Comparative genomics of industrial yeasts

A DISSERTATION PRESENTED
BY
BRIGIDA GALLONE
TO

GHENT UNIVERSITY & UNIVERSITY OF LEUVEN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY
IN THE SUBJECTS OF

BIOCHEMISTRY AND BIOTECHNOLOGY
GHENT UNIVERSITY
GHENT, BELGIUM

AND

BIOSCIENCE ENGINEERING
UNIVERSITY OF LEUVEN
LEUVEN, BELGIUM

GHENT UNIVERSITY

KU LEUVEN

Thesis commitee

Promoters

Prof. Dr. Ir. Steven Maere

Department of Plant Biotechnology and Bioinformatics
Ghent University

Prof. Dr. Kevin J. Verstrepen

Faculty of Bioscience Engineering
University of Leuven

Other members

Prof. Dr. Geert De Jaeger, Ghent University

Prof. Dr. Vera van Noort, University of Leuven

Prof. Dr. Joseph Schacherer, University of Strasbourg

Dr. Gianni Liti, University of Nice

This dissertation is dedicated to all the people
who devoted their lives to
the conservation of biodiversity,
the most precious thing on our planet.

1

Summary

Production of beer, wine and bread, can be considered the earliest examples of biotechnology together with plant and animal domestication. The most ancient, chemically confirmed barley beer to date, ca. 3400-3000 BC, was discovered at the archeological site of Godin Tepe, western Iran. Fermented products played an important role in the emerging human societies and are a direct consequence of the domestication of cereals and the development of ceramics. Historically these fermentations developed from spontaneous processes that rely on natural, uncontrolled inoculation of the medium with complex mixture of microbes. The conversion of sugars into ethanol and desirable flavor compounds performed by yeast during the fermentation process, allowed for obtaining foods and beverages with a prolonged shelf-life, enriched sensorial palate and improved digestibility. *Saccharomyces cerevisiae* is the predominant yeast species in certain types of fermentations, e.g. in ale beers and wine, and likely played a crucial function in ancient fermentations too. The role of yeast in the fermentation process was elucidated only in the 19th century and consequently systems to isolate pure cultures and reuse the same yeast became available.

Nowadays, the majority of industrial fermentations relies on a multitude of well-defined starter cultures, mostly consisting of specialised *S. cerevisiae* strains applied to highly controlled processes. Industrial fermentations vary greatly according to the specific application, but they all share the presence of several environmental stressors such as extreme temperatures, high alcohol concentrations, fluctuating osmolarity, variable pH and limited nutrients. As a consequence, the transition from natural to industrial environments imposed strong selective pressures on *Saccharomyces* yeasts, often divergent according to the specific industry considered. Several studies

investigated *S. cerevisiae* evolution focusing primarily on isolates from wild and clinical habitats while including only a limited number of industrial strains mainly from wine/vineyard environments.

In this project we aimed to determine the nature, extent and origin of industrial *Saccharomyces* yeast diversity by undertaking a whole-genome sequencing approach combined with high-throughput phenotyping. We investigated how hundreds of years of unknowing selection shaped the genomes and phenomes of industrial yeasts compared to non-industrial isolates and across different industrial applications. We describe an in depth genetic and phenotypic analysis of more than 200 *Saccharomyces* yeasts directly used in the industrial production of multiple fermentation products (wine, beer, saké, bioethanol, bread and spirits) or simply associated with the industrial environments (e.g. brewing contaminants). We show that today's industrial yeasts are genetically and phenotypically separated from wild strains and originated from only a few ancestors through complex patterns of domestication and local divergence. Furthermore, we observed strong industry-specific selection for stress tolerance, sugar utilization and aroma production, while the sexual cycle and phenotypes related to survival in nature show decay. Signatures of domestication differ according to the specific industrial environment and are most clear in beer yeasts, likely because they continuously live in their niche, allowing only limited genetic admixture with wild stocks and minimal contact with natural environments. In addition, we describe how niche adaptation in industrial yeasts has been facilitated by interspecific hybridisation between various *Saccharomyces* species. Hybrids originated and thrived in industrial environments as a result of their unique combination of key traits from each parental species. This process was accompanied by extensive structural rearrangements and chimaerisation between the sub-genomes.

Taken together, our results provide a comprehensive analysis of the evolutionary history of *Saccharomyces* industrial yeasts and demonstrate how we can apply the rich biodiversity of industrial fermentations to study microbial domestication and, more generally, genetic mechanisms underlying swift adaptation to new niches.

2

Samenvatting

Samen met de domesticatie van planten en dieren kan het brouwen van bier, maken van wijn en bakken van brood worden beschouwd als één van de eerste voorbeelden van biotechnologie. Het oudste (chemisch geverifieerde) bewijs van de productie van gerstenbier werd aangetroffen op de archeologische site van Godin Tepe in West-Iran, en werd gedateerd op 3400-3000 v. C. Gefermenteerde producten hebben een belangrijke rol gespeeld in het ontstaan van de menselijke samenleving en zijn een direct gevolg van de domesticatie van granen en de ontwikkeling van keramiek. De fermentaties waren initieel een spontaan proces, waarin een natuurlijk, niet-gecontroleerde inoculatie met een complexe microbiële cocktail de fermentatie in gang zette. De door gist uitgevoerde biochemische omzetting van suikers in ethanol en flavorcomponenten tijdens het fermentatieproces resulteerde in producten met een verhoogde houdbaarheid, verrijkt aromaprofiel en verbeterde verteerbaarheid voor de mens. *Saccharomyces cerevisiae* is vandaag één van de meest voorkomende gistspecies in veel industriële fermentatieprocessen, zoals bv. bieren van hoge gisting en wijn, en vervulde waarschijnlijk vroeger ook al een cruciale rol in spontane fermentatiereacties. De rol van gist in het fermentatieproces werd pas in de 19de eeuw ontrafeld. Deze baanbrekende ontdekking opende de deur tot de ontwikkeling van technieken om pure gistculturen te isoleren en specifieke giststammen voor fermentaties te selecteren.

Sindsdien wordt de meerderheid van industriële fermentaties gecontroleerd door het gebruik van starterculturen (meestal één pure *S. cerevisiae* giststam), wat de reproduceerbaarheid en kwaliteit sterk verhoogt. Industriële fermentaties kunnen (afhankelijk van de specifieke applicatie) zeer divers zijn, maar hebben gemeen dat de omgevingscondities zeer stresserend kunnen zijn voor

de gist. Extreme temperature (zowel hoge als lage), hoge alcoholconcentraties, fluctuerende osmolariteit, variabele pH en nutriëntenschaarste zijn slechts enkele van de stressoren die in een fermentatieproces kunnen voorkomen. Door de aanwezigheid van deze suboptimale omgevingscondities ging voor *S. cerevisiae* de overgang van hun natuurlijke, wilde habitat naar de fermentatieomgeving gepaard met een sterke selectiedruk, die sterk verschilde naargelang het specifieke fermentatieproces. En hoewel verschillende wetenschappelijke studies de evolutie van *S. cerevisiae* hebben bestudeerd, beperken deze studies zich voornamelijk tot de 'wilde' en 'klinische' isolaten, en worden slechts sporadisch industriële gisten onderzocht.

Deze doctoraatsthesis heeft tot doel de diversiteit van industriële *S. cerevisiae* gisten in kaart te brengen en hun oorsprong en evolutie te onderzoeken. Hiervoor werden de genomen van deze gisten volledig gesequeneerd, en hun karakteristieken via high-throughput fenotyperingen in detail bepaald. Deze analyses geven inzicht in hoe door honderden jaren van onbewuste selectie door mensen het genoom en fenoom van gisten uit verschillende industriële fermentatieprocessen geëvolueerd is, en hoe zich dit verhoudt ten opzichte van hun 'wilde' voorouders. Meer specifiek werden meer dan 200 *Saccharomyces* gisten die als startercultuur in verscheiden fermentatieprocessen (bv. wijn, bier, saké, bioethanol, brood, sterke dranken) worden gebruikt, of op een andere manier geassocieerd zijn met de fermentatieomgeving (bv. contaminanten van de brouwerij), in detail onderzocht. We tonen aan dat de huidige industriële giststammen genetisch en fenotypisch volledig gescheiden zijn van stammen geïsoleerd uit niet-fermentie omgevingen. Daarnaast tonen onze analyses aan dat industriële gisten onstaan zijn uit slechts enkele voorouders en dat de huidige biodiversiteit het gevolg is van complexe domesticatiepatronen en lokale divergentie. Daarnaast hebben we bewijzen gevonden van sterke industrie-specifieke selectie voor stress-tolerantie, suikermetabolisme en aromaproductie, terwijl hun sexuele cyclus en fenotypes gerelateerd aan overleving in een niet-fermentatie-omgeving sterk verval vertonen. De exacte kenmerken van domesticatie verschillen naargelang het specifieke fermentatieproces, maar zijn het duidelijkst merkbaar in biergisten. Dit is waarschijnlijk het gevolg van de continue associatie met hun fermentatie-omgeving, wat het contact met meer 'natuurlijke' omgevingen limiteert en de kans op genetische inmenging van andere, wilde giststammen sterk reduceert. Daarnaast tonen we ook aan hoe de adaptatie van industriële gisten aan hun nieuwe niche kan gefaciliteerd worden door interspecifieke hybridizatie tussen verschillende *Saccharomyces* species. Deze hybriden zijn ontstaan in (of in de nabijheid van) industriële fermentaties en zijn perfect aan deze condities aangepast door de unieke combinatie van karakteristieken van beide ouderspecies. Dit hybridizatieproces heeft geleid tot veelvuldige structurele herschikkingen in het genoom en chimaerisatie tussen de twee subgenomen.

Samengevat schetsen onze resultaten een gedetailleerd beeld van de evolutionaire geschiedenis van industriële *Saccharomyces* gisten, en tonen we aan hoe deze rijke biodiversiteit een perfect model is om microbiële domesticatie (en meer algemeen de genetische mechanismen die verantwoordelijk zijn voor niche-adaptatie) te bestuderen.

Contents

1 Summary **7**

2 Samenvatting **9**

3 List of abbreviations **19**

4 Introduction **23**
4.1 Abstract 24
4.2 Introduction 24
4.3 From phenotyping to Whole-Genome Sequencing 29
4.4 *Saccharomyces cerevisiae* an alcohol-producer tuned to perfection 31
4.5 *Saccharomyces pastorianus* the odd one out of the *Saccharomyces* genus 43
4.6 *Brettanomyces bruxellensis* the locomotive of spontaneous beer fermentation . . . 51
4.7 Conclusions 60
4.8 References 62

5 Scientific problem addressed and aims of the project **77**
5.1 References 80

6 Domestication and Divergence of *Saccharomyces cerevisiae* Beer Yeasts **83**
6.1 Abstract 84
6.2 Introduction 85
6.3 Results 86
6.4 Discussion 104
6.5 Methods 111
6.6 References 128

7 Origins, Evolution, Domestication and Diversity of *Saccharomyces* Beer Yeasts **133**
7.1 Abstract 135
7.2 Introduction 135

7.3 Domestication of *Saccharomyces cerevisiae* ale beer yeasts . . . 136
7.4 Domestication of *Saccharomyces pastorianus* lager beer yeasts . . . 142
7.5 Conclusions and outlook . . . 145
7.6 References . . . 147

8 Interspecific Hybridization Facilitates Niche Adaptation in Beer Yeasts 151
8.1 Abstract . . . 152
8.2 Introduction . . . 152
8.3 Results & Discussion . . . 153
8.4 Concluding remarks . . . 176
8.5 Supplementary figures . . . 177
8.6 Methods . . . 186
8.7 Supplementary tables . . . 195
8.8 References . . . 196

9 Conclusions and future prospects 203
9.1 **General conclusions** . . . 203
9.2 **Finding meaning in complexity** . . . 205
9.3 References . . . 209

10 Funding, Author Contributions and Conflict of Interest 213
10.1 Funding . . . 213
10.2 Contribution . . . 213
10.3 Conflict of interest . . . 214

11 Thanks 215

12 Resume 217
12.1 Contact Information . . . 217
12.2 Research Interests . . . 217
12.3 Education . . . 218
12.4 Research Experience . . . 218
12.5 Conference Presentations . . . 219
12.6 Honors & Awards . . . 219
12.7 Editorial Work . . . 219
12.8 Teaching & Mentoring . . . 219
12.9 References . . . 220
12.10 List of Publications . . . 221

Listing of figures

4.1 ***Saccharomyces cerevisiae*** **phylogenetics and population structure.** . . . 36
4.2 ***Saccharomyces cerevisiae*** **phylogenetics and population structure.** . . . 37
4.3 **The origin of lager yeast.** . . . 47
4.4 ***Brettanomyces bruxellensis*** **phylogenetics.** . . . 57

6.1 **Graphical Abstract.** . . . 84
6.2 **Phylogeny and population structure of industrial *S. cerevisiae* strains. Part 1.** . . . 87
6.3 **Phylogeny and population structure of industrial *S. cerevisiae* strains. Part 2.** . . . 88
6.4 **Ploidy and Copy-Number Variation in industrial *S. cerevisiae* strains. Part 1.** . . . 91
6.5 **Ploidy and Copy-Number variation in industrial *S. cerevisiae* strains. Part 2.** . . . 92
6.6 **Trait variation of industrial *S. cerevisiae* strains. Part 1.** . . . 94
6.7 **Trait variation of industrial *S. cerevisiae* strains. Part 2.** . . . 95
6.8 **The reproductive lifestyle of industrial *S. cerevisiae* strains. Part 1.** . . . 96
6.9 **The Reproductive lifestyle of industrial *S. cerevisiae* strains. Part 2.** . . . 97
6.10 **Production of 4-Vinyl Guaiacol by industrial *S. cerevisiae* strains. Part 1.** . . . 100
6.11 **Production of 4-Vinyl Guaiacol by industrial *S. cerevisiae* strains. Part 2.** . . . 101
6.12 **Production of 4-Vinyl Guaiacol by industrial *S. cerevisiae* Strains. Part 3.** . . . 102
6.13 **Phylogeny and population structure of industrial *S. cerevisiae* Strains, related to Figures 6.2-6.3. Part 1.** . . . 106
6.14 **Phylogeny and population structure of industrial *S. cerevisiae* strains, related to Figures 6.2-6.3. Part 2.** . . . 107

6.15 **Copy number variability across the phylogenetic tree, related to Figures 6.4-6.5.** . . . 108
6.16 **Trait variation in industrial *S. cerevisiae* strains, related to Figures 6.6-6.7.** . . . 109
6.17 **Heterozygosity of industrial yeasts, related to Figures 6.8-6.9.** . . . 110

7.1 **Graphical Abstract.** . . . 134
7.2 **Genetic and phenotypic diversity of industrial *S. cerevisiae* strains: Maltotriose utilization.** . . . 137
7.3 **Genetic and phenotypic diversity of industrial *S. cerevisiae* strains: Production of 4-vinyl guaiacol (4VG).** . . . 138
7.4 **Genetic and phenotypic diversity of industrial *S. cerevisiae* strains: Sporulation efficiency.** . . . 139
7.5 **Ploidy variation of industrial yeast genomes.** . . . 141
7.6 **Current models for the origin of Frohberg and Saaz lineages.** . . . 144

8.1 **Tanglegram depicting the relationships between *Saccharomyces* pure species and interspecific hybrids.** . . . 156
8.2 **Genome structure of *S. cerevisiae* x *S. kudriavzevii* interspecific hybrids.** 158
8.3 **Genome structure of *S. uvarum* x *S. eubayanus* interspecific hybrids.** . 159
8.4 **Genome structure of *S. cerevisiae* x *S. eubayanus* (Saaz) interspecific hybrids.** . . . 160
8.5 **Genome structure of *S. cerevisiae* x *S. eubayanus* (Frohberg) interspecific hybrids.** . . . 161
8.6 **Distribution of chimeric breakpoints across *S. cerevisiae* x *S. kudriavzevii* hybrids** . . . 163
8.7 **Distribution of chimeric breakpoints across *S. uvarum* x *S. eubayanus* hybrids** . . . 164
8.8 **Distribution of chimeric breakpoints across *S. cerevisiae* x *S. eubayanus* hybrids** . . . 165
8.9 **Trait variation of interspecific hybrids.** . . . 168
8.10 **Interspecific hybridization leads to novel phenotypic combinations that allow colonization of new niches.** . . . 169
8.11 **Distribution of 4-VG production across *Saccharomyces* pure species and hybrids** . . . 170
8.12 **Schematic representation of the genetic basis of loss of 4-VG production in *S. cer* x *S. eub* hybrids.** . . . 172
8.13 **Time-calibrated phylogeny of *S. cerevisiae* Beer 1 clade.** . . . 175
8.14 **Detailed phylogeny of *S. cerevisiae* genomes and subgenomes.** . . . 178

8.15 **Detailed phylogeny of *S. eubayanus* genomes and subgenomes.** 179
8.16 **Detailed phylogeny of *S. uvarum* genomes and subgenomes.** 180
8.17 **Detailed phylogeny of *S. kudriavzevii* genomes and subgenomes.** 181
8.18 **Genomic instability and copy-number heterogeneity in the sequenced interspecific hybrids.** 182
8.19 **Detailed view of chimeric breakpoints identified in four representative hybrids.** 183
8.20 **Chimeric breakpoints overlap low nucleotide identity regions between subgenomes.** 184
8.21 **Potential introgressive hybridization in a *S. cer* x *S. eub* strain.** 185

List of Tables

4.1 **Overview of phenotypic differences between the lager yeast *S. pastorianus* and the ale yeast *S. cerevisiae*.** . . . 30
4.2 **Examples of (suspected) domestication traits of *S. cerevisiae*.** . . . 33
4.3 **Distribution of 13 clustered genes across different *Brettanomyces bruxellensis* strains originating from beer.** . . . 58
4.4 **Distribution of 13 clustered genes across different *Brettanomyces bruxellensis* strains originating from wine.** . . . 59

6.1 **Genetic diversity within each subpopulation of industrial *S. cerevisiae* strains.** 90
6.2 **Distance between subpopulations, determined by the average number of nucleotide substitutions per site (d_{XY})** . . . 126
6.3 ***S. cerevisiae* strains included in this study.** . . . 127
6.4 **Referenced overview of all strains included in the multilocus phylogenetic analysis.** . . . 127
6.5 **Gene Ontology Enrichment Analysis of genes with increased and decreased copy-number variation in all strains of the collection.** . . . 127
6.6 **Gene Ontology Enrichment Analysis of genes with increased and decreased copy-number variation in individual subpopulations.** . . . 127
6.7 **Phenotypic variation of industrial *S. cerevisiae* strains.** . . . 127
6.8 **Statistical analysis of the phenotypic behavior of the different subpopulations in environmental stress conditions, ethanol accumulation capacity, and maltotriose utilization.** . . . 127
6.9 **Distribution of *MAL11* alleles in the sequenced strains.** . . . 127

8.1 ***Saccharomyces* strains included in this study.** . . . 195
8.2 **Detailed overview of chimeric breakpoints identified in each hybrid** . . . 195
8.3 **Phenotypic variation across *Saccharomyces* hybrids and *Saccharomyces* pure species.** . . . 195

Large tables for chapter 6 and chapter 8 are provided at: https://bit.ly/2O6C8Zk

3

List of abbreviations

Abbreviation	**Explanation**
°P	Degrees Plato ; 1°P = 1 % w w^{-1} soluble extract
4-VG	4-vinylguaiacol
5-FOA	5-fluoroorotic acid
AAB	Acetic acid bacteria
AAM	Acetic acid medium
ADH	Alcohol dehydrogenase
AGT1	alpha-glucoside transporter
ATF1/2	Alcohol acetyltransferase 1/2
au	Arbitrary units
BCE	Before common era
Bel/Ger	Belgium/Germany
BIC	Bayesian information criterion
bp	base pairs
BS	Bootstrapping
BSA	Bulk segregant analysis
BWA	Burrows-Wheeler transform
cAMP/PKA	Cyclic adenosine monophosphate/Protein kinase A

Abbreviation	Explanation
CFU	Colony forming units
chr	chromosome
CNV	Copy Number Variation
CV	Coefficient of variation
dH2O	Distilled water
DNA	Deoxyribonucleic acid
EA	Ethyl acetate
EMS	Ethyl methanesulfonate
ESS	Effective sample size
ExaML	Exascale Maximum Likelihood
EXPLoRA	Extraction of over-represented alleles in BSA
FACS	Fluorescence activated cell sorting
FDC1	Ferulic acid decarboxylase
FDR	false discovery rate
FGM	Fermentable growth medium-induced
GATK	Genome analysis toolkit
GM	Genetically modified
GWAS	Genome Wide Association Study
gTME	Global transcription machinery engineering
HGT	Horizontal gene transfer
HML	Hidden MAT left
HMM	hidden markov model
HMR	Hidden MAT right
HPD	High Posterior Density
HS-GC-FID	Headspace - Gas Chromatograph- Flame Ionization Detector
IA	Isoamyl acetate
InDels	Insertions and deletions
ITS	Internally transcribed spacer
LAB	Lactic acid bacteria
LCA	last common ancestor
LOH	Loss of heterozygosity
MAL	Maltose
MCC	maximum clade credibility

Abbreviation	Explanation
MCMC	Markov chain Monte Carlo
ML	Maximum-likelihood
MPD	Mean pairwise distance
MRCA	Most recent common ancestor
MRE	majority rule extended
MRS	de Man, Rogosa and Sharpe
MSA	multi sequence alignment
NA	Not available
OD600	Optical density at 600nm
ORF	Open reading frame
PAD1	Phenylacrylic acid decarboxylase
PBS	Phosphate buffered saline
PCA	Principal component analysis
PCR	Polymerase chain reaction
PI	Propidium iodide
POF	Phenolic off-flavor
Ppm	Parts per million
Psi	Pound-force per square inch
QTL	Quantitative trait locus
RAxML	Randomized Axelerated Maximum Likelihood
rDNA	Ribosomal DNA
REML	restricted maximum likelihood
RF	Robinson-Foulds
RHA	Reciprocal Hemizygosity Analysis
RLC	Random local clock
rpm	Revolutions per minute
S.	*Saccharomyces*
SC	Synthetic Complete (medium)
Scer	*Saccharomyces cerevisiae*
Seub	*Saccharomyces eubayanus*
Skud	*Saccharomyces kudriavzevii*
SNP	Single nucleotide polymorphism
Spar	*Saccharomyces paradoxus*

Abbreviation	**Explanation**
Suva	*Saccharomyces uvarum*
Tukey HSD	Tukey's Honest Significant Difference test
UCLD	uncorrelated relaxed molecular clock (log normal distribution)
UPGMA	Unweighted Pair Group Method with Arithmetic Mean
UV	Ultra violet
WGD	Whole genome duplication
WLN	Wallerstein Laboratory Nutrient
WRF	Weighted Robinson-Foulds
YAC	Yeast articificial chromosome
YPGlu	Yeast extract peptone Glucose
α-AA	α-aminoadipic acid

4

Introduction

This chapter is adapted from:

Brigida Gallone, Stijn Mertens, Sam Crauwels, Bart Lievens, Kevin J. Verstrepen and Jan Steensels. Genomics and Evolution of Beer Yeasts. *Brewing Microbiology*, 145-178, 2017.

Jan Steensels, Brigida Gallone, Karin Voordeckers and Kevin J. Verstrepen. On microbes and men. *Primer Current Biology*, in preparation.

Author contributions:

Conceptualization: B.G., J.S., S.C., B.L., K.J.V; Data Visualisation: B.G., S.M., S.C.; Writing: B.G., J.S., S.M., S.C., B.L., K.J.V; Supervision: J.S., B.L., K.J.V.

4.1 Abstract

The evolutionary adaptation of organisms to a specific niche is one of the most fascinating processes in biology. Classic Darwinian theory explains how the interplay between (genetic and phenotypic) variation on one hand, and selection on the other, drives evolution. For certain traits, evolutionary adaptation is forced by man-mediated selection, which results in domestication or the adaptation of organisms to man-made niches, as is commonly observed in crops, livestock, and pet animals. Yeasts serve as a very interesting model organism for adaptive evolution, since their small and compact genomes provide a very attractive and powerful model for comparative genomics and genome-evolution studies. Moreover, several species, such as the traditional baker's or brewer's yeast *Saccharomyces cerevisiae*, have been subjected to natural as well as human selection, both of which shaped their genotypes and phenotypes. The emergence of whole-genome sequencing technologies resulted in an overwhelming amount of high-quality and highly detailed yeast genome sequences, allowing researchers to investigate how these yeasts interacted with (and adapted to) their environment. In this Chapter, we describe how evolutionary processes shaped the genome and phenome of three intriguing yeast species associated with beer brewing: *S. cerevisiae*, *Saccharomyces pastorianus*, and *Brettanomyces bruxellensis*.

4.2 Introduction

Darwin's theory of evolution describes how species gradually change over time through variation and selection. When the gene flow between individuals of the same species is reduced, for example because of geographical isolation, evolution can result in genetically and phenotypically distinct subpopulations, which can eventually lead to speciation [1, 2]. The driving force behind evolution is mutation on which natural selection can act on [3]. However, selection is not always natural. Instead, humans often interfere with the natural process by deliberately limiting gene flow to specific subpopulations, by selecting traits that do not necessarily bestow higher fitness in natural environments, and by creating unnatural, man-made niches. This process, dubbed 'domestication', explains how some species have become completely accustomed and even dependent on humans and have evolved features that are useful or pleasing to humans, often at the cost of fitness in natural settings [4, 5]. For example, present-day bulldogs have been bred to have a short nose and flat face, giving them a more human-like expression. However, this feature is clearly not advantageous in a more natural setting, as the shorter and flatter snout causes severe breathing issues [6]. Similarly, the Belgian Blue cattle race is a very muscular race that was bred for its higher feed conversion rate and increased muscular mass, but the hyperplasmia of the muscles results in an obstructed birth canal (dystocia) and makes reproduction of the animals dependent on caesarian sections [7–9]. During domestication, specific features useful for or desired by humans are selected for. These features can range from yield (e.g. beef cattle [10] and agricultural crops

[11]) and stress tolerance (e.g. drought-resistant crops [12]) to behavioral (e.g. companion animals [13]) and esthetical (e.g. flowers [14, 15]) aspects. The emergence of domestication is intertwined with the First Agricultural revolution (the 'Neolithic Revolution'), the period around 12,000 years ago when humans changed their way of life from hunting and gathering to one of agriculture and settlement [16]. They started cultivating plants and animals, and started selecting offspring of these organisms that best met their specific needs, thereby effectively taking over the natural processes of selection and gene flow. For example, a mutation that prevents wheat to shatter and fall to the ground when ripe was selected for by humans already very early after initial wheat cultivation. This mutation, which makes wheat harvesting much easier, still persists in present-day wheat varieties [17, 18]. The general concept of domestication was already described in 1868 by Charles Darwin [19]. He was the first to recognize the striking morphological resemblance between different domestic animals, despite the lack of close evolutionary relationships between their wild ancestors. Smaller brains, smaller teeth, increased passivity and docility, floppy ears, shorter curly tails and lighter and blotchy coats are all examples of phenotypes that hallmark many animal domesticates, and which Darwin therefore referred to as the 'domestication syndrome'. Experimental evidence for this theory was obtained in an audacious experiment in Siberia in the 1950s, known as the 'Russian farm-fox experiment' [20]. In these experiments, Soviet scientists performed an elaborate, multi-generation breeding effort with red foxes, each generation choosing only the tamest individuals for the next breeding round. Within 10 generations, the foxes did not only become more docile and friendly towards humans, they also started to show the same array of phenotypes described by Darwin, even though these were not directly selected for [21]. This new lineage still exists to this day and is now even available as a pet. More importantly, this experiment highlights the power of artificial selection to transform species, even over short evolutionary timescales.

Compared to plants and animals, the emergence of domesticated microbes has been studied to a much lesser extent. Nevertheless, it is hard to imagine a world without bread, cheese, beer, wine, sake, yoghurt, soy sauce or spirits; all of which exist in their present forms because of domesticated microbes. Interestingly, whereas the domestication of animals and crops was mostly the result of a conscious human effort, the domestication of microbes was largely unconscious. Indeed, the very existence of microbes was only discovered by Antonie van Leeuwenhoek in the late 17th century and the role of microbes in fermentations was only unraveled in the 19th century by the pioneering work of Louis Pasteur [22]. Perhaps most importantly, the first pure culture of microbes, more specifically lager beer yeast, was only isolated in 1883 by Emil Christian Hansen at the Carlsberg brewery [23]. Before this breakthrough, fermentations occurred spontaneously and were driven by contaminating microbes present in the environment. However, after the switch from a nomadic to sedentary society in the Neolithic revolution, humans also started to change the way they performed fermentations, motivated by the clear advantages of process control on reproducibility, safety, taste and flavor of the product. The most drastic practice that triggered microbial domestication is probably 'backslopping', a process in which artisans continuously pas-

saged materials (and thus also microbes) from the previous fermentation product to start a new batch. As a result, these early artisans unconsciously promoted adaptation of the microbes to the human-made fermentation environment and generated subpopulations that were isolated from their natural habitats and started adapting to the human-made niches associated with the production of fermented beverage and foods. While conceptually similar, it is important to note that there are several fundamental differences between the domestication of microbes and that of higher eukaryotes. First, since microbial populations are typically very large and individuals can only be isolated using modern equipment, selection occurred on populations of millions of cells, leading to a less-controlled domestication compared to plant and animals where meticulous selection happened at the level of individuals. Second, animal and plant breeding largely act on genetic pathways that affect growth, morphology and social behavior, while microbial domestication centers on altering metabolism, tweaking a microbe's machinery for producing energy and chemicals [24]. Third, during microbial domestication, there is a continuous and fierce competition between different variants (e.g. species, strains, mutants) in each fermentation round. In a way, it is 'survival of the fittest' in a human-controlled environment. Therefore, much of the microbial domestication characteristics were selected on the basis of increased fitness in specific fermentation environments. Interestingly, nowadays, new biotechnological techniques provide opportunities to perform more 'traditional' domestication in microbes, in which specific parental strains with desired characteristics can be crossed to develop offspring with new, advantageous properties, such as the production of exotic flavors [25]. Although microbial domestication is closely linked with our daily lives, scientists are still puzzled by several seemingly simple questions: how, why and when did microbial domestication happen? Studying microbial domestication is challenging for a number of reasons. Most importantly, in contrast to some domesticated animals or plants, no fossils or ancient DNA is available to reconstruct patterns of evolution. Moreover, geographic origins and information about the life history of individual strains are often not documented. Recently however, several studies used large-scale whole-genome sequencing and phenotyping to illuminate some of the origins, ecological context and mechanisms driving the transition from wild microbial strains to domesticates [26–31].

Over the past decades, numerous microbial species have been linked to domestication. However, it is important to bear in mind that not all food-associated microbes have been subjected to domestication, simply because not all fermentation setups would induce domestication. Domestication requires continuous cultivation and evolution of the microbe in human-associated environments. Therefore, it is important to critically assess whether a food-associated microbial organism meets the basic domestication criteria, or whether they are merely haphazardly present during the fermentation process because they are the best-adapted natural variant. The key features to evaluate whether a microbe is domesticated include (i) clear genetic differentiation from the wild ancestor, indicative of limited gene flow required to obtain a clearly separated species, subpopulation, strain, (ii) partial dependency on and adaptation to the human environment, (iii)

suboptimal performance in natural settings caused by loss of genes/metabolic pathways no longer required in the human environment ('genome decay') and (iv) reduction of genetic diversity due to repeated bottlenecks and intensive selection during the backslopping process. This reduction might be followed by further population expansion and diversification in the man-made niche.

Lactic acid bacteria (LAB) are a class of bacteria that produce lactic acid from sugars. These bacteria naturally occur in both plant and dairy environments. Today, many LAB play a crucial role in the production of various foods and beverages, where they contribute to flavor, texture, safety and quality of fermented foods. Domestication of LAB resulted in two important distinct lineages of starter cultures used for fermented food production [32]. *Lactobacillus* and *Lactococcus* are two LAB genera used to make fermented dairy products, including soft and hard cheeses, sour cream and buttermilk. In both genera, domestication is associated with species-specific gene loss in metabolic pathways, sporulation and oxidative stress responses, as well as acquisition of genes specifically required for optimal growth in the dairy environment, such as genes involved in utilization of milk protein (casein) and milk sugar (lactose) [33, 34]. For example, domesticated *Lactococci* often have lost the ability to synthesize specific amino acids, and instead have acquired an elaborate system of proteases that allows them to break down the milk protein casein into amino acids. Many of these novel functions can be found on large plasmids (so-called plasmidomes) acquired through horizontal gene transfer [35]. Another key difference between domesticated and wild *Lactococci* is the ability to degrade plant-derived sugars. Plant material contains many different carbohydrates, whereas milk mainly contains lactose. Wild isolates of *Lactococci* can grow on a variety of plant-derived carbohydrates, but domesticated isolates no longer can, due to pseudogenization or transposon insertion in genes encoding for example in carbohydrate transporters [36].

The fungus *Aspergillus oryzae* is important for the production of fermented traditional Japanese foods such as miso (soy bean paste), su (vinegar) and shoyu (soy sauce). *A. oryzae* is also involved in the production of sake (wine rice), where it breaks down the starch present in rice into sugars that can subsequently be fermented by yeast. *A. oryzae* was domesticated from *Aspergillus flavus*, an agricultural pest that produces the carcinogenic toxin aflatoxin. The *A. oryzae* genome has accumulated inactivating mutations in the genes encoding for aflatoxin and hence no longer produces this toxin. In fact, it has been postulated that *A. oryzae* was domesticated from an *A. flavus* variant that no longer produced this toxin [37, 38]. Compared to its wild relative, *A. oryzae* contains more copies of the α-amylase gene, responsible for the breakdown of starch. Additionally, the genome of A. oryzae has multiple genes encoding secreted hydrolytic enzymes, responsible for the breakdown of proteins and complex carbohydrates present in the outer layers of grains, allowing the fungus access to the starch-rich interior. The origin of these genes is still debated, but horizontal gene transfer has been proposed as the most likely mechanism [39]. However, the most intensively studied microbial domesticate is probably the baker's and brewer's yeast *Saccharomyces cerevisiae*, the main driver of many industrial fermentation processes performed today.

Few people realize that yeast serves as an important eukaryotic model organism for basic research. Between 2001 and 2016, five Nobel prizes (in Physiology or Medicine and Chemistry) have been awarded to researchers who used yeast to understand fundamental aspects of (human) biology, including cell cycle regulation, gene expression, cell aging, autophagy and protein secretion (http://www.nobelprize.org). Moreover, in 1996, the baker's yeast *S. cerevisiae* (strain S288c), was the first eukaryotic organism to have its genome fully sequenced. This multi-million-dollar project was driven by an international consortium of researchers from 19 countries working in 94 laboratories, using several different sequencing methods and technologies [40, 41]. Now, two decades later, as whole-genome sequencing (WGS) becomes an almost trivial procedure, we are faced with a burst of high-quality and highly detailed yeast genome sequences; in-depth sequencing of a yeast genome now takes only a few days and costs a few hundred dollars [42, 43]. Consequently, to date, the whole genome sequence of over 40 yeast species has been published (http://www.genolevures.org/), with thousands of additional genomes in the pipeline. This rich dataset makes it possible to systematically explore the interplay between genetic variation (i.e. changes in DNA sequences) and phenotypic variation (i.e. changes in properties and behavior).

In the last 20 years, our basic knowledge of the physiology and evolution of yeasts has increased drastically. This growing body of information is largely rooted in the recent development of high-throughput techniques and analytical tools, often called "Omic" methodologies, that allow simultaneous investigation of thousands of genes (genomics), transcripts (transcriptomics), proteins (proteomics), and metabolites (metabolomics), providing an in-depth analysis of cellular processes at a global and systems level. Using these groundbreaking techniques, several research groups have invested in large-scale characterization of diverse yeast strains [26, 28, 44–51]. Moreover, while initial genomic studies merely focused on a small number of laboratory isolates (e.g., S288c, a *S. cerevisiae* lab strain bred for easy laboratory use and therefore often not representative of natural or industrial yeasts [52], there is a growing interest in "natural" or "wild" isolates (i.e. isolated from environments not associated with human activity, such as soil or tree bark) and isolates originating from industrial niches (e.g., beer or wine fermentations). Although these studies only emerged in the past couple of years, they already yielded valuable information on genetic diversity, genome evolution, and population history and structure of industrially relevant yeast species, such as *S. cerevisiae*, *S. pastorianus*, and *B. bruxellensis*, which are heavily influenced by their association with man-made fermentation processes.

This chapter will start with a chronological overview of how yeast characterization, identification, taxonomy, and phylogeny evolved from basic phenotyping to advanced sequencing approaches. Next, we summarize the current knowledge of the evolution and genomic features of the three main beer-related yeast species (*S. cerevisiae*, *S. pastorianus*, and *B. bruxellensis*), and illustrate how these observations are related to their behavior in a brewing environment.

4.3 From phenotyping to Whole-Genome Sequencing

This section describes how techniques used for beer yeast typing and characterization evolved from low-resolution, error-prone, and labor-intensive phenotype-based genotyping to WGS. We focus on three species closely associated with beer brewing: the common brewer's yeast *S. cerevisiae* (used in ale production), the lager yeast *S. pastorianus*, and *Brettanomyces (Dekkera) bruxellensis*, an organism involved in spontaneous beer fermentation.

Before the emergence of molecular typing techniques, early yeast classification and speciation was based on the yeast's morphology and physiological properties [53, 54]. This is at the least remarkable, as the definition of species relies on the basic principle of genetic isolation, and not on the subjective appraisal and weighting of phenotypic properties [55]. Morphological differences between beer yeasts are rather limited; they are all unicellular fungi [56], the cellular morphology of which can vary from spheroidal to ovoidal, with multilateral budding observed when clonally reproducing. Cells occur isolated, in pairs, or sometimes as short chains or clusters, and in some conditions even as pseudohyphae [53, 54, 57, 58]. Besides highly similar cell morphology, brewing yeasts also generally form highly similar white, smooth colonies when plated on agar medium.

The shortcoming of brewing yeast classification based solely on morphological properties was already described by Pasteur in his "Études sur la bière" [22] and he clearly stressed the need for other classification criteria. Therefore, several new and more detailed typing methods were developed for the identification and classification of yeasts in the early 20th century. These tests were based on phenotypic identification procedures in which several physiological and biochemical tests were used, typically targeting the yeast's ability to ferment different sugars, assimilate (aerobic growth) carbon and nitrogen compounds and grow in different stressful conditions (e.g. high or low temperatures or vitamin-free medium) (reviewed in [53, 54]).

Since there is often a need in the brewing industry to quickly and effectively distinguish ale (*S. cerevisiae*) and lager (*S. pastorianus*) yeasts, specific and rapid phenotypic tests to discriminate these two species were developed **(Table 4.1, p. 30)**. First, these species were shown to differ in their flocculation behavior. Flocculation is the ability of yeasts to forms flocs (clumps of cells) after the fermentation process. The flocculation process involves flocculins, which are lectin-like proteins that are associated with the cell wall of flocculating cells. These flocculins selectively bind mannose residues present in the cell wall of adjacent yeast cells and are activated by calcium ions present in the medium. Interestingly, *S. cerevisiae* ale yeast tends to adhere to the ascending CO2 bubbles towards the surface of the fermenting wort, whereas *S. pastorianus* lager yeast will sediment to the bottom of the fermentation vessel. This property has been ascribed to the different properties of the *FLO1*/Lg-*FLO1* gene. The lager yeast-specific Lg-*FLO1* gene can not only bind mannose residues, but also has a high affinity towards glucose residues, and is therefore responsible for the NewFLo phenotype of lager yeasts, and makes them to floc out to the bottom of the fermentor towards the end of the fermentation [59–61]. For this reason, *S. cerevisiae* yeast was dubbed top-fermenting

yeast and *S. pastorianus* is generally known as the bottom-fermenting yeast. Second, they differ in their carbon metabolism. Most notably, *S. pastorianus* yeasts produce the extracellular enzyme α-galactosidase (melibiase, encoded by the MEL1 gene), which enables the hydrolysis of melibiose into the readily assimilable sugars galactose and glucose, while *S. cerevisiae* is unable to do so [53, 57, 62, 63]. Moreover, it was also shown that *S. cerevisiae* yeast only can partly ferment the trisaccharide raffinose [57]. Third, ale and lager yeasts show differences in temperature tolerance. Whereas the optimal fermentation temperature of lager yeasts was shown to be below 15°C, it is generally higher for ale yeasts, typically between 20-30 °C [56]. Additionally, ale yeasts are more tolerant towards high temperatures, and can grow at up to 41 or even 42 °C, whereas lager yeasts can only grow at temperatures up to 32-34 °C [57, 64–66].

Table 4.1: Overview of phenotypic differences between the lager yeast *S. pastorianus* and the ale yeast *S. cerevisiae* (adapted from [57]).

Characteristics	Lager strains	Ale strains
Mode of flocculation	Bottom	Top
Fermentation temperature	4-15°C	15-24°C
Max growth temperature	32-34°C	38-42°C
Utilization of maltotriose	More complete	Less efficient
Utilization of melibiose	Yes	No
SO2 production	$>$ 4mg/l	$<$ 2mg/l
Fructose transport	Active proton symport	Facilitated diffusion
Sporulation	No	Yes

Since *Brettanomyces* and *Saccharomyces* yeasts are genetically so distinct and their characteristics often differ widely, development of quick screening tests to distinguish between the two was relatively straightforward. Especially in the wine industry, where *Brettanomyces* is a vicious spoilage organism, much research has been dedicated to developing reliable methods for detection and identification of *Brettanomyces* in the fermentation environment. One of the most widely used methods involves a plating assay on semi-selective media with ethanol as carbon source, combined with bromocresol green and/or phenolic precursors such as hydrocynnamic acids to distinguish the genus *Brettanomyces* from other yeasts after a long period of cultivation [67].

However, it soon became apparent that yeast classification solely based on morphological and phenotypic characteristics was insufficient to deal with the wide variety of yeast used and found in the fermentation industry. Moreover, diverse process developments and changes in the beer brewing industry have undermined some of the earlier mentioned physiologic classification markers. For example, the use of large cylindroconical fermentation vessels induces ale yeast to sediment after the main fermentation to the cone, a property characteristic of lager yeast [53]. Therefore, new techniques, based on the genomic rather than phenotypic features of the yeast, were developed for

the detection, identification, and classification of yeasts [55, 68]. Moreover, these techniques led to a new area for ecological surveys, and enabled researchers to have a closer look into the population dynamics of fermentative yeasts [69].

The first developed molecular method for the differentiation between lager and ale beer yeast was based on mitochondrial DNA (mtDNA) restriction profiling [70]. Gel electrophoresis of the resulting genomic fragments revealed that there were some clear and consistent differences between ale and lager yeasts. Moreover, the obtained patterns of beer yeasts (both ale and lager) were in turn very different to the patterns of non-beer yeasts, suggesting that it also is a good technique to detect possible contaminations.

In the following decades, the portfolio of molecular typing techniques for beer yeast differentiation was further expanded: DNA-DNA homology [71–73], electrophoretic karyotyping [71, 74], random amplified polymorphic DNA analysis (RAPD) [75], amplification of interdelta regions [76], and ribosomal RNA coding DNA restriction fragment length polymorphism (RFLP) [77–79], or a combination of different techniques [55]. Later, more advanced DNA-based techniques with higher resolution, such as microsatellite comparison [80–83]; restriction site-associated sequencing (RAD-seq) [84]; multilocus sequence typing [85–88]; tiling array hybridization [50] and ultimately WGS [49] were developed and established for yeast characterization. The advent of WGS revolutionized the way to investigate and characterize genetic and phenotypic diversity in yeast. The analysis of whole genomes rapidly progressed from the study of one or a handful of yeast isolates to simultaneous investigation of tens or even hundreds of individuals, enabling the development of a population genetic perspective. Examining genome-wide patterns of sequence variation within and between closely related species is providing the first comprehensive view of the evolutionary history of *S. cerevisiae*, *S. pastorianus*, and *B. bruxellensis*.

4.4 *Saccharomyces cerevisiae* AN ALCOHOL-PRODUCER TUNED TO PERFECTION

In order to thrive in a fermentation environment, *Saccharomyces* spp. in general (and *S. cerevisiae* in particular) possess several phenotypic features that make them the ultimate fermentation specialists they are today. These features arose gradually during evolution, both inside and outside man-made fermentation environments.

4.4.1 NATURAL SELECTION SHAPED THE *Saccharomyces* GENOME

One of the most striking attributes of *Saccharomyces* spp. is their perfect adaptation to sugar-rich, oxygen-limited environments. It was hypothesized that the emergence of fruit-bearing plants (a new niche that provided a rich, but highly competitive source of ready-to-use sugars) approx. 80-150 million years ago triggered the selection of a cascade of specific genetic adaptations, all targeted towards colonization of these new niches. Therefore, it seems obvious that these adaptations happened by natural selection, and no man-mediated (artificial) selection was involved.

One of the most striking examples is the emergence of the so-called "Crabtree effect" in *Saccharomyces*. Crabtree-positive yeasts show a metabolism in which glucose (above a certain threshold concentration) represses respiration, so that even when oxygen is still available, cells will favor fermentation. This persistent fermentative behavior has several advantages over respiration. First, fermentation enables a higher carbon flux and faster production of energy. Second, the main end-product of a fermentative metabolism is ethanol, which can serve as an effective antimicrobial agent, to which *Saccharomyces* itself is highly tolerant. Therefore, the Crabtree effect fits a make-accumulate-consume strategy, an ecological strategy in which ethanol is first produced and accumulated to high concentrations to inhibit the growth of other microbes and later consumed again when all fermentable sugars have been converted [89]. Apart from the Crabtree effect, there are several other physiological features that provide *Saccharomyces* spp. with a competitive advantage in fermentation-like environments (e.g. rotting fruit). They have evolved a high tolerance to several environmental stresses (such as high temperatures and a high concentration of osmolytes), a very high glycolytic flux, and the ability to grow in both aerobic and anaerobic conditions [90–92]. It is interesting to note that these individual properties are also present in various other yeasts, but they are only uniquely combined in a few species, including *S. cerevisiae* and its closest relatives, providing a strong competitive advantage over other wild yeasts (and bacteria) in many fermentation environments [92].

While most of these properties are now common knowledge, the underlying genetics were only investigated recently. While many questions remain unsolved, these studies led to the first hints towards the evolutionary pathways *Saccharomyces* spp. went through in response to these newly emerged niches. For example, it was shown that the duplication of several key genes, such as those encoding alcohol dehydrogenase [89, 93], hexose transporters [94], and enzymes linked to glycolysis [90], as well as global rewiring of the transcriptional network after whole-genome duplication [95], played a major role in the evolution of *Saccharomyces*.

4.4.2 *Saccharomyces cerevisiae* as the first domesticated microbe

While adaptation to the presence of fruit-bearing plants occurred millions of years prior, as described above, new opportunities arose for *Saccharomyces* yeasts relatively more recently (approximately 8,000-10,000 years ago). Humankind abandoned its hunter-gatherer lifestyle and introduced a horticultural tradition during the Neolithic evolution. After that, it didn't take long before people realized that exposing fruits and grains to the environment (sometimes) positively changed the characteristics of these products significantly, and additionally prolonged their shelf life. From then on, beer and other fermented beverages such as wine, sake, cider or mead were an inherent part of the human diet, as they served as a source of nutrition, as medicine, and as a vital supply of uncontaminated water [96–99]. Later, skilled artisans found out that it paid to keep a small sample of fermented dough or beer sediment and mix this sample with a new, unfermented batch. In this

way, without realizing it, they transferred specific, well-adapted microbes from one fermentation cycle to the next, thereby introducing the concept of starter cultures. During these consecutive fermentation steps, several novel superior yeast mutants and variants emerged through (mainly unintentional) artificial selection by breeding and directed evolution. It is now hypothesized that this practice induced the gradual adaptation of *Saccharomyces* yeasts to man-made conditions, resulting in organisms specialized in specific fermentation environments but behaving suboptimal in most other, more "natural" habitats, thereby making *S. cerevisiae* one of the oldest domesticated organisms on the planet [49, 100, 101].

Although these initial steps in microbial domestication happened haphazardly and are rarely documented, genetic analysis of current biodiversity enables the identification of specific adaptation. Not unexpectedly, these characteristics are often unique for specific types of industrial fermentation processes, e.g., brewing, baking, and winemaking [101–104] **(Table 4.2, p. 33)**. However, it is important to note that some of these trait divergences might be due rather to genetic drift and in fact pre-date or coincide with the emergence of synthetic fermentation environments (as suggested in [52]).

Table 4.2: Examples of (suspected) domestication traits of *S. cerevisiae*. Domestication traits are characteristics that have diverged between the domesticated strains and their wild ancestors.

Trait	Industry	Responsible gene(s)	Reference
Stress tolerance			
Copper tolerance	Wine	*CUP1*	[49, 52, 105]
Molasses toxin tolerance	Beer, Distillery	*RTM1*	[76, 102]
Sulfite tolerance	Wine	*SSU1*	[106]
Nutrient utilization			
Fructose utilization	Wine	*FSY1*, *HXT3*	[107, 108]
Malto(trio)se utilization	Beer	*AGT1*	[58, 109–113]
Xylose utilization	Wine	*XDH1*	[114]
Sensory quality			
General wine aroma	Wine	Unknown	[115]
Acetate ester production	Fermented beverages	ND	[113]
Phenolic off-flavor (POF) production	Beer	*PAD1*, *FDC1*	[26, 28, 111, 116, 117]
Other			
Flocculation/flor-formation	Sherry, beer	*FLO* genes	[118, 119]
Lag phase	Wine, Bakery	*ARO8*, *ADE5,7*, *VBA3*	[120]
Mesophilic behavior	Lager beer	ND	[111]
Vitamin biosynthesis	Biofuel	*SNO*, *SNZ*	[112]

4.4.3 Beer fermentation as highly selective niche

Yeasts used in the beer industry provide an excellent model to investigate the effects of artificial selection. For example, they can be recovered and reused after the fermentation process (unlike, e.g., bread yeasts) and are employed continuously throughout the year (unlike, e.g., wine yeasts, where the fermentation scheme is tightly linked to the grape harvest season). Therefore, many current beer yeast strains can be considered to be the result of a centuries-long evolution experiment, performed by brewers in a highly selective niche, the brewing environment. It is therefore not surprising that multiple genetic adaptations to the beer-making process have been described.

First and foremost, beer yeasts show a remarkable efficiency in utilization of maltose, the prime carbon source in beer wort. Through extensive duplication and subsequent functional divergence of subtelomeric genes involved in maltose metabolism, present-day brewer's yeast is capable of hydrolyzing a variety of sugars much more efficiently than its ancestors [26, 28, 58, 109–111, 121]. Another trait in which beer yeasts often excel is flocculation, and more specifically the timing of the onset of flocculation. Flocculation is the ability of cells to stick to each other and form aggregates that rapidly sediment to the bottom, or rise to the top, of the fermentation medium. This is an important trait in the beer industry, since it provides an easy and cheap way to separate the yeast cells from the finished beverage [122]. However, early flocculation leads to inefficient or even stuck fermentation processes. Therefore, many brewers' yeasts have been selected to flocculate at the exact moment when all fermentable sugars have been converted into carbon dioxide and ethanol. Moreover, some reports suggest that brewers have fine-tuned the flocculation behavior of their yeast strain by selecting specific layers of yeast sediment for re-inoculation of a subsequent fermentation batch [123]. The genetic basis of this phenotype, and the remarkable speed at which yeasts are able to switch their flocculation behavior, has been studied intensively [119, 122]. It was shown that flocculation behavior is controlled by the *FLO* genes, encoding flocculins, and that the instability of the tandem repeats present in these genes enables relatively rapid expansions and contractions in the gene size, thereby allowing for the fast isolation of spontaneous mutants with altered flocculation characteristics [124]. Lastly, recent evidence also pinpoints 4-VG production (or rather the absence thereof) as an important feature of brewing yeasts. Indeed, genetic analysis shows that disruptive genetic mutations in the causative genes (*PAD1* and *FDC1*) were heavily selected for in brewing yeasts, while this was never encountered in wild strains [26, 28]. Interestingly, this strong adaptation to the beer environment came at a cost. General stress resistance (temperature tolerance, ethanol tolerance, salt tolerance), which is vital for survival in nature, is often impaired in brewing yeast. This loss of "survival skills" is typical for domesticated organisms (imagine releasing a chihuahua in the wilderness), and is one of the clearest signs of human interference with the organism's evolution. In conclusion, *S. cerevisiae* combines several natural features that allow it to thrive in industrial fermentation processes. Moreover, these features were further enhanced or specialized during domestication.

4.4.4 *Saccharomyces cerevisiae* PHYLOGENETICS AND POPULATION STRUCTURE

With their small and compact genomes, *Saccharomyces* yeasts represent a very attractive and powerful model for comparative genomics and genome evolution. Despite the current knowledge on the life history of wild and industrial strains being limited, recent population genetic studies provide the first steps in understanding how the *Saccharomyces* population is structured and stratified, how this evolved over time, and to what extent these processes are influenced by human interactions.

In 2005, Fay and coworkers analyzed the genetic diversity of 81 strains of *S. cerevisiae*, isolated from a variety of human and natural fermentation environments, as well as sources unrelated to alcoholic beverage production (such as tree exudates and immunocompromised patients), at five unlinked genetic loci. This study provided for the first time genetic evidence for domestication events, and established that diverged populations of wild *S. cerevisiae* exist independently of domesticated isolates. This latter observation was particularly interesting, since it was commonly believed that *S. cerevisiae* is a domesticated species with no truly natural strains existing, and that isolates from the wild were simply escaped industrial strains [125, 126]. The authors further concluded that all biodiversity of industrial *S. cerevisiae* strains could be traced back to (at least) two independent domestication events; one leading to the current wine, beer, and bread yeasts (most likely in Europe or the Middle East), and one leading to the current sake yeasts (in Asia). In a later study, Legras and coworkers confirmed this theory, but additionally hypothesized that since up to 28% of the genetic diversity within yeasts from the same industry was associated with geographical origin, strains were further domesticated locally in the past millennia [83].

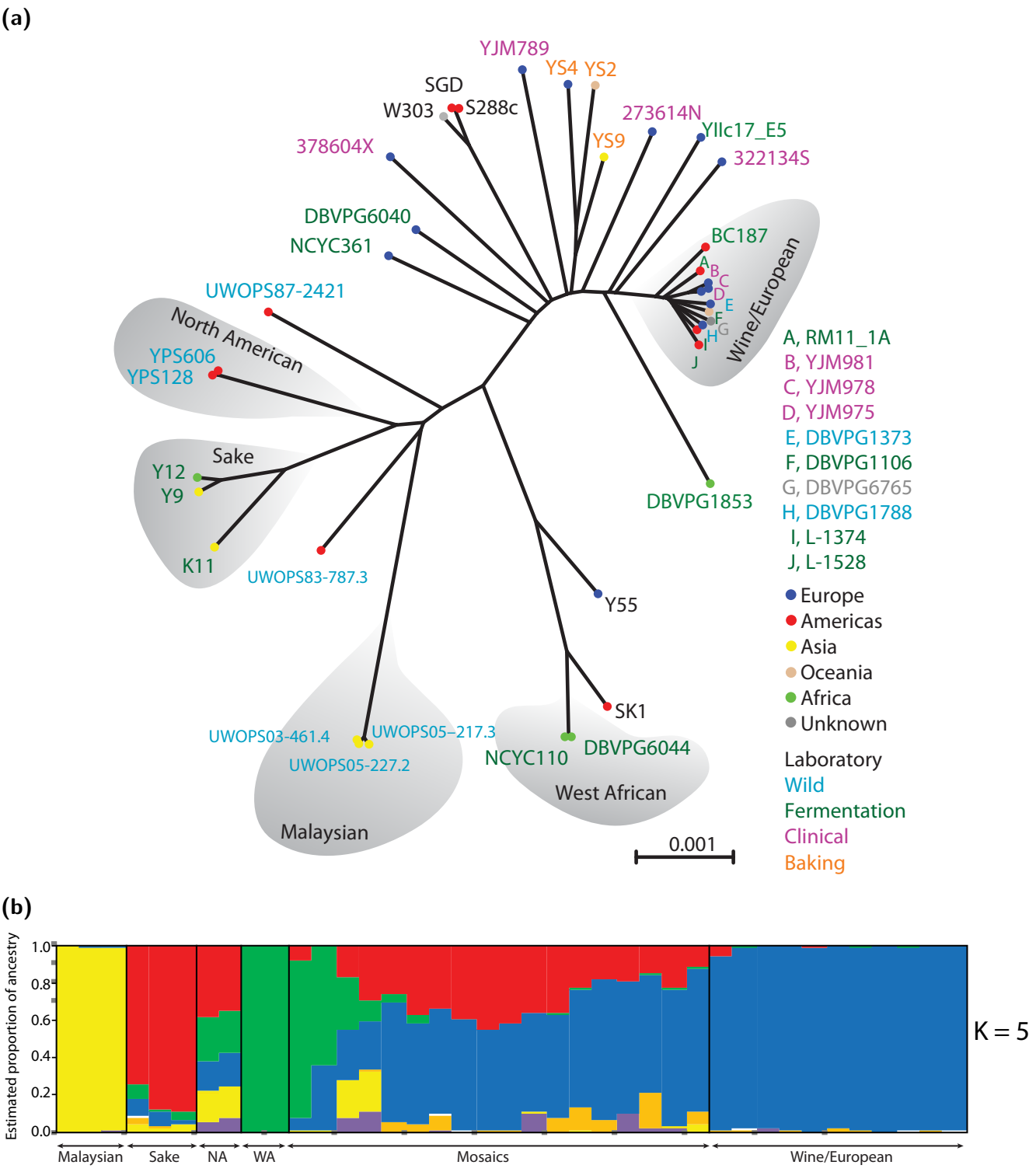
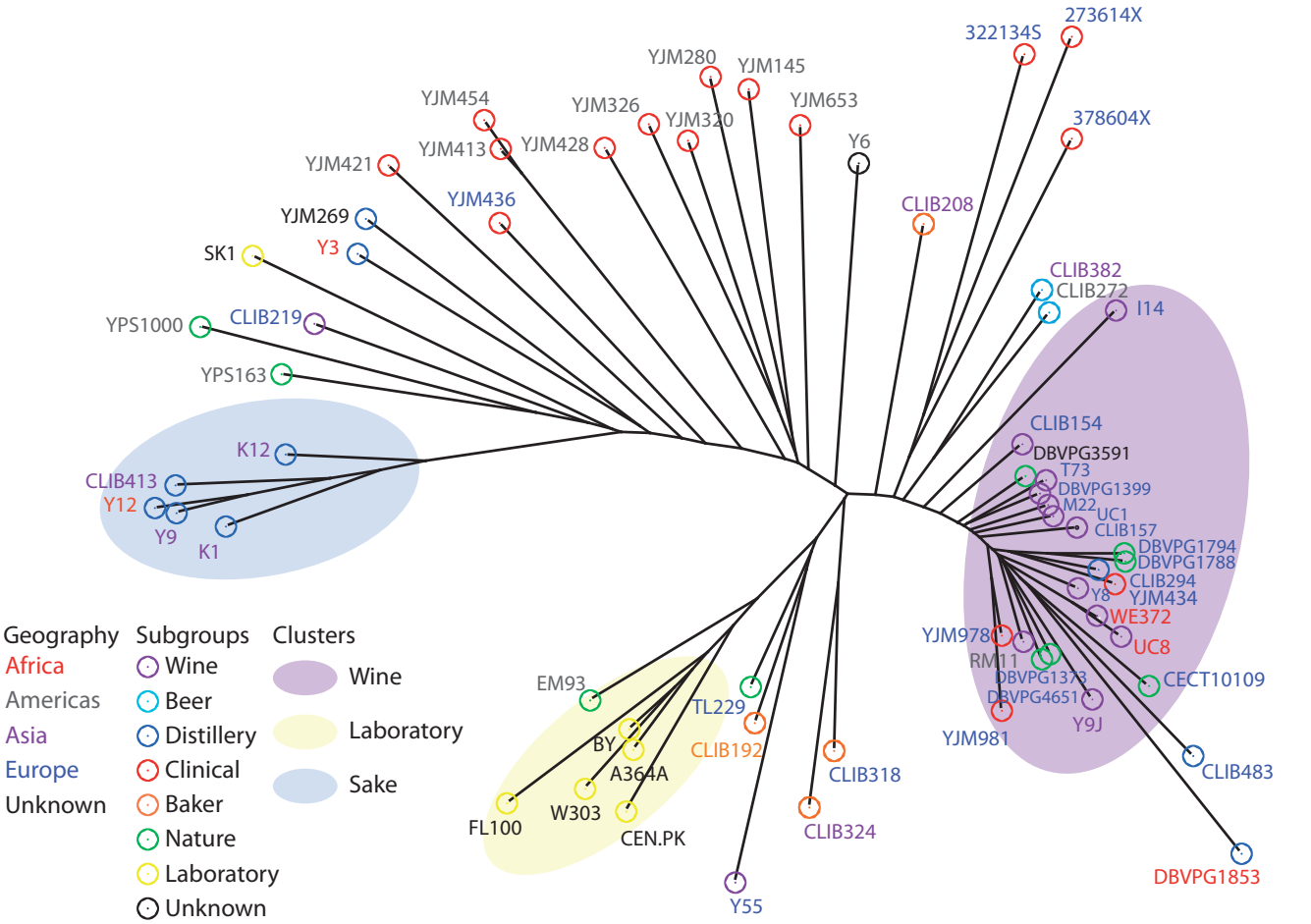


Figure 4.1: *Saccharomyces cerevisiae* phylogenetics and population structure. a: Neighbour-joining tree based on SNP differences of the *S. cerevisiae* strains sequenced in [127]. The five clean lineages are highlighted in grey, color denotes ecological niche (font) or geographic origin (dots); scale bar indicate frequencies of base-pair differences; **b:** Inferred population structure of *S. cerevisiae* strains sequenced in [127], based on SNP data. The most likely number of inferred subpopulations is five (estimated ancestry K=5). Each strain is represented by a single vertical bar which is partitioned into colored segments indicating the estimated coefficient of membership to each subpopulation. All panels are adapted from [127].

(a)



(b)

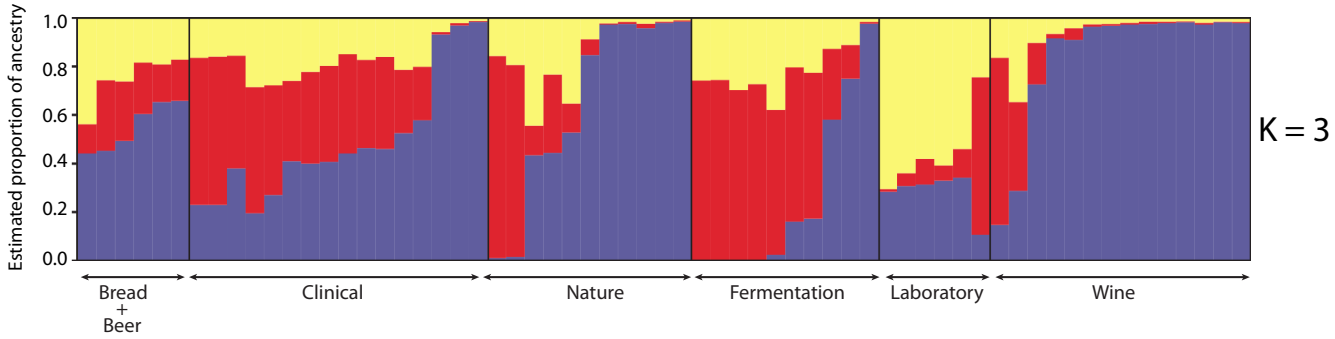


Figure 4.2: *Saccharomyces cerevisiae* phylogenetics and population structure. a: Neighbour-joining tree based on SNP differences of *S. cerevisiae* strains sequenced in [50]. The three main lineages are highlighted with colored ellipses. Color denotes geographic origin (font) or ecological niche (dots). **b:** Inferred population structure of *S. cerevisiae* strains sequenced in [50], based on SNP data. All panels are adapted from [50].

In 2009, two seminal papers showed for the first time genome-wide maps at the nucleotide level of large collections of *S. cerevisiae* isolates, sampled from diverse ecological (wine, beer, spirits, bread, immunocompromised individuals, soil, etc.) and geographical sources [50, 127]. The first study employed a high-density Affimetrix Yeast Tilling Microarray (YTM) to genetically characterize 63 diverse *S. cerevisiae* strains and compare them to the genome of the lab strain S288c [50]. The second study, that gave rise to the *Saccharomyces* Genome Resequencing Project (SGRP), used low-coverage (oneto four-fold) WGS to study the haploid derivatives of 36 *S. cerevisiae* and 35 *S. paradoxus* strains. The inclusion of *S. paradoxus* in this study is especially interesting, since this species is the closest known relative of *S. cerevisiae*, but has never been associated with industrial fermentation processes, and can thus be regarded as a completely non-domesticated species. In both studies, the authors could identify several clean subpopulations ("lineages") in *S. cerevisiae*. Liti et al., 2009, identified five (specific to geographic location or ecological niches; "Malaysian", "West African", "Sake", "North America", and "Wine/European") **(Figure 4.1a, p. 36)**, while Schacherer et al., 2009, identified three ("Sake", "Wine", and "Laboratory") **(Figure 4.2a, p. 37)**. Each clean lineage was monomorphic for the majority of segregating sites that were private within the lineage. However, it has to be noted that the strain collections used in these studies are not exhaustive, since intensive sampling revealed that at least eight more wild lineages are present in China alone [88]. Nevertheless, these studies provided valuable insight into general *S. cerevisiae* genetic diversity, population history, and evolution. First, it was established that most wine strains, although they were collected from dispersed geographical locations, were members of a single subpopulation in both studies. Recently, a new non-industrial sublineage has been identified within the wine group that includes Mediterranean oak isolates. Comparative genomics and demographic analysis suggested that this wild population included the ancestral genetic stock of today's wine yeasts [128]. Furthermore, the wine subpopulation showed in all studies a remarkably low level of polymorphisms and harbored an excess of low-frequency single-nucleotide polymorphisms (SNPs), all features that can be explained by a population bottleneck associated to a single domestication event in wine strains, followed by human direct or accidental dispersal around the world. In 2016 WGS of 212 *S. cerevisiae* strains, including 106 commercial wine starter isolates, corroborated previous findings and confirmed the presence of high level of inbreeding and substantial strain redundancy across the full catalogue of commercial wine strains [47]. As a result, the genetic diversity within these industrial strains is rather limited compared to the full spectrum of *S. cerevisiae* biodiversity; it was calculated that while the nucleotide diversity (π, the average number of nucleotide differences per site between any two DNA sequences chosen randomly from the sample population) was 0.56×10^{-3} in 14 representative wine yeasts, a more elaborate study of 138 strains (including both industrial and wild strains) revealed a sequence diversity that was more than one order of magnitude larger (7.27×10^{-3}) [88]. Interestingly, this diversity is much broader than observed in the human population, which is estimated to be "only" 1×10^{-3} [129]. Second, it became clear that there has been extensive transition of strains from one industry to

the other, as well as "back to nature" events after domestication [50]. For instance, the wine yeast lineage also included strains originating from distilleries, immunocompromised patients, and natural environments. Third, it became evident that in both studies the majority of strains (including many beer, bread, and clinical isolates) did not harbor the genome of one clean lineage, but rather included genomic fractions of various lineages, and can thus be considered as "recombinant" (mosaic) strains **(Figures 4.1b and 4.2b, p. 36 and 37, respectively)**. Interestingly, this peculiar population structure was not observed in the wild species *S. paradoxus*, supporting the idea that close association of *S. cerevisiae* with human activity facilitated crossbreeding of geographically isolated lineages and thus generation of new combinations of pre-existing variations, resulting in phenotypically divergent variants that were in turn spread across the globe. Whereas in theory these results support the hypothesis about two independent domestication events (as raised by Fay et al., 2005), Liti et al. (2009) also proposed an alternative scenario. They argue that it is equally possible that human-activity may have used existing natural strains with specific fermentation capabilities, providing opportunities for out-crossing and recombination between isolates originating from disparate environments [127, 130]. Unfortunately, beer yeasts are poorly represented in these studies, which are often biased towards natural and/or wine strains. Moreover, these studies also suffer from several technical constraints, which results in the loss of valuable genetic information. For example, both studies used S288c as a reference strain, either for genome mapping [127] or array design [50]. Therefore, in 2011, Borneman and coworkers documented the first de novo WGS of six industrial *S. cerevisiae* strains in their natural ploidy [45]. While this paper also mainly focused on wine strains (Lalvin QA23, AWRI769, Vin13, and VL3), two Australian ale-brewing strains (FostersO and FostersB) were included. These six strains were sequenced at high coverage (average 20-fold, compared to the one to four-fold in Liti et al., 2009), with a combination of shotgun and paired-end sequencing, which resulted in high-quality genomic assemblies. Phylogenetic analysis revealed that the wine and ale strains clustered in two separate populations; the wine strains clustered with the previously established "Wine/European" lineage, while the beer strains were not part of one of the previously defined clean lineages. However, since these clusters were shown to be much more closely related to each other than to strains with origin outside Europe, it was hypothesized that two distantly related, but both European *S. cerevisiae* lineages are the ancestors of the wine and beer subgroups.

