabolites showing differences during the 6th and 12th month.

Figure S35: GC-MS log2 values of first 25 out of 48 metabolites for cellar ICVV. Acronyms BAN and BAO refer to new and old barrel respectively, FML to final malolactic fermentation stage and BTN and BTO to bottle-aged wine from new and old barrel respectively. Barplots of metabolites with blue background represent the group of metabolites whose concentration has an increasing trend whereas in orange background the group of metabolites with decreasing tendency.

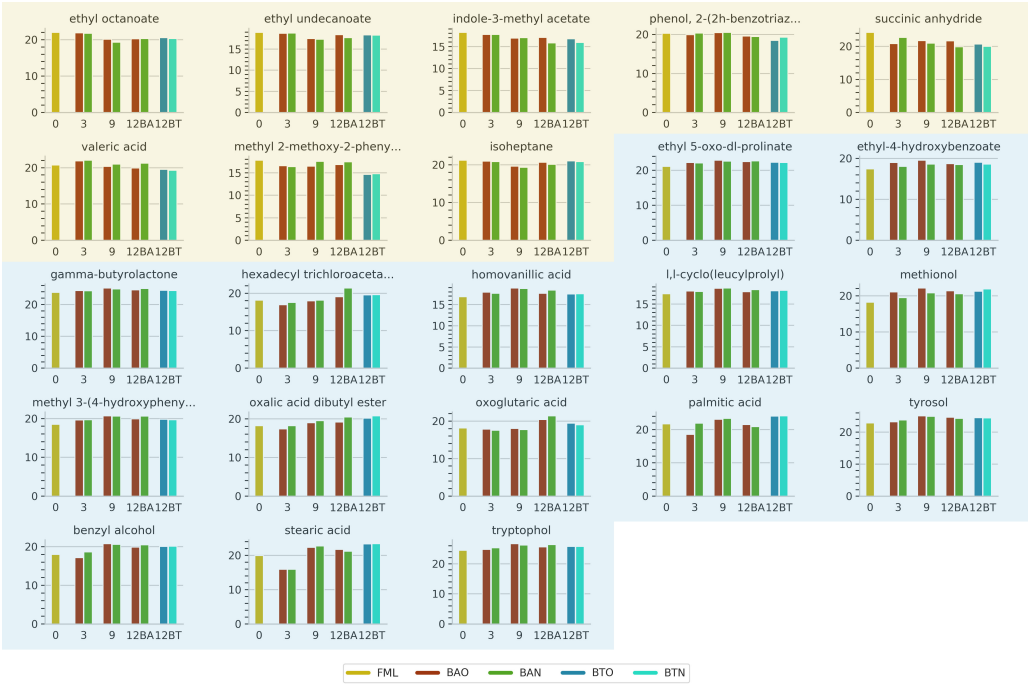


Figure S36: GC-MS log2 values of last 23 out of 48 metabolites for cellar ICVV. Acronyms BAN and BAO refer to new and old barrel respectively. Barplots of metabolites with blue background represent the group of metabolites whose concentration has an increasing trend whereas in orange background the group of metabolites with decreasing tendency.

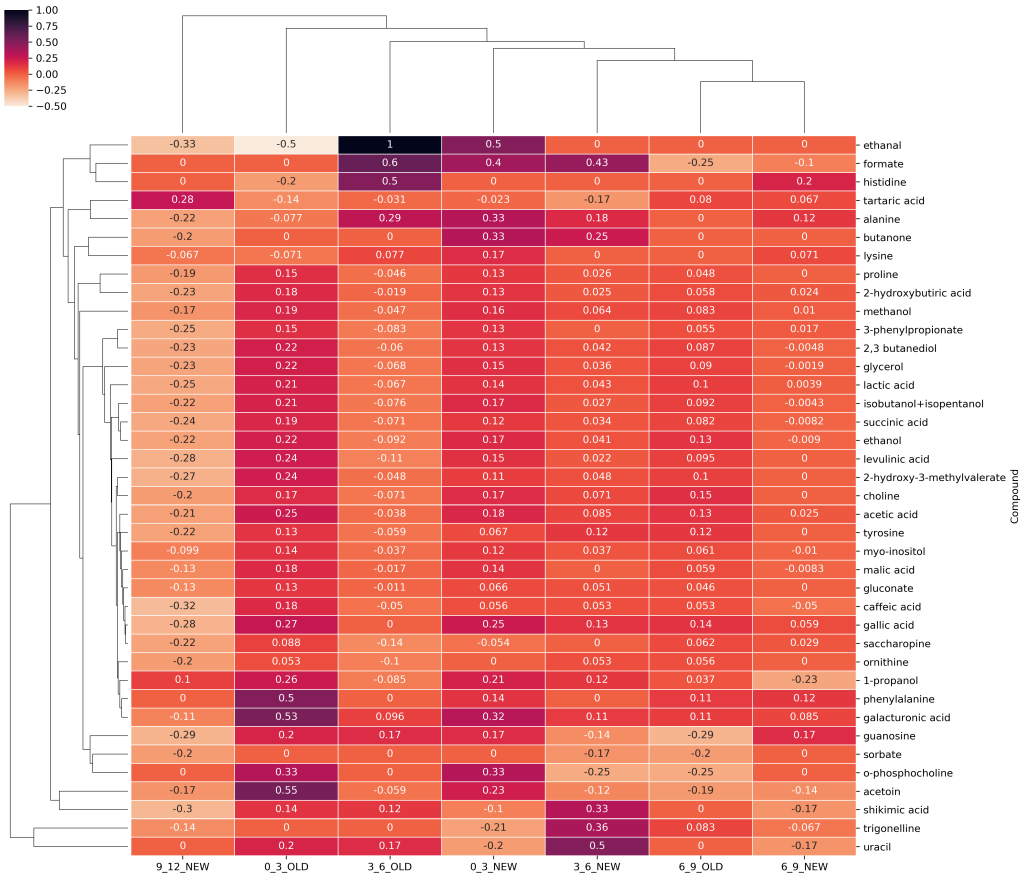


Figure S37: NMR growth rates heatmap for periods 0-3, 3-6, 6-9, and 9-12 based on barrel-type (OLD or NEW) for cellar FB. Values represent percentage of increase or decrease for each metabolite for the given period.

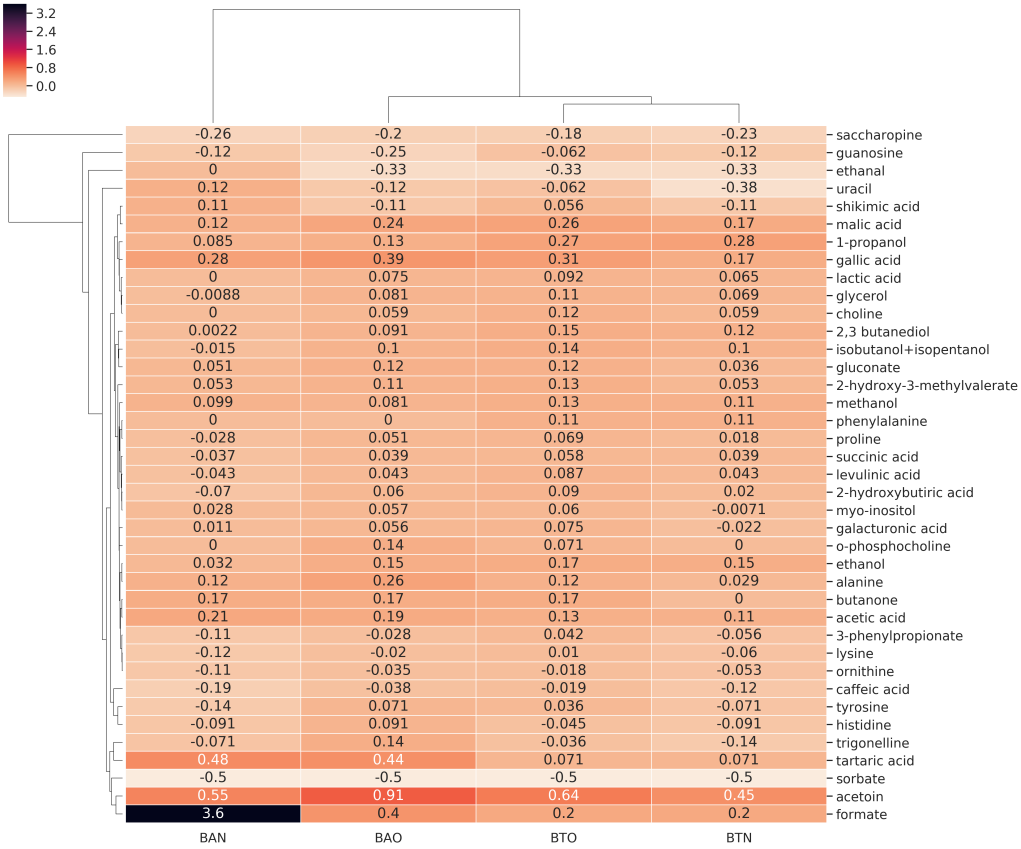


Figure S38: NMR growth rates for period 0-12 of barrel-aged wine from new (BAN) and old (BAO) barrel as well as of bottle-aged wine from new (BTN) and old (BTO) barrel for cellar ICVV. Values represent percentage of increase or decrease for each metabolite for the given period.

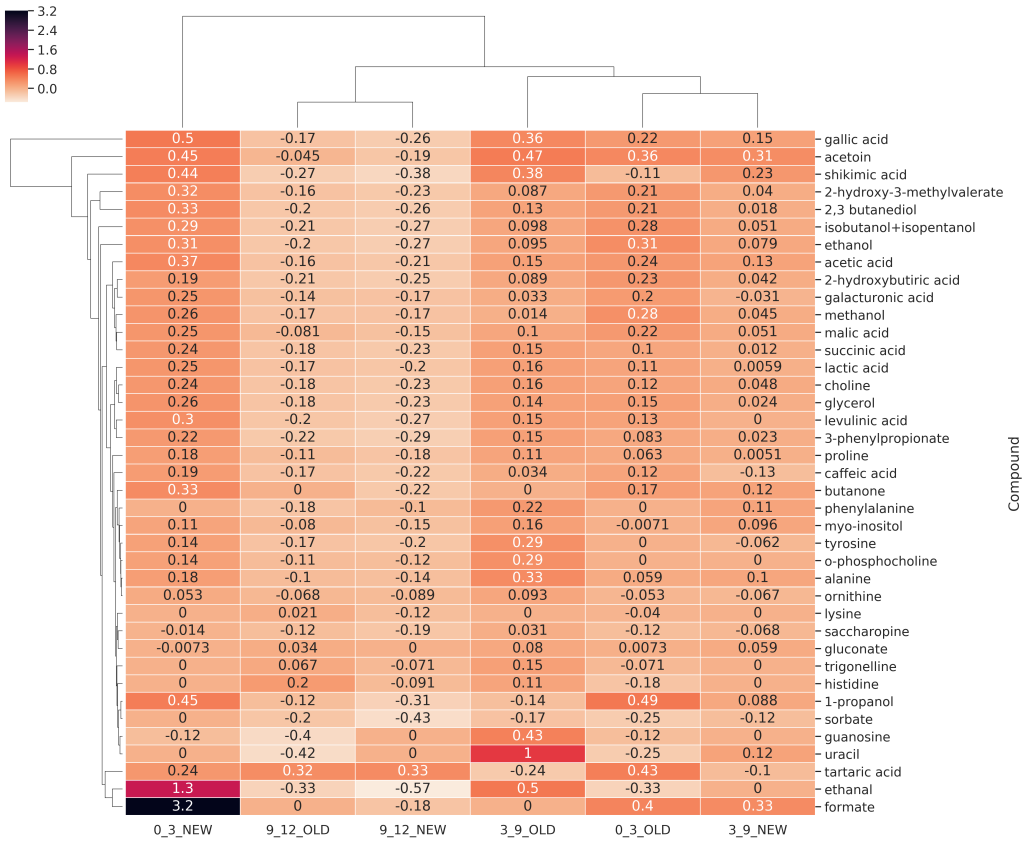


Figure S39: NMR growth rates for period 0-3, 3-9 and 9-12 based on barrel-type (OLD or NEW) for cellar ICVV. Values represent percentage of increase or decrease for each metabolite for the given period.

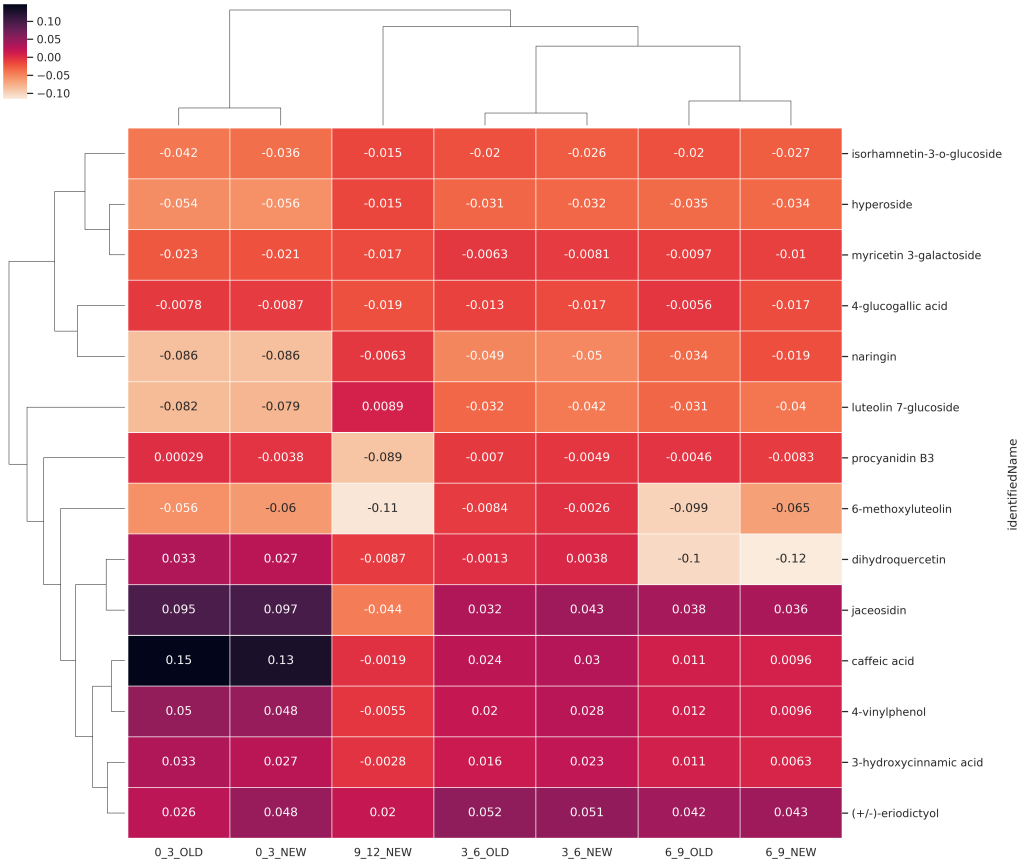


Figure S40: LC-MS growth rates heatmap for periods 0-3, 3-6, 6-9, and 9-12 based on barrel-type (OLD or NEW) for cellar FB. Values represent percentage of increase or decrease for each metabolite for the given period.

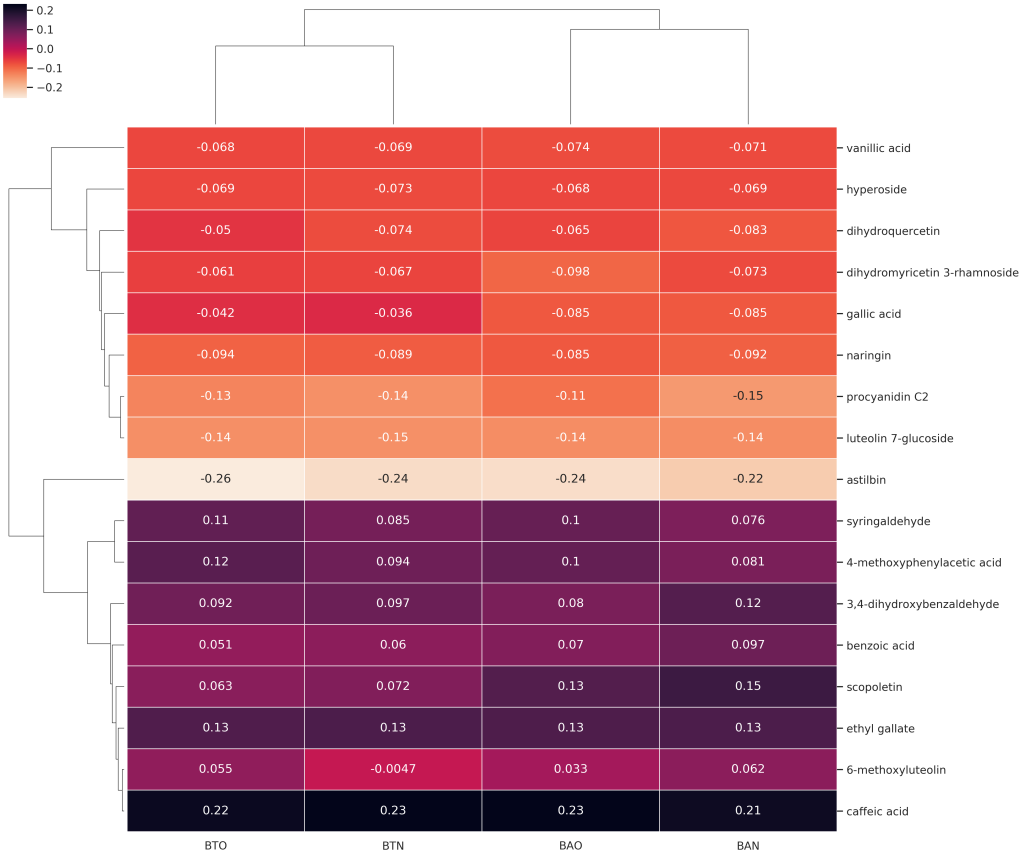


Figure S41: LC-MS growth rates for period 0-12 of barrel-aged wine from new (BAN) and old (BAO) barrel as well as of bottle-aged wine from new (BTN) and old (BTO) barrel for cellar ICVV. Values represent percentage of increase or decrease for each metabolite for the given period.