In 2016, Gallone, Steensels and coworkers reported WGS and phenotyping of a large collection of 157 industrial yeasts including 102 commercial beer *S. cerevisiae* strains, providing a comprehensive insight into the evolution and diversification of brewing isolates [26] (see **Chapter 6 Domestication and Divergence of *Saccharomyces cerevisiae* Beer Yeasts**). Surprisingly, the pool of commercial *S. cerevisiae* beer isolates was not exclusively composed by ale-brewing strains, but also included 10 *S. cerevisiae* strains used for the industrial production of lager beers, of which it was assumed that they were solely brewed using strains of *S. pastorianus* **(see Section 4.5, p. 43)**. Phylogenetic analysis revealed the presence of 5 distinct lineages including only

industrial strains, which were genetically and phenotypically separated from wild strains. Wine and sake strains clustered in the well-established "Wine" and "Asia" lineages [127]. The majority of beer yeasts clustered in two only distantly related lineages, "Beer 1" and "Beer 2". The last industrial lineage, dubbed "Mixed" lineage, included all bread strains and a large portion of the "atypical" beer yeasts used for bottle refermentation of strong Belgian ales. These findings have the exciting implication that multiple domestication events are responsible for the current-day ale yeast biodiversity. Moreover, the large set of sequencing data allowed the authors to estimate the origin of these Beer yeast lineages. They found that the Beer 1 lineage originated around 1573-1604AD, while the Beer 2 lineage is more recent and originated about 50 years later. This means that beer yeast domestication started well after the first reported beer production (3000-4000 BC), but before the first yeast isolation by Emil Christian Hansen, or even the discovery of microbes by Louis Pasteur, in the 19th century. Interestingly, the calculated dates coincide with the gradual switch from home-centered beer brewing where every family produced their own beer, to more professional large-scale brewing, first in pubs and monasteries, and later also in breweries [97].

The paper further indicates that whereas the specific industrial niche (wine, beer, sake, bread, etc.) was shown to have a major influence in the diversification of industrial yeasts, the influence of geographical isolation was also observed. Most notably, the Beer 1 lineage consisted of three geographically distinct groups: "Belgium/Germany", "Britain" and "US". The Beer 2 lineage on the other hand, showed absence of geographical substructure and included strains originating from Eastern Europe, Belgium, Germany, UK and US. Maybe not surprisingly, subclustering based on beer-styles was generally not observed. This is in line with brewer's common practice of reusing the same yeast strain for a wide variety of beers. Remarkable exceptions to this trend were identified in strains used for the production of specialty beers traditional of a specific geographic area (such as Hefeweizen, English stout and Belgian Saison).

While industrial *S. cerevisiae* yeasts show clear hallmarks of domestication (see further), the exact origin and onset of domestication is not yet fully understood. In a recent study, Duan and coworkers suggested that one 'out-of Asia' event represents the origin of the current diversity of domesticated *S. cerevisiae* strains [29]. First, far East Asia and more specifically China harbors the largest genetic diversity not only of wild lineages [27, 88] but also of domesticated ones. Second, genetic analyses show that both European and Asian domesticates stem from wild lineages isolated in Chinese primeval forest. It still unclear, however, whether domestication started in Asia and the domesticated strains were later on introduced to Europe, or whether wild *Saccharomyces* found its way from Asia to Europe, and domestication was initiated independently for the European lineages. A thorough and systematic investigation of wild and *S. cerevisiae* lineages across Asia and Western Europe, and especially in regions that bridge the two, could further elucidate the onset of domestication. In 2018 the most comprehensive population genomic study on *Saccharomyces* was published, bringing the number of total sequenced strains to over 1000 and providing a genomic resource for comparative analyses of unprecedented levels. The study supported a single 'out-of-

China' event as the origin of the genetic diversity observed for this species, followed by several independent domestication events [27].

In conclusion, the overall picture gathered by recent studies shows how the genetic diversity of *S. cerevisiae* has been influenced by complex dynamics related to its close association to human technology through history, as well as by genetic drift and migration, leading to progressively genetically differentiated populations. While the history of commercial beer strains was obscure for a very long time, various recent papers [26, 28, 45] unraveled the interesting population structure and evolution of these industrial yeasts.

4.4.5 The complex genome structure of *S. cerevisiae* ale yeasts

The recent genetic analysis of ale beer yeasts in their natural ploidy and the unbiased, de novo mapping approach exposed various genetic elements of industrial yeast strains that remained concealed in previous investigations. Major alterations of ploidy, large chromosomal copy number variations (CNVs), intra-strain variation between homologous chromosomes, gene content variation across strains, and novel strain-specific gene clusters were identified, further emphasizing the complexity of industrial *S. cerevisiae* genomes [26, 28, 45].Many of these observations are likely a direct consequence of the lack of a sexual lifecycle in ale yeasts: continuous clonal (vegetative) reproduction, especially in a stressful environment, is known to lead to heterozygosity, gross chromosomal rearrangements and poly- and aneuploidies over time [131–133]. Wine and natural strains usually possess a fully functional sexual lifecycle and are mostly clean diploids. In contrast, ale strains are often obligate asexual or show little or no spore viability. For instance, Gallone, Steensels and coworkers observed that almost half of the strains from the Beer 1 lineage were unable to sporulate, while this trait was rarely detected in the other lineages. Furthermore, strains from the Beer 1 lineage that sporulated, produced in 80% of the cases only unviable spores textbfChapter 6 (Domestication and Divergence of *Saccharomyces cerevisiae* Beer Yeasts.

Partial and whole chromosome copy number modifications are frequently observed in laboratory strains of *S. cerevisiae* [134] as well as in industrial [135] and wild isolates [136, 137]. The extent of these modifications can vary greatly according to the origin of the strains. Ale-strains typically possess a more irregular genomic structure and many of them are polyploid or aneuploid [26, 83, 138]. Interestingly, laboratory strains often show growth deficiencies when forced to acquire extra copies of chromosomes, while non-laboratory (e.g., industrial) yeast strains are more tolerant to aneuploidy and display for some gene classes gene-dosage compensation effect at the transcriptional level [136]. Indeed, Borneman and coworkers for example, observed large chromosomal amplifications (a 600 kb region chrII, a 200 kb region of chrX, and the whole chrIII of FostersO, and the whole chrIII, chrV, and chrXV of FostersB) and deletions (a 400kb region in chrVII and the whole chrXIV of FostersO) in beer strains, but to a lesser extent in wine and wild strains.

Copy number variations (gene duplications or deletions) can have substantial phenotypic effects and emerging evidence describes aneuploidy/polyploidy as an evolutionary capacitor with a crucial role in the process of rapid adaptation to harsh environments and genetic perturbations [135, 139–141]. For instance, in many industrial strains, such variations are more frequent in specific regions such as sub-telomeric sites. For instance, sub-telomeric gene families regularly subjected to contraction or expansion include sugar utilization genes, vitamins metabolism, ion transport, stress resistance, DNA recombination, and regulation of meiotic cell cycle and reproduction **(see Section 4.4.3, p. 34)**. This suggests that in many industrial strains aneuploidies and copy number variations might be adaptive and/or linked to desirable industrial features (resistance to stress, high sugar content, extreme pH, nutrient limitations, etc.). One clear example of adaptive copy number variations is the strongly increased number of *MAL* genes, responsible for the utilization of maltose, in ale beer yeasts [26, 28].

Many industrial *S. cerevisiae* strains contain a high level of heterozygous SNPs compared to their wild counterparts. However, the proportion of these sites was shown to vary immensely. Beer yeasts typically exhibited a higher degree of heterozygosity compared to wine yeasts, with a 5.10-fold and 2.04-fold increase of heterozygous sites detected in strains from Beer 1 and Beer 2 lineages, respectively [26]. The level of intra-strain allelic differences (i.e., heterozygosity) is an important indicator of the life history and provides insights into the role of asexual vs sexual reproduction, mitotic recombination, and out-crossing [142]. The high level of (heterozygous) SNPs in beer yeasts compared to wine yeasts is in line with the "genome renewal" theory posed by Mortimer, in which cells with a fully functional sexual lifecycle (such as the majority of wine and wild strains) can eliminate deleterious mutations (accumulated during clonal reproduction) by one round of sporulation and self-mating [138, 143]. In this context, highly heterozygous strains could represent isolates that are less likely to undergo sexual cycles, preserving high degree of intra-specific variability. However, an alternative hypothesis suggests that heterozygosity can also be generated through (recent) out-crossing, and human activities might have favored recombination between strains with different genetic backgrounds. Indeed, recent population genomic studies **(see Section 4.4.4, p. 35)** demonstrated the presence of high degrees of admixture between lineages within the global population of *S. cerevisiae*, with many mosaic strains identified, specifically among industrial and clinical isolates [26, 28, 50, 51, 127].

4.4.6 Ale yeasts: probably the oldest microbial pets known to mankind

Taken together these findings shed a light on how the combination of life history and niche adaptation shaped the genome of *S. cerevisiae* ale yeasts. It becomes increasingly clear that today's industrial *S. cerevisiae* yeasts are genetically and phenotypically separated from wild stocks due to human selection and trafficking. Maybe even more interesting is that the thousands of industrial yeasts that are available today seem to stem from only a few ancestral strains that made their

way into food fermentations and subsequently evolved into separate lineages, each used for specific industrial applications. Within each cluster, strains are sometimes further subdivided along geographical boundaries, as is the case for the Beer 1 clade, which is divided into three main subgroups. However, it is important to note that strains from different fermentation environments experienced fundamentally different evolutionary paths. Continuous growth in man-made, rich beer medium led to large changes in the genome (aneuploidies, major chromosomal rearrangements, ...) and the loss of survival skills outside this specialized niche. This is in sharp contrast to for example wine yeasts, which experience the grape must environment only for a short period during the year and persist the rest of the time in and around vineyards or in gut of insects. Therefore, wine strains are exposed more often to natural, nutrient-poor environments, and have consequently characteristics which are more similar to the strains encountered in the wild. Moreover, these frequent "back to nature" events likely induced sporulation (sexual reproduction) and thus favoring hybridization with wild yeasts. In addition, the different common practices for wine and brewing industries have a strong influence on the effective population size of yeast populations and as a consequence, on the patterns of molecular evolution and variation. Because beer is produced throughout the year and beer yeasts are recycled for a few batches of fermentation each time, trillions of cells are transferred when a new batch is inoculated. By contrast, for wine yeasts only a relatively small amount of cells will contribute to the next harvest season grape must. This resulted in a high genetic diversity within beer yeasts compared to the more uniform wine yeast population. Together, this makes ale strains the most domesticated *S. cerevisiae* strains around, and explains why they are such a perfect fit for beer fermentations. However, this also implies that these yeasts fail to perform well in more innovative brewing conditions (for example "Very High Gravity" (VHG) brewing), as they encounter new stresses for which they are usually not adapted, while other strains from different lineages often are.

4.5 *Saccharomyces pastorianus* THE ODD ONE OUT OF THE *Saccharomyces* GENUS

Lagers currently account for more than 90% of the global beer market (http://www.statista.com/-statistics/270275/worldwide-beer-production/). They are typically fermented at a lower temperature (8 to 15°C), after which a period of cold storage (i.e., lagering, a traditional practice vital for sensorial quality) is performed. It is believed that the lager beer production process originally was introduced in the 15th century in Bavaria (Germany), when brewing became legally restricted to wintertime (at colder temperatures) to minimize the microbial spoilage of Bavarian beers. Later, the advent of refrigeration in the 19th century enabled lager brewing throughout the whole year [56, 62, 98]. Because of the high appreciation of this type of beer, it quickly spread around the globe. The yeasts used in this practice typically sink to the bottom (and not rise to the top) of the fermentation vessel towards the end of the fermentation, and are therefore often

called "bottom-fermenting yeasts". Interestingly, this bottom-fermenting phenotype was described very quickly after the dawn of lager beer brewing in Nuremberg, a town in the state of Bavaria [62, 97, 98, 144, 145]. In the 19th century, groundbreaking work by Louis Pasteur established that this phenomenon was caused by yeasts [22, 146–149], and the species name *S. pastorianus* was first coined by the German scientist Max Rees in 1870 as a tribute to Pasteur's work in the field. Emil Christian Hansen isolated shortly thereafter the first pure yeast cultures from lager beer fermentation during his work in the Carlsberg laboratory in Copenhagen. He classified the three isolated pure yeast lineages as separate species; one as *S. pastorianus*, one as *Saccharomyces carlsbergenesis* ("Unterhefe Nr. 1"), and one as *Saccharomyces monacensis* ("Unterhefe Nr. 2") [146, 147, 150]. After Hansen's findings, starter cultures became general practice in the breweries and bottom-fermenting yeast was classified as *S. carlsbergensis*, disregarding the earlier classification of bottom fermenting yeast as *S. pastorianus* by Max Rees. However, genetic analysis of the various isolates in 1985 showed that the type strains of *S. carlsbergensis*, *S. monacensis*, and *S. pastorianus* (the original isolates from Hansen) were almost identical, resulting in the reclassification of all bottom-fermenting yeasts to *S. pastorianus* [62, 98, 144, 145, 151, 152]. Furthermore, recent research has revealed that *S. pastorianus* is not a true species at all, but instead an interspecific hybrid of *S. cerevisiae* x *Saccharomyces eubayanus* **(see Section 4.5.1 below)**; nevertheless, in practice the species name *S. pastorianus* is still used to denote this lineage of interspecific hybrids. Because of its industrial importance, much research has been dedicated to the characterization of the lager yeast genome. This led to novel insights in the peculiar genome of this species, provided clues about its origin and shed light on some evolutionary processes that enabled this species to thrive in lager beer fermentation.

4.5.1 Hybrid nature of lager yeast

The physiology of lager yeasts differs fundamentally from the physiology of other brewing yeasts [22, 146, 149]. One of the most peculiar differences is the inability of *S. pastorianus* to sporulate and form viable spores, a property that is still present in many ale-type (and other *S. cerevisiae*) yeast strains [25, 98, 153, 154]. This inability to form viable offspring is a trait typically encountered in interspecific cross-breeding, e.g., mules (horse x donkey) or ligers (lion x tiger). Indeed, early genetic analysis showed that *S. pastorianus* harbored genetic material of (at least) two different species, and was thus not a clean yeast lineage, but rather the result of a hybridization event between *S. cerevisiae* and another (non-*cerevisiae*) *Saccharomyces* species. The first molecular evidence of the hybrid nature of lager yeast was obtained by a technique, called kar-mediated single chromosome transfer [155], and revealed that the chromosomes of lager yeasts could be divided into three types: i) homologous (*cerevisiae*-like) chromosomes, ii) homeologous (non-*cerevisiae* like) chromosomes, and iii) mosaic chromosomes, i.e. chromosomes composed of both homologous and homeologous segments. Later, these findings were confirmed by several DNA hybridization

experiments (e.g. by using southern blot or *S. cerevisiae*-specific gene arrays) [111, 156–158].

By comparing the DNA sequence of 11 independent loci, Dunn and Sherlock further suggested that the *S. cerevisiae* parent of *S. pastorianus* was closely related to ale-type *S. cerevisiae* strains, and not to wild isolates or strains used in other fermentation industries [111]. This result was in line with the previous work in which allelic variation in 12 microsatellite loci of 651 diverse *S. cerevisiae* and 15 *S. pastorianus* strains was investigated [83].

The first research suggesting a potential origin of the non-*cerevisiae* part of the lager genome was published in 1985 [72]. Using DNA-DNA hybridization methods, the authors revealed a similarity of 72% between the non-*cerevisiae* moiety of the lager yeast CBS1513 (at that time classified as *S. carlsbergensis*) and *S. bayanus*, a cold tolerant species commonly encountered in wine fermentation environments. This finding was later confirmed by PCR/RFLP analysis of 48 genes of the same lager yeast strain [159]. Interestingly, this hypothesis was later questioned when it was established that part of the *S. pastorianus MET2* gene showed a significant sequence difference with the corresponding *S. bayanus* sequence, suggesting a closely related *Saccharomyces* species, rather than *S. bayanus*, as the non-*cerevisiae* parent [160]. In 2009, the analysis of the first whole-genome sequence of a lager strain (Weihenstephan 34/70) confirmed this hypothesis. By aligning annotated Open Reading Frames (ORFs) of the lager brewing strain Weihenstephan 34/70 to annotated ORFs of *S. cerevisiae* S288c and *S. bayanus* CBS7001 reference genomes, the *cerevisiae*-type subgenome showed very high similarity to *S. cerevisiae* S288c (> 99 %) as opposed to the *bayanus*-type subgenome, exhibiting lower sequence identity with *S. bayanus* CBS7001 (average of 92.7%) [161]. Moreover, the authors identified eight genes in the lager yeast genome that were not present in the genomes of the *S. cerevisiae* yeast S288c nor in the genome of the *S. bayanus* yeast CBS7001, further indicating that probably a different, yet closely related species is the second lager yeast parent.

In 2011, Argentinean researchers sampling for cryotolerant *Saccharomyces* yeasts in the Patagonian forest stumbled upon a new *Saccharomyces* species [99]. The draft genome obtained through WGS of this species (dubbed *Saccharomyces eubayanus*), showed a remarkable high degree of similarity (99.56%) to the non-*cerevisiae* portion of the lager yeast genome, indicating that this species is very likely the missing link in the *S. pastorianus* origin. The authors further suggest a possible scenario where the initial hybridization event between a diploid *S. cerevisiae* cell and a diploid *S. eubayanus* cell gave rise to an allotetraploid hybrid (the original *S. pastorianus* strain), which was subsequently subjected to extensive genome rearrangement and mitotic recombination, resulting in loss of heterozygosity and recombinant chimeric chromosomes. Since these adaptations occurred in the highly selective and man-made environment of (lager) beer fermentation, they considered this the "domestication" of lager yeasts **(Figure 4.3, p. 47)**. However, several questions still remain unanswered. While *S. eubayanus* was originally discovered in Argentina, it is rather unlikely that lager yeasts originated in South America. Initially, Libkind and coworkers hypothesized that the South American *S. eubayanus* strain was introduced in Europe via the transatlantic travel

between Europe and America [99, 162]. However, while transatlantic travel was only established after Columbus' first voyage to the new world and the first reports of the bottom-fermenting phenotype predate Columbus' travels, lager brewing yeasts most likely originated earlier (probably in the early 1400s in Bavaria) [62, 97, 98, 144, 145]. More recent discoveries of genetically distinct lineages of *S. eubayanus* in other parts of the world (North America and China) suggest that *S. eubayanus* is not unique to South America [85, 162]. Moreover, genetic evidence suggests that the non-cerevisiae moiety of the Weihenstephan 34/70 is more closely related to a *S. eubayanus* lineage isolated in Tibet (sequencing of 12 loci indicated a 99.82% similarity to the non-cerevisiae moiety of lager yeast of the Asian *S. eubayanus* isolate, compared to the 99.56% similarity obtained by WGS with the Argentinean isolate described by [99]), suggesting that this lineage is more likely the direct ancestor of this lager yeast [85]. Therefore, it is now hypothesized that *S. eubayanus* made his way to Europe via the 2000-year-old Silk Road. Nevertheless, *S. eubayanus* is up till now not yet discovered in Europe and it could be that this yeast species occupies a highly specific niche in Europe and still awaits discovery [62].

However, the question of when and how *S. eubayanus* got into contact with the lager brewing environment remains open. It is hypothesized that this event took place five to six hundred years ago, triggered by a law enforcing brewing at cold temperatures in Bavaria. The non-*cerevisiae* parent (*S. eubayanus*) is thought to occured as a wild yeast contaminant around the brewing environment, and being better equipped to withstand the fermentation conditions in cold temperatures compared to the native ale yeasts. However, phenotypic analysis of the first two *S. eubayanus* strains isolated showed an inferior fermentation profile of *S. eubayanus* compared to *S. cerevisiae* (e.g., it is unable to ferment maltotriose, shows a lower ethanol tolerance, and produces an inferior aroma profile), prohibiting its use as starter cultures for lager beer fermentation. An interspecific hybridization event (between the *S. eubayanus* contaminant and the ale-type *S. cerevisiae*), that probably happened within the brewing tank, resolved the shortcomings of both species, and resulted in a hybrid species (*S. pastorianus*) that possessed the combined advantage of cold tolerance and fermentation capacity. This species was therefore able to outcompete its parental strains in lager beer fermentation, and was in this way (unintentionally) selected by the brewers, as traditionally part of the fermented beer was used to inoculate the next batch [62].

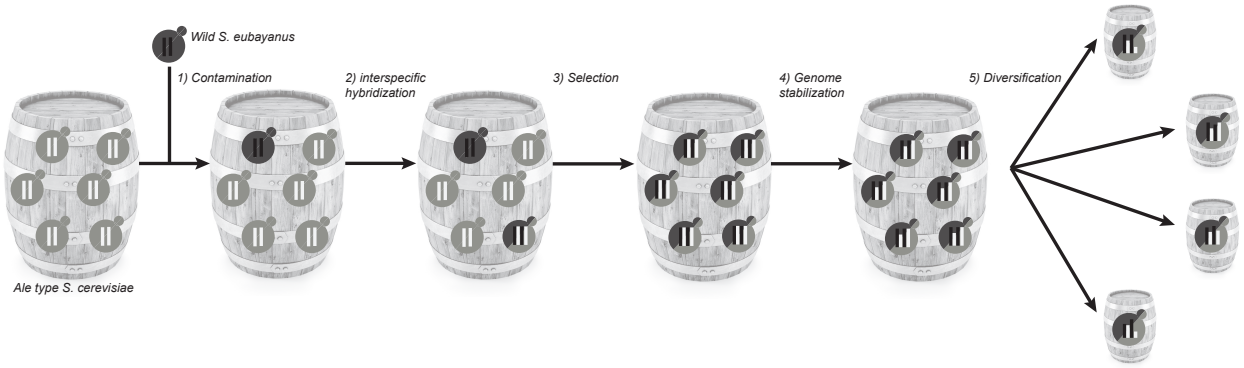


Figure 4.3: The origin of lager yeast. Current hypothesis about the origin of lager yeast involves (1) a contamination of ancient Bavarian fermentation (suggested to be originally conducted by ale yeast *S. cerevisiae*) by a wild *S. eubayanus* yeast contaminant; (2) a rare interspecific hybridization event between both yeast species; (3) selection of the interspecific hybrid due to its likely ability to combine the fermentation capacity and ethanol tolerance of its *S. cerevisiae* parent with the cold tolerance of its *S. eubayanus* parent; (4-5) followed by genome stabilization and diversification of the ancient *S. pastorianus* yeast to its current conformation. This whole process happened at least twice in history, giving rise to the two groups of lager yeasts (Saaz and Frohberg types).

4.5.2 Lager yeasts can be divided into two genetically and phenotypically distinct lineages

Over the years, the interspecific hybridization between *S. cerevisiae* and *S. eubayanus* yeasts probably occurred numerous times. However, to date, only two lineages of lager yeast remain present in industry. These two archetypes are referred to as "Saaz" or "group I" and "Frohberg" or "group II" type *S. pastorianus* yeasts [62, 111, 163–166]. These names refer to the work of Paul Lindner (1909), who reported the isolation of two individual *S. pastorianus* yeast lineages, which he named "Saaz" and "Frohberg" after the locations in Bohemia and Germany in which these strains were originally used [62, 164, 167]. Today, a clear trend exists between the lager yeast archetype and the country in which these yeasts are used. Group I (Saaz type) are mainly used in Czech breweries, as well as the Carlsberg brewery in Denmark, whereas Group II (Frohberg type) are more widespread in other European and North-American breweries [111]. While these archetypes show many similarities, there are some interesting genetic and phenotypic differences as well.

4.5.3 Genetic differences

The separation of *S. pastorianus* yeasts into two genetically distinct groups was firstly suggested based on RFLP analysis and the analysis of transposon sequence distribution among different *S. pastorianus* yeasts [165, 168]. Later, this hypothesis was confirmed by the use of array Comparative Genomic Hybridization (array CGH) and DNA sequencing of the adjacent and intronic regions of 11 intron-containing genes [111]. In the same experiments, it also became clear that

both archetypes did not share a common ancestry. Initially, it was predicted that Saaz-type yeasts originated from a haploid-haploid hybridization event, whereas Frohberg-type yeasts derived from a diploid-haploid hybridization event, where the diploid moiety originates from the *S. cerevisiae* parent [111, 144]. However, recent WGS analysis revealed that both Weihenstephan 34/70 (a Frohberg-type yeast) and CBS1513 (Unterhefe I, a Saaz-type yeast) are allotetraploid strains, both originating from a diploid-diploid rare mating event between *S. cerevisiae* and *S. eubayanus* [161, 166]. Interestingly, the two strains differed remarkably in the ratio in which both parental strains were retained in the genome: while Weihenstephan 34/70 harbored a tetraploid genome of 23.6Mb, with an approx. 1:1 ratio of both parental strains (36 different chromosome structures, 64 chromosomes in total), the genome of CBS1513 was much smaller (19.5Mb), and the *S. cerevisiae* parental strain was underrepresented compared to *S. eubayanus* (allotriploid (3n-1) with 29 different chromosome structures, 47 chromosomes in total). This interesting parental imbalance in the CBS1513 genome was shown to be caused by a large loss of the *S. cerevisiae* genome, including three complete *S. cerevisiae* chromosomes (VI, XI and XII). This is in line with previous findings based on PCR-RFLP [159] and array CGH [111]. Additionally, the CBS1513 genome was characterized by numerous large regions of loss of heterozygosity in chromosomes originating from *S. cerevisiae*, resulting in homozygous sequences derived from the *S. eubayanus* parent. Indeed, it was shown that approximately 1.44 Mb of *S. cerevisiae*-derived DNA got replaced by its *S. eubayanus* complement on four different chromosomes (chromosome IV, XIII, XV and XVI), whereas the opposite was only true for 0.22 Mb.

More recently, Okuno et al. (2016) analyzed in more detail the genetic composition of ten lager yeasts (five Saaz type: CBS1503, CBS1513, CBS1538, CBS1174, CBS2440 and five Frohberg type W34/70, CBS1483, CBS1484, CBS2156, CBS5832) using Illumina next generation-sequencing. The ploidy of the strains was estimated by mapping the obtained paired-end reads onto the genomes of both parental species. *Saccharomyces cerevisiae* derived chromosomes in Saaz type yeasts were haploid or missing, whereas most *S. eubayanus* derived chromosomes appeared to be diploid or triploid. Moreover, partial or whole deletions of *S. cerevisiae* derived chromosomes were observed frequently (e.g. deletions of the right arm of chromosome IV, left arm of chromosome XIII and the entire chromosome XII were common in all five sequenced Saaz type lager yeasts). In contrast, the genome of Frohberg type lager yeasts was composed of a haploid or diploid set of *S. cerevisiae* derived chromosomes whereas the ploidy of *S. eubayanus* derived chromosomes ranged from haploid to triploid. The observed chromosomal imbalance implied that the five Saaz type *S. pastorianus* strains, together with Frohberg type lager yeast CBS2156 can be regarded as being triploids, with a total genome size ranging from 14.4 Mb to 19.2 Mb and that Frohberg type lager yeasts (except for CBS2156) are tetraploids [169]. Analysis of the genome structure of *S. pastorianus* pinpoints to a complex evolutionary history and allows inference on the origin of the new species.

Throughout the years, there has been some uncertainty about whether both *S. pastorianus* archetypes originate from a single hybridization between *S. cerevisiae* and *S. eubayanus*, or if the

Saaz-type and Frohberg-type lager yeasts are derived from two independent hybridization events. Based on the recent identification of three shared translocation sites (*HSP82*, *XRN1*/*KEM1*, and *MAT*) between the Weihenstephan 34/70 and the CBS1513 yeast, it was suggested that both types of lager yeasts share a joint history and a common ancestor, backing up previous work in which the close relationship between Weihenstephan 34/70 and CBS1513 was appointed based on single gene analysis [166, 170]. More recent work rejected this hypothesis of a shared ancestor for both types of lager yeasts, and restored the hypothesis that the two lager yeast lineages originate from two separate hybridization events [171]. Monerawela's study argued that the Frohberg strains originated from a stout-type yeast, while the *S. cerevisiae* moiety of the Saaz strains had the highest similarity with Foster O-like ale strains (an Australian ale yeast with European roots).

However, the sequencing of ten lager yeasts by Okuno and coworkers identified novel shared interchromosomal translocations between *S. cerevisiae*-type and *S. eubayanus*-type genomes in both groups of lager yeasts, favoring the presence of at least one shared hybridization event [169]. For instance, this hypothesis was supported by the presence of two translocations in *HSP82* and *KEM1* loci, occurring in eight out of the ten tested lager yeasts (for yeasts CBS1538 and CBS1174 the *HSP82* locus got lost via a chromosomal deletion).

In addition, the mtDNA of *S. pastorianus* showed some interesting trends. In a study on 22 different lager yeasts by RFLP analysis (using four different restriction enzymes), all lager yeasts tested showed a similar uni-parental inheritance of the mtDNA of their non-*cerevisiae* parent [172]. These findings were later confirmed in the work of Dunn and Sherlock (where none of the 17 tested lager yeasts seemed to contain *S. cerevisiae*-derived mtDNA; [111]) and the WGS of Weihenstephan 34/70 [161]. This suggests that *S. eubayanus* mtDNA might harbor one or more genes that provide a competitive advantage for *S. pastorianus* in a lager beer environment. This theory finds further support in the work of Gonzalez and coworkers, who also observed a similar uni-parental inheritance of mtDNA in natural interspecific hybrids between *S. cerevisiae* and the cold-tolerant species *S. bayanus* and *S. kudriavzevii*, discovered in European wine fermentation environments [173]. However, the genes causing this trend are yet to be identified.

4.5.4 Phenotypic differences

The two lager yeast archetypes were originally described in 1909 [167], but the first systematic phenotypic screening of these groups was only published recently [62, 66, 166, 168]. In these studies, several remarkable differences were revealed. First, Saaz-type yeasts showed a higher growth capacity at 10°C than Frohberg-type yeasts, suggesting that Saaz-type yeasts harbor a higher tolerance towards cold temperatures [63, 166]. Second, Frohberg-type yeasts showed a faster fermentation profile and higher degree of attenuation in 22°P fermentation at 15°C [63], 14°P fermentation at 14°C [166], and 12°P fermentation at 16°C [66]. This remarkable difference was explained by the incapability of Saaz-type lager yeasts to efficiently metabolize maltotriose [63].

Third, Saaz-type yeasts showed lower cell viability and formed more respiration-deficient 'petite' cells at the end of fermentation. These phenotypes are generally undesired in lager yeasts, since lager yeasts are traditionally reused for seven to twenty-one consecutive fermentation batches, and thus require high phenotypic stability. Finally, the aroma profile of both lager yeast archetypes also differed significantly. In general, Frohberg-type yeasts produced higher concentrations of ethyl acetate, isoamyl acetate, and isoamyl alcohol, but less acetaldehyde [63, 66, 166]. These phenotypic differences may partly explain why Frohberg-type lager yeasts are generally preferred over Saaz-type lager yeasts in today's beer industry.

4.5.5 Future prospects of lager yeasts

The earlier described limited genetic and phenotypic diversity of lager yeasts (especially when compared to the immense genetic and aromatic diversity of ale *S. cerevisiae* yeast strains [25]) has inspired several researchers to develop new and more diverse lager yeasts (for the most recent review, see [62]). Moreover, the recent discovery and isolation of *S. eubayanus* lent the opportunity to develop new interspecific hybrids between *S. cerevisiae* and *S. eubayanus* in the lab, providing a powerful tool to generate strains with superior brewing properties [64, 66, 174]. Indeed, the development of interspecific hybrids has proven to be an efficient approach to develop novel yeast variants with enhanced characteristics for wine and beer-making. Typically, an industrial strain of *S. cerevisiae* yeast is crossed with a wild, non-cerevisiae member of the *Saccharomyces* sensu stricto complex, such as *S. bayanus* [175–178], *S. kudriavzevii* [179, 180], *S. uvarum* [181, 182], *S. mikatae* [183], or *S. eubayanus* [64, 66, 174]. The first attempt to generate new lager yeasts was established through pair-wise mass mating of ale type *S. cerevisiae* yeasts and a *S. bayanus* yeast strain. Resulting hybrids showed a higher fermentation capacity than both their respective parental strains, and some showed a similar fermentation capacity to their control bottom-fermenting yeast, already highlighting the potential of newly formed interspecific yeast hybrids for lager beer brewing [176]. More recently, two papers were published in which a similar, auxotrophic marker-assisted mass-mating strategy was used to generate four interspecific yeast hybrids between *S. eubayanus* type strain CBS12357 (wild isolate isolated in Patagonia; [99]) and *S. cerevisiae* strains IMK439 [64] or VTT-A81062 [174]. The hybrids inherited interesting properties of both parental strains (cold tolerance, maltotriose utilization, and strong flocculation) and showed hybrid vigor for several traits, such as fermenting speed and fermentation capacity in lab-scale lager beer fermentation tests [64, 174]. In a more elaborate study, a set of 31 novel interspecific yeast hybrids were developed, resulting from large-scale robot-assisted selection and breeding between six different *S. cerevisiae* and two different *S. eubayanus* strains, with the aim to increase the aromatic diversity in lager beers [66]. Many of these new hybrids produced an aromatic profile significantly different from those produced by currently available lager yeast in both lab and pilot scale fermentation tests, therefore providing a source of lager yeasts able to produce new, aromatically diverse lager beers.

4.6 *Brettanomyces bruxellensis* THE LOCOMOTIVE OF SPONTANEOUS BEER FERMENTATION

A yeast that received considerably less scientific attention compared to *Saccharomyces*, but which is strongly linked with the production process of specific beer styles, is *B. bruxellensis*. While the vast majority of beers are brewed by pure starter cultures of *S. cerevisiae* or *S. pastorianus* yeasts (see above), this species plays a key role in beer fermentation processes relying on a natural inoculum, such as the American Coolship Ales [184] and the lambic and gueuze beers produced in the surroundings of Brussels, Belgium [185, 186]. In these complex microbial matrices, *Brettanomyces* lives in perfect harmony with various other microbial groups, such as lactic acid bacteria, and accounts for many of the typical organoleptic characteristics of the beer. Moreover, the incredible potential and unique aromatic properties of this species are increasingly recognized, with more and more artisans adding it deliberately to their fermentations, either as a pure culture or in combination with more traditional brewing strains.

Interestingly, *Brettanomyces* yeasts were amongst the first yeasts ever to be isolated and described in detail. The etymological origin of the *Brettanomyces* genus lies in Great Britain, where it was first isolated in 1904 by Niels Hjelte Claussen, a younger colleague of the famous Emile Christian Hansen at the Carlsberg brewery. He retrieved the yeasts from strong English stock ales, and was inspired by this niche when naming his new discovery (Brettano = British, Myces = fungus). Claussen quickly recognized the peculiar flavor characteristics of these yeasts compared to more conventional *Saccharomyces* strains, and therefore protected the application of *Brettanomyces* in beer fermentations with a patent (UK patent GB190328184). Interestingly, this is the first patented microorganism in history.

One century later, however, the role of *Brettanomyces* in the food industry is confounded and ambiguous. While their presence is still imperative in some (mainly spontaneously fermented) beers, they are considered to be some of the worst spoilage microbes in wine, mostly due to their typical aroma profile, which can be described as 'burnt plastic', 'barnyard', 'horse sweat', and 'leather' (also referred to as the co-called "Brett character" **(Licker et al., 1999;Wedral et al., 2010)**. Mainly driven by their role as a villain in wine production, several research groups adopted this yeast for a more in-depth study. The advent of next-generation sequencing (see above) recently provided these researchers with the tools necessary for the in-depth analysis of *Brettanomyces* (population) genomics, yielding fascinating data on the genomic build-up of this evolutionarily intriguing yeast species. In this section, we will provide a concise overview of the current knowledge of the *Brettanomyces* genomics, evolution, and population structure. More elaborate studies on this yeast are given elsewhere [187–189].

4.6.1 Brettanomyces taxonomy and phylogeny

Over the years, a plethora of different *Brettanomyces* species were suggested and the names of these species were freely used in scientific publications. However, *Brettanomyces* taxonomy has been changed regularly in the past decades and there have been many reclassifications over the years, making direct comparisons between old and more recent papers often challenging. A first attempt to describe the *Brettanomyces* genus comprehensively was performed by Mathieu Custers in 1940. He defined a classification based on a few asexually reproducing (anamorphic) variants [190, 191]. A few decades later, in 1960, the formation of ascospores was observed in some strains and the genus *Dekkera* was introduced in the taxonomy as the teleomorphic (sexual) counterpart of Brettanomyces. In the first edition of their manual on yeast characteristics and identification, Barnett and co-workers described the following 9 *Brettanomyces* and *Dekkera* species: *Brettanomyces abstinens*, *Brettanomyces anomalus*, *Brettanomyces claussenii*, *Brettanomyces custersianus*, *Brettanomyces custersii*, *Brettanomyces lambicus*, *Brettanomyces naardenensis*, *Dekkera bruxellensis* and *Dekkera intermedia* [192]. Today, five species are described, based on molecular analysis of the genera: the anamorphs *B. bruxellensis*, *B. anomalus*, *B. custersianus*, *B. naardenensis*, and *Brettanomyces nanus*, with teleomorphs existing for the first two species, *D. bruxellensis* and *D. anomala*. However, according to the recent guidelines of the International Code of Nomenclature for algae, fungi, and plants (the Melbourne Code), fungal species should be assigned only a single valid name. Since the name *Brettanomyces* is well-known and used more commonly in the food and beverage industries, it will likely be prioritized over Dekkera [193]. Moreover, after the first description of spore formation, spores have not been reported again [188], suggesting the teleomorphic state is very rare (see further).

Despite that *Brettanomyces* spp. have several traits in common with *S. cerevisiae*, these two genera are genetically only distantly related [194]. Research papers focusing on the phylogeny of *B. bruxellensis* and related species remain scarce, but it is shown that *Brettanomyces* is part of a phylogenetic clade comprising methylotrophic yeast species of the genus *Ogataea*, such as *Ogataea angusta*, *Ogataea glucozyma*, *Ogataea parapolymorpha*, and *Ogataea polymorpha*, as well as other methylotrophs such as *Kuraishia capsulata*, *Candida boidinii*, and *Komagetaella pastoris* (previously *Pichia pastoris*), although the inclusion of the latter species in this clade is uncertain. This clade seems to have diverged from the "CTG-clade" (containing, e.g., *Candida albicans*, *Debaryomyces hansenii*, and *Scheffersomyces stipitis*) progenitor after sharing a common ancestor with *S. cerevisiae*. Using these phylogenetic relationships, it was calculated that *B. bruxellensis* and *S. cerevisiae* separated roughly estimated 200 million years ago [195]. Nevertheless, both species share several interesting traits. For example, they have independently acquired resistance to high ethanol concentrations, as well as the ability to produce ethanol even in the presence of oxygen (the Crabtree effect, see **Section 4.4.1, p. 31**), which enables both species to thrive in alcoholic fermentation processes (see further). This is a clear example of parallel evolution, since

these traits arose independently in different lineages that share a similar niche [195].

4.6.2 A first glimpse at the peculiar genome and population structure of *Brettanomyces bruxellensis*

To date, genome assemblies of six *B. bruxellensis* strains have been published, with several more on the way. Five of these six strains originate from the wine industry [46, 196–199], while one was isolated from lambic beer [200]. In 2007, Woolfit and coworkers published the first exploratory genome survey of the French wine spoilage strain CBS2499, providing a first glimpse of the surprising features of the *B. bruxellensis* genome (Woolfit et al., 2007). Initially, CBS2499 was believed to be haploid, but this was quickly debunked by a follow-up study in which a more in-depth de-novo assembly was performed, showing that the strain was actually diploid [197]. This latter study described a total assembly size of 13.4 MB and was able to identify 5600 genes, from which 75% could be functionally annotated. Around the same time, a second strain was sequenced (AWRI 1499), yielding a 12.7 MB assembly and 4969 predicted genes [196]. ST05.12/22, the only non-wine isolate sequenced to date, was shown to be 13.0 Mbp and 5255 genes were predicted [200]. Recently, the first South American isolate was sequenced [198], and Borneman and coworkers published an in-depth genomic comparison of four wine isolates: the previously sequenced AWRI 1499 and CBS2499 and two newly sequenced strains, AWRI 1608 and AWRI 1613 [46]. Analysis of the sequences revealed some interesting similarities of, but also peculiar differences between, the different strains.

First, it was shown that the ploidy of *B. bruxellensis* is variable: while CBS2499, AWRI 1613, and ST05.12/22 were shown to be diploid, haplotyping analysis revealed that AWRI 1499 and AWRI 1608 consist of two moderately heterozygous sets of chromosomes (a core, diploid fraction) and a third haploid set that is divergent [46, 200]. Fascinatingly, the variable haploid fraction was shown to be phylogenetically distant, suggesting that multiple independent hybridization events were involved in the emergence of these allotriploids. It is not yet clear if the divergent haploid set originates from a separate species or distantly related *B. bruxellensis* strains.

Second, it was shown that the karyotype varies drastically from strain to strain [201]. *B. bruxellensis* strains can contain between 4 and 9 chromosomes, and the size of these chromosomes can range from 1 to 6Mbp. This is peculiar, since chromosome configuration is usually well preserved among populations belonging to the same species; e.g. different strains of *S. cerevisiae* are collinear and consist of 16 chromosomes. However, since newly formed hybrid genomes tend to be very unstable (as often shown for *Saccharomyces* hybrids; [181, 182, 202]), mechanisms that drive genome stabilization could at least explain some of the extreme karyotype variability observed. Additionally, the karyotype variability could suggest that *B. bruxellensis* might employ frequent variations in chromosome structure to increase their genome variability and competitiveness. However, although genomic mutability is beneficial for the adaptability of the species, it can impede sexual

reproduction and drive speciation [203]. This could explain the absence (or very low frequency) of sexual reproduction in *B. bruxellensis* (see further).

Third, even though *Brettanomyces* did not undergo a whole genome duplication (WGD) event, many localized duplications (copy number variations) and even opposing copy number changes in the same region between strains, were observed [46, 200, 204]. Since gene duplication is a major source of new genes, it is a central factor influencing genome evolution [205, 206]. The observed CNVs were frequent in subtelomeric regions and often included genes involved in nutrient (sugar) metabolism, indicating that they might affect utilization of specific carbon sources [46]. However, a more in-depth study on the paralogs is required to draw strong conclusion.

4.6.3 *Brettanomyces* as the well-adapted fermentation scavenger

The occurrence of *Brettanomyces* in natural habitats is only sporadically described [207], but it has been shown that they frequently colonize man-made ecological niches, such as alcoholic fermentation processes (wine, beer, bioethanol, cider, etc), soft drinks, dairy products, kombucha, and sourdough [189]. A common thread in these niches is the presence of harsh environmental conditions that are lethal for many microbes: high ethanol concentrations, low pH, the absence of readily fermentable nitrogen and carbon sources, low oxygen, etc. While resistance to these stressors is not uncommon in microbes, there are few species that combine all of these traits, thus withstanding such challenging environments. In this section, we will further elaborate on how *Brettanomyces* evolved to become such a highly specialized fermentation organism and how it has acquired different mechanisms to outcompete, or rather outlive, its main fungal competitor in alcoholic fermentations, *S. cerevisiae*.

Comparative analysis of the *B. bruxellensis* genomes revealed some interesting properties that could be linked to their behavior and ecological niches. First, similar to *S. cerevisiae*, *B. bruxellensis* seems to have evolved a mechanism that allows them to accumulate and be highly tolerant to high concentrations of a toxic metabolite (ethanol), even in the presence of oxygen, a strategy that heavily contributes to their dominance over ethanol-sensitive microbes in sugar-rich environments, and is called the "make-accumulate-consume" strategy [92]. Recently, a study performed by Rozpedowska and coworkers [195] deciphered how *B. bruxellensis* has evolved this phenotype similarly to, but independently of, *Saccharomyces* yeasts: both lineages used the same strategy relying on global promoter rewiring to change the expression pattern of respiration-associated genes. Interestingly, *B. bruxellensis* seems to have evolved an additional strategy to outcompete other microbes. Besides ethanol, they are also capable of producing, accumulating, and later consuming acetic acid in aerobic conditions, and withstand the resulting low-pH environment. The (lineage-specific) duplication of oxidoreductase genes might explain the capacity to produce acetate under aerobic conditions [196, 197]. It is important to note, however, that not all *Brettanomyces* species share this trait. For example, *B. naardenensis*, which separated approximately 100 million years

ago from *B. bruxellensis*, is unable to grow in the absence of oxygen and their metabolism is completely respiratory (and thus no ethanol or acetic acid is formed), indicating that they are Crabtree-negative [195].

Second, it was shown that *B. bruxellensis* is well-equipped to withstand nutrient-poor environments. For example, many *B. bruxellensis* strains are equipped with a gene cluster containing a nitrate transporter, nitrate reductase, nitrite reductase, and two Zn(II)2 Cys6-type transcription factors, which enables the utilization of nitrate as a sole nitrogen source [46, 199, 200, 204]. This trait might provide an important fitness advantage over other species, such as *S. cerevisiae* (typically unable to utilize nitrate), in low-nitrogen environments like molasses, soft drinks, or late stages of beer and wine fermentations. Additionally, *B. bruxellensis* lacks (or only shows very limited) glycerol 3-phosphate phosphatase activity, resulting in no (or very low) glycerol production [191, 208]. This trait can be disadvantageous in certain conditions (e.g. it reduces growth speed in environments deprived from oxygen, since glycerol is an important factor in maintaining the redox balance in anaerobiosis), but it provides a competitive advantage over *S. cerevisiae* in nutrient-limiting environments (where energy efficiency is pivotal), since glycerol production is an energy-consuming process. All these adaptations to harsh, nutrient-limiting environments support the scavenging lifestyle of *B. bruxellensis*.

Third, gene content analysis revealed a relative enrichment in cell membrane-related genes compared to two genetically closely related species (*Pichia angusta* and *Komagetaella pastoris*) and *S. cerevisiae* [196]. Interestingly, many of the enriched membrane-related genes (e.g., *FIG2*, *FLO1*, *FLO5*, *FLO9, HKR1*, and *MUC1*) might be advantageous for survival in wine or beer matured in oak barrels, where they could mediate the adhesion of the cells to the internal wall of the barrel and protect them from washing out during high-pressure cleaning [60, 119].

Finally, *B. bruxellensis* shows some interesting trends in carbon-source utilization, which can sometimes even be variable amongst different isolates (see further). Compared to *S. cerevisiae*, the genome of *B. bruxellensis* shows a significant enrichment for membrane-associated nutrient transporters and genes involved in the metabolism of alternative carbon sources (such as chitin, N-acetylglucosamine, galactose, mannose, and lactose) [196, 199]. Moreover, it shows very efficient sucrose utilization, accounted for by the expression of a high-efficiency sucrose transporter (for which no homologues exist in *S. cerevisiae*). This trait might be key for the high competitiveness of *B. bruxellensis* in sucrose-based fermentations such as certain types of molasses [208, 209]. Additionally, *B. bruxellensis* shows a higher affinity for glucose in carbon-limiting conditions, possibly mediated by an orthologue of the Candida albicans HGT1 gene, encoding a high-affinity H+-symport glucose transporter [210].

In short, the behavior of *B. bruxellensis* shows strong similarities with *Saccharomyces* species, which is reflected in the numerous niches that they share. However, *B. bruxellensis* seems to have evolved an extensive specialization in more challenging conditions (nutrient-poor environments with high ethanol and low pH), which are for example created during the course of beverage

fermentation processes.

4.6.4 Intraspecific *Brettanomyces bruxellensis* variability: adaptations to wine and beer environment?

Genetic diversity studies using classic DNA fingerprinting techniques have revealed significant genotypic intraspecific variability of *B. bruxellensis* **(Figure 4.4, p. 57)** [200, 211–215]. Moreover, much like in *S. cerevisiae*, some of these studies reported a correlation between genotype groups of *B. bruxellensis* and their source of isolation (e.g. beer or wine) [200, 211, 215], suggesting niche adaptation. Furthermore, recently, this correlation was also established for *B. bruxellensis* phenotypes [204]. In this latter study, the authors performed high-throughput phenotyping experiments (using Biolog Phenotype Microarrays), and identified several consistent differences between strains from different origins, such as wine and beer fermentations. For example, the ability to assimilate particular α and β-glycosides as well as α and β-substituted monosaccharides was shown to be highly variable amongst isolates, but consistent within strains from the same origin. While strains isolated from wine were able to utilize D-galactose, this is not the case for beer isolates. Coinciding with these observations, strains unable to grow on galactose were found to lack at least one of the genes involved in the Leloir pathway of galactose metabolism. Further, brewing strains are, in contrast to wine strains, not capable of using the β-glycoside disaccharides cellobiose and gentiobiose as well as the β-substituted monosaccharides β-methyl-glucoside and arbutine, suggesting that these strains lack the enzyme(s) responsible for the breakage of specific β-bonded sugars. Indeed, WGS revealed that while wine strains contain two (distinct) β-glucosidase genes, beer strains lack one of these genes [204]. Interestingly, these β-glucosidases are industrially very relevant, since they also play an important role in flavor development of beer and wine [216]: they enable the hydrolysis of locked, glycosidically bound flavor compounds and can thus enrich the flavor profile of these fermented beverages [217]. However, further research is needed to investigate the exact role of these β-glucosidases in flavoring capability of *B. bruxellensis* strains.

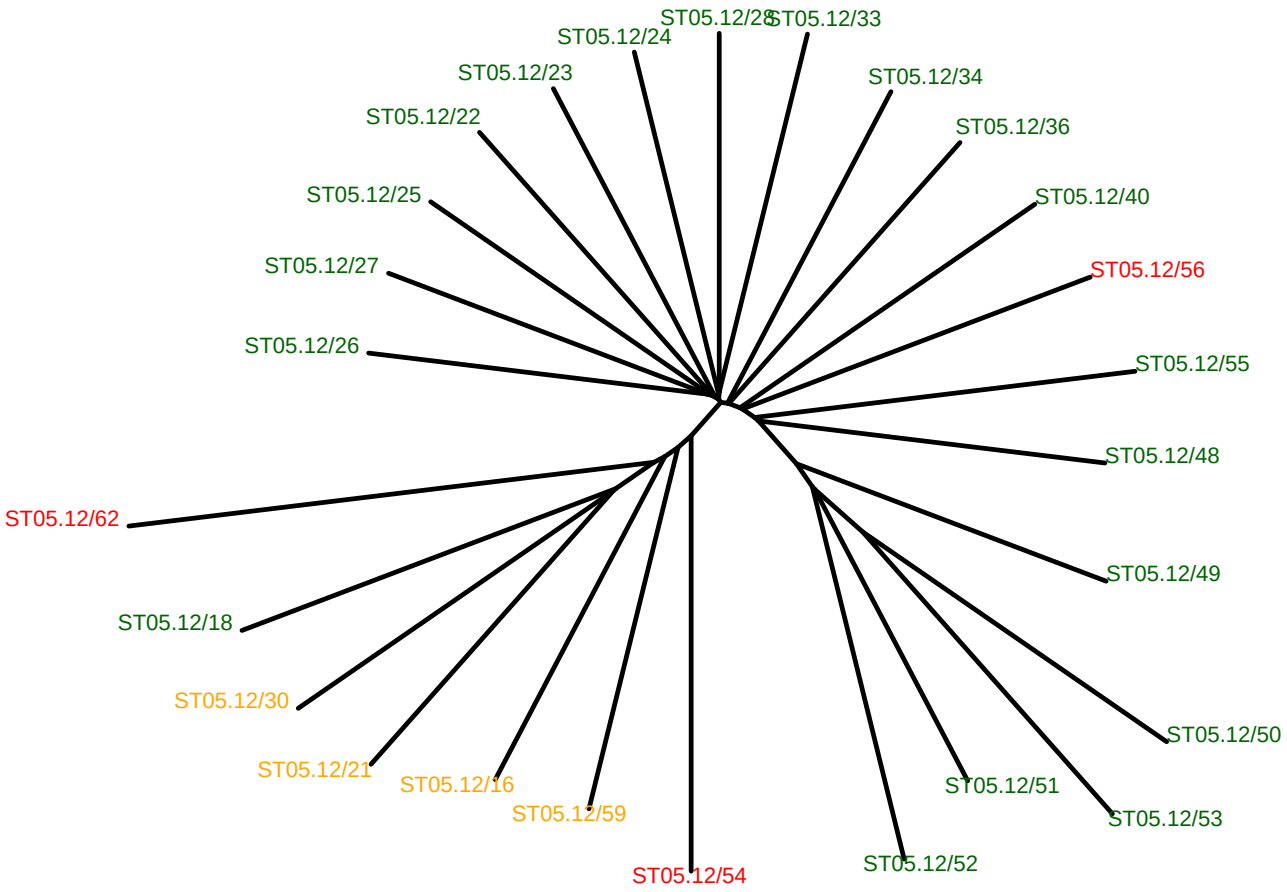


Figure 4.4: *Brettanomyces bruxellensis* phylogenetics. Neighbour-joining tree of *B. bruxellensis* strains from diverse ecological niches, including beer (indicated in green), soft drinks (orange), and wine (red). Studied strains were genotyped using seven established DNA fingerprinting techniques. Data analysis was performed on the combined dataset as described in [200]

Interestingly, the genes involved in the Leloir pathway as well as the above-mentioned β-glucosidase genes are clustered in a 36 kb region encompassing 13 genes, the majority of which are involved in carbon metabolism. This region was found to be completely absent in the beer strain ST05/12.22 [200]. Moreover, a more thorough investigation revealed that this cluster has been gradually lost over time in beer strains: some lack only a few genes, other lack all 13 genes, but all beer strains lack the β-glucosidase gene. In contrast, this cluster of genes was entirely present in wine strains (at least in eight of the nine studied strains) **(Tables 4.3 and 4.4, p. 58 and 59, respectively)** [204]. Interestingly, this genomic region is also prone to copy number variations and loss-of-heterozygosity [204]. Based on these findings, it may be speculated that this gene cluster carries a fitness cost for *B. bruxellensis* in certain fermentation systems such as beer brewing, thereby providing a selective pressure for its loss. These observations are reminiscent of the concerted loss of the galactose catabolism cluster in Japanese *Saccharomyces kudriavzevii* isolates compared to European isolates [218]. However, further research is required to draw firm

conclusions.

It is also interesting to note the possibility that the ploidy level may be linked to the yeast's ecological niche. More particularly, triploidy seems to be predominant in the Australian *B. bruxellensis* population, since it is observed in 92% of all isolates from Australian wines [46, 219]. Moreover, microsatellite typing hints towards the existence of similar populations in French and South African wine isolates [220]. In contrast, the majority of *B. bruxellensis* beer strains are found to be diploid [204]. This interesting allotriploid genome structure is not rare in fungi and is evocative of the interspecific hybrids identified in the *Saccharomyces* sensu stricto clade, such as the lager yeast *S. pastorianus* and the *S. cerevisiae*/*S. kudriavzevii* hybrids isolated from wine and ale beer fermentations [173, 221]. In the case of *Saccharomyces* hybrids, it was suggested that the additional set of chromosomes confers a selective advantage in an industrial environment (see above), but it remains to be determined whether similar evolutionary driving forces are at play in the allotriploid *B. bruxellensis* strains. It was suggested that the ability of wine strains to withstand high levels of sulphite, the main antimicrobial agent in wine fermentations, might be at least partially explained by the triploidy state [46, 196], but further research is required.

Table 4.3: Distribution of 13 clustered genes[a] across different *Brettanomyces bruxellensis* strains originating from beer[b,c]. **[a]:** Clustered genes in a region of 36 kb, corresponding to a part of *B. bruxellensis* AWRI 1499 (ST05.12/62) contig AHIQ01000280. **[b]:** Determined by PCR amplification using primers targeting the almost complete ORF, followed by agarose gel electrophoresis. Grey, target band detected; black, target band not detected. **[c]:** Adapted from [204].

GenBank Accession N°	Function	22c	33	25	27	34	23	26	48	49	50	51	52	53	24	28	40	18
EIF45404	Mfs drug transporter																	
EIF45405	Putative mfs-mdr transporter																	
EIF45406	Mfs multidrug																	
EIF45407	High-affinity glucose transporter																	
EIF45408	Galactose-1-phosphate uridylyltransferase																	
EIF45409	Galactokinase																	
EIF45410	Gal10 bifunctional protein																	
EIF45411	Dtdp-glucose -dehydratase																	
EIF45412	Hexose transporter																	
EIF45413	Maltase																	
EIF45414	Multidrug resistance regulator 1																	
EIF45415	β-glucosidase																	
EIF45416	Hexose transporter																	

Table 4.4: Distribution of 13 clustered genes[a] across different *Brettanomyces bruxellensis* strains originating from wine[b,c]. [a]: Clustered genes in a region of 36 kb, corresponding to a part of *B. bruxellensis* AWRI 1499 (ST05.12/62) contig AHIQ01000280. **[b]:** Determined by PCR amplification using primers targeting the almost complete ORF, followed by agarose gel electrophoresis. Grey, target band detected; black, target band not detected. **[c]:** Adapted from [204].

GenBank Accession N°	Function	70	54	56	62	63	66	67	68	69
EIF45404	Mfs drug transporter									
EIF45405	Putative mfs-mdr transporter									
EIF45406	Mfs multidrug									
EIF45407	High-affinity glucose transporter									
EIF45408	Galactose-1-phosphate uridylyltransferase									
EIF45409	Galactokinase									
EIF45410	Gal10 bifunctional protein									
EIF45411	Dtdp-glucose -dehydratase									
EIF45412	Hexose transporter									
EIF45413	Maltase									
EIF45414	Multidrug resistance regulator 1									
EIF45415	β-glucosidase									
EIF45416	Hexose transporter									

4.6.5 *Brettanomyces* GENOMICS: CATCHING UP WITH *Saccharomyces*

All in all, our knowledge of *Brettanomyces* is still very limited compared to the vast amount of data available for *S. cerevisiae*, the model fungus par excellence. However, several research groups currently are trying to bridge this gap and unravel the peculiar genomic properties of this industrially highly relevant yeast. However, several important challenges remain.

First, despite the fact that some research groups reported ascospore formation (albeit at low frequency) of *B. bruxellensis* (e.g. CBS74 and CBS4914; [222]) and *B. anomalus* (e.g. CBS8139; [223]) strains, it still remains unclear to what extent *Brettanomyces* can reproduce sexually, how often this occurs, and what the influence of these events is on the evolution of the species. The extreme karyotypic variability and frequently observed allotriploidy might hint towards the lack (or at least the rarity) of sexual reproduction. In line with this theory, an extensive genetic investigation of the *Brettanomyces* population within a winery (spanning several vintages) revealed that the same persistent clonal *Brettanomyces* population was responsible for wine spoilage in this winery for over half a century [220].

Second, the genetic and molecular toolbox for *Brettanomyces* should be expanded in order to facilitate more in-depth genetic analysis. For example, an optimal transformation protocol is still lacking, auxotrophic mutants are scarce, and only a limited amount of dedicated commercial tools, such as plasmid vectors or gene arrays, are available [188]. However, some initial attempts to develop molecular tools and protocols were conducted. Miklenić and coworkers developed various transformation protocols (based on spheroplast transformation, electroporation, and the LiAc/PEG method) and described the first (reported) genetic transformation of *B. bruxellensis*, using a kanMX4 marker cassette that allowed for selection on agar plates containing geneticin [224]. Additionally, Schifferdecker and colleagues developed a URA3-deficient mutant strain and

a plasmid (named P892, derived from pUC57) containing a functional *URA3* gene of CBS2499 [188]. Using these tools, an auxotrophic transformation system and an expression vector was developed, allowing overexpression of individual genes with a constitutive TEF1 promotor [225]. While commercial microarray kits for *Brettanomyces* are still lacking, a first large-scale transcriptomic analysis of *B. bruxellensis* was recently performed [208]. Using whole transcriptome sequencing, the authors managed to detect the expression of 3715 out of the 4861 annotated genes of CBS11270, and the results elucidated survival strategies of this yeast in harsh environments. For example, they observed a low expression of genes involved in glycerol production and a high number of expressed sugar transporter genes, two mechanisms possibly accounting for the high competitiveness of *B. bruxellensis* in oxygen-limited and nutrient-deprived environments [208].

Third, to fully grasp the *Brettanomyces* population structure and evolutionary history, analysis of a broader collection of ecologically and geographically diverse strains is necessary. While the main focus thus far is put on wine spoilers (mainly originating from Australia), preliminary investigation of strains from different niches and geographical origins already revealed a large degree of intraspecific (phenotypic and genetic) diversity, and sometimes revealed links between phenotypic behavior and source of isolation [200, 204]. While some of these correlations might be due to genetic drift and shared population histories of strains originating from the same environment, some might reflect adaptive mechanisms developed in industrial niches. However, more elaborate studies, targeting the genetic and phenotypic characterization of large collections of diverse strains, will yield more information on some fundamental questions, and shed light on the currently obscure *Brettanomyces* evolutionary history. For example, it will be interesting to see if (and to what extent) *Brettanomyces* is domesticated, and how the *Brettanomyces* population is structured.

4.7 Conclusions

With their small and compact genome, short replication time and easy handling in the lab, yeasts pose the ultimate model organism for genetics and population genomics. However, surprisingly, the number of sequenced yeast strains currently available is still limited, especially compared to other eukaryotic model organisms, such as the fruit fly, *Arabidopsis thaliana* and humans, which have more complex and much larger genomes [226]. However, current technological developments provide the opportunity to obtain whole-genome sequences for just a few 100 dollars, resulting in a recent burst of yeast genomics data [45–47, 51, 128, 169, 227, 228]. However, these studies only cover the tip of the iceberg of yeast biodiversity, mostly focusing on wild and clinical isolates of *S. cerevisiae*. Therefore, future studies that include a larger number of yeast strains associated with beer brewing (including non-cerevisiae species such as *B. bruxellensis*), will further elucidate the evolutionary path these yeasts went through during their association with human-made fermentation environments. Furthermore, these studies will give valuable insights in genotype-phenotype associations, providing opportunities for targeted strain improvement strategies, such

as marker-assisted breeding.

4.8 References

[1] O. Seehausen, R. K. Butlin, I. Keller, C. E. Wagner, J. W. Boughman, P. A. Hohenlohe, C. L. Peichel, G.-P. Saetre, C. Bank, Å. Brännström, A. Brelsford, C. S. Clarkson, F. Eroukhmanoff, J. L. Feder, M. C. Fischer, A. D. Foote, P. Franchini, C. D. Jiggins, F. C. Jones, A. K. Lindholm, K. Lucek, M. E. Maan, D. A. Marques, S. H. Martin, B. Matthews, J. I. Meier, M. Möst, M. W. Nachman, E. Nonaka, D. J. Rennison, J. Schwarzer, E. T. Watson, A. M. Westram, and A. Widmer, "Genomics and the origin of species," *Nature Reviews Genetics*, vol. 15, no. 3, pp. 176–192, 2014.

[2] B. J. Shapiro, J.-B. Leducq, and J. Mallet, "What Is Speciation?," *PLOS Genetics*, vol. 12, no. 3, p. e1005860, 2016.

[3] G. P. Wagner, "The Changing Face of Evolutionary Thinking," *Genome Biology and Evolution*, vol. 5, no. 10, pp. 2006–2007, 2013.

[4] J. W. Bull and M. Maron, "How humans drive speciation as well as extinction.," *Proceedings. Biological sciences*, vol. 283, no. 1833, p. 20160600, 2016.

[5] C. D. Thomas, "Rapid acceleration of plant speciation during the Anthropocene," *Trends in Ecology & Evolution*, vol. 30, no. 8, pp. 448–455, 2015.

[6] R. M. A. Packer, A. Hendricks, M. S. Tivers, and C. C. Burn, "Impact of Facial Conformation on Canine Health: Brachycephalic Obstructive Airway Syndrome," *PLOS ONE*, vol. 10, no. 10, p. e0137496, 2015.

[7] L. O. Fiems, S. de Campeneere, W. van Caelenbergh, and C. V. Boucqué, "Relationship between dam and calf characteristics with regard to dystocia in Belgian Blue double-muscled cows," *Animal Science*, vol. 72, no. 2, pp. 389–394, 2001.

[8] L. Grobet, L. J. Royo Martin, D. Poncelet, D. Pirottin, B. Brouwers, J. Riquet, A. Schoeberlein, S. Dunner, F. Ménissier, J. Massabanda, R. Fries, R. Hanset, and M. Georges, "A deletion in the bovine myostatin gene causes the double–muscled phenotype in cattle," *Nature Genetics*, vol. 17, no. 1, pp. 71–74, 1997.

[9] R. Kambadur, M. Sharma, T. P. Smith, and J. J. Bass, "Mutations in myostatin (GDF8) in double-muscled Belgian Blue and Piedmontese cattle," *Genome research*, vol. 7, no. 9, pp. 910–6, 1997.

[10] P. Wiener and S. Wilkinson, "Deciphering the genetic basis of animal domestication," *Proceedings. Biological sciences*, vol. 278, no. 1722, pp. 3161–70, 2011.

[11] A. Shomura, T. Izawa, K. Ebana, T. Ebitani, H. Kanegae, S. Konishi, and M. Yano, "Deletion in a gene associated with grain size increased yields during rice domestication," *Nature Genetics*, vol. 40, no. 8, pp. 1023–1028, 2008.

[12] A. Fita, A. Rodríguez-Burruezo, M. Boscaiu, J. Prohens, and O. Vicente, "Breeding and Domesticating Crops Adapted to Drought and Salinity: A New Paradigm for Increasing Food Production," *Frontiers in plant science*, vol. 6, p. 978, 2015.

[13] C. A. Driscoll, D. W. Macdonald, and S. J. O'Brien, "From wild animals to domestic pets, an evolutionary view of domestication," *Proceedings of the National Academy of Sciences of the United States of America*, no. Supplement 1, pp. 9971–8, 2009.

[14] O. Raymond, J. Gouzy, J. Just, H. Badouin, M. Verdenaud, A. Lemainque, P. Vergne, S. Moja, N. Choisne, C. Pont, S. Carrère, J.-C. Caissard, A. Couloux, L. Cottret, J.-M. Aury, J. Szécsi, D. Latrasse, M.-A. Madoui, L. François, X. Fu, S.-H. Yang, A. Dubois, F. Piola, A. Larrieu, M. Perez, K. Labadie, L. Perrier, B. Govetto, Y. Labrousse, P. Villand, C. Bardoux, V. Boltz, C. Lopez-Roques, P. Heitzler, T. Vernoux, M. Vandenbussche, H. Quesneville, A. Boualem, A. Bendahmane, C. Liu, M. Le Bris, J. Salse, S. Baudino, M. Benhamed, P. Wincker, and M. Bendahmane, "The Rosa genome provides new insights into the domestication of modern roses," *Nature Genetics*, vol. 50, no. 6, pp. 772–777, 2018.

[15] C. Yan, P. Lin, T. Lyu, Z. Hu, Z. Fan, X. Li, X. Yao, J. Li, and H. Yin, "Unraveling the Roles of Regulatory Genes during Domestication of Cultivated Camellia: Evidence and Insights from Comparative and Evolutionary Genomics," *Genes*, vol. 9, no. 10, 2018.

[16] I. Lazaridis, D. Nadel, G. Rollefson, D. C. Merrett, N. Rohland, S. Mallick, D. Fernandes, M. Novak, B. Gamarra, K. Sirak, S. Connell, K. Stewardson, E. Harney, Q. Fu, G. Gonzalez-Fortes, E. R. Jones, S. A. Roodenberg, G. Lengyel, F. Bocquentin, B. Gasparian, J. M. Monge, M. Gregg, V. Eshed, A.-S. Mizrahi, C. Meiklejohn, F. Gerritsen, L. Bejenaru, M. Blüher, A. Campbell, G. Cavalleri, D. Comas, P. Froguel, E. Gilbert, S. M. Kerr, P. Kovacs, J. Krause, D. McGettigan, M. Merrigan, D. A. Merriwether, S. O'Reilly, M. B. Richards, O. Semino, M. Shamoon-Pour, G. Stefanescu, M. Stumvoll, A. Tönjes, A. Torroni, J. F. Wilson, L. Yengo, N. A. Hovhannisyan, N. Patterson, R. Pinhasi, and D. Reich, "Genomic insights into the origin of farming in the ancient Near East," *Nature*, vol. 536, no. 7617, pp. 419–424, 2016.