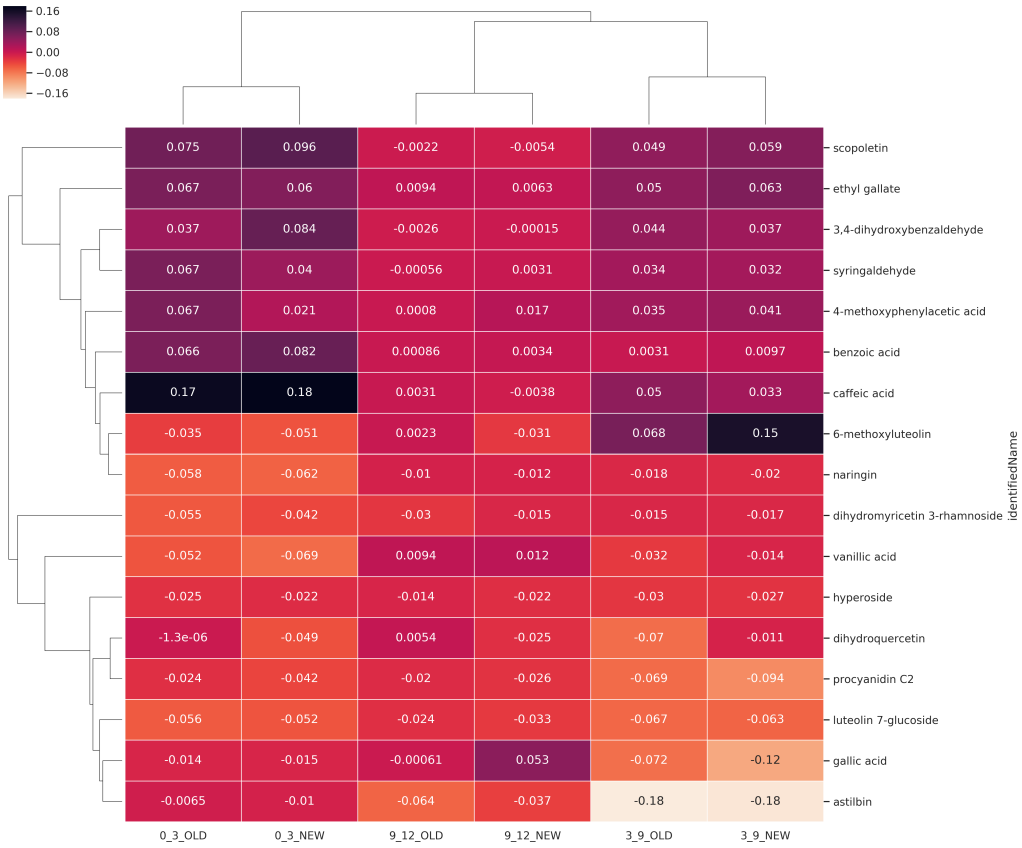


Figure S42: LC-MS growth rates for period 0-3, 3-9 and 9-12 based on barrel-type (OLD or NEW) for cellar ICVV. Values represent percentage of increase or decrease for each metabolite for the given period.

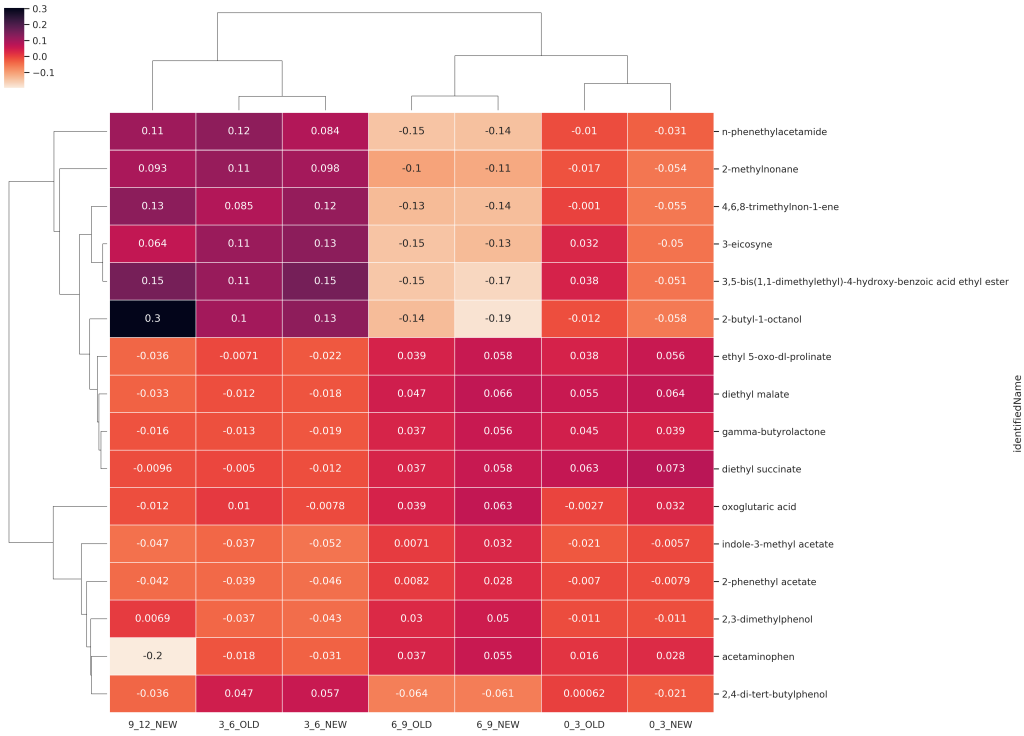


Figure S43: GC-MS growth rates heatmap for periods 0-3, 3-6, 6-9, and 9-12 based on barrel-type (OLD or NEW) for cellar FB. Values represent percentage of increase or decrease for each metabolite for the given period.

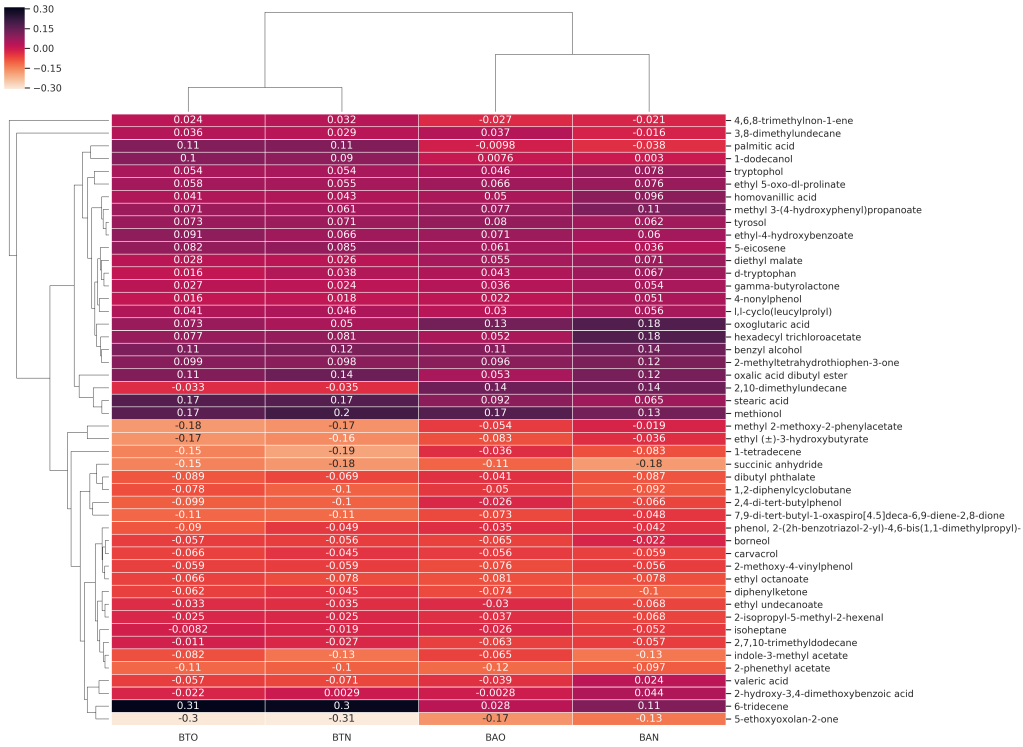


Figure S44: GC-MS growth rates for period 0-12 of barrel-aged wine from new (BAN) and old (BAO) barrel as well as of bottle-aged wine from new (BTN) and old (BTO) barrel for cellar ICVV. Values represent percentage of increase or decrease for each metabolite for the given period.

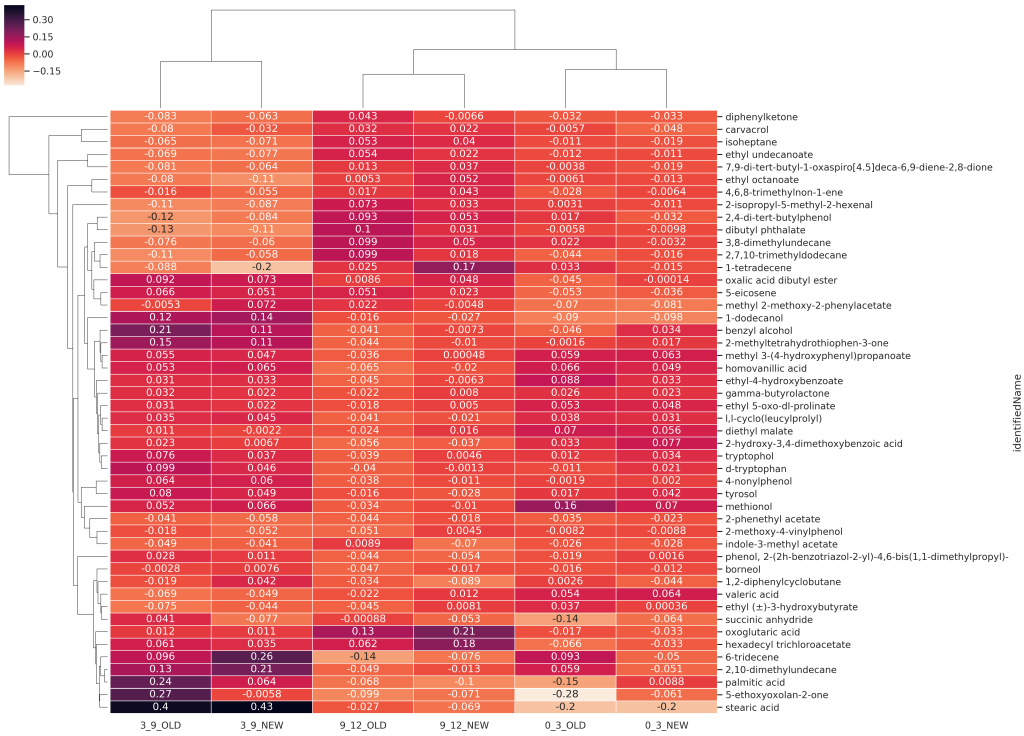


Figure S45: GC-MS growth rates for period 0-3, 3-9 and 9-12 based on barrel-type (OLD or NEW) for cellar ICVV. Values represent percentage of increase or decrease for each metabolite for the given period.

GENERAL DISCUSSION

Winemaking is an intricate process involving a series of stages from the grape harvesting till the wine bottling and commercialization. During the whole process the microbial communities appear quite dynamic, and in each stage various factors and microbial interactions influence the microbial diversity in a way that could potentially confer beneficial or detrimental organoleptic characteristics to the wine quality. For years the study of the microbial dynamics has been performed with conventional methodologies, accompanied by empirical strategies for controlling undesirable microbial activity and growth. Nevertheless, the known limitations of these approaches along with the acknowledgement of the multifactorial basis of the wine spoilage have created the need for innovative methodologies that could allow an in-depth analysis of the microbial diversity. Next-generation-sequencing appears to be the technology that promises such capability, and its implementation aims at providing better understanding of the underlying microbial interactions and ultimately a better control over the final wine product.

The overall work of this dissertation could be summarized as an endeavor to evaluate the performance of NGS technology in comparison to conventional methodologies, as well as to establish a NGS-based bioinformatic and statistical framework for the studying of microbial dynamics during fermentation and winematuration. Moreover, various factors related to the microbial diversity and winemaking process have been assessed, such as grape health and fermentation stage in chapter 2, geographical origin, grape variety and grape maturation state in chapter 4, as well as time, barrel-type and bottled-wine in chapters 5 and 6. All these extensions are referred as NGS-based and biological aspects of the main objectives, respectively.

Evaluation of the NGS-based aspects

The literature review on the state-of-art of the NGS metataxonomic analysis concerning the alcoholic fermentation revealed the potential of this technology since it allows the observation of higher microbial diversity, with various studies confirming previous NGS insights or complementing them with new findings. Nevertheless, it also showed the necessity of a systematic benchmarking and evaluation of the available bioinformatic frameworks, due to the variety of laboratory protocols and the plethora of algorithms and parameters to choose and tweak respectively.

In this work, QIIME appeared to be a convenient and useful bioinformatic framework, as it incorporates numerous algorithms dedicated to the various stages of the bioinformatic and statistical analysis of NGS-based metataxonomic data. Although, these algorithms may be written in different programming languages, they are wrapped in python code providing this way a unified user interface. However, not every aspect of the wrapped algorithms are exposed to the user, making external implementations necessary in cases where higher level of customization is desirable.

The first NGS-based bioinformatic pipeline implemented in chapter 2, corroborated the claims over the enhanced performance of the NGS technology compared to the conventional methodologies. From the latter, plate culturing provided the least informative insights regarding the microbial communities, since the labor intensity of the culturing and the fact tha only certain species are able to grow under laboratory conditions [13], led to the monotoring of only culturable microorganisms, whereas qPCR [22, 150] was limited to the detection of targeted species of the communities. On the other hand, PCR-DGGE exhibited better performance from the side of the conventional methodologies. However the limitation regarding its inability of detecting populations in low concentrations [22, 150] became apparent, since it failed to detect species, such as *Saccharomyces cerevisiae*, that were in low abundance in the must and the beginning of the alcoholic fermentation. Overall, the trends of the microbial dynamics demonstrated by NGS were in congruence with those shown by the conventional methodologies, however NGS exhibited higher sensitivity at detecting low abundance microbial populations, revealed higher bacterial diversity than any other method, and 70% of the yeasts that it identified were also detected by other methods.

Although, in chapter 2 the phylogenetic diversity and weighted UniFrac distance were useful metrics for the evaluation of the alpha and beta diversity respectively, an alternative was sought for the study of these two types of diversity at genus level. Even though, both metrics can be calculated at genus level, their current implementation in QIIME renders the process of collapsing at this level a complicated process due to the

way that the phylogenetic tree is constructed. By default, QIIME assigns taxonomy to the maximum level that is possible, which is the species level. Although, an algorithm that aggregates the OTU table at genus level and merges the OTU counts exists in QIIME, currently there is no direct way of collapsing the representative sequences of the grouped OTUs. This is a known limitation and has already been discussed in QIIME's forums [4]. Since the phylogenetic tree is constructed using the representative sequences, the resulted diversity metrics cannot be collapsed at a given taxonomic level without resorting to elaborate external procedures. This limitation along with the fact that the UniFrac distance may lead to misleading results with ITS amplicons [275], leave Shannon index and Bray-Curtis distance as better metrics for studying alpha and beta diversity, respectively, at genus level in QIIME.

In chapter 3, numerous aspects related to the laboratory and bioinformatic workflows were examined, however their evaluation had a qualitative character since only a single mock community was utilized in the study. Nevertheless, certain findings corroborated previous results, such as those from the study of Van Der Pol *et al.* [31] regarding the similar performance of OTU (with 99% similarity threshold) and ASV clustering methods, and Trtkova *et al.* [193] concerning the higher specificity of the ITS amplicon compared to the 18S amplicon. Furthermore, as it has been pointed-out by Ahn *et al.* [27], the number of PCR cycles should be retained as low as possible, since small increase does not confer any observed improvement compared to a previously established value. As far the rest of the results are concerned, Illumina platform is a preferable choice especially in cases where machine learning taxonomic algorithms of QIIME are to be implemented, since these algorithms are trained with Illumina generated metataxonomic data [34]. Moreover, the Phred quality threshold should be set as high as possible, with Q20 being a suggested minimum, by this work, as it represents 99% base call accuracy [29], and led to a better taxonomic performance than Q10. Overall, the established QIIME's default parameters exhibited good performance, and even though a significant amount of mock communities is necessary for an in-depth modification of these parameters, the usage of a single mock community is suggested, by this work, as a form of an initial elementary performance evaluation of any bioinformatic pipeline.

The statistical analysis framework GNEISS represents a non-parametric alternative that utilizes all the available metataxonomic data, and is not restricted to assumptions regarding homogeneous and randomly distributed microbial diversity within the samples, as rarefaction assumes [276]. This kind of assumptions may affect the effectiveness of setting a rarefaction threshold based on rarefaction curves, since the observed microbial

[4] https://forum.qiime2.org/t/genus-level-gene-trees/11369/4

diversity, after the quality filtering of the raw sequencing, cannot be considered as true microbial diversity. Moreover, inferences based on GNEISS do not run the risk of being underpowered due to small number of samples as with Kruskal-Wallis [277], since GNEISS provides balances and not p-values. Therefore, detached from these limitations, the calculated balances consist a normalized version of the OTU counts that could be used for differential or factor analysis. Despite the benefits of using GNEISS, its main drawback concerns the lack of direct information associated with the taxa that drive the observed differences among the samples, making further analysis necessary for that purpose.

Evaluation of the biological aspects

Despite the lack of common basis regarding the biological aspects of the main objectives, due to the differences and peculiarities of the experimental design in each chapter, the overall evaluation of these aspects concerns the impact of various factors on the observed microbial diversity, as well as the insights provided by the NGS-based metataxonomic implementation. These insights are assessed based on their level of alignment to previous studies and the new information that they provide.

In chapter 2, the health status of the grapes seemed to represent a factor that captured better the changes of the bacterial dynamics, with the bacterial communities being significantly different between healthy and damaged grapes. This demonstrates their opportunistic nature, where availability of nutrients may cause bacterial growth in early stages [2, 14], and underlines the need of proper handling of the grapes during harvesting. Although, NGS identified higher bacterial diversity than conventional methods, the majority was below 1% abundance and considered as noise. From the rest, the most abundant taxa, *Acetobacter*, *Gluconobacter*, *Gluconoacetobacter* and *Oenococcus* have also been identified by other NGS studies [49, 79], and the results of chapter 2 indicated *Oenococcus* as the taxa with the ability to draw distinctions between healthy and damaged grapes. Even though, this inference was based on statistical analysis of the relative abundances, it was also denoted by the weighted UniFrac PCoA (Figure 6B) where a clear separation of the final fermentation stage could be observed between the healthy and damaged grapes. These findings proved once again the well-adapted nature of *Oenococcus* to the alcoholic fermentation compared to other LAB [2]. However, the fact that its thriving appeared to be benefited only by the healthy grapes, raises doubts about the suggested lack of rivalry between AAB and LAB [16, 17].

On the other hand, the fermentation stage was the factor that had greater discriminatory power on the fungal communities, with their populations at the early fermentation

stages being significantly different than those at the late stages. This factor, revealed the previously described fungal dynamics with the rise and dominance of the *Saccharomyces cerevisiae* towards the end of the fermentation, along with the parallel decline of the non-*Saccharomyces* yeasts and the aerobic filamentous fungi [3]. Overall, NGS managed to identify genera described also by other studies [42, 44, 111], and more specifically it identified previously reported taxa that are associated with damaged grapes such as *Botrytis*, *Kazachstania*, *Zygosaccharomyces* and *Issatchenkia* [53, 136, 137]. Additionally, the potential wine spoilers *Aspergillus* and *Penicillium* were detected at the very early stages of the fermentation in the healthy grapes, highlighting the capacity of NGS to provide a more refined monitor of the quality-threatening microbial diversity.

In chapter 4, the results verified previous studies regarding the concept of "*terroir*" and its connection to the wine microbiota [42–49], as well as the influence of the harvest time [22, 44, 50, 52–54] and grape variety [55] on the observed microbial diversity. All three factors managed to capture 75% of the observed variance among the fungal communities, with geographical origin having greater impact (63%) than grape variety (11%) and maturation state (1%). This could be explained from the fact that terroir represents a broader factor, whose impact is derived from the cumulative effect of additional factors such as soil, climate, and human field interventions, each one having its own influence on the microbial diversity. Regarding the observed fungal diversity, the difficulty of isolating *Saccharomyces cerevisiae* from the grape surface [3] was confirmed, and the fungal populations were consisted of non-*Saccharomyces* and filamentous fungi. Concerning the latter, GNEISS pointed-out the previously reported on the grape surface *Aspergillus*, *Botrytis*, *Colletotrichium*, *Rhodotorula*, and *Penicillium* [278–280] as the genera that mainly influenced the observed differences among the vineyards. Although, filamentous fungi have received less attention regarding their beneficial contribution to the final wine product than *Saccharomyces* and non-*Saccharomyces* yeasts, organoleptic characteristics such as earthy or musty aromas have been attributed to their activity and have been described as characteristics that are potentially vineyard dependant [281].

As in chapter 2, the limitations of the conventional methodologies could also be observed in chapter 5, complemented by the inability to culture AAB in laboratory conditions due to their VBNC state [19, 241]. Regarding the studied factors, the barrels demonstrated similar microbial diversities, implying an insignificant impact of the factor barrel-type on the microbiota when it concerns a new barrel and a barrel with 1-year of prior use. On the other hand, the microbial diversity in the bottles was similar to that at the end of FML, when the wine was introduced into the barrels, whereas higher diversity was observed at the 12-month barrel samples (factor bottled-wine). This underlines

the impact that the oak wood has on the microbial communities, since its porosity and absorbent properties may allow the establishment of microorganisms that could be nourished and grow once they come in contact with the wine [50]. Furthermore, the factor time revealed a stable microbial diversity during the ageing process, and a more dynamic bacterial communities that showed differences between the early and late maturation periods. These findings are aligned with previous studies showing stability of the microbial diversity during wine ageing [150], and possibility of microbial growth due to alterations in the wine caused by the characteristics of barrel's wood, fluctuations of the storage conditions and cell autolysis [13].

From the bacterial communities, one of the most abundant genera, in the barrels, was *Oenococcus* and from the fungal communities *Saccharomyces*. This demonstrated the well adapted nature, to the alcoholic and malolactic fermentation, of the former and the dominating attitude of the latter [2, 3]; behaviours that were observed in both cellars. Additionally, despite the differences regarding the grape variety and storage conditions between the FB and ICVV cellars, the bacterial genera *Acetobacter*, *Oenococcus* and *Lactobacillus*, as well as the fungal genera *Saccharomyces*, *Hanseniaspora*, *Aspergillus* and *Malassezia* were detected in both cellars. These findings urge the need for further investigation, in order to infer whether these genera consist a regular microbial composition during winematuration.

The metabolomic analysis in chapter 6 did not allow the extraction of direct conclusions regarding the properties that the identified metabolites could potentially confer to the wine, due to the lack of dedicated and curated databases for this purpose. Furthermore, the analysis of the metabolites at the chemical class level was hindered by the different trends the metabolites demonstrate within a given chemical class. Therefore, the growth rates of the metabolites were calculated for different maturation periods, and were used to perform hierarchical clustering of the samples. This process revealed that the clustering of the samples, based on the untargeted metabolic methods, was mainly influenced by two groups of metabolites; one with increasing and the other with decreasing tendency. Based on these two groups, untargeted GC-MS was found to be adequate for showing differences based on the factor bottled-wine and LC-MS based on the factor time. The latter, was the method also used by Fang *et al.* [282] for studying the flavonoid content of red wines stored in three oak barrels, with the results suggesting great association between flavonoid levels and wine ageing.

Regarding the targeted metabolomic methods, NMR did not manage to reveal any informative pattern based on this dataset. Nevertheless, this result should not be considered as doubt towards its usefulness, since there are many studies that have demonstrated the

benefits of its application [252, 259, 260]. On the other hand, targeted GC-MS confirmed the connection of 5-MF to the barrel's toasting process and intensity [283, 284], making it ideal for drawing distinctions for the factor barrel-type. Moreover, the differences of 5-MF in the bottled wines showed that this compound could be transferred, at detectable and stable for 1-year levels, to bottled wines from a new barrel after a short maturation period. The stability of the 5-MF inside the bottles could be extrapolated to the rest of the volatile compounds, and explain thus the ability of the untargeted GC-MS to show differences for the factor bottled-wine. These conclusions indicate that the combination of targeted and untargeted metabolomic methods is necessary so as to obtain a more spherical view of the wine metabolome based on different factors.

Future research and limitations

Next-generation-sequencing appears a very promising technology, and based on the results presented by this work it seems capable of surpassing the limitations of the conventional methodologies and provide an enhanced view into the microbial diversity. Moreover, the derived results managed to corroborate previous findings associated with the winemaking process, suggest improvements on the bioinformatic and statistical analyses, and provide new insights regarding the various factors that influence the vinification and winematuration process. Therefore, based on these considerations we may deduce that the main hypothesis of this dissertation holds true.

Despite its potential, NGS requires further improvement in order to reach the level of controlling the final wine product based on microbial interactions. From statistical standpoint, the known shortcomings of hypothesis testing and the resulting p-value [285, 286], urge the need of developing methodologies that could show the size of an effect associated with a discovery. In this work, the log2-transformation of the OTU counts has been implemented, as it can easily be translated to fold-change. However, this transformation has to be performed on the rarefied OTU table, with rarefaction carrying its own assumptions as already being discussed. Furthermore, an accurate estimation of the effect size is hindered by the difficulty of determining the amount of the true microbial variation that is accounted by the observed variation of the OTU counts, since the extracted DNA quantity is usually inconsistent [287] and the microbial diversity is altered during the filtering of low quality raw sequences.

In order to understand the microbial interactions, we need to incorporate metatranscriptomic data into the wine analysis, since this type of data can provide useful information regarding the kind of microorganisms that are active on a given moment, and potentially allow to discard assigned taxonomies originated from dead cells. However,

this presupposes the ability of accurately identifying microorganisms at species level and the availability of corresponding reference genomes. Additionally, such implementation is impeded by the difficulty of extracting RNA from wine samples at adequate quantity and quality, due to the presence of wine molecules such as phenols and enzymatic inhibitors [288].

Therefore, machine learning approaches consist an alternative solution for the control of the wine final product, with the incorporation of metataxonomic and metabolomic data. As mentioned in chapter 1, such approximation has been applied by Bokulich *et al.* [109], where the authors implemented the random forest classifier so as to predict the wine metabolome using the wine microbiota. There are many supervised learning solutions to achieve this goal, with neural networks being the most flexible machine learning application. Nevertheless, the amount of input necessary for the training process of such algorithms, prohibits even the consideration of such solutions due to the current high financial burden of NGS. Therefore, more meta-analysis is necessary to provide the necessary data pool, and on the other hand to corroborate findings presented either in this dissertation or other studies.

Overall, in this work we have shown that the establishment of a NGS-based implementation for a metataxonomic analysis, requires an in-depth look and evaluation of the available bioinformatic tools and methods. Therefore, according to the presented general results, the analytical framework that this dissertation has established represents a solid foundation upon which further research can be based in order to provide additional improvements and insights.

GENERAL CONCLUSIONS

1. NGS-based metataxonomic analysis exhibits higher sensitivity regarding low abundance microbial populations, and reveals higher microbial diversity than conventional methodologies.

2. The dynamics of the fungal communities are more pronounced based on the fermentation stage, whereas the bacterial communities are influenced more by the grape health.

    - *Oenococcus* abundance appears to be affected negatively by damaged grapes.

3. QIIME is a useful NGS-based bioinformatic and statistical framework for metataxonomic analysis. However, a systematic evaluation and benchmarking of its available algorithms and default parameters is necessary in order to achieve robust performance.

    - Q20 represents the suggested minimum threshold for filtering raw metataxonomic data based on Phred quality scores, since it represents 99% base call accuracy and provides better taxonomic classification results.
    - Shannon index and Bray-Curtis distance are better metrics, compared to phylogenetic diversity and UniFrac distance, for studying alpha and beta diversity, respectively, at genus level in QIIME.
    - ASV and OTU clustering methods have similar performance at genus level, as long as the similarity threshold of the latter is set at 99%.

4. The multifactorial nature of the concept of terroir, causes the factor geographical origin to have greater impact on the fungal communities than grape variety and berry maturation state.

- The populations of the filamentous fungi on the grape surface appear quite diverse in different vineyards, due to the geographical characteristics of the latter.
- GNEISS represents a promising non-parametric statistical framework for an unbiased inference of the microbial dynamics.

5. During a 1-year maturation period, the microbial diversity remains relative stable with sporadic microbial growth that could be attributed to the material of the barrel, fluctuations of the storage conditions and cell autolysis events.

   - Similar microbial diversity should be expected between a new barrel and a barrel with 1-year of prior use.
   - Within a 1-year ageing period, the material and characteristics of an oak barrel confers changes to the microbial diversity that would not be observed if the wine was stored in glass bottles.

6. A combination of targeted and untargeted metabolomic methods is necessary in order to infer changes of the wine metabolome based on different factors.

   - LC-MS appears suitable technique for revealing differences based on the factor time, targeted GC-MS on the factor barrel-type and untargeted GC-MS on the factor bottled-wine.
   - The detected levels of 5-methylfurfural by targeted GC-MS verified its association with the barrel's toasting process and intensity, and demonstrated that this compound can be transferred at detectable and stable levels to bottled wines from new barrels and after a short storage period.

REFERENCES

[1] Padilla, B.; Gil, J.V.; Manzanares, P. Past and future of non-Saccharomyces yeasts: from spoilage microorganisms to biotechnological tools for improving wine aroma complexity. *Frontiers in microbiology* **2016**, *7*, 411.

[2] Ribéreau-Gayon, P.; Dubourdieu, D.; Donèche, B.; Lonvaud, A. *Handbook of enology, the microbiology of wine and vinifications*; Vol. 1, John Wiley & Sons, 2006.

[3] Mas, A.; Torija, M.J.; Beltran, G.; Sengun, I.Y. Winemaking: Microbiology. In *Wine Making: Basics and Applied Aspects*; Joshi, V.; Ray, R.C., Eds.; Taylor and Francis, 2020; chapter 9, pp. 224–253.

[4] Jolly, N.P.; Varela, C.; Pretorius, I.S. Not your ordinary yeast: non-Saccharomyces yeasts in wine production uncovered. *FEMS yeast research* **2014**, *14*, 215–237.

[5] Mas, A.; Padilla, B.; Esteve-Zarzoso, B.; Beltran, G.; Reguant, C.; Bordons, A. Taking advantage of natural biodiversity for wine making: The WILDWINE Project. *Agric. Agric. Sci. Procedia* **2016**, *8*, 4–9.

[6] Pretorius, I.; Van der Westhuizen, T.; Augustyn, O. Yeast biodiversity in vineyards and wineries and its importance to the South African wine industry. A review. *South African Journal of Enology and Viticulture* **1999**, *20*, 61–70.

[7] Torija, M.J.; Rozes, N.; Poblet, M.; Guillamón, J.M.; Mas, A. Yeast population dynamics in spontaneous fermentations: comparison between two different wine-producing areas over a period of three years. *Antonie Van Leeuwenhoek* **2001**, *79*, 345–352.

[8] Beltran, G.; Torija, M.J.; Novo, M.; Ferrer, N.; Poblet, M.; Guillamón, J.M.; Rozès, N.; Mas, A. Analysis of yeast populations during alcoholic fermentation: a six year follow-up study. *Systematic and Applied Microbiology* **2002**, *25*, 287–293.

[9] Combina, M.; Elía, A.; Mercado, L.; Catania, C.; Ganga, A.; Martinez, C. Dynamics of indigenous yeast populations during spontaneous fermentation of wines from Mendoza, Argentina. *International journal of food microbiology* **2005**, *99*, 237–243.

[10] Nurgel, C.; Erten, H.; Canbas, A.; Cabaroglu, T.; Selli, S. Yeast flora during the fermentation of wines made from vitis viniferaL. cv. emir and kalecik karasi grown in anatolia. *World Journal of Microbiology and Biotechnology* **2005**, *21*, 1187–1194.

[11] Di Maro, E.; Ercolini, D.; Coppola, S. Yeast dynamics during spontaneous wine fermentation of the Catalanesca grape. *International Journal of Food Microbiology* **2007**, *117*, 201–210.

[12] Ocón, E.; Gutiérrez, A.; Garijo, P.; Tenorio, C.; López, I.; López, R.; Santamaría, P. Quantitative and qualitative analysis of non-Saccharomyces yeasts in spontaneous alcoholic fermentations. *European Food Research and Technology* **2010**, *230*, 885–891.

[13] Guillamón, J.M.; Mas, A. Acetic acid bacteria. In *Biology of Microorganisms on Grapes, in Must and in Wine*; König, H.; Unden, G.; Fröhlich, J., Eds.; Springer, 2017; pp. 43–64.

[14] Fleet, G.H. Wine. In *Food microbiology fundamentals and frontiers, 2nd edn.*; Doyle, M.; Beuchat, L.; Montville, T., Eds.; ASM, Washington DC, 2001; pp. 267–288.

[15] Narendranath, N.V.; Thomas, K.C.; Ingledew, W.M. Acetic acid and lactic acid inhibition of growth of Saccharomyces cerevisiae by different mechanisms. *Journal of the American Society of Brewing Chemists* **2001**, *59*, 187–194.

[16] Joyeux, A.; Lafon-Lafourcade, S.; Ribéreau-Gayon, P. Evolution of acetic acid bacteria during fermentation and storage of wine. *Applied and Environmental Microbiology* **1984**, *48*, 153–156.

[17] Guillamón, J.; González, A.; Poblet, M.; Mas, A. Development of molecular techniques for the analysis of acetic acid bacteria in winemaking. Yeast-Bacteria Interactions. Lallemand Tecnical Meetings, 2002, Vol. 10, pp. 45–49.

[18] Bokulich, N.A.; Ohta, M.; Richardson, P.M.; Mills, D.A. Monitoring seasonal changes in winery-resident microbiota. *PloS one* **2013**, *8*, e66437.

[19] Torija, M.; Mateo, E.; Guillamón, J.; Mas, A. Identification and quantification of acetic acid bacteria in wine and vinegar by TaqMan–MGB probes. *Food Microbiology* **2010**, *27*, 257–265.

[20] Valera, M.J.; Torija, M.J.; Mas, A.; Mateo, E. Acetobacter malorum and Acetobacter cerevisiae identification and quantification by real-time PCR with TaqMan-MGB probes. *Food microbiology* **2013**, *36*, 30–39.

[21] Andorrà, I.; Landi, S.; Mas, A.; Esteve-Zarzoso, B.; Guillamn, J.M. Effect of fermentation temperature on microbial population evolution using culture-independent and dependent techniques. *Food Research International* **2010**, *43*, 773 – 779.

[22] Prakitchaiwattana, C.J.; Fleet, G.H.; Heard, G.M. Application and evaluation of denaturing gradient gel electrophoresis to analyse the yeast ecology of wine grapes. *FEMS yeast research* **2004**, *4*, 865–877.

[23] Cocolin, L.; Bisson, L.; Mills, D. Direct profiling of the yeast dynamics in wine fermentations. *FEMS microbiology letters* **2000**, *189*, 81–87.

[24] Hesham, A.E.L.; Khan, S.; Liu, X.; Zhang, Y.; Wang, Z.; Yang, M. Application of PCR–DGGE to analyse the yeast population dynamics in slurry reactors during degradation of polycyclic aromatic hydrocarbons in weathered oil. *Yeast* **2006**, *23*, 879–887.

[25] Loureiro, V.; Malfeito-Ferreira, M. Spoilage yeasts in the wine industry. *International journal of food microbiology* **2003**, *86*, 23–50.

[26] Breitwieser, F.P.; Lu, J.; Salzberg, S.L. A review of methods and databases for metagenomic classification and assembly. *Briefings in bioinformatics* **2017**.

[27] Ahn, J.H.; Kim, B.Y.; Song, J.; Weon, H.Y. Effects of PCR cycle number and DNA polymerase type on the 16S rRNA gene pyrosequencing analysis of bacterial communities. *Journal of Microbiology* **2012**, *50*, 1071–1074.

[28] Edgar, R. UCHIME2: improved chimera prediction for amplicon sequencing. *BioRxiv* **2016**, p. 074252.

[29] Ewing, B.; Green, P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome research* **1998**, *8*, 186–194.

[30] Callahan, B.J.; McMurdie, P.J.; Holmes, S.P. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *The ISME journal* **2017**, *11*, 2639.

[31] Van Der Pol, W.J.; Kumar, R.; Morrow, C.D.; Blanchard, E.E.; Taylor, C.M.; Martin, D.H.; Lefkowitz, E.J.; Muzny, C.A. In Silico and Experimental Evaluation of Primer Sets for Species-Level Resolution of the Vaginal Microbiota Using 16S Ribosomal RNA Gene Sequencing. *The Journal of infectious diseases* **2018**, *219*, 305–314.

[32] Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; others. Scikit-learn: Machine learning in Python. *Journal of machine learning research* **2011**, *12*, 2825–2830.

[33] Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T.L. BLAST+: architecture and applications. *BMC bioinformatics* **2009**, *10*, 421.

[34] Bokulich, N.A.; Kaehler, B.D.; Rideout, J.R.; Dillon, M.; Bolyen, E.; Knight, R.; Huttley, G.A.; Caporaso, J.G. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2s q2-feature-classifier plugin. *Microbiome* **2018**, *6*, 90.

[35] Faith, D.P. Conservation evaluation and phylogenetic diversity. *Biological conservation* **1992**, *61*, 1–10.