[17] R. Avni, M. Nave, O. Barad, K. Baruch, S. O. Twardziok, H. Gundlach, I. Hale, M. Mascher, M. Spannagl, K. Wiebe, K. W. Jordan, G. Golan, J. Deek, B. Ben-Zvi, G. Ben-Zvi, A. Himmelbach, R. P. MacLachlan, A. G. Sharpe, A. Fritz, R. Ben-David, H. Budak, T. Fahima, A. Korol, J. D. Faris, A. Hernandez, M. A. Mikel, A. A. Levy, B. Steffenson, M. Maccaferri, R. Tuberosa, L. Cattivelli, P. Faccioli, A. Ceriotti, K. Kashkush, M. Pourkheirandish, T. Komatsuda, T. Eilam, H. Sela, A. Sharon, N. Ohad, D. A. Chamovitz, K. F. X. Mayer, N. Stein, G. Ronen, Z. Peleg, C. J. Pozniak, E. D. Akhunov, and A. Distelfeld, "Wild emmer genome architecture and diversity elucidate wheat evolution and domestication.," *Science (New York, N.Y.)*, vol. 357, no. 6346, pp. 93–97, 2017.

[18] K.-I. Tanno and G. Willcox, "How fast was wild wheat domesticated?," *Science (New York, N.Y.)*, vol. 311, no. 5769, p. 1886, 2006.

[19] T. R. Gregory, "Artificial Selection and Domestication: Modern Lessons from Darwin's Enduring Analogy," *Evolution: Education and Outreach*, vol. 2, no. 1, pp. 5–27, 2009.

[20] D. K. Belyaev, "Domestication of animals," *Sci. J.*, vol. 5, pp. 47–52, 1969.

[21] A. V. Kukekova, J. L. Johnson, X. Xiang, S. Feng, S. Liu, H. M. Rando, A. V. Kharlamova, Y. Herbeck, N. A. Serdyukova, Z. Xiong, V. Beklemischeva, K.-P. Koepfli, R. G. Gulevich, A. V. Vladimirova, J. P. Hekman, P. L. Perelman, A. S. Graphodatsky, S. J. O'Brien, X. Wang, A. G. Clark, G. M. Acland, L. N. Trut, and G. Zhang, "Red fox genome assembly identifies genomic regions associated with tame and aggressive behaviours," *Nature Ecology & Evolution*, vol. 2, no. 9, pp. 1479–1491, 2018.

[22] L. Pasteur, "Etudes sur la bière ses maladies, causes qui les provoquent, procédé pour la rendre inaltérable, avec une théorie nouvelle de la fermentation," 1876.

[23] E. Hansen, "Recherches sur la physiologie et la morphologie des ferments alcooliques V. Methodes pour obtenir des cultures pures de Saccharomyces et de mikroorganismes analogues.," *C. R. Trav. Lab. Carlsberg*, vol. 2, pp. 92–105, 1883.

[24] A. Rokas, "The effect of domestication on the fungal proteome," *Trends Genet.*, vol. 25, no. 2, pp. 60–63, 2009.

[25] J. Steensels, T. Snoek, E. Meersman, M. P. Nicolino, K. Voordeckers, and K. J. Verstrepen, "Improving industrial yeast strains: exploiting natural and artificial diversity," *FEMS microbiology reviews*, pp. 1–49, 2014.

[26] B. Gallone, J. Steensels, T. Prahl, L. Soriaga, V. Saels, B. Herrera-Malaver, A. Merlevede, M. Roncoroni, K. Voordeckers, L. Miraglia, C. Teiling, B. Steffy, M. Taylor, A. Schwartz, T. Richardson, C. White, G. Baele, S. Maere, and K. J. Verstrepen, "Domestication and Divergence of *Saccharomyces cerevisiae* Beer Yeasts," *Cell*, vol. 166, no. 6, pp. 1397–1410.e16, 2016.

[27] J. Peter, M. De Chiara, A. Friedrich, J. X. Yue, D. Pflieger, A. Bergstrom, A. Sigwalt, B. Barre, K. Freel, A. Llored, C. Cruaud, K. Labadie, J. M. Aury, B. Istace, K. Lebrigand, P. Barbry, S. Engelen, A. Lemainque, P. Wincker, G. Liti, and J. Schacherer, "Genome evolution across 1,011 Saccharomyces cerevisiae isolates," *Nature*, vol. 556, no. 7701, pp. 339–344, 2018.

[28] M. Goncalves, A. Pontes, P. Almeida, R. Barbosa, M. Serra, D. Libkind, M. Hutzler, P. Goncalves, and J. P. Sampaio, "Distinct Domestication Trajectories in Top-Fermenting Beer Yeasts and Wine Yeasts," *Curr. Biol.*, vol. 26, no. 20, pp. 2750–2761, 2016.

[29] S. F. Duan, P. J. Han, Q. M. Wang, W. Q. Liu, J. Y. Shi, K. Li, X. L. Zhang, and F. Y. Bai, "The origin and adaptive evolution of domesticated populations of yeast from Far East Asia," *Nat Commun*, vol. 9, no. 1, p. 2690, 2018.

[30] J. L. Legras, V. Galeote, F. Bigey, C. Camarasa, S. Marsit, T. Nidelet, I. Sanchez, A. Couloux, J. Guy, R. Franco-Duarte, M. Marcet-Houben, T. Gabaldon, D. Schuller, J. P. Sampaio, and S. Dequin, "Adaptation of S. cerevisiae to Fermented Food Environments Reveals Remarkable Genome Plasticity and the Footprints of Domestication," *Mol. Biol. Evol.*, vol. 35, no. 7, pp. 1712–1727, 2018.

[31] S. Marsit, J.-B. Leducq, É. Durand, A. Marchant, M. Filteau, and C. R. Landry, "Evolutionary biology through the lens of budding yeast comparative genomics," *Nature Reviews Genetics*, vol. 18, no. 10, pp. 581–598, 2017.

[32] K. S. Makarova and E. V. Koonin, "Evolutionary genomics of lactic acid bacteria," *Journal of bacteriology*, vol. 189, no. 4, pp. 1199–208, 2007.

[33] K. Makarova, A. Slesarev, Y. Wolf, A. Sorokin, B. Mirkin, E. Koonin, A. Pavlov, N. Pavlova, V. Karamychev, N. Polouchine, V. Shakhova, I. Grigoriev, Y. Lou, D. Rohksar, S. Lucas, K. Huang, D. M. Goodstein, T. Hawkins, V. Plengvidhya, D. Welker, J. Hughes, Y. Goh, A. Benson, K. Baldwin, J.-H. Lee, I. Diaz-Muniz, B. Dosti, V. Smeianov, W. Wechter, R. Barabote, G. Lorca, E. Altermann, R. Barrangou, B. Ganesan, Y. Xie, H. Rawsthorne, D. Tamir, C. Parker, F. Breidt, J. Broadbent, R. Hutkins, D. O'Sullivan, J. Steele, G. Unlu, M. Saier, T. Klaenhammer, P. Richardson, S. Kozyavkin, B. Weimer, and D. Mills, "Comparative genomics of the lactic acid bacteria," *Proceedings of the National Academy of Sciences*, vol. 103, no. 42, pp. 15611–15616, 2006.

[34] R. J. Siezen, B. Renckens, I. van Swam, S. Peters, R. van Kranenburg, M. Kleerebezem, and W. M. de Vos, "Complete Sequences of Four Plasmids of *Lactococcus lactis* subsp. cremoris SK11 Reveal Extensive Adaptation to the Dairy Environment," *Applied and Environmental Microbiology*, vol. 71, no. 12, pp. 8371–8382, 2005.

[35] S. Mills, O. E. McAuliffe, A. Coffey, G. F. Fitzgerald, and R. P. Ross, "Plasmids of *lactococci* – genetic accessories or genetic necessities?," *FEMS Microbiology Reviews*, vol. 30, no. 2, pp. 243–273, 2006.

[36] H. Bachmann, M. J. C. Starrenburg, D. Molenaar, M. Kleerebezem, and J. E. T. van Hylckama Vlieg, "Microbial domestication signatures of Lactococcus lactis can be reproduced by experimental evolution," *Genome Research*, vol. 22, no. 1, pp. 115–124, 2012.

[37] P. Barbesgaard, H. Heldt-Hansen, and B. Diderichsen, "On the safety of *Aspergillus oryzae*: a review," *Applied Microbiology and Biotechnology*, vol. 36, no. 5, pp. 569–572, 1992.

[38] J. Gibbons, L. Salichos, J. Slot, D. Rinker, K. McGary, J. King, M. Klich, D. Tabb, W. McDonald, and A. Rokas, "The Evolutionary Imprint of Domestication on Genome Variation and Function of the Filamentous Fungus *Aspergillus oryzae*," *Current Biology*, vol. 22, no. 15, pp. 1403–1409, 2012.

[39] T. Kobayashi, K. Abe, K. Asai, K. Gomi, P. R. Juvvadi, M. Kato, K. Kitamoto, M. Takeuchi, and M. Machida, "Genomics of *Aspergillus oryzae*," *Bioscience, Biotechnology, and Biochemistry*, vol. 71, no. 3, pp. 646–670, 2007.

[40] A. Goffeau, "Four years of post-genomic life with 6000 yeast genes," *FEBS Letters*, vol. 480, pp. 37–41, 2000.

[41] a. Goffeau, B. G. Barrell, H. Bussey, R. W. Davis, B. Dujon, H. Feldmann, F. Galibert, J. D. Hoheisel, C. Jacq, M. Johnston, E. J. Louis, H. W. Mewes, Y. Murakami, P. Philippsen, H. Tettelin, and S. G. Oliver, "Life with 6000 genes.," *Science (New York, N.Y.)*, vol. 274, no. NOVEMBER, pp. 546, 563–567, 1996.

[42] S. R. Engel and J. M. Cherry, "The new modern era of yeast genomics: community sequencing and the resulting annotation of multiple *Saccharomyces cerevisiae* strains at the *Saccharomyces* Genome Database," *Database : the journal of biological databases and curation*, vol. 2013, p. bat012, 2013.

[43] S. R. Engel, F. S. Dietrich, D. G. Fisk, G. Binkley, R. Balakrishnan, M. C. Costanzo, S. S. Dwight, B. C. Hitz, K. Karra, R. S. Nash, S. Weng, E. D. Wong, P. Lloyd, M. S. Skrzypek, S. R. Miyasato, M. Simison, and J. M. Cherry, "The reference genome sequence of *Saccharomyces cerevisiae*: then and now.," *G3 (Bethesda, Md.)*, vol. 4, no. March, pp. 389–98, 2014.

[44] A. Bergström, J. T. Simpson, F. Salinas, B. Barré, L. Parts, A. Zia, A. N. Nguyen Ba, A. M. Moses, E. J. Louis, V. Mustonen, J. Warringer, R. Durbin, and G. Liti, "A high-definition view of functional genetic variation from natural yeast genomes," *Molecular biology and evolution*, vol. 31, no. 4, pp. 872–88, 2014.

[45] A. R. Borneman, B. a. Desany, D. Riches, J. P. Affourtit, A. H. Forgan, I. S. Pretorius, M. Egholm, and P. J. Chambers, "Whole-genome comparison reveals novel genetic elements that characterize the genome of industrial strains of *Saccharomyces cerevisiae*," *PLoS Genetics*, vol. 7, no. 2, 2011.

[46] A. R. Borneman, R. Zeppel, P. J. Chambers, and C. D. Curtin, "Insights into the *Dekkera bruxellensis* genomic landscape: comparative genomics reveals variations in ploidy and nutrient utilisation potential amongst wine isolates.," *PLoS genetics*, vol. 10, no. 2, p. e1004161, 2014.

[47] A. R. Borneman, A. H. Forgan, R. Kolouchova, J. A. Fraser, and S. A. Schmidt, "Whole Genome Comparison Reveals High Levels of Inbreeding and Strain Redundancy Across the Spectrum of Commercial Wine Strains of *Saccharomyces cerevisiae*.," *G3 (Bethesda, Md.)*, 2016.

[48] B. Dunn and G. Sherlock, "Reconstruction of the genome origins and evolution of the hybrid lager yeast *Saccharomyces pastorianus*," *Genome Research*, vol. 18, pp. 1610–1623, 2008.

[49] G. Liti, D. M. Carter, A. M. Moses, J. Warringer, L. Parts, S. a. James, R. P. Davey, I. N. Roberts, A. Burt, V. Koufopanou, I. J. Tsai, C. M. Bergman, D. Bensasson, M. J. T. O'Kelly, A. van Oudenaarden, D. B. H. Barton, E. Bailes, A. N. Nguyen, M. Jones, M. a. Quail, I. Goodhead, S. Sims, F. Smith, A. Blomberg, R. Durbin, and E. J. Louis, "Population genomics of domestic and wild yeasts.," *Nature*, vol. 458, no. 7236, pp. 337–41, 2009.

[50] J. Schacherer, J. a. Shapiro, D. M. Ruderfer, and L. Kruglyak, "Comprehensive polymorphism survey elucidates population structure of *Saccharomyces cerevisiae*.," *Nature*, vol. 458, no. 7236, pp. 342–5, 2009.

[51] P. K. Strope, D. a. Skelly, S. G. Kozmin, G. Mahadevan, E. a. Stone, P. M. Magwene, F. S. Dietrich, J. H. Mccusker, N. Carolina, B. Sciences, and N. Carolina, "The 100-genomes strains, a *S . cerevisiae* resource that illuminates its natural phenotypic and genotypic variation and emergence as an opportunistic pathogen," *Genome Res.*, pp. 1–13, 2015.

[52] J. Warringer, E. Zörgö, F. a. Cubillos, A. Zia, A. Gjuvsland, J. T. Simpson, A. Forsmark, R. Durbin, S. W. Omholt, E. J. Louis, G. Liti, A. Moses, and A. Blomberg, "Trait variation in yeast is defined by population history.," *PLoS genetics*, vol. 7, no. 6, p. e1002111, 2011.

[53] C. Boulton and D. Quain, *Brewing Yeast and fermentation*. Blackwell Science, 2009.

[54] C. P. Kurtzman and J. Fell, *The Yeasts: a taxonomic study*. Amsterdam: Elsevier B.V., 4th ed., 1997.

[55] J. Tornai-Lehoczki and D. Dlauchy, "Delimination of brewing yeast strains using different molecular techniques," *International Journal of Food Microbiology*, vol. 62, pp. 37–45, 2000.

[56] A. Querol and U. Bond, "The complex and dynamic genomes of industrial yeasts: MINIREVIEW," *FEMS Microbiology Letters*, vol. 293, pp. 1–10, 2009.

[57] T. Deak, *Handbook of Food Spoilage Yeasts*. CRC Press, 2008.

[58] K. Voordeckers, C. a. Brown, K. Vanneste, E. van der Zande, A. Voet, S. Maere, and K. J. Verstrepen, "Reconstruction of Ancestral Metabolic Enzymes Reveals Molecular Mechanisms Underlying Evolutionary Innovation through Gene Duplication," *PLoS Biology*, vol. 10, no. 12, 2012.

[59] Y. L. Jin and R. Alex Speers, "Flocculation of *Saccharomyces cerevisiae*," *Food Research International*, vol. 31, no. 6-7, pp. 421–440, 1998.

[60] K. J. Verstrepen and F. M. Klis, "Flocculation, adhesion and biofilm formation in yeasts," *Molecular Microbiology*, vol. 60, pp. 5–15, 2006.

[61] K. J. Verstrepen, G. Derdelinckx, H. Verachtert, and F. R. Delvaux, "Yeast flocculation: what brewers should know," *Applied microbiology and biotechnology*, vol. 61, no. 3, pp. 197–205, 2003.

[62] B. Gibson and G. Liti, "Saccharomyces pastorianus: genomic insights inspiring innovation for industry.," *Yeast (Chichester, England)*, vol. 28, no. 28, pp. 17–27, 2014.

[63] B. R. Gibson, J. Londesborough, J. Rautio, L. Mattinen, and V. Vidgren, "Transcription of α-glucoside transport and metabolism genes in the hybrid brewing yeast *Saccharomyces pastorianus* with respect to gene provenance and fermentation temperature," *Journal of the Institute of Brewing*, vol. 119, no. April, pp. 23–31, 2013.

[64] M. Hebly, A. Brickwedde, I. Bolat, M. R. M. Driessen, E. a. F. de Hulster, M. van den Broek, J. T. Pronk, J.-M. Geertman, J.-M. Daran, and P. Daran-Lapujade, "*S. cerevisiae* x *S. eubayanus* interspecific hybrid, the best of both worlds and beyond," *FEMS Yeast Research*, vol. 15, no. 3, p. fov005, 2015.

[65] E. Meersman, "Selectie van specifieke gisten voor een optimale cacaofermentatie," *Dissertation*, 2011.

[66] S. Mertens, J. Steensels, V. Saels, G. De Rouck, G. Aerts, and K. J. Verstrepen, "A Large Set of Newly Created Interspecific *Saccharomyces* Hybrids Increases Aromatic Diversity in Lager Beers," *Applied and Environmental Microbiology*, vol. 81, no. 23, pp. 8202–8214, 2015.

[67] M. E. Rodríguez, L. Pérez-Través, M. P. Sangorrín, E. Barrio, and C. a. Lopes, "*Saccharomyces eubayanus* and *Saccharomyces uvarum* associated with the fermentation of Araucaria araucana seeds in Patagonia.," *FEMS yeast research*, vol. 14, pp. 948–965, 2014.

[68] I. Campbell, "Numerical analysis of the genera *Saccharomyces* and *Kluyveromyces*," *Journal of general microbiology*, vol. 73, no. 2, pp. 279–301, 1972.

[69] J. L. Legras and F. Karst, "Optimisation of interdelta analysis for *Saccharomyces cerevisiae* strain characterisation," *FEMS Microbiology Letters*, vol. 221, pp. 249–255, 2003.

[70] M. Aigle, D. Erbs, and M. Moll, "Some Molecular Structures in the Genome of Lager Brewing Yeasts," *Am. Soc. Brew. Chem.*, no. 42, pp. 1–7, 1984.

[71] J. Tornai-Lehoczki, G. Péter, D. Dlauchy, and T. Deák, "Some remarks on "a taxonomic key for the genus Saccharomyces" (Vaughan Martini and Martini 1993).," *Antonie van Leeuwenhoek*, vol. 69, no. 3, pp. 229–233, 1996.

[72] A. Vaughan Martini and C. P. Kurtzman, "Deoxyribonucleic Acid Relatedness among Species of the Genus *Saccharomyces Sensu Stricto*," *International Journal of Systematic Bacteriology*, vol. 35, no. 31, pp. 508–511, 1985.

[73] a. Vaughan Martini and a. Martini, "Three newly delimited species of *Saccharomyces sensu stricto*," *Antonie van Leeuwenhoek*, vol. 53, no. 2, pp. 77–84, 1987.

[74] F. Vezinhet, B. Blondin, and J.-N. Hallet, "Chromosomal DNA patterns and mitochondrial DNA polymorphism as tools for identification of enological strains of *Saccharomyces cerevisiae*," *Applied Microbiology and Biotechnology*, vol. 32, no. 5, pp. 568–571, 1990.

[75] M. M. Baleiras Couto, J. M. B. Van der Vossen, H. Hofstra, and J. H. J. Huis in 't Veld, "RAPD analysis: A rapid technique for differentiation of spoilage yeasts," *International Journal of Food Microbiology*, vol. 24, pp. 249–260, 1994.

[76] F. Ness, F. Lavallee, D. Dubourdieu, M. Aigle, and L. Dulau, "Identification Of Yeast Strains Using the Polymerase Chain Reaction," *Journal of the Science of Food and Agriculture*, no. 62, pp. 89–94, 1993.

[77] M. M. Baleiras Couto, B. Eijsma, H. Hofstra, J. H. Huis in't Veld, and J. M. van der Vossen, "Evaluation of molecular typing techniques to assign genetic diversity among *Saccharomyces cerevisiae* strains.," *Applied and environmental microbiology*, vol. 62, pp. 41–46, 1996.

[78] R. Messner and H. Prillinger, "Saccharomyces species assignment by long range ribotyping," *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology*, vol. 67, no. 4, pp. 363–370, 1995.

[79] S. Smole Mozina, D. Dlauchy, T. Deak, and P. Raspor, "Identification of *Saccharomyces sensu stricto* and *Torulaspora* yeasts by PCR ribotyping.," *Letters in applied microbiology*, vol. 24, no. 4, pp. 311–315, 1997.

[80] M. R. Goddard, N. Anfang, R. Tang, R. C. Gardner, and C. Jun, "A distinct population of *Saccharomyces cerevisiae* in New Zealand: Evidence for local dispersal by insects and human-aided global dispersal in oak barrels," *Environmental Microbiology*, vol. 12, pp. 63–73, 2010.

[81] T. Katz Ezov, E. Boger-Nadjar, Z. Frenkel, I. Katsperovski, S. Kemeny, E. Nevo, a. Korol, and Y. Kashi, "Molecular-genetic biodiversity in a natural population of the yeast *Saccharomyces cerevisiae* from "Evolution Canyon": Microsatellite polymorphism, ploidy and controversial sexual status," *Genetics*, vol. 174, no. November, pp. 1455–1468, 2006.

[82] J. L. Legras, O. Ruh, D. Merdinoglu, and F. Karst, "Selection of hypervariable microsatellite loci for the characterization of *Saccharomyces cerevisiae* strains," *International Journal of Food Microbiology*, vol. 102, pp. 73–83, 2005.

[83] J. L. Legras, D. Merdinoglu, J. M. Cornuet, and F. Karst, "Bread, beer and wine: *Saccharomyces cerevisiae* diversity reflects human history," *Molecular Ecology*, vol. 16, pp. 2091–2102, 2007.

[84] G. A. Cromie, K. E. Hyma, C. L. Ludlow, C. Garmendia-Torres, T. L. Gilbert, P. May, A. a. Huang, A. M. Dudley, and J. C. Fay, "Genomic sequence diversity and population structure of *Saccharomyces cerevisiae* assessed by RAD-seq.," *G3 (Bethesda, Md.)*, vol. 3, no. 12, pp. 2163–71, 2013.

[85] J. Bing, P.-J. Han, W.-Q. Liu, Q.-M. Wang, and F.-Y. Bai, "Evidence for a Far East Asian origin of lager beer yeast.," *Current biology : CB*, vol. 24, no. 10, pp. R380–1, 2014.

[86] J. C. Fay and J. a. Benavides, "Evidence for domesticated and wild populations of *Saccharomyces cerevisiae*," *PLoS Genetics*, vol. 1, no. 1, pp. 0066–0071, 2005.

[87] M. Ramazzotti, L. Berná, I. Stefanini, and D. Cavalieri, "A computational pipeline to discover highly phylogenetically informative genes in sequenced genomes: application to *Saccharomyces cerevisiae* natural strains.," *Nucleic acids research*, vol. 40, no. 9, pp. 3834–48, 2012.

[88] Q.-M. Wang, W.-Q. Liu, G. Liti, S.-A. Wang, and F.-Y. Bai, "Surprisingly diverged populations of *Saccharomyces cerevisiae* in natural environments remote from human activity.," *Molecular ecology*, vol. 21, no. 22, pp. 5404–17, 2012.

[89] J. M. Thomson, E. A. Gaucher, M. F. Burgan, D. W. De Kee, T. Li, J. P. Aris, and S. A. Benner, "Resurrecting ancestral alcohol dehydrogenases from yeast," *Nat. Genet.*, vol. 37, no. 6, pp. 630–635, 2005.

[90] G. C. Conant and K. H. Wolfe, "Increased glycolytic flux as an outcome of whole-genome duplication in yeast," *Molecular systems biology*, vol. 3, no. 129, p. 129, 2007.

[91] M. R. Goddard, "Quantifying the complexities of *Saccharomyces cerevisiae*'s ecosystem engineering via fermentation," *Ecology*, vol. 89, no. 8, pp. 2077–2082, 2008.

[92] J. Piškur, E. Rozpedowska, S. Polakova, A. Merico, and C. Compagno, "How did *Saccharomyces* evolve to become a good brewer?," *Trends in Genetics*, vol. 22, no. xx, pp. 183–186, 2006.

[93] A. Hagman, T. Säll, C. Compagno, and J. Piskur, "Yeast "Make-Accumulate-Consume" Life Strategy Evolved as a Multi-Step Process That Predates the Whole Genome Duplication," *PLoS ONE*, vol. 8, no. 7, 2013.

[94] Z. Lin and W. H. Li, "Expansion of hexose transporter genes was associated with the evolution of aerobic fermentation in yeasts," *Molecular Biology and Evolution*, vol. 28, no. 1, pp. 131–142, 2011.

[95] J. Ihmels, S. Bergmann, M. Gerami-Nejad, I. Yanai, M. McClellan, J. Berman, and N. Barkai, "Rewiring of the yeast transcriptional network through the evolution of motif usage," *Science (New York, N.Y.)*, vol. 309, no. 2005, pp. 938–940, 2005.

[96] B. Gibson and G. Liti, "*Saccharomyces pastorianus*: genomic insights inspiring innovation for industry," *Yeast (Chichester, England)*, 2014.

[97] I. S. Hornesey, *A History of Beer and Brewing*. Cambridge: The Royal Society of Chemistry, 1 ed., 2003.

[98] Y. Kodama, M. C. Kielland-Brandt, and J. Hansen, "Lager brewing yeast," *Topics in Current Genetics*, vol. 15, no. January, pp. 145–164, 2006.

[99] D. Libkind, C. T. Hittinger, E. Valério, C. Gonçalves, J. Dover, M. Johnston, P. Gonçalves, and J. P. Sampaio, "Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, pp. 14539–14544, 2011.

[100] J. C. Fay and J. a. Benavides, "Evidence for domesticated and wild populations of *Saccharomyces cerevisiae*," *PLoS Genetics*, vol. 1, no. 1, pp. 0066–0071, 2005.

[101] D. Sicard and J. L. Legras, "Bread, beer and wine: Yeast domestication in the *Saccharomyces sensu stricto* complex," *Comptes Rendus - Biologies*, vol. 334, no. 3, pp. 229–236, 2011.

[102] A. R. Borneman, B. A. Desany, D. Riches, J. P. Affourtit, A. H. Forgan, I. S. Pretorius, M. Egholm, and P. J. Chambers, "Whole-genome comparison reveals novel genetic elements that characterize the genome of industrial strains of *Saccharomyces cerevisiae*.," *PLoS genetics*, vol. 7, no. 2, p. e1001287, 2011.

[103] A. Spor, T. Nidelet, J. Simon, A. Bourgais, D. de Vienne, and D. Sicard, "Niche-driven evolution of metabolic and life-history strategies in natural and domesticated populations of *Saccharomyces cerevisiae*.," *BMC evolutionary biology*, vol. 9, p. 296, 2009.

[104] J. L. Will, H. S. Kim, J. Clarke, J. C. Painter, J. C. Fay, and A. P. Gasch, "Incipient balancing selection through adaptive loss of aquaporins in natural *Saccharomyces cerevisiae* populations," *PLoS Genetics*, vol. 6, no. 4, 2010.

[105] J. Fay, H. McCullough, P. Sniegowski, and M. Eisen, "Population genetic variation in gene expression is associated with phenotypic variation in *Saccharomyces cerevisiae*," *Genome Biology*, vol. 5, no. 4, p. R26, 2004.

[106] J. E. Pérez-Ortín, A. Querol, S. Puig, and E. Barrio, "Molecular characterization of a chromosomal rearrangement involved in the adaptive evolution of yeast strains," *Genome research*, vol. 12, no. 10, pp. 1533–9, 2002.

[107] V. Galeote, M. Novo, M. Salema-Oom, C. Brion, E. Valério, P. Gonçalves, and S. Dequin, "FSY1, a horizontally transferred gene in the *Saccharomyces cerevisiae* EC1118 wine yeast strain, encodes a high-affinity fructose/H+ symporter," *Microbiology (Reading, England)*, vol. 156, no. Pt 12, pp. 3754–61, 2010.

[108] M. Novo, F. Bigey, E. Beyne, V. Galeote, F. Gavory, S. Mallet, B. Cambon, J.-L. Legras, P. Wincker, S. Casaregola, and S. Dequin, "Eukaryote-to-eukaryote gene transfer events revealed by the genome sequence of the wine yeast *Saccharomyces cerevisiae* EC1118," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 38, pp. 16333–8, 2009.

[109] C. A. Brown, A. W. Murray, and K. J. Verstrepen, "Rapid expansion and functional divergence of subtelomeric gene families in yeasts," *Current biology : CB*, vol. 20, no. 10, pp. 895–903, 2010.

[110] M. J. Charron and C. a. Michels, "The naturally occurring alleles of *MAL1* in *Saccharomyces* species evolved by various mutagenic processes including chromosomal rearrangement," *Genetics*, vol. 120, pp. 83–93, 1988.

[111] B. Dunn and G. Sherlock, "Reconstruction of the genome origins and evolution of the hybrid lager yeast *Saccharomyces pastorianus*," *Genome Research*, vol. 18, pp. 1610–1623, 2008.

[112] B. U. Stambuk, B. Dunn, S. L. Alves, E. H. Duval, and G. Sherlock, "Industrial fuel ethanol yeasts contain adaptive copy number changes in genes involved in vitamin B1 and B6 biosynthesis.," *Genome research*, vol. 19, no. 12, pp. 2271–8, 2009.

[113] J. Steensels and K. J. Verstrepen, "Taming Wild Yeast: Potential of Conventional and Nonconventional Yeasts in Industrial Fermentations," *Annual Review of Microbiology*, vol. 68, no. 1, pp. 61–80, 2014.

[114] J. W. Wenger, K. Schwartz, and G. Sherlock, "Bulk segregant analysis by high-throughput sequencing reveals a novel xylose utilization gene from *Saccharomyces cerevisiae*.," *PLoS genetics*, vol. 6, no. 5, p. e1000942, 2010.

[115] K. E. Hyma, S. M. Saerens, K. J. Verstrepen, and J. C. Fay, "Divergence in wine characteristics produced by wild and domesticated strains of *Saccharomyces cerevisiae*.," *FEMS yeast research*, vol. 11, no. 7, pp. 540–51, 2011.

[116] N. Mukai, K. Masaki, T. Fujii, M. Kawamukai, and H. Iefuji, "PAD1 and FDC1 are essential for the decarboxylation of phenylacrylic acids in *Saccharomyces cerevisiae*.," *Journal of bioscience and bioengineering*, vol. 109, no. 6, pp. 564–9, 2010.

[117] N. Mukai, K. Masaki, T. Fujii, and H. Iefuji, "Single nucleotide polymorphisms of PAD1 and FDC1 show a positive relationship with ferulic acid decarboxylation ability among industrial yeasts used in alcoholic beverage production.," *Journal of bioscience and bioengineering*, vol. 118, no. 1, pp. 50–5, 2014.

[118] M. Fidalgo, R. R. Barrales, J. I. Ibeas, and J. Jimenez, "Adaptive evolution by mutations in the FLO11 gene.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 30, pp. 11228–33, 2006.

[119] J. F. Christiaens, S. E. Van Mulders, J. Duitama, C. A. Brown, M. G. Ghequire, L. De Meester, J. Michiels, T. Wenseleers, K. Voordeckers, and K. J. Verstrepen, "Functional divergence of gene duplicates through ectopic recombination," *EMBO reports*, vol. 13, no. 12, pp. 1145–1151, 2012.

[120] D. Carmona-Gutierrez, A. Alavian-Ghavanini, L. Habernig, M. A. Bauer, A. Hammer, C. Rossmann, A. S. Zimmermann, C. Ruckenstuhl, S. Büttner, T. Eisenberg, W. Sattler, E. Malle, and F. Madeo, "The cell death protease Kex1p is essential for hypochlorite-induced apoptosis in yeast.," *Cell cycle (Georgetown, Tex.)*, vol. 12, no. 11, pp. 1704–12, 2013.

[121] K. Pougach, A. Voet, F. a. Kondrashov, K. Voordeckers, J. F. Christiaens, B. Baying, V. Benes, R. Sakai, J. Aerts, B. Zhu, P. Van Dijck, and K. J. Verstrepen, "Duplication of a promiscuous transcription factor drives the emergence of a new regulatory network," *Nature Communications*, vol. 5, p. 4868, 2014.

[122] K. J. Verstrepen, G. Derdelinckx, J. P. Dufour, J. Winderickx, J. M. Thevelein, I. S. Pretorius, and F. R. Delvaux, "Flavor-active esters: Adding fruitiness to beer," *Journal of Bioscience and Bioengineering*, vol. 96, no. 2, pp. 110–118, 2003.

[123] C. D. Powell, D. E. Quain, and K. a. Smart, "The impact of sedimentation on cone yeast heterogeneity," *Journal of the American Society of Brewing Chemists*, vol. 62, no. 15, pp. 8–17, 2004.

[124] K. J. Verstrepen, A. Jansen, F. Lewitter, and G. R. Fink, "Intragenic tandem repeats generate functional variability," *Nature Genetics*, vol. 37, no. 9, pp. 986–990, 2005.

[125] A. Martini, "Origin and domestication of the wine yeast *Saccharomyces cerevisiae*," *Journal of Wine Research*, vol. 4, no. December 2014, pp. 165–176, 1993.

[126] G. Naumov, "Genetic identification of biological species in theSaccharomyces sensu stricto complex," *Journal of Industrial Microbiology & Biotechnology*, vol. 17, pp. 295–302, 1996.

[127] G. Liti, D. M. Carter, A. M. Moses, J. Warringer, L. Parts, S. a. James, R. P. Davey, I. N. Roberts, A. Burt, V. Koufopanou, I. J. Tsai, C. M. Bergman, D. Bensasson, M. J. T. O'Kelly, A. van Oudenaarden, D. B. H. Barton, E. Bailes, A. N. Nguyen, M. Jones, M. a. Quail, I. Goodhead, S. Sims, F. Smith, A. Blomberg, R. Durbin, and E. J. Louis, "Population genomics of domestic and wild yeasts.," *Nature*, vol. 458, no. 7236, pp. 337–341, 2009.

[128] P. Almeida, R. Barbosa, P. Zalar, Y. Imanishi, K. Shimizu, B. Turchetti, J.-L. Legras, M. Serra, S. Dequin, A. Couloux, J. Guy, D. Bensasson, P. Gonçalves, and J. P. Sampaio, "A population genomics insight into the Mediterranean origins of wine yeast domestication.," *Molecular ecology*, vol. 24, no. 21, pp. 5412–27, 2015.

[129] L. Jorde and S. Wooding, "Genetic variation, classification and'race'," *Nature genetics*, vol. 36, no. 11, pp. 28–33, 2004.

[130] M. R. Goddard and D. Greig, "*Saccharomyces cerevisiae*: a nomadic yeast with no niche?," *FEMS yeast research*, p. fov009, 2015.

[131] M. J. Dunham, H. Badrane, T. Ferea, J. Adams, P. O. Brown, F. Rosenzweig, and D. Botstein, "Characteristic genome rearrangements in experimental evolution of *Saccharomyces cerevisiae*.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 25, pp. 16144–16149, 2002.

[132] J. Masel and D. N. Lyttle, "The consequences of rare sexual reproduction by means of selfing in an otherwise clonally reproducing species," *Theoretical Population Biology*, vol. 80, no. 4, pp. 317–322, 2011.

[133] A. Selmecki, M. Gerami-Nejad, C. Paulson, A. Forche, and J. Berman, "An isochromosome confers drug resistance in vivo by amplification of two genes, ERG11 and TAC1," *Molecular Microbiology*, vol. 68, no. March, pp. 624–641, 2008.

[134] E. M. Torres, T. Sokolsky, C. M. Tucker, L. Y. Chan, M. Boselli, M. J. Dunham, and A. Amon, "Effects of aneuploidy on cellular physiology and cell division in haploid yeast.," *Science (New York, N.Y.)*, vol. 317, no. August, pp. 916–924, 2007.

[135] B. Dunn, C. Richter, D. J. Kvitek, T. Pugh, and G. Sherlock, "Analysis of the *Saccharomyces cerevisiae* pan-genome reveals a pool of copy number variants distributed in diverse yeast strains from differing industrial environments," *Genome Research*, vol. 22, pp. 908–924, 2012.

[136] J. Hose, C. M. Yong, M. Sardi, Z. Wang, M. A. Newton, and A. P. Gasch, "Dosage compensation can buffer copy- number variation in wild yeast," *eLife*, pp. 1–27, 2015.

[137] G. Liti, A. N. Nguyen Ba, M. Blythe, C. a. Müller, A. Bergström, F. a. Cubillos, F. Dafhnis-Calas, S. Khoshraftar, S. Malla, N. Mehta, C. C. Siow, J. Warringer, A. M. Moses, E. J. Louis, and C. a. Nieduszynski, "High quality de novo sequencing and assembly of the *Saccharomyces arboricolus* genome.," *BMC genomics*, vol. 14, p. 69, 2013.

[138] R. K. Mortimer, "Evolution and variation of the yeast (*Saccharomyces*) genome," *Genome Research*, vol. 10, pp. 403–409, 2000.

[139] R. M. Ames, B. M. Rash, K. E. Hentges, D. L. Robertson, D. Delneri, and S. C. Lovell, "Gene duplication and environmental adaptation within yeast populations," *Genome Biology and Evolution*, vol. 2, pp. 591–601, 2010.

[140] A. M. Selmecki, Y. E. Maruvka, P. A. Richmond, M. Guillet, N. Shoresh, A. L. Sorenson, S. De, R. Kishony, F. Michor, R. Dowell, and D. Pellman, "Polyploidy can drive rapid adaptation in yeast," *Nature*, vol. 519, no. 7543, pp. 349–352, 2015.

[141] A. H. Yona, Y. S. Manor, R. H. Herbst, G. H. Romano, A. Mitchell, M. Kupiec, Y. Pilpel, and O. Dahan, "Chromosomal duplication is a transient evolutionary solution to stress.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, no. 51, pp. 21010–5, 2012.

[142] P. M. Magwene, Ö. Kayıkçı, J. a. Granek, J. M. Reininga, Z. Scholl, and D. Murray, "Outcrossing, mitotic recombination, and life-history trade-offs shape genome evolution in *Saccharomyces cerevisiae.*," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 5, pp. 1987–92, 2011.

[143] R. K. Mortimer, P. Romano, G. Suzzi, and M. Polsinelli, "Genome renewal: A new phenomenon revealed from a genetic study of 43 strains of *Saccharomyces cerevisiae* derived from natural fermentation of grape musts," *Yeast*, vol. 10, pp. 1543–1552, 1994.

[144] U. Bond, *Chapter 6 The Genomes of Lager Yeasts*, vol. 69. Elsevier Inc., 1 ed., 2009.

[145] K. A. Smart, "Brewing yeast genomes and genome wide expression and proteome profiling during fermentation 2007," *Yeast (Chichester, England)*, vol. 24, pp. 993–1013, 2007.

[146] J. A. Barnett, "A history of research on yeasts 2: Louis Pasteur and his contemporaries, 1850-1880," *Yeast*, vol. 16, pp. 755–771, 2000.

[147] F. G. Meussdoerffer, *A Comprehensive History of Beer Brewing*. Wiley-VCH Verlag GmbH & Co. KGaA, 2009.

[148] L. Pasteur, *études sur le vin, ses maladies, causes qui les provoquent, Procédés Nouveaux, pour le conserver et pour le vieillir*. Paris: F. Savy, 2 ed., 1873.

[149] M. Rees, *Botanische Untersuchungen über die Alkoholgährungspilze*. Leipzig: A. Felix, Leipzig, 1870.

[150] B. Regenberg and J. Hansen, "A history of research on yeasts 3: Emil Fischer Eduard Buchner and their contemporaries, 1880-1900," *Yeast*, vol. 18, pp. 363–388, 2001.

[151] J. Polaina, "Brewer's Yeast : Genetics and Biotechnology," *Applied Mycology and Biotechnology*, vol. 2, 2002.

[152] J. Wendland, "Lager yeast comes of age," *Eukaryotic Cell*, no. August, 2014.

[153] A. R. J., C. J. F., C. D., and H. J. R., "Dimethyl sulphide as a feature of lager flavour," *Journal of the Institute of Brewing*, vol. 81, no. 3, pp. 208–213, 1975.

[154] T. Snoek, M. Picca Nicolino, S. Van den Bremt, S. Mertens, V. Saels, A. Verplaetse, J. Steensels, and K. J. Verstrepen, "Large-scale robot-assisted genome shuffling yields industrial *Saccharomyces cerevisiae* yeasts with increased ethanol tolerance," *Biotechnology for Biofuels*, vol. 8, no. 1, pp. 1–32, 2015.

[155] T. Nilsson-Tillgren, C. Gjermansen, M. C. Kielland-Brandt, J. G. L. Petersen, and S. Holmberg, "Genetic differences between *Saccharomyces carlsbergensis* and *S. cerevisiae*. Analysis of chromosome III by single chromosome transfer," *Carlsberg Research Communications*, vol. 46, pp. 65–76, 1981.

[156] G. Paul Casey, "Molecular and genetic analysis of chromosomes X in *Saccharomyces carlsbergensis*," *Carlsberg Research Communications*, vol. 51, no. 5, pp. 343–362, 1986.

[157] Y. Tamai, T. Momma, H. Yoshimoto, and Y. Kaneko, "Co-existence of two types of chromosome in the bottom fermenting yeast, *Saccharomyces pastorianus*," *Yeast*, vol. 14, no. 10, pp. 923–933, 1998.

[158] H. Yamagishi and T. Ogata, "Chromosomal structures of bottom fermenting yeasts.," *Systematic and applied microbiology*, vol. 22, no. 3, pp. 341–353, 1999.

[159] S. Rainieri, Y. Kodama, Y. Kaneko, K. Mikata, Y. Nakao, and T. Ashikari, "Pure and mixed genetic lines of *Saccharomyces bayanus* and *Saccharomyces pastorianus* and their contribution to the lager brewing strain genome," *Applied and Environmental Microbiology*, vol. 72, pp. 3968–3974, 2006.

[160] J. Hansen and M. C. Kielland-Brandt, "*Saccharomyces carlsbergensis* contains two functional MET2 alleles similar to homologues from *S. cerevisiae* and *S. monacensis.*," *Gene*, vol. 140, no. 1, pp. 33–40, 1994.

[161] Y. Nakao, T. Kanamori, T. Itoh, Y. Kodama, S. Rainieri, N. Nakamura, T. Shimonaga, M. Hattori, and T. Ashikari, "Genome sequence of the lager brewing yeast, an interspecies hybrid," *DNA Research*, vol. 16, pp. 115–129, 2009.

[162] D. Peris, K. Sylvester, D. Libkind, P. Gonçalves, J. P. Sampaio, W. G. Alexander, and C. T. Hittinger, "Population structure and reticulate evolution of *Saccharomyces eubayanus* and its lager-brewing hybrids," *Molecular Ecology*, vol. 23, pp. 2031–2045, 2014.

[163] F. Fabian and N. McCullough, "Dissociation in yeasts," *Michigan Academy of Science*, no. 169, 1933.

[164] B. R. Gibson, E. Storgårds, K. Krogerus, and V. Vidgren, "Comparative physiology and fermentation performance of Saaz and Frohberg lager yeast strains and the parental species *Saccharomyces eubayanus*," *Yeast*, vol. 30, no. 7, pp. 255–266, 2013.

[165] G. Liti, A. Peruffo, S. a. James, I. N. Roberts, and E. J. Louis, "Inferences of evolutionary relationships from a population survey of LTR-retrotransposons and telomeric-associated sequences in the *Saccharomyces sensu stricto* complex," *Yeast*, vol. 22, no. 3, pp. 177–192, 2005.

[166] A. Walther, A. Hesselbart, and J. Wendland, "Genome sequence of *Saccharomyces carlsbergensis*, the world's first pure culture lager yeast.," *G3 (Bethesda, Md.)*, vol. 4, no. May, pp. 783–93, 2014.

[167] P. Lindner, *Mikroskopische Betriebskontrolle in den Garungswerben*. Berlin: Paul Parey, 6th ed., 1909.

[168] M. D. B. Lopes, M. D. B. Lopes, J. R. Bellon, J. R. Bellon, N. J. Shirley, N. J. Shirley, P. F. Ganter, and P. F. Ganter, "Evidence for multiple interspecicc hybridization in *Saccharomyces sensu stricto* species," *FEMS Yeast Research*, vol. 1, 2002.

[169] M. Okuno, R. Kajitani, R. Ryusui, H. Morimoto, Y. Kodama, and T. Itoh, "Next-generation sequencing analysis of lager brewing yeast strains reveals the evolutionary history of interspecies hybridization," *DNA Research*, vol. 23, no. 1, p. dsv037, 2016.

[170] H. V. Nguyen, J. L. Legras, C. Neuvéglise, and C. Gaillardin, "Deciphering the hybridisation history leading to the lager lineage based on the mosaic genomes of *Saccharomyces bayanus* strains NBRC1948 and CBS380 T," *PLoS ONE*, vol. 6, no. 10, 2011.

[171] C. Monerawela, T. C. James, K. H. Wolfe, and U. Bond, "Loss of lager specific genes and subtelomeric regions define two different *Saccharomyces cerevisiae* lineages for *Saccharomyces pastorianus* Group I and II strains," *FEMS Yeast Research*, vol. 15, no. 2, 2015.

[172] S. Rainieri, Y. Kodama, Y. Nakao, A. Pulvirenti, and P. Giudici, "The inheritance of mtDNA in lager brewing strains," *FEMS Yeast Research*, vol. 8, no. 4, pp. 586–596, 2008.

[173] S. S. González, E. Barrio, J. Gafner, and A. Querol, "Natural hybrids from *Saccharomyces cerevisiae*, *Saccharomyces bayanus* and *Saccharomyces kudriavzevii* in wine fermentations," *FEMS Yeast Research*, vol. 6, no. 8, pp. 1221–1234, 2006.

[174] K. Krogerus, F. Magalhães, V. Vidgren, and B. Gibson, "New lager yeast strains generated by interspecific hybridization," *Journal of Industrial Microbiology & Biotechnology*, vol. 42, no. 42, pp. 769–778, 2015.

[175] G. Marinoni, M. Manuel, R. F. Petersen, J. Hvidtfeldt, P. Sulo, and J. Piskur, "Horizontal Transfer of Genetic Material among *Saccharomyces* Yeasts," *Journal of Bacter*, vol. 181, no. 20, pp. 6488–6496, 1999.

[176] M. Sato, M. Kishimoto, J. Watari, and M. Takashio, "Breeding of brewer's yeast by hybridization between a top-fermenting yeast *Saccharomyces cerevisiae* and a cryophilic yeast *Saccharomyces bayanus*," *Journal of Bioscience and Bioengineering*, vol. 93, no. 5, pp. 509–511, 2002.

[177] F. Sebastiani, C. Barberio, E. Casalone, D. Cavalieri, and M. Polsinelli, "Crosses between *Saccharomyces cerevisiae* and *Saccharomyces bayanus* generate fertile hybrids," *Research in Microbiology*, vol. 153, pp. 53–58, 2002.

[178] A. Serra, P. Strehaiano, and P. Taillandier, "Influence of temperature and pH on *Saccharomyces bayanus var. uvarum* growth; impact of a wine yeast interspecific hybridization on these parameters," *International Journal of Food Microbiology*, vol. 104, pp. 257–265, 2005.

[179] J. R. Bellon, J. M. Eglinton, T. E. Siebert, A. P. Pollnitz, L. Rose, M. De Barros Lopes, and P. J. Chambers, "Newly generated interspecific wine yeast hybrids introduce flavour and aroma diversity to wines," *Applied Microbiology and Biotechnology*, vol. 91, pp. 603–612, 2011.

[180] L. Pérez-Través, C. a. Lopes, E. Barrio, and A. Querol, "Evaluation of different genetic procedures for the generation of artificial hybrids in *Saccharomyces* genus for winemaking," *International Journal of Food Microbiology*, vol. 156, no. 2, pp. 102–111, 2012.

[181] B. Dunn, T. Paulish, A. Stanbery, J. Piotrowski, G. Koniges, E. Kroll, E. J. Louis, G. Liti, G. Sherlock, and F. Rosenzweig, "Recurrent Rearrangement during Adaptive Evolution in an Interspecific Yeast Hybrid Suggests a Model for Rapid Introgression," *PLoS Genetics*, vol. 9, no. 3, p. e1003366, 2013.

[182] J. S. Piotrowski, S. Nagarajan, E. Kroll, A. Stanbery, K. E. Chiotti, A. L. Kruckeberg, B. Dunn, G. Sherlock, and F. Rosenzweig, "Different selective pressures lead to different genomic outcomes as newly-formed hybrid yeasts evolve," *BMC Evolutionary Biology*, vol. 12, no. 1, p. 46, 2012.

[183] J. R. Bellon, F. Schmid, D. L. Capone, B. L. Dunn, and P. J. Chambers, "Introducing a New Breed of Wine Yeast: Interspecific Hybridisation between a Commercial *Saccharomyces cerevisiae* Wine Yeast and *Saccharomyces mikatae*," *PLoS ONE*, vol. 8, no. 4, p. e62053, 2013.

[184] N. A. Bokulich, C. W. Bamforth, and D. a. Mills, "Brewhouse-resident microbiota are responsible for multi-stage fermentation of American coolship ale," *PLoS ONE*, vol. 7, no. 4, 2012.

[185] N. A. Bokulich, C. W. Bamforth, and W. Bamforth, "The microbiology of malting and brewing.," *Microbiology and molecular biology reviews : MMBR*, vol. 77, no. 2, pp. 157–72, 2013.

[186] H. Martens, "Microbiological Aspects of a Mixed Yeast—Bacterial Fermentation in the Production of a Special Belgian Acidic Ale," *Journal of the Institute of …*, vol. 103, no. April, pp. 85–91, 1997.

[187] C. D. Curtin and I. S. Pretorius, "Genomic insights into the evolution of industrial yeast species *Brettanomyces bruxellensis*.," *FEMS yeast research*, vol. 14, pp. 997–1005, 2014.

[188] A. J. Schifferdecker, S. Dashko, O. P. Ishchuk, and J. Piskur, "The Wine and Beer Yeast *Dekkera bruxellensis*," *Yeast*, vol. 31, no. 31, pp. 323–332, 2014.

[189] J. Steensels, L. Daenen, P. Malcorps, G. Derdelinckx, H. Verachtert, and K. J. Verstrepen, "*Brettanomyces* yeasts - From spoilage organisms to valuable contributors to industrial fermentations," *International Journal of Food Microbiology*, vol. 206, pp. 24–38, 2015.

[190] M. Custers, *Onderzoekingen over het gistgeslacht Brettanomyces*. PhD thesis, Delft, 1940.

[191] M. R. Wijsman, J. P. van Dijken, B. H. van Kleeff, and W. a. Scheffers, "Inhibition of fermentation and growth in batch cultures of the yeast *Brettanomyces* intermedius upon a shift from aerobic to anaerobic conditions (Custers effect).," *Antonie van Leeuwenhoek*, vol. 50, pp. 183–192, 1984.

[192] J. A. Barnett, R. W. Payne, and D. Yarrow, *Yeasts: characteristics and identification*. Cambridge University Press, 1983.

[193] H. M. Daniel, M. A. Lachance, and C. P. Kurtzman, "On the reclassification of species assigned to *Candida* and other anamorphic ascomycetous yeast genera based on phylogenetic circumscription," *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology*, vol. 106, pp. 67–84, 2014.

[194] C. P. Kurtzman and C. J. Robnett, "Relationships among genera of the *Saccharomycotina* (Ascomycota) from multigene phylogenetic analysis of type species," *FEMS Yeast Research*, vol. 13, no. 1985, pp. 23–33, 2013.

[195] E. Rozp dowska, L. Hellborg, O. P. Ishchuk, F. Orhan, S. Galafassi, A. Merico, M. Woolfit, C. Compagno, and J. Piskur, "Parallel evolution of the make-accumulate-consume strategy in *Saccharomyces* and *Dekkera* yeasts.," *Nature communications*, vol. 2, no. May, p. 302, 2011.

[196] C. D. Curtin, A. R. Borneman, P. J. Chambers, and I. S. Pretorius, "De-novo assembly and analysis of the heterozygous triploid genome of the wine spoilage yeast *Dekkera bruxellensis* AWRI1499.," *PloS one*, vol. 7, no. 3, p. e33840, 2012.

[197] J. Piškur, Z. Ling, M. Marcet-Houben, O. P. Ishchuk, A. Aerts, K. LaButti, A. Copeland, E. Lindquist, K. Barry, C. Compagno, L. Bisson, I. V. Grigoriev, T. Gabaldón, and T. Phister, "The genome of wine yeast *Dekkera bruxellensis* provides a tool to explore its food-related properties," *International Journal of Food Microbiology*, vol. 157, pp. 202–209, 2012.

[198] J. Valdes, P. Tapia, V. Cepeda, J. Varela, L. Godoy, F. a. Cubillos, E. Silva, C. Martinez, and M. A. Ganga, "Draft genome sequence and transcriptome analysis of the wine spoilage yeast *Dekkera bruxellensis* LAMAP2480 provides insights into genetic diversity, metabolism and survival," *FEMS Microbiology Letters*, vol. 361, pp. 104–106, 2014.

[199] M. Woolfit, E. Rozp dowska, J. Piškur, and K. H. Wolfe, "Genome survey sequencing of the wine spoilage yeast *Dekkera (Brettanomyces) bruxellensis*," *Eukaryotic Cell*, vol. 6, pp. 721–733, 2007.

[200] S. Crauwels, B. Zhu, J. Steensels, P. Busschaert, G. De Samblanx, K. Marchal, K. a. Willems, K. J. Verstrepen, and B. Lievens, "Assessing genetic diversity among Brettanomyces yeasts by DNA fingerprinting and whole-genome sequencing," *Applied and Environmental Microbiology*, vol. 80, no. May, pp. 4398–4413, 2014.

[201] L. Hellborg and J. Piškur, "Complex nature of the genome in a wine spoilage yeast, *Dekkera bruxellensis*," *Eukaryotic Cell*, vol. 8, no. August, pp. 1739–1749, 2009.

[202] Z. Antunovics, H. V. Nguyen, C. Gaillardin, and M. Sipiczki, "Gradual genome stabilisation by progressive reduction of the *Saccharomyces Uvarum* genome in an interspecific hybrid with *Saccharomyces cerevisiae*," *FEMS Yeast Research*, vol. 5, pp. 1141–1150, 2005.

[203] G. Fischer, S. a. James, I. N. Roberts, S. G. Oliver, and E. J. Louis, "Chromosomal evolution in *Saccharomyces*," *Nature*, vol. 405, pp. 451–454, 2000.

[204] S. Crauwels, A. Van Assche, R. de Jonge, A. Borneman, C. Verreth, P. Troels, G. De Samblanx, K. Marchal, Y. Van de Peer, K. Willems, K. Verstrepen, C. Curtin, and B. Lievens, "Comparative phenomics and targeted use of genomics reveals variation in carbon and nitrogen assimilation amongst different *Brettanomyces bruxellensis strains*," *Appl. Microbiol. Biotechnol.*, 2015.

[205] S. Ohno, *Evolution by Gene Duplication*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1970.

[206] K. H. Wolfe and W.-H. Li, "Molecular evolution meets the genomics revolution.," *Nature genetics*, vol. 33 Suppl, no. march, pp. 255–265, 2003.

[207] V. Renouf and A. Lonvaud-Funel, "Development of an enrichment medium to detect *Dekkera/Brettanomyces bruxellensis*, a spoilage wine yeast, on the surface of grape berries," *Microbiological Research*, vol. 162, pp. 154–167, 2007.

[208] I. a. Tiukova, M. E. Petterson, C. Tellgren-Roth, I. Bunikis, T. Eberhard, O. V. Pettersson, and V. Passoth, "Transcriptome of the Alternative Ethanol Production Strain *Dekkera bruxellensis* CBS 11270 in Sugar Limited, Low Oxygen Cultivation," *PLoS ONE*, vol. 8, no. 3, pp. 2–8, 2013.

[209] W. De Barros Pita, F. C. B. Leite, A. T. De Souza Liberal, D. A. Simões, and M. A. De Morais, "The ability to use nitrate confers advantage to *Dekkera bruxellensis* over *S. cerevisiae* and can explain its adaptation to industrial fermentation processes," *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology*, vol. 100, pp. 99–107, 2011.

[210] F. C. B. Leite, T. O. Basso, W. D. B. Pita, A. K. Gombert, D. A. Simões, and M. A. De Morais, "Quantitative aerobic physiology of the yeast *Dekkera bruxellensis*, A major contaminant in bioethanol production plants," *FEMS Yeast Research*, vol. 13, pp. 34–43, 2013.

[211] L. Conterno, C. M. L. Joseph, T. J. Arvik, T. Henick-Kling, and L. F. Bisson, "Genetic and physiological characterization of *Brettanomyces bruxellensis* strains isolated from wines," *American Journal of Enology and Viticulture*, vol. 57, pp. 139–147, 2006.

[212] P. Martorell, A. Barata, M. Malfeito-Ferreira, M. T. Fernández-Espinar, V. Loureiro, and A. Querol, "Molecular typing of the yeast species *Dekkera bruxellensis* and *Pichia guilliermondii* recovered from wine related sources," *International Journal of Food Microbiology*, vol. 106, pp. 79–84, 2006.

[213] C. Miot-Sertier and a. Lonvaud-Funel, "Development of a molecular method for the typing of *Brettanomyces bruxellensis* (*Dekkera bruxellensis*) at the strain level," *Journal of Applied Microbiology*, vol. 102, pp. 555–562, 2007.

[214] C. M. Mitrakul, T. Henick-Kling, and C. M. Egli, "Discrimination of *Brettanomyces/Dekkera* yeast isolates from wine by using various DNA finger- printing methods," *Food Microbiology*, vol. 16, pp. 3–14, 1999.

[215] I. Vigentini, G. De Lorenzis, C. Picozzi, S. Imazio, A. Merico, S. Galafassi, J. Piškur, and R. Foschino, "Intraspecific variations of *Dekkera/Brettanomyces bruxellensis* genome studied by capillary electrophoresis separation of the intron splice site profiles," *International Journal of Food Microbiology*, vol. 157, pp. 6–15, 2012.

[216] L. Daenen, B. Vanderhaegen, H. Verachtert, and G. Derdelinckx, "Flavour enhancement in beer by yeast beta-glucosidase activity.," *Communications in agricultural and applied biological sciences*, vol. 69, no. 2, pp. 73–6, 2004.

[217] E. Pogorzelski and A. Wilkowska, "Flavour enhancement through the enzymatic hydrolysis of glycosidic aroma precursors in juices and wine beverages: a review," *Flavour and Fragrance Journal*, vol. 22, no. 4, pp. 251–254, 2007.

[218] C. T. Hittinger, P. Gonçalves, J. P. Sampaio, J. Dover, M. Johnston, and A. Rokas, "Remarkably ancient balanced polymorphisms in a multi-locus gene network.," *Nature*, vol. 464, no. March, pp. 54–58, 2010.

[219] C. D. Curtin, J. R. Bellon, P. a. Henschke, P. W. Godden, and M. a. De Barros Lopes, "Genetic diversity of *Dekkera bruxellensis* yeasts isolated from Australian wineries," *FEMS Yeast Research*, vol. 7, pp. 471–481, 2007.

[220] W. Albertin, A. Panfili, C. Miot-Sertier, A. Goulielmakis, A. Delcamp, F. Salin, A. Lonvaud-Funel, C. Curtin, and I. Masneuf-Pomarede, "Development of microsatellite markers for the rapid and reliable genotyping of *Brettanomyces bruxellensis* at strain level," *Food Microbiology*, vol. 42, pp. 188–195, 2014.

[221] S. S. González, E. Barrio, and A. Querol, "Molecular characterization of new natural hybrids of *Saccharomyces cerevisiae* and *S. kudriavzevii* in brewing.," *Applied and environmental microbiology*, vol. 74, no. 8, pp. 2314–20, 2008.

[222] J. P. van Der Walt, "A new genus of the *Saccharomycetaceae*," *Antonie van Leeuwenhoek*, vol. 30, pp. 273–280, 1964.

[223] M. Smith and A. van Grinsven, "*Dekkera anomala* sp. nov., the teleomorph of *Brettanomyces anomalus*, recovered from spoiled soft drinks," *Antonie van Leeuwenhoek*, vol. 50, pp. 143–148, 1984.

[224] Miklenić, Marina, A. Štafa, A. Baji, B. Žunar, B. Lisni, and I. K. Svetec, "Genetic transformation of the yeast *Dekkera/Brettanomyces bruxellensis* with non-homologous DNA," *Journal of Microbiology and Biotechnology*, vol. 23, pp. 674–680, 2013.

[225] A. J. Schifferdecker, J. Siurkus, M. R. Andersen, D. Joerck-Ramberg, Z. Ling, N. Zhou, J. E. Blevins, A. A. Sibirny, J. Piškur, and O. P. Ishchuk, "Alcohol dehydrogenase gene ADH3 activates glucose alcoholic fermentation in genetically engineered *Dekkera bruxellensis* yeast," *Appl. Microbiol. Biotechnol.*, vol. 100, pp. 3219–3231, Apr 2016.

[226] J. Peter and J. Schacherer, "Population genomics of yeasts: towards a comprehensive view across a broad evolutionary scale," *Yeast*, vol. 33, pp. 73–81, Mar 2016.

[227] P. Almeida, C. Gonçalves, S. Teixeira, D. Libkind, M. Bontrager, I. Masneuf-Pomarède, W. Albertin, P. Durrens, D. J. Sherman, P. Marullo, C. T. Hittinger, P. Gonçalves, and J. P. Sampaio, "A Gondwanan imprint on global diversity and domestication of wine and cider yeast *Saccharomyces uvarum*.," *Nature communications*, vol. 5, no. May, p. 4044, 2014.

[228] D. C. Jeffares, C. Rallis, A. Rieux, D. Speed, M. Převorovský, T. Mourier, F. X. Marsellach, Z. Iqbal, W. Lau, T. M. K. Cheng, R. Pracana, M. Mülleder, J. L. D. Lawson, A. Chessel, S. Bala, G. Hellenthal, B. O'Fallon, T. Keane, J. T. Simpson, L. Bischof, B. Tomiczek, D. A. Bitton, T. Sideri, S. Codlin, J. E. E. U. Hellberg, L. van Trigt, L. Jeffery, J.-J. Li, S. Atkinson, M. Thodberg, M. Febrer, K. McLay, N. Drou, W. Brown, J. Hayles, R. E. Carazo Salas, M. Ralser, N. Maniatis, D. J. Balding, F. Balloux, R. Durbin, and J. Bähler, "The genomic and phenotypic diversity of *Schizosaccharomyces* pombe.," *Nature genetics*, vol. 47, no. 3, pp. 235–41, 2015.

5

Scientific problem addressed and aims of the project

The evolutionary adaptation of organisms to specific environments is one of the most fascinating processes in biology. Diversity arises from the interaction between variation and selection, both natural and human-mediated. Human selection and breeding (unintentional or deliberate) of wild species to obtain cultivated variants that thrive in man-made environments, but behave sub-optimally in nature, characterise the process of domestication. Domestication may co-occur or be followed by further diversification that results in variants adapted to different man-made conditions and applications. In response to domestication, certain traits beneficial to humans or as a by-product of the artificial environmental conditions increase in frequency and often reach fixation. These heritable traits describe the so called "domestication syndrome" and distinguish the new variants from their progenitor wild variants.

Domestication is a complex evolutionary process and, more importantly, long-term and continuous. This implies that defining the how, when, where and how many times is nor straightforward neither completely correct. The relationship between wild and domesticated is often seen as unidirectional, obscuring the fact that it represents a long-term transformation involving multiple "back and forth" events and transition states [1]. Over the past 13,000 years, hundreds of plants and animals have been independently domesticated into crops, livestock and pets across the globe, affecting human history at unprecedented levels [2, 3]. Domestication is a well-studied process in plants and animals. Multiple genetic and phenotypic changes have been reported as signatures of

this process: genome decay, polyploidy, presence of adaptive horizontally transferred genes, chromosomal rearrangements, gene duplications and phenotypes resulting from human-driven selection [4–10].

The creation of multiple, novel and highly specialised man-made environments also shaped microbial diversity, allowing for rapid adaptation and diversification of different microorganisms. For example, the filamentous fungus *Aspergillus oryzae* is used for the production of saké, miso and soy sauce [11]; multiple *Penicillium* species are applied to the production of cheese and cured meat [12]; the baker's yeast, *Saccharomyces cerevisiae*, arguably the most iconic example, is used for the production of fermented food and beverages such as wine, beer, sake, spirits, bread and for the production of cellulosic biofuel. For thousands of years humans have exploited the capacity of *S. cerevisiae* to convert sugars into ethanol and desirable flavour compounds [13–15]. While the use of pure cultures only started well after the work of Pasteur and Hansen in the mid-late 19th century, early brewers, winemakers and bakers had already learned that inoculating unfermented foods with a small portion of fermented product resulted in fast and more predictable fermentations. This practice, called "backslopping", might have resulted in yeast lineages that grew continuously in man-made environments and lost contact with their natural niches, providing a perfect setting for domestication. However, strong evidence for this hypothesis was still missing and it was poorly understood whether industrial yeast diversity is shaped by selection and niche adaptation or neutral divergence caused by geographic isolation and limited dispersal [16–18].

Several studies recently investigated *S. cerevisiae* evolution by sequencing the genomes of hundreds of different strains, providing a first glimpse of the complex evolution of this species [19–25]. However, most of these studies focus primarily on yeasts from wild and clinical habitats, and often include only a limited set of industrial strains, mainly originating from wine. Moreover, most studies use haploid derivatives instead of natural strains and can therefore not explore typical patterns of domestication like poly- and aneuploidy and heterozygosity. The use of haploids also excludes a large fraction of industrial *S. cerevisiae* strains that have lost the ability to sporulate, such as the vast majority of beer yeasts. Nevertheless, signatures of domestication have already been revealed in wine strains such as an increased resistance to copper (present in grapevine pesticides) and sulphite (used as a preservative in wine) [17, 26, 27]. However, an in-depth investigation on the origin of the rich biodiversity of industrial yeasts we observe today was still lacking at the beginning of this project.

In this project, we performed whole-genome sequencing of a collection of more than 200 *Saccharomyces* yeasts directly used in the industrial production of multiple fermentation products (wine, beer, saké, bioethanol, bread and spirits) or associated with industrial environments (e.g. brewing contaminants and yeasts isolated from old bottles). We charted their phenotypic diversity by measuring more than 90 distinct traits such as environmental stresses, growth on different sugars, aroma production, mating behaviour and fermentation performance. Our overall goals included:

1. Investigate the genetic and phenotypic relationships of *Saccharomyces* industrial yeasts compared to natural and clinical isolates and across different niches of isolation.

2. Trace the origin and evolutionary history of *Saccharomyces* industrial yeasts.

3. Define the impact of ecological factors in the diversification of *Saccharomyces* industrial yeasts.

4. Identify genetic and phenotypic signatures of adaptation to man-made environments.

5. Provide a model of domestication for *Saccharomyces* industrial yeasts, specifically, and for microbial domestication generally.

Strains were obtained from historical yeast collections mainly of the VIB Laboratory for Systems Biology (KU Leuven, Belgium) and White Labs (USA).

The first goal of our project was to identify the species of the yeasts sequenced. Whereas *S. cerevisiae* is the predominant species in many types of fermentations, it is well known that lager beers are fermented by *S. pastorianus*, an interspecific hybrid between *S. cerevisiae* and *S. eubayanus* [28]. Many brewers and winemakers use a specific yeast strain based primarily on historical reasons and the genetic background is often unknown or poorly documented. For example, we identified 10 *S. cerevisiae* strains employed for the production of lager beers and 1 *S. pastorianus* strain employed for the production of whiskey. Most interestingly, the sequenced collection included more than 70 interspecific hybrids such as the aforementioned *S. pastorianus* but also *S. cerevisiae* x *S. kudriavzevii* hybrids (mainly in Trappist beers) and *S. uvarum* x *S. eubayanus* hybrids (in Lambic beers and as contaminant of the brewing environment).