[36] Forman, R. Land Mosaics: The ecology of landscapes and regions (1995). *The Ecological Design and Planning Reader. IslandPress, Washington, DC* **2014**, pp. 217–234.

[37] Lozupone, C.; Lladser, M.E.; Knights, D.; Stombaugh, J.; Knight, R. UniFrac: an effective distance metric for microbial community comparison. *The ISME journal* **2011**, *5*, 169–172.

[38] Lozupone, C.A.; Hamady, M.; Kelley, S.T.; Knight, R. Quantitative and qualitative β diversity measures lead to different insights into factors that structure microbial communities. *Appl. Environ. Microbiol.* **2007**, *73*, 1576–1585.

[39] Beals, E.W. Bray-Curtis ordination: an effective strategy for analysis of multivariate ecological data. In *Advances in ecological research*; Elsevier, 1984; Vol. 14, pp. 1–55.

[40] Morton, J.T.; Sanders, J.; Quinn, R.A.; McDonald, D.; Gonzalez, A.; Vázquez-Baeza, Y.; Navas-Molina, J.A.; Song, S.J.; Metcalf, J.L.; Hyde, E.R.; others. Balance trees reveal microbial niche differentiation. *MSystems* **2017**, *2*, e00162–16.

[41] Steel, C.C.; Blackman, J.W.; Schmidtke, L.M. Grapevine bunch rots: impacts on wine composition, quality, and potential procedures for the removal of wine faults. *Journal of agricultural and food chemistry* **2013**, *61*, 5189–5206.

[42] Bokulich, N.A.; Thorngate, J.H.; Richardson, P.M.; Mills, D.A. Microbial biogeography of wine grapes is conditioned by cultivar, vintage, and climate. *Proceedings of the National Academy of Sciences* **2014**, *111*, E139–E148.

[43] Pinto, C.; Pinho, D.; Sousa, S.; Pinheiro, M.; Egas, C.; Gomes, A.C. Unravelling the diversity of grapevine microbiome. *PLoS One* **2014**, *9*, e85622.

[44] Setati, M.E.; Jacobson, D.; Andong, U.C.; Bauer, F. The vineyard yeast microbiome, a mixed model microbial map. *PLoS One* **2012**, *7*.

[45] Gilbert, J.A.; van der Lelie, D.; Zarraonaindia, I. Microbial terroir for wine grapes. *Proceedings of the National Academy of Sciences* **2014**, *111*, 5–6.

[46] Taylor, M.W.; Tsai, P.; Anfang, N.; Ross, H.A.; Goddard, M.R. Pyrosequencing reveals regional differences in fruit-associated fungal communities. *Environmental microbiology* **2014**, *16*, 2848–2858.

[47] Morrison-Whittle, P.; Lee, S.A.; Goddard, M.R. Fungal communities are differentially affected by conventional and biodynamic agricultural management approaches in vineyard ecosystems. *Agriculture, Ecosystems & Environment* **2017**, *246*, 306–313.

[48] Burns, K.N.; Bokulich, N.A.; Cantu, D.; Greenhut, R.F.; Kluepfel, D.A.; O'Geen, A.T.; Strauss, S.L.; Steenwerth, K.L. Vineyard soil bacterial diversity and composition revealed by 16S rRNA genes: differentiation by vineyard management. *Soil Biology and Biochemistry* **2016**, *103*, 337–348.

[49] del Carmen Portillo, M.; Franquès, J.; Araque, I.; Reguant, C.; Bordons, A. Bacterial diversity of Grenache and Carignan grape surface from different vineyards at Priorat wine region (Catalonia, Spain). *International journal of food microbiology* **2016**, *219*, 56–63.

[50] RENOUF, V.; CLAISSE, O.; Lonvaud-Funel, A. Understanding the microbial ecosystem on the grape berry surface through numeration and identification of yeast and bacteria. *Australian Journal of Grape and Wine Research* **2005**, *11*, 316–327.

[51] Martins, G.; Vallance, J.; Mercier, A.; Albertin, W.; Stamatopoulos, P.; Rey, P.; Lonvaud, A.; Masneuf-Pomarède, I. Influence of the farming system on the epiphytic yeasts and yeast-like fungi colonizing grape berries during the ripening process. *International journal of food microbiology* **2014**, *177*, 21–28.

[52] Jolly, N.; Augustyn, O.; Pretorius, I. The occurrence of non-Saccharomyces cerevisiae yeast species over three vintages in four vineyards and grape musts from four production regions of the Western Cape, South Africa. *South African Journal of Enology and Viticulture* **2003**, *24*, 35–42.

[53] Barata, A.; Malfeito-Ferreira, M.; Loureiro, V. The microbial ecology of wine grape berries. *International journal of food microbiology* **2012**, *153*, 243–259.

[54] Martins, G.; Miot-Sertier, C.; Lauga, B.; Claisse, O.; Lonvaud-Funel, A.; Soulas, G.; Masneuf-Pomarède, I. Grape berry bacterial microbiota: impact of the ripening process and the farming system. *International journal of food microbiology* **2012**, *158*, 93–100.

[55] Kondrashov, A.; Ševčík, R.; Benáková, H.; Koštířová, M.; Štípek, S. The key role of grape variety for antioxidant capacity of red wines. *e-SPEN, the European e-Journal of Clinical Nutrition and Metabolism* **2009**, *4*, e41–e46.

[56] Alañón, M.; Pérez-Coello, M.; Marina, M. Wine science in the metabolomics era. *TrAC Trends in Analytical Chemistry* **2015**, *74*, 1–20.

[57] Pinu, F.R. Grape and wine metabolomics to develop new insights using untargeted and targeted approaches. *Fermentation* **2018**, *4*, 92.

[58] Revenue of the wine market worldwide by country. `https://www.statista.com/forecasts/758149/revenue-of-the-wine-market-worldwide-by-country`.

[59] Yan, M.; Pamp, S.J.; Fukuyama, J.; Hwang, P.H.; Cho, D.Y.; Holmes, S.; Relman, D.A. Nasal microenvironments and interspecific interactions influence nasal microbiota complexity and S. aureus carriage. *Cell host & microbe* **2013**, *14*, 631–640.

[60] Ghaisas, S.; Maher, J.; Kanthasamy, A. Gut microbiome in health and disease: Linking the microbiome–gut–brain axis and environmental factors in the pathogenesis of systemic and neurodegenerative diseases. *Pharmacology & therapeutics* **2016**, *158*, 52–62.

[61] Carrau, F.M.; Medina, K.; Boido, E.; Farina, L.; Gaggero, C.; Dellacassa, E.; Versini, G.; Henschke, P.A. De novo synthesis of monoterpenes by Saccharomyces cerevisiae wine yeasts. *FEMS microbiology letters* **2005**, *243*, 107–115.

[62] Varela, C.; Siebert, T.; Cozzolino, D.; Rose, L.; McLean, H.; Henschke, P. Discovering a chemical basis for differentiating wines made by fermentation with wildindigenous and inoculated yeasts: role of yeast volatile compounds. *Australian Journal of Grape and Wine Research* **2009**, *15*, 238–248.

[63] Ciani, M.; Comitini, F.; Mannazzu, I.; Domizio, P. Controlled mixed culture fermentation: a new perspective on the use of non-Saccharomyces yeasts in winemaking. *FEMS Yeast Research* **2010**, *10*, 123–133.

[64] Belda, I.; Zarraonaindia, I.; Perisin, M.; Palacios, A.; Acedo, A. From Vineyard Soil to Wine Fermentation: Microbiome Approximations to Explain the terroir Concept. *Frontiers in Microbiology* **2017**.

[65] Bokulich, N.A.; Ohta, M.; Richardson, P.M.; Mills, D.A. Monitoring Seasonal Changes in Winery-Resident Microbiota. *PLoS ONE* **2013**, *8*, e66437.

[66] Spano, G.; Torriani, S. Editorial: Microbiota of Grapes: Positive and Negative Role on Wine Quality. *Frontiers in Microbiology* **2016**, *7*, 2036.

[67] Amann, R.I.; Ludwig, W.; Schleifer, K.H. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiological reviews* **1995**, *59*, 143–169.

[68] Curtis, T.P.; Sloan, W.T.; Scannell, J.W. Estimating prokaryotic diversity and its limits. *Proceedings of the National Academy of Sciences* **2002**, *99*, 10494–10499.

[69] Franzosa, E.A.; Hsu, T.; Sirota-Madi, A.; Shafquat, A.; Abu-Ali, G.; Morgan, X.C.; Huttenhower, C. Sequencing and beyond: integrating molecular'omics' for microbial community profiling. *Nature Reviews Microbiology* **2015**, *13*, 360–372.

[70] Sanger, F.; Nicklen, S.; Coulson, A.R. DNA sequencing with chain-terminating inhibitors. *Proceedings of the national academy of sciences* **1977**, *74*, 5463–5467.

[71] Rusk, N. Torrents of sequence. *Nature Methods* **2011**, *8*, 44–44.

[72] Quail, M.A.; Smith, M.; Coupland, P.; Otto, T.D.; Harris, S.R.; Connor, T.R.; Bertoni, A.; Swerdlow, H.P.; Gu, Y. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC genomics* **2012**, *13*, 341.

[73] Gupta, A.K.; Gupta, U. Next generation sequencing and its applications **2014**. pp. 345–367.

[74] Margulies, M.; Egholm, M.; Altman, W.E.; Attiya, S.; Bader, J.S.; Bemben, L.A.; Berka, J.; Braverman, M.S.; Chen, Y.J.; Chen, Z.; others. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **2005**, *437*, 376–380.

[75] Morozova, O.; Marra, M.A. Applications of next-generation sequencing technologies in functional genomics. *Genomics* **2008**, *92*, 255–264.

[76] Pettersson, E.; Lundeberg, J.; Ahmadian, A. Generations of sequencing technologies. *Genomics* **2009**, *93*, 105–111.

[77] Bleidorn, C. Third generation sequencing: technology and its potential impact on evolutionary biodiversity research. *Systematics and biodiversity* **2016**, *14*, 1–8.

[78] Morgan, H.H.; Du Toit, M.; Setati, M.E. The grapevine and wine microbiome: Insights from high-throughput amplicon sequencing. *Frontiers in microbiology* **2017**, *8*.

[79] Bokulich, N.A.; Joseph, C.L.; Allen, G.; Benson, A.K.; Mills, D.A. Next-generation sequencing reveals significant bacterial diversity of botrytized wine. *PloS one* **2012**, *7*, e36357.

[80] Campisano, A.; Antonielli, L.; Pancher, M.; Yousaf, S.; Pindo, M.; Pertot, I. Bacterial endophytic communities in the grapevine depend on pest management. *PLoS One* **2014**, *9*, e112763.

[81] Perazzolli, M.; Antonielli, L.; Storari, M.; Puopolo, G.; Pancher, M.; Giovannini, O.; Pindo, M.; Pertot, I. Resilience of the natural phyllosphere microbiota of the grapevine to chemical and biological pesticides. *Applied and environmental microbiology* **2014**, *80*, 3585–3596.

[82] Sundquist, A.; Bigdeli, S.; Jalili, R.; Druzin, M.L.; Waller, S.; Pullen, K.M.; El-Sayed, Y.Y.; Taslimi, M.M.; Batzoglou, S.; Ronaghi, M. Bacterial flora-typing with targeted, chip-based Pyrosequencing. *BMC microbiology* **2007**, *7*, 108.

[83] Liu, Z.; DeSantis, T.Z.; Andersen, G.L.; Knight, R. Accurate taxonomy assignments from 16S rRNA sequences produced by highly parallel pyrosequencers. *Nucleic acids research* **2008**, *36*, e120–e120.

[84] Chakravorty, S.; Helb, D.; Burday, M.; Connell, N.; Alland, D. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *Journal of microbiological methods* **2007**, *69*, 330–339.

[85] David, V.; Terrat, S.; Herzine, K.; Claisse, O.; Rousseaux, S.; Tourdot-Maréchal, R.; Masneuf-Pomarede, I.; Ranjard, L.; Alexandre, H. High-throughput sequencing of amplicons for monitoring yeast biodiversity in must and during alcoholic fermentation. *Journal of industrial microbiology & biotechnology* **2014**, *41*, 811–821.

[86] Holland, T.C.; Bowen, P.; Bogdanoff, C.; Hart, M. Arbuscular mycorrhizal fungal communities associated with Vitis vinifera vines under different frequencies of irrigation. *American Journal of Enology and Viticulture* **2014**, pp. ajev–2014.

[87] Bokulich, N.A.; Mills, D.A. Improved selection of internal transcribed spacer-specific primers enables quantitative, ultra-high-throughput profiling of fungal communities. *Applied and environmental microbiology* **2013**, *79*, 2519–2526.

[88] Stefanini, I.; Albanese, D.; Cavazza, A.; Franciosi, E.; De Filippo, C.; Donati, C.; Cavalieri, D. Dynamic changes in microbiota and mycobiota during spontaneous Vino Santo Trentinofermentation. *Microbial biotechnology* **2016**, *9*, 195–208.

[89] Caporaso, J.G.; Kuczynski, J.; Stombaugh, J.; Bittinger, K.; Bushman, F.D.; Costello, E.K.; Fierer, N.; Pena, A.G.; Goodrich, J.K.; Gordon, J.I.; others. QIIME allows analysis of high-throughput community sequencing data. *Nature methods* **2010**, *7*, 335.

[90] Schloss, P.D.; Westcott, S.L.; Ryabin, T.; Hall, J.R.; Hartmann, M.; Hollister, E.B.; Lesniewski, R.A.; Oakley, B.B.; Parks, D.H.; Robinson, C.J.; others. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and environmental microbiology* **2009**, *75*, 7537–7541.

[91] Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *Journal of molecular biology* **1990**, *215*, 403–410.

[92] Caporaso, J.G.; Bittinger, K.; Bushman, F.D.; DeSantis, T.Z.; Andersen, G.L.; Knight, R. PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* **2009**, *26*, 266–267.

[93] Wang, Q.; Garrity, G.M.; Tiedje, J.M.; Cole, J.R. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and environmental microbiology* **2007**, *73*, 5261–5267.

[94] Price, M.N.; Dehal, P.S.; Arkin, A.P. FastTree 2–approximately maximum-likelihood trees for large alignments. *PloS one* **2010**, *5*, e9490.

[95] Edgar, R.C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **2010**, *26*, 2460–2461.

[96] Meyer, F.; Paarmann, D.; D'Souza, M.; Olson, R.; Glass, E.M.; Kubal, M.; Paczian, T.; Rodriguez, A.; Stevens, R.; Wilke, A.; others. The metagenomics RAST server–a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC bioinformatics* **2008**, *9*, 386.

[97] Plummer, E.; Twin, J.; Bulach, D.M.; Garland, S.M.; Tabrizi, S.N. A comparison of three bioinformatics pipelines for the analysis of preterm gut microbiota using 16S rRNA gene sequencing data. *Journal of Proteomics & Bioinformatics* **2015**, *8*, 283.

[98] DeSantis, T.Z.; Hugenholtz, P.; Larsen, N.; Rojas, M.; Brodie, E.L.; Keller, K.; Huber, T.; Dalevi, D.; Hu, P.; Andersen, G.L. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and environmental microbiology* **2006**, *72*, 5069–5072.

[99] Quast, C.; Pruesse, E.; Yilmaz, P.; Gerken, J.; Schweer, T.; Yarza, P.; Peplies, J.; Glöckner, F.O. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic acids research* **2012**, *41*, D590–D596.

[100] Abarenkov, K.; Henrik Nilsson, R.; Larsson, K.H.; Alexander, I.J.; Eberhardt, U.; Erland, S.; Høiland, K.; Kjøller, R.; Larsson, E.; Pennanen, T.; others. The UNITE database for molecular identification of fungi–recent updates and future perspectives. *New Phytologist* **2010**, *186*, 281–285.

[101] Cole, J.R.; Wang, Q.; Fish, J.A.; Chai, B.; McGarrell, D.M.; Sun, Y.; Brown, C.T.; Porras-Alfaro, A.; Kuske, C.R.; Tiedje, J.M. Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic acids research* **2013**, *42*, D633–D642.

[102] Portillo, M.; Mas, A. Analysis of microbial diversity and dynamics during wine fermentation of Grenache grape variety by high-throughput barcoding sequencing. *LWT-Food Science and Technology* **2016**, *72*, 317–321.

[103] Du Toit, W.; Lambrechts, M. The enumeration and identification of acetic acid bacteria from South African red wine fermentations. *International journal of food microbiology* **2002**, *74*, 57–64.

[104] González, Á.; Hierro, N.; Poblet, M.; Mas, A.; Guillamón, J.M. Application of molecular methods to demonstrate species and strain evolution of acetic acid bacteria population during wine production. *International Journal of Food Microbiology* **2005**, *102*, 295–304.

[105] Bokulich, N.A.; Swadener, M.; Sakamoto, K.; Mills, D.A.; Bisson, L.F. Sulfur dioxide treatment alters wine microbial diversity and fermentation progression in a dose-dependent fashion. *American Journal of Enology and Viticulture* **2015**, *66*, 73–79.

[106] Godálová, Z.; Kraková, L.; Puškárová, A.; Bučková, M.; Kuchta, T.; Piknová, L.; Pangallo, D. Bacterial consortia at different wine fermentation phases of two typical Central European grape varieties: Blaufränkisch (Frankovka modrá) and Grüner Veltliner (Veltlínske zelené). *International journal of food microbiology* **2016**, *217*, 110–116.

[107] Campanaro, S.; Treu, L.; Vendramin, V.; Bovo, B.; Giacomini, A.; Corich, V. Metagenomic analysis of the microbial community in fermented grape marc reveals that Lactobacillus fabifermentans is one of the dominant species: insights into its genome structure. *Applied microbiology and biotechnology* **2014**, *98*, 6015–6037.

[108] Zarraonaindia, I.; Owens, S.M.; Weisenhorn, P.; West, K.; Hampton-Marcell, J.; Lax, S.; Bokulich, N.A.; Mills, D.A.; Martin, G.; Taghavi, S.; others. The soil microbiome influences grapevine-associated microbiota. *MBio* **2015**, *6*, e02527–14.

[109] Bokulich, N.A.; Collins, T.S.; Masarweh, C.; Allen, G.; Heymann, H.; Ebeler, S.E.; Mills, D.A. Associations among wine grape microbiome, metabolome, and fermentation behavior suggest microbial contribution to regional wine characteristics. *MBio* **2016**, *7*, e00631–16.

[110] Knight, S.; Klaere, S.; Fedrizzi, B.; Goddard, M.R. Regional microbial signatures positively correlate with differential wine phenotypes: evidence for a microbial aspect to terroir. *Scientific reports* **2015**, *5*, 14233.

[111] Pinto, C.; Pinho, D.; Cardoso, R.; Custódio, V.; Fernandes, J.; Sousa, S.; Pinheiro, M.; Egas, C.; Gomes, A.C. Wine fermentation microbiome: a landscape from different Portuguese wine appellations. *Frontiers in microbiology* **2015**, *6*.

[112] Setati, M.E.; Jacobson, D.; Bauer, F.F. Sequence-based analysis of the Vitis vinifera L. cv Cabernet Sauvignon grape must mycobiome in three South African vineyards employing distinct agronomic systems. *Frontiers in microbiology* **2015**, *6*.

[113] Lleixà, J.; Martín, V.; Portillo, M.d.C.; Carrau, F.; Beltran, G.; Mas, A. Comparison of fermentation and wines produced by inoculation of Hanseniaspora vineae and Saccharomyces cerevisiae. *Frontiers in microbiology* **2016**, *7*.

[114] Couto, J.A.; Neves, F.; Campos, F.; Hogg, T. Thermal inactivation of the wine spoilage yeasts Dekkera/Brettanomyces. *International journal of food microbiology* **2005**, *104*, 337–344.

[115] Suárez, R.; Suárez-Lepe, J.; Morata, A.; Calderón, F. The production of ethylphenols in wine by yeasts of the genera Brettanomyces and Dekkera: a review. *Food Chemistry* **2007**, *102*, 10–21.

[116] Selişteanu, D.; Petre, E.; Roman, M.; Ionete, C.; Popescu, D. Estimation strategies for kinetic parameters of an alcoholic fermentation bioprocess. SICE Annual Conference 2010, Proceedings of. IEEE, 2010, pp. 3560–3565.

[117] Killian, E.; Ough, C. Fermentation estersformation and retention as affected by fermentation temperature. *American Journal of Enology and Viticulture* **1979**, *30*, 301–305.

[118] Mc Lellan, M.; others. The effect of fermentation temperature on chemical and sensory characteristics of wines from seven white grape cultivars grown in New York State. *American journal of enology and viticulture* **1986**, *37*, 190–194.

[119] Torija, M.J.; Beltran, G.; Novo, M.; Poblet, M.; Guillamón, J.M.; Mas, A.; Rozes, N. Effects of fermentation temperature and Saccharomyces species on the cell fatty acid composition and presence of volatile compounds in wine. *International journal of food microbiology* **2003**, *85*, 127–136.

[120] Dubourdieu, D.; Tominaga, T.; Masneuf, I.; des Gachons, C.P.; Murat, M.L. The role of yeasts in grape flavor development during fermentation: the example of Sauvignon blanc. *American Journal of Enology and Viticulture* **2006**, *57*, 81–88.

[121] Eglinton, J.M.; McWILLIAM, S.J.; FOGARTY, M.W.; FRANCIS, I.L.; KWIATKOWSKI, M.J.; HØJ, P.B.; HENSCHKE, P.A. The effect of Saccharomyces bayanus-mediated fermentation on the chemical composition and aroma profile of Chardonnay wine. *Australian Journal of Grape and Wine Research* **2000**, *6*, 190–196.

[122] Cramer, A.C.; Vlassides, S.; Block, D.E. Kinetic model for nitrogen-limited wine fermentations. *Biotechnology and Bioengineering* **2002**, *77*, 49–60.

[123] Birch, R.M.; Ciani, M.; Walker, G.M. Magnesium, calcium and fermentative metabolism in wine yeasts. *Journal of Wine Research* **2003**, *14*, 3–15.

[124] McBryde, C.; Gardner, J.M.; de Barros Lopes, M.; Jiranek, V. Generation of novel wine yeast strains by adaptive evolution. *American Journal of Enology and Viticulture* **2006**, *57*, 423–430.

[125] Soden, A.; Francis, I.; Oakey, H.; Henschke, P. Effects of co-fermentation with Candida stellata and Saccharomyces cerevisiae on the aroma and composition of Chardonnay wine. *Australian Journal of Grape and Wine Research* **2000**, *6*, 21–30.

[126] De Filippis, F.; La Storia, A.; Stellato, G.; Gatti, M.; Ercolini, D. A selected core microbiome drives the early stages of three popular Italian cheese manufactures. *PLoS One* **2014**, *9*, e89680.

[127] Chaillou, S.; Chaulot-Talmon, A.; Caekebeke, H.; Cardinal, M.; Christieans, S.; Denis, C.; Desmonts, M.H.; Dousset, X.; Feurer, C.; Hamon, E.; others. Origin and ecological selection of core and food-specific bacterial communities associated with meat and seafood spoilage. *The ISME journal* **2015**, *9*, 1105–1118.

[128] Bokulich, N.A.; Rideout, J.R.; Mercurio, W.G.; Shiffer, A.; Wolfe, B.; Maurice, C.F.; Dutton, R.J.; Turnbaugh, P.J.; Knight, R.; Caporaso, J.G. mockrobiota: a public resource for microbiome bioinformatics benchmarking. *MSystems* **2016**, *1*, e00062–16.

[129] Bokulich, N.A.; Rideout, J.R.; Kopylova, E.; Bolyen, E.; Patnode, J.; Ellett, Z.; McDonald, D.; Wolfe, B.; Maurice, C.F.; Dutton, R.J.; others. A standardized, extensible framework for optimizing classification improves marker-gene taxonomic assignments. Technical report, PeerJ PrePrints, 2015.

[130] Fleet, G.H. Yeast interactions and wine flavour. *International journal of food microbiology* **2003**, *86*, 11–22.

[131] Huber, C.; McFadden-Smith, W.; Inglis, D. Management and etiology of grape sour rot in the Niagara region. Phytopathology, 2011, Vol. 101, pp. S259–S259.

[132] Barata, A.; Santos, S.C.; Malfeito-Ferreira, M.; Loureiro, V. New insights into the ecological interaction between grape berry microorganisms and Drosophila flies during the development of sour rot. *Microbial ecology* **2012**, *64*, 416–430.

[133] Nigro, F.; Schena, L.; Ligorio, A.; Pentimone, I.; Ippolito, A.; Salerno, M.G. Control of table grape storage rots by pre-harvest applications of salts. *Postharvest Biology and Technology* **2006**, *42*, 142–149.

[134] Barata, A.; Seborro, F.; Belloch, C.; Malfeito-Ferreira, M.; Loureiro, V. Ascomycetous yeast species recovered from grapes damaged by honeydew and sour rot. *Journal of applied microbiology* **2008**, *104*, 1182–1191.

[135] Mateo, E.; Torija, M.; Mas, A.; Bartowsky, E. Acetic acid bacteria isolated from grapes of South Australian vineyards. *International journal of food microbiology* **2014**, *178*, 98–106.

[136] Nisiotou, A.A.; Spiropoulos, A.E.; Nychas, G.J.E. Yeast community structures and dynamics in healthy and Botrytis-affected grape must fermentations. *Appl. Environ. Microbiol.* **2007**, *73*, 6705–6713.

[137] Barata, A.; Malfeito-Ferreira, M.; Loureiro, V. Changes in sour rotten grape berry microbiota during ripening and wine fermentation. *International journal of food microbiology* **2012**, *154*, 152–161.

[138] Rantsiou, K.; Urso, R.; Iacumin, L.; Cantoni, C.; Cattaneo, P.; Comi, G.; Cocolin, L. Culture-dependent and-independent methods to investigate the microbial ecology of Italian fermented sausages. *Appl. Environ. Microbiol.* **2005**, *71*, 1977–1986.

[139] Millet, V.; Lonvaud-Funel, A. The viable but non-culturable state of wine microorganisms during storage. *Letters in applied microbiology* **2000**, *30*, 136–141.

[140] Cocolin, L.; Campolongo, S.; Alessandria, V.; Dolci, P.; Rantsiou, K. Culture independent analyses and wine fermentation: an overview of achievements 10 years after first application. *Annals of Microbiology* **2011**, *61*, 17–23.

[141] Nocker, A.; Burr, M.; Camper, A.K. Genotypic microbial community profiling: a critical technical review. *Microbial ecology* **2007**, *54*, 276–289.

[142] Nisiotou, A.A.; Rantsiou, K.; Iliopoulos, V.; Cocolin, L.; Nychas, G.J.E. Bacterial species associated with sound and Botrytis-infected grapes from a Greek vineyard. *International journal of food microbiology* **2011**, *145*, 432–436.

[143] Ercolini, D. High-throughput sequencing and metagenomics: moving forward in the culture-independent analysis of food microbial ecology. *Appl. Environ. Microbiol.* **2013**, *79*, 3148–3155.

[144] De Man, J.; Rogosa, d.; Sharpe, M.E. A medium for the cultivation of lactobacilli. *Journal of applied Bacteriology* **1960**, *23*, 130–135.

[145] De Angelo, J.; Siebert, K. A new medium for the detection of wild yeast in brewing culture yeast. *Journal of the American Society of Brewing Chemists* **1987**, *45*, 135–140.

[146] Heard, G.; Fleet, G. Evaluation of selective media for enumeration of yeasts during wine fermentation. *Journal of applied bacteriology* **1986**, *60*, 477–481.

[147] Jolly, N.; Augustyn, O.; Pretorius, I. The role and use of non-Saccharomyces yeasts in wine production. *South African Journal of Enology and Viticulture* **2006**, *27*, 15–39.

[148] Hierro, N.; Esteve-Zarzoso, B.; González, Á.; Mas, A.; Guillamón, J.M. Real-time quantitative PCR (QPCR) and reverse transcription-QPCR for detection and enumeration of total yeasts in wine. *Appl. Environ. Microbiol.* **2006**, *72*, 7148–7155.

[149] Hierro, N.; Esteve-Zarzoso, B.; Mas, A.; Guillamón, J.M. Monitoring of Saccharomyces and Hanseniaspora populations during alcoholic fermentation by real-time quantitative PCR. *FEMS Yeast Research* **2007**, *7*, 1340–1349.

[150] Andorrà, I.; Landi, S.; Mas, A.; Esteve-Zarzoso, B.; Guillamón, J.M. Effect of fermentation temperature on microbial population evolution using culture-independent and dependent techniques. *Food Research International* **2010**, *43*, 773–779.

[151] Zott, K.; Claisse, O.; Lucas, P.; Coulon, J.; Lonvaud-Funel, A.; Masneuf-Pomarede, I. Characterization of the yeast ecosystem in grape must and wine using real-time PCR. *Food microbiology* **2010**, *27*, 559–567.

[152] Díaz, C.; Molina, A.M.; Nähring, J.; Fischer, R. Characterization and dynamic behavior of wild yeast during spontaneous wine fermentation in steel tanks and amphorae. *BioMed research international* **2013**, *2013*.

[153] Suarez, M.B.; Walsh, K.; Boonham, N.; ONeill, T.; Pearson, S.; Barker, I. Development of real-time PCR (TaqMan®) assays for the detection and quantification of Botrytis cinerea in planta. *Plant Physiology and Biochemistry* **2005**, *43*, 890–899.

[154] González, Á.; Hierro, N.; Poblet, M.; Mas, A.; Guillamón, J.M. Enumeration and detection of acetic acid bacteria by real-time PCR and nested PCR. *FEMS Microbiology Letters* **2006**, *254*, 123–128.

[155] Neeley, E.T.; Phister, T.G.; Mills, D.A. Differential real-time PCR assay for enumeration of lactic acid bacteria in wine. *Appl. Environ. Microbiol.* **2005**, *71*, 8954–8957.

[156] Meroth, C.B.; Hammes, W.P.; Hertel, C. Identification and population dynamics of yeasts in sourdough fermentation processes by PCR-denaturing gradient gel electrophoresis. *Appl. Environ. Microbiol.* **2003**, *69*, 7453–7461.

[157] Muyzer. G, de Waal EC, Uitterlinden A G. Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction amplified genes coding for 16S rRNA. *Applied and Environmental Microbiology* **1993**, *59*, 695.

[158] Caporaso, J.G.; Lauber, C.L.; Walters, W.A.; Berg-Lyons, D.; Lozupone, C.A.; Turnbaugh, P.J.; Fierer, N.; Knight, R. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the national academy of sciences* **2011**, *108*, 4516–4522.

[159] Prevost-Boure, N.C.; Christen, R.; Dequiedt, S.; Mougel, C.; Lelievre, M.; Jolivet, C.; Shahbazkia, H.R.; Guillou, L.; Arrouays, D.; Ranjard, L. Validation and application of a PCR primer set to quantify fungal communities in the soil environment by real-time quantitative PCR. *PloS one* **2011**, *6*, e24166.

[160] Lozupone, C.; Knight, R. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* **2005**, *71*, 8228–8235.

[161] Ciani, M.; Beco, L.; Comitini, F. Fermentation behaviour and metabolic interactions of multistarter wine yeast fermentations. *International journal of food microbiology* **2006**, *108*, 239–245.

[162] Magyar, I.; Tóth, T. Comparative evaluation of some oenological properties in wine strains of Candida stellata, Candida zemplinina, Saccharomyces uvarum and Saccharomyces cerevisiae. *Food Microbiology* **2011**, *28*, 94–100.

[163] Andorrà, I.; Landi, S.; Mas, A.; Guillamón, J.M.; Esteve-Zarzoso, B. Effect of oenological practices on microbial populations using culture-independent techniques. *Food Microbiology* **2008**, *25*, 849–856.

[164] Llauradó, J.; Rozes, N.; Bobet, R.; Mas, A.; Constantí, M. Low temperature alcoholic fermentations in high sugar concentration grape musts. *Journal of Food Science* **2002**, *67*, 268–273.

[165] Muyzer, G.; Smalla, K. Application of denaturing gradient gel electrophoresis (DGGE) and temperature gradient gel electrophoresis (TGGE) in microbial ecology. *Antonie van Leeuwenhoek* **1998**, *73*, 127–141.

[166] Wang, C.; García-Fernández, D.; Mas, A.; Esteve-Zarzoso, B. Fungal diversity in grape must and wine fermentation assessed by massive sequencing, quantitative PCR and DGGE. *Frontiers in microbiology* **2015**, *6*, 1156.

[167] Barbe, J.C.; De Revel, G.; Joyeux, A.; Bertrand, A.; Lonvaud-Funel, A. Role of botrytized grape micro-organisms in SO2 binding phenomena. *Journal of Applied Microbiology* **2001**, *90*, 34–42.

[168] Ciani, M.; Comitini, F. Yeast interactions in multi-starter wine fermentation. *Current Opinion in Food Science* **2015**, *1*, 1–6.

[169] Wang, C.; Mas, A.; Esteve-Zarzoso, B. The interaction between Saccharomyces cerevisiae and non-Saccharomyces yeast during alcoholic fermentation is species and strain specific. *Frontiers in microbiology* **2016**, *7*, 502.

[170] Mills, D.A.; Johannsen, E.A.; Cocolin, L. Yeast diversity and persistence in botrytis-affected wine fermentations. *Appl. Environ. Microbiol.* **2002**, *68*, 4884–4893.

[171] Constantí, M.; Poblet, M.; Arola, L.; Mas, A.; Guillamón, J.M. Analysis of yeast populations during alcoholic fermentation in a newly established winery. *American Journal of Enology and Viticulture* **1997**, *48*, 339–344.

[172] La Guerche, S.; Dauphin, B.; Pons, M.; Blancard, D.; Darriet, P. Characterization of some mushroom and earthy off-odors microbially induced by the development of rot on grapes. *Journal of agricultural and food chemistry* **2006**, *54*, 9193–9200.

[173] Padilla, B.; García-Fernández, D.; González, B.; Izidoro, I.; Esteve-Zarzoso, B.; Beltran, G.; Mas, A. Yeast biodiversity from DOQ Priorat uninoculated fermentations. *Frontiers in microbiology* **2016**, *7*, 930.

[174] Bartowsky, E.J.; Henschke, P.A. Acetic acid bacteria spoilage of bottled red winea review. *International journal of food microbiology* **2008**, *125*, 60–70.

[175] Joyeux, A.; Lafon-Lafourcade, S.; Ribéreau-Gayon, P. Metabolism of acetic acid bacteria in grape must. Consequences on alcoholic and malolactic fermentation. *Sciences des Aliments* **1984**, *4*, 247–255.

[176] Bokulich, N.A.; Bergsveinson, J.; Ziola, B.; Mills, D.A. Mapping microbial ecosystems and spoilage-gene flow in breweries highlights patterns of contamination and resistance. *Elife* **2015**, *4*, e04634.

[177] Bolyen, E.; Rideout, J.R.; Dillon, M.R.; Bokulich, N.A.; Abnet, C.; Al-Ghalith, G.A.; Alexander, H.; Alm, E.J.; Arumugam, M.; Asnicar, F.; others. QIIME 2: Reproducible, interactive, scalable, and extensible microbiome data science. Technical report, PeerJ Preprints, 2018.

[178] Bokulich, N.A.; Subramanian, S.; Faith, J.J.; Gevers, D.; Gordon, J.I.; Knight, R.; Mills, D.A.; Caporaso, J.G. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nature methods* **2013**, *10*, 57.

[179] Yuan, S.; Cohen, D.B.; Ravel, J.; Abdo, Z.; Forney, L.J. Evaluation of methods for the extraction and purification of DNA from the human microbiome. *PloS one* **2012**, *7*, e33865.

[180] Yeh, Y.C.; Needham, D.M.; Sieradzki, E.T.; Fuhrman, J.A. Taxon Disappearance from Microbiome Analysis Reinforces the Value of Mock Communities as a Standard in Every Sequencing Run. *MSystems* **2018**, *3*, e00023–18.

[181] Lleixà, J.; Kioroglou, D.; Mas, A.; del Carmen Portillo, M. Microbiome dynamics during spontaneous fermentations of sound grapes in comparison with sour rot and Botrytis infected grapes. *International journal of food microbiology* **2018**, *281*, 36–46.

[182] Herlemann, D.P.; Labrenz, M.; Jürgens, K.; Bertilsson, S.; Waniek, J.J.; Andersson, A.F. Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. *The ISME journal* **2011**, *5*, 1571.

[183] White, T.J.; Bruns, T.; Lee, S.; Taylor, J.; others. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR protocols: a guide to methods and applications* **1990**, *18*, 315–322.

[184] Andrews, S.; others. FastQC: a quality control tool for high throughput sequence data **2010**.

[185] Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120.

[186] Magoč, T.; Salzberg, S.L. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **2011**, *27*, 2957–2963.

[187] Callahan, B.J.; McMurdie, P.J.; Rosen, M.J.; Han, A.W.; Johnson, A.J.A.; Holmes, S.P. DADA2: high-resolution sample inference from Illumina amplicon data. *Nature methods* **2016**, *13*, 581.