For this reason in **Chapter 6 (Domestication and Divergence of *Saccharomyces cerevisiae* Beer Yeasts, p. 83)** we provide in depth genetic and phenotypic analyses for the *S. cerevisiae* strains and in **Chapter 8 (Interspecific Hybridization Facilitates Niche Adaptation in Beer Yeasts, p. 151)** we expand our analyses to the interspecific hybrids and to the role of interspecific hybridization in the evolution of *Saccharomyces* yeasts in man-made environments. **Chapter 7 (Origins, Evolution, Domestication and Diversity of *Saccharomyces* Beer Yeasts, p. 133)** represents a transition between **Chapter 6** and **Chapter 8** where we discuss how, based on our findings, the brewing environment represents the perfect model to study microbial domestication. In **Chapter 9, (Conclusions and future prospects, p. 203)** we end with an overall discussion of the main findings reported in this dissertation and we provide an overview of current and future challenges to be addressed in the field.

5.1 References

[1] K. Dobney and G. Larson, "Genetics and animal domestication: new windows on an elusive process," *Journal of Zoology*, vol. 269, no. 2, pp. 261–271, 2006.

[2] J. Diamond, "Evolution, consequences and future of plant and animal domestication," *Nature*, vol. 418, no. 6898, pp. 700–707, 2002.

[3] G. Larson, D. R. Piperno, R. G. Allaby, M. D. Purugganan, L. Andersson, M. Arroyo-Kalin, L. Barton, C. Climer Vigueira, T. Denham, K. Dobney, A. N. Doust, P. Gepts, M. T. P. Gilbert, K. J. Gremillion, L. Lucas, L. Lukens, F. B. Marshall, K. M. Olsen, J. C. Pires, P. J. Richerson, R. Rubio de Casas, O. I. Sanjur, M. G. Thomas, and D. Q. Fuller, "Current perspectives and the future of domestication studies.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 111, no. 17, pp. 6139–46, 2014.

[4] J. Y. Choi, A. E. Platts, D. Q. Fuller, Y.-I. Hsing, R. A. Wing, and M. D. Purugganan, "The rice paradox: Multiple origins but single domestication in Asian rice," *Molecular Biology and Evolution*, vol. 34, no. 4, pp. 969–979, 2017.

[5] C. A. Driscoll, M. Menotti-Raymond, A. L. Roca, K. Hupe, W. E. Johnson, E. Geffen, E. H. Harley, M. Delibes, D. Pontier, A. C. Kitchener, N. Yamaguchi, S. J. O'brien, and D. W. Macdonald, "The Near Eastern origin of cat domestication.," *Science (New York, N.Y.)*, vol. 317, no. 5837, pp. 519–23, 2007.

[6] C. A. Driscoll, D. W. Macdonald, and S. J. O'Brien, "From wild animals to domestic pets, an evolutionary view of domestication.," *Proceedings of the National Academy of Sciences of the United States of America*, no. Supplement 1, pp. 9971–8, 2009.

[7] L. A. F. Frantz, V. E. Mullin, M. Pionnier-Capitan, O. Lebrasseur, M. Ollivier, A. Perri, A. Linderholm, V. Mattiangeli, M. D. Teasdale, E. A. Dimopoulos, A. Tresset, M. Duffraisse, F. McCormick, L. Bartosiewicz, E. Gál, É. A. Nyerges, M. V. Sablin, S. Bréhard, M. Mashkour, A. Bălăşescu, B. Gillet, S. Hughes, O. Chassaing, C. Hitte, J.-D. Vigne, K. Dobney, C. Hänni, D. G. Bradley, and G. Larson, "Genomic and archaeological evidence suggest a dual origin of domestic dogs.," *Science (New York, N.Y.)*, vol. 352, no. 6290, pp. 1228–31, 2016.

[8] J. Ramos-Madrigal, B. D. Smith, J. V. Moreno-Mayar, S. Gopalakrishnan, J. Ross-Ibarra, M. T. P. Gilbert, and N. Wales, "Genome Sequence of a 5,310-Year-Old Maize Cob Provides Insights into the Early Stages of Maize Domestication," *Current Biology*, vol. 26, no. 23, pp. 3195–3201, 2016.

[9] A. Salman-Minkov, N. Sabath, I. Mayrose, N. C. Ellstrand, and L. H. Rieseberg, "Whole-genome duplication as a key factor in crop domestication," *Nature Plants*, vol. 2, no. 8, p. 16115, 2016.

[10] Z. Zhou, Y. Jiang, Z. Wang, Z. Gou, J. Lyu, W. Li, Y. Yu, L. Shu, Y. Zhao, Y. Ma, C. Fang, Y. Shen, T. Liu, C. Li, Q. Li, M. Wu, M. Wang, Y. Wu, Y. Dong, W. Wan, X. Wang, Z. Ding, Y. Gao, H. Xiang, B. Zhu, S.-H. Lee, W. Wang, and Z. Tian, "Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean," *Nature Biotechnology*, vol. 33, no. 4, pp. 408–414, 2015.

[11] J. Gibbons, L. Salichos, J. Slot, D. Rinker, K. McGary, J. King, M. Klich, D. Tabb, W. McDonald, and A. Rokas, "The Evolutionary Imprint of Domestication on Genome Variation and Function of the Filamentous Fungus *Aspergillus oryzae*," *Current Biology*, vol. 22, no. 15, pp. 1403–1409, 2012.

[12] K. Cheeseman, J. Ropars, P. Renault, J. Dupont, J. Gouzy, A. Branca, A.-L. Abraham, M. Ceppi, E. Conseiller, R. Debuchy, F. Malagnac, A. Goarin, P. Silar, S. Lacoste, E. Sallet, A. Bensimon, T. Giraud, and Y. Brygoo, "Multiple recent horizontal transfers of a large genomic region in cheese making fungi," *Nature Communications*, vol. 5, no. 1, p. 2876, 2014.

[13] P. J. Chambers and I. S. Pretorius, "Fermenting knowledge: the history of winemaking, science and yeast research," *EMBO Rep.*, vol. 11, no. 12, pp. 914–920, 2010.

[14] P. E. McGovern, A. Mirzoian, and G. R. Hall, "Ancient Egyptian herbal wines," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 106, no. 18, pp. 7361–7366, 2009.

[15] J. Steensels and K. J. Verstrepen, "Taming Wild Yeast: Potential of Conventional and Nonconventional Yeasts in Industrial Fermentations," *Annual Review of Microbiology*, vol. 68, no. 1, pp. 61–80, 2014.

[16] M. R. Goddard and D. Greig, "*Saccharomyces cerevisiae*: a nomadic yeast with no niche?," *FEMS yeast research*, p. fov009, 2015.

[17] J. Warringer, E. Zörgö, F. a. Cubillos, A. Zia, A. Gjuvsland, J. T. Simpson, A. Forsmark, R. Durbin, S. W. Omholt, E. J. Louis, G. Liti, A. Moses, and A. Blomberg, "Trait variation in yeast is defined by population history.," *PLoS genetics*, vol. 7, no. 6, p. e1002111, 2011.

[18] E. Zörgö, A. Gjuvsland, F. a. Cubillos, E. J. Louis, G. Liti, A. Blomberg, S. W. Omholt, and J. Warringer, "Life history shapes trait heredity by accumulation of loss-of-function alleles in yeast.," *Molecular biology and evolution*, vol. 29, no. 7, pp. 1781–9, 2012.

[19] P. Almeida, R. Barbosa, P. Zalar, Y. Imanishi, K. Shimizu, B. Turchetti, J.-L. Legras, M. Serra, S. Dequin, A. Couloux, J. Guy, D. Bensasson, P. Gonçalves, and J. P. Sampaio, "A population genomics insight into the Mediterranean origins of wine yeast domestication.," *Molecular ecology*, vol. 24, no. 21, pp. 5412–27, 2015.

[20] A. R. Borneman, B. A. Desany, D. Riches, J. P. Affourtit, A. H. Forgan, I. S. Pretorius, M. Egholm, and P. J. Chambers, "Whole-genome comparison reveals novel genetic elements that characterize the genome of industrial strains of *Saccharomyces cerevisiae*.," *PLoS genetics*, vol. 7, no. 2, p. e1001287, 2011.

[21] A. R. Borneman, A. H. Forgan, R. Kolouchova, J. A. Fraser, and S. A. Schmidt, "Whole Genome Comparison Reveals High Levels of Inbreeding and Strain Redundancy Across the Spectrum of Commercial Wine Strains of *Saccharomyces cerevisiae*.," *G3 (Bethesda, Md.)*, vol. 6, no. 4, pp. 957–971, 2016.

[22] M. Gonçalves, A. Pontes, P. Almeida, R. Barbosa, M. Serra, D. Libkind, M. Hutzler, P. Gonçalves, J. Sampaio, W. Heijne, and et Al., "Distinct Domestication Trajectories in Top-Fermenting Beer Yeasts and Wine Yeasts," *Current Biology*, vol. 26, no. 20, pp. 2750–2761, 2016.

[23] J.-L. Legras, V. Galeote, F. Bigey, C. Camarasa, S. Marsit, T. Nidelet, I. Sanchez, A. Couloux, J. Guy, R. Franco-Duarte, M. Marcet-Houben, T. Gabaldon, D. Schuller, J. P. Sampaio, and S. Dequin, "Adaptation of *S. cerevisiae* to Fermented Food Environments Reveals Remarkable Genome Plasticity and the Footprints of Domestication," *Molecular Biology and Evolution*, vol. 35, no. 7, pp. 1712–1727, 2018.

[24] J. Peter, M. De Chiara, A. Friedrich, J.-X. Yue, D. Pflieger, A. Bergström, A. Sigwalt, B. Barre, K. Freel, A. Llored, C. Cruaud, K. Labadie, J.-M. Aury, B. Istace, K. Lebrigand, P. Barbry, S. Engelen, A. Lemainque, P. Wincker, G. Liti, and J. Schacherer, "Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates," *Nature*, vol. 556, no. 7701, pp. 339–344, 2018.

[25] P. K. Strope, D. a. Skelly, S. G. Kozmin, G. Mahadevan, E. a. Stone, P. M. Magwene, F. S. Dietrich, J. H. Mccusker, N. Carolina, B. Sciences, and N. Carolina, "The 100-genomes strains, an *S. cerevisiae* resource that illuminates its natural phenotypic and genotypic variation and emergence as an opportunistic pathogen," *Genome Res.*, vol. 25, no. 5, pp. 762–774, 2015.

[26] M. Novo, F. Bigey, E. Beyne, V. Galeote, F. Gavory, S. Mallet, B. Cambon, J.-L. Legras, P. Wincker, S. Casaregola, and S. Dequin, "Eukaryote-to-eukaryote gene transfer events revealed by the genome sequence of the wine yeast *Saccharomyces cerevisiae* EC1118.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 38, pp. 16333–8, 2009.

[27] J. E. Pérez-Ortín, A. Querol, S. Puig, and E. Barrio, "Molecular characterization of a chromosomal rearrangement involved in the adaptive evolution of yeast strains.," *Genome research*, vol. 12, no. 10, pp. 1533–9, 2002.

[28] D. Libkind, C. T. Hittinger, E. Valério, C. Gonçalves, J. Dover, M. Johnston, P. Gonçalves, and J. P. Sampaio, "Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 35, pp. 14539–44, 2011.

6

Domestication and Divergence of *Saccharomyces cerevisiae* Beer Yeasts

This chapter is adapted from:

Brigida Gallone, Jan Steensels, Troels Prahl, Leah Soriaga, Veerle Saels, Beatriz Herrera-Malaver, Adriaan Merlevede, Miguel Roncoroni, Karin Voordeckers, Loren Miraglia, Clotilde Teiling, Brian Steffy, Maryann Taylor, Ariel Schwartz, Toby Richardson, Christopher White, Guy Baele, Steven Maere and Kevin J. Verstrepen. Domestication and Divergence of *Saccharomyces cerevisiae* Beer Yeasts. *Cell*, 166, 1397-1410, September 8, 2016.

Author contributions:

Conceptualization: B.G., J.S., T.P., L.M., S.M., K.J.V; Data collection: B.G., J.S.; Formal analysis: B.G., J.S., L.S., M.R.; Investigation: B.G., J.S., T.P., A.M., L.S., V.S., B.H., K.V., M.T.; Data Visualisation: B.G.; Resources: L.M., B.S., C.T., C.W., K.J.V; Writing: B.G., J.S., S.M., K.J.V; Supervision: T.P., T.R., A.S., C.W., G.B., S.M., K.J.V.

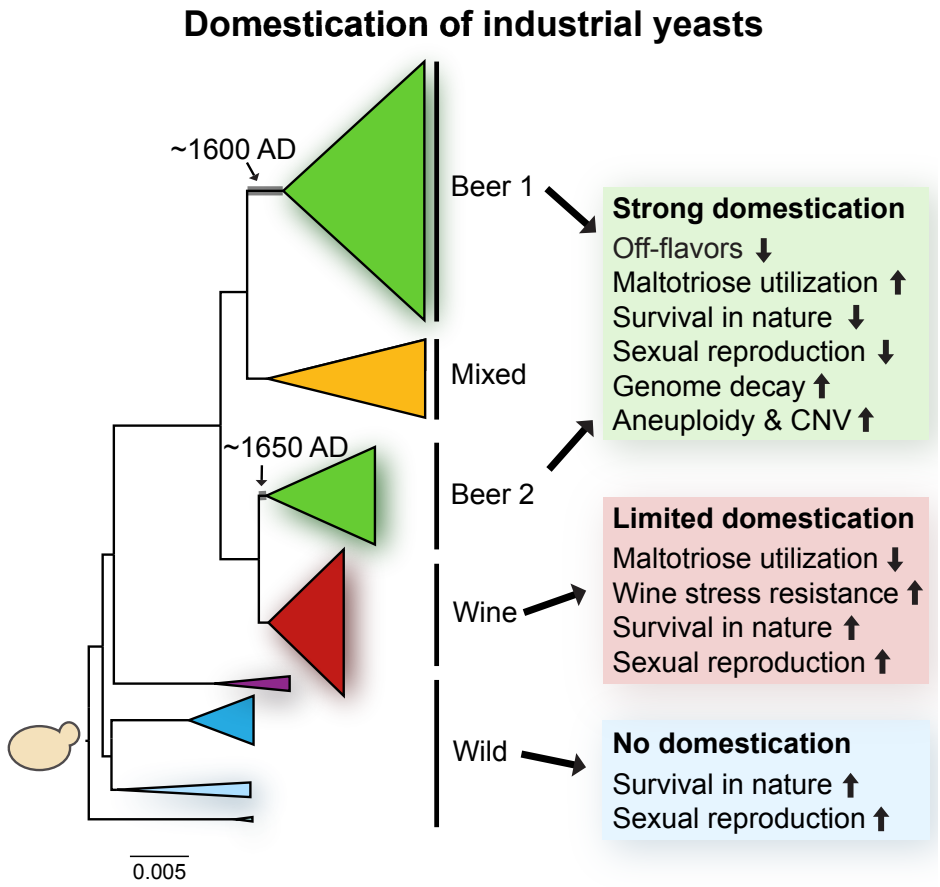


Figure 6.1: Graphical Abstract. Domestication of industrial yeasts

6.1 Abstract

Whereas domestication of livestock, pets, and crops is well documented, it is still unclear to what extent microbes associated with the production of food have also undergone human selection and where the plethora of industrial strains originates from. Here, we present the genomes and phenomes of 157 industrial *Saccharomyces cerevisiae* yeasts. Our analyses reveal that today's industrial yeasts can be divided into five sublineages that are genetically and phenotypically separated from wild strains and originate from only a few ancestors through complex patterns of domestication and local divergence. Large-scale phenotyping and genome analysis further show strong industry-specific selection for stress tolerance, sugar utilization, and flavor production, while the sexual cycle and other phenotypes related to survival in nature show decay, particularly in beer yeasts. Together, these results shed light on the origins, evolutionary history, and phenotypic diversity of industrial yeasts and provide a resource for further selection of superior strains.

6.2 Introduction

Since prehistoric times, humans have exploited the capacity of the common baker's yeast *Saccharomyces cerevisiae* to convert sugars into ethanol and desirable flavor compounds to obtain foods and beverages with a prolonged shelf-life, enriched sensorial palate, improved digestibility, and an euphoriant effect due to the presence of ethanol [1, 2]. Whereas the use of pure cultures started well after the pioneering work of Pasteur and Hansen in the 19th century, early brewers, winemakers, and bakers had already learned that inoculating unfermented foods with a small portion of fermented product resulted in fast and more predictable fermentations. This so-called "backslopping" might have resulted in yeast lineages that grew continuously in these man-made environments and lost contact with their natural niches, providing a perfect setting for domestication. However, strong evidence for this hypothesis is still missing and it remains unclear whether industrial yeast diversity is shaped by selection and niche adaptation (domestication) or neutral divergence caused by geographic isolation and limited dispersal [3, 4].

Domestication is defined as human selection and breeding of wild species to obtain cultivated variants that thrive in man-made environments, but behave suboptimally in nature. Typical signs of domestication, including genome decay, polyploidy, chromosomal rearrangements, gene duplications, and phenotypes resulting from human-driven selection, have been reported in crops, livestock, and pets [5, 6]. Several studies have recently investigated the *S. cerevisiae* population by sequencing the genomes of hundreds of different strains, providing a first glimpse of the complex evolution of this species [7–13]. However, most of these studies focused primarily on yeasts from wild and clinical habitats and often include only a limited set of industrial strains, mainly originating from wine. Moreover, most studies use haploid derivatives instead of natural strains and can therefore not explore typical patterns of domestication like polyploidy, aneuploidy, and heterozygosity. The use of haploids also excludes a large fraction of industrial *S. cerevisiae* strains that have lost the ability to sporulate, such as the vast majority of beer yeasts. Nevertheless, some studies already revealed signs of domestication in wine strains, such as an increased resistance to copper (present in grapevine pesticides) and sulfite (used as a preservative in wine) [4, 14]. An in-depth investigation of strains originating from other industrial niches is still lacking.

Here, we describe the high-quality sequencing, de novo assembly, annotation, and extensive phenotyping of 157 *S. cerevisiae* strains used for the industrial production of beer, wine, bread, spirits, saké, and bioethanol, in their natural ploidy. Our data reveal that industrial yeasts are genetically and phenotypically distinct from wild strains and stem from only a limited set of ancestral strains that have been adapting to man-made environments. They further diversified into five clades: one including Asian strains such as saké yeasts, one mostly containing wine yeasts, a mixed clade that contains bread and other yeasts, and two separate families of beer yeasts. While most clades lack strong geographical substructure, one of the beer clades contains geographically isolated subgroups of strains used in continental Europe (Belgium/Germany), the United Kingdom,

and a recent sublineage of United States beer yeasts that diverged from the British subclade during colonization. Interestingly, these beer yeast lineages exhibit clear and profound hallmarks of domestication, more so than the other lineages. The shift from variable, complex, and often harsh environments encountered in nature to more stable and nutrient-rich beer medium favored specialized adaptations in beer yeasts, but also led to genome decay, aneuploidy, and loss of a functional sexual cycle. Specifically, we find evidence for active human selection, demonstrated by convergent evolution for efficient fermentation of beer-specific carbon sources, mainly through mutations and duplications of the *MAL* (maltose) genes, as well as nonsense mutations in *PAD1* and *FDC1*, which are involved in the production of 4-vinyl guaiacol (4-VG), an undesirable off-flavor in beer. Our results further suggest that beer yeast domestication was initiated hundreds of years ago, well after the first reported beer production, but before the discovery of microbes. Together, our results reveal how today's industrial yeasts are the outcome of centuries of human domestication and provide a new resource for further selection and breeding of superior variants.

6.3 Results

6.3.1 Niche and Geography Drive Diversification

To examine the evolutionary history of industrial yeasts, we sequenced the genomes of 157 *S. cerevisiae* isolates originating from various sources in their natural ploidy to a median coverage of 1353 (min = 263, max = 4033). This collection includes 102 industrial beer strains, 19 wine strains, 11 spirit strains, 7 saké strains, 7 strains isolated from spontaneous fermentations, 5 bioethanol strains, 4 bread strains, and 2 laboratory strains **(Table 6.3, p. 127)**. Interestingly, ten of these *S. cerevisiae* beer strains are used for commercial production of lager beers, which were believed to be exclusively produced by strains of the genetically related *Saccharomyces pastorianus*. After de novo assembly of each of the genomes, we inferred a maximum-likelihood phylogenetic tree based on codon alignments for 2,020 concatenated single copy nuclear genes shared by each of the 157 isolates and the outgroup species *Saccharomyces paradoxus* **(Figure 6.13a, p. 106)**. Additionally, we included a representative set of 24 previously sequenced strains belonging to the main established lineages of the *S. cerevisiae* phylogeny [10, 13], extending the number of strains to 181 **(Figure 6.2, p. 87)**. Trees constructed from the original and extended datasets are congruent and show five main lineages that contain the majority of industrial yeasts: Wine (bootstrap support 100%), Beer 1 (86%), Beer 2 (56%), Asia (100%), and a Mixed lineage (99%) containing yeasts used in different industries. Three of these lineages (Beer 1, Beer 2, and Mixed) were not previously described.

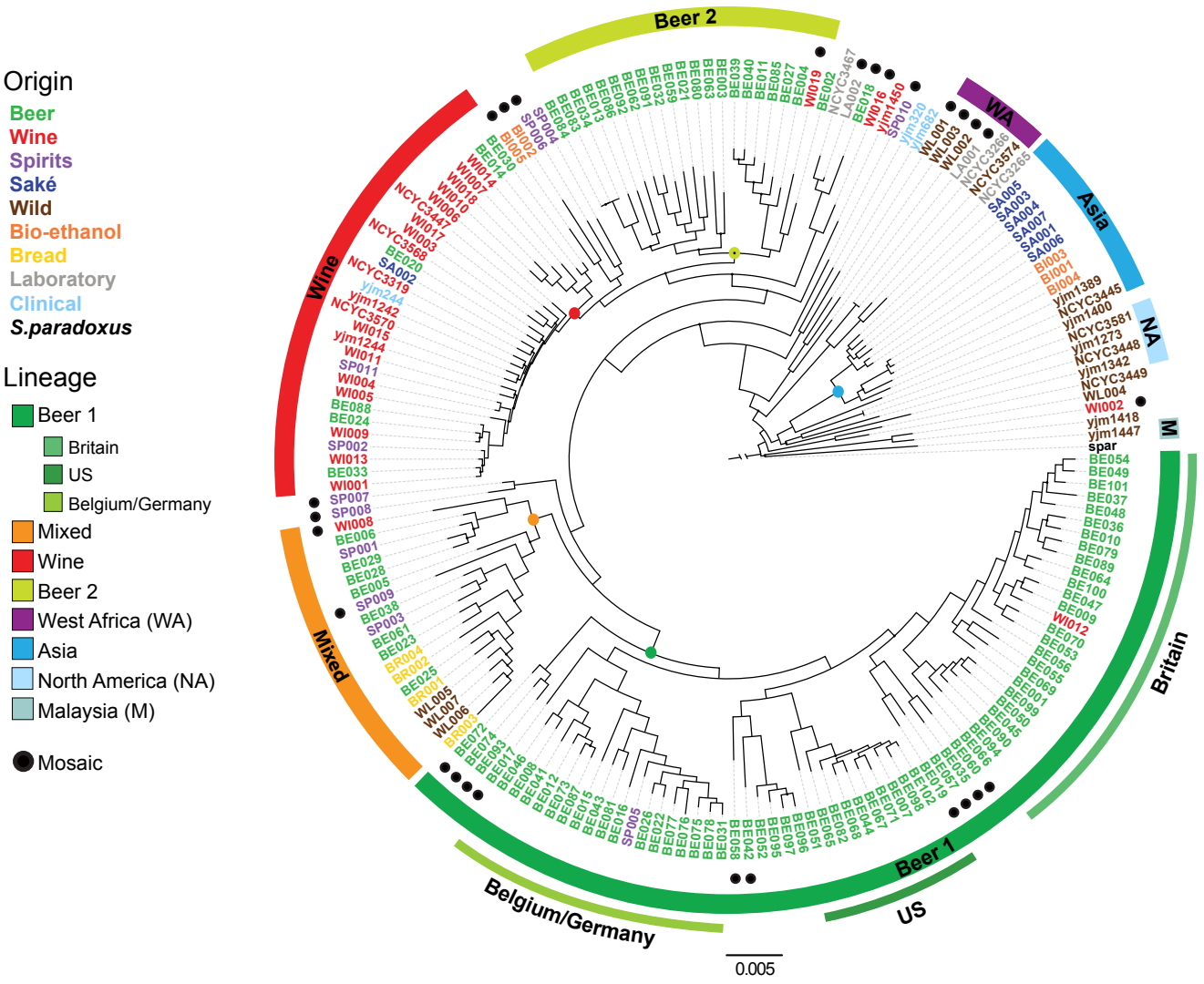


Figure 6.2: Phylogeny and population structure of industrial *S. cerevisiae* strains. Part 1. Maximum likelihood phylogenetic tree of all S. cerevisiae strains sequenced in this project supplemented with a representative set of 24 previously sequenced strains [10, 13] and using *Saccharomyces paradoxus* as an outgroup. Black dots on nodes indicate bootstrap support values <70%. Color codes indicate origin (names) and lineage (circular bands). The basal splits of the five industrial lineages are indicated with colored dots. Mosaic strains identified in this study are indicated with black dots next to the strain codes. Branch lengths reflect the average number of substitutions per site. Scale bar, 0.005 substitutions per site.

Next, we studied the population structure in a filtered set of 53,929 polymorphic sites accounting for 2,454,052 SNPs across all strains, using the Bayesian model-based clustering approach implemented in fastStructure [15] **(Figures 6.3a and 6.13b, p. 88 and p. 106, respectively)**. This analysis yields a population structure that is highly consistent with the major lineages defined in the phylogeny and identifies mosaicism in 17% of the strains (in which the estimated ancestry $Q < 0.8$ for $K = 8$ ancestral populations). The population structure is further supported by a principal component analysis (PCA) on the same SNP data **(Figure 6.3b, p. 88)**.

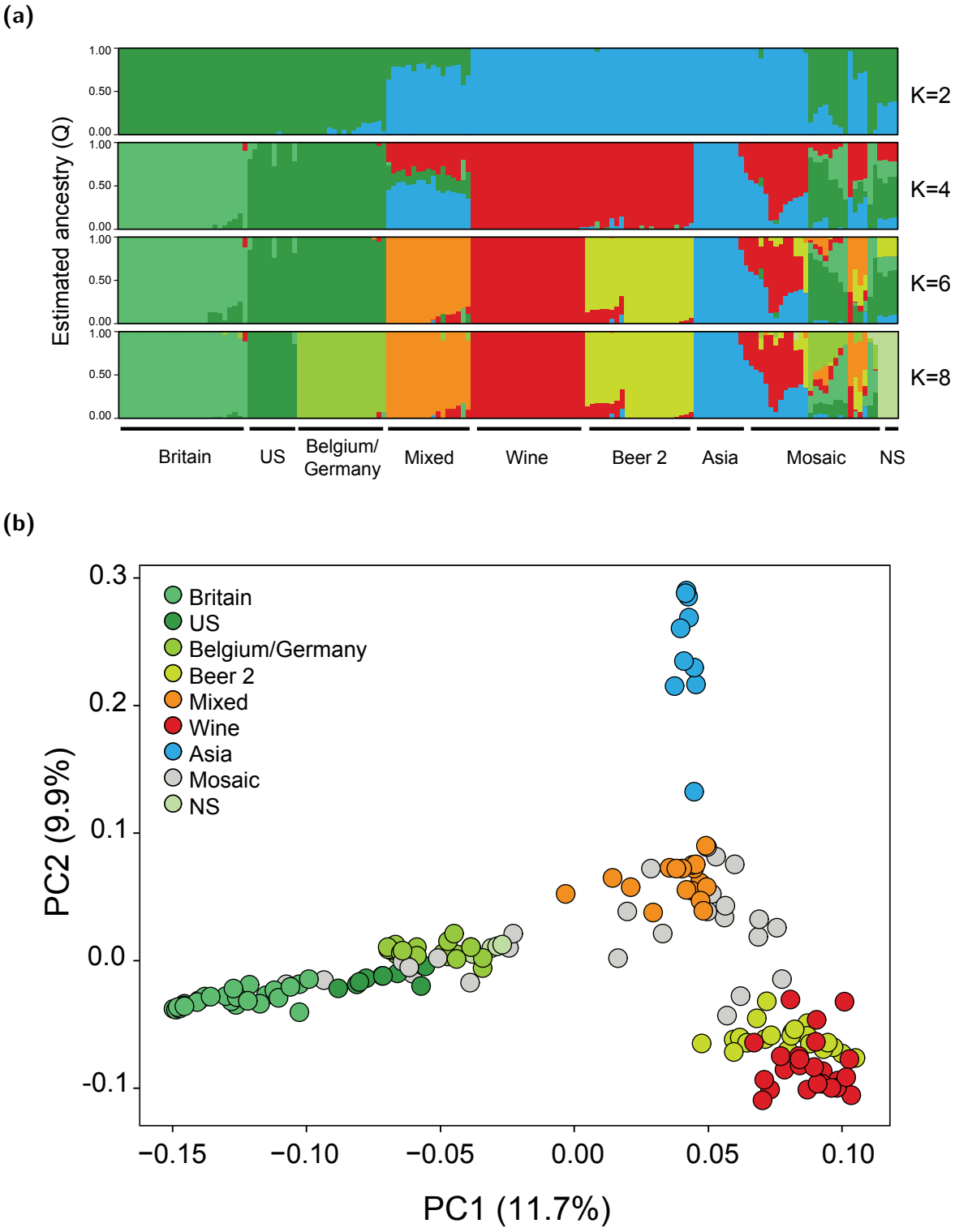


Figure 6.3: Phylogeny and population structure of industrial *S. cerevisiae* strains. Part 2. a: Population structure identified in the 157 surveyed strains. The vertical axis depicts the fractional representation of resolved populations (colors) within each strain (horizontal axis, strains listed in **Figure 6.14, p. 107**) for K = 2, 4, 6, and 8 assumed ancestral populations (where K = 8 maximizes the marginal likelihood and best explains the data structure). Mosaic strains (i.e., strains that possess <80% ancestry from a single population) are visualized as a separate group. **b:** Principal component projection, using the same set of SNPs as in Figure 6.3a. Colors represent different populations. WA, West Africa; NA, North America; M, Malaysia; NS, not specified.

Further analysis of the phylogeny and population structure reveals that the evolutionary divergence of industrial yeasts is shaped by both their industrial application and geographical origin. First, most yeasts cluster together according to the industry in which they are used and are clearly separated from the wild or clinical yeasts that have previously been sequenced. This was further confirmed by constructing a larger phylogeny, based on nine genomic regions, that includes the vast majority of all sequenced *S. cerevisiae* strains, 450 isolates in total **(Figure 6.14, p. 107; Table 6.4, p. 127)**. Wine and saké yeasts cluster in the previously identified Wine and Asia lineages [10]. The majority of beer yeasts (85.3%) are found in two main lineages (Beer 1 and Beer 2) that are only distantly related. The Mixed clade harbors 7.8% of all beer strains (most of which are atypical beer yeasts that are used for bottle refermentation of strong Belgian ales) and contains all bread strains. Interestingly, spirit strains lack this clear phylogenetic relationship, as they are highly mosaic and scattered throughout the tree, suggesting that these strains might be the result of breeding by modern-day yeast companies that sell yeasts for spirits production. Moreover, because spirit yeasts are typically not re-used after fermentation, they likely had less opportunity to diverge into a separate clade.

Within and between the lineages, we also observed geographical patterns. For example, most saké yeasts form a monophyletic group and cluster together with wild isolates and bioethanol strains from China, while South American bioethanol strains are closely related to strains used to produce cachaça, a Brazilian sugarcane spirit. Moreover, the Beer 1 clade consists of three separate subpopulations, each reflecting geographically distinct groups: Belgium/Germany, Britain, and the United States. The absence of genetic admixture among these subpopulations indicates that these strains diverged allopatrically after the initial split **(Figure 6.3a, p. 88)**. Moreover, the high nucleotide diversity within each of the Beer 1 sublineages exceeds that within the Wine population, suggesting that the split did not happen recently **(Table 6.1, p.90)**. Compared to Beer 1, Beer 2 is more closely related to the Wine lineage and includes 20.6% of all brewing strains. However, in contrast to the Beer 1 group, the Beer 2 lineage lacks geographic structure and contains yeasts originating from Belgium, the United Kingdom, the United States, Germany, and Eastern Europe. The presence of two major genetically distinct sources of beer yeasts hints toward two independent European domestication events, one of which is at the origin of both the Wine and Beer 2 clade.

Table 6.1: Genetic diversity within each subpopulation of industrial *S. cerevisiae* strains. The number of strains per subpopulation, the amount of analyzed and segregating sites, as well as nucleotide diversity (π) and population mutation rate (Watterson's θ, θw) are indicated.

Subpopulation	Number of Strains	Analyzed Sites	Segregating Sites	π	Θw
Britain	26	12,018,937	101,881	3.13E-03	1.88E-03
United States	10	11,973,239	72,559	2.31E-03	1.72E-03
Belgium/Germany	18	12,017,007	108,560	3.12E-03	2.19E-03
Mixed	17	12,043,532	132,188	4.35E-03	2.69E-03
Wine	24	12,052,956	114,133	1.59E-03	2.15E-03
Beer 2	21	12,063,361	142,745	2.95E-03	2.77E-03
Asia	10	12,035,745	99,879	2.39E-03	2.36E-03

6.3.2 Remarkable structural variation in beer yeasts

Variation in genome structure, such as polyploidy, aneuploidy, large segmental duplications, and copy-number variations (CNVs), have repeatedly been found in association with domestication and adaptation to specific niches in experimentally evolved microbes [8, 16–22] and in association with domestication of higher organisms [6].

Sequencing the yeast strains in their natural ploidy allowed analysis of gross chromosomal rearrangements and aneuploidies **(Figure 6.4, p. 91)**. We detected a staggering 15,288 deletion and amplification events across all strains, covering on average 1.57 Mb per strain. The size of the regions ranges from complete chromosomes (resulting in aneuploidies) to small local variations of a few kilobases (kb), all of which we will refer to as "CNVs." The extent of deletions significantly exceeds that of amplifications, respectively 1.07 Mb and 0.50 Mb on average per strain (2.15-fold difference, Wilcoxon signed rank test, $p < 0.001$). We observed significant variation among strains originating from different industries in the total frequency of CNV events (ANOVA F test, $p < 0.001$) and the fraction of the genome affected (ANOVA F-test, $p < 0.001$) **(Figure 6.15, p. 108)**. Pairwise comparisons of subpopulations and industries show no significant differences in the load of amplifications between strains from different industries or subpopulations, but we detected significant differences in the load of deletions between strains from the wine (median = 0.51 Mb) and beer (median = 0.94 Mb) industry (Tukey honest significant difference [HSD], $p < 0.05$) **(Figures 6.5a-6.5d, p. 92)**. This high incidence of CNV in beer strains goes together with a high incidence of polyploidy and aneuploidy ($R2 \sim 0.14$, $p < 0.001$; average genome content of 3.52, SD = 0.67, **Figures 6.4, 6.5e, and 6.5f, p. 91, 92, and 92, respectively**), which is linked to extensive chromosomal loss and general genome instability [23].

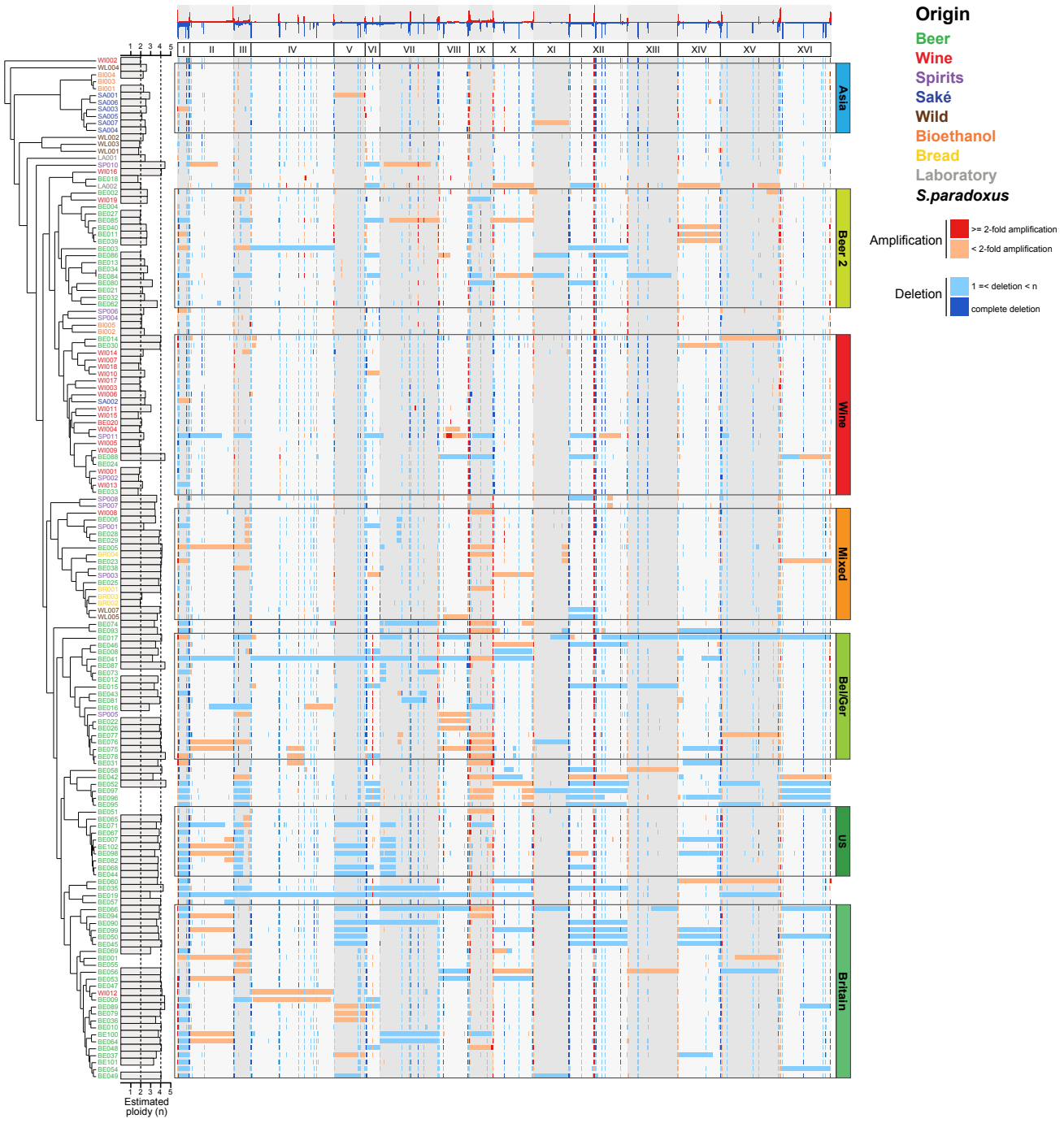


Figure 6.4: Ploidy and Copy-Number Variation in industrial *S. cerevisiae* strains. Part 1. Genome-wide visualization of copy-number variation (CNV) profiles, with the aggregate profile across all strains depicted on the top. Estimates for the nominal ploidy (n) values of the strains are represented by a bar chart next to the strain codes. Heat map colors reflect amplification (red shades) or deletion (blue shades) of genomic fragments. A distinction is made between completely deleted fragments (dark blue) and fragments of which at least one copy is still present (light blue). Similarly, highly amplified fragments (copy number R2-fold the basal ploidy) are depicted in dark red, while low and moderately amplified fragments (copy number <2-fold the basal ploidy) are depicted in orange. For strains with no estimated ploidy available, colors are only indicative of the presence of amplifications (orange) or deletions (light blue). Roman numbers indicate chromosome number. Strains are clustered according to their genetic relatedness as determined in **Figure 6.13a, p. 106**. Origin (name colors) and population (colored rectangles) are indicated on the figure.

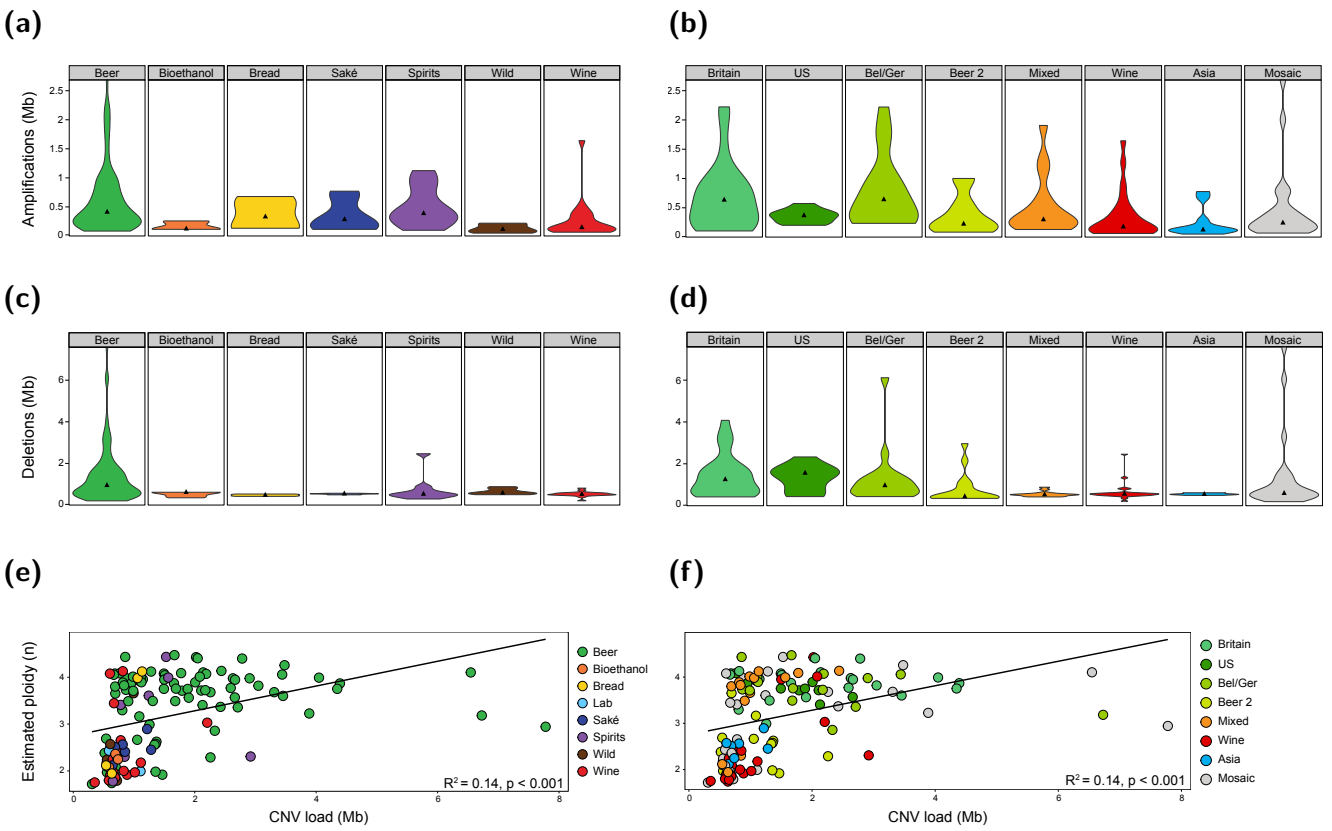


Figure 6.5: Ploidy and Copy-Number variation in industrial *S. cerevisiae* strains. Part 2. a-d: Violin plots describing the density of amplifications and deletions across different industries and subpopulations. Triangles indicate the median within each group. **e and f:** Correlations between levels of CNV load (Mb) and estimated ploidy (n), by industry and subpopulations.

CNVs are not uniformly spread across the genome. Considering subtelomere lengths of 33 kb [24], on average 39.7% of subtelomeric nucleotide positions are affected by CNV events compared to 9.54% of non-subtelomeric nucleotide positions (4.1-fold difference, Wilcoxon signed-rank test, $p < 0.001$). However, not all subtelomeres are equally prone to CNV: most variability is detected in ChrI, ChrVII, ChrVIII, ChrIX, ChrX, ChrXII, ChrXV, and ChrXVI **(Figure 6.4, p. 91)**. Gene ontology (GO) enrichment analysis reveals that genes involved in nitrogen and carbon metabolism, ion transport, and flocculation are most heavily influenced by CNVs **(Table 6.5, p. 127)**, which is in line with previous results[16, 18]. Interestingly, some CNVs seem linked to specific environments **(Table 6.6, p. 127)**, suggesting that CNVs may underlie niche adaptation. For example, many genes involved in uptake and breakdown of maltose (present in saké medium, main carbon source in beer, but absent from grape must) are amplified in beer and saké -related subpopulations, while they are often lost in strains from the Wine subpopulation (false discovery rate [FDR] q value < 0.001).

6.3.3 Relaxed Selection on Sex and Survival in Nature

Apart from selection for industrial traits, domestication is also characterized by relaxed selection and potential loss of costly traits that are not beneficial in the man-made environment [5]. In order to chart the phenome of our collection and investigate signs of selection for some traits and loss of others, 82 phenotypes, such as aroma production, sporulation characteristics, and tolerance to osmolytes, acids, ethanol, and low and high temperatures, were measured in all strains **(Figures 6.6 and 6.16, p. 94 and p. 109, respectively; Table 6.7, p. 127)**. Hierarchical clustering of the phenotypes resolves the main phylogenetic lineages and reveals a moderate correlation between genotype and phenotype distances between strains (Spearman correlation ~ 0.33), which is further increased (Spearman correlation ~ 0.36) when mosaic strains, for which genetic distance has no straightforward evolutionary interpretation, are omitted **(Figure 6.6, p. 94)**. Moreover, the clustering splits the collection into two main phenotypic subgroups: one largely overlapping with the Beer 1 clade that contains the majority of the Belgium/Germany, United States, and Britain beer yeasts as well as mosaic strains containing major genome fractions of these subpopulations and a second one where the remaining genetic subpopulations are strongly overrepresented (Fisher's exact test, Bonferroni corrected $p < 0.001$). Overall, strains from the Beer 1 clade perform poorly in general stress conditions that are not usually encountered in the brewing environment **(Figure 6.16, p. 109; Table 6.9, p. 127)**. In contrast, strains from the Wine subpopulation show superior performance in general stress conditions, which likely reflects the high-sugar and high-alcohol environments encountered in wine-making, as well as survival in potentially nutrient-poor and harsh natural environments in between the grape harvest seasons.

Saccharomyces cerevisiae is a facultative sexual organism. While its main mode of reproduction is clonal, sporadic sporulation can help to survive periods of stress [25]. It has been shown that in yeast, sexual reproduction is beneficial when adapting to new, harsh niches, but plays a lesser role in more favorable environments [26, 27]. Our data show that there are large systematic differences in the reproductive lifestyle of yeasts inhabiting different industrial niches: 44.4% of the Beer 1 population is obligate asexual, while this trait ranges between 0% and 21% in the other populations **(Figure 6.8a, p. 96)** and is absent in wild strains. Furthermore, over 80% of the non-mosaic Beer 1 strains that are able to sporulate show little or no spore viability **(Figure 6.8b, p. 96)**. Additionally, beer yeast lineages generally show a high level of heterozygosity, especially Beer 1. Compared to the Wine clade for example, strains from the Beer 1 and Beer 2 clade have on average 5.10-fold (Tukey HSD, $p < 0.001$) and 2.04-fold (Tukey HSD, $p = 0.06$) more heterozygous sites, respectively **(Figures 6.8c, 6.17a, and 6.17b, p. 96, 110, and 110, respectively)**. The lack of genetic admixture suggests that this heterozygosity was acquired during long periods of asexual reproduction, rather than through outbreeding. Further analysis of the correlation between sexual lifestyle and genome structure shows that spore viability is weakly anticorrelated with the heterozygosity level ($R2 \sim 0.17$; $p < 0.001$) and the fraction of the genome associated with large

(>20 kb) amplifications and deletions (R2 $\sim$ 0.16; p $<$ 0.001), while sporulation efficiency is only significantly anticorrelated with the latter (R2 $\sim$ 0.19; p $<$ 0.001) **(Figure 6.9a - 6.9d, p. 97)**.

Together, this indicates that the genome of beer yeasts, but not wine yeasts, show signs of decay and loss of survival skills outside a specific man-made environment, probably caused by their long (estimated >75,000 generations) and uninterrupted growth in rich medium.

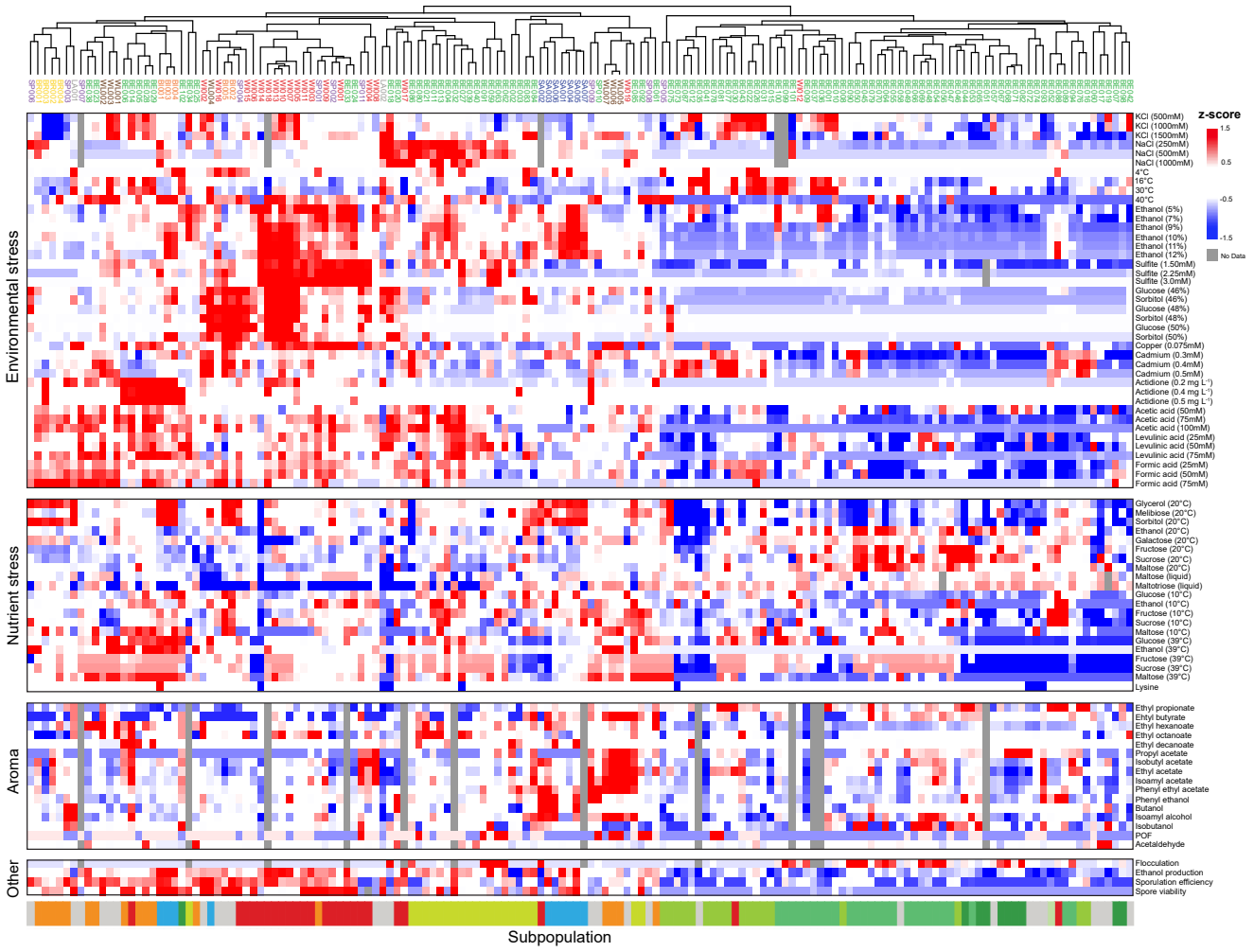


Figure 6.6: Trait variation of industrial *S. cerevisiae* strains. Part 1. Heat map representation of phenotypic diversity within industrial *S. cerevisiae* strains. Phenotypic values are calculated as Z-scores (normalized values) and colored according to the scale on the right. Missing values are represented by gray shadings. Strains are hierarchically clustered based on phenotypic behavior. Strain names are colored according to geographical origin, as in **Figure 6.2, p. 87**. The corresponding subpopulation of each strain is indicated by the colored bar below the figure, according to the color code of **Figure 6.3a, p. 88**.

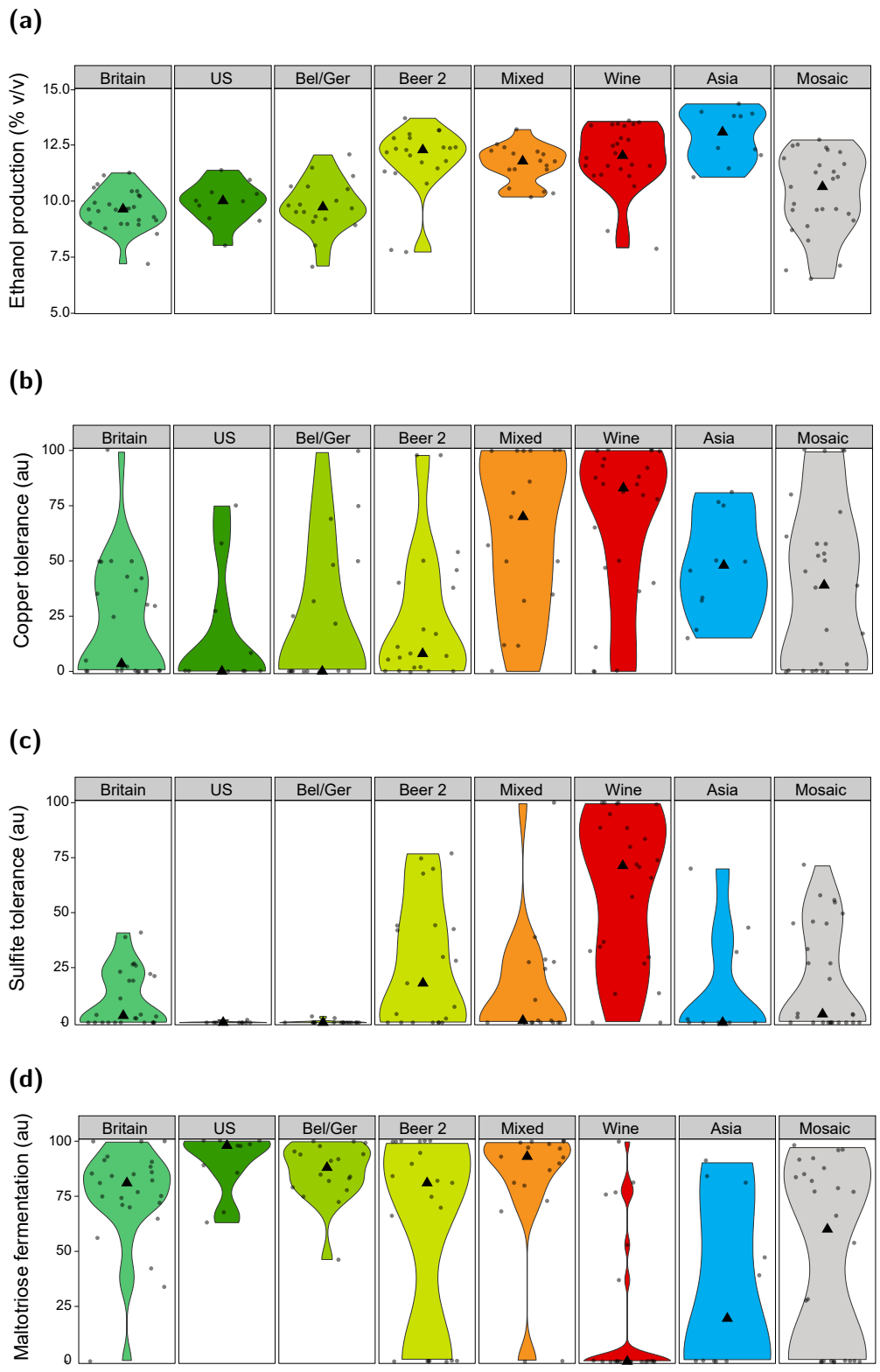


Figure 6.7: Trait variation of industrial *S. cerevisiae* strains. Part 2. a: Ethanol production (depicted as % v/v 1) of all strains from different subpopulations in fermentation medium containing 35% glucose. **b:** Growth of all strains from different subpopulations on medium supplemented with 0.075 mM copper, relative to growth on medium without copper. **c:** Growth of all strains from different subpopulations on medium supplemented with 2.25 mM sulfite, relative to growth on medium without sulfite. **d:** Growth of all strains from different subpopulations in medium containing 1% w v 1 maltotriose as the sole carbon source, relative to growth on medium with 1% w/v 1 glucose. au, arbitrary units; Bel/Ger, Belgium/Germany.

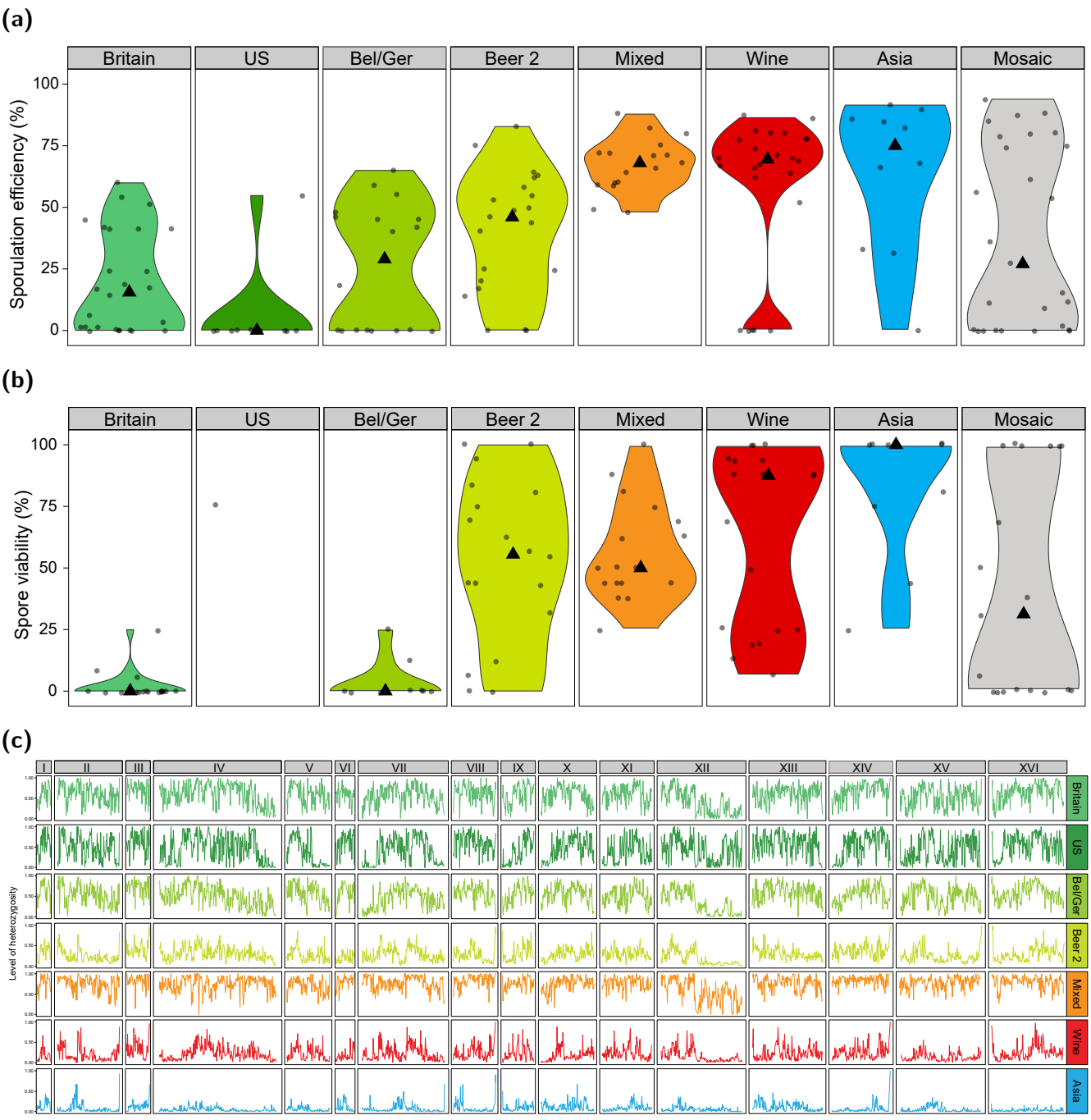


Figure 6.8: The reproductive lifestyle of industrial *S. cerevisiae* Strains. Part 1. **a:** Violin plots depicting sporulation efficiency of all strains from different subpopulations. **b:** Violin plots depicting spore viability of all sporulating strains from different subpopulations. **c:** Visualization of the level of heterozygosity across the genome of the different subpopulations, calculated as the ratio of heterozygous/homozygous SNPs in 10 kb windows.

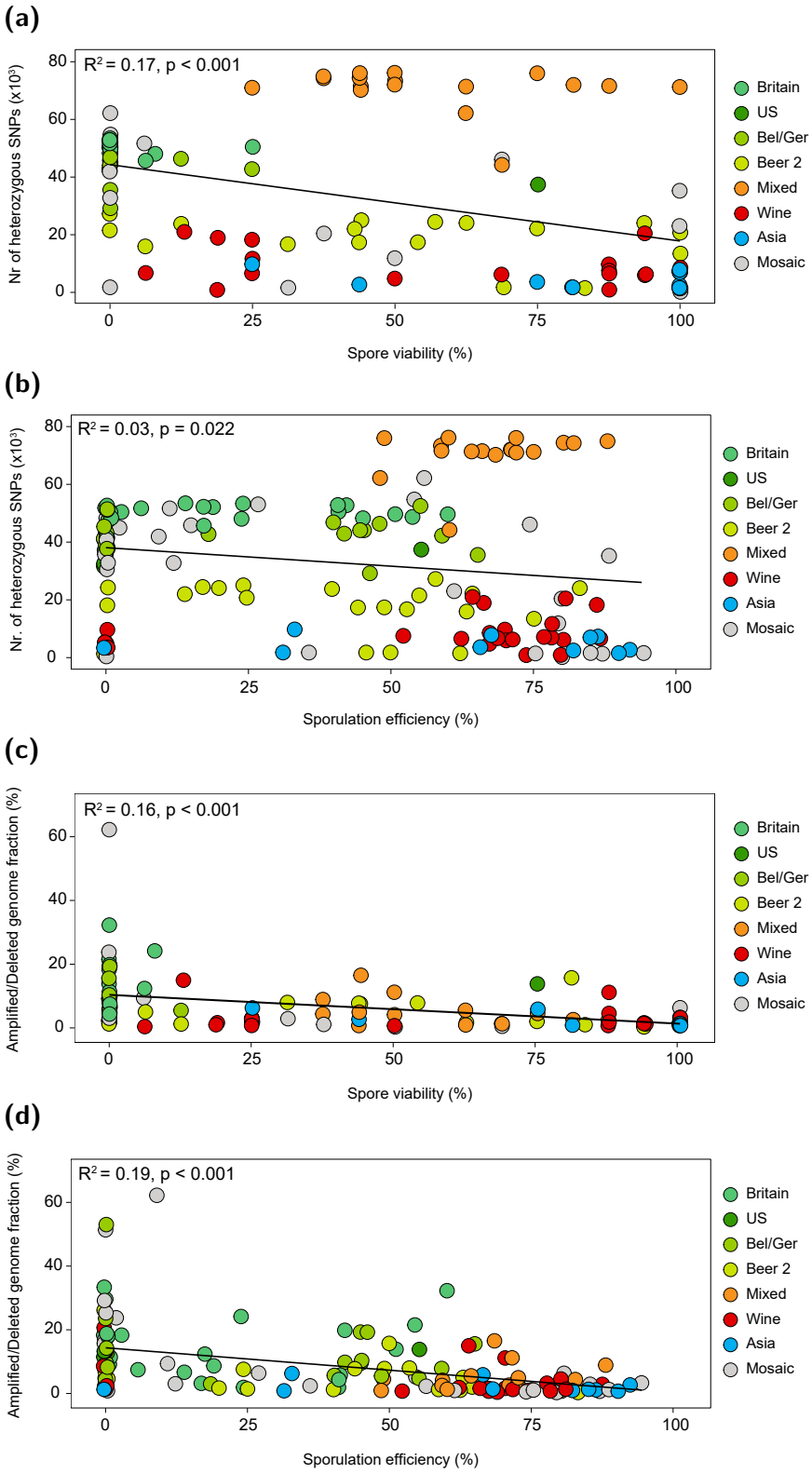


Figure 6.9: The reproductive lifestyle of industrial *S. cerevisiae* strains. Part 2. Scatter plots depicting the correlation between the number of heterozygous loci and spore viability **(a)** or sporulation efficiency **(b)**, and the correlation between the fraction of the genome subjected to large (>20 kb) structural variation and spore viability **(c)** or sporulation efficiency **(d)**. Dot colors indicate subpopulations similar to the color code of **Figure 6.3a, p. 88**.

6.3.4 Selection for Industrial Phenotypes

A key hallmark of domestication is phenotypic adaptation to artificial, man-made niches and accentuation of traits desirable for humans. Phenotypic evaluation of the strains for industrially relevant traits (including aroma production, ethanol production, and fermentation performance) shows that many strains harbor phenotypic signatures linked to their industrial application. The ability to accumulate high concentrations of ethanol, for example, seems tightly linked to industrial niche. Beer 1 strains typically generate only 7.5%-10% v/v of ethanol, while strains used for the production of high-alcohol products like saké, spirits, wine, and especially bioethanol, can produce up to 14.5% v/v **(Figure 6.7a, p. 95; Table 6.9, p. 127)**.

With the exception of a few wine yeast characteristics (see earlier) **(Figures 6.7b and 6.7c, p. 95 and 95, respectively)**, it remains unclear whether genetic and phenotypic variation between *S. cerevisiae* lineages is primarily caused by human-driven selection and domestication, or if neutral genetic drift or non-human selection are involved. To assess this further, we compared the phenotypic behavior of different subpopulations for two industrially relevant traits for which the genetic underpinnings are largely known, namely maltotriose fermentation and the production of 4-vinyl guaiacol (4-VG), the main compound responsible for phenolic off-flavors (POF). Beer yeasts show a significantly higher capacity to metabolize maltotriose, a carbon source specifically found in beer medium **(Figure 6.7d, p. 95; Table 6.9, p. 127)**. Efficient utilization of maltotriose correlates with the presence of a specific allele (*AGT1*) of the sugar transporter *MAL11*, known to show high affinity for maltotriose (phenotypic variability explained by SNPs in *MAL11* $\sim$ 77.40%, SE 0.5%). This allele is only present in Beer 1 subpopulations and some mosaic strains, while the complete *MAL1* locus (including the *MAL11* gene) is absent in the Wine subpopulation **(Table 6.9, p. 127)**. Interestingly, strains of the Beer 2 subpopulation are generally able to ferment maltotriose but contain various frameshift mutations in *MAL11* and show a reduced CNV for the complete *MAL1* locus, suggesting that other, yet unknown mechanisms facilitate maltotriose uptake in this lineage, and maltotriose metabolism evolved convergently in the Beer 1 and Beer 2 lineages.