[188] Rognes, T.; Flouri, T.; Nichols, B.; Quince, C.; Mahé, F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **2016**, *4*, e2584.

[189] Rideout, J.R.; He, Y.; Navas-Molina, J.A.; Walters, W.A.; Ursell, L.K.; Gibbons, S.M.; Chase, J.; McDonald, D.; Gonzalez, A.; Robbins-Pianka, A.; others. Subsampled open-reference clustering creates consistent, comprehensive OTU definitions and scales to billions of sequences. *PeerJ* **2014**, *2*, e545.

[190] Giordano, C.; Spennati, F.; Mori, G.; Munz, G.; Vannini, C. The microbial community in a moving bed biotrickling filter operated to remove hydrogen sulfide from gas streams. *Systematic and applied microbiology* **2018**.

[191] Nilsson, R.H.; Larsson, K.H.; Taylor, A.F.S.; Bengtsson-Palme, J.; Jeppesen, T.S.; Schigel, D.; Kennedy, P.; Picard, K.; Glöckner, F.O.; Tedersoo, L.; others. The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications. *Nucleic acids research* **2018**, *47*, D259–D264.

[192] Sentausa, E.; Fournier, P.E. Advantages and limitations of genomics in prokaryotic taxonomy. *Clinical Microbiology and Infection* **2013**, *19*, 790–795.

[193] Trtkova, J.; Raclavsky, V.; others. Molecular-genetic approaches to identification and typing of pathogenic Candida yeasts. *BIOMEDICAL PAPERS-PALACKY UNIVERSITY IN OLOMOUC* **2006**, *150*, 51.

[194] Vacher, C.; Hampe, A.; Porté, A.J.; Sauer, U.; Compant, S.; Morris, C.E. The phyllosphere: microbial jungle at the plant–climate interface. *Annual review of ecology, evolution, and systematics* **2016**, *47*, 1–24.

[195] Belda, I.; Zarraonaindia, I.; Perisin, M.; Palacios, A.; Acedo, A. From vineyard soil to wine fermentation: microbiome approximations to explain the terroir concept. *Frontiers in microbiology* **2017**, *8*, 821.

[196] Pérez-Magariño, S.; González-San José, M.L. Polyphenols and colour variability of red wines made from grapes harvested at different ripeness grade. *Food chemistry* **2006**, *96*, 197–208.

[197] Tsilimigras, M.C.; Fodor, A.A. Compositional data analysis of the microbiome: fundamentals, tools, and challenges. *Annals of epidemiology* **2016**, *26*, 330–335.

[198] Weiss, S.; Xu, Z.Z.; Peddada, S.; Amir, A.; Bittinger, K.; Gonzalez, A.; Lozupone, C.; Zaneveld, J.R.; Vázquez-Baeza, Y.; Birmingham, A.; others. Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome* **2017**, *5*, 27.

[199] Bolyen, E.; Rideout, J.R.; Dillon, M.R.; Bokulich, N.A.; Abnet, C.C.; Al-Ghalith, G.A.; Alexander, H.; Alm, E.J.; Arumugam, M.; Asnicar, F.; others. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature biotechnology* **2019**, *37*, 852–857.

[200] Tonietto, J.; Carbonneau, A. A multicriteria climatic classification system for grape-growing regions worldwide. *Agricultural and Forest Meteorology* **2004**, *124*, 81–97.

[201] Zhang, J.; Kobert, K.; Flouri, T.; Stamatakis, A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **2014**, *30*, 614–620.

[202] Edgar, R.C. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature methods* **2013**, *10*, 996.

[203] Kõljalg, U.; Larsson, K.H.; Abarenkov, K.; Nilsson, R.H.; Alexander, I.J.; Eberhardt, U.; Erland, S.; Høiland, K.; Kjøller, R.; Larsson, E.; others. UNITE: a database providing web-based methods for the molecular identification of ectomycorrhizal fungi. *New Phytologist* **2005**, *166*, 1063–1068.

[204] Anderson, M.J. A new method for non-parametric multivariate analysis of variance. *Austral ecology* **2001**, *26*, 32–46.

[205] Seabold, S.; Perktold, J. Statsmodels: Econometric and statistical modeling with python. Proceedings of the 9th Python in Science Conference. Scipy, 2010, Vol. 57, p. 61.

[206] Jones, E.; Oliphant, T.; Peterson, P.; others. SciPy: Open source scientific tools for Python, 2001.

[207] McKinney, W.; others. Data structures for statistical computing in python. Proceedings of the 9th Python in Science Conference. Austin, TX, 2010, Vol. 445, pp. 51–56.

[208] Čadež, N.; Zupan, J.; Raspor, P. The effect of fungicides on yeast communities associated with grape berries. *FEMS Yeast Research* **2010**, *10*, 619–630.

[209] Šuklje, K.; Carlin, S.; Antalick, G.; Blackman, J.W.; Deloire, A.; Vrhovsek, U.; Schmidtke, L.M. Regional discrimination of Australian Shiraz wine volatome by two-dimensional gas chromatography coupled to Time-of-Flight mass spectrometry. *Journal of agricultural and food chemistry* **2019**, *67*, 10273–10284.

[210] Šuklje, K.; Carlin, S.; Stanstrup, J.; Antalick, G.; Blackman, J.W.; Meeks, C.; Deloire, A.; Schmidtke, L.M.; Vrhovsek, U. Unravelling wine volatile evolution during Shiraz grape ripening by untargeted HS-SPME-GC× GC-TOFMS. *Food chemistry* **2019**, *277*, 753–765.

[211] Young, P.R.; Eyeghe-Bickong, H.A.; du Plessis, K.; Alexandersson, E.; Jacobson, D.A.; Coetzee, Z.; Deloire, A.; Vivier, M.A. Grapevine plasticity in response to an altered

microclimate: Sauvignon Blanc modulates specific metabolites in response to increased berry exposure. *Plant physiology* **2016**, *170*, 1235–1254.

[212] Gao, F.; Chen, J.; Xiao, J.; Cheng, W.; Zheng, X.; Wang, B.; Shi, X. Microbial community composition on grape surface controlled by geographical factors of different wine regions in Xinjiang, China. *Food research international* **2019**, *122*, 348–360.

[213] Ding, S.; Li, N.; Cao, M.; Huang, Q.; Chen, G.; Xie, S.; Zhang, J.; Cheng, G.; Li, W. Diversity of epiphytic fungi on the surface of Kyoho grape berries during ripening process in summer and winter at Nanning region, Guangxi, China. *Fungal biology* **2019**, *123*, 283–289.

[214] Dissanayake, A.J.; Purahong, W.; Wubet, T.; Hyde, K.D.; Zhang, W.; Xu, H.; Zhang, G.; Fu, C.; Liu, M.; Xing, Q.; others. Direct comparison of culture-dependent and culture-independent molecular approaches reveal the diversity of fungal endophytic communities in stems of grapevine (Vitis vinifera). *Fungal Diversity* **2018**, *90*, 85–107.

[215] Moreno-Arribas, M.V.; Polo, M.C. Occurrence of lactic acid bacteria and biogenic amines in biologically aged wines. *Food microbiology* **2008**, *25*, 875–881.

[216] Ortega-Heras, M.; González-Sanjosé, M.L.; González-Huerta, C. Consideration of the influence of aging process, type of wine and oenological classic parameters on the levels of wood volatile compounds present in red wines. *Food chemistry* **2007**, *103*, 1434–1448.

[217] Cadahía, E.; de Simón, B.F.; Sanz, M.; Poveda, P.; Colio, J. Chemical and chromatic characteristics of Tempranillo, Cabernet Sauvignon and Merlot wines from DO Navarra aged in Spanish and French oak barrels. *Food chemistry* **2009**, *115*, 639–649.

[218] Garde-Cerdán, T.; Ancín-Azpilicueta, C. Review of quality factors on wine ageing in oak barrels. *Trends in Food Science & Technology* **2006**, *17*, 438–447.

[219] Gómez-Plaza, E.; Cano-López, M. A review on micro-oxygenation of red wines: Claims, benefits and the underlying chemistry. *Food Chemistry* **2011**, *125*, 1131–1140.

[220] Spillman, P.J.; Pollnitz, A.P.; Liacopoulos, D.; Skouroumounis, G.K.; Sefton, M.A. Accumulation of vanillin during barrel-aging of white, red, and model wines. *Journal of Agricultural and Food Chemistry* **1997**, *45*, 2584–2589.

[221] Du Toit, W.; Pretorius, I.; Lonvaud-Funel, A. The effect of sulphur dioxide and oxygen on the viability and culturability of a strain of Acetobacter pasteurianus and a strain of Brettanomyces bruxellensis isolated from wine. *Journal of Applied Microbiology* **2005**, *98*, 862–871.

[222] Capozzi, V.; Di Toro, M.R.; Grieco, F.; Michelotti, V.; Salma, M.; Lamontanara, A.; Russo, P.; Orrù, L.; Alexandre, H.; Spano, G. Viable But Not Culturable (VBNC) state of Brettanomyces bruxellensis in wine: New insights on molecular basis of VBNC behaviour using a transcriptomic approach. *Food microbiology* **2016**, *59*, 196–204.

[223] Wedral, D.; Shewfelt, R.; Frank, J. The challenge of Brettanomyces in wine LWT Food Sci Technol, 43. *COI* **2010**, *1*, 1474–1479.

[224] Guzzon, R.; Bernard, M.; Barnaba, C.; Bertoldi, D.; Pixner, K.; Larcher, R. The impact of different barrel sanitation approaches on the spoilage microflora and phenols composition of wine. *Journal of food science and technology* **2017**, *54*, 810–821.

[225] Izquierdo-Cañas, P.; López-Martín, R.; García-Romero, E.; González-Arenzana, L.; Mínguez-Sanz, S.; Chatonnet, P.; Palacios-García, A.; Puig-Pujol, A. Effect of kaolin silver complex on the control of populations of Brettanomyces and acetic acid bacteria in wine. *Journal of food science and technology* **2018**, *55*, 1823–1831.

[226] Martorell, P.; Barata, A.; Malfeito-Ferreira, M.; Fernández-Espinar, M.T.; Loureiro, V.; Querol, A. Molecular typing of the yeast species Dekkera bruxellensis and Pichia guilliermondii recovered from wine related sources. *International journal of food microbiology* **2006**, *106*, 79–84.

[227] Nisiotou, A.; Gibson, G. Isolation of culturable yeasts from market wines and evaluation of the 5· 8S-ITS rDNA sequence analysis for identification purposes. *Letters in applied microbiology* **2005**, *41*, 454–463.

[228] Renouf, V.; Claisse, O.; Lonvaud-Funel, A. Inventory and monitoring of wine microbial consortia. *Applied Microbiology and Biotechnology* **2007**, *75*, 149–164.

[229] Andorrà, I.; Esteve-Zarzoso, B.; Guillamón, J.M.; Mas, A. Determination of viable wine yeast using DNA binding dyes and quantitative PCR. *International journal of food microbiology* **2010**, *144*, 257–262.

[230] Tofalo, R.; Schirone, M.; Corsetti, A.; Suzzi, G. Detection of Brettanomyces spp. in red wines using real-time PCR. *Journal of food science* **2012**, *77*, M545–M549.

[231] Marzano, M.; Fosso, B.; Manzari, C.; Grieco, F.; Intranuovo, M.; Cozzi, G.; Mulè, G.; Scioscia, G.; Valiente, G.; Tullo, A.; others. Complexity and dynamics of the winemaking bacterial communities in berries, musts, and wines from Apulian grape cultivars through time and space. *PloS one* **2016**, *11*, e0157383.

[232] Mezzasalma, V.; Sandionigi, A.; Guzzetti, L.; Galimberti, A.; Grando, M.S.; Tardaguila, J.; Labra, M. Geographical and cultivar features differentiate grape microbiota in Northern Italy and Spain vineyards. *Frontiers in microbiology* **2018**, *9*, 946.

[233] Piao, H.; Hawley, E.; Kopf, S.; DeScenzo, R.; Sealock, S.; Henick-Kling, T.; Hess, M. Insights into the bacterial community and its temporal succession during the fermentation of wine grapes. *Frontiers in microbiology* **2015**, *6*, 809.

[234] Salvetti, E.; Campanaro, S.; Campedelli, I.; Fracchetti, F.; Gobbi, A.; Tornielli, G.B.; Torriani, S.; Felis, G.E. Whole-metagenome-sequencing-based community profiles of Vitis vinifera L. cv. Corvina berries withered in two post-harvest conditions. *Frontiers in microbiology* **2016**, *7*, 937.

[235] Sirén, K.; Mak, S.S.T.; Melkonian, C.; Carøe, C.; Swiegers, J.H.; Molenaar, D.; Fischer, U.; Gilbert, M.T.P. Taxonomic and functional characterization of the microbial community during spontaneous in vitro fermentation of riesling must. *Frontiers in microbiology* **2019**, *10*, 697.

[236] Arapitsas, P.; Guella, G.; Mattivi, F. The impact of SO 2 on wine flavanols and indoles in relation to wine style and age. *Scientific reports* **2018**, *8*, 1–13.

[237] Azevedo, J.; Fernandes, I.; Lopes, P.; Roseira, I.; Cabral, M.; Mateus, N.; Freitas, V. Migration of phenolic compounds from different cork stoppers to wine model solutions: antioxidant and biological relevance. *European Food Research and Technology* **2014**, *239*, 951–960.

[238] Kioroglou, D.; Mas, A.; Portillo, M.d.C. Evaluating the Effect of QIIME Balanced Default Parameters on Metataxonomic Analysis Workflows With a Mock Community. *Frontiers in microbiology* **2019**, *10*, 1084.

[239] Phister, T.G.; Mills, D.A. Real-time PCR assay for detection and enumeration of Dekkera bruxellensis in wine. *Applied and Environmental Microbiology* **2003**, *69*, 7430–7434.

[240] Bokulich, N.; Collins, T.; Masarweh, C.; Allen, G.; Heymann, H.; Ebeler, S.; others. Fermentation behavior suggest microbial contribution to regional wine characteristics. *MBio* **2016**, *7*, 1–12.

[241] Navarro, D.; Mateo, E.; Torija, M.J.; Mas, A. Acetic acid bacteria in grape must. *Acetic Acid Bacteria* **2013**, *2*.

[242] Aiken, J.W.; Noble, A. Comparison of the aromas of oak-and glass-aged wines. *American journal of enology and viticulture* **1984**, *35*, 196–199.

[243] Rubio, P.; Garijo, P.; Santamaría, P.; López, R.; Martínez, J.; Gutiérrez, A. Influence of oak origin and ageing conditions on wine spoilage by Brettanomyces yeasts. *Food Control* **2015**, *54*, 176–180.

[244] Grangeteau, C.; Roullier-Gall, C.; Rousseaux, S.; Gougeon, R.D.; Schmitt-Kopplin, P.; Alexandre, H.; Guilloux-Benatier, M. Wine microbiology is driven by vineyard and winery anthropogenic factors. *Microbial biotechnology* **2017**, *10*, 354–370.

[245] Takahashi, M.; Masaki, K.; Mizuno, A.; Goto-Yamamoto, N. Modified COLD-PCR for detection of minor microorganisms in wine samples during the fermentation. *Food microbiology* **2014**, *39*, 74–80.

[246] Arbulu, M.; Sampedro, M.; Gómez-Caballero, A.; Goicolea, M.; Barrio, R. Untargeted metabolomic analysis using liquid chromatography quadrupole time-of-flight mass spectrometry for non-volatile profiling of wines. *Analytica Chimica Acta* **2015**, *858*, 32–41.

[247] Rossouw, D.; Bauer, F.F. Comparing the transcriptomes of wine yeast strains: toward understanding the interaction between environment and transcriptome during fermentation. *Applied microbiology and biotechnology* **2009**, *84*, 937.

[248] Fotakis, C.; Kokkotou, K.; Zoumpoulakis, P.; Zervou, M. NMR metabolite fingerprinting in grape derived products: An overview. *Food research international* **2013**, *54*, 1184–1194.

[249] Arapitsas, P.; Speri, G.; Angeli, A.; Perenzoni, D.; Mattivi, F. The influence of storage on the chemical age of red wines. *Metabolomics* **2014**, *10*, 816–832.

[250] Fernández de Simón, B.; Cadahía, E.; Jalocha, J. Volatile compounds in a Spanish red wine aged in barrels made of Spanish, French, and American oak wood. *Journal of agricultural and food chemistry* **2003**, *51*, 7671–7678.

[251] Belda, I.; Ruiz, J.; Esteban-Fernández, A.; Navascués, E.; Marquina, D.; Santos, A.; Moreno-Arribas, M. Microbial contribution to wine aroma and its intended use for wine quality improvement. *Molecules* **2017**, *22*, 189.

[252] Consonni, R.; Cagliani, L.R.; Guantieri, V.; Simonato, B. Identification of metabolic content of selected Amarone wine. *Food chemistry* **2011**, *129*, 693–699.

[253] Lloyd, N.; Johnson, D.; Herderich, M. Metabolomics approaches for resolving and harnessing chemical diversity in grapes, yeast and wine. *Australian journal of grape and wine research* **2015**, *21*, 723–740.

[254] Nakagawa, S.; Cuthill, I.C. Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biological reviews* **2007**, *82*, 591–605.

[255] Jewison, T.; Knox, C.; Neveu, V.; Djoumbou, Y.; Guo, A.C.; Lee, J.; Liu, P.; Mandal, R.; Krishnamurthy, R.; Sinelnikov, I.; others. YMDB: the yeast metabolome database. *Nucleic acids research* **2012**, *40*, D815–D820.

[256] Wishart, D.S.; Tzur, D.; Knox, C.; Eisner, R.; Guo, A.C.; Young, N.; Cheng, D.; Jewell, K.; Arndt, D.; Sawhney, S.; others. HMDB: the human metabolome database. *Nucleic acids research* **2007**, *35*, D521–D526.

[257] Mock, A.; Warta, R.; Dettling, S.; Brors, B.; Jäger, D.; Herold-Mende, C. MetaboDiff: an R package for differential metabolomic analysis. *Bioinformatics* **2018**, *34*, 3417–3418.

[258] scikit-bio development team, T. scikit-bio: A Bioinformatics Library for Data Scientists, Students, and Developers, 2020.

[259] Ali, K.; Maltese, F.; Toepfer, R.; Choi, Y.H.; Verpoorte, R. Metabolic characterization of Palatinate German white wines according to sensory attributes, varieties, and vintages using NMR spectroscopy and multivariate data analyses. *Journal of Biomolecular NMR* **2011**, *49*, 255–266.

[260] Cassino, C.; Tsolakis, C.; Bonello, F.; Gianotti, V.; Osella, D. Effects of area, year and climatic factors on Barbera wine characteristics studied by the combination of 1H-NMR metabolomics and chemometrics. *Journal of wine research* **2017**, *28*, 259–277.

[261] Choi, H.K.; Yoon, J.H.; Kim, Y.S.; Kwon, D.Y. Metabolomic profiling of Cheongguk-jang during fermentation by 1H NMR spectrometry and principal components analysis. *Process Biochemistry* **2007**, *42*, 263–266.

[262] Pérez-Prieto, L.J.; López-Roca, J.M.; Martínez-Cutillas, A.; Pardo-Mínguez, F.; Gómez-Plaza, E. Extraction and formation dynamic of oak-related volatile compounds from different volume barrels to wine and their behavior during bottle storage. *Journal of Agricultural and Food Chemistry* **2003**, *51*, 5444–5449.

[263] Cerdán, T.G.; Mozaz, S.R.; Azpilicueta, C.A. Volatile composition of aged wine in used barrels of French oak and of American oak. *Food Research International* **2002**, *35*, 603–610.

[264] Zhang, Z.; Li, J.; Fan, L. Evaluation of the composition of Chinese bayberry wine and its effects on the color changes during storage. *Food chemistry* **2019**, *276*, 451–457.

[265] Johnson, J.; Maher, P.; Hanneken, A. The flavonoid, eriodictyol, induces long-term protection in ARPE-19 cells through its effects on Nrf2 activation and phase 2 gene expression. *Investigative ophthalmology & visual science* **2009**, *50*, 2398–2406.

[266] Kim, A.R.; Zou, Y.N.; Park, T.H.; Shim, K.H.; Kim, M.S.; Kim, N.D.; Kim, J.D.; Bae, S.J.; Choi, J.S.; Chung, H.Y. Active components from Artemisia iwayomogi displaying ONOO- scavenging activity. *Phytotherapy Research: An International Journal Devoted to Pharmacological and Toxicological Evaluation of Natural Product Derivatives* **2004**, *18*, 1–7.

[267] Ai-li, J.; Chang-hai, W. Antioxidant properties of natural components from Salvia plebeia on oxidative stability of ascidian oil. *Process Biochemistry* **2006**, *41*, 1111–1116.

[268] Laqui-Estaña, J.; López-Solís, R.; Peña-Neira, Á.; Medel-Marabolí, M.; Obreque-Slier, E. Wines in contact with oak wood: the impact of the variety (Carménère and Cabernet Sauvignon), format (barrels, chips and staves), and aging time on the phenolic composition. *Journal of the Science of Food and Agriculture* **2019**, *99*, 436–448.

[269] Kalua, C.M.; Boss, P.K. Evolution of volatile compounds during the development of Cabernet Sauvignon grapes (Vitis vinifera L.). *Journal of agricultural and food chemistry* **2009**, *57*, 3818–3830.

[270] Alañón, M.; Castro-Vázquez, L.; Díaz-Maroto, M.; Gordon, M.; Pérez-Coello, M. A study of the antioxidant capacity of oak wood used in wine ageing and the correlation with polyphenol composition. *Food Chemistry* **2011**, *128*, 997–1002.

[271] Dombre, C.; Guillard, V.; Chalier, P. Protection of methionol against oxidation by oxygen scavenger: an experimental and modelling approach in wine model solution. *Food Packaging and Shelf Life* **2015**, *3*, 76–87.

[272] Herrero, M.; García, L.A.; Díaz, M. The effect of SO2 on the production of ethanol, acetaldehyde, organic acids, and flavor volatiles during industrial cider fermentation. *Journal of agricultural and food chemistry* **2003**, *51*, 3455–3459.

[273] Cao, Y.; Xie, G.; Wu, C.; Lu, J. A study on characteristic flavor compounds in traditional Chinese rice wineGuyue Longshan rice wine. *Journal of the Institute of Brewing* **2010**, *116*, 182–189.

[274] Tumanov, S.; Pinu, F.R.; Greenwood, D.; Villas-Bôas, S.G. Effect of free fatty acids and lipolysis on Sauvignon Blanc fermentation. *Australian journal of grape and wine research* **2018**, *24*, 398–405.

[275] Halwachs, B.; Madhusudhan, N.; Krause, R.; Nilsson, R.H.; Moissl-Eichinger, C.; Högenauer, C.; Thallinger, G.G.; Gorkiewicz, G. Critical issues in mycobiota analysis. *Frontiers in microbiology* **2017**, *8*, 180.

[276] Wong, R.G.; Wu, J.R.; Gloor, G.B. Expanding the UniFrac toolbox. *PloS one* **2016**, *11*, e0161196.

[277] Kruskal, W.H.; Wallis, W.A. Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association* **1952**, *47*, 583–621.

[278] Bau, M.; Bragulat, M.; Abarca, M.; Minguez, S.; Cabañes, F. Ochratoxigenic species from Spanish wine grapes. *International Journal of Food Microbiology* **2005**, *98*, 125–130.

[279] Bejaoui, H.; Mathieu, F.; Taillandier, P.; Lebrihi, A. Black aspergilli and ochratoxin A production in French vineyards. *International journal of food microbiology* **2006**, *111*, S46–S52.

[280] Serra, R.; Braga, A.; Venâncio, A. Mycotoxin-producing and other fungi isolated from grapes for wine production, with particular emphasis on ochratoxin A. *Research in Microbiology* **2005**, *156*, 515–521.

[281] Diguta, C.; Vincent, B.; Guilloux-Benatier, M.; Alexandre, H.; Rousseaux, S. PCR ITS-RFLP: a useful method for identifying filamentous fungi isolates on grapes. *Food Microbiology* **2011**, *28*, 1145–1154.

[282] Fang, F.; Li, J.M.; Pan, Q.H.; Huang, W.D. Determination of red wine flavonoids by HPLC and effect of aging. *Food Chemistry* **2007**, *101*, 428–433.

[283] Chatonnet, P.; Cutzach, I.; Pons, M.; Dubourdieu, D. Monitoring toasting intensity of barrels by chromatographic analysis of volatile compounds from toasted oak wood. *Journal of Agricultural and Food Chemistry* **1999**, *47*, 4310–4318.

[284] Prida, A.; Chatonnet, P. Impact of oak-derived compounds on the olfactory perception of barrel-aged wines. *American Journal of Enology and Viticulture* **2010**, *61*, 408–413.

[285] Yaddanapudi, L.N. The American Statistical Association statement on P-values explained. *Journal of anaesthesiology, clinical pharmacology* **2016**, *32*, 421.

[286] Grabowski, B. P$<$ 0.05 might not mean what you think: American Statistical Association clarifies P values. *JNCI: Journal of the National Cancer Institute* **2016**, *108*.

[287] Jara, C.; Mateo, E.; Guillamón, J.; Torija, M.; Mas, A. Analysis of several methods for the extraction of high quality DNA from acetic acid bacteria in wine and vinegar for characterization by PCR-based methods. *International journal of food microbiology* **2008**, *128*, 336–341.

[288] Beltramo, C.; Desroche, N.; Tourdot-Maréchal, R.; Grandvalet, C.; Guzzo, J. Real-time PCR for characterizing the stress response of Oenococcus oeni in a wine-like medium. *Research in Microbiology* **2006**, *157*, 267–274.

UNIVERSITAT ROVIRA I VIRGILI
ANALYSIS OF MICROBIAL POPULATIONS IN WINES THROUGH NGS METHODOLOGIES
Dimitrios Kioroglou

APPENDIX A

Research stay outline

The 3-months research stay was realized at the Technical University of Denmark and hosted in the research group of the associate professor Dr. Jose Maria Gonzalez Izarzugaza at the Center for Biological Sequence Analysis (CBS). The main focus of CBS is the research in the field of bioinformatics and systems biology, and the past years it has produced numerous computational methods that can be found in the following link:

http://www.cbs.dtu.dk/services/

The main activity of the stay concerned the RNASeq analysis of a dataset derived from a trascriptomics experiment. The design and implementation of this experiment was performed by Dr. Motlhalamme as part of her PhD thesis "***Characterization of melatonin production and physiological functions in yeast***".

Dr. Motlhalamme's PhD was conducted at the south African grape and wine research institute of the faculty of AgriSciences of the Stellenbosch university under the supervision of the professors Florian Bauer and Bernard Prior. Her PhD thesis can be found at:

https://scholar.sun.ac.za/handle/10019.1/107867

The RNASeq experiment and the derived dataset is attached to chapter 4 (pages 75-110), "***Melatonin modulates the transcriptional response of S. cerevisiae to hydrogen peroxide oxidative stress***", of Dr. Motlhalamme's dissertation.

The RNASeq analysis and results presented here in the Appendix A, represent an independent bioinformatic analysis of Dr. Motlhalamme's transcriptomic data and the derived results are not connected to the chapter 4 of Dr. Motlhalamme's dissertation

by any mean. The results included in the Appendix A have not been published and the whole RNASeq analysis was performed for purely educational purposes, with the aim of a potential similar implementation in this dissertation.

Experimental background

Introduction

The following passage has been taken from the chapter 4 of Dr. Motlhalamme's dissertation (pages 77-78):

> "*Several studies have documented the production of melatonin by S. cerevisiae and non-Saccharomyces yeast strains. However, the biosynthetic pathway and functions of this indoleamine in these microorganisms remain obscure. The protective role of melatonin against H_2O_2-induced oxidative stress and UV stress in yeast was recently described (Vzquez et al., 2017; Bisquert et al., 2018). In these batch culture studies, melatonin reduced reactive oxygen species (ROS) accumulation, enhanced the expression of endogenous antioxidant genes in H_2O_2 stressed S. cerevisiae and increased cell viability after exposure to UV and oxidative stress. However, the exact mechanisms behind the protective effects observed remain unclear. This study aimed to understand the transcriptional response of S. cerevisiae IWBT Y805 to exogenous melatonin treatment and the temporal response of melatonin treated cultures to H_2O_2-induced oxidative stress. Contrarily to previous studies, a steady state continuous culture set-up was used which reduced the changes to the physico-chemical conditions associated with batch culture systems. This constant environmental milieu focused the transcriptional response on melatonin-specific mechanisms. Experiments were conducted in nitrogen-limited continuous culture in order to control the cell growth rate and environmental conditions.*"

Methods

The following passage has been taken from the chapter 4 of Dr. Motlhalamme's dissertation (page 79):

> ***4.3.4 Hydrogen peroxide induced stress in continuous culture conditions***

"*Steady state cultures were stressed with 0.6 and 1 mM H_2O_2 in order to determine a sublethal concentration. H_2O_2 was injected aseptically into the bioreactor and sampling was conducted at 0, 15 min, 30 min, 60 min, and 120 min. Growth was monitored spectrophotometrically at 600 nm. To assess the impact of melatonin on oxidative stress, this indoleamine was injected into the feed reservoir to a final concentration of 300 μM. This concentration was shown to protect S. cerevisiae yeast strains against redox and oxidative stressors on plate assays reported in Chapter 5. In addition, a study by Tan et al. (2000) found that 300 μM melatonin was able to scavenge H_2O_2. The melatonin containing minimal medium was fed into the bioreactor by a previously calibrated peristaltic pump for one residence time before induction of H_2O_2 stress as described above. The bioreactor, feed reservoir and all tubes were covered in aluminium foil to limit exposure of melatonin to light.*"

References of passages

Bisquert, R., Muñiz-Calvo, S., and Guillamón, J. M. (2018). Protective role of intracellular melatonin against oxidative stress and UV radiation in Saccharomyces cerevisiae. Frontiers in Microbiology 9, 318. doi:10.3389/fmicb.2018.00318.

Tan, D. X., Manchester, L. C., Reiter, R. J., Plummer, B. F., Limson, J., Weintraub, S. T., Qi, W. (2000). Melatonin directly scavenges hydrogen peroxide: A potentially new metabolic pathway of melatonin biotransformation. Free Radical Biology and Medicine 29, 11771185. doi:10.1016/S0891-5849(00)00435-4.

Vázquez, J., González, B., Sempere, V., Mas, A., Torija, M. J., and Beltran, G. (2017). Melatonin reduces oxidative stress damage induced by hydrogen peroxide in Saccharomyces cerevisiae. Frontiers in Microbiology 8, 1066. doi:10.3389/fmicb.2017.01066

Research stay RNASeq analysis and results

Main objectives

- Identify differentially expressed (DE) genes among the experimental conditions.
- From the DE genes, identify genes with at least 2-fold over or under-expression due to melatonine compared to the H_20_2 unstressed and melatonine untreated samples. These genes are named **MDE** genes in this analysis.

- Perform enrichement analysis on the MDE genes in relation to the oxidative stress.

Experimental conditions and notation

4 conditions with 2 biological replicates per condition (a,b) and 4 technical replicates per biological replicate:

- 1st condition: Melatonine untreated - H_20_2 unstressed (denoted as 0_0)
  timepoint T0: a_0_0_t0 - b_0_0_t0
  timepoint 1 hour: a_0_0_1hr - b_0_0_1hr

- 2nd condition: Melatonine treated - H_20_2 unstressed (denoted as 1_0)
  timepoint T0: a_1_0_t0 - b_1_0_t0
  timepoint 1 hour: a_1_0_1hr - b_1_0_1hr

- 3rd condition: Melatonine untreated - H_20_2 stressed (denoted as 0_1)
  timepoint 15min: a_0_1_15min - b_0_1_15min
  timepoint 30min: a_0_1_30min - b_0_1_30min
  timepoint 1hr: a_0_1_1hr - b_0_1_1hr
  timepoint 2hr: a_0_1_2hr - b_0_1_2hr

- 4th condition: Melatonine treated - H_20_2 stressed (denoted as 1_1)
  timepoint 15min: a_1_1_15min - b_1_1_15min
  timepoint 30min: a_1_1_30min - b_1_1_30min
  timepoint 1hr: a_1_1_1hr - b_1_1_1hr
  timepoint 2hr: a_1_1_2hr - b_1_1_2hr

Data filtering and depth coverage

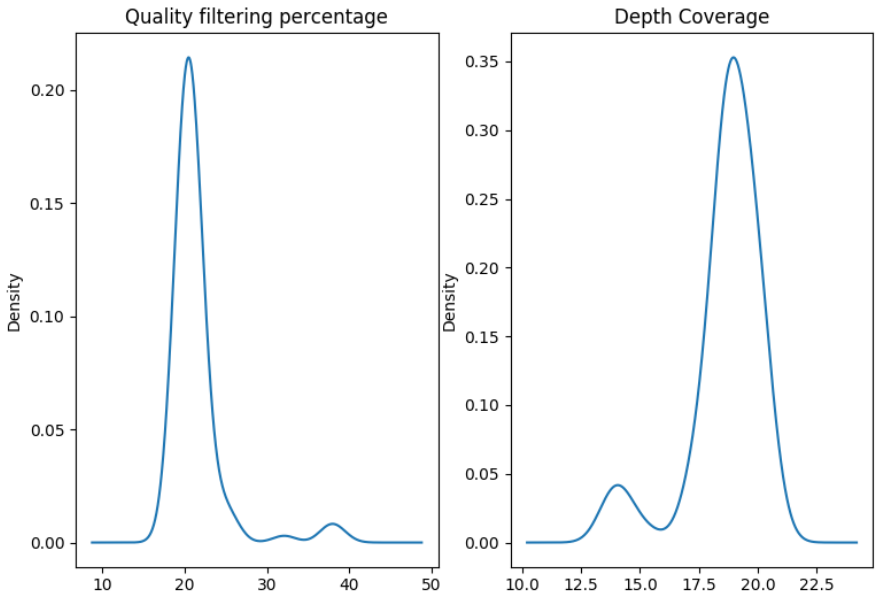


Figure S46: **Left plot:**The filtering of the raw fastq files was around 20%. **Right plot:** The depth after the mapping process was estimated around 20X.

Excluded Samples

The following samples have been excluded from the analysis because the biological replicates were not demonstrating strong correlation.

- a_1_0_1hr and b_1_0_1hr (Figure S47)
- a_1_1_1hr and b_1_1_1hr (Figure S48)

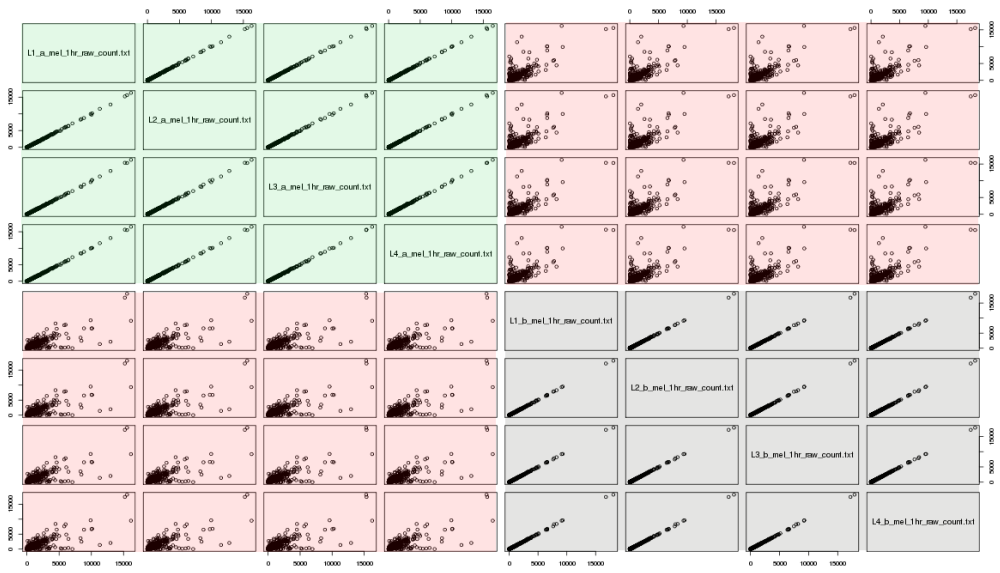


Figure S47: **Red plots:** Correlation plots between biological replicats. **Green plots:** Correlation plots between technical replicates for biological replicate **a**. **Grey plots:** Correlation plots between technical replicates for biological replicate **b**.