Yeasts used for the production of alcoholic beverages ideally should not produce undesirable aromas. Although tolerated in some specialty beers, the presence of 4-VG, a compound with a spicy, clove-like aroma, is generally undesired in saké , wine, and most beer styles. Two genes, phenylacrylic acid decarboxylase (*PAD1*) and ferulic acid decarboxylase (*FDC1*), both located as a cluster in the subtelomeric region of ChrIV, control 4-VG production. *PAD1* encodes a flavin prenyltransferase that catalyzes the formation of a flavin-derived cofactor, which is required by Fdc1 for decarboxylation of the precursor ferulic acid [28]. Pad1 and Fdc1 help to detoxify phenylacrylic acids found in plant cell walls [29]. Therefore, it would be expected that, unless there is counterselection, activity of these genes is retained. Interestingly, phenotypic profiling reveals that many industrial strains have lost the ability to produce 4-VG, while it is generally

retained in wild strains, as well as in bakery and bioethanol strains **(Figure 6.11a, p. 101)**. In these cases, 4-VG production is likely less detrimental, either because the flavor disappears during baking, or the product is not destined for consumption. Sequence analysis shows that many industrial strains, especially beer and saké strains, acquired loss-of-function mutations (SNPs and/or frameshift InDels) in *PAD1* and/or *FDC1*, while this was never observed in strains from natural environments or bioethanol production **(Figure 6.10, p. 100)**. Moreover, different sublineages acquired different disruptive mutations, hinting to the presence of diverse convergent adaptive strategies in response to human selection against 4-VG production. To investigate the origin and the maintenance of the phenotypic diversity in 4-VG production, we used Bayesian inference to reconstruct the ancestral phenotypic state in the two key genes *PAD1* and *FDC1* using BEAST [30] **(Figure 6.11b, p. 101)**. Shifts from 4-VG+ to 4-VG and vice versa occurred frequently after the initial split from *S. paradoxus*. In both the *PAD1* and the *FDC1* trees, an early subclade containing most Beer 1 strains acquired loss-of-function mutations at the base of the clade, suggesting that already very early during domestication of the Beer 1 lineage, a 4-VG variant was derived from the 4-VG+ ancestor. Several other loss and gain-of-function mutations occurred across both trees, most notably the loss-of-function mutation in *FDC1* of the Asian saké (but not bioethanol) strains. Interestingly, a strong incongruence between single gene trees and the strain phylogeny is present for three beer strains used in the production of German Hefeweizen beers (BE072, BE074, and BE093). Hefeweizen (wheat) beer is a traditional German beer style and one of the few styles where a high 4-VG level is desirable because it contributes to the typical smoky, spicy aroma of these beers. Phylogenetically, Hefeweizen yeasts cluster within the Beer 1 lineage, but they are shown to be highly mosaic, containing genomic fragments of all three Beer 1 subclades (mainly from Belgium/Germany). Only a small fraction (8%-13%) of the genome originates from the Wine subpopulation, but this fraction includes the subtelomeric region of ChrIV, containing a functional *PAD1* and *FDC1* allele. This suggests that hybridization between different domesticated subpopulations yielded variants combining the typical traits of beer yeasts, including maltotriose fermentation, with a particular trait from a wine strain (4-VG production) that is only desirable in special beer styles.

Figure 6.10: Production of 4-Vinyl Guaiacol by industrial *S. cerevisiae* strains. Part 1. Distribution of loss-of-function SNPs and frame-shift mutations in *FDC1* and *PAD1* of industrial *S. cerevisiae* strains. Gray boxes indicate the presence of the loss-of-function mutation, diagonal bars indicate heterozygosity at this site. Strains are clustered according to the strain phylogeny and strain names are colored according to their origin. Basal splits of the five industrial lineages are indicated with colored dots **(see Figure 6.2, p. 87)**.

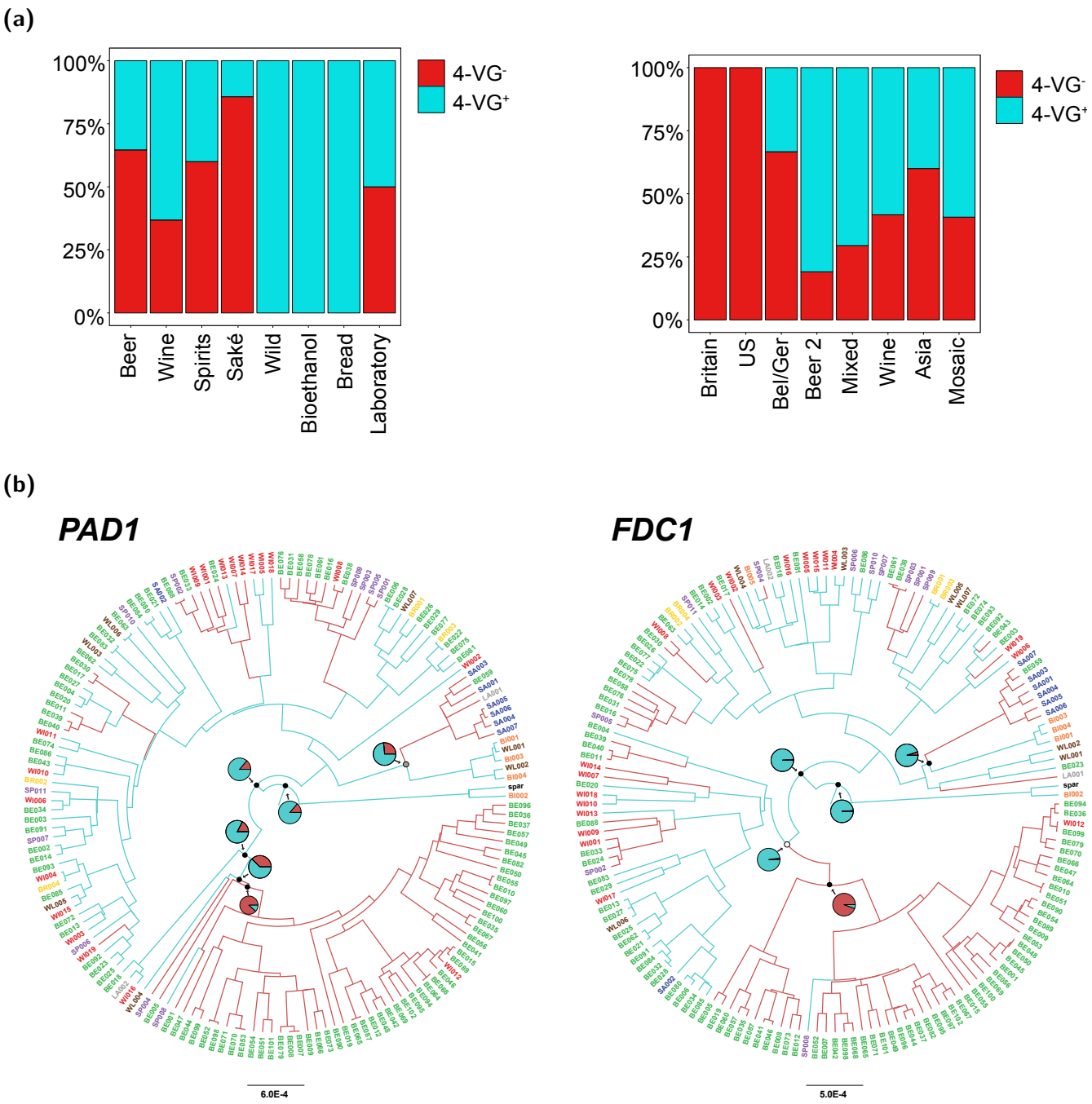


Figure 6.11: Production of 4-Vinyl guaiacol by industrial *S. cerevisiae* strains. Part 2. a: Percentage of strains within each origin (left) and population (right) capable of producing 4-vinyl guaiacol (4-VG). Red, 4-VG ; turquoise, 4-VG+. **b:** Phylogenetic trees and ancestral trait reconstruction of *PAD1* and *FDC1* genes. Branches are colored according to the most probable state of their ancestral nodes, turquoise (4-VG+) or red (4-VG). Pie charts indicate probabilities of each state at specific nodes, turquoise (4-VG+) or red (4-VG); posterior probability for the same nodes is indicated by a dot: black dot, 90% - 100%; gray, 70%; white, 42%. Branch lengths reflect the average numbers of substitutions per site (compare scale bars).

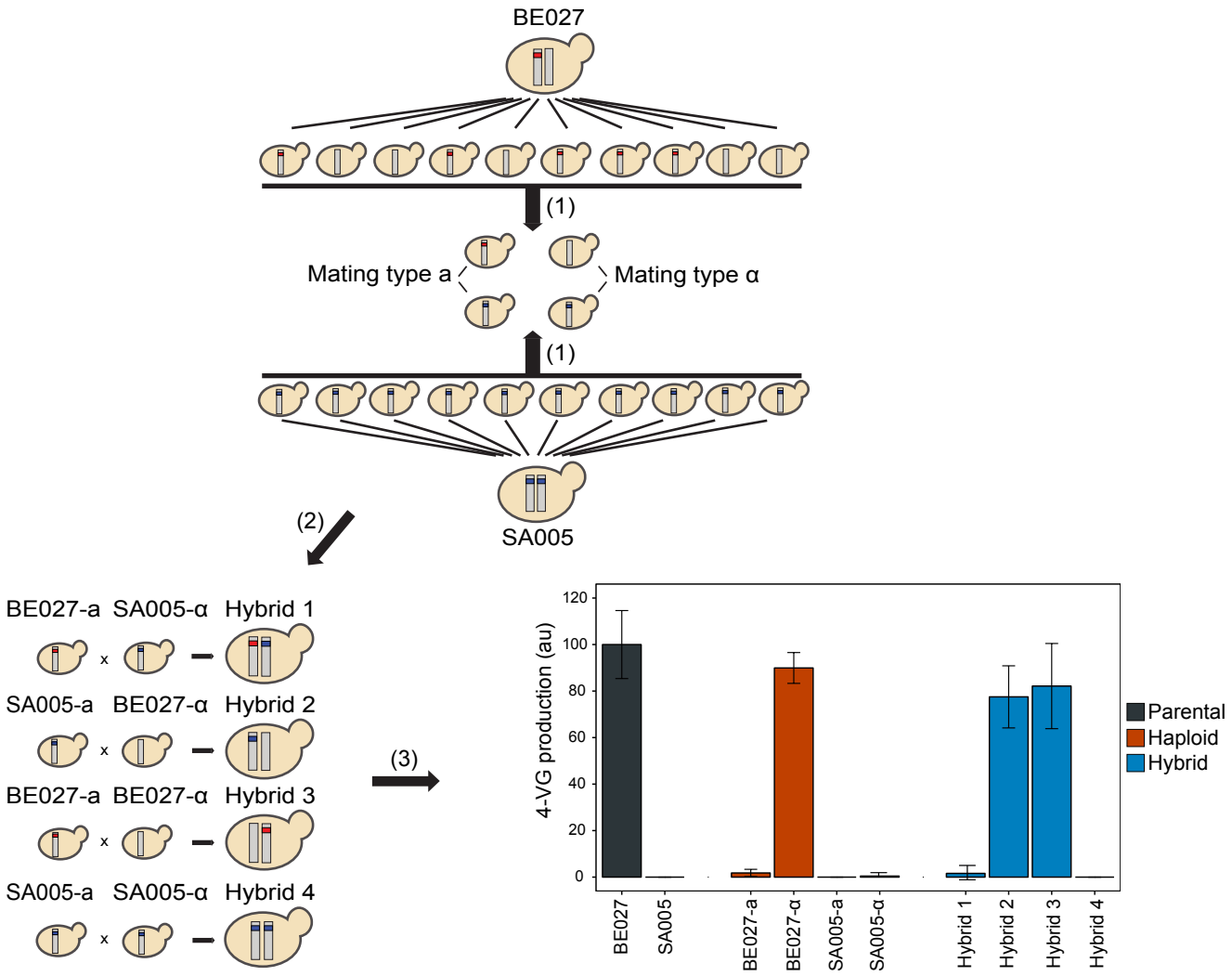


Figure 6.12: Production of 4-Vinyl Guaiacol by Industrial *S. cerevisiae* Strains. Part 3. Development of new yeast variants with specific phenotypic features by marker-assisted breeding. Two parent strains (BE027 and SA005) were sporulated and, using genetic markers, segregants with the desired genotype were selected (1). Next, breeding between segregants from different parents (outbreeding) or the same parent (inbreeding) were performed (2). This breeding scheme yields hybrids with altered aromatic properties that can directly be applied in industrial fermentations (3). 4-VG production is shown relative to the production of BE027. Yeast genomes are represented by gray bars, loss-of-function mutations in *FDC1* as red (W497*) and blue (K54*) boxes within the gray bars. Error bars represent one SD from the mean.

6.3.5 Creating superior hybrid yeasts through marker-assisted breeding

Apart from yielding insight into the origins of today's industrial yeasts, our results also open new routes for the creation of new superior strains. The availability of genomic data and the increasing number of polymorphisms that are known to contribute to industrially relevant phenotypes enables rapid DNA-based selection of superior segregants and hybrids in large-scale breeding schemes. Such marker-assisted breeding is already intensively used for crop and livestock breeding, because

it circumvents labor-intensive and time-consuming phenotyping. As proof-of-concept, we combined our genomic and phenotypic data to obtain new hybrids with altered aromatic properties using marker-assisted breeding. Specifically, a 4-VG producing beer strain harboring a heterozygous loss-of-function mutation in *FDC1* (strain BE027) was selected and sporulated to obtain segregants. Next, the *FDC1* allele of the segregants was genotyped using mismatch PCR. Two segregants, one harboring the loss-of-function allele and one harboring the functional allele, were crossed with segregants of SA005, an Asian saké strain with a homozygous nonfunctional *FDC1* allele, resulting in hybrids with good beer fermentation characteristics but drastically different aroma profiles (4-VG+ versus 4-VG) that suit specific beer styles **(Figure 6.12, p. 102)**.

6.3.6 Domestication Predates Microbe Discovery

Despite its wide use in industry and as a model organism, little is known about the ecology and evolutionary history of *S. cerevisiae*. Moreover, because early brewers, winemakers, and bakers were unaware of the existence of yeast, there is no record of how yeasts made their way into these processes, nor how yeasts were propagated and shared. As a result, it has proven difficult to estimate when specific industrial lineages originated. Moreover, current demographic and molecular clock models of *S. cerevisiae* employ the experimentally determined mutation rate of the haploid lab strain S288c in rich growth medium [31], while it is known that the mutation rate is heavily influenced by the genetic background [32], ploidy [23], growth speed [33], and environmental stress [22], factors that are likely very different for industrial, wild, and lab yeasts. However, our dataset, and specifically the Beer 1 clade, provides a strong tool for dating beer yeast divergence. First, given the absence of a functional sexual cycle and lack of admixture, exclusively clonal reproduction can be assumed. Second, our data show that United States beer yeasts are related closest to European beer yeasts, suggesting that they were imported from Europe during colonization, rather than stemming from indigenous wild United States yeasts. More specifically, United States beer yeasts seem phylogenetically most closely related to British beer yeasts **(Figure 6.2, p. 87)**, which is to wine, beer is not produced seasonally but throughout the whole year, which provides fermenting yeast with a predictable and stable growth environment. Yeast cells undergo about three doublings during one batch of beer fermentation, which takes $\sim$ 1 week. Moreover, brewers typically recycle yeasts from a finished fermentation to inoculate a new batch, which implies that beer yeasts are continuously growing in their industrial niche. Together, these facts make it possible to estimate the number of generations to be $\sim$ 150/year.

Based on estimates of the number of generations per year and the divergence time between United Kingdom and United States beer strains, we calculated the average mutation rate in a brewing environment to be 1.61-1.73E-08/bp/generation. While this value differs from previous assumptions, it is similar to the measured mutation rate in a diploid yeast strain that was subjected to 2 years of artificial evolution in a high-ethanol environment [22]. Moreover, mutations likely

also occur in the second phase of beer fermentations, when cells are no longer dividing, which implies that the mutation rate per generation in industrial conditions should be higher than what is measured under conditions where the cells are dividing frequently, as is usually the case in laboratory experiments [34]. Using these data, the last common ancestor of the three major Beer 1 subclades (Belgium/Germany, United Kingdom, and the United States) is estimated to date from AD 1573-1604, suggesting that domestication started around this time. Interestingly, this coincides with the gradual switch from home-centered beer brewing where every family produced their own beer, to more professional large-scale brewing, first in pubs and monasteries and later also in breweries [35]. The last common ancestor of Beer 2 is estimated to be more recent, between AD 1645-1671. This suggests that beer yeast domestication started before the discovery of microbes and the isolation of the first pure yeast cultures by Emil Hansen in the Carlsberg brewery in 1883, but well after the invention of beer production, estimated to have occurred as early as 3000 BC [1]. Although it is difficult to assess how many different yeast strains were domesticated and in which industrial context these domestications occurred, the limited number of clades of industrial yeasts and the clear segregation of wild and industrial yeasts suggests that today's industrial yeasts originated from a limited set of ancestral strains, or closely related groups of ancestral strains.

6.4 Discussion

Together, our results show that today's industrial *S. cerevisiae* yeasts are genetically and phenotypically separated from wild stocks due to human selection and trafficking. Specifically, the thousands of industrial yeasts that are available today seem to stem from only a few ancestral strains that made their way into food fermentations and subsequently evolved into separate lineages, each used for specific industrial applications. Within each cluster, strains are sometimes further subdivided along geographical boundaries, as is the case for the Beer 1 clade, which is divided into three main subgroups. However, further subclustering of beer yeasts according to beer style was generally not observed, which may not be surprising as it is common practice for brewers to use only one yeast strain within their brewery for the production of a wide array of different beers. Notable exceptions are yeasts associated with the few beers that largely depend on very specific yeasts characteristics, such as Hefeweizen beers. Another exception may include those beers for which production is restricted to a specific geographic area, such as Belgian Saisons or British Stouts.

We further show that industrial yeasts were clearly subjected to domestication, which is reflected in their genomes and phenomes. Interestingly, domestication seems strongest in beer yeasts, which demonstrate domestication hallmarks such as decay of sexual reproduction and general stress resistance, as well as convergent evolution of desirable traits like maltotriose utilization. Yeasts from the Beer 1 clade show the clearest signs of domestication, possibly because Beer 2 only diverged more recently from other sublineages. Many of these domestication features may have

simply been the result of the yeasts' adaptation to their new industrial niches. However, for some traits, it is likely that humans actively intervened, e.g., by selecting strains that do not produce undesirable off-flavors, which our analysis identifies as *PAD1* or *FDC1* nonsense mutants.

The presence of a strong domestication signature in beer yeast genomes agrees well with the common practices in the brewing industry. Beer yeasts are typically recycled after each fermentation batch, and because beer is produced throughout the year, this implies that beer yeasts are continuously growing in their industrial niche. By contrast, wine yeasts can only grow in wine must for a short period every year, spending the rest of their lives in and around the vineyards or in the guts of insects [36–38]. During these nutrient-poor periods, wine yeasts likely undergo few mitotic doublings, yet they may undergo sexual cycles and even hybridize with wild yeasts [38]. Moreover, only a very small portion of the yeasts may find their way back into the grape must when the next harvest season arrives, while trillions of cells are being transferred to the next batch during backslopping in beer production. This results in large effective population sizes for beer, but not wine, yeasts. The differences in industrial practices between beer brewing and winemaking likely had three important consequences. First, beer yeasts evolved faster than wine strains **(Figure 6.2, p. 87)**. This resulted in a large genetic diversity within beer yeasts, while wine yeasts are genetically more homogeneous **(Table 6.1, p.90)**. Second, after the initial domestication event, some beer yeasts were contained in the brewery and diverged allopatrically, leading to geographically defined subpopulations mirroring human traffic and colonization. Third, beer strains generally lost their ability to reproduce sexually. This, combined with continuous cultivation in a mild growth environment, made them susceptible to genetic drift and fixation of deleterious alleles that would otherwise be purged by evolutionary competition in harsh conditions. Hence, these asexual populations continuously accumulated deleterious mutations in an irreversible manner, a process known as Muller's ratchet [39]. We propose that continuous clonal reproduction and relaxed selection for general stress resistance and famine likely allowed genome decay in beer yeasts and resulted in yeasts specialized in thriving in a man-made niche like beer fermentations, but not in natural environments. Both these characteristics (genome decay and niche specialization) are considered to be key characteristics of domestication.

Our study does not only provide insight into the domestication origin of industrial yeasts, it may also help to select and breed new superior strains. The genome sequences, phylogenetic tree, and phenome data can be used to set up marker-assisted breeding schemes similar to those routinely used for the breeding of superior crops and livestock [40].

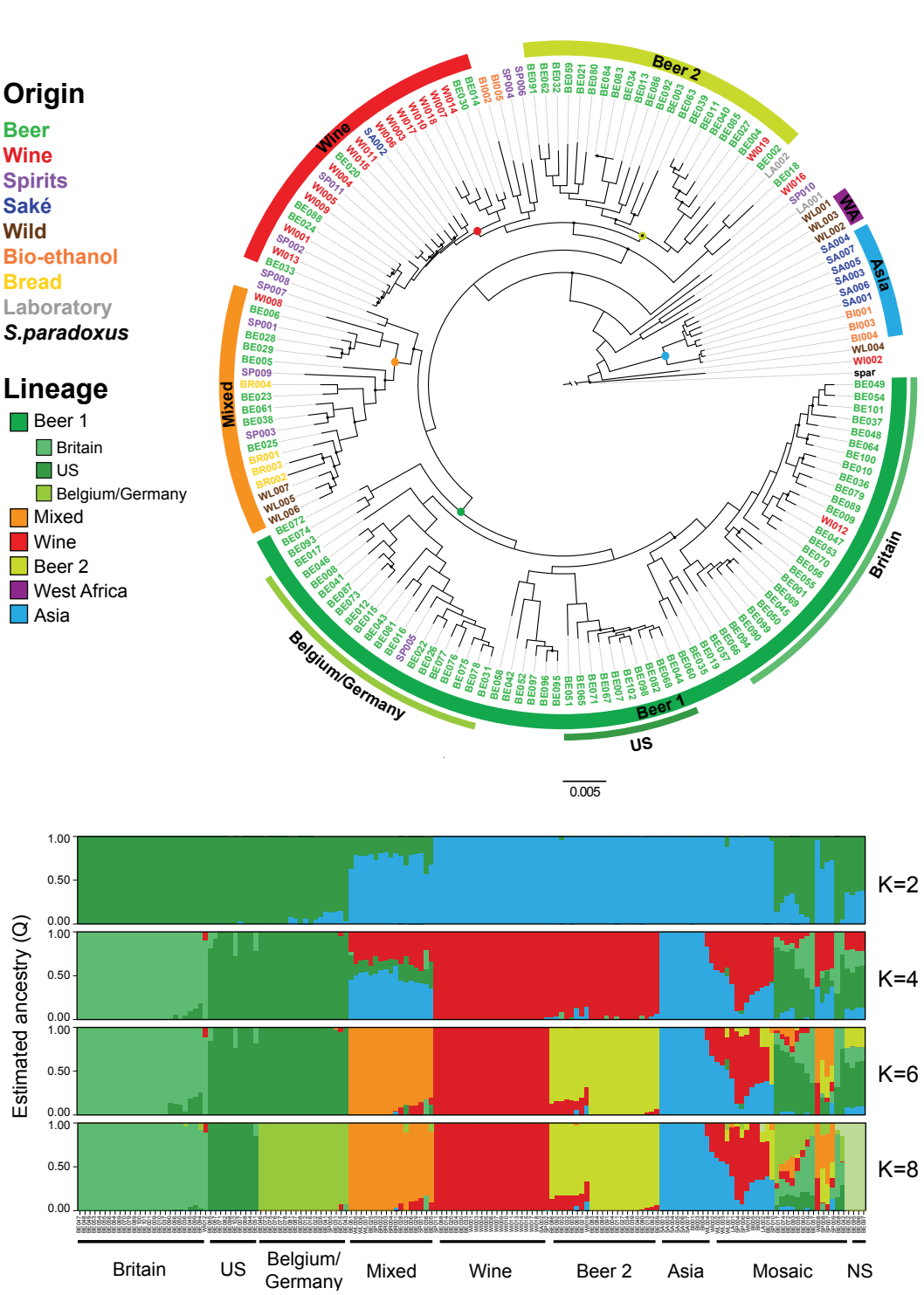


Figure 6.13: Phylogeny and population structure of the industrial *S. cerevisiae* strains, Related to Figures 6.2-6.3. Part 1. a: Phylogenetic tree of strains sequenced in this study, using *Saccharomyces paradoxus* as an outgroup. The tree was inferred from the concatenation matrix of 2,020 single copy orthologs. Black dots on nodes indicate bootstrap support values $< 70\%$. Color codes indicate origin (names) and lineage (circular bands). The basal splits of the five industrial lineages are indicated with a colored dot. Branch lengths reflect the average number of substitutions per site (scale bar = 0.005 substitutions per site). **b:** Population structure plot, with strain codes indicated.

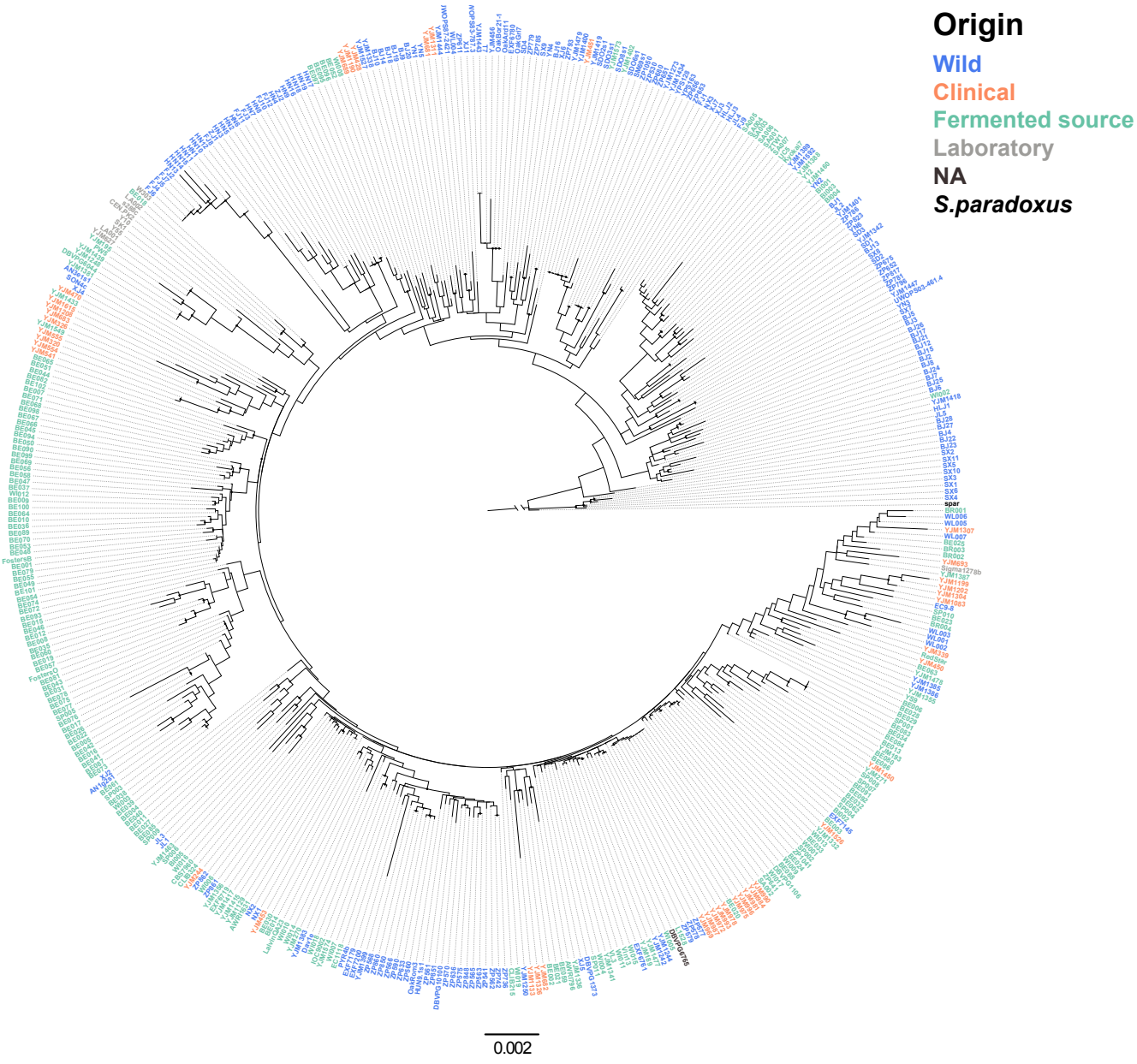


Figure 6.14: Phylogeny and population structure industrial *S. cerevisiae* strains, related to Figures 6.2-6.3. Part 2. Maximum likelihood phylogenetic tree inferred from a concatenated alignment of nine partial genes of 450 *S. cerevisiae* isolates, using *S. paradoxus* as an outgroup. Strains are colored according to origin. For a list of the included strains and corresponding references, see Table 6.4, p. 127. Branch lengths reflect the average number of substitutions per site (scale bar = 0.002 substitutions per site). Dots indicate nodes with bootstrap support values > 50%. Font color codes indicate origin: wild (blue), clinical (orange), fermented source (green), laboratory (gray), not available (NA) (dark gray), *S. paradoxus* (black).

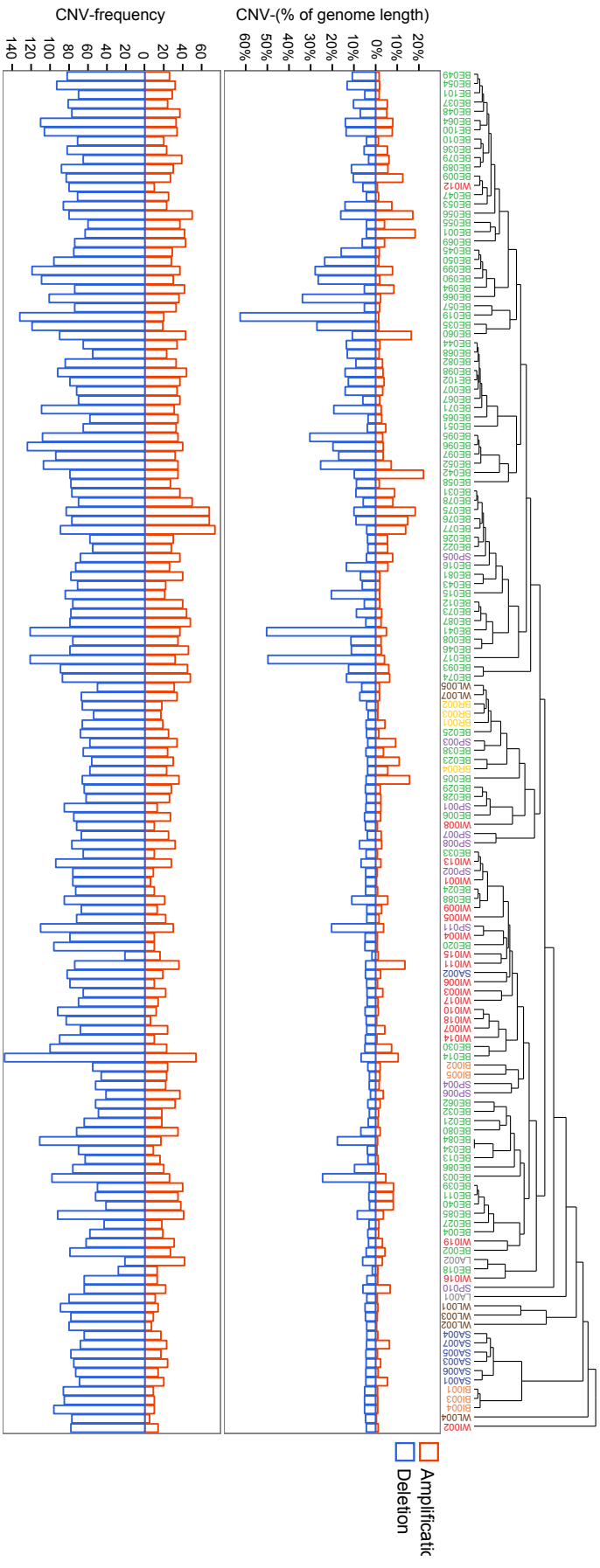


Figure 6.15: Copy number variability across the phylogenetic tree, related to Figures 6.4-6.5. Distribution of amplifications (red) and deletions (blue) in each strain, expressed in percentage of the genome affected (top) and in number of CNV events (bottom). The phylogenetic tree is described in **Figure 6.13a**.

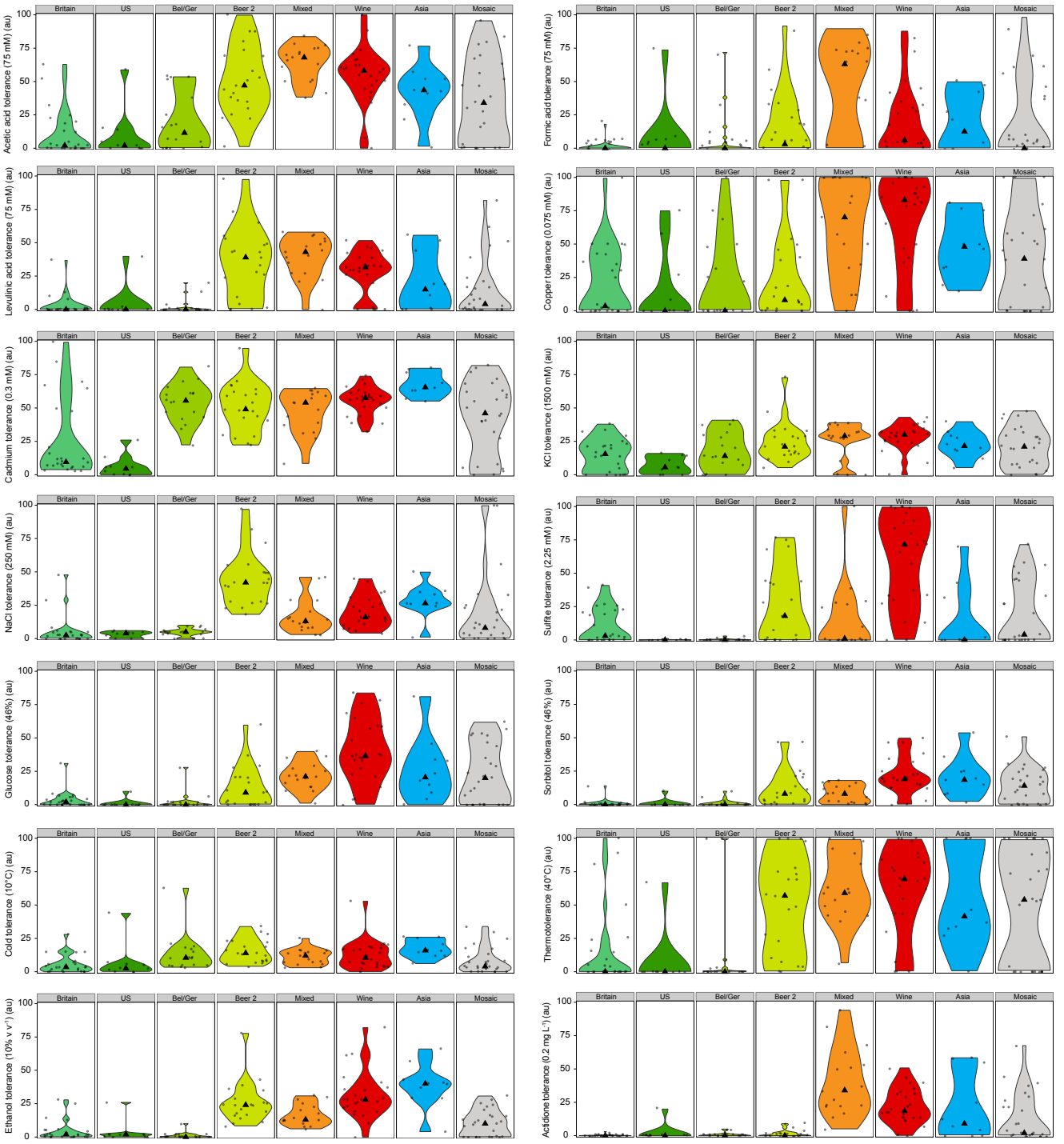


Figure 6.16: Trait variation in industrial *S. cerevisiae* strains, selated to Figures 6.6-6.7. Graphic representation (violin plots) of trait variation within and between the subpopulations for different environmental stressors. Triangles represent median values for each subpopulation. All values are depicted as relative growth compared to growth on medium without the stressor. Statistical analysis for each trait is given in Table 6.9, p. 127. au = arbitrary units.

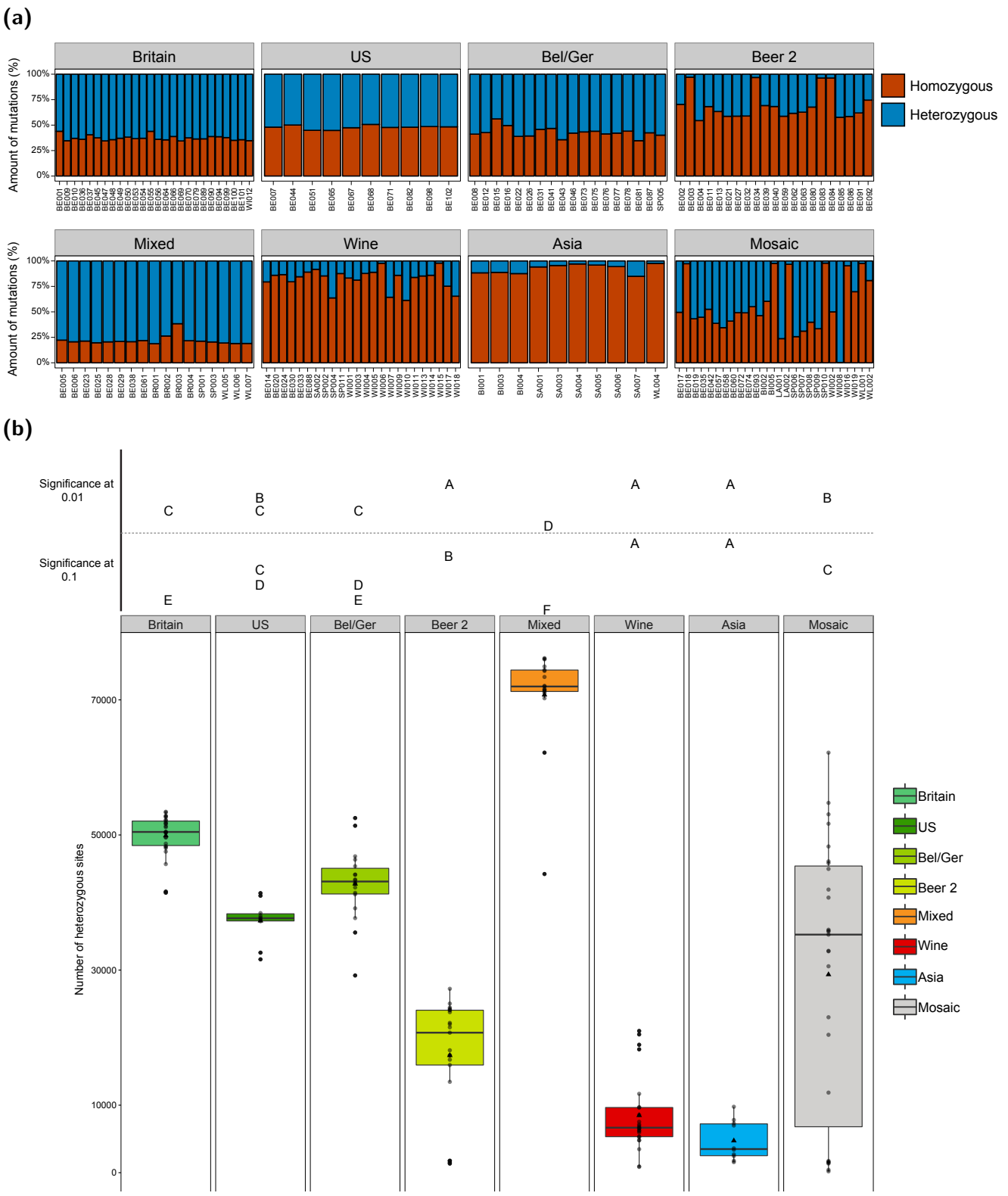


Figure 6.17: Heterozygosity of industrial yeasts, related to Figures 6.8-6.9. **a:** Percentage of total SNPs identified as heterozygous or homozygous in each strain. Boxes depict subpopulations and bar colors indicate the percentage of homozygous (red) and heterozygous (blue) SNPs. **b:** Box plots depicting the total number of heterozygous sites per subpopulation. The mean number of heterozygous sites for each comparison group is indicated by a triangle and the median by a horizontal line. Groups sharing the same letter (top) are not significantly different at the 10% or 1% level.

6.5 Methods

6.5.1 Strain Collection

For this study, a set of 157 *Saccharomyces cerevisiae* strains was sequenced, phenotyped and analyzed. Strains were obtained from historical yeast collections of the VIB Laboratory for Systems Biology (KU Leuven, Belgium) and White Labs (USA). While detailed background information on many of these strains is limited, their geographical origin was in most cases documented and is listed in Table 6.3, p. 127. Beer strains mainly originate from the main fermentation (82/102) or bottle conditioning (10/102) of ale beers. While lager beer fermentations are usually carried out by *S. pastorianus*, an interspecific hybrid of *S. cerevisiae* and *S. eubayanus*, we identified 10 *S. cerevisiae* strains typically employed in lager fermentations and included them in the selection. All strains are long-term stored in 80 °C using a glycerol-based standard storage medium (peptone 1% w v^{-1}, yeast extract 0.5% w v^{-1}, glucose 1% w v^{-1}, glycerol 25% v v^{-1}).

6.5.2 DNA Extraction

For strains BE001-043, BI001-005, BR001-004, LA001, SA001-007, SP001-007, NA001-004 and WI001-018, genomic DNA was prepared using the QIAGEN genomic tip kit (QIAGEN, Germany) according to recommended protocols. Final DNA concentrations were measured using Qubit (Thermo Fisher Scientific, USA). For the other strains, genomic DNA was extracted with the MasterPureTM Yeast DNA Purification Kit (Epicenter, USA), but with some modifications to the recommended protocols. Three mL of each of the liquid yeast cultures were pelleted by centrifugation at 20,800 x g for 5min in 1.7mL micro centrifuge tubes. 300 ml of yeast cell lysis solution was added to each micro centrifuge tube along with 1 ml of 5g L^{-1} RNase A. The cells were resuspended by vortex mixing each micro centrifuge tube for 10 s. Each tube was incubated at 65 °C for 15 min and was then chilled on ice for 5 min. Next, 150 ml of protein precipitation reagent was added to each tube and the tubes were vortexed for 10 s. The suspensions were then centrifuged for 10 min at 20,800 x g to pellet the cellular debris. The supernatants were transferred to clean 1.7 mL micro centrifuge tubes. Next, 500 ml of isopropanol was added to each tube. Each tube was then inverted several times to precipitate the DNA, which was then pelleted by centrifugation at 20,800 x g for 10 min. The supernatants were discarded and the pellets were washed with 500 ml of ice-cold 70% ethanol and briefly centrifuged. The ethanol was removed by pipetting. The DNA pellets were centrifuged again for 10 min at 20,800 x g to remove any remaining ethanol. Each DNA pellet was then suspended in 35 ml of TE buffer and stored at 4 °C. After isolation, the purified DNA was quantified using fluorimetric methods and diluted to the optimal concentration for library construction.

6.5.3 Library Prep and Whole Genome Sequencing

For strains BE001-043, BI001-005, BR001-004, LA001, SA001-007, SP001-007, NA001-004 and WI001-018, paired-end sequencing libraries (100bp) with a mean insert size of 300bp were prepared and run according to the manufacturer's instructions on an Illumina HiSeq2000 at the EMBL GeneCore facility, Heidelberg (http://genecore3.genecore.embl.de/genecore3/). For the other strains, libraries were prepared using the Nextera XT sample preparation kit. A total of 50 ng of yeast DNA was fragmented and tagged with DNA adapters by the Nextera transposome resulting in adaptor-ligated DNA fragments. The DNA was purified and PCR-amplified to add the dual indexes as well as the common adapters required for cluster generation and sequencing. All samples were pooled together and clustered on board a HiSeq 2500 instrument at Illumina (San Diego, USA). Samples were sequenced in both Rapid Run Mode and High Output Mode using 2 x 100 bp paired-end reads.

6.5.4 *De novo* Assembly

For each library, low-quality and ambiguous reads were trimmed using Trimmomatic (v0.30) [41]. After k-mer based read correction with musket [42], reads were assembled using idba_ud [43]. De novo assemblies were evaluated by mapping back the reads and also by checking BLAST matches of assembled contigs to the SILVA database for rDNA classification. Each de novo assembly was scaffolded against the *S. cerevisiae* S288c reference genome assembly (http:// downloads.yeastgenome.org/sequence/S288C_reference/genome_releases/S288C_reference_genome_-R64-1-1_20110203.tgz). The liftOver workflow [44] was used to determine the coordinates of contigs from each newly assembled strain relative to the reference strain (http://genomewiki.ucsc.edu/-index.php/Minimal_Steps_For_LiftOver). Scaffolded contigs mapped to each reference strain chromosome were combined into a "pseudo-molecule," with the placed contigs stitched together with gaps indicated by "N." Unplaced contigs (including alternative lower scoring matches) were kept. Unplaced contigs less than 300 nucleotides were not included in the final assembly **(Table 6.3, p. 127)**.

6.5.5 Annotation

The genome annotation of the *S. cerevisiae* S288c reference genome (nr. of genes = 6,692) was downloaded from the UCSC (version Apr2011/sacCer3) Table Browser in GPR format. FASTA records were renamed to match the chromosome naming convention in the GPR file. The liftOver workflow was used to create a coordinate conversion file (chain file) between the *S. cerevisiae* S288c genome and each newly scaffolded assembly. Using the chain file, the coordinates of the *S. cerevisiae* S288c genes were "lifted" to each new genome assembly. Lifted genes were considered valid if they did not contain internal stop codons. Independently, the gene prediction tool AUGUSTUS v2.5 [45] was used to predict genes for each new strain using the provided training set/model for

S. cerevisiae S288c with the following parameters (-noInFrameStop = true, maxDNAPieceSize = 1000000, progress = false, uniqueGeneId = true, keep_viterbi = false). The annotated and predicted genes, using liftOver and AUGUSTUS respectively, were combined with priority given to the liftOver annotation when the predictions overlapped **(Table 6.3, p. 127)**.

6.5.6 Core Genome Analysis and Identification of Single Copy Genes

Across the 157 annotated *S. cerevisiae* genomes, 986,179 genes were annotated and predicted in total, with an average of 6,281 genes per genome (min = 6,099 genes, max = 6,655 genes). CD-HIT (v4.6) was used to approximate a non-redundant set of putative translations across the 157 genomes (parameters, -c 0.7 -M 3200 -T 0 -d 60) [46]. Using a 70% amino acid identity threshold, the collection of 986,179 translated genes was reduced to 8,410 clusters. A total of 3,519 clusters contained exactly one gene from each of the 157 assembled genomes. The 3,519 clusters represent an approximation of the *S. cerevisiae* core genome across the 157 genomes evaluated. A conservative set of single copy genes was identified across the 157 genome assemblies, the *Saccharomyces paradoxus* genome and an additional set of 24 *S. cerevisiae* strains assembled in recent studies **(Table 6.3, p. 127)** [10, 13, 16]. First, one-to-one ortholog pairs were extracted from previously identified orthologs between *S. paradoxus* (NRRL-Y17217) and *S. cerevisiae* S288c (https://portals.broadinstitute.org/regev/orthogroups/orthologs/Scer-Spar-orthologs.txt) [47]. A total of 5,096 one-to-one ortholog pairs were identified, with a subset of 5,084 annotations mapped to the *S. cerevisiae* S288c reference gene set used. The set of 5,084 genes was filtered to a smaller subset of 2,417 genes based on i) inclusion in the set of 3,333 *S. cerevisiae* S288c ORFS that could be mapped by liftOver across all 157 genomes, and ii) inclusion in the set of 3,519 clusters uniquely represented in each of the 157 strains based on CD-HIT results. Lastly, the presence and the single-copy status of the 2,417 genes were investigated in the additional 24 previously sequenced *S. cerevisiae* strains (http://www.moseslab.csb.utoronto.ca/sgrp/download.html and http://www.ncbi.nlm.nih.gov/genbank with accession numbers as reported in [13] last access June 2015), further reducing the selection to a conservative set of 2,026 genes. The final set included 2,020 single-copy genes after removing six highly fragmented sequences.

6.5.7 Reference-Based Alignments and Variant Calling

To identify single nucleotide polymorphisms (SNPs) and short insertions and deletions (InDels), reads were pre-processed by filtering low quality and ambiguous reads, adapters and PhiX contaminations, using Trimmomatic (v0.30) [41]. Clean reads were mapped to the *S. cerevisiae* reference genome S288c (R64-1-1, EF4-Ensemble Release 74) with the Burrows-Wheeler Aligner (BWA, v0.6.1) using default parameters except for -q 10 [48]. Non-primary alignments and non-properly paired reads were filtered out and duplicate reads were marked using Picard Tools (v1.56) (http://picard.sourceforge.net). Before variant calling, reads were locally realigned in order to

eliminate false positives due to misalignment of reads, which was followed by a base quality score recalibration step, using the Broad Institute Genome Analysis Toolkit (GATK v2.7.2) [49]. SNP and InDel discovery and genotyping was performed across all 157 strains simultaneously, to minimize false positive calls, with a minimum base quality score of 20, a standard minimum confidence threshold for calling of 50 and a standard emitted confidence of 20. Sites with total quality by depth < 2.00 and mapping quality < 40, genotype quality < 30 and genotype depth < 5 were filtered out using GATK Variant Filtration. For SNP calling, sites overlapping InDels, sites with more than 50% missing genotypes and multiallelic sites were filtered out using VcfTools (v0.1.10;v0.1.14) [50]. The final set of SNPs included a total of 421,361 biallelic segregating sites accounting for a total of 10,576,934 SNPs across all strains. SnpEff (v3.3) [51] was used to annotate and predict the effect of the variants.

6.5.8 Phylogenetic Analyses

Phylogenetic Tree for the Sequenced Collection, Figure 6.13a, p. 106

Multiple sequence alignments (MSAs) for the 2,020 amino acid sequences identified above were generated using MAFFT (v7.187), with default settings and 1,000 refinement iterations [52]. Codon alignments were obtained from MSAs of predicted amino acid sequences and the corresponding DNA sequences by the PAL2NAL program (v14) [53]. Quality checks and format conversions were performed using trimAl (v1.2) [54]. The full set of codon alignments were concatenated into a supermatrix using FASconCAT (v1.0) [55]. The resulting supermatrix included 158 taxa and 2,782,494 positions, 99.174% nucleotides, 0.826% gaps and 0% ambiguities. The matrix was partitioned based on all 2,020 gene blocks and all three codon positions within each block, resulting in 6,060 distinct data partitions accounting for 144,171 distinct alignment patterns. Twenty completely random starting trees and 20 randomized stepwise-addition parsimony starting trees were obtained using RAxML (v8.1.3) [56]. Robinson-Foulds (RF) distances were computed between all trees in both the fully random and parsimony tree sets, to avoid systematic bias due to low diversity in starting trees. Because the stepwise addition algorithm generated a set of starting trees with low diversity, all the subsequent analyses were conducted with fully random starting trees. Twenty maximum-likelihood (ML) tree searches were performed on each of the 20 fully random starting trees under the GTRGAMMA model (4 discrete rate categories) using ExaML (v3) and the rapid hill climbing algorithm (-f d) [57]. During the ML search, the alpha parameter of the model of rate heterogeneity and the rates of the GTR model of nucleotide substitutions were optimized independently for each partition. The branch lengths were optimized jointly across all partitions. For each starting tree, the best tree was selected based on the highest log-likelihood score. Parameters and branch lengths were reoptimized on the best 20 topologies with ExaML (-f E) using the median of the four categories for the discrete approximation of the GAMMA model of rate heterogeneity (-a). The tree with the best overall log-likelihood score of all 20 tree infer-

ences was considered the final ML tree. Non-parametric bootstrap analysis was performed on the concatenated matrix using RaxML (v8.1.3). The a posteriori boot-stopping criterion [58] (MRE bootstrapping convergence criterion) was applied to define the number of replicates. After every 50 replicates, the set of bootstrapped trees generated so far is repeatedly (100 permutations) split in two equal subsets, and the Weighted Robinson-Foulds (WRF) distance is calculated between the majority-rule consensus trees of both subsets (for each permutation). Low WRF distances ($< 3\%$) for $>= 99\%$ of permutations were used to indicate bootstrapping convergence. Convergence was reached after 250 replicates: average weighted Robinson-Foulds distance (WRF) = 1.86%, percentage of permutations in which the WRF was % 3.00 = 100%. The tree was visualized and rooted in FigTree (v1.4.2) using *S. paradoxus* as the outgroup (http://tree.bio.ed.ac.uk/software/figtree/).

Phylogenetic Tree for the Extended Collection, Figure 6.2, p. 87

In order to compare our strain collection with previously sequenced strains, we included 24 additional isolates, previously described in [10] (re-sequenced in [16]) and [13] **(Table 6.3, p. 127)**. Generation of MSAs and construction of the concatenation matrix was performed as described earlier. The resulting supermatrix included 2,785,239 positions, 99.077% nucleotides, 0.922% gaps and 0.001% ambiguities. The matrix was partitioned based on all 2,020 gene blocks and all three codon positions within each block, resulting in 6,060 distinct data partitions, accounting for 163,920 distinct alignments patterns. The ML searches and re-optimization were run on 30 fully random starting trees as described above, using RAxML (v8.1.3) and ExaML (v3). Non-parametric bootstrap analysis was performed as described above. Convergence was reached after 250 replicates: average weighted Robinson-Foulds distance (WRF) = 2.10%, percentage of permutations in which the WRF was % 3.00 = 99%. The tree was visualized and rooted in FigTree using *S. paradoxus* as the outgroup.

Multi-locus Phylogeny, Figure 6.14, p. 107

Nine partial genes previously used to genetically characterize 99 Chinese isolates of *S. cerevisiae* [59] were recovered from 194 previously sequenced genomes **(Table 6.4, p. 127)** and from the 157 isolates sequenced in this study, for a total of 450 strains. Each gene was aligned with MAFFT (v7.187) and the final MSAs were concatenated with FASconCAT (v1.0). The concatenated alignment was trimmed using trimAl (v1.2) with the automated1 option, optimized for ML tree reconstruction. The resulting supermatrix included 19,254 positions, 83.348% nucleotides and 16.652% gaps. The matrix was partitioned based on gene blocks in nine distinct partitions with joint branch length optimization. ML search on 30 fully random starting trees and non-parametric bootstrap analysis were performed as described above. Convergence was reached after 550 replicates: average weighted Robinson-Foulds distance (WRF) = 2.10%, percentage of permutations in which the WRF was % 3.00 = 99%. The tree was visualized and rooted in FigTree using *S.*

paradoxus as the outgroup.

6.5.9 Population Structure and Diversity Analysis

The model-based Bayesian algorithm fastSTRUCTURE (v1.0) was used to detect and quantify the number of populations and the degree of admixture in the 157 sequenced genomes [15]. The set of 421,361 biallelic segregating sites identified above was filtered further by removing SNPs with a minor allele frequency (MAF) < 0.05 and SNPs in linkage-disequilibrium, using PLINK (v1.07) [60] with a window of size 50 SNPs advanced by 5 SNPs at a time and an r2 threshold of 0.5. fastSTRUCTURE was run on a filtered set of 53,929 segregating sites, varying the number of ancestral populations (K) between 1 and 10 using the simple prior implemented in fastSTRUCTURE. The number of iterations varied from 10 at K = 1 up to 80 at K = 10. K = 8 was found to be optimal, i.e., scoring the highest marginal likelihood (log-marginal likelihood: 0.7298459788, total iterations: 60). Analysis of estimated ancestry (Q) matrices and plotting were performed in R (v3.1) [61]. The same set of 53,929 SNPs was used to perform a principal component analysis (PCA) as implemented in the SNPrelate package (v1.6.4) [62]. Whole-genome levels of polymorphism were estimated using Variscan (v2.0.3) [63], considering only sites with valid high-quality alleles (> Q40) in at least 80% of strains within the group (defined with the NumNuc parameter adjusted for each group together with CompleteDeletion = 0 and FixNum = 0).

6.5.10 Time Divergence Estimate

In order to estimate the timing of the split leading to the Beer 1 and Beer 2 clade, the mutation rate of yeasts in a beer fermentation environment was calculated. This calculation was based on four assumptions:

US Beer Yeasts Originate from UK Beer Yeasts

Phylogenetic analysis of strains sequenced in this study and previously sequenced wild isolates reveals that US beer yeasts are genetically closely related to European beer yeasts, but not to strains isolated from natural sources in the US. This strongly suggests migration of strains from Europe to the US after colonization. Moreover, the average per-site nucleotide divergence (d_{XY}) further indicates that these US strains likely originate from the UK **(Table 6.2, p. 126)**. The per-site nucleotide diversity between subpopulations was calculated on the set of 2,020 genes used for the inference of the strain phylogeny using the R package PopGenome [64].

The Split between US and UK Beer Yeasts Happened between 1607 and 1637 AD

The first permanent English settlement in North America was established in 1607 in Jamestown, Virginia. In 1609, American "help wanted" advertisements appear in London seeking brewers for

this colony, indicating the importance of beer brewing in early colonial America. In 1637, the well-known colonist Captain Sedgwick founded the first authoritatively recorded brewery in the Massachusetts Bay Colony. Therefore, it is likely that even in early colonial America, beer was produced using yeasts that were brought in by English settlers.

Beer Yeasts Reproduce Clonally during Beer Brewing

During beer brewing, yeasts do not face long periods of nutrient starvation. Nutrient starvation is generally required to induce sporulation and thus initiate sexual reproduction. Indeed, our analyses of the sexual lifestyle of beer yeasts revealed that most beer strains, especially from Beer 1, lost the ability to produce viable spores (only observed in $\sim 6\%$ of the strains, and only in response to severe nutrient-poor conditions), further indicating that sexual reproduction is not favorable, and definitely not common, for beer yeasts during brewing.

Beer Yeasts Undergo around 150 Doublings per Year

As beer fermentations take around one week, and yeasts on average undergo a bit under three doublings per fermentation, it can be estimated that they undergo about 150 generations per year.

Using these parameters, an estimated mutation rate per site per generation was calculated from the formula $k = 2\mu t$, where k is the average per-site nucleotide divergence between US and UK strains (1.97E-03, see Table S8), μ is the mutation rate per site per generation and t is the time in generations, assuming 150 generations per year and divergence of US and UK strains between 1607 and 1637. This calculation yields a mutation rate of 1.61-1.73E-08/bp/generation, a value approximately 50x greater than the value typically calculated in haploid laboratory strains in non-stressful conditions [31]. While this value might seem high, it is not unreasonable for several reasons.

First, the mutation rate estimates obtained here are comparable to those measured in a directed evolution experiment performed in 6%-12% ethanol [22]. Although the conditions that were used in this directed evolution experiment differ from real beer fermentations, they do show that ethanol has a drastic effect on the mutation rate. Moreover, a 6%-12% ethanol concentration should be fairly comparable to the ethanol concentrations encountered in beer brewing over the past 400 years. It is a common misunderstanding that the alcohol percentage of ale beer in the past centuries used to be much lower than it is now. Indeed, in the time span that we considered in our calculations (the past 400 years), high-alcohol beers were being produced (Stan Hieronymus, Anders Kissmeyer, Martyn Cornell and Ron Pattinson, personal communication). The generally low ethanol tolerance of beer yeasts (as compared to wine and spirits yeasts, for example) and the presence of other stress factors during industrial fermentation (nutrient starvation for example) might further contribute to an increased mutation rate.

Second, yeast only undergoes about three cell divisions during beer fermentations, which gener-

ally take place in the first 48 hr of the fermentation. After this, the yeast cells are further exposed to high ethanol concentrations for several days, and it has been shown for several microbes that in this state of quiescence, mutations can still accumulate [34]. Hence, mutations can also occur in the second phase of fermentation, when the cells are not dividing, which implies that the mutation rate (per generation) in industrial growing conditions should be higher than what is measured under conditions where the cells are dividing frequently, as is usually the case in laboratory experiments.

Given the mutation rate estimate μ = 1.61-1.73E-08/bp/generation, an average of 150 generations/year and an average divergence d_{xy} = 2.14E-03 substitutions/site between the UK/US and Belgium/Germany subclades in the Beer1 lineage, the last common ancestor of the major Beer 1 subclades is calculated to have existed until d_{xy} / (2 μ * 150) = $\sim$ 443-412 years ago. A similar calculation for Beer 2 (d_{xy} = 1.79E-03 substitutions/site between the earliest diverging Beer 2 subclades) suggests that the last common ancestor of Beer 2 existed until $\sim$ 371-345 years ago. Given the limited amount of information that could be used for dating, both ages should be considered only rough approximations.

6.5.11 Copy-Number Variation Analysis

Copy-number variations (CNVs) were identified on the reference-based alignments. Initial read depth profiles were obtained for each isolate based on the average read depth calculated in non-overlapping windows of 1000bp. In 68 samples (BE044-BE102, LA002, SP008-SP011, NA005-NA007, WI019), a deviation in read depth was detected: instead of fluctuating around a constant line, the read depth profile showed a convex trend with high depth at the terminal regions of the chromosomes that gradually decreased toward the center. These samples also showed high local variance. This bias in coverage is further referred to as a "smiley pattern". Since conventional methods for CNV detection rely on read depth as a proxy for copy number, these methods were not applicable on the "smiley pattern" strains. To tackle this problem, a custom-built algorithm was developed, dubbed Splint (available upon request), which instead measures the size of discontinuities in read depth by using a discontinuous spline regression technique. In Splint, the data were modeled as the product of the bias and the copy number of each region, plus error. Here, the bias was assumed to be a continuous curve (expected depth as a function of chromosomal location), modeled as a smoothing spline. The copy number on the other hand is a piecewise constant function, with discontinuities at breakpoints in between regions of constant copy number. This was modeled as a sum of indicator functions, one for each region. After regression, the fitted value of the coefficient of each indicator function is proportional to the copy number in the corresponding region. The regression method requires the locations of the discontinuities as input values. Initially, these are located in a rough manner by comparing the 50kb regions to the left and right of each 1000 bp window. If the difference between the median depth in the left and right regions is small, the frame is not likely to contain a copy-number breakpoint; if the

difference is large, it may contain a breakpoint. This measure is smoothed (by moving average) and corrected for linear bias by subtracting a linear trendline. Peaks that exceed 2.5 times the sample-wide median, in absolute value, are annotated as breakpoints. However, this method only gives rough coordinates of discontinuities, delimiting large regions of constant copy number. After this rough estimation, an initial regression was run, and a hidden Markov model (HMM) was used to find regions where the regressed values are significantly different from the data. The HMM accepts deviance of the estimated curve from the data as input signals (15% greater than, 15% lesser than, or approximately equal), and aggregates high densities of deviant signals into output states (under-estimation, over-estimation or correct estimation of copy number; better results were obtained when a special state was reserved for total deletions). The windows where the state changes are seen as likely breakpoints. The regression and HMM were re-evaluated until no more deviating regions could be found. The regression coefficients of the piecewise constant function in the final regression are proportional to the copy number in the corresponding regions, but the proportionality constant depends on the shape and scale of the continuous (spline) factor in the regression, which is different for each chromosome. The form of the spline is such that its value is always 1 in the left telomere for each chromosome. Using the regression coefficients of the piecewise function as a proportional proxy for the copy numbers implicitly assumes that the bias is the same for each chromosome at the left telomere. We observe that the smiley pattern is generally similar on both sides of the chromosomes, so we repeat the regression setting the spline value at the right telomere at 1, and instead use the means of the two sets of regression coefficients to estimate the copy number. Splint was run using frames of 1000bp and 500bp. Shorter frames will result in higher resolution of the CNV calls at the cost of an increased rate of false positive calls. Because results that depend on the window size were not deemed robust, only CNVs found in both 1000bp and 500bp window analyses were used in the final results. The functional enrichment analysis of CNV-driven genes was carried out using the Gorilla database [65] using the complete set of *S. cerevisiae* genes as the reference. False discovery rate (FDR) Benjamini & Hochberg adjusted q % 0.05 were considered significant.

6.5.12 Character Evolution Analysis

Ancestral character states for the production of the phenolic off-flavor 4-vinyl guaiacol (4-VG) were estimated based on the sequences of the two key genes *PAD1* and *FDC1*. The protein-coding nucleotide sequences of *PAD1* and *FDC1* were retrieved from the 157 *de novo* assemblies obtained in this study and from the outgroup species *S. paradoxus*. Because the annotation procedure described above excluded all genes including internal stop codons, a local BLAST database was set up for all the genomes and BLASTN searches were performed (1E-04E-value cut-off) using the *PAD1* and *FDC1* coding sequences from the *S. cerevisiae* strain S288c reference genome (R64-2-1). We found that in one bioethanol strain (BI002), both genes resulted from an introgression

event from *S. paradoxus*. *S. paradoxus* introgression events involving *PAD1* and *FDC1* have been previously reported for the Brazilian bioethanol strain BG1 [18]. In another bioethanol strain (BI006), *PAD1* was shown to be a chimeric gene between the *S. cerevisiae* and *S. paradoxus* allele. This latter strain was therefore excluded from the analysis. Considering the presence of stop codons and frameshift mutations in some of the sequences, the sets of protein-coding sequences for either *PAD1* or *FDC1* were aligned with MACSE (v1.0b), a tool that prevents the disruption of the underlying codon structure when aligning non-functional sequences [66]. Each MSA was partitioned based on codon positions according to three schemes: i) CP111 - all codon positions are combined; ii) CP112 - first and second codon positions are combined and third position is independent; iii) CP123 all codon positions are independent. The optimal partitioning scheme and the best-fit nucleotide substitution model for each partition of the two MSAs were estimated using the software PartitionFinder (v1.1.1) [67]. For all analyses, branch lengths were linked between partitions and 24 substitutions models were considered for each partition. Model and partitioning scheme were selected based on the Bayesian information criterion (BIC). For *PAD1*, the best scheme was obtained with CP112, and K80 and HKY+G (gamma-distributed rate heterogeneity across sites using four rate categories) were the best-fit nucleotide substitution models assigned for the two partitions respectively. For *FDC1* the best scheme was obtained with CP123, and HKY was selected as the best-fit nucleotide substitution model for each of the three partitions. All phylogenetic analyses and ancestral state reconstructions were performed in BEAST (v1.8.2) [30]. The trait was treated as discrete and one of two character states were assigned to each isolate based on GC analysis (see further): production of 4-VG, state = 1; no production of 4-VG, state = 0. Monophyly was imposed on all the *S.cerevisiae* strains with the exception of strain BI002. The phylogenetic tree and ancestral state for all internal nodes were simultaneously inferred for each gene, to account for phylogenetic uncertainty. The partitioned MSAs were examined under three different clock models: i) a global molecular clock with fixed evolutionary rates; ii) an uncorrelated relaxed molecular clock with an underlying lognormal distribution (UCLD) on the evolutionary rates; iii) a random local molecular clock (RLC) that allows different evolutionary rates in sub-regions of the phylogenetic tree [68]. Additionally, each clock model was tested in combination with an asymmetric versus a symmetric model of trait evolution [69]. A pure-birth Yule speciation prior and a random starting tree were used for all the analyses. The suitable number of iterations to allow convergence and proper mixing of the Markov chain Monte Carlo (MCMC) runs was determined for each MSA-molecular clock-trait model combination using Tracer (v1.6) [70]. Effective Sample Size (ESS) > 100 was reached for all parameters in each run. Model selection was performed in BEAST by comparing marginal likelihoods estimated using path sampling and stepping-stone sampling with a chain length of 2 million generations sampling every 200 steps [71, 72]. Model comparison showed strong support for the RLC model and asymmetry in the evolution of the trait for both *PAD1* and *FDC1*. A second independent run for both *PAD1* and *FDC1* was performed under the most supported scheme to ensure convergence to the same topology. LogCombiner [30]

was used to remove burn-in trees (10%) and to resample to a frequency of 10,000. The final Maximum Clade Credibility (MCC) trees were obtained in TreeAnnotator [30] on 19,998 trees in total for each gene. MCC trees were visualized in FigTree (v1.4.2).

6.5.13 Determination of Cell Ploidy

DNA content of the sequenced strains (a measure for the ploidy level) was determined by staining cells with propidium iodide (PI) and analysis of 50,000 stained cells by flow cytometry on a BD Influx (BD Biosciences, USA). The fluorescent signal of previously established haploid (BY4742), diploid (BY4743) and tetraploid (BR001) strains was used to generate a calibration curve. Highly flocculent strains were excluded from the analysis (missing bar charts **Figure 6.4, p. 91)**.

6.5.14 Phenotypic Analysis

Flavor Production and Flocculation in Fermentation Conditions

To assess the production of aroma-active compounds, lab-scale fermentation experiments were performed. These fermentations were performed in rich growth medium (YPGlu 10%; peptone 2% w v^{-1}, yeast extract 1% w v^{-1}, glucose 10% w v^{-1}). Yeast precultures were shaken overnight at 30 °C in test tubes containing 5mL of yeast extract (1% w v^{-1}), peptone (2% w v^{-1}) and glucose (4% w v^{-1}) medium (YPGlu 4%). After 16 hr of growth, 0.5 mL of the preculture was used to inoculate 50 mL of YPGlu 4% medium in 250 mL Erlenmeyer flasks, and this second preculture was shaken at 30 °C for 16 hr. This second preculture was used for inoculation of the fermentation medium (YPGlu 10%) at an initial optical density (at 600 nm; OD 600) of 0.5, roughly equivalent to 107 cells mL^{-1}. The fermentations, performed in 250 mL Schott bottles with a water lock placed on each bottle, were incubated statically for 7 days at 20 °C. Weight loss was measured daily to estimate fermentation progress. After 7 days, the fermentations were stopped, filtered (0.15 mm paper filter) and samples for chromatographic analysis, density and ethanol measurements were taken.

Headspace gas chromatography coupled with flame ionization detection (HS-GC-FID) (Agilent Technologies, USA) was used for the quantification of yeast aroma production. The GC was calibrated for 16 important aroma compounds, including esters (ethyl acetate, isobutyl acetate, propyl acetate, isoamyl acetate, phenyl ethyl acetate, ethyl propionate, ethyl butyrate, ethyl hexanoate, ethyl octanoate, ethyl decanoate), higher alcohols (isoamyl alcohol, isobutanol, butanol, phenyl ethanol), acetaldehyde and 4-VG. The GC was equipped with a headspace autosampler (PAL system, CTC analytics, Switzerland) and contained a DB-WAXETER column (length, 30 m; internal diameter, 0.25 mm; layer thickness, 0.5 mm, Agilent Technologies, USA) and N2 was used as the carrier gas. Samples were heated for 25 min at 70 °C in the autosampler. The injector block and FID temperatures were both kept constant at 250 °C. Samples of 5mL filtered fermentation medium were collected in 15 ml glass tubes containing 1.75 g of sodium chloride each. These

tubes were immediately closed and cooled to minimize evaporation of volatile compounds. The oven temperature was held at 50 °C for 5 min, after which it increased to 80 °C at 4 °C min^{-1}. Next, it increased to 200 °C at 5 C min-1 and was held at 200 °C for 3 min. Results were analyzed with the Agilent Chemstation software (Agilent Technologies, USA).

Additionally, after fermentation, the flocculation character of each strain was scored visually from 1 (not flocculent) to 6 (extremely flocculent, big flocs).

Ethanol Accumulation Capacity

To assess the maximal ethanol accumulation capacity of all strains, fermentation tests were performed in rich medium containing 35% w v^{-1} glucose. Yeast precultures were shaken overnight at 30 °C in test tubes containing 5mL YPGlu 4%. After 16 hr of growth, 0.5 mL of the preculture was used to inoculate 50 mL of YPGlu 4% medium in 250 mL Erlenmeyer flasks, and this second pre-culture was shaken at 30 °C for 16 hr. This preculture was used for inoculation of the fermentation medium (peptone 2% w v^{-1}, yeast extract 1% w v^{-1}, glucose 35% w v^{-1}; YPGlu 35%) at an initial OD600 of 0.5, roughly equivalent to 107 cells mL^{-1}. The fermentations, performed in 250 ml Schott bottles with a water lock placed on each bottle, were incubated statically for 14 days at 30 °C. Weight loss was measured daily to estimate fermentation progress. After 14 days, the fermentations were stopped, filtered (0.15 mm paper filter) and samples for ethanol measurements [performed with the Alcolyzer Beer DMA 4500M (Anton Paar, Austria)] were taken.

Screening for Environmental and Nutrient Stress Tolerance

All strains were tested in robot-assisted, high-throughput spotting assays in several conditions. All isolates were evaluated on YPGlu 2% agar (Yeast Extract 1% w v^{-1}, Peptone 2% w v^{-1}, Glucose 2% w v^{-1}, agar 2% w v^{-1}) for (i) temperature tolerance (4 °C 16 °C 30 °C 40 °C), (ii) sugarand/or osmotolerance using increasing concentrations of glucose and sorbitol (final osmolyte concentration of 46 48 50% w v^{-1}), (iii) acid tolerance using increasing concentrations of acetic (50 75 100 mM), levulinic (25 50 75 mM) and formic acid (50 75 mM), (iv) sulphite tolerance using increasing concentrations of SO2 (1.50 2.25 3.00 mM), (v) ethanol tolerance using increasing concentrations of ethanol (5 7 9 10 11 12 13% v v^{-1}), (vi) actidione (= cycloheximide) tolerance using increasing concentrations of actidione (0.2 0.4 mgL^{-1}), (vii) halotolerance using increasing concentrations of NaCl (250 500 100 mM) and KCl (500 1000 - 1500 mM) and (viii) metal tolerance using 0.075 mM copper and increasing concentrations of cadmium (0.3 0.4 0.5 mM). Stressor concentrations were selected based on previous pilot experiments with a broader concentration range (data not shown). Additionally, temperature tolerance (10 °C 39 °C) on three different carbon sources (glucose, fructose, sucrose, ethanol and maltose) was assessed on YP agar supplemented with 2% w v^{-1} of one of the carbon sources (or, in case of ethanol, 2% v v^{-1}). For each of these experiments, growth on YPGlu 2% agar on 20 °C was used as a control condition.

For utilization of different carbon sources, experiments were performed on SC (Synthetic Complete) agar containing 2% w v^{-1} galactose, glycerol, melibiose, sorbitol, ethanol, fructose, sucrose or maltose as sole carbon source. Additionally, maltose and maltotriose were assessed in liquid medium (see further). Growth on SC 2% glucose at 20 °C was used as a control.

Prior to the experiment, the 96-well microtiter plates containing the isolates (stored at 80 °C) were thawed and spotted on YPGlu 2% agar and incubated at 30 °C for 48 hr. Next, 96-well plates containing 150ml of YPGlu 2% in each well were inoculated with the isolates and incubated overnight at 30 °C on a microtiter plate shaking platform (Heidolph Instruments, Germany) at 600rpm, allowing the cells to reach stationary phase. Then, the OD600 of all wells was measured using a microtiter plate reader (Molecular Devices, Sunnyvale, USA). Subsequently, the cell density was manually adjusted to OD600 z0.1 in a second 96-well microtiter plate using sterile deionized water in order to standardize the starting cell density for all isolates. This plate was used as the source plate for spotting the test media. In order to maximize throughput and reproducibility, a high-density array robot (Singer Instruments, UK) was used for all spotting or replication steps. After spotting, all plates were sealed using plastic paraffin film and all plates (except for plates used in the thermotolerance assays) were incubated at 20 °C. After 4-14 days of incubation (depending on the experiment), all plates were scanned using a high-definition scanner (Seiko Epson, Japan). Scanned images were processed using ImageJ combined with the ScreenMill macro [73]. Data were processed by calculating relative growth compared to the control condition, and subsequent normalization by conversion to z-scores **(Table S5, included in the data package supplied with this thesis)**. Heat maps were obtained using the Gene-E software (http://www.broadinstitute.org/cancer/software/GENE-E/). Strains were hierarchically clustered based on phenotypic behavior using a centered Pearson correlation metric and average linkage mapping.

Maltose and Maltotriose Fermentation Capacity in Liquid Medium

For maltose and maltotriose fermentation capacity, experiments were performed in 96 well plates with 150 ml SC liquid medium containing 1% (w v^{-1}) of maltose or maltotriose, supplemented with 3 mg L^{-1} antimycin to block respiration. Pregrowth was performed as described above, and cells were inoculated at OD600. OD600 was assessed after 4 days of growth at 20 °C (shaken, 900 rpm). Growth in SC liquid medium containing 1% glucose at 20 °C was used as a control.

4-VG Production

Screening for 4-VG production was assessed by measuring 4-VG production of each strain in medium enriched in the 4-VG precursor, ferulic acid. Strains were pregrown on YPGlu 2% agar, a single colony was picked and directly inoculated in a GC vial filled with 5 mL of test medium (YPGlu 2% supplemented with 100 mg L^{-1} ferulic acid). Vials were capped (but not completely closed) and wrapped with plastic paraffin film and statically incubated at 30 °C for 3 days. Next,

4-VG concentration was measured using GC-FID as described earlier. Strains were scored as 4-VG+ if the concentration produced was significantly higher compared to the non-inoculated fermentation medium.

Investigation of the Yeast's Sexual Life Cycle

Sporulation was induced on minimal sporulation medium [1% (w v^{-1}) KAc, 0.05% (w v^{-1}) amino acids, 2% (w v^{-1}) agar] at 23 °C after pre-growth in YPGlu 2%. Tetrad dissection of 4 tetrads of each strains was carried out using a Singer micromanipulator (Singer Instruments, UK), and mating-type determination of all germinated spores was carried out by mating-type PCR.

6.5.15 Development of Artificial Hybrids

Sporulation, tetrad dissection, and mating type characterizations were performed as described earlier. For SNP genotyping, PCR primers were developed for the W497* stop-gained mutation in *FDC1*. To detect segregants carrying the stop-gained mutation, following primers were used: W497-FW (50-TGCAGATCAGATGGCTTTTG-30), W497-RV-STOP (50-GCAATTATTTATATCCGTACCT TTTT-30). To detect the alternative allele (without the stop-gained mutation), following primers were used: W497-FW (50-TGCAGATC AGATGGCTTTTG-30), W497-RV-ALT (50-GCAATTATTTATATCCGTACCTTTTC-30).

To hybridize haploid segregants, a direct mating approach as described in [74], was performed. The segregants were first streaked to single colonies on a YPGlu 2% agar plate. One colony of each segregant was picked, and both were mixed on a second YPGlu 2% agar plate. Ten microliters of distilled water were added to the mixed cell cultures to increase mixing efficiency. The plates were dried and incubated at room temperature for 10-12 hr. A small fraction of the spot was picked with a toothpick and streaked to single colonies on a fresh YPGlu 2% agar plate. After 48 hr of incubation, the diploid status of the resulting colonies was verified by mating type PCR, the presence of both genomes in the hybrid by Interdelta Analysis [75].

6.5.16 Quantification and statistical analysis

Standard statistical analyses were conducted in RStudio (v.0.98.994) (https://www.rstudio.com/) with custom scripts.

Preprocessing of the phenotypic data to produce **Figure 6.6, p. 94** and Table S5 consisted of a conversion to z-scores, calculated as follows: z-score = $(X_i\ \mu)/\sigma$, where X_i is the data value for strain i, μ the mean of all strains and σ the standard deviation across all strains. To facilitate direct comparison between different strains for a specific trait, a reference condition for each environmental stressor related trait was determined. This reference condition is defined as the most stringent condition (i.e., the condition with the highest stressor concentration or most extreme

temperature) where around 50% of the investigated strains still managed to reach a colony area greater than 10% of their colony area in the control condition (YPGlu 2% agar, 20 °C).

The phenotypic variability explained by the set of mutations identified in *MAL11* gene was computed as following: first, all SNPs and InDels were searched in pairwise comparison for high correlations (> 0.90). A Ward clustering was additionally performed to retain or exclude mutations according to the clusters identified. Second, a linear regression analysis was performed on individual mutations; the mutation was retained if significantly associated with the phenotype, with $p < 0.001$ considered as significant. Last, REML (restricted maximum likelihood) analysis with completely random effects was carried out on the final set of SNPs in SAS (v9.4).

6.5.17 Data avalability

The accession numbers for the *de novo* assembly data reported in this paper are DDBJ/ENA/GenBank: Bioproject ID, PRJNA323691, Biosample ID SAMN05190362−SAMN05190518, and MBUB00000000 − MCAB00000000 **(Table 6.3, p. 127)**.

6.5.18 Supplementary tables

Table 6.2: Distance between subpopulations, determined by the average number of nucleotide substitutions per site, d_{XY}, calculated in 2,020 genes.

	Britain	US	Belgium/Germany	Mixed	Wine	Beer 2
US	1.97E-03					
Belgium/Germany	2.10E-03	2.26E-03				
Mixed	2.34E-03	2.51E-03	2.41E-03			
Wine	2.39E-03	2.51E-03	2.54E-03	2.34E-03		
Beer 2	2.53E-03	2.68E-03	2.66E-03	2.46E-03	1.68E-03	
Asia	3.65E-03	3.75E-03	3.60E-03	3.18E-03	4.33E-03	4.00E-03

Table 6.3: Additional Information on the *S. cerevisiae* strains investigated in this chapter and strains added as reference strains to the phylogenetic analyses. Related to **Figure 6.2, p. 87**

This table is provided at:
https://www.dropbox.com/s/0tav1e17t08yn5x/Table%20S1.xlsx?dl=0

Table 6.4: Referenced overview of all strains included in the multilocus phylogenetic analysis. Related to **Figure 6.14, p. 107**

This table is provided at:
https://www.dropbox.com/s/isn3r95hclni0av/Table%20S2.xlsx?dl=0

Table 6.5: Gene Ontology Enrichment Analysis of genes with increased and decreased copy-number variation in all strains of the collection. Related to **Figure 6.4, p. 91**

This table is provided at:
https://www.dropbox.com/s/ih7q01m7esh5ekv/Table%20S3.xlsx?dl=0

Table 6.6: Gene Ontology Enrichment Analysis of genes with increased and decreased copy-number variation in individual subpopulations. Related to **Figure 6.4, p. 91**

This table is provided at:
https://www.dropbox.com/s/v73b6eyc54d5vbc/Table%20S4.xlsx?dl=0

Table 6.7: Phenotypic variation of Industrial *S. cerevisiae* strains for a selection of 82 traits, depicted as z-scores. Related to **Figure 6.6, p. 94**. Traits were selected for their industrial relevance, and/or their link to survival. Additionally, to document the absolute variability of each trait across strains, the theoretical range, minimum, maximum, median, mean and standard deviation value of the raw data are given below each column.

This table is provided at:
https://www.dropbox.com/s/gd8cet2xxdceuwk/Table%20S5.xlsx?dl=0

Table 6.8: Statistical analysis of the phenotypic behavior of the different subpopulations in environmental stress conditions, ethanol accumulation capacity, and maltotriose utilization. Related to **Figure 6.7a,6.7b,6.7c,6.7d, p.95 and Figure 6.16, p.109**. For each environmental stressor, a representative condition was selected (defined as the conditions where approximately 50% of the investigated isolates are still able to grow). Statistical analyses were performed using Unbalanced ANOVA with Post Hoc Tukey tests: F, ANOVA F-ratio (degrees of freedom between groups, degrees of freedom within groups); MSE, mean squared error. Asterisks indicate the level of significance ("***" 0, "**" 0.001, "*" 0.01, "." 0.05).

This table is provided at:
https://www.dropbox.com/s/jee1tw9dd6cane4/Table%20S6.xlsx?dl=0

Table 6.9: Distribution of *MAL11* alleles in the sequenced strains. Related to **Figure 6.7d, p. 95**. "0/0" indicates absence of the corresponding deleterious frameshift, "0/1" its heterozygous presence, "1/1" its homozygous presence and "./." the complete absence of the MAL11 gene. The phenotype represents growth of the strains in 150 μL liquid SC medium containing 1% (w v-1) maltotriose, relative to growth on SC medium containing 1% (w/v) glucose instead of maltotriose.

This table is provided at:
https://www.dropbox.com/s/ep7u7l85n4n4j4b/Table%20S7.xlsx?dl=0

6.6 References

[1] R. Michel and P. McGovern, "Chemical evidence for ancient beer," *Nature*, pp. 145–150, 1992.

[2] J. Steensels and K. J. Verstrepen, "Taming wild yeast: potential of conventional and nonconventional yeasts in industrial fermentations.," *Annual review of microbiology*, vol. 68, pp. 61–80, 2014.

[3] M. R. Goddard and D. Greig, "*Saccharomyces cerevisiae*: a nomadic yeast with no niche?," *FEMS yeast research*, vol. 15, no. 3, 2015.

[4] J. Warringer, E. Zorgo, F. A. Cubillos, A. Zia, A. Gjuvsland, J. T. Simpson, A. Forsmark, R. Durbin, S. W. Omholt, E. J. Louis, G. Liti, A. Moses, and A. Blomberg, "Trait variation in yeast is defined by population history.," *PLoS genetics*, vol. 7, no. 6, p. e1002111, 2011.

[5] C. A. Driscoll, D. W. Macdonald, and S. J. O'Brien, "From wild animals to domestic pets, an evolutionary view of domestication.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106 Suppl 1, pp. 9971–9978, 2009.

[6] M. D. Purugganan and D. Q. Fuller, "The nature of selection during plant domestication.," *Nature*, vol. 457, no. 7231, pp. 843–848, 2009.

[7] P. Almeida, R. Barbosa, P. Zalar, Y. Imanishi, K. Shimizu, B. Turchetti, J.-L. Legras, M. Serra, S. Dequin, A. Couloux, J. Guy, D. Bensasson, P. Goncalves, and J. P. Sampaio, "A population genomics insight into the Mediterranean origins of wine yeast domestication.," *Molecular ecology*, vol. 24, no. 21, pp. 5412–5427, 2015.

[8] A. R. Borneman, B. A. Desany, D. Riches, J. P. Affourtit, A. H. Forgan, I. S. Pretorius, M. Egholm, and P. J. Chambers, "Whole-genome comparison reveals novel genetic elements that characterize the genome of industrial strains of *Saccharomyces cerevisiae*.," *PLoS genetics*, vol. 7, no. 2, p. e1001287, 2011.

[9] A. R. Borneman, A. H. Forgan, R. Kolouchova, J. A. Fraser, and S. A. Schmidt, "Whole Genome Comparison Reveals High Levels of Inbreeding and Strain Redundancy Across the Spectrum of Commercial Wine Strains of *Saccharomyces cerevisiae*.," *G3 (Bethesda, Md.)*, vol. 6, no. 4, pp. 957–971, 2016.

[10] G. Liti, D. M. Carter, A. M. Moses, J. Warringer, L. Parts, S. A. James, R. P. Davey, I. N. Roberts, A. Burt, V. Koufopanou, I. J. Tsai, C. M. Bergman, D. Bensasson, M. J. T. O'Kelly, A. van Oudenaarden, D. B. H. Barton, E. Bailes, A. N. Nguyen, M. Jones, M. A. Quail, I. Goodhead, S. Sims, F. Smith, A. Blomberg, R. Durbin, and E. J. Louis, "Population genomics of domestic and wild yeasts.," *Nature*, vol. 458, no. 7236, pp. 337–341, 2009.

[11] P. M. Magwene, O. Kayikci, J. A. Granek, J. M. Reininga, Z. Scholl, and D. Murray, "Outcrossing, mitotic recombination, and life-history trade-offs shape genome evolution in *Saccharomyces cerevisiae*.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 5, pp. 1987–1992, 2011.

[12] J. Schacherer, J. A. Shapiro, D. M. Ruderfer, and L. Kruglyak, "Comprehensive polymorphism survey elucidates population structure of *Saccharomyces cerevisiae*.," *Nature*, vol. 458, no. 7236, pp. 342–345, 2009.

[13] P. K. Strope, D. A. Skelly, S. G. Kozmin, G. Mahadevan, E. A. Stone, P. M. Magwene, F. S. Dietrich, and J. H. McCusker, "The 100-genomes strains, an *S. cerevisiae* resource that illuminates its natural phenotypic and genotypic variation and emergence as an opportunistic pathogen.," *Genome research*, vol. 25, no. 5, pp. 762–774, 2015.

[14] J. E. Perez-Ortin, A. Querol, S. Puig, and E. Barrio, "Molecular characterization of a chromosomal rearrangement involved in the adaptive evolution of yeast strains.," *Genome research*, vol. 12, no. 10, pp. 1533–1539, 2002.

[15] A. Raj, M. Stephens, and J. K. Pritchard, "fastSTRUCTURE: variational inference of population structure in large SNP data sets.," *Genetics*, vol. 197, no. 2, pp. 573–589, 2014.

[16] A. Bergstrom, J. T. Simpson, F. Salinas, B. Barre, L. Parts, A. Zia, A. N. Nguyen Ba, A. M. Moses, E. J. Louis, V. Mustonen, J. Warringer, R. Durbin, and G. Liti, "A high-definition view of functional genetic variation from natural yeast genomes.," *Molecular biology and evolution*, vol. 31, no. 4, pp. 872–888, 2014.

[17] M. J. Dunham, H. Badrane, T. Ferea, J. Adams, P. O. Brown, F. Rosenzweig, and D. Botstein, "Characteristic genome rearrangements in experimental evolution of *Saccharomyces cerevisiae*.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 25, pp. 16144–16149, 2002.

[18] B. Dunn, C. Richter, D. J. Kvitek, T. Pugh, and G. Sherlock, "Analysis of the *Saccharomyces cerevisiae* pan-genome reveals a pool of copy number variants distributed in diverse yeast strains from differing industrial environments.," *Genome research*, vol. 22, no. 5, pp. 908–924, 2012.

[19] N. Pavelka, G. Rancati, J. Zhu, W. D. Bradford, A. Saraf, L. Florens, B. W. Sanderson, G. L. Hattem, and R. Li, "Aneuploidy confers quantitative proteome changes and phenotypic variation in budding yeast.," *Nature*, vol. 468, no. 7321, pp. 321–325, 2010.

[20] G. Rancati, N. Pavelka, B. Fleharty, A. Noll, R. Trimble, K. Walton, A. Perera, K. Staehling-Hampton, C. W. Seidel, and R. Li, "Aneuploidy underlies rapid adaptive evolution of yeast cells deprived of a conserved cytokinesis motor.," *Cell*, vol. 135, no. 5, pp. 879–893, 2008.

[21] A. M. Selmecki, K. Dulmage, L. E. Cowen, J. B. Anderson, and J. Berman, "Acquisition of aneuploidy provides increased fitness during the evolution of antifungal drug resistance.," *PLoS genetics*, vol. 5, no. 10, p. e1000705, 2009.

[22] K. Voordeckers, J. Kominek, A. Das, A. Espinosa-Cantu, D. De Maeyer, A. Arslan, M. Van Pee, E. van der Zande, W. Meert, Y. Yang, B. Zhu, K. Marchal, A. DeLuna, V. Van Noort, R. Jelier, and K. J. Verstrepen, "Adaptation to High Ethanol Reveals Complex Evolutionary Pathways.," *PLoS genetics*, vol. 11, no. 11, p. e1005635, 2015.

[23] J. M. Sheltzer, H. M. Blank, S. J. Pfau, Y. Tange, B. M. George, T. J. Humpton, I. L. Brito, Y. Hiraoka, O. Niwa, and A. Amon, "Aneuploidy drives genomic instability in yeast.," *Science (New York, N.Y.)*, vol. 333, no. 6045, pp. 1026–1030, 2011.

[24] C. A. Brown, A. W. Murray, and K. J. Verstrepen, "Rapid expansion and functional divergence of subtelomeric gene families in yeasts.," *Current biology : CB*, vol. 20, no. 10, pp. 895–903, 2010.

[25] P. Briza, M. Breitenbach, A. Ellinger, and J. Segall, "Isolation of two developmentally regulated genes involved in spore wall maturation in *Saccharomyces cerevisiae*.," *Genes & development*, vol. 4, no. 10, pp. 1775–1789, 1990.

[26] M. R. Goddard, H. C. J. Godfray, and A. Burt, "Sex increases the efficacy of natural selection in experimental yeast populations.," *Nature*, vol. 434, no. 7033, pp. 636–640, 2005.

[27] M. J. McDonald, D. P. Rice, and M. M. Desai, "Sex speeds adaptation by altering the dynamics of molecular evolution.," *Nature*, vol. 531, no. 7593, pp. 233–236, 2016.

[28] M. D. White, K. A. P. Payne, K. Fisher, S. A. Marshall, D. Parker, N. J. W. Rattray, D. K. Trivedi, R. Goodacre, S. E. J. Rigby, N. S. Scrutton, S. Hay, and D. Leys, "UbiX is a flavin prenyltransferase required for bacterial ubiquinone biosynthesis.," *Nature*, vol. 522, no. 7557, pp. 502–506, 2015.

[29] N. Mukai, K. Masaki, T. Fujii, M. Kawamukai, and H. Iefuji, "PAD1 and FDC1 are essential for the decarboxylation of phenylacrylic acids in *Saccharomyces cerevisiae*.," *Journal of bioscience and bioengineering*, vol. 109, no. 6, pp. 564–569, 2010.

[30] A. J. Drummond, M. A. Suchard, D. Xie, and A. Rambaut, "Bayesian phylogenetics with BEAUti and the BEAST 1.7.," *Molecular biology and evolution*, vol. 29, no. 8, pp. 1969–1973, 2012.

[31] M. Lynch, W. Sung, K. Morris, N. Coffey, C. R. Landry, E. B. Dopman, W. J. Dickinson, K. Okamoto, S. Kulkarni, D. L. Hartl, and W. K. Thomas, "A genome-wide view of the spectrum of spontaneous mutations in yeast.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 27, pp. 9272–9277, 2008.

[32] M. Filteau, V. Hamel, M.-C. Pouliot, I. Gagnon-Arsenault, A. K. Dube, and C. R. Landry, "Evolutionary rescue by compensatory mutations is constrained by genomic and environmental backgrounds.," *Molecular systems biology*, vol. 11, no. 10, p. 832, 2015.

[33] D. van Dijk, R. Dhar, A. M. Missarova, L. Espinar, W. R. Blevins, B. Lehner, and L. B. Carey, "Slow-growing cells within isogenic populations have increased RNA polymerase error rates and DNA damage.," *Nature communications*, vol. 6, p. 7972, 2015.

[34] L. Loewe, V. Textor, and S. Scherer, "High deleterious genomic mutation rate in stationary phase of *Escherichia coli*.," *Science (New York, N.Y.)*, vol. 302, no. 5650, pp. 1558–1560, 2003.

[35] I. S. Hornsey, *History of Beer and Brewing*. Cambridge: Cambridge: Royal Society of Chemistry., 2003.

[36] N. A. Bokulich, J. H. Thorngate, P. M. Richardson, and D. A. Mills, "Microbial biogeography of wine grapes is conditioned by cultivar, vintage, and climate.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 111, no. 1, pp. E139–48, 2014.

[37] J. F. Christiaens, L. M. Franco, T. L. Cools, L. De Meester, J. Michiels, T. Wenseleers, B. A. Hassan, E. Yaksi, and K. J. Verstrepen, "The fungal aroma gene ATF1 promotes dispersal of yeast cells through insect vectors.," *Cell reports*, vol. 9, no. 2, pp. 425–432, 2014.

[38] I. Stefanini, L. Dapporto, L. Berna, M. Polsinelli, S. Turillazzi, and D. Cavalieri, "Social wasps are a *Saccharomyces* mating nest.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 113, no. 8, pp. 2247–2251, 2016.

[39] H. J. Muller, "The relation of recombination to mutational advance.," *Mutation research*, vol. 106, pp. 2–9, 1964.

[40] S. Takeda and M. Matsuoka, "Genetic approaches to crop improvement: responding to environmental and population changes.," *Nature reviews. Genetics*, vol. 9, no. 6, pp. 444–457, 2008.

[41] A. M. Bolger, M. Lohse, and B. Usadel, "Trimmomatic: a flexible trimmer for Illumina sequence data.," *Bioinformatics (Oxford, England)*, vol. 30, no. 15, pp. 2114–2120, 2014.

[42] Y. Liu, J. Schroder, and B. Schmidt, "Musket: a multistage k-mer spectrum-based error corrector for Illumina sequence data.," *Bioinformatics (Oxford, England)*, vol. 29, no. 3, pp. 308–315, 2013.

[43] Y. Peng, H. C. M. Leung, S. M. Yiu, and F. Y. L. Chin, "Idba – a practical iterative de bruijn graph de novo assembler," in *Research in Computational Molecular Biology* (B. Berger, ed.), (Berlin, Heidelberg), pp. 426–440, Springer Berlin Heidelberg, 2010.

[44] R. M. Kuhn, D. Karolchik, A. S. Zweig, H. Trumbower, D. J. Thomas, A. Thakkapallayil, C. W. Sugnet, M. Stanke, K. E. Smith, A. Siepel, K. R. Rosenbloom, B. Rhead, B. J. Raney, A. Pohl, J. S. Pedersen, F. Hsu, A. S. Hinrichs, R. A. Harte, M. Diekhans, H. Clawson, G. Bejerano, G. P. Barber, R. Baertsch, D. Haussler, and W. J. Kent, "The UCSC genome browser database: update 2007.," *Nucleic acids research*, vol. 35, no. Database issue, pp. D668–73, 2007.

[45] M. Stanke and B. Morgenstern, "AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints.," *Nucleic acids research*, vol. 33, no. Web Server issue, pp. W465–7, 2005.

[46] L. Fu, B. Niu, Z. Zhu, S. Wu, and W. Li, "CD-HIT: accelerated for clustering the next-generation sequencing data.," *Bioinformatics (Oxford, England)*, vol. 28, no. 23, pp. 3150–3152, 2012.

[47] I. Wapinski, A. Pfeffer, N. Friedman, and A. Regev, "Natural history and evolutionary principles of gene duplication in fungi.," *Nature*, vol. 449, no. 7158, pp. 54–61, 2007.

[48] H. Li and R. Durbin, "Fast and accurate short read alignment with Burrows-Wheeler transform.," *Bioinformatics (Oxford, England)*, vol. 25, no. 14, pp. 1754–1760, 2009.

[49] A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, and M. A. DePristo, "The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data.," *Genome research*, vol. 20, no. 9, pp. 1297–1303, 2010.

[50] P. Danecek, A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, and R. Durbin, "The variant call format and VCFtools.," *Bioinformatics (Oxford, England)*, vol. 27, no. 15, pp. 2156–2158, 2011.

[51] P. Cingolani, A. Platts, L. L. Wang, M. Coon, T. Nguyen, L. Wang, S. J. Land, X. Lu, and D. M. Ruden, "A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3.," *Fly*, vol. 6, no. 2, pp. 80–92, 2012.

[52] K. Katoh and D. M. Standley, "MAFFT multiple sequence alignment software version 7: improvements in performance and usability.," *Molecular biology and evolution*, vol. 30, no. 4, pp. 772–780, 2013.

[53] M. Suyama, D. Torrents, and P. Bork, "PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments.," *Nucleic acids research*, vol. 34, no. Web Server issue, pp. W609–12, 2006.

[54] S. Capella-Gutierrez, J. M. Silla-Martinez, and T. Gabaldon, "trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses.," *Bioinformatics (Oxford, England)*, vol. 25, no. 15, pp. 1972–1973, 2009.

[55] P. Kuck and K. Meusemann, "FASconCAT: Convenient handling of data matrices.," *Molecular phylogenetics and evolution*, vol. 56, no. 3, pp. 1115–1118, 2010.

[56] A. Stamatakis, "RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies.," *Bioinformatics (Oxford, England)*, vol. 30, no. 9, pp. 1312–1313, 2014.

[57] A. M. Kozlov, A. J. Aberer, and A. Stamatakis, "ExaML version 3: a tool for phylogenomic analyses on supercomputers.," *Bioinformatics (Oxford, England)*, vol. 31, no. 15, pp. 2577–2579, 2015.

[58] N. D. Pattengale, M. Alipour, O. R. P. Bininda-Emonds, B. M. E. Moret, and A. Stamatakis, "How many bootstrap replicates are necessary?," *Journal of computational biology : a journal of computational molecular cell biology*, vol. 17, no. 3, pp. 337–354, 2010.

[59] Q.-M. Wang, W.-Q. Liu, G. Liti, S.-A. Wang, and F.-Y. Bai, "Surprisingly diverged populations of *Saccharomyces cerevisiae* in natural environments remote from human activity.," *Molecular ecology*, vol. 21, no. 22, pp. 5404–5417, 2012.

[60] S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, and P. C. Sham, "PLINK: a tool set for whole-genome association and population-based linkage analyses.," *American journal of human genetics*, vol. 81, no. 3, pp. 559–575, 2007.

[61] R. R Development Core Team, "R: A Language and Environment for Statistical Computing," 2011.

[62] X. Zheng, D. Levine, J. Shen, S. M. Gogarten, C. Laurie, and B. S. Weir, "A high-performance computing toolset for relatedness and principal component analysis of SNP data.," *Bioinformatics (Oxford, England)*, vol. 28, no. 24, pp. 3326–3328, 2012.

[63] S. Hutter, A. J. Vilella, and J. Rozas, "Genome-wide DNA polymorphism analyses using VariScan.," *BMC bioinformatics*, vol. 7, p. 409, 2006.

[64] B. Pfeifer, U. Wittelsburger, S. E. Ramos-Onsins, and M. J. Lercher, "PopGenome: an efficient Swiss army knife for population genomic analyses in R.," *Molecular biology and evolution*, vol. 31, no. 7, pp. 1929–1936, 2014.

[65] E. Eden, R. Navon, I. Steinfeld, D. Lipson, and Z. Yakhini, "GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists.," *BMC bioinformatics*, vol. 10, p. 48, 2009.

[66] V. Ranwez, S. Harispe, F. Delsuc, and E. J. P. Douzery, "MACSE: Multiple Alignment of Coding SEquences accounting for frameshifts and stop codons.," *PloS one*, vol. 6, no. 9, p. e22594, 2011.

[67] R. Lanfear, B. Calcott, S. Y. W. Ho, and S. Guindon, "Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses.," *Molecular biology and evolution*, vol. 29, no. 6, pp. 1695–1701, 2012.

[68] A. J. Drummond and M. A. Suchard, "Bayesian random local clocks, or one rate to rule them all.," *BMC biology*, vol. 8, p. 114, 2010.

[69] P. Lemey, A. Rambaut, A. J. Drummond, and M. A. Suchard, "Bayesian phylogeography finds its roots.," *PLoS computational biology*, vol. 5, no. 9, p. e1000520, 2009.

[70] A. Rambaut, A. J. Drummond, D. Xie, G. Baele, and M. A. Suchard, "Posterior summarization in bayesian phylogenetics using tracer 1.7," *Systematic Biology*, p. syy032, 2018.

[71] G. Baele and P. Lemey, "Bayesian evolutionary model testing in the phylogenomics era: matching model complexity with computational efficiency.," *Bioinformatics (Oxford, England)*, vol. 29, no. 16, pp. 1970–1979, 2013.

[72] G. Baele, P. Lemey, T. Bedford, A. Rambaut, M. A. Suchard, and A. V. Alekseyenko, "Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty.," *Molecular biology and evolution*, vol. 29, no. 9, pp. 2157–2167, 2012.

[73] J. C. Dittmar, R. J. Reid, and R. Rothstein, "ScreenMill: a freely available software suite for growth measurement, analysis and visualization of high-throughput screen data.," *BMC bioinformatics*, vol. 11, p. 353, 2010.

[74] J. Steensels, E. Meersman, T. Snoek, V. Saels, and K. J. Verstrepen, "Large-scale selection and breeding to generate industrial yeasts with superior aroma production.," *Applied and environmental microbiology*, vol. 80, no. 22, pp. 6965–6975, 2014.

[75] J.-L. Legras and F. Karst, "Optimisation of interdelta analysis for *Saccharomyces cerevisiae* strain characterisation.," *FEMS microbiology letters*, vol. 221, no. 2, pp. 249–255, 2003.

[76] A. R. Borneman, A. H. Forgan, I. S. Pretorius, and P. J. Chambers, "Comparative genome analysis of a *Saccharomyces cerevisiae* wine strain.," *FEMS yeast research*, vol. 8, no. 7, pp. 1185–95, 2008.

[77] M. Novo, F. Bigey, E. Beyne, V. Galeote, F. Gavory, S. Mallet, B. Cambon, J.-L. Legras, P. Wincker, S. Casaregola, and S. Dequin, "Eukaryote-to-eukaryote gene transfer events revealed by the genome sequence of the wine yeast *Saccharomyces cerevisiae* EC1118.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 38, pp. 16333–8, 2009.

7

Origins, Evolution, Domestication and Diversity of *Saccharomyces* Beer Yeasts

This chapter is adapted from:

Brigida Gallone, Stijn Mertens, Jonathan L Gordon, Steven Maere, Kevin J. Verstrepen and Jan Steensels. Origins, evolution, domestication and diversity of *Saccharomyces* beer yeasts. *Current Opinion in Biotechnology*, 2018, 49:148-155.

Author contributions:

Conceptualization: B.G., J.S., S.M., J.L.G., S.M., K.J.V; Data collection: B.G.; Formal analysis: B.G.; Data Visualisation: B.G.; Writing: B.G., J.S., S.M., J.L.G., S.M., K.J.V; Supervision: S.M., K.J.V.

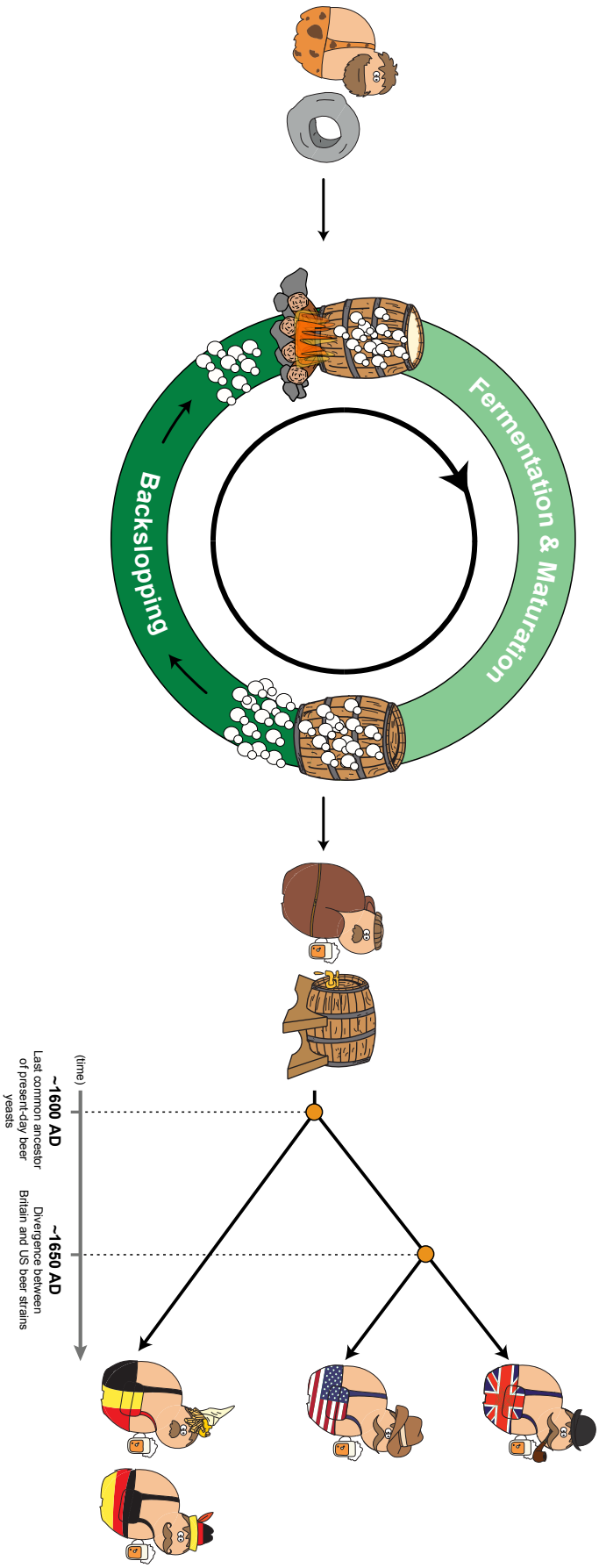


Figure 7.1: Graphical Abstract. Domestication of beer yeasts

7.1 Abstract

Yeasts have been used for food and beverage fermentations for thousands of years. Today, numerous different strains are available for each specific fermentation process. However, the nature and extent of the phenotypic and genetic diversity and specific adaptations to industrial niches have only begun to be elucidated recently. In *Saccharomyces*, domestication is most pronounced in beer strains, likely because they continuously live in their industrial niche, allowing only limited genetic admixture with wild stocks and minimal contact with natural environments. As a result, beer yeast genomes show complex patterns of domestication and divergence, making both ale (*S. cerevisiae*) and lager (*S. pastorianus*) producing strains ideal models to study domestication and, more generally, genetic mechanisms underlying swift adaptation to new niches.

7.2 Introduction

"Domestication" is a term that refers to artificial selection and breeding of wild species to obtain cultivated variants with enhanced desirable features that thrive in man-made environments, often at the cost of suboptimal fitness in natural settings. Several genotypic and phenotypic signatures of domestication have been described in crops, livestock and pets. These include genome decay, polyploidy, chromosomal rearrangements, gene amplifications and deletions, horizontal gene transfer and loss of genetic diversity due to bottlenecking [1, 2]. Interestingly, similar phenomena are also observed in various microbial species, both prokaryotic and eukaryotic, that are linked to human food production. Perhaps the most well studied model is the common brewer's and baker's yeast, *Saccharomyces cerevisiae*, which is the main driver in many industrial fermentations. However, studies focusing on the evolution of industrial *Saccharomyces* strains often use the terms 'adaptive evolution' or 'domestication' too freely. For example, both terms are commonly used to explain phenotypic divergence from wild ancestors, overlooking alternative explanations such as random genetic drift [3]. Only recently, more elaborate studies have reported clear genome-wide signatures of domestication as well as convergent evolution of industrially relevant traits in separate lineages. These observations provide conclusive evidence that industrial yeast diversity is not solely shaped by genetic drift caused by bottlenecking and small isolated populations, but also as a result of selection and niche adaptation. In wine yeasts for example, adaptive horizontal gene transfer events [4–6] and copy number variations [7–10] have been described that increase sugar and nitrogen metabolic activity, conferring competitive advantages during grape must fermentation and providing better tolerance to chemicals used in vineyards (e.g. copper sulphate) and in wine [11] (e.g. sulphite) (For a review see [4]). Interestingly however, the strongest genetic and phenotypic signatures of domestication are found in yeasts used for beer production. Several distinctive features

make traditional beer production an ideal setting for microbial domestication. Firstly, beer yeasts are harvested and re-used after the fermentation process to initiate the next fermentation batch, a process called 'backslopping'. This continuous growth in a very specific industrial niche has resulted in continuous selection imposed by the brewing environment. Secondly, beer is produced year-round, causing a near-complete isolation from wild isolates. In contrast, wine is seasonal and wine yeasts spend most of the year in and around the vineyards or in the guts of insects, where nutrient limitation can trigger sexual cycles and hybridization with wild yeasts [12]. Therefore, present-day beer yeasts can be considered the result of a centuries-long evolution experiment in a highly selective niche. In this review, we will highlight new insights into beer yeast evolution and domestication. We will discuss *S. cerevisiae* and *S. pastorianus*, both involved in production of specific beer types, which underwent a different route to domestication.

7.3 Domestication of *Saccharomyces cerevisiae* ale beer yeasts

Saccharomyces cerevisiae is the main microbial workhorse for the production of ale beers, which includes beer styles such as stouts, pale ales, doubles, triples and quadruples. As with all domesticated organisms, in *S. cerevisiae* the phenotypes of domesticated strains are a combination of enhanced selectable traits inherently present in *S. cerevisiae* (e.g. adaptation to sugar-rich, oxygen-limited environments and high tolerance to ethanol), and traits acquired during interaction with humans (e.g. efficient maltotriose utilization). In this review, we will expound on the latter aspect, and we refer to other review papers for the former [13, 14].

7.3.1 Phylogenetics and population structure

Many studies of *S. cerevisiae* population structure focused on wine, wild and/or clinical strains, neglecting the broad diversity of beer yeasts. However, two recent studies, sequencing more than 100 ale beer strains, have provided the first comprehensive insight into their evolution and diversification [15, 16]. Both studies found that the majority of beer yeasts are genetically distinct from known wild stocks, and cluster into two independent lineages **(Figures 7.2, 7.3 and 7.4, p. 137, 138 and 139, respectively)**. It has been estimated that the last common ancestor (LCA) of each lineage occurred around 1600-1700AD, well after the first reported beer production (3000-4000 BC), but before the discovery of microbes in the 19th century. Interestingly, the estimated period of occurrence of these LCAs coincides with the gradual switch from domestic brewing in private households, to more professional large-scale brewing, first in pubs and monasteries, and later in breweries. This suggests that true domestication of yeast occurred far more recently than the first leveraging of yeast for the production of fermented food and beverages, which likely happened several thousand years ago [17].

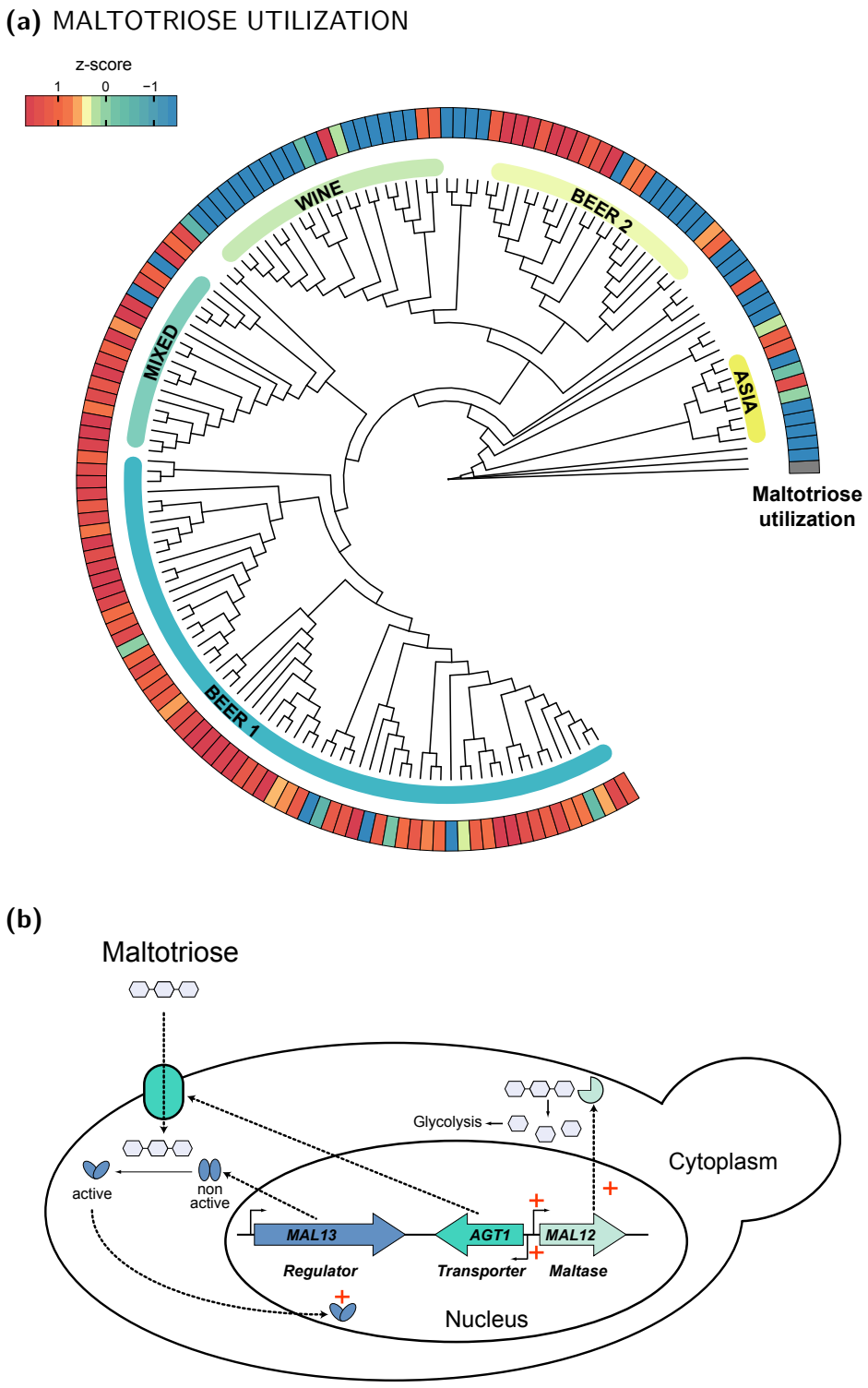


Figure 7.2: Genetic and phenotypic diversity of industrial *S. cerevisiae* strains: Maltotriose utilization. Beer yeasts show a significantly higher capacity to metabolize maltotriose, a prominent carbon source in beer medium. *AGT1*, a homolog of the sugar transporter MAL11, encodes for a permease with increased affinity to maltotriose compared to the wild type allele. *AGT1* is present in the Beer 1 and the Mixed clade, but absent and/or non-functional in the Wine and the Beer 2 clade. **a:** A cladogram depicting genetic relationship and a heatmap (surrounding the cladogram) depicting phenotypic behaviour of each strain is given. **b:** A schematic depiction of the phenotypic traits.

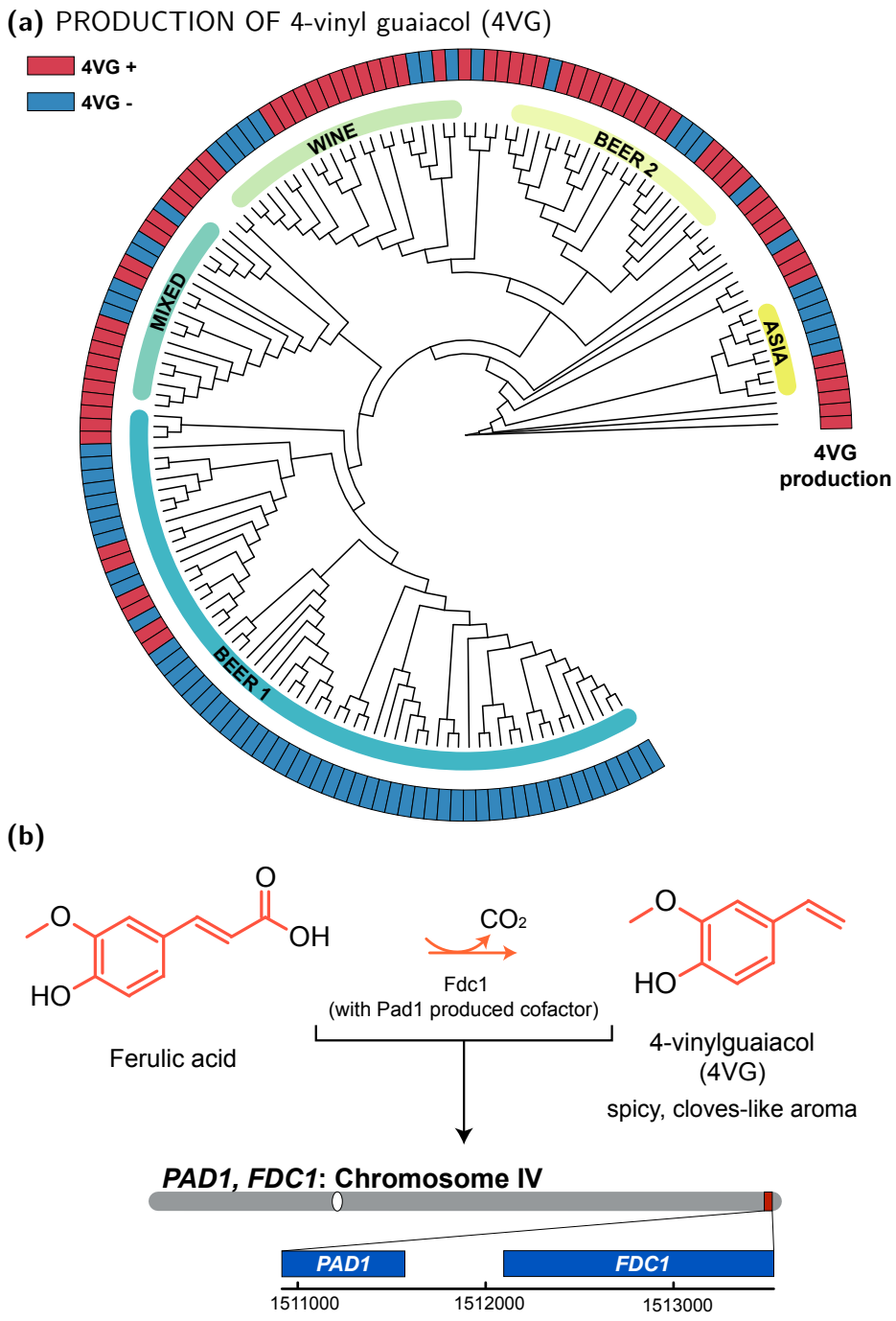


Figure 7.3: Genetic and phenotypic diversity of industrial *S. cerevisiae* strains: Production of 4-vinyl guaiacol (4VG). 4VG is a spicy clove-like aroma compound, generally undesired in most fermented beverages. It is produced by yeast by the decarboxylation of ferulic acid, an abundant phenolic compound found in many plant cell walls, by the *FDC1*-encoded decarboxylase. This decarboxylase requires a flavin-derived cofactor produced by PAD1. FDC1 and PAD1 are clustered in the subtelomeric region of the right arm of chromosome IV. Many industrial yeasts, most notably the yeast in the Beer 1 clade, acquired loss-of-function mutations in *PAD1* and/or *FDC1*, resulting in a loss of the ability to produce 4VG [15, 16]. **a:** A cladogram depicting genetic relationship and a heatmap (surrounding the cladogram) depicting phenotypic behaviour of each strain is given. **b:** A schematic depiction of the phenotypic traits.

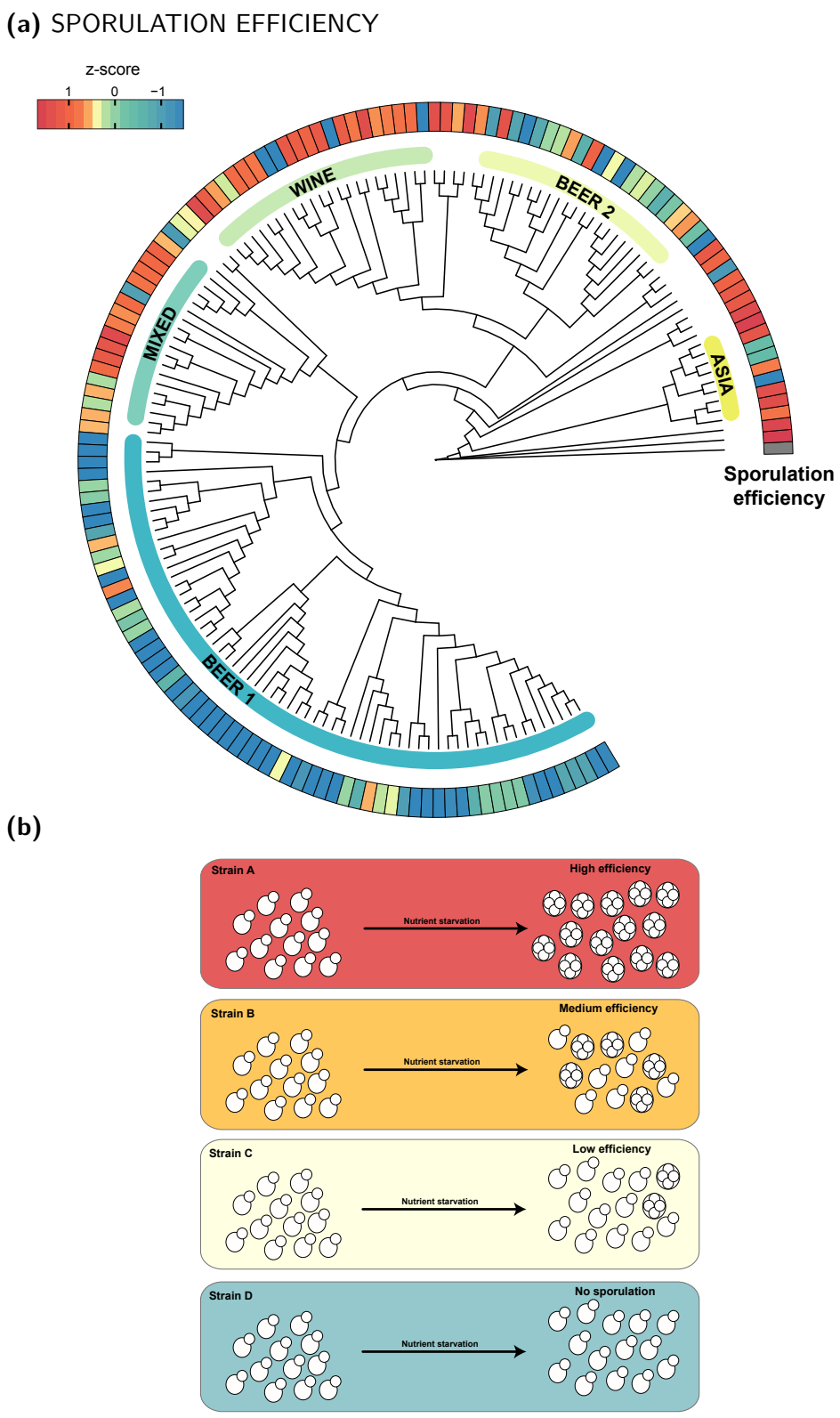


Figure 7.4: Genetic and phenotypic diversity of industrial *S. cerevisiae* strains: Sporulation efficiency. Yeasts from the main beer lineages are generally obligate asexual or show low sporulation efficiency compared to strains from other clades. **a:** A cladogram depicting genetic relationship and a heatmap (surrounding the cladogram) depicting phenotypic behaviour of each strain is given. **b:** A schematic depiction of the phenotypic traits.

7.3.2 Genome structure

Variation in genome structure, including changes in ploidy and large segmental duplications or copy number variations (CNVs), have repeatedly been found in association with adaptation to specific niches in experimentally evolved microbes [18, 19]. Perhaps not surprisingly, similar chromosomal changes are also a recurrent theme in domestication of higher organisms, especially in crop species, for which polyploidization promoted the proper genetic circumstances to domestication [20]. Similarly, *S. cerevisiae* beer yeasts show large-scale genome structure variations. While most wild *S. cerevisiae* strains are clean diploids with very few large segmental duplications, the ploidy of the vast majority of beer strains exceeds 2n, with most close to 4n **(Figure 7.5a, p. 141)**. However, most ale yeasts also show aneuploidies, and are almost never perfectly diploid or tetraploid **(Figure 7.5b, p. 141)**. Previous studies have shown that aneuploidies and polyploidies can provide an adaptive advantage under selection [18], but that they are often transient and are maintained until a more costeffective adaptive strategy has evolved [21]. It has also been shown that small structural genome variations (e.g. duplication, deletion, recombination, gene conversion and rearrangement) are frequently located in telomeric and subtelomeric regions, which are known hotspots for evolution. These regions are functionally enriched for genes involved in nitrogen and carbon metabolism, ion transport and flocculation, and likely play a role in niche adaptation [15, 22, 23].

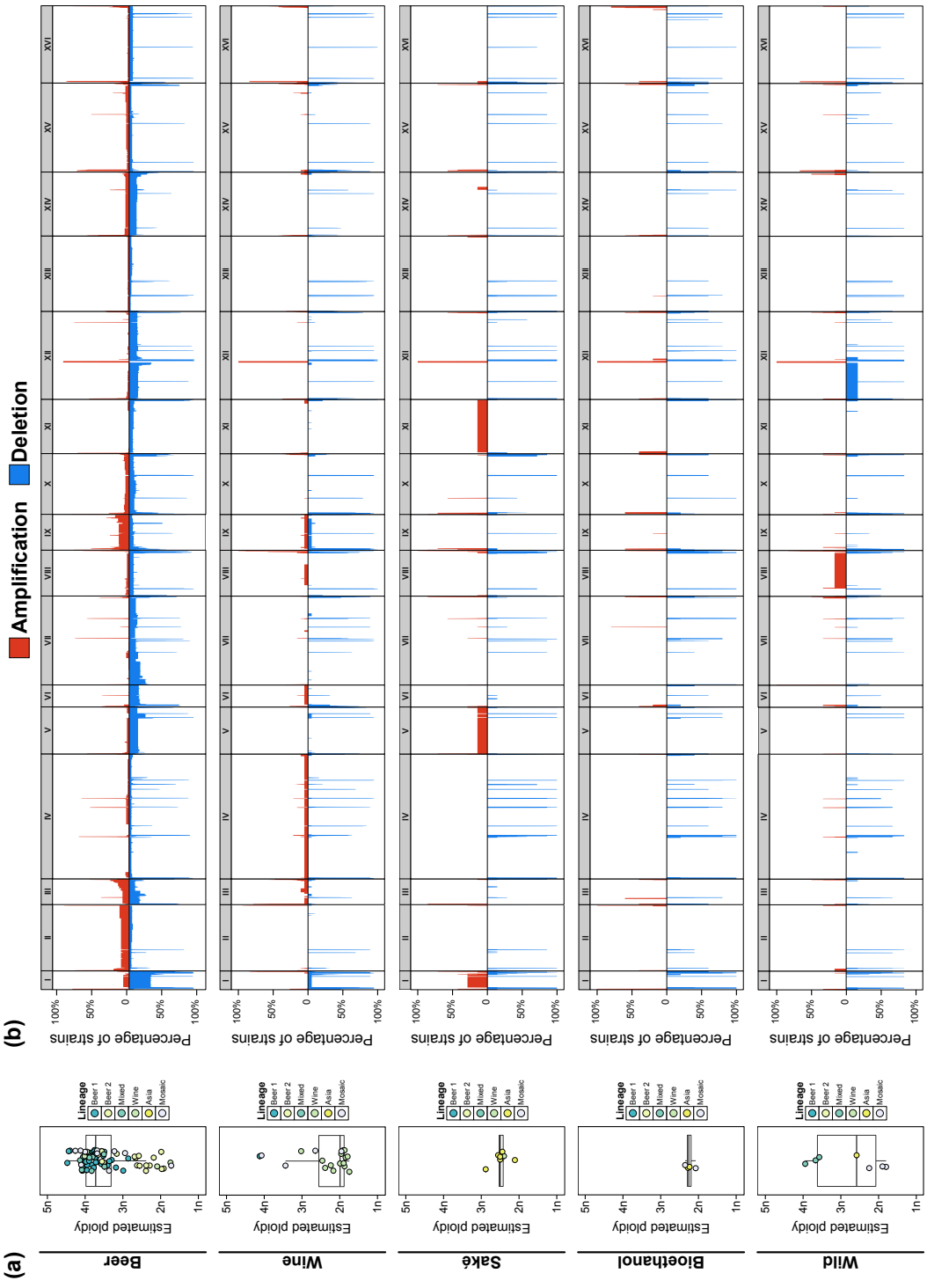


Figure 7.5: Ploidy variation of industrial yeast genomes. a: Box plots representing the estimated ploidy of *S. cerevisiae* strains grouped by niche of isolation (beer, wine, saké , bioethanol, wild); colours indicate phylogenetic relationship (lineage); strains with admixed ancestries, not belonging to any specific clean lineage, are indicated as 'mosaic'. Black lines represent median value and box edges are the 25th and 75th percentiles. Data obtained from [15]. **b:** Genome-wide visualization of CNV profiles compared to the reference strain S228C, aggregated across all strains originating from different niches; colours indicate amplification (red) and deletion (blue). Data obtained from [15].

7.3.3 Brewing phenotypes

The most obvious sign of adaptation to a specific industrial niche is arguably the accentuation of traits desirable for humans that are a burden for the organism in a natural setting. Closer examination of the genetic underpinnings of specific traits have provided strong evidence that human selection indeed underlies certain industrially relevant traits in beer yeasts. A prime example of a domestication trait in beer yeast is their ability to ferment maltotriose, an important carbon source in beer wort, but not generally found in high concentrations in natural yeast environments. Efficient metabolism of maltotriose imposes a selective advantage in brewing environments where it is present at high concentrations because it opens the door to a previously underor poorly utilized energy source. This trait has evolved independently and through different genetic pathways in the two main beer lineages, suggesting strong selection pressure [15] **(Figure 7.2, p. 137)**. In one of the beer groups a homolog of the maltose transporter (called *AGT1*) with an increased affinity for maltotriose is present [24]. Interestingly, in the second beer group, the *AGT1* allele is non-functional, but the majority of isolates are able to efficiently ferment maltotriose, suggesting the presence of a distinct, yet unknown mechanism for the maltotriose uptake in this lineage. Another well-documented domestication trait is the selection against production of 4-vinyl guaiacol (4VG), an unpleasant aromaactive compound that is derived from ferulic acid, a cell wall component of barley. Yeast requires two genes for decarboxylation of ferulic acid to 4VG: *PAD1* and *FDC1*. Various independent nonsense mutations in these genes have been found in many industrial (and especially beer) yeasts, but not in biofuel or non-industrial isolates, suggesting that the selection of 4VG-free fermentations has favoured the spread of domesticated beer yeasts unable to produce this specific off-flavour **(Figure 7.3, p. 138)** [15, 16].

Apart from human-driven selection for specific traits, domestication is often also accompanied by relaxation of selective constraints on traits that are not advantageous or too costly in man-made environments. This relaxed selection can result in gene loss or pseudogenisation of genes that are no longer required for survival, a process dubbed 'genome decay' [1]. In ale yeasts this is reflected by their inability to survive environmental and nutrient stress conditions not encountered during continuous growth in the nutrient-rich wort medium [15]. One likely example of this phenomenon is represented by the lack of a functional sexual cycle in most ale beer isolates [15] **(Figure 7.4, p. 139)**, a trait with a crucial role in speeding adaptation under new, harsh conditions, but with limited added value in benign environments [25].

7.4 Domestication of *Saccharomyces pastorianus* lager beer yeasts

With over 90% market share, lager beer is by far the most popular beer style globally. The production process differs in many ways form ale beer brewing, but arguably the most profound distinction is that a different yeast species is used. Lager is brewed using *S. pastorianus*, which

originates from the interspecific hybridization between *S. cerevisiae* and *S. eubayanus*, a closely related species that is not typically associated with industrial fermentations [26–28].

7.4.1 Phylogenetics and population structure

Lager strains are divided in two main distinct lineages, most commonly referred to as "Saaz" (Type 1) and "Frohberg" (Type 2). Whereas the existence of these two lineages suggests some degree of convergent domestication, the precise ancestry and evolution of the two *S. pastorianus* lineages is still controversial (for a review see [29]). Three main hypotheses for the *S. pastorianus* origin have been proposed, and different analyses have provided support for each. The most widespread hypothesis involves two completely independent hybridization events, each involving a different domesticated ale-type *S. cerevisiae* and a different wild *S. eubayanus* strain **(Figure 7.6a, p. 144)**. This hypothesis is supported by phylogenetic analysis, where the relative branch lengths for the *S. cerevisiae* and the *S. eubayanus* subgenomes are significantly different between the groups [27], as well as by the pattern of loss or retention of subtelomeric regions in the cerevisiae part of the lager yeast genome, which are very different in Saaz and Frohberg yeasts [30]. However, other studies have identified several *S. eubayanus*/*S. cerevisiae* translocations that share identical breakpoints within the subgenomes of both Saaz and Frohberg lineages [31, 32] **(Figure 7.6b, p. 144)**. Although it has been argued that identical breakpoints could have resulted from i) independent events at recombination hotspots or fragile sites in *Saccharomyces* chromosomes, or ii) events that occurred in one of the parental strains prior to the hybridization [27], they might also indicate a shared hybridization event prior the divergence in the distinct lager lineages. A further hypothesis suggests a combination of the two scenarios; a single hybridization event between a haploid *S. cerevisiae* and a diploid *S. eubayanus* that led to an ancestral Saaz-like hybrid, followed by a second hybridization with a distinct haploid *S. cerevisiae* isolate that led to a Frohberg-like ancestral hybrid [33] **(Figure 7.6c, p. 144)**.