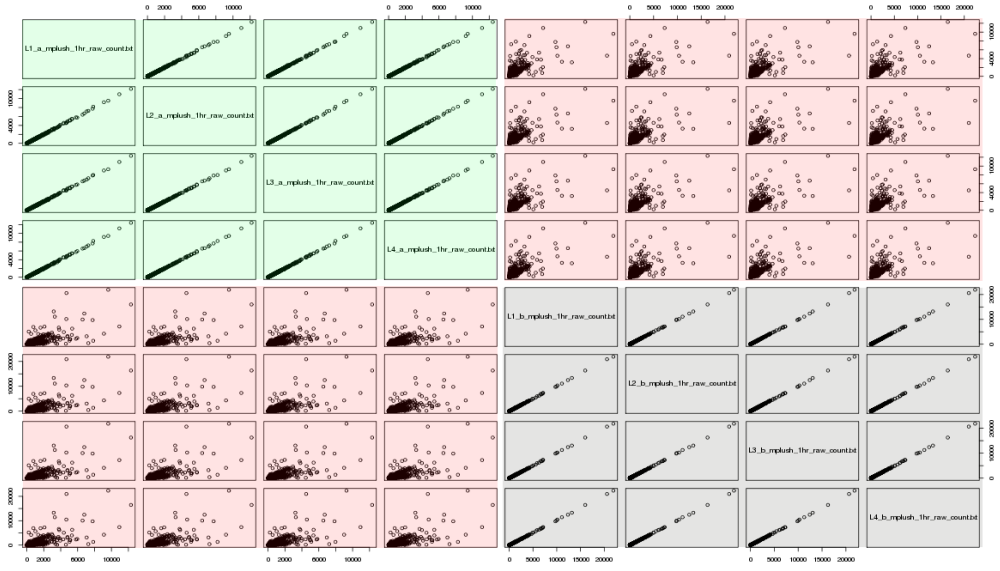


Figure S48: **Red plots:** Correlation plots between biological replicats. **Green plots:** Correlation plots between technical replicates for biological replicate **a**. **Grey plots:** Correlation plots between technical replicates for biological replicate **b**.

Diffential gene expression analysis

- Analysis performed with DESeq2 [289].
- **868** diffentially expressed genes were identified with p-value < 0.01 (p-values adjusted with Benjamini and Hochberg correction.)
- PCA was performed taking into consideration only the 868 differentially expressed genes. (Figure S49)

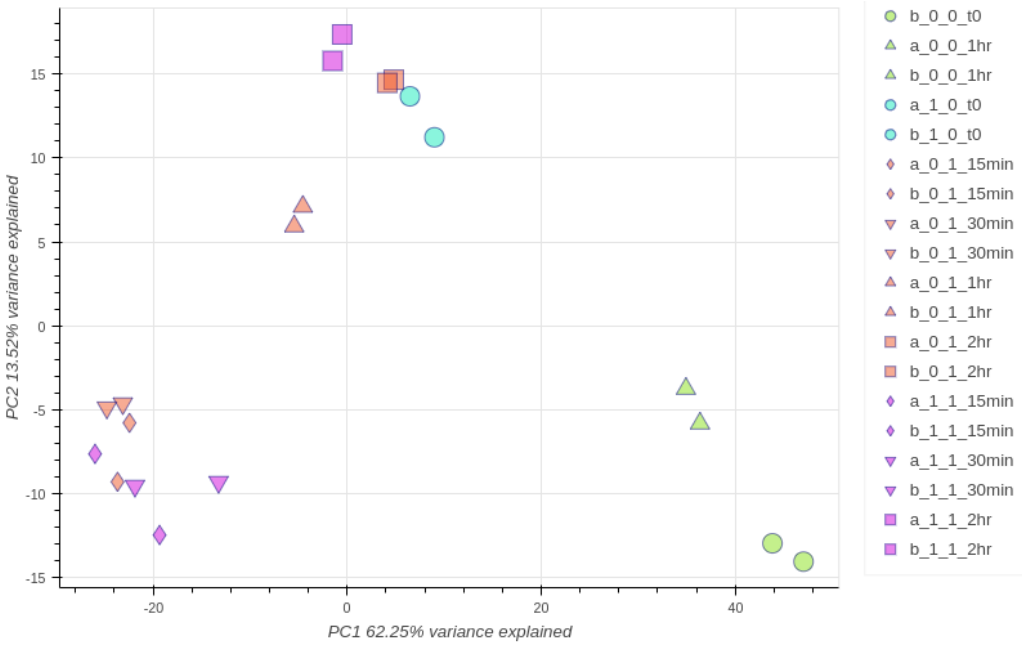


Figure S49: Principal component analysis capturing 75.77% of the observed variance and performed by taking into consideration only the 868 differentially expressed genes that were identified.

PCA remarks

- There is clear separation between the unstressed and H_20_2 stressed samples.
- From the H_20_2 stressed samples, the timepoints 15min and 30min clustered together whereas the 1hr timepoint is quite close to 2hr timepoint in both cases of metalotine treated and untreated samples.
- Overall, for the H_20_2 stressed samples, the melatonine treated and untreated samples exhibited the same pattern.

- Melatonine treated and unstressed samples clustered together with the H_20_2 stressed samples of 1 and 2 hours. This suggests that melatonine has a potential effect equal to that of a H_20_2 stressed samples after 1 or 2 hours.

Genes expression profile of 868 DE genes

Genes over-expressed due to melatonin

Without oxidative stress, melatonin caused an at least 2-fold over-expression at **56** genes. From these 56 genes, the biological replicates did not show the same behaviour of $\geq$2-fold over-expression for 26 genes.

In Figure S50 are the plots for the rest of the 30 genes for which the biological replicates showed $\geq$2-fold over-expression:

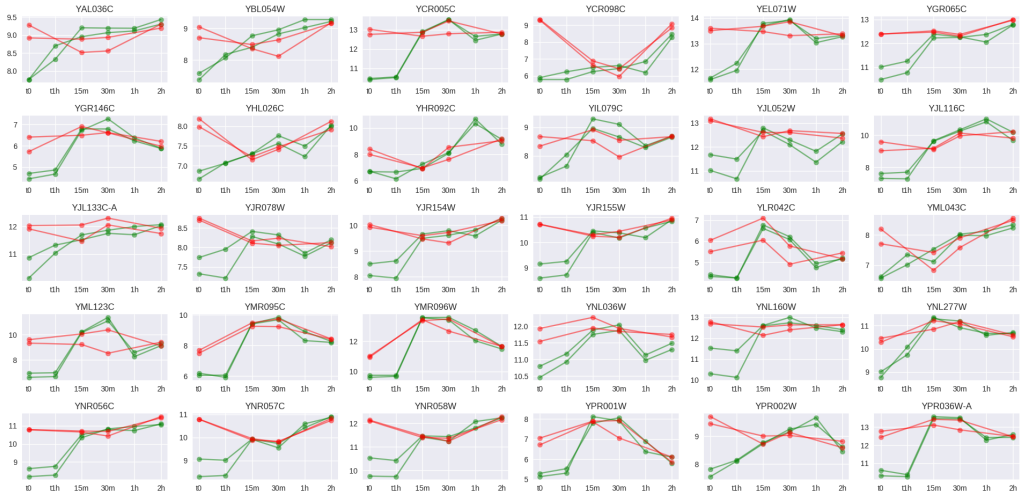


Figure S50: **Green** lines represent biological replicates (a,b) without melatonin. **Red** lines represent biological replicates (a,b) with melatonin. Time-points **t0** and **t1hr** indicate no oxidative stress. Expression levels are expressed in **log2** scale.

Genes under-expressed due to melatonin

Without oxidative stress, melatonin caused an at least 2-fold under-expression at **87** genes. From these 87 genes, the biological replicates did not show the same behaviour

of ≥2-fold under-expression for 47 genes.

Below are the plots (Figures S51, S52) for the rest of the 40 genes for which the biological replicates showed ≥2-fold under-expression:

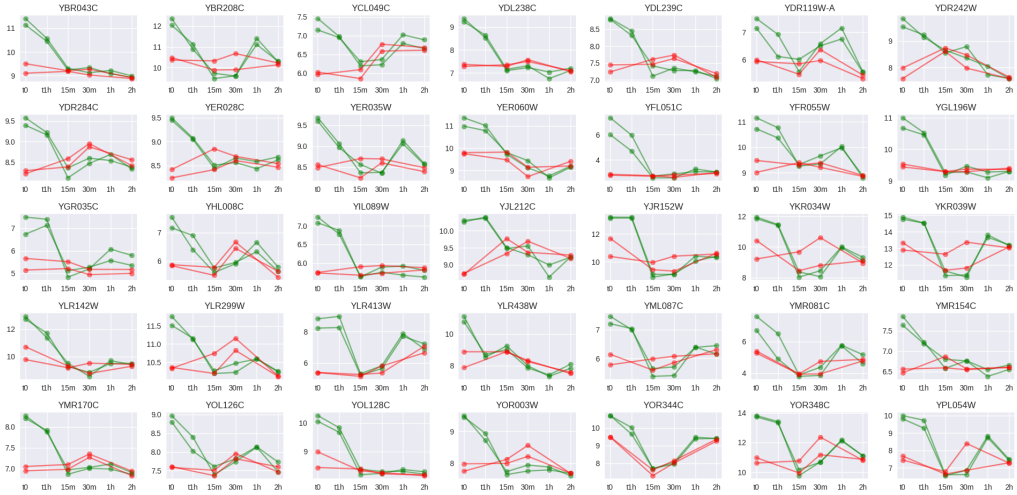


Figure S51: **Green** lines represent biological replicates (a,b) without melatonin. **Red** lines represent biological replicates (a,b) with melatonin. Time-points **t0** and **t1hr** indicate no oxidative stress. Expression levels are expressed in **log2** scale.

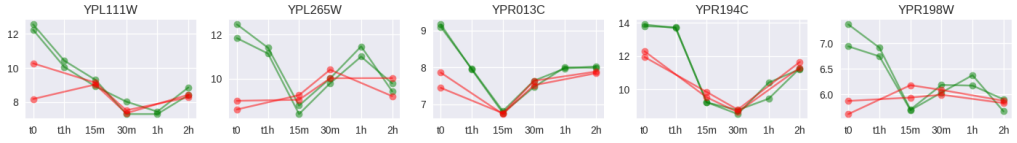


Figure S52: **Green** lines represent biological replicates (a,b) without melatonin. **Red** lines represent biological replicates (a,b) with melatonin. Time-points **t0** and **t1hr** indicate no oxidative stress. Expression levels are expressed in **log2** scale.

Genes unaffected by melatonin

Melatonine did not affect (that is <2-fold over or under-expression) the expression levels of **725** genes. In Figure S53 are the expression profiles of 35 randomly selected genes to demonstrate this.

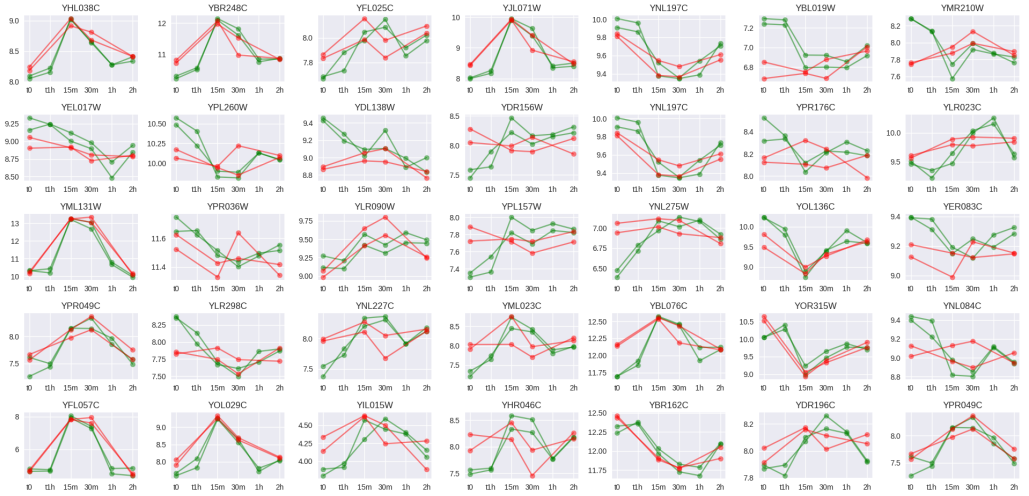


Figure S53: **Green** lines represent biological replicates (a,b) without melatonin. **Red** lines represent biological replicates (a,b) with melatonin. Time-points **t0** and **t1hr** indicate no oxidative stress. Expression levels are expressed in **log2** scale.

Gene profiling conclusions

Addition of melatonin have caused significant and similar changes to the expression levels as the oxidative stress for **143** out of the **868** DE genes (56 over-expressed + 87 under-expressed). More specifically, the expression levels caused by the oxidative stress at the time-point **2hr** are quite close to the expression levels caused by melatonin prior to stress. For that reason in the PCA plot the melatonin replicates at time-point **t0** clustered together with the ones at time-point **2hr**. Additionally in most cases of these 143 MDE genes, melatonin caused the expression levels to remain relatively stable, meaning that melatonin prepared the yeast for the stress. From energy stand-point, prior to stress the increase of energy demand for the over-expression of the 56 genes seems to have been counter-balanced by the under-expression of the 87 genes.

Enrichment analysis report of 143 MDE genes

Oxidative stress related genes not found in the 143 MDE genes

The following genes have been found to be significantly (p-value <0.01) differentially expressed due to oxidative stress and not melatonin:

- catalases: CTT1, CTA1
- peroxidases: CCP1. TSA1, TSA2

- reductases: MXR1, TRR1
- glutathione: GSH1, GSH2
- thioredoxin: TRX2
- glutaredoxins: GRX2
- superoxide dismutase: SOD2

Over-expressed genes due to melatonin

Melatonin has caused an over-expression of genes involved in the following pathways:

- **Biotin biosynthesis**: BIO3, BIO4, BIO5
  The following enzymes are biotin-dependent and related to oxidative stress as they are involved in production of oxaloacetate: PYC1, PYC2
  Both enzymes were found to be significantly differentially expressed after applying oxidative stress but not over-expressed by melatonin prior to stress.
- **Vitamin B6 biosynthesis**: SNZ1, SNZ3, SNO1
  Vitamin B6 has been identified as a potent antioxidant with a high ability to quench reactive oxygen species (ROS) [290].
- **Inhibition of mating type switching initiation**:
  Melatonin caused an over-expression of RPL22 which is required for the translation of the ASH1 gene [291] which in turn inhibits the transcription of HO gene.
  RPL22 is also required by IME1 which activates transcription of early meiotic genes.
  Perhaps these events serve as a way for *S. cerevisiae* to conserve energy.
- **Genes encoding GAPDH enzymes**: GAPDH enzymes participate in many reactions and are encoded by three genes: TDH1, TDH2, TDH3.
  Melatonin causes an over-expression of TDH1 gene which has been found of being over-expressed in low concentrations of H_20_2 [292].

Under-expressed genes due to melatonin

Melatonin has caused an under-expression of genes involved in the following pathways:

- **Glutathione degradation**: ECM38
- **Arginine metabolism**: CAR1
- **Nitrogen degradation**: DAL80
- **Allantoin degradation**: DAL1, DAL2
  Allantoin degradation pathway is shown in Figure S54.

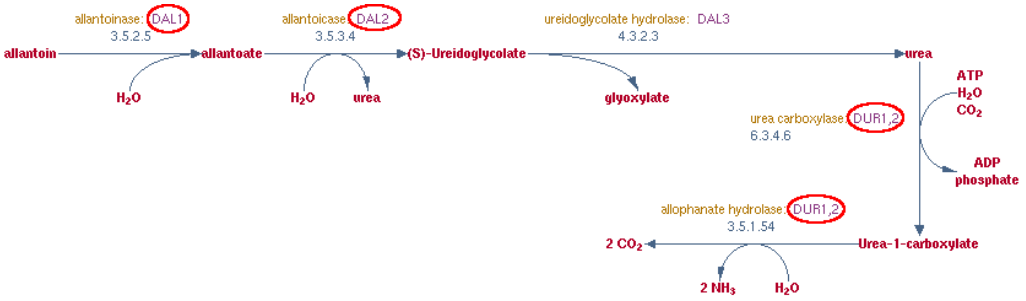


Figure S54: Allantoin degradation pathway. Enzymes in red circles represent under-expressed genes due to melatonin. The DUR1,2 enzyme is biotin-dependent.

Allantoin has been proposed as a biomarker of oxidative stress [293]. Moreover, allantoin participates in the bidirectional reaction of Figure S55.

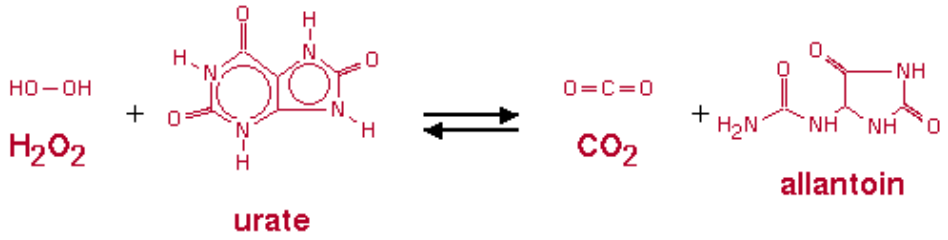


Figure S55: Allantoin, urate bidirectional reaction.

Therefore increase of allantoin's concentration may direct the reaction towards the production of urate which is an important antioxidant.

Conclusions from the enrichment analysis of MDE genes

It seems that melatonin does not have a direct effect on the genes that are directly related to oxidative stress. It rather has a synergetic role preparing and helping *S. cerevisiae* to cope with the oxidative stress.

REFERENCES APPENDIX A

[289] Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* **2014**, *15*, 550.

[290] Samsatly, J.; Chamoun, R.; Gluck-Thaler, E.; Jabaji, S. Genes of the de novo and Salvage Biosynthesis Pathways of Vitamin B6 are Regulated under Oxidative Stress in the Plant Pathogen Rhizoctonia solani. *Frontiers in microbiology* **2016**, *6*, 1429.

[291] Abrhámová, Katerina and Nemčko, Filip and Libus, Jiří and Převorovskỳ, Martin and Hàlovà, Martina and Pŭta, František and Folk, Petr. Introns provide a platform for intergenic regulatory feedback of RPL22 paralogs in yeast. *PloS one* **2018**, *13*, e0190685.

[292] Cyrne, L.; Antunes, F.; Sousa-Lopes, A.; Diaz-Bérrio, J.; Marinho, H.S. Glyceraldehyde-3-phosphate dehydrogenase is largely unresponsive to low regulatory levels of hydrogen peroxide in Saccharomyces cerevisiae. *BMC biochemistry* **2010**, *11*, 49.

[293] Yardim-Akaydin, S.; Sepici, A.; Özkan, Y.; Şimşek, B.; Sepici, V. Evaluation of allantoin levels as a new marker of oxidative stress in Behçet's disease. *Scandinavian journal of rheumatology* **2006**, *35*, 61–64.

UNIVERSITAT ROVIRA I VIRGILI
ANALYSIS OF MICROBIAL POPULATIONS IN WINES THROUGH NGS METHODOLOGIES
Dimitrios Kioroglou

UNIVERSITAT ROVIRA I VIRGILI
ANALYSIS OF MICROBIAL POPULATIONS IN WINES THROUGH NGS METHODOLOGIES
Dimitrios Kioroglou