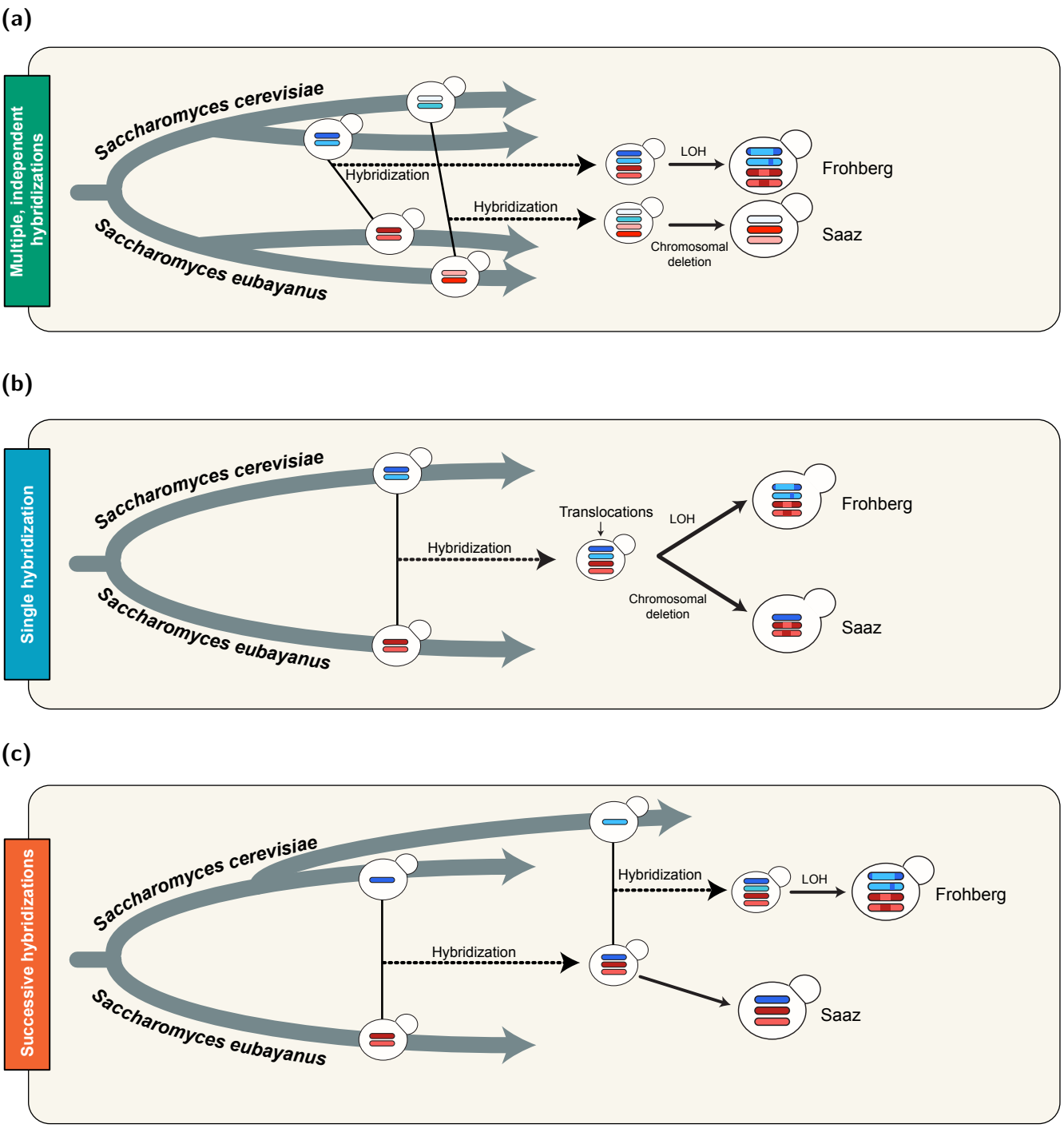


Figure 7.6: Current models for the origin of Frohberg and Saaz lineages. a: Frohberg and Saaz groups originated from at least two independent hybridization events between distinct diploid *S. cerevisiae* and diploid *S. eubayanus* parental strains. **b:** Frohberg and Saaz groups originated from a single hybridization event between a diploid *S. cerevisiae* and a diploid *S. eubayanus*. Translocations occurred in the ancestral hybrid prior to the divergence of the Saaz and Frohberg lineages and are shared between the two groups. After hybridization, the Frohberg lineage experienced loss of variation between intra-homologous chromosomes in the *S. cerevisiae* subgenome via loss of heterozygosity [33] and the Saaz lineage lost roughly half of the *S. cerevisiae* derived chromosomes. **c:** Frohberg and Saaz groups originated from at least one shared hybridization event between a haploid *S. cerevisiae* and a diploid *S. eubayanus*. The triploid ancestral hybrid further diverged into the Saaz lineage, and the Frohberg lineage arose by another hybridization event with a distinct haploid *S. cerevisiae* [33].

7.4.2 Genome structure interaction between subgenomes

Experiments where new interspecific hybrids are generated in the laboratory have shown that the genomes of newly formed hybrids tend to be highly unstable and undergo progressive genomic evolution until a more stable karyotype is reached. Hybrid genomic instability is probably due to an interplay between the relaxed selection on regions of genetic redundancy introduced by the hybridization event, selection acting on gene dosage balance, the potential co-adaptation of genes from the same parental genome and selection imposed by their growth environment [34–37]. Genomic changes include copy number variations [38] and partial or total chromosome loss [38], but also rearrangements between both subgenomes, resulting in chimaeric chromosomes [32]. Many of these mechanisms have been shown to allow adaptation in experimentally evolving interspecific hybrids. For example, loss-of-heterozygosity in newly developed interspecific hybrids is shown to be a reproducible adaptive strategy in low-nutrient environments, highlighting that hybrid genome resolution can be driven by positive selection acting on existing heterozygosity [39]. Chimaerism can be adaptive on the level of individual genes, for example, new interspecific *Saccharomyces* hybrids evolved in stringent nitrogen limitation recurrently evolve chimeric *MEP1* alleles [40]. In lager yeasts, more than 20 chimaeric genes have been identified so far, including chimaeric variants of *ALD2* and *TDH2*, both involved in ethanol metabolism [32].

7.4.3 Brewing phenotypes

The dominance of *S. pastorianus* in the lager brewing industry suggests a strong selective advantage of the interspecific hybrid over their respective parental species. It has been argued that some parts of the *S. eubayanus* genome confer enhanced cold-tolerance, while the *S. cerevisiae* subgenome holds the advantage of other brewing adaptations, such as efficient fermentation, including the use of maltotriose. However, maltotriose transporters from *S. eubayanus* and not *S. cerevisiae* enable maltotriose utilization in some Saaz-type lager yeasts [41]. Conversely, certain *S. cerevisiae* strains show adaptation to cold environments [42], which implies that the cold tolerance does not necessarily originate from *S. eubayanus*, and that the origin of lager yeasts may have a different foundation. Recent papers suggest that increased fitness of interspecific yeast hybrids can also be due to genetic incompatibilities that perturb safeguard mechanisms that would normally limit growth in the parental strains, leading to hybrids that divide more (and thus have a higher fitness) in stressful environments, such as beer wort at low temperatures [43]. However, more research is required to untangle the specific roles of lager yeast subgenomes in the brewing environment.

7.5 Conclusions and outlook

Recent studies have demonstrated that beer yeasts have been domesticated by enduring growth in man-made fermentation environments. The strong selective pressure imposed over many genera-

tions has contributed to the emergence of desirable phenotypes, but has also dramatically affected the genomic structure and genome stability of the domesticates. Interestingly, continuous 'back-slopping' is not common practice anymore in most of today's commercial breweries. Instead, brewers dispose of their yeast culture after a few consecutive fermentations to start a new brew with a frozen stock culture. This continual reversion to the same yeast stock ensures consistency of their product, but prevents further evolution of the beer yeast. Therefore evolution and domestication of beer yeasts within breweries may have practically halted. However, the process has now moved to specialized labs, where the expanding experimental toolbox and the wealth of "omics" data available for *Saccharomyces* yeasts opens new avenues to generate novel and superior industrial variants. Specifically, experimental evolution, similar to the process in traditional brewing, can be used in conjunction with techniques like crossing, marker-assisted breeding and mutagenesis to effectively generate and test new phenotypic variants and combinations [44, 45]. Thus, several centuries after the dawn of beer yeast domestication by commercial-scale brewing, a second revolution, sparked by biotechnology, is now driving a new era of beer yeast evolution and domestication.

7.6 References

[1] C. A. Driscoll, D. W. Macdonald, and S. J. O'Brien, "From wild animals to domestic pets, an evolutionary view of domestication.," *Proceedings of the National Academy of Sciences of the United States of America*, no. Supplement 1, pp. 9971–8, 2009.

[2] M. D. Purugganan and D. Q. Fuller, "The nature of selection during plant domestication," *Nature*, vol. 457, no. 7231, pp. 843–848, 2009.

[3] E. Zörgö, A. Gjuvsland, F. A. Cubillos, E. J. Louis, G. Liti, A. Blomberg, S. W. Omholt, and J. Warringer, "Life history shapes trait heredity by accumulation of loss-of-function alleles in yeast.," *Molecular biology and evolution*, vol. 29, no. 7, pp. 1781–9, 2012.

[4] S. Marsit, A. Mena, F. Bigey, F.-X. Sauvage, A. Couloux, J. Guy, J.-L. Legras, E. Barrio, S. Dequin, and V. Galeote, "Evolutionary Advantage Conferred by an Eukaryote-to-Eukaryote Gene Transfer Event in Wine Yeasts," *Molecular Biology and Evolution*, vol. 32, no. 7, pp. 1695–1707, 2015.

[5] P. Almeida, R. Barbosa, P. Zalar, Y. Imanishi, K. Shimizu, B. Turchetti, J.-L. Legras, M. Serra, S. Dequin, A. Couloux, J. Guy, D. Bensasson, P. Gonçalves, and J. P. Sampaio, "A population genomics insight into the Mediterranean origins of wine yeast domestication.," *Molecular ecology*, vol. 24, no. 21, pp. 5412–27, 2015.

[6] S. Marsit, I. Sanchez, V. Galeote, and S. Dequin, "Horizontally acquired oligopeptide transporters favour adaptation of *Saccharomyces cerevisiae* wine yeast to oenological environment," *Environmental Microbiology*, vol. 18, no. 4, pp. 1148–1161, 2016.

[7] S. Marsit and S. Dequin, "Diversity and adaptive evolution of *Saccharomyces* wine yeast: a review," *FEMS Yeast Research*, vol. 15, no. 7, p. fov067, 2015.

[8] B. Dunn, C. Richter, D. J. Kvitek, T. Pugh, and G. Sherlock, "Analysis of the *Saccharomyces cerevisiae* pan-genome reveals a pool of copy number variants distributed in diverse yeast strains from differing industrial environments.," *Genome research*, vol. 22, no. 5, pp. 908–24, 2012.

[9] J. Steenwyk and A. Rokas, "Extensive Copy Number Variation in Fermentation-Related Genes Among *Saccharomyces cerevisiae* Wine Strains," *G3 (Bethesda)*, vol. 7, no. 5, pp. 1475–1485, 2017.

[10] P. K. Strope, D. a. Skelly, S. G. Kozmin, G. Mahadevan, E. a. Stone, P. M. Magwene, F. S. Dietrich, J. H. Mccusker, N. Carolina, B. Sciences, and N. Carolina, "The 100-genomes strains, an *S. cerevisiae* resource that illuminates its natural phenotypic and genotypic variation and emergence as an opportunistic pathogen," *Genome research*, pp. 1–13, 2015.

[11] P. Almeida, R. Barbosa, D. Bensasson, P. Gonçalves, and J. P. Sampaio, "Adaptive divergence in wine yeasts and their wild relatives suggests a prominent role for introgressions and rapid evolution at noncoding sites," *Molecular Ecology*, vol. 26, no. 7, pp. 2167–2182, 2017.

[12] I. Stefanini, L. Dapporto, L. Berná, M. Polsinelli, S. Turillazzi, and D. Cavalieri, "Social wasps are a *Saccharomyces* mating nest.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 113, no. 8, pp. 2247–51, 2016.

[13] J. Piskur, E. Rozpedowska, S. Polakova, A. Merico, and C. Compagno, "How did *Saccharomyces* evolve to become a good brewer?," *Trends in Genetics*, vol. 22, no. 4, pp. 183–186, 2006.

[14] J. Steensels and K. J. Verstrepen, "Taming Wild Yeast: Potential of Conventional and Nonconventional Yeasts in Industrial Fermentations," *Annual Review of Microbiology*, vol. 68, no. 1, pp. 61–80, 2014.

[15] B. Gallone, J. Steensels, T. Prahl, L. Soriaga, V. Saels, B. Herrera-Malaver, A. Merlevede, M. Roncoroni, K. Voordeckers, L. Miraglia, C. Teiling, B. Steffy, M. Taylor, A. Schwartz, T. Richardson, C. White, G. Baele, S. Maere, and K. Verstrepen, "Domestication and Divergence of *Saccharomyces cerevisiae* Beer Yeasts," *Cell*, vol. 166, no. 6, pp. 1397–1410.e16, 2016.

[16] M. Gonçalves, A. Pontes, P. Almeida, R. Barbosa, M. Serra, D. Libkind, M. Hutzler, P. Gonçalves, J. Sampaio, W. Heijne, and et Al., "Distinct Domestication Trajectories in Top-Fermenting Beer Yeasts and Wine Yeasts," *Current Biology*, vol. 26, no. 20, pp. 2750–2761, 2016.

[17] I. S. Hornsey, *History of Beer and Brewing*. Cambridge: Cambridge: Royal Society of Chemistry., 2003.

[18] A. M. Selmecki, Y. E. Maruvka, P. A. Richmond, M. Guillet, N. Shoresh, A. L. Sorenson, S. De, R. Kishony, F. Michor, R. Dowell, and D. Pellman, "Polyploidy can drive rapid adaptation in yeast," *Nature*, vol. 519, no. 7543, pp. 349–352, 2015.

[19] K. Voordeckers, J. Kominek, A. Das, A. Espinosa-Cantú, D. De Maeyer, A. Arslan, M. Van Pee, E. van der Zande, W. Meert, Y. Yang, B. Zhu, K. Marchal, A. DeLuna, V. Van Noort, R. Jelier, and K. J. Verstrepen, "Adaptation to High Ethanol Reveals Complex Evolutionary Pathways," *PLOS Genetics*, vol. 11, no. 11, p. e1005635, 2015.

[20] A. Salman-Minkov, N. Sabath, I. Mayrose, N. C. Ellstrand, and L. H. Rieseberg, "Whole-genome duplication as a key factor in crop domestication," *Nature Plants*, vol. 2, no. 8, p. 16115, 2016.

[21] A. H. Yona, Y. S. Manor, R. H. Herbst, G. H. Romano, A. Mitchell, M. Kupiec, Y. Pilpel, and O. Dahan, "Chromosomal duplication is a transient evolutionary solution to stress," *Proceedings of the National Academy of Sciences*, vol. 109, no. 51, pp. 21010–21015, 2012.

[22] C. A. Brown, A. W. Murray, and K. J. Verstrepen, "Rapid expansion and functional divergence of subtelomeric gene families in yeasts.," *Current biology : CB*, vol. 20, no. 10, pp. 895–903, 2010.

[23] J.-X. Yue, J. Li, L. Aigrain, J. Hallin, K. Persson, K. Oliver, A. Bergström, P. Coupland, J. Warringer, M. C. Lagomarsino, G. Fischer, R. Durbin, and G. Liti, "Contrasting evolutionary genome dynamics between domesticated and wild yeasts," *Nature Genetics*, 2017.

[24] S. L. Alves, R. A. Herberts, C. Hollatz, D. Trichez, L. C. Miletti, P. S. de Araujo, and B. U. Stambuk, "Molecular Analysis of Maltotriose Active Transport and Fermentation by *Saccharomyces cerevisiae* Reveals a Determinant Role for the AGT1 Permease," *Applied and Environmental Microbiology*, vol. 74, no. 5, pp. 1494–1501, 2008.

[25] M. R. Goddard, H. C. J. Godfray, and A. Burt, "Sex increases the efficacy of natural selection in experimental yeast populations," *Nature*, vol. 434, no. 7033, pp. 636–640, 2005.

[26] D. Libkind, C. T. Hittinger, E. Valério, C. Gonçalves, J. Dover, M. Johnston, P. Gonçalves, and J. P. Sampaio, "Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 35, pp. 14539–44, 2011.

[27] E. Baker, B. Wang, N. Bellora, D. Peris, A. B. Hulfachor, J. A. Koshalek, M. Adams, D. Libkind, and C. T. Hittinger, "The Genome Sequence of *Saccharomyces eubayanus* and the Domestication of Lager-Brewing Yeasts," *Molecular Biology and Evolution*, vol. 32, no. 11, pp. 2818–2831, 2015.

[28] D. Peris, Q. K. Langdon, R. V. Moriarty, K. Sylvester, M. Bontrager, G. Charron, J.-B. Leducq, C. R. Landry, D. Libkind, and C. T. Hittinger, "Complex Ancestries of Lager-Brewing Hybrids Were Shaped by Standing Variation in the Wild Yeast *Saccharomyces eubayanus*," *PLOS Genetics*, vol. 12, no. 7, p. e1006155, 2016.

[29] C. Monerawela and U. Bond, "Brewing up a storm: The genomes of lager yeasts and how they evolved," *Biotechnology Advances*, vol. 35, no. 4, pp. 512–519, 2017.

[30] C. Monerawela, T. C. James, K. H. Wolfe, and U. Bond, "Loss of lager specific genes and subtelomeric regions define two different *Saccharomyces cerevisiae* lineages for *Saccharomyces pastorianus* Group I and II strains," *FEMS Yeast Research*, vol. 15, no. 2, 2015.

[31] A. Walther, A. Hesselbart, and J. Wendland, "Genome sequence of *Saccharomyces carlsbergensis*, the world's first pure culture lager yeast.," *G3 (Bethesda, Md.)*, vol. 4, no. 5, pp. 783–93, 2014.

[32] S. K. Hewitt, I. Donaldson, S. C. Lovell, D. Delneri, and U. Kingdom, "Sequencing and characterisation of rearrangements in three *S. pastorianus* strains reveals the presence of chimeric genes and gives evidence of breakpoint reuse," *PLOS ONE*, pp. 1–71, 2013.

[33] M. Okuno, R. Kajitani, R. Ryusui, H. Morimoto, Y. Kodama, and T. Itoh, "Next-generation sequencing analysis of lager brewing yeast strains reveals the evolutionary history of interspecies hybridization," *DNA Research*, vol. 23, no. 1, p. dsv037, 2016.

[34] G. C. Conant, J. A. Birchler, and J. C. Pires, "Dosage, duplication, and diploidization: clarifying the interplay of multiple models for duplicate gene evolution over time," *Current Opinion in Plant Biology*, vol. 19, pp. 91–98, 2014.

[35] L. Morales and B. Dujon, "Evolutionary role of interspecies hybridization and genetic exchanges in yeasts.," *Microbiology and molecular biology reviews : MMBR*, vol. 76, no. 4, pp. 721–39, 2012.

[36] W. Albertin and P. Marullo, "Polyploidy in fungi: evolution after whole-genome duplication.," *Proceedings. Biological sciences*, vol. 279, no. 1738, pp. 2497–509, 2012.

[37] M. R. Sanchez, A. W. Miller, I. Liachko, A. B. Sunshine, B. Lynch, M. Huang, E. Alcantara, C. G. DeSevo, D. A. Pai, C. M. Tucker, M. L. Hoang, and M. J. Dunham, "Differential paralog divergence modulates genome evolution across yeast species.," *PLoS genetics*, vol. 13, no. 2, p. e1006585, 2017.

[38] M. van den Broek, I. Bolat, J. F. Nijkamp, E. Ramos, M. A. H. Luttik, F. Koopman, J. M. Geertman, D. de Ridder, J. T. Pronk, and J.-M. Daran, "Chromosomal Copy Number Variation in *Saccharomyces pastorianus* Is Evidence for Extensive Genome Dynamics in Industrial Lager Brewing Strains," *Applied and Environmental Microbiology*, vol. 81, no. 18, pp. 6253–6267, 2015.

[39] C. S. Smukowski Heil, C. G. DeSevo, D. A. Pai, C. M. Tucker, M. L. Hoang, and M. J. Dunham, "Loss of Heterozygosity Drives Adaptation in Hybrid Yeast," *Molecular Biology and Evolution*, 2017.

[40] B. Dunn, T. Paulish, A. Stanbery, J. Piotrowski, G. Koniges, E. Kroll, E. J. Louis, G. Liti, G. Sherlock, and F. Rosenzweig, "Recurrent rearrangement during adaptive evolution in an interspecific yeast hybrid suggests a model for rapid introgression.," *PLoS genetics*, vol. 9, no. 3, p. e1003366, 2013.

[41] F. Magalhães, V. Vidgren, L. Ruohonen, and B. Gibson, "Maltose and maltotriose utilisation by group I strains of the hybrid lager yeast *Saccharomyces pastorianus*," *FEMS Yeast Research*, vol. 16, no. 5, p. fow053, 2016.

[42] E. García-Ríos, M. Morard, L. Parts, G. Liti, and J. M. Guillamón, "The genetic architecture of low-temperature adaptation in the wine yeast *Saccharomyces cerevisiae*," *BMC Genomics*, vol. 18, no. 1, p. 159, 2017.

[43] R. H. Herbst, D. Bar-Zvi, S. Reikhav, I. Soifer, M. Breker, G. Jona, E. Shimoni, M. Schuldiner, A. A. Levy, and N. Barkai, "Heterosis as a consequence of regulatory incompatibility," *BMC Biology*, vol. 15, no. 1, p. 38, 2017.

[44] J. Steensels, T. Snoek, E. Meersman, M. P. Nicolino, K. Voordeckers, and K. J. Verstrepen, "Improving industrial yeast strains: exploiting natural and artificial diversity.," *FEMS microbiology reviews*, vol. 38, no. 5, pp. 947–95, 2014.

[45] S. Mertens, J. Steensels, V. Saels, G. De Rouck, G. Aerts, and K. J. Verstrepen, "A Large Set of Newly Created Interspecific *Saccharomyces* Hybrids Increases Aromatic Diversity in Lager Beers," *Applied and Environmental Microbiology*, vol. 81, no. 23, pp. 8202–8214, 2015.

8

Interspecific Hybridization Facilitates Niche Adaptation in Beer Yeasts

This chapter is adapted from a manuscript in preparation for Nature Ecology and Evolution:

Brigida Gallone, Jan Steensels, Stijn Mertens, Maria C. Dzialo, Jonathan L. Gordon, Ruben Wauters, Florian A. Thesseling, Francesca Bellinazzo, Veerle Saels, Beatriz Herrera-Malaver, Mathias Hutzler, Franz Meussdoerffer, Philippe Malcorps, Ben Souffriau, Luk Daenen, Guy Baele, Steven Maere and Kevin J. Verstrepen. Interspecific Hybridization Facilitates Niche Adaptation in Beer Yeasts. 2018, *submitted.*

Author contributions:

Conceptualization: B.G., J.S., S.M., K.J.V; Data collection: B.G., J.S.; Formal analysis: B.G., J.S., G.B.; Investigation: B.G., J.S., S.M., J.L.G., R.W., F.A.T., F.B., V.S., B.H., G.B.; Data visualisation: B.G.; Resources: M.H., F.M., P.M., B.S., L.D., K.J.V; Writing: B.G., J.S., M.C.D.,S.M., K.J.V; Supervision: S.M., K.J.V.

8.1 Abstract

Hybridization of different species is a rare event that often leads to inviable or infertile offspring, yet examples of interspecific hybrids have been reported in all kingdoms of life. Such events presumably yield novel phenotypic features that facilitate colonization of ecological niches not accessible to the parental species. However, the ecological circumstances, molecular details, and evolutionary role of interspecific hybridization remain largely unknown. In this study, we sequenced and phenotyped a rich biodiversity of hybrid yeasts isolated from breweries to uncover the consequences of interspecific hybridization in yeast adaptation and domestication in the context of industrial beer production. Our analyses demonstrate that hybrids between various *Saccharomyces* species originated in unique industrial environments. Large-scale phenotyping suggests that the newly formed hybrids colonized various brewery niches by combining key traits of each parental species. Furthermore, post-hybridization evolution within each hybrid lineage often led to sub-specialization and adaptation to highly specific beer styles. This process was accompanied by extensive structural rearrangements and chimerization between the subgenomes. Our results demonstrate that yeast interspecific hybridizations allowed for swift adaptation to industrial beer brewing environments by creating novel combinations of phenotypes primed for surviving in these new niches.

8.2 Introduction

A canonical view of Darwinian evolution asserts that new species arise during extended periods of gradual evolution and selection. However, amendments to Darwin's original theory have been proposed to incorporate other mechanisms for evolution and speciation, including the emergence of new species through interspecific hybridisation [1–3]. Hybridisation provides a way to rapidly combine distinct phenotypic features from established populations that converge in an unfamiliar ecological niche. Sometimes, a hybrid's unique combination of phenotypes can enable it to thrive in the new environment and outcompete its parental species [4–6]. Examples of rapid niche specialisation via interspecific hybridisation are found across all kingdoms of life. A new hybrid lineage of Darwin's finches with competitive beak morphology emerged in only three generations [7]. Hybridisation of the sunflowers *Helianthus annus* and *Heliantus petiolaris* gave rise to three novel species capable of colonizing previously untapped environments. The resulting species *Helianthus anomalus*, *Helianthus deserticola*, and *Helianthus paradoxus* thrive in sand dunes, basin deserts, and desert salt marshes, respectively, extreme environments that are inaccessible to the parental species [8–10]. The emerging opportunistic fungal pathogen *Candida metapsilosis* arose from a single hybridisation event between two non-pathogenic parental species [11]. The *Saccharomyces*

species complex provides numerous examples of interspecific hybridisation. Despite the high sequence divergence between the species within this genus (up to 20% [12]), prezygotic barriers are generally weak and species can successfully crossbreed [13, 14]. The resulting hybrids are typically infertile, yet viable, and can reproduce asexually by budding [14–17]. Wild *Saccharomyces* interspecific hybrids are occasionally encountered [18, 19] but the best-known example, *Saccharomyces pastorianus*, was isolated from an industrial environment [20–23]. Humans have historically utilised the capacity of *Saccharomyces* to convert high concentrations of sugars into ethanol and carbon dioxide to produce a wide variety of fermented products [24–26]. This long-term domestication process has resulted in hundreds of different industrial strains, each with characteristics suitable for specific industrial processes [27–39]. Strains used in the production of lager (pilsner) beer generally belong to the species *S. pastorianus*, a hybrid of *S. cerevisiae* and *Saccharomyces eubayanus* [20–23, 40–43]. Two archetypical S. pastorianus sublineages exist, dubbed 'Frohberg' and 'Saaz', each with its own distinct fermentation properties [44]. Several other industrial interspecific hybrids have been isolated, including *S. cerevisiae* x *Saccharomyces uvarum* x *S. eubayanus* triple hybrids in wine and cider, and *S. cerevisiae* x *Saccharomyces kudriavzevii* hybrids in ale beers and wine fermentation vessels [45–51]. A comprehensive analysis of the prevalence, molecular details, and the ecological and evolutionary context of interspecific hybridisation in yeasts is lacking. Here, we report how whole-genome sequencing of more than 200 industrial yeasts revealed that a surprisingly large fraction (25%) proved to be interspecific hybrids derived from four parental species: *S. cerevisiae*, *S. kudriavzevii*, *S. eubayanus*, and *S. uvarum*. The ubiquity of these hybrids and the defined industrial environments they are isolated from make them an excellent model for studying the role of hybridisation in microbial evolution and adaptation. By combining large-scale phenotyping with our knowledge of industrial niches and beer brewing history, we provide evidence that these hybrids originated in industrial environments and are highly niche-specific. Additionally, we characterised the genomic changes that occurred during domestication of these hybrids and describe the genetic mechanism leading to a key domestication phenotype, namely flavour production. Our results demonstrate that interspecific hybridisation is an evolutionary strategy that allows swift adaptation to novel niches and opens new avenues for the development of superior industrial yeasts.

8.3 Results & Discussion

8.3.1 Origins and diversity of *Saccharomyces* interspecific hybrids in beer

For several years we have been collecting and sequencing yeasts isolated from different industrial niches, including beer, wine, bread, sake, chocolate and liquor fermentations. As expected, the majority of isolates were *S. cerevisiae*, the yeast species most commonly associated with the production of fermented foods and beverages [33]. However, our analysis revealed that a surprisingly

large fraction of sequenced isolates were interspecific hybrids (∼25%). While some were isolated from lager beer fermentations, known to be driven by hybrid yeasts, many were collected from other beer niches, such as Trappist beers, spontaneously fermented 'Lambic' beers, and old beer bottles or equipment **Table 8.1, p. 195**. We identified three distinct species compositions across the hybrids investigated: *S. cerevisiae* x *S. eubayanus* (*S.cer* x *S.eub*), *S. cerevisiae* x *S. kudriavzevii* (*S.cer* x *S.kud*) and *S. eubayanus* x *S. uvarum* (*S.eub* x *S.uva*). Phylogenetic trees were built to position the hybrid subgenomes in the context of the currently available strain diversity of their parental species **Figure 8.1 p. 156 and Figures 8.14-8.17 p. 178-181**.

The *S. cerevisiae* subgenomes of the *S.cer* x *S.eub* (*S. pastorianus*, lager) hybrids belong to a well-defined monophyletic clade in the Beer 1 lineage, which mainly contains strains isolated from ale beers [31, 33, 35] **Figures 8.1, p. 156 and 8.14, p. 178**. Within Beer 1, the lager subclade is most closely related to the Belgium/Germany lineage and to Hefeweizen (German wheat beer) strains, indicating a Western European origin of lager yeasts. The lager clade of Beer 1 is further separated into the archetypical 'Frohberg' and 'Saaz' lineages [21, 52, 53], both hallmarked by very low nucleotide diversity (π=1.18E-03 and π=3.65E-04, respectively). The *S. eubayanus* subgenomes of the *S.cer* x *S.eub* hybrids similarly form a monophyletic group within the previously identified Holarctic (northern continents) clade [43] **Figures 8.1, p. 156 and 8.15, p. 179**, with clearly defined 'Frohberg' and 'Saaz' subclades. In accordance with previous reports, the *S. eubayanus* progenitor(s) of lager yeasts appear to be most closely related to Tibetan *S. eubayanus* strains, which may have reached Europe through Silk Road trading [20, 53]. Interestingly, the origin of domesticated barley (another major beer ingredient) in Northwestern Europe was also traced to Tibet [54]. However, considerable outcrossing and incomplete lineage sorting among Holarctic *S. eubayanus* strains makes determination of the exact geographical origin of the *S. eubayanus* ancestor(s) of lager yeasts difficult [43].

All but one of the *S. cerevisiae* subgenomes of the *S.cer* x *S.kud* hybrids belong to a monophyletic clade, closely related to the industrial yeast clade Beer 2 (labelled 'traditional Belgian beer' in **Figure 8.1, p. 156 and Figure 8.14, p. 178**, reflecting the niche origin of the isolates in this group). Similar to lager yeasts, the 'traditional Belgian beer' clade further divides into two subgroups containing hybrids isolated from either Lambic beers ('spontaneous beer fermentations') or Trappist beers ('Trappist ales'). One hybrid bread strain clusters within the clade, likely reflecting the historical relationship between brewers and bakers [55]. A close association between bread and ale beer strains was previously found among pure S. cerevisiae strains as well ('Mixed' clade [33] **Figure 8.14, p. 178**). The only *S.cer* x *S.kud* strain not belonging to the monophyletic clade is the VIN7 strain (a hybrid used in the wine industry), whose *S. cerevisiae* subgenome clusters within the Wine clade. This indicates that VIN7 originated independently from the ale and bread hybrids. The phylogenetic relationships of the *S. kudriavzevii* subgenomes in the *S.cer* x *S.kud* hybrids mirror the *S. cerevisiae* subgenome counterparts (**Figure 8.1, p. 156 and Figure 8.17, p. 181**). Traditional Belgian beer hybrids form a monophyletic clade

with 'Trappist ales' and 'spontaneous beer fermentations' subgroups, with the wine hybrid VIN7 as a sister group and European pure species as an outgroup.

The phylogenetic structure of the hybrid subgenomes offers insight into the origins of these interspecific hybrids. First, the *S. cerevisiae* subgenomes of the lager beer lineage stems from Beer 1 whereas the traditional Belgian beer lineage forms a sister clade to Beer 2 indicating that both major beer lineages were involved in the emergence of interspecific beer yeast hybrids. Secondly, the monophyletic clustering of each beer hybrid type suggests that the present-day strains are the result of only one hybridization event per hybrid clade, or very few events involving very similar strains [56–60]. This implies that establishment of new hybrid strains would have been largely due to the spread and diversification of existing hybrids rather than multiple, independent emergence and selection events. Third, the *S. cerevisiae* progenitors seem to come from industrial niches closely associated with the ones from which the hybrids were isolated. For example, the *S. cerevisiae* subgenomes of the *S.cer* x *S.kud* beer hybrids are most closely related to the Beer 2 clade, while the *S. cerevisiae* subgenome of VIN7 is located within the Wine clade (**Figure 8.14, p. 178**). Therefore successful industrial yeast hybrids likely formed in or close to the fermentation environments in which they are used, followed by a strong selection of the hybrids in their respective specific industrial niche (sour beers, wine, strong trappist beers, etc.).

The *S. eubayanus* subgenomes of the *S.eub* x *S.uva* hybrids form a monophyletic sister clade to those of the *S.cer* x *S.eub* lager strains indicating that the *S. eubayanus* parents were closely related. The subgenomes clearly separate according to geographical isolation (Germany vs. Belgium) which suggests that a single or few closely related European *S. eubayanus* strains formed the *S. eub* x *S. uva* hybrids and diversification is only due to spatial isolation. All *S. uvarum* hybrid subgenomes are also closely related to European strains **Table 8.1, p. 195**, and thus the hybrids likely originated in Europe. In contrast to the monophyly of the *S. eubayanus* subgenomes, the *S. uvarum* subgenomes are spread out and indicate that at least three hybridisation events gave rise to the *S.eub* x *S.uva* hybrid diversity in our collection (**Figure 8.16, p. 180**). The genetic diversity across the *S. uvarum* subgenomes is significantly higher than across the corresponding *S. eubayanus* subgenomes (average nucleotide diversity π=2.11E-03 and π=4.66E-04 respectively, one-sided Mann-Whitney U test, P $<$ 2.2 10^{-}16) and there is no clear niche substructure, suggesting that *S. uvarum* strains move more freely across environments. Indeed, *S. uvarum* has been isolated all over the world from a wide array of niches, natural and man-made [61] and is a known contaminant in brewing environments [41]. To date, no wild European *S. eubayanus* has been isolated. Coupled with the fact that all of our *S.eub* x *S.uva* hybrids were isolated from spontaneous fermentation environments or old bottles and brewing equipment, it is likely that the hybridisations occurred within the brewing environment.

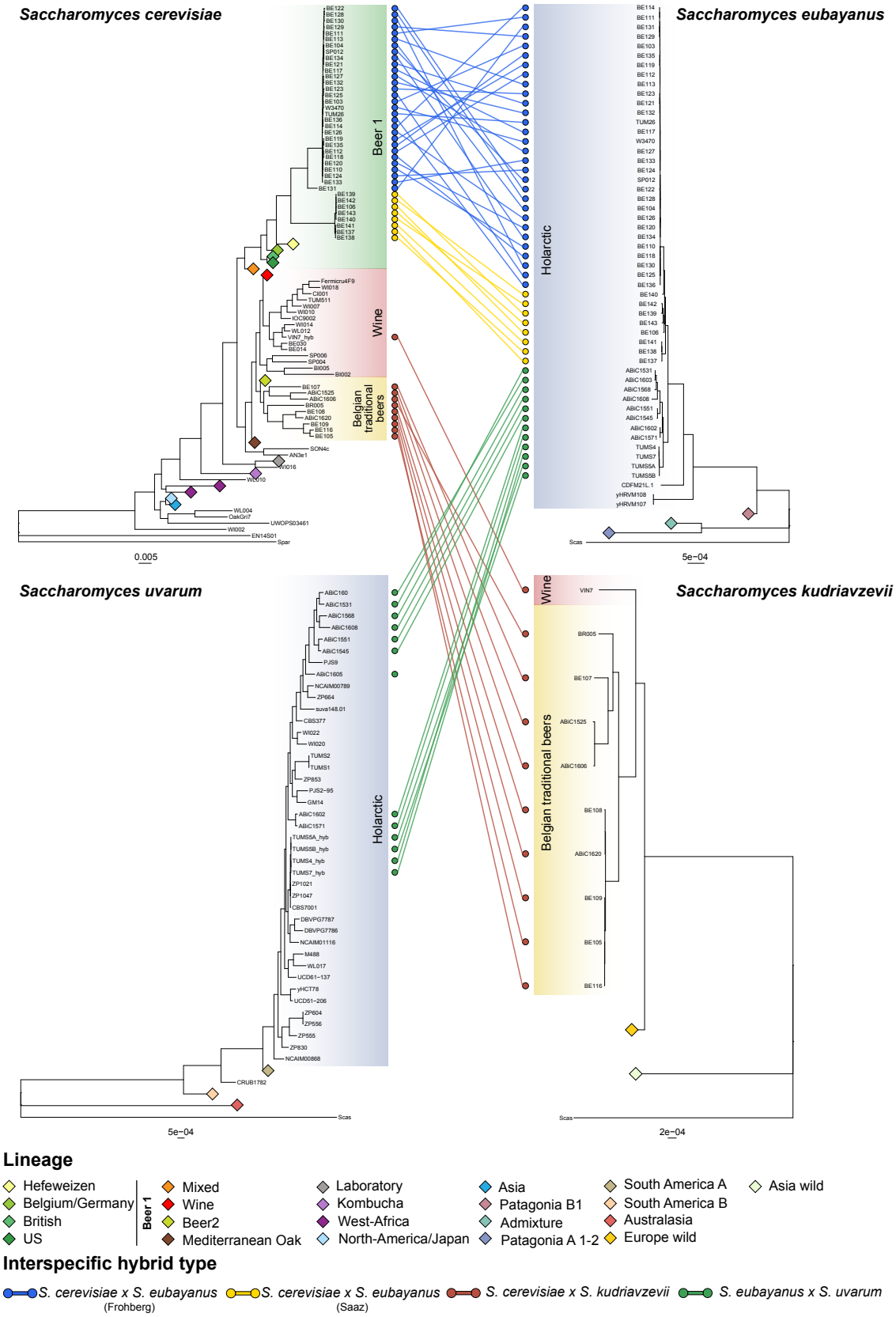


Figure 8.1: Tanglegram depicting the relationships between *Saccharomyces* pure species and interspecific hybrids. The tanglegram includes phylogenetic trees of *Saccharomyces cerevisiae*, *Saccharomyces eubayanus*, *Saccharomyces uvarum*, and *Saccharomyces kudriavzevii*. Lines connect distinct species identified within the same strain and are colored according to hybrid type. Lineages that do not contain hybrid strains are collapsed (diamonds). Expanded trees are reported in **Figures 8.14-8.17, p. 178-181**.

8.3.2 **Interspecific hybrids genomes are hallmarked by allopolyploidy and chimerism**

Table 8.1, p. 195 Figure 8.14, p. 178 Figures 8.1-8.15, p. 156-179. Newly formed hybrids experience extensive genome reorganization resulting in aneuplodies and chimeric chromosomes [62, 63]. We found significant variation within and between hybrid types in overall ploidy, copy number of large chromosomal fragments and full chromosomes, as well as in the degree of parental species contribution to the hybrid genome **Figures 8.2-8.5, p. 158-161**.

The *S.cer* x *S.kud* hybrids are overall triploid, with diploid *S. cerevisiae* and haploid *S. kudriavzevii* subgenomes (**Figure 8.2, p. 158**). Notable exceptions are two Belgian spontaneous beer fermentation strains, ABI1606 and ABI1525, with a total ploidy of roughly 4n (triploid *S. cerevisiae* and haploid *S. kudriavzevii*).

The *S.uva* x *S.eub* strains are divided into several subgroups with respect to subgenome content (**Figure 8.3, p. 159**). The first subgroup, containing a subset of Belgian Lambic strains, exhibits a diploid *S. uvarum* subgenome and an extremely fragmented *S. eubayanus* subgenome. The second subgroup, consisting of German brewing contaminant strains, presents a more uniform 1:1 *S. uvarum*: *S. eubayanus* ratio, with multiple partial or complete chromosome deletions and a few duplicated chromosomal segments. The last subgroup contains two Lambic strains ABIC1571 and ABIC1602 which exhibit an intermediate genome composition. Even though the subgroups likely derive from different hybridisation events (**Figure 8.16, p. 180**), the differences in genome composition might reflect different selective pressures or evolutionary speeds in their respective niches. For example, the German brewing contaminant isolates clearly evolved more slowly than the Lambic strains in the first subgroup as indicated by variations in branch length between the two subgroups **Figure 8.3, p. 159 and Figure 8.16, p. 180**.

In *S.cer* x *S.eub* strains, there is a clear distinction in genomic composition between the Saaz and Frohberg lineages (**Figures 8.4-8.5 p. 161-161**). Saaz strains are typically triploid (haploid *S. cerevisiae* and diploid *S. eubayanus*) as observed in other reported Saaz sequences [23, 59]. Frohberg strains are generally tetraploid to pentaploid with a basal 2n:2n ratio of the parental subgenomes. This high ploidy level is also in line with previously reported Frohberg genomes [21, 56, 58, 59, 62]. Both types demonstrate dramatic deletions and amplifications of large segments and even full chromosomes; Saaz strains mostly harbour losses in *S. cerevisiae* and amplifications of *S. eubayanus* whereas Frohberg strains demonstrate losses in *S. eubayanus* and amplifications of *S. cerevisiae*. Most chromosomal regions exhibit integer ploidy changes but we did find instances of strains within each hybrid type with intermediate ploidy changes (e.g. regions of 0.5 ploidy increments) (**Figure 8.18 p. 182**). Given that sequencing was performed from single cell colonies, we would have expected relatively isogenic populations. However, intermediate ploidies indicate that some populations carried alternative, unfixed genomic rearrangements, demonstrating the overall instability of the hybrid genome.

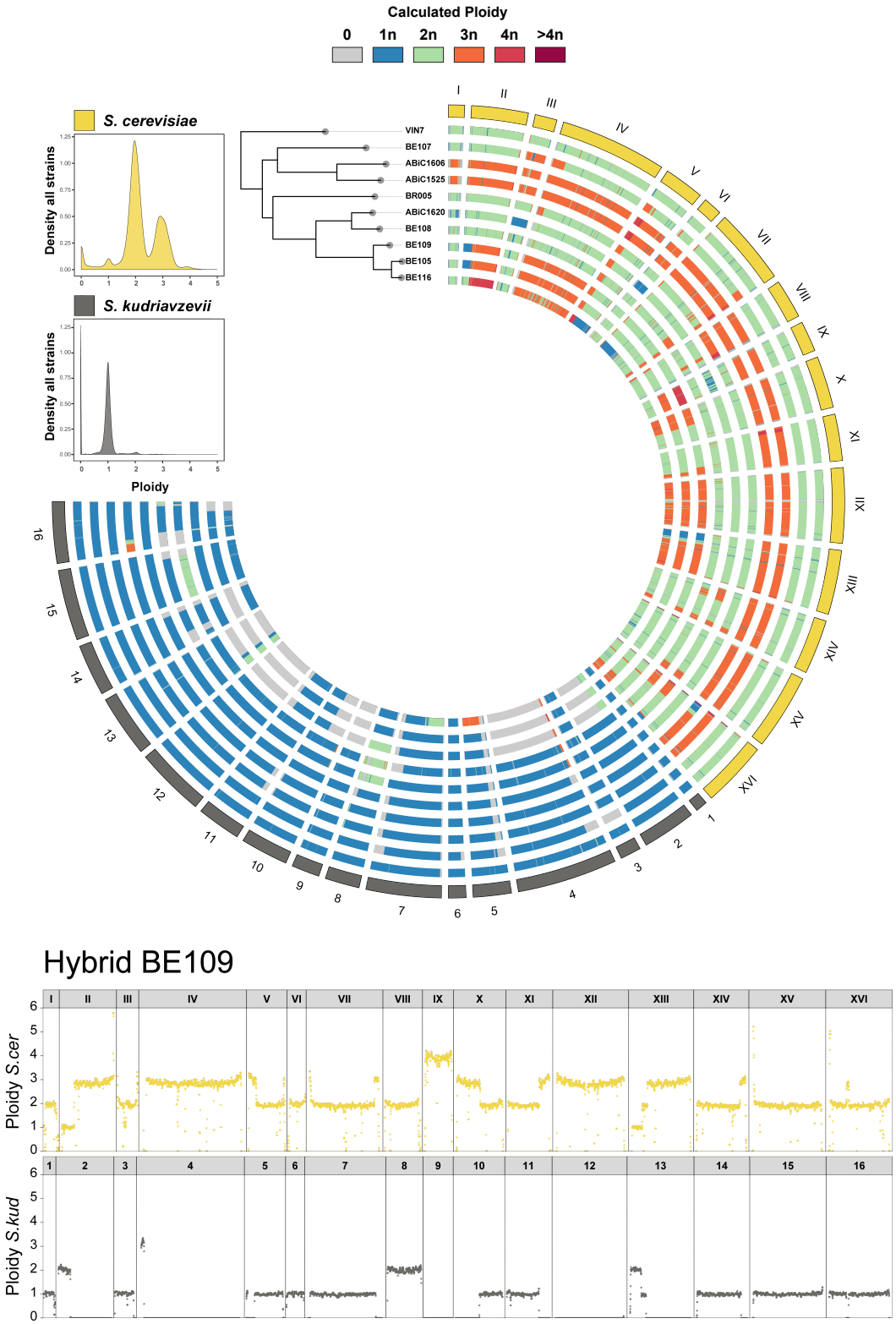


Figure 8.2: Genome structure of *S. cerevisiae* x *S. kudriavzevii* interspecific hybrids. Ploidy profiles of the *S. cerevisiae* x *S. kudriavzevii* hybrids. Top: strains are sorted based on the phylogeny of *S. cerevisiae* parental species **(see Figure 8.14, p. 178)**. Chromosomes are colored based on the calculated ploidy (complete deletion: gray, haploid: blue, diploid: green, triploid: orange, tetraploid: red, higher than tetraploid: maroon). Density plots represent the per-species ploidy distribution aggregated across all the strains of each hybrid type. Bottom: detailed representation of the genomic contribution of the two parental species in one selected hybrid.

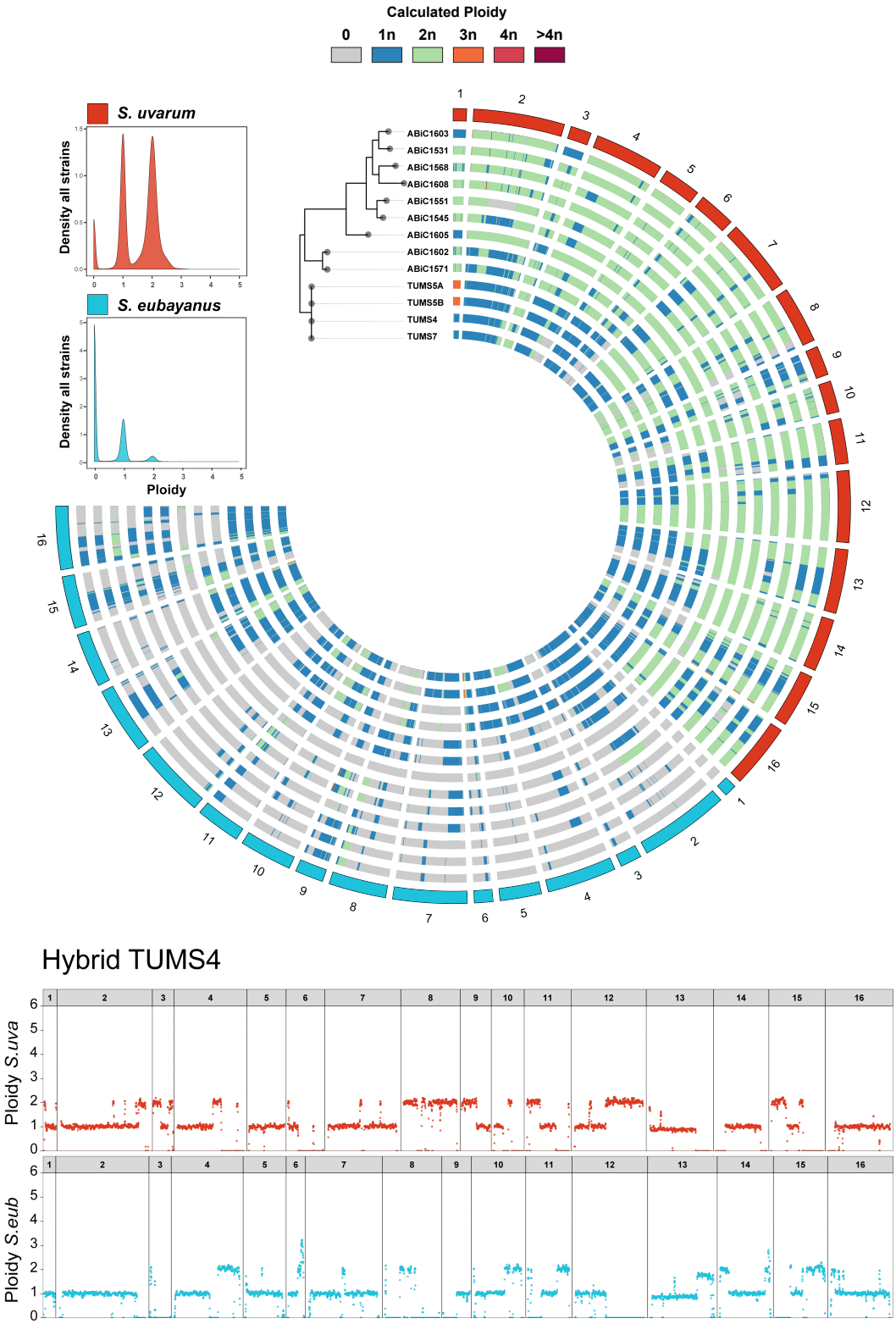


Figure 8.3: Genome structure of *S. uvarum* x *S. eubayanus* interspecific hybrids. Ploidy profiles of the *S. uvarum* x *S. eubayanus* hybrids. Top: strains are sorted based on the phylogeny of *S. uvarum* **(see Figure 8.16, p. 180)**. Chromosomes are colored based on the calculated ploidy (complete deletion: gray, haploid: blue, diploid: green, triploid: orange, tetraploid: red, higher than tetraploid: maroon). Density plots represent the per-species ploidy distribution aggregated across all the strains of each hybrid type. Bottom: detailed representation of the genomic contribution of the two parental species in one selected hybrid.

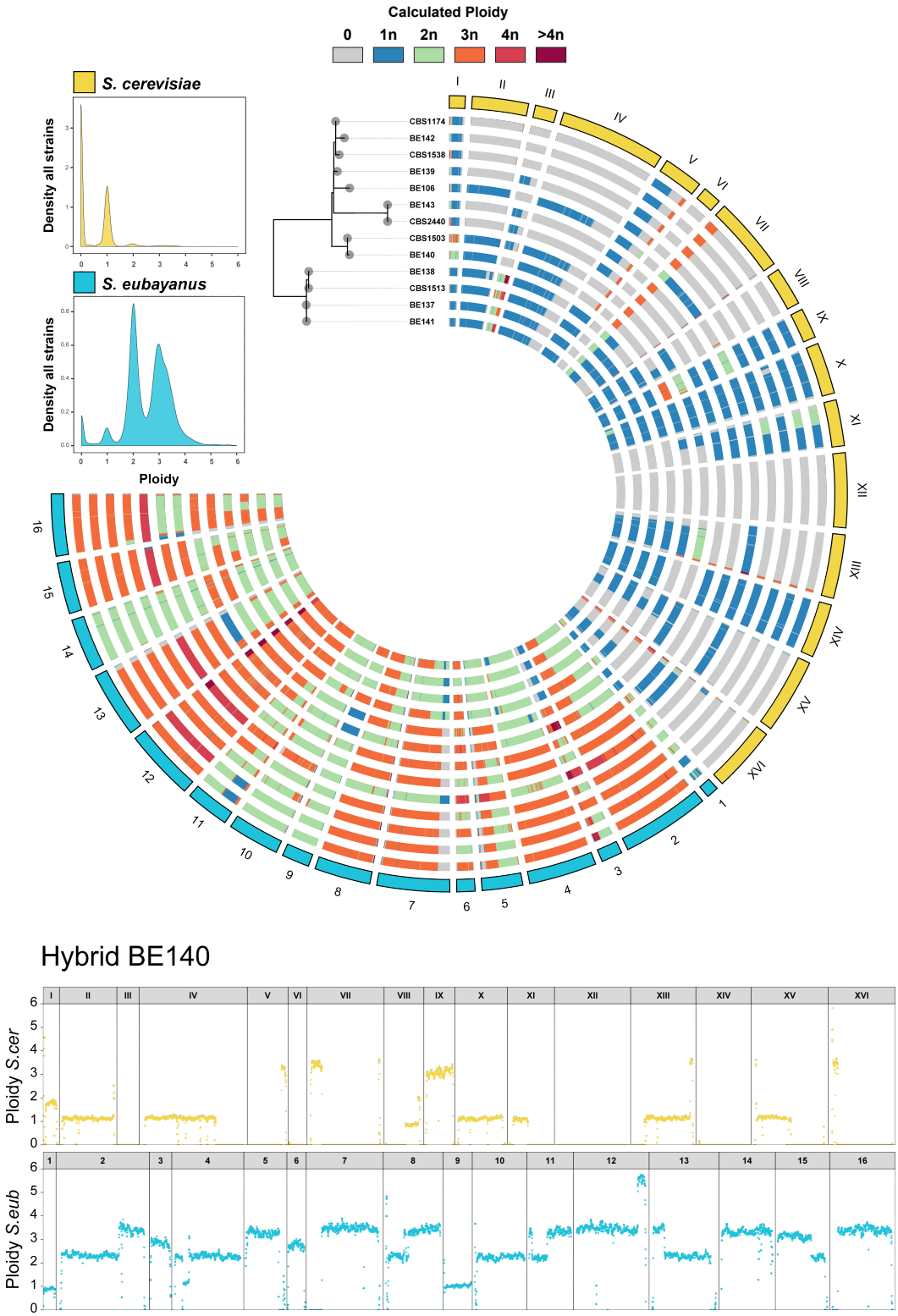


Figure 8.4: Genome structure of *S. cerevisiae* x *S. eubayanus* (Saaz) interspecific hybrids. Ploidy profiles of the *S. cerevisiae* x *S. eubayanus* hybrids from the Saaz lineage. Top: strains are sorted based on the phylogeny of *S. cerevisiae* **(see Figure 8.14, p. 178)**. Chromosomes are colored based on the calculated ploidy (complete deletion: gray, haploid: blue, diploid: green, triploid: orange, tetraploid: red, higher than tetraploid: maroon). Density plots represent the per-species ploidy distribution aggregated across all the strains of each hybrid type. Bottom: detailed representation of the genomic contribution of the two parental species in one selected hybrid.

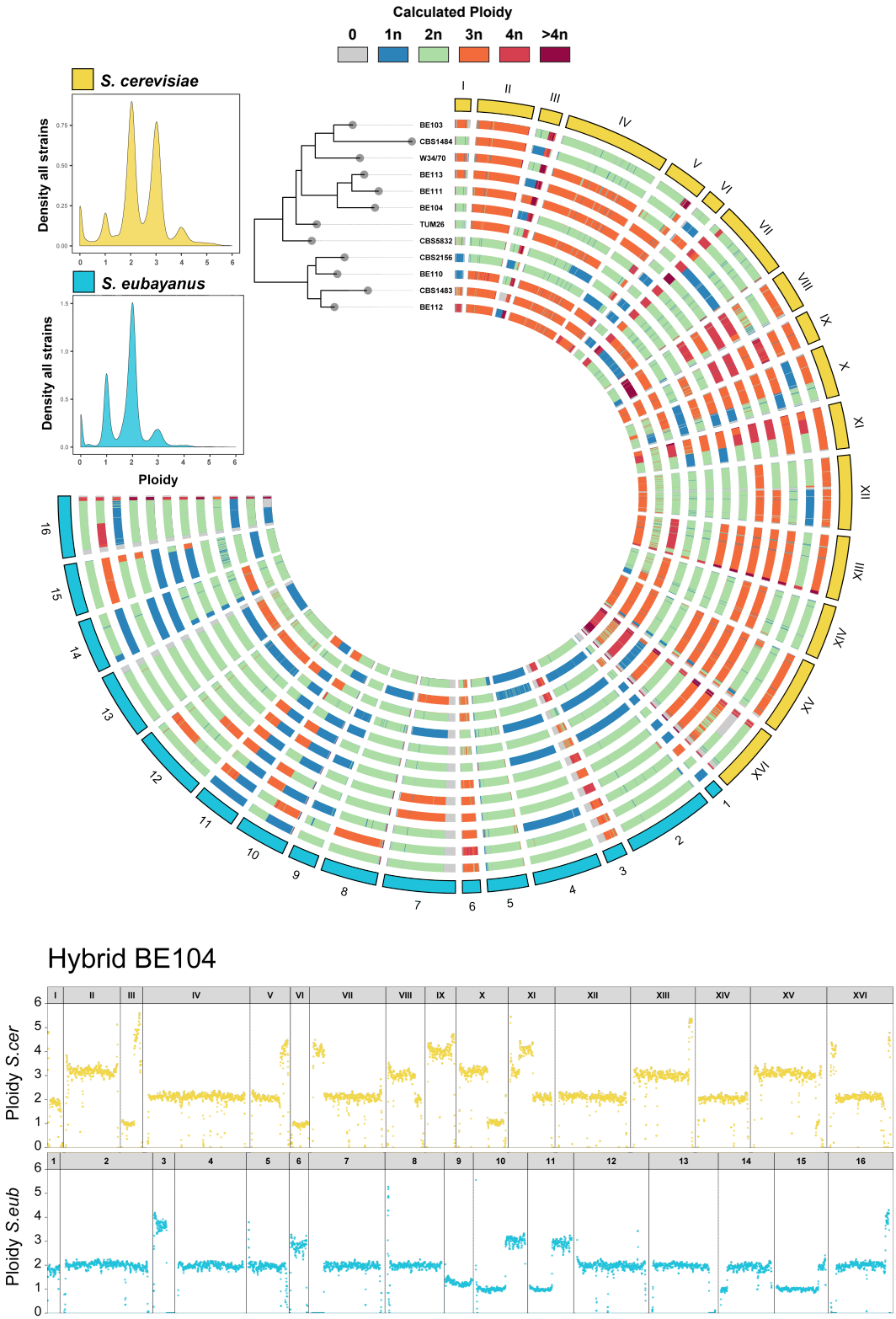


Figure 8.5: Genome structure of *S. cerevisiae* x *S. eubayanus* (Frohberg) interspecific hybrids. Ploidy profiles of the *S. cerevisiae* x *S. eubayanus* hybrids from the Frohberg lineage. Top: strains are sorted based on the phylogeny of *S. cerevisiae* **(see Figure 8.14, p. 178)**. Chromosomes are colored based on the calculated ploidy (complete deletion: gray, haploid: blue, diploid: green, triploid: orange, tetraploid: red, higher than tetraploid: maroon). Density plots represent the per-species ploidy distribution aggregated across all the strains of each hybrid type. Bottom: detailed representation of the genomic contribution of the two parental species in one selected hybrid.

Within the same hybrid type, we also observed striking differences in the copy number of full chromosomes and large chromosomal fragments between strains. These copy number changes are often shared by some but not all strains originating from the same hybridisation event, indicating that post-hybridisation genome structural rearrangements are still ongoing in most strains. In many cases, a copy number change in one subgenome is compensated by an opposite copy number change on the homoeologous portion of the other subgenome, suggesting the large-scale occurrence of genomic rearrangements leading to chimeric chromosomes and often to loss of heterozygosity (LOH).

Mapping of these chimeric regions revealed more than 300 breakpoints amongst the hybrids (77 in *S.cer* x *S.kud*, 84 in *S.cer* x *S.eub*, and 150 in *S.uva* x *S.eub*, **Figures 8.6-8.8, p. 163-165 and Figure 8.19**). The higher occurrence of breakpoints in S.uva x S.eub hybrids could be due to the lower nucleotide divergence between the two subgenomes (average whole-genome nucleotide identity 91.7%) compared to *S.cer* x *S.kud* (84.9%) and *S.cer* x *S.eub* (84.5%), which leads to a higher frequency of homoeologous recombination [14, 57, 58]. In fact, DNA sequence homology between subgenomes is significantly higher in breakpoint regions compared to non-breakpoint regions within stretches ranging from 50bp (microhomology) up to 1kb (**Figure 8.20, p. 184**).

Using the identified breakpoints, we performed a multi-strain comparative analysis of breakpoint occurrence to investigate which hybrids share a common origin and predict the trajectory of genome reorganization during diversification (**Figures 8.6-8.8, p. 163-165**). Given the larger number of strains and higher coverage of the *S.cer* x *S.eub* hybrids, we focused our analysis on this group (**Figure 8.8, p. 165**). Out of the 84 identified breakpoints, 2 are found in all analysed strains. An additional 7 breakpoints are found across all three subclades (Frohberg, Saaz 1 and Saaz 2) but the majority (78.5%) are subclade-specific (33 Frohberg, 11 Saaz 1, 22 Saaz 2). The sharing of breakpoints across the subclades supports a common origin of all lager yeasts. However, the exact trajectory and relationship of the subclades is difficult to disentangle given that several breakpoints may have been present at one time but are currently undetectable due to chromosomal losses (e.g. the complete loss of *S. cerevisiae* chromosome XII in all Saaz strains). However, it is also possible that some or all of these sites are more susceptible to rearrangements (fragile sites) and that shared breakpoints might have arisen independently in different hybrid lineages [56, 58–60].

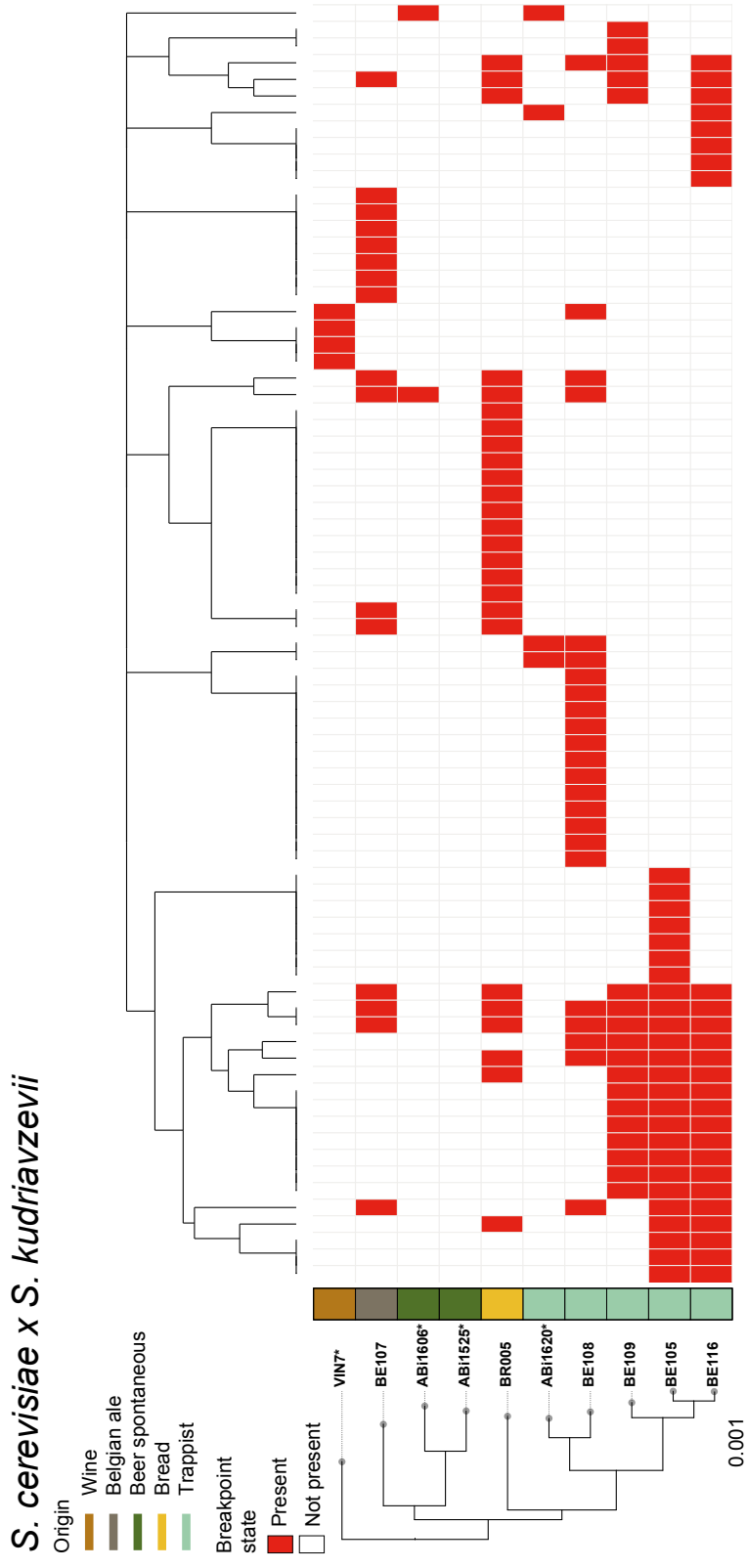


Figure 8.6: Distribution of chimeric breakpoints across interspecific hybrids. *S. cerevisiae* x *S. kudriavzevii* hybrids. Strains (rows) are sorted phylogenetically according to the *S. cerevisiae* subgenome **(see Figure 8.14, p. 178)**. Breakpoints (columns) are hierarchically clustered based on their presence (red square) or absence (white square) across strains. Strains for which low sequencing coverage level negatively affected the detection of breakpoints are indicated with an asterisk.

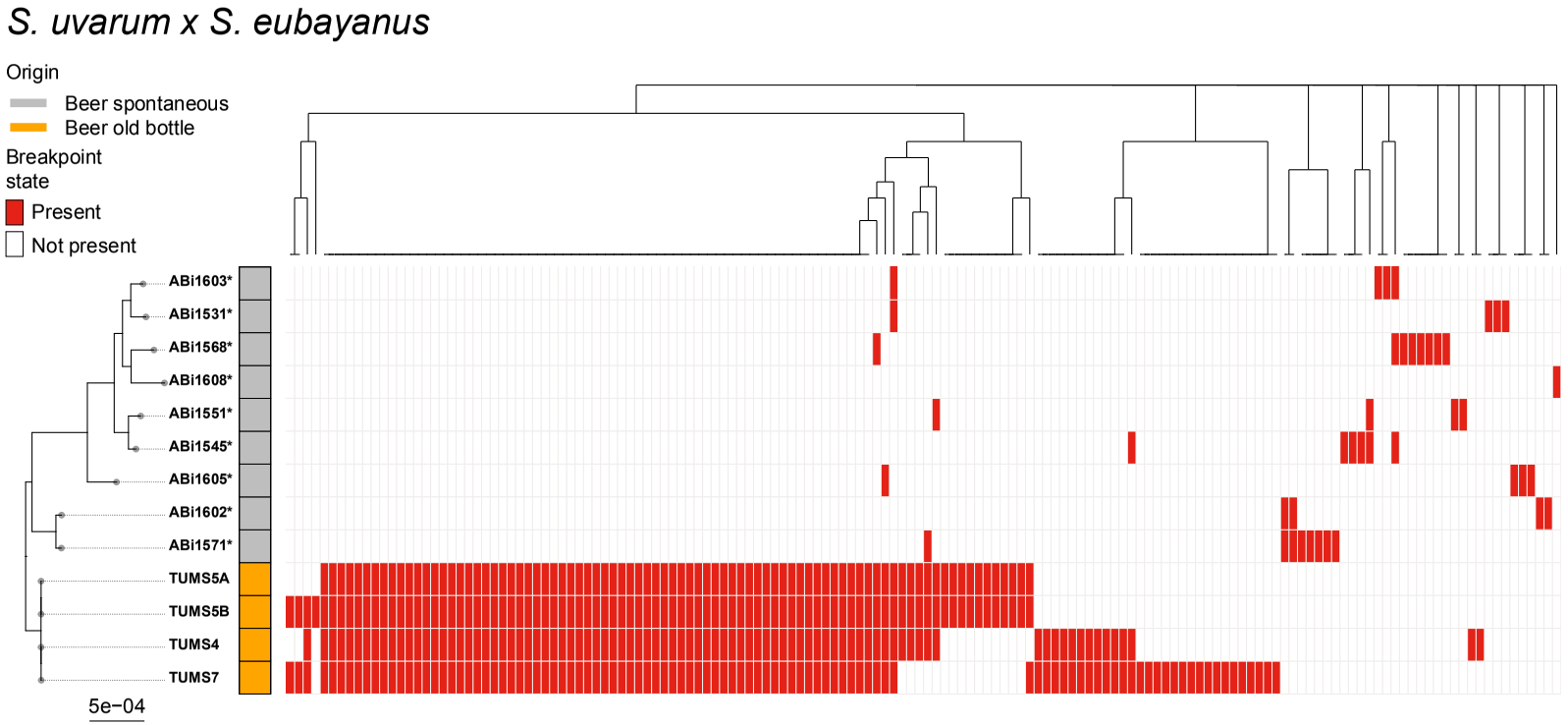


Figure 8.7: Distribution of chimeric breakpoints across interspecific hybrids. *S. uvarum* x *S. eubayanus* hybrids. Strains (rows) are sorted phylogenetically according to the *S. uvarum* subgenome **(see Figure 8.14, p. 178)**. Breakpoints (columns) are hierarchically clustered based on their presence (red square) or absence (white square) across strains. Strains for which low sequencing coverage level negatively affected the detection of breakpoints are indicated with an asterisk.

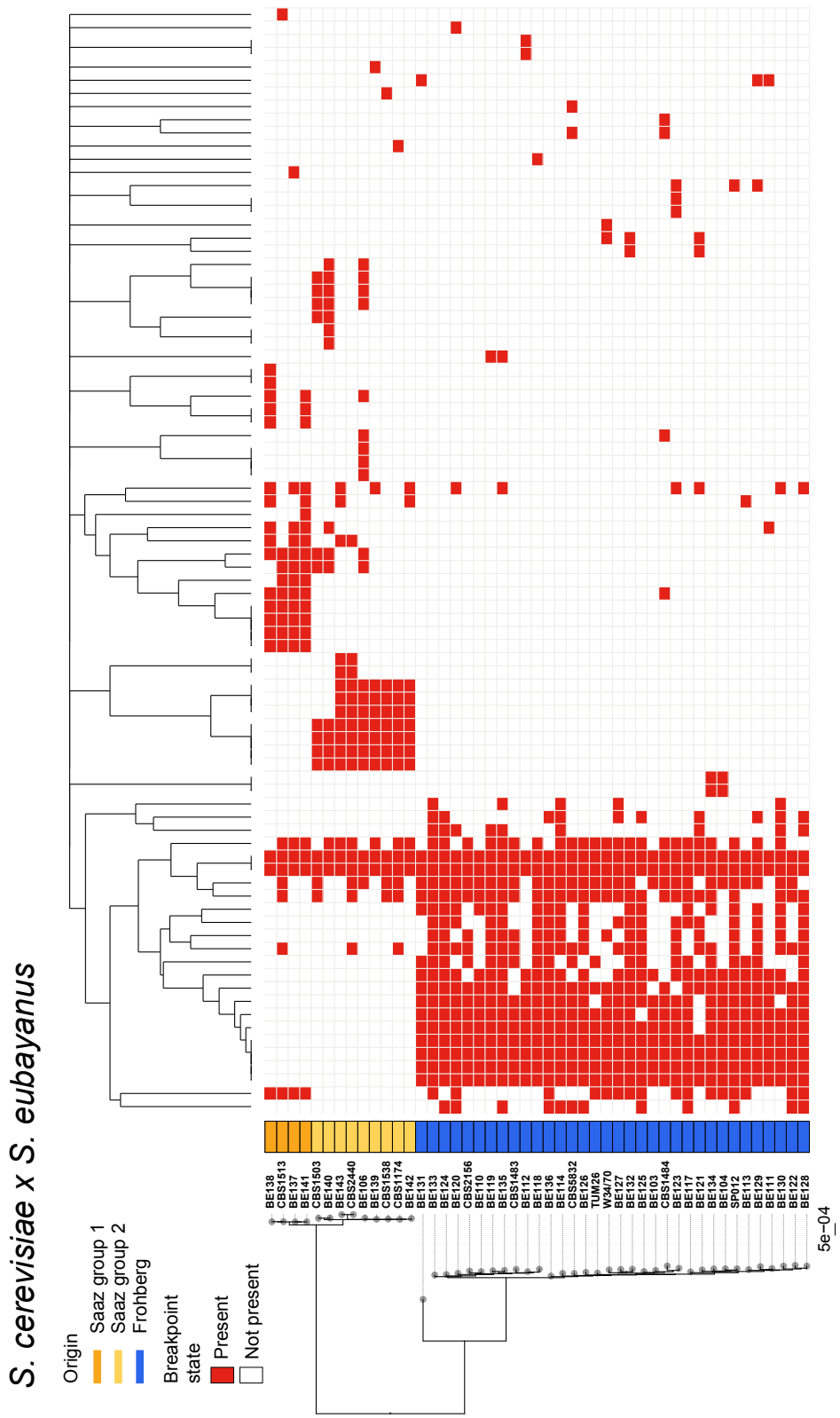


Figure 8.8: Distribution of chimeric breakpoints across interspecific hybrids. *S. cerevisiae* x *S. eubayanus* hybrids. Strains (rows) are sorted phylogenetically according to the *S. cerevisiae* subgenome **(see Figure 8.14, p. 178)**. Breakpoints (columns) are hierarchically clustered based on their presence (red square) or absence (white square) across strains. Strains for which low sequencing coverage level negatively affected the detection of breakpoints are indicated with an asterisk.

8.3.3 Hybrid beer yeasts exhibit unique phenotypic features that reflect niche adaptation

Large-scale changes in genome content and structure are intrinsically linked to phenotypic changes, which may confer novel fitness advantages [64–66]. To assess the extent of phenotypic changes in our hybrids, we extensively phenotyped the sequenced isolates and multiple pure species. Assays covered several industrially relevant traits including stress tolerances, metabolite production, growth as well as traits important for survival in nature such as sporulation efficiency and spore viability.

Based on overall phenotypic behaviour, strains cluster into three major groups with a clear genetic origin and industrial niche substructure (**Figure 8.9, p. 168**). Group A contains all pure *S. cerevisiae* strains and *S.cer* x *S.kud* hybrids plus two *S.cer* x *S.eub* hybrids, which form several subgroups with distinct phenotypic profiles. One of the most distinguishing features between these subgroups is the division between beer-like traits and wine-like traits; subgroups A3 and A4 (Trappist *S.cer* x *S.kud* hybrids and *S. cerevisiae* ale strains) exhibit weaker environmental stress resistance and sporulation efficiency than subgroups A1, A2, and A5 (strains used in wine, sake, spirits, cider, bioethanol, and Lambic brewing). This is indicative of a strong domestication signature (genome decay) in A3 and A4 which is common in ale strains [33, 35]. Group B includes the non-*cerevisiae* pure species as well as all but two *S.uva* x *S.eub* hybrids. Except for the pure *S. kudriavzevii* strains in subgroup B1, these strains demonstrate high cold, osmo-, and desiccation tolerance. Group C contains all but three lager hybrids and further subdivides in Frohberg (C1) and Saaz (C3). They can tolerate lower temperatures than strains from Group A. Tolerance to extreme cold (4 °C) is limited compared to strains from Group B. Group C exhibits overall lower environmental stress tolerance than yeasts from non-beer or spontaneous fermentation environments (primarily in Groups A and B). Cold tolerance (a common trait in non-*cerevisiae* pure species) and the ability to efficiently use maltotriose (a beer-specific carbon source fermentable by many *S. cerevisiae* strains but typically not by non-*cerevisiae* pure species) are commonly combined in interspecific hybrids (**Figure 8.10a, p. 169**). This combination has been proposed as the reason why these hybrids persisted and flourished in cold brewing environments [44]. Only one hybrid (BE131) appears to have lost its cold tolerance, potentially a result of a unique introgressive hybridisation to the *S. cerevisiae* Beer 1 ?Britain? clade (**Figure 8.21, p. 185**). The combination of cold tolerance and efficient maltotriose utilisation could either be the result of post-hybridisation adaptation from niche-specific selective pressures or a direct consequence of the hybridisation process itself. To directly evaluate these possibilities, we phenotyped several newly developed artificial *S.cer* x *S.eub* hybrids and compared them to their parental strains (**Figure 8.10b, p. 169**). The artificial hybrids demonstrated improved cold tolerance compared to the *S.cer* parental strain (ANOVA $F=8.88$, $p < 0.001$) and improved maltotriose utilisation compared to the *S.eub* parental strain (ANOVA $F= 75.53$, $p < 0.001$), indicating that a hybridisation event by itself can

generate immediate fitness advantages in novel niches such as cold beer fermentation. Although hybridisation can impart immediate fitness advantages, the instability of interspecific hybrids likely facilitates further adaptation to specific beer niches, best exemplified by the 'traditional Belgian beer' lineage. This lineage likely originated from a single hybridisation event (see **Figure 8.1, p. 156**) and split into two distinct subgroups, 'Lambic' and 'Trappist' strains, which function in two profoundly different beer production processes. During Lambic beer production, the presence of acid-producing bacteria leads to considerable concentrations of acetic and lactic acid [67, 68], whereas in Trappist beer production, these bacteria are much less prominent. Despite their shared origin, tolerance to organic acids is observed in the 'Lambic', but not in the 'Trappist' subgroup, suggesting adaptation after the hybridisation event (**Figure 8.10c, p. 169**).

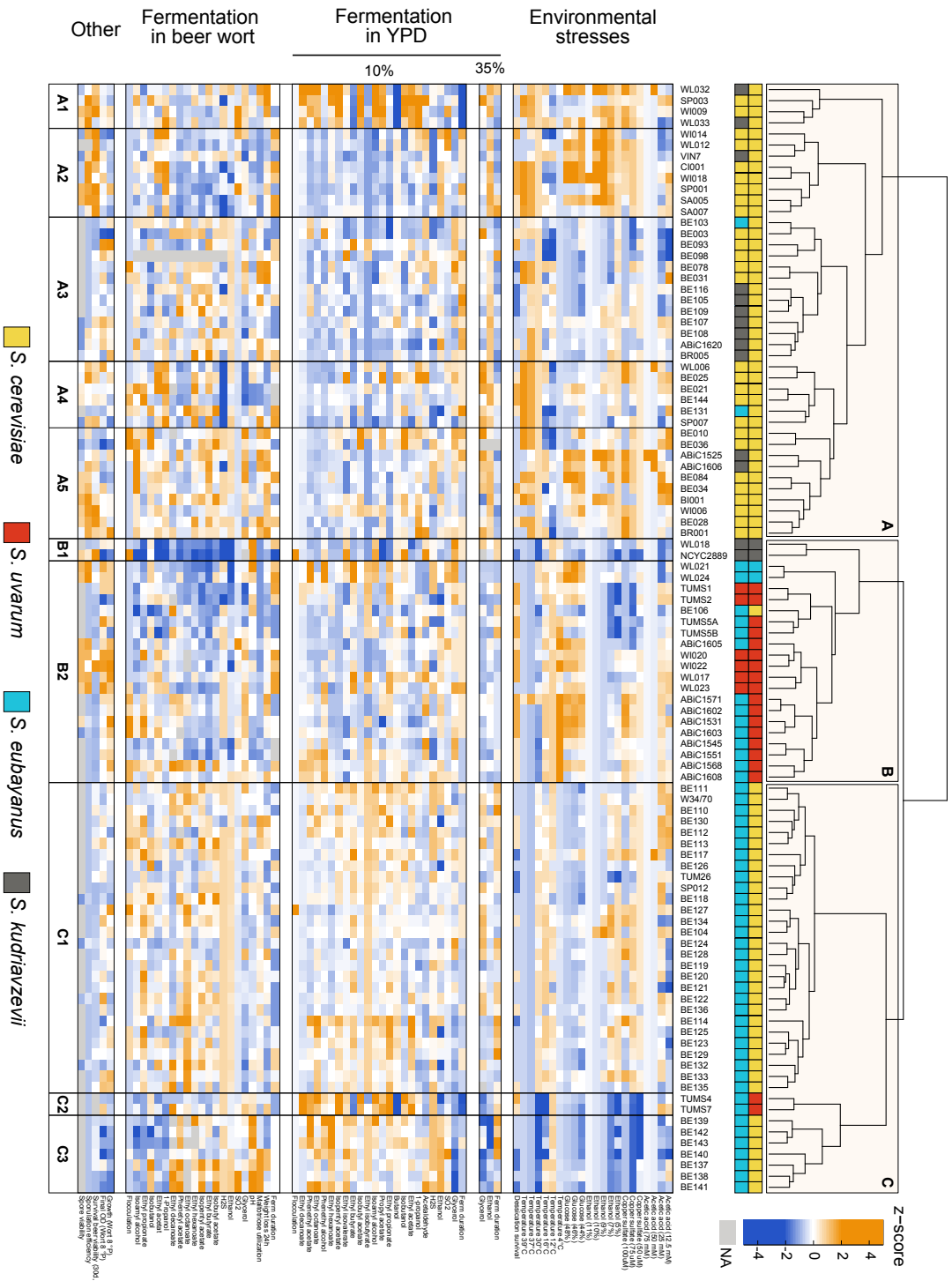


Figure 8.9: Trait variation of interspecific hybrids. Heat map representation of phenotypic diversity within interspecific hybrids and pure species. Phenotypic values are calculated as z-scores (normalized values) and coloured according to the scale on the top right. Missing values are represented by gray shadings. Strains (columns) are hierarchically clustered based on phenotypic behaviour. Phenotypes (rows) are sorted based on five major categories (black boxes). The coloured bar on the top of the figure indicates the species.

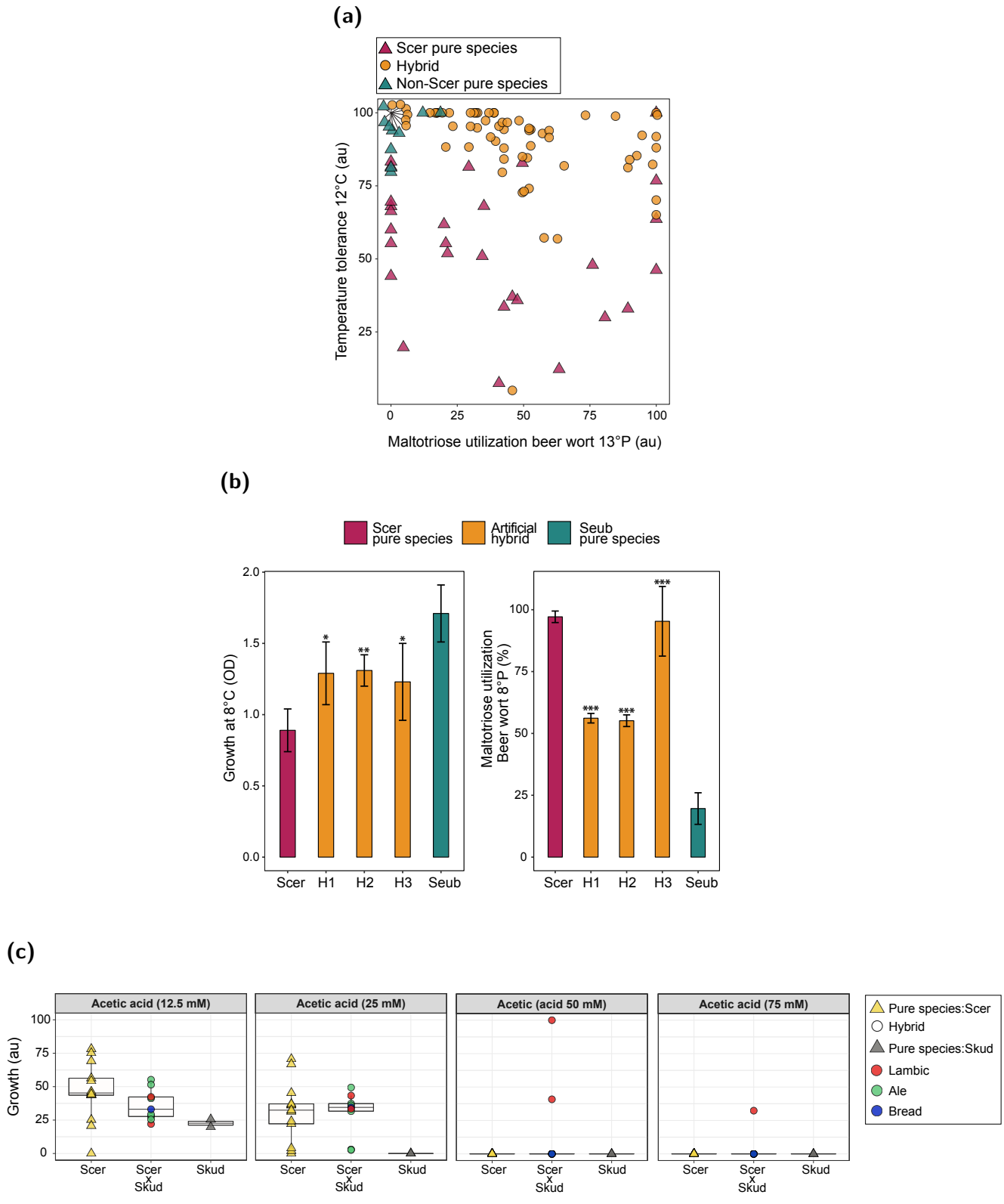


Figure 8.10: Interspecific hybridization leads to novel phenotypic combinations that allow colonization of new niches. a: Scatter plot describing cold tolerance vs. maltotriose utilization of hybrids (circles) and pure species (triangles). **b:** Performance of artificial hybrids and parental species in beer wort fermentations at cold temperatures. Bars indicate mean ± sd of four biological replicates. Statistical significance determined by ANOVA between hybrids and the inferior parental species: *($P \leq 0.05$), **($P \leq 0.01$), ***($P \leq 0.001$) **c:** Post-hybridization adaptation of a single hybrid lineage, ("Traditional belgian beers", **see Figure 8.14, p. 178**) to the production of different beer styles. The Lambic subgroup is better adapted to acid stress.

8.3.4 Structural rearrangements underlie phenotypic convergence to low off-flavour production in lager yeast

Beer flavour and aroma diversity can often be attributed to the yeast (and sometimes, bacteria) used during the fermentation process [69]. The style of beer dictates which specific aroma compounds are desirable and has thus influenced selection over time. The presence of 4-vinyl guaiacol (4-VG) is tolerated in some specialty beers such as wheat and saison beers, but is strongly undesired in most beer styles, including lagers. This compound produces a spicy, clove-like aroma referred to as 'phenolic off-flavour' (POF). Production of 4-VG depends on functional copies of two subtelomeric genes, phenylacrylic acid decarboxylase (*PAD1*) and ferulic acid decarboxylase (*FDC1*) [70]. Our phenotype analysis demonstrated that the non-*cerevisiae* strains (*S. eubayanus*, *S. kudriavzevii*, and *S. uvarum*) produce 4-VG (POF+) **(Figure 8.11, p. 170)**. As a dominant trait, newly formed hybrids with a non-*cerevisiae* parent should be POF+ regardless of the phenotype of the other parental strain. However, all *S.cer* x *S.eub* hybrids tested were found to be POF-, suggesting that 4-VG production was lost after hybridisation in lager strains.

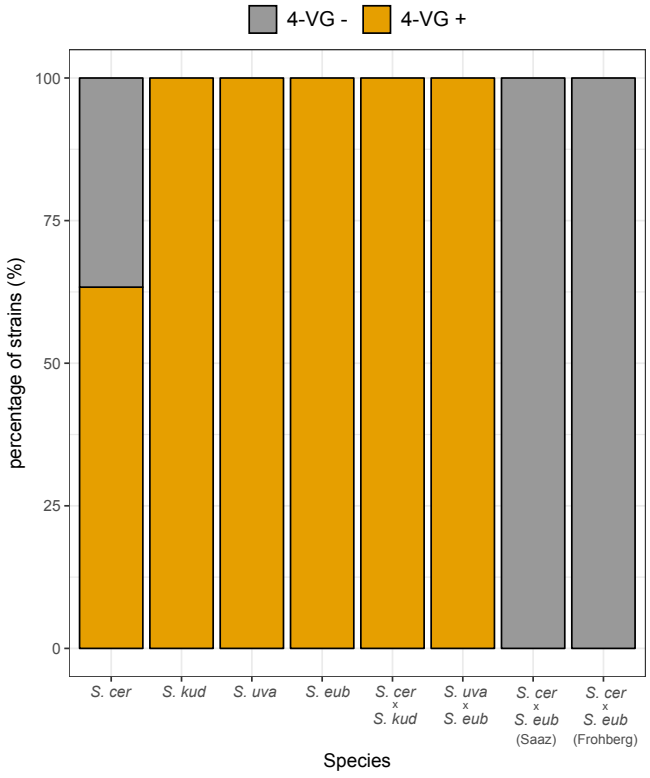


Figure 8.11: Distribution of 4-VG production across *Saccharomyces* pure species and hybrids

To understand the genetic foundation of this loss, we further evaluated the subgenomes of the lager hybrids. The *S. cerevisiae* subgenomes of these strains localise within the Beer 1 clade; the majority of strains from this clade acquired disruptive mutations in *PAD1* and/or *FDC1* [33, 35, 71], which suggests that the hybrids should have inherited an inactive POF pathway from their *S. cerevisiae* ancestor. Our analysis shows that Frohberg strains do indeed harbour the same disruptive mutation(s) as other Beer 1 *S. cerevisiae* **(Figure 8.12, p. 172)**). Saaz strains however have complete deletions of the *S. cerevisiae* PAD1-FDC1 gene cluster **(Figure 8.12, p. 172)**). Nevertheless, the hybrids must also have inherited an inactive POF pathway from the *S. eubayanus* parent to render them POF-. Three distinct genomic changes resulted in the loss of the PAD1-FDC1 gene cluster in the *S. eubayanus* subgenomes of the Saaz 1, Saaz 2 and Frohberg lineages, each involving a different chimeric breakpoint between the terminal regions of *S. cerevisiae* ChrXIII and *S. eubayanus Chr13* **(Figure 8.12, p. 172)**). Thus, each lineage experienced an independent loss of heterozygosity (LOH) event, further supporting a post-hybridisation loss of 4-VG production, most likely after the divergence of Frohberg and Saaz. Interestingly, this was the only region in the *S. eubayanus* subgenome that was lost three times independently in lager strains, signifying the strong selection against the POF phenotype.

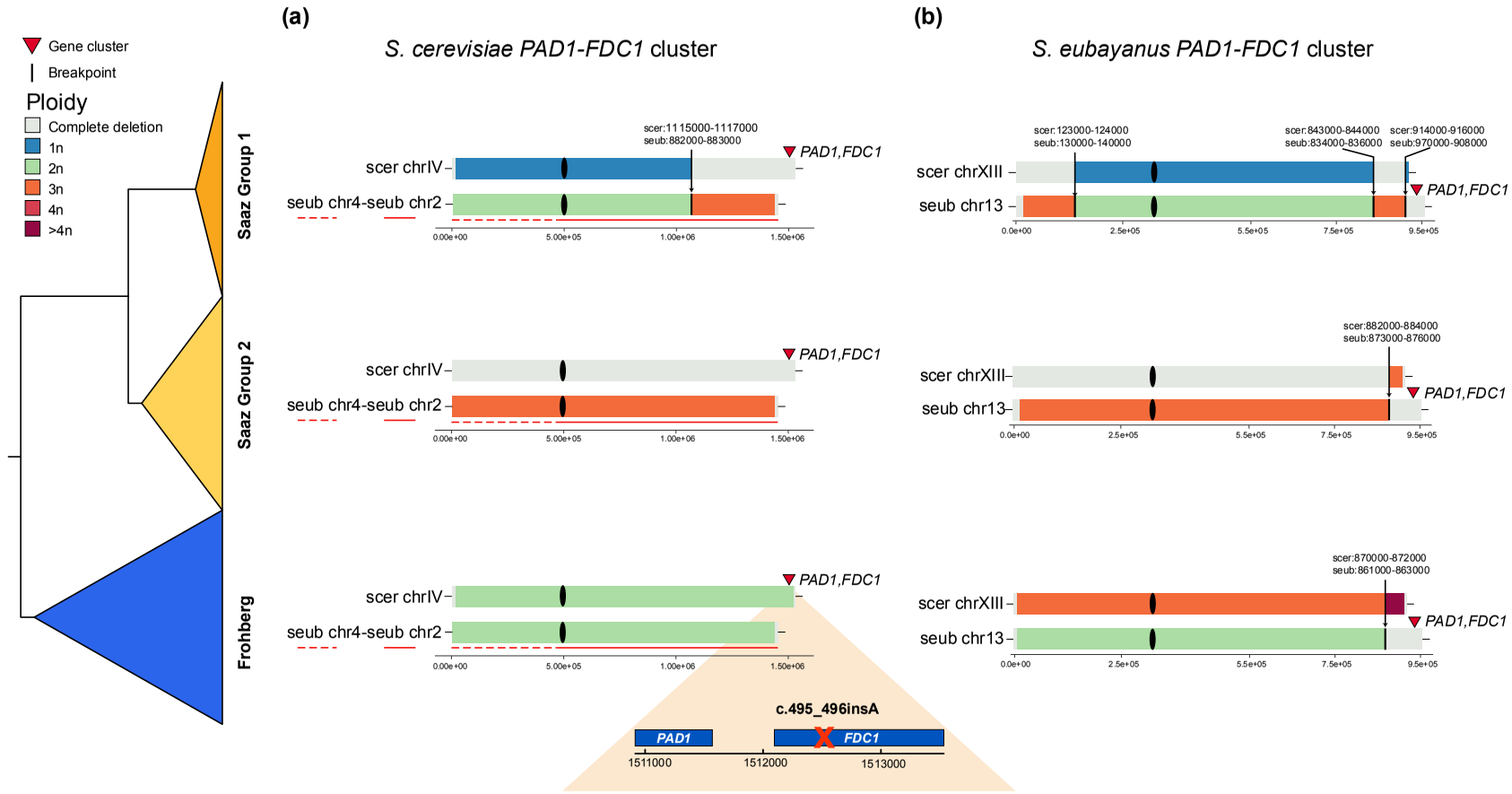


Figure 8.12: Schematic representation of the genetic basis of loss of 4-VG production in *S. cer* x *S. eub* hybrids. Three distinct mechanisms explain the loss or inactivation of the *PAD1-FDC1* gene cluster across the three *S. cer* x *S. eub* subgroups sorted phylogenetically on the left (Saaz group 1-dark yellow, Saaz group 2-light yellow, Frohberg-blue). **(a)** In the *S. cerevisiae* subgenome, the *PAD1-FDC1* gene cluster (red triangle) is located at the terminal region of *S. cer* chromosome IV. Saaz group 1: the right part of *S. cer* chromosome IV that contains *PAD1-FDC1* gene cluster is lost and overwritten by the syntenic region of *S.eub* chromosome 2. Black arrow indicates the location of the chimerisation breakpoint. Saaz group 2: *S.cer* chromosome IV is completely lost. Frohberg: the *PAD1-FDC1* gene cluster is present but *FDC1* is inactive due to the presence of a loss-of-function mutation. **(b)** In the *S. eubayanus* subgenome, the *PAD1-FDC1* gene cluster is located at the terminal region of *S.eub* chromosome 13. The *PAD1-FDC1* gene cluster is deleted in all the three groups and replaced by the terminal region of *S. cer* chromosome XIII. The location of the chimerisation breakpoints (black arrows) is distinct for each group. Chromosome are colored according to ploidy levels of one representative strain for each group.

8.3.5 Historical context of the origin and evolution of lager yeasts

The data we present here may offer a rare opportunity to place evolutionary events in industrial yeast hybrid lineages into a historical perspective, and to gain insight into the potential driving forces behind yeast hybrid domestication. It is important however to note that absolute dating of evolutionary events in industrial yeasts is complicated by a lack of solid calibration points (for example, fossils) and the difficulty of tracing historical movements of industrial yeasts, partly due to the late discovery of yeasts as the living entities responsible for fermentation and partly due to imperfect historical records. One biogeographical event related to the "Beer 1" clade may serve as a calibration point to date divergence times within the clade. As previously reported, the split between British and US ale-producing strains can be traced back to the import of ale (and the yeast they carried) in North America by British settlers in the early 17th century [33, 72]. More specifically, the split between US and UK Beer 1 yeasts is thought to have occurred around between 1607 and 1637 AD. We used this calibration in a Bayesian molecular clock analysis to date evolutionary events in the lager strain lineage and match these molecular events with historical records. It should however be noted that, given the paucity of calibration data available, our dating results can only be considered rough approximations.

According to our dating results, the most recent common ancestor (MRCA) of present-day ale and lager beer strains dates to the mid-16th century, sometime between 1547-1578 AD (see yellow highlight in **Figure 8.12, p. 172**). The lager yeasts lineage splits off from the Belgium/Germany clade within the same time frame. This divergence dates shortly after the enactment of the Bavarian purity law ("Reinheitsgebot") in 1516 AD which established strict beer brewing practices [21, 73]. Historically, this restriction had a large, reductive impact on the variety of German beer styles with many local beers and brewing traditions disappearing in order to conform to a more controlled beer production process, which served as a basis for present-day lager brewing practices [73]. The law established rules regarding ingredients and restricted brewing to winter. While the primary objective was to reduce bacterial contaminations, brewing beer in cold temperatures inadvertently selected for cold tolerant strains, a trait characteristic of *S. eubayanus*. From this point on, hybrids containing *S. eubayanus* would have a significant competitive edge over other brewing strains.

Perhaps the most thoroughly discussed divergence within the lager lineage is that between the classic archetypes Frohberg and Saaz [56, 59, 74]. Our dating results indicate that the split between these two sublineages occurred between 1693-1722 AD see blue highlight in **Figure 8.12, p. 172**). The most important lager-related historical event within this time frame is likely the change of ownership of the Spaten lager brewery in Munich, which had a profound impact on lager yeast dissemination and cultivation of the first pure lager yeast strains a century later. There is no concrete possible explanation for this split other than geographical separation. The most notable difference between the lager and ale clades is the sudden, dramatic reduction of evolutionary rates

of the Frohberg and Saaz lineages approximately 150 years after the split. This rate reduction is coupled to the onset of diversification of both lineages suggesting that the (low) genetic diversity we observe in today's commercial lager strains is a consequence of stringent selection and subsequent local divergence at the turn of the 20th century. Historically, this period coincides with three important events that revolutionized (lager) beer production. First, Louis Pasteur described the importance of yeast in alcoholic fermentations in 1857 AD [75]. Once brewers became aware of the true significance of their sediment and the possible economic implications, greater care was taken to maintain their brewing cultures. Second, mechanical refrigeration was introduced into breweries in 1873 AD [23, 76]. This technological advance allowed for year-round lager fermentations as well as proper cooled storage of successful yeast cultures. Additionally, it allowed for brewing consistently at even colder temperatures, fostering selection for very cold-tolerant strains. These colder fermentation temperatures and cold storage of yeast cultures could at least partly explain the reduction in evolutionary rate observed in the lager lineages. Third, the isolation of the first pure yeast culture in 1883 AD (now shown to be a *S.cer* x *S.eub* hybrid) by E. C. Hansen at the Carlsberg brewery (Denmark) gave rise to a wave of brewers isolating and sharing their strains [23, 76]. For example, the first culture isolated by Hansen (dubbed "Unterhefe nr. 1") was disseminated across many breweries in central Europe, and rapidly implemented in their production process. Together, these three events led to a drastic standardization of industrial lager production. The clear parallels between our molecular dating and several important historical brewing events demonstrate that today's lager yeast biodiversity was shaped by multiple waves of human selection. Most importantly, these waves were initiated by dramatic alterations or improvements to brewing practices and technology.

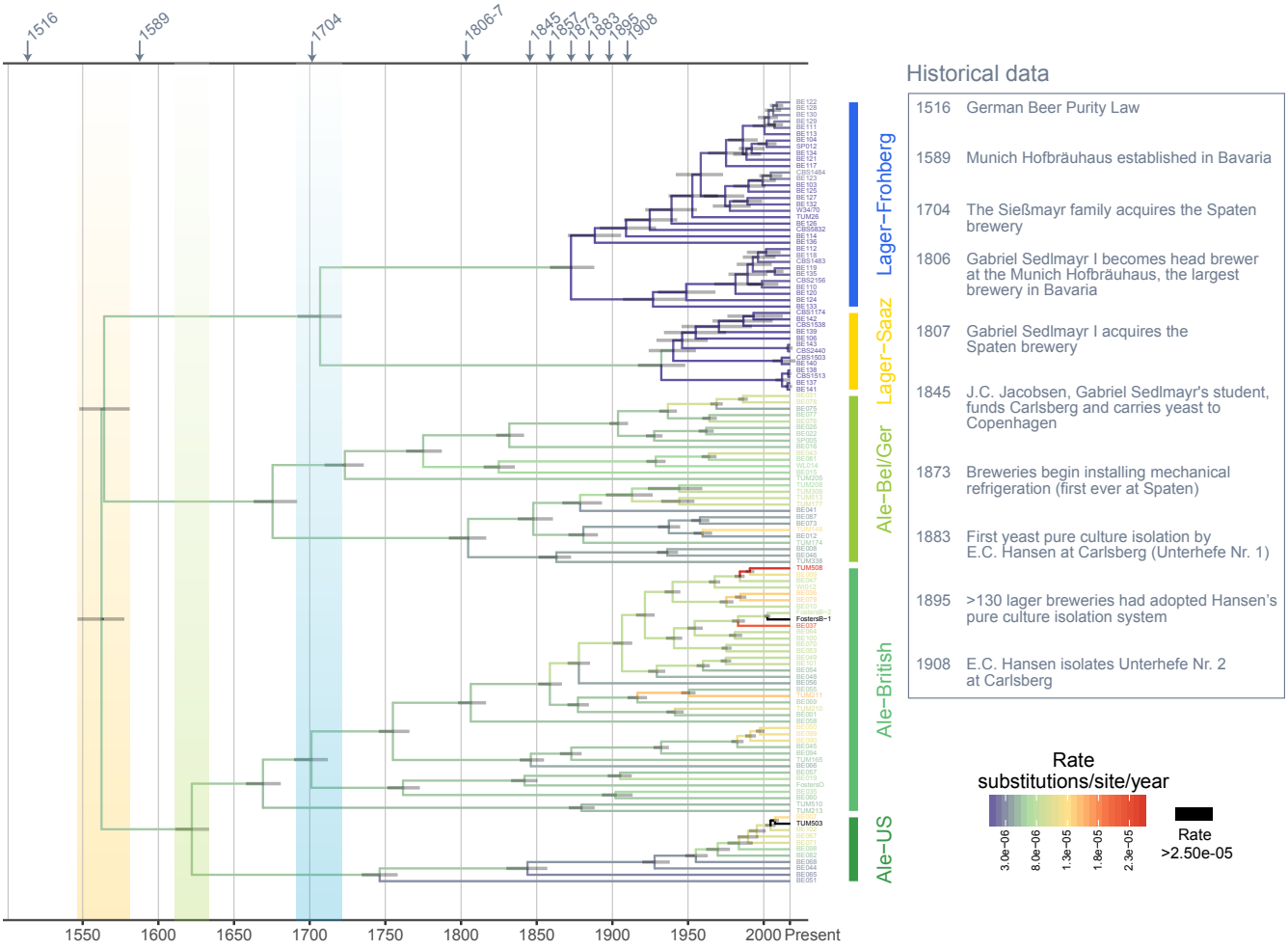


Figure 8.13: Time-calibrated phylogeny of *S. cerevisiae* Beer 1 clade. Divergence times have been inferred through Bayesian inference, as implemented in the BEAST software package [77], using a random local clock model (RLC). Node bars represent 95% Highest Posterior Density (HPD) intervals. Colored bars and text on the right indicate the five subclades identified in Beer1 **(see Figure 8.14, p. 178)**. Shaded boxes illustrate major splits within the *S. cerevisiae* Beer1 clade: Most recent common ancestor (MRCA) of present day ale and lager yeasts (yellow); introduction of European ale yeasts in US (green); MRCA of lager yeast Saaz and Frohberg clades (light blue). Branches are colored according to their branch-specific average posterior rate (purple-slowest, red-fastest). Slowest rates are observed in the two lager yeast subclades (Saaz and Frohberg). Dates on the top indicate relevant/major events in lager beer history summarized in the text box on the left.

8.4 Concluding remarks

Industrial, man-made environments challenge microbes with completely novel ecological niches and therefore offer insight into the processes that allow colonization in such conditions. Our results show that interspecific hybridisation is an important and common route towards diversification and adaptation to novel niches. Moreover, our genome analyses and phenotyping suggest that adaptation is likely fuelled both by the direct selective advantages of the new hybrids, as well as their genomic plasticity, allowing for swift adaptation to specific niches. Beer yeasts provide a perfect example of this phenomenon as interspecific hybridisation yielded new variants that combine the fermentation capacity of *S. cerevisiae* with the cold tolerance of other species. The instability and plasticity of the hybrid genomes likely allowed further adaptation and yielded variants that lost undesirable properties, such as the production of 4-vinyl guaiacol.

Our discovery of a large proportion of interspecific hybrids in Belgian specialty beers, many of which were previously thought to be pure *S. cerevisiae* strains, suggests that interspecific hybridisation played an important role in the history of industrial fermentations. Over the last century, the adoption of new brewing technologies and highly controlled single-yeast fermentation processes likely contributed to the decline of the natural beer yeast diversity, including interspecific hybrids. The low diversity of present-day lager yeasts exemplifies this trend. Although the first *S.cer* x *S.eub* hybrids originated around the 16th century AD, today's lager yeasts can be traced back to only a few lineages that were pure-cultured, cold-stored, and dispersed across multiple breweries in the late 19th century.

Belgium represents a notable exception to this biodiversity decline. Traditional Belgian beer styles harbour a remarkably diverse array of yeasts, likely as a result of the continued use of old beer brewing practices. For example, the production of Belgian Lambic beers, the most traditional of modern beer styles, has remained unchanged for centuries. The use of medieval brewing technologies such as open-air inoculation and long-term fermentation and storage in barrels, promotes the survival of unique yeast hybrids presented in this study, and could therefore provide a source of new biodiversity for industrial applications.

The study of the genetic and phenotypic makeup of naturally-occurring yeast hybrids may further aid in the development of innovative new hybrids for industry. Recent studies have shown that recreation of interspecific hybridisation events that occurred in nature or development of new combinations of species result in hybrids with a phenotypic landscape beyond that of the strains usually employed in industrial fermentations [78–80]. Some of these hybrids produce new aromas that cater to changing consumer demands and demonstrate superior performance in challenging production environments. Detailed insight into how these hybrids evolved in different niches and which phenotypic features are retained from the parental species can aid in selecting suitable strains for development of industrially relevant hybrids.

8.5 Supplementary figures

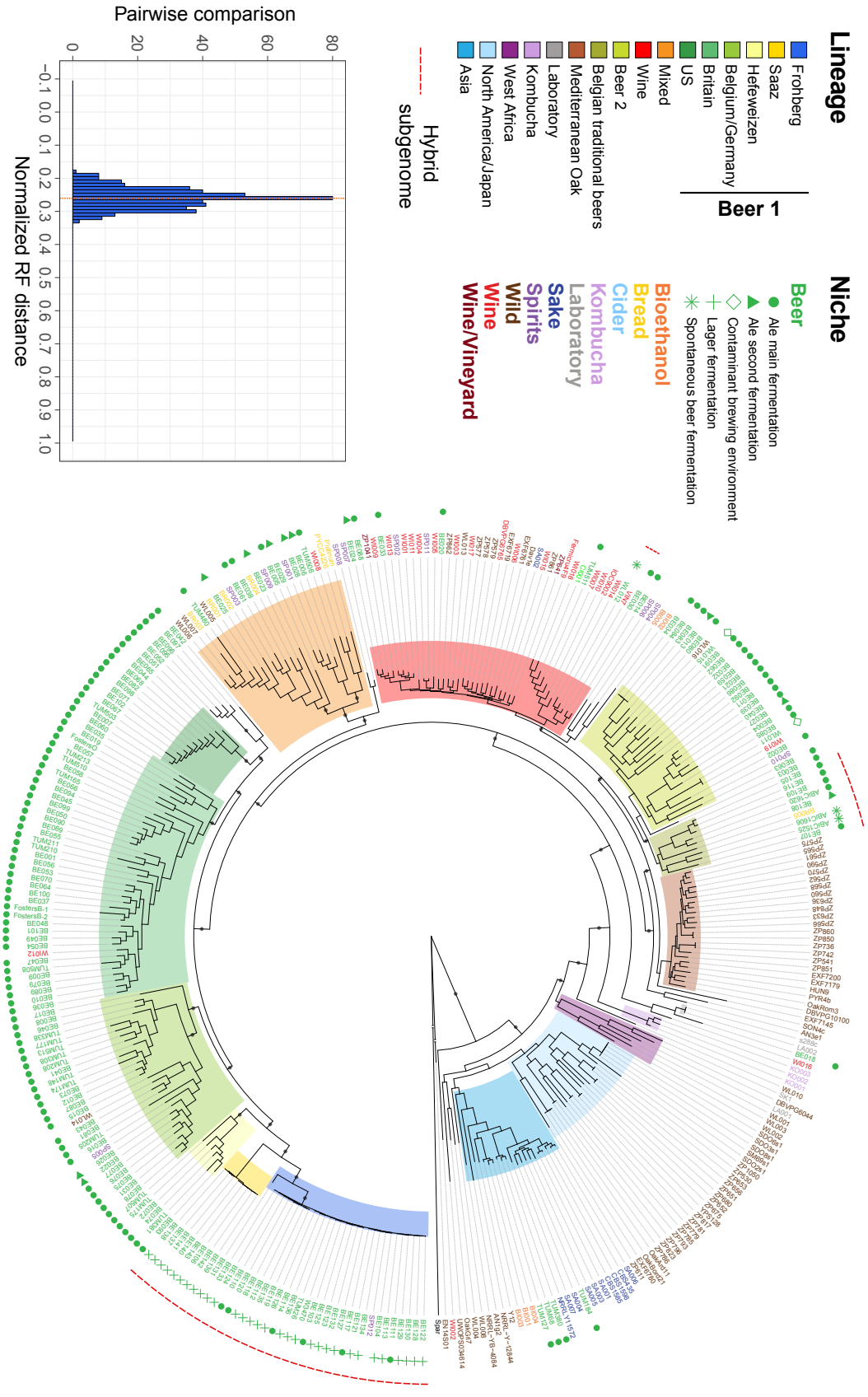


Figure 8.14: Detailed phylogeny of *S. cerevisiae* genomes and subgenomes. Best maximum likelihood (ML) phylogenetic tree inferred for the *S. cerevisiae* (sub)genomes. Color codes indicate origin (font) and lineage (bands). Origin of beer strains is additionally dissected in multiple categories (shape). Bar graphs (bottom left of the tree) represent pairwise comparisons across all the best scoring ML trees obtained through multiple ML searches expressed as Robinson-Foulds (RF) distance. Deep branches supported by more than 70% of bootstrap replicates are indicated by a grey dot. The number of bootstrap replicates is based on the majority rule (MR) bootstopping criterion.

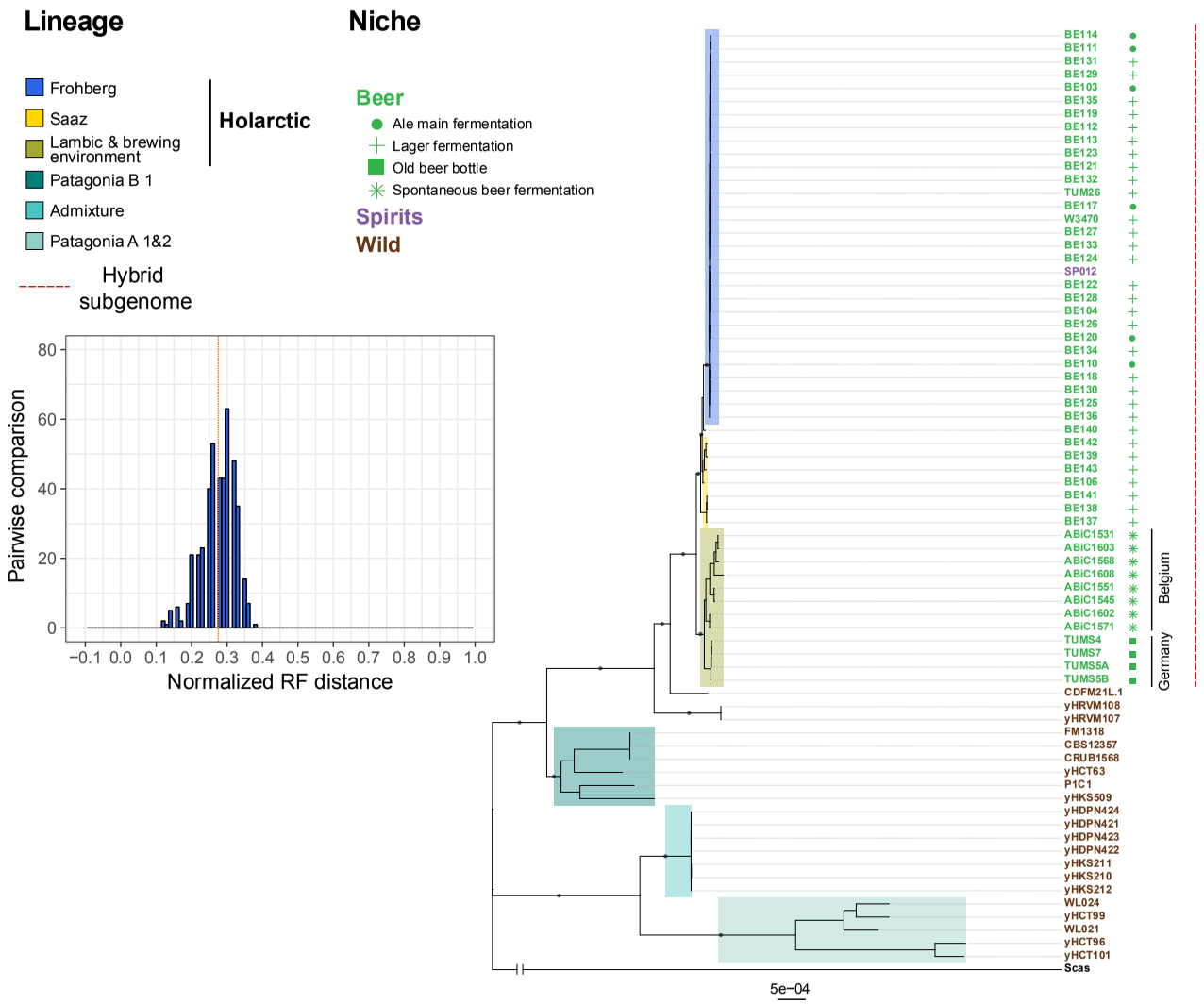


Figure 8.15: Detailed phylogeny of *S. eubayanus* genomes and subgenomes. Best maximum likelihood (ML) phylogenetic tree inferred for the *S. eubayanus* (sub)genomes. Color codes indicate origin (font) and lineage (bands). Origin of beer strains is additionally dissected in multiple categories (shape). Bar graphs (bottom left of each tree) represent pairwise comparisons across all the best scoring ML trees obtained through multiple ML searches expressed as Robinson-Foulds (RF) distance. Deep branches supported by more than 70% of bootstrap replicates are indicated by a grey dot. The number of bootstrap replicates is based on the majority rule (MR) bootstopping criterion.

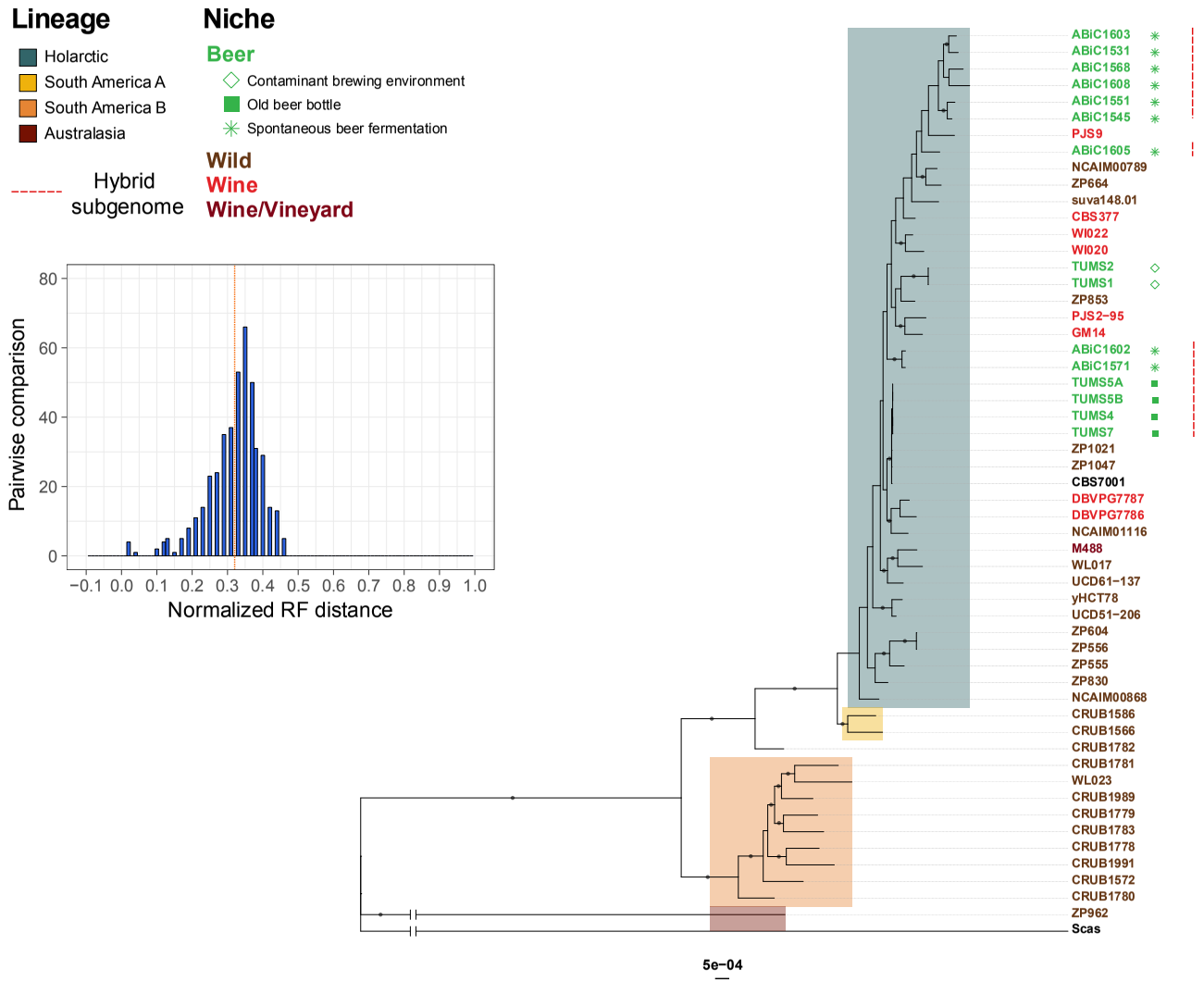


Figure 8.16: Detailed phylogeny of *S. uvarum* genomes and subgenomes. Best maximum likelihood (ML) phylogenetic tree inferred for *S. uvarum* (sub)genomes. Color codes indicate origin (font) and lineage (bands). Origin of beer strains is additionally dissected in multiple categories (shape). The location of subgenome taxa is highlighted by red, dashed lines. Bar graphs (bottom left of the tree) represent pairwise comparisons across all the best scoring ML trees obtained through multiple ML searches expressed as Robinson-Foulds (RF) distance. Deep branches supported by more than 70% of bootstrap replicates are indicated by a grey dot. The number of bootstrap replicates is based on the majority rule (MR) bootstopping criterion.

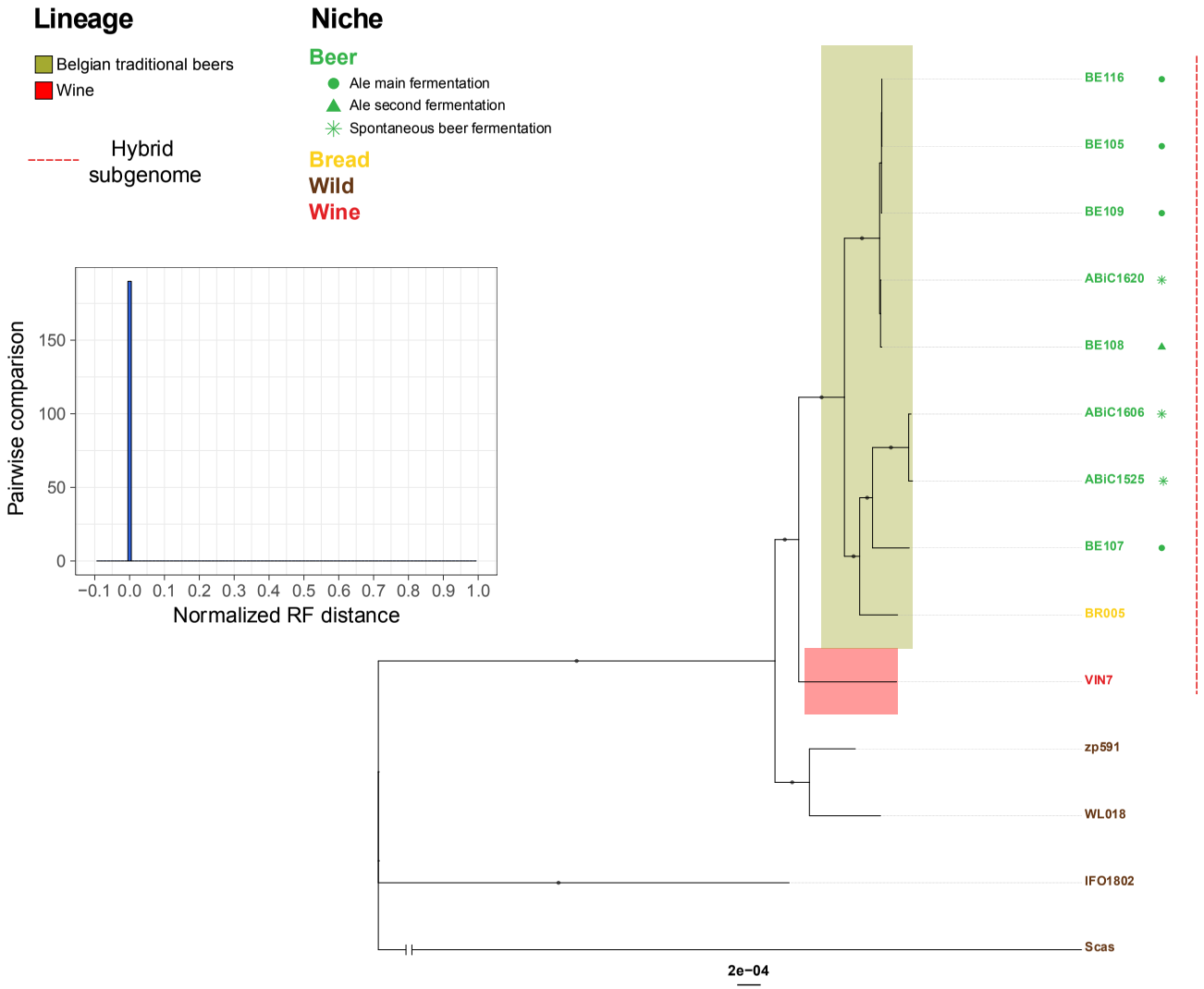


Figure 8.17: Detailed phylogeny of *S. kudriavzevii* genomes and subgenomes. Best maximum likelihood (ML) phylogenetic tree inferred for the *S. kudriavzevii* (sub)genomes. Color codes indicate origin (font) and lineage (bands). Origin of beer strains is additionally dissected in multiple categories (shape). Bar graphs (bottom left of the tree) represent pairwise comparisons across all the best scoring ML trees obtained through multiple ML searches expressed as Robinson-Foulds (RF) distance. Deep branches supported by more than 70% of bootstrap replicates are indicated by a grey dot. The number of bootstrap replicates is based on the majority rule (MR) bootstopping criterion.

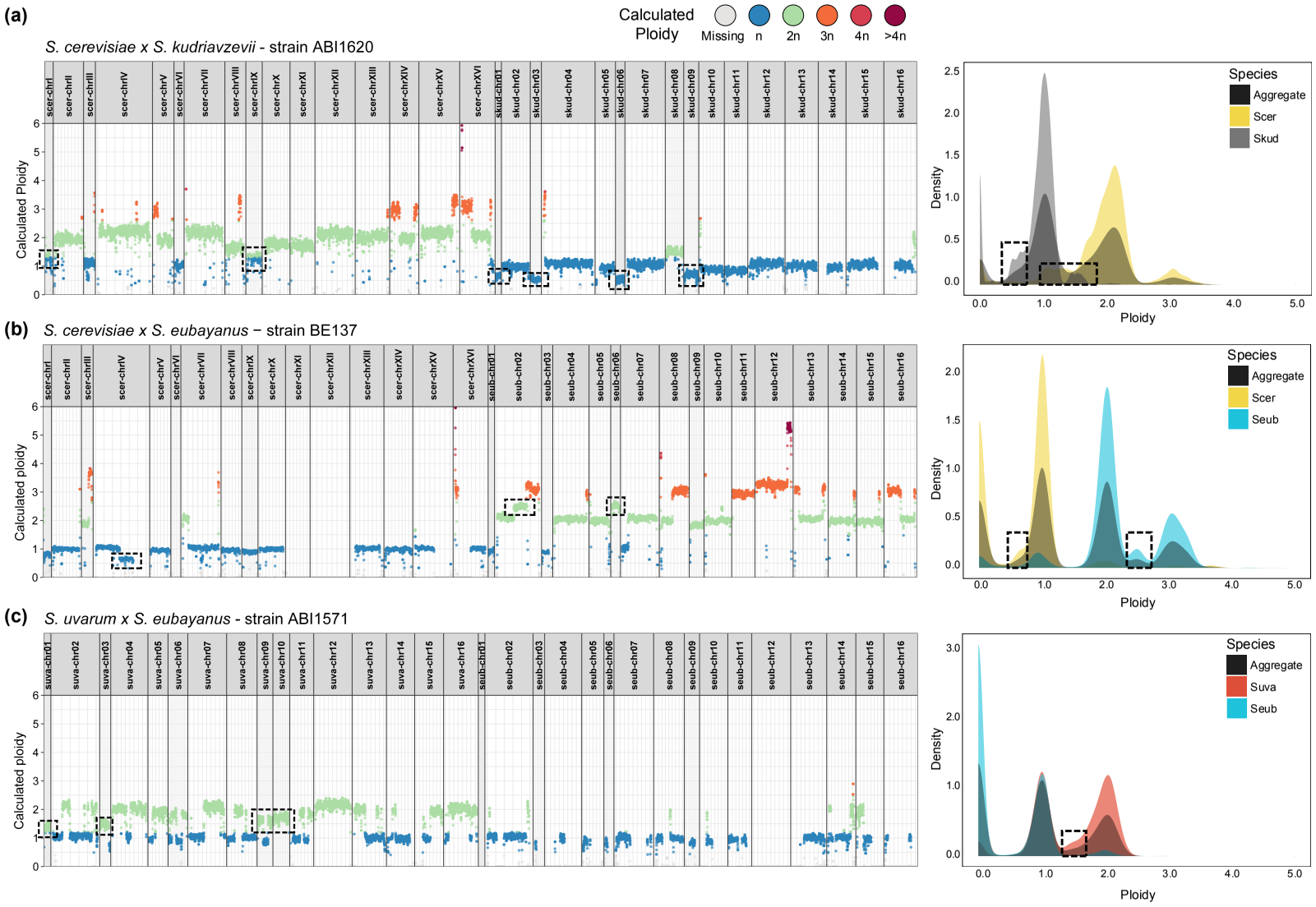


Figure 8.18: Genomic instability and copy-number heterogeneity in the sequenced interspecific hybrids. Examples of copy-number heterogeneity in three representative strains for each hybrid type: **a:** *S. cerevisiae* x *S. kudriavzevii* (strain ABI1620); **b:** *S. cerevisiae* x *S. eubayanus* (strain BE137); **c:** *S. uvarum* x *S. eubayanus* (strain ABI1571). Left: chromosomal level ploidy calculated in 1kb regions based on read depth (see methods). Right: total density of calculated ploidy per subgenome and aggregated for the full hybrid. Black, dashed squares highlight the presence of regions with an intermediate copy number level relative to the calculated distribution, , which indicate copy number heterogeneity in the population of cells sequenced for a particular hybrid.

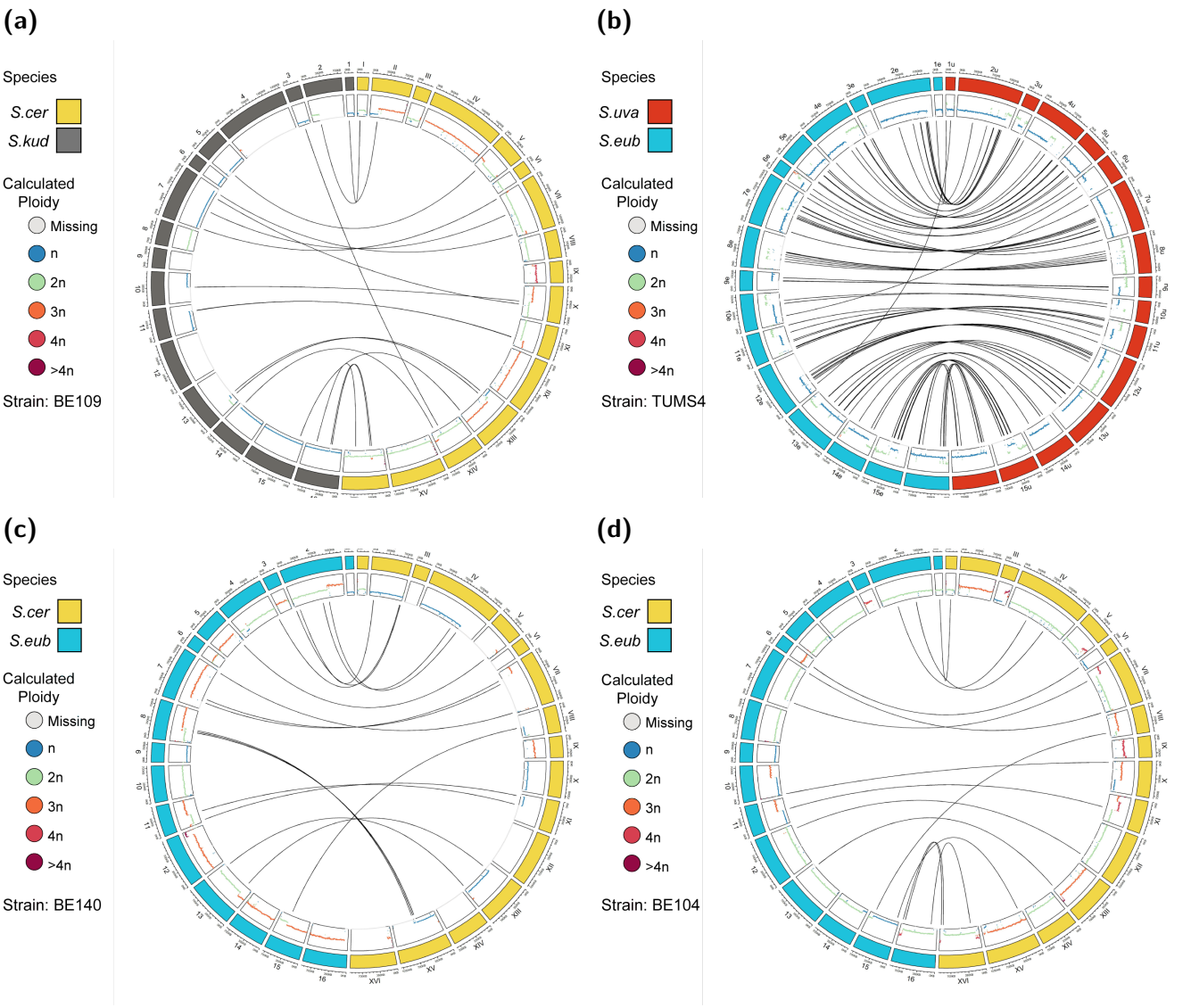


Figure 8.19: Detailed view of chimeric breakpoints identified in four representative hybrids. a: *S. cerevisiae* x *S. kudriavzevii* hybrid (strain BE109); **b:** *S. uvarum* x *S. eubayanus* (strain TUMS4); **c:** *S. cerevisiae* x *S. eubayanus* - Saaz (strain BE140); **d:** *S. cerevisiae* x *S. eubayanus* - Frohberg (strain BE104); Outer tracks of the circular plots represent reference chromosomes colored according to the species. Inner tracks represent chromosomal ploidy levels of the hybrid genome for each reference chromosome. Black lines connect chimeric breakpoint locations across two subgenomes.

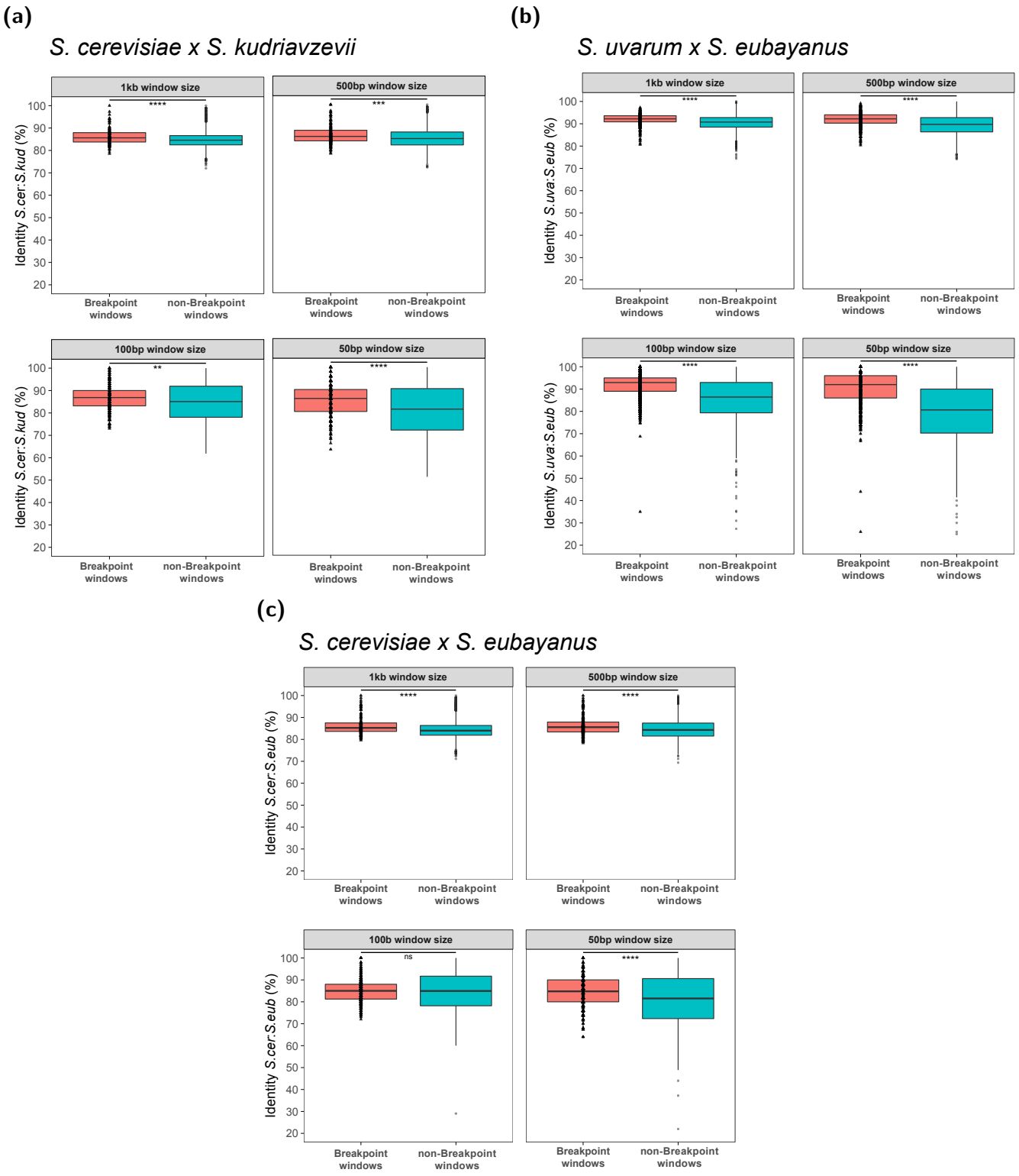


Figure 8.20: Chimeric breakpoints overlap low nucleotide identity regions between subgenomes. Nucleotide identity (%) between species (subgenomes) calculated in 1kb, 500b, 100bp and 50bp windows within **a:** *S. cerevisiae* x *S. kudriavzevii* hybrids, **a:** *S. uvarum* x *S. eubayanus* hybrids and **a:** *S. cerevisiae* x *S. eubayanus* hybrids. Box plots show the comparison between windows that contain chimeric breakpoints and windows that do not contain chimeric breakpoints. Statistical significance from Wilcoxon Signed Rank Test: ns (P-value $>$ 0.05), * (P-value $\leq$ 0.05), ** (P-value $\leq$ 0.01), *** (P-value $\leq$ 0.001) and **** (P-value $\leq$ 0.0001).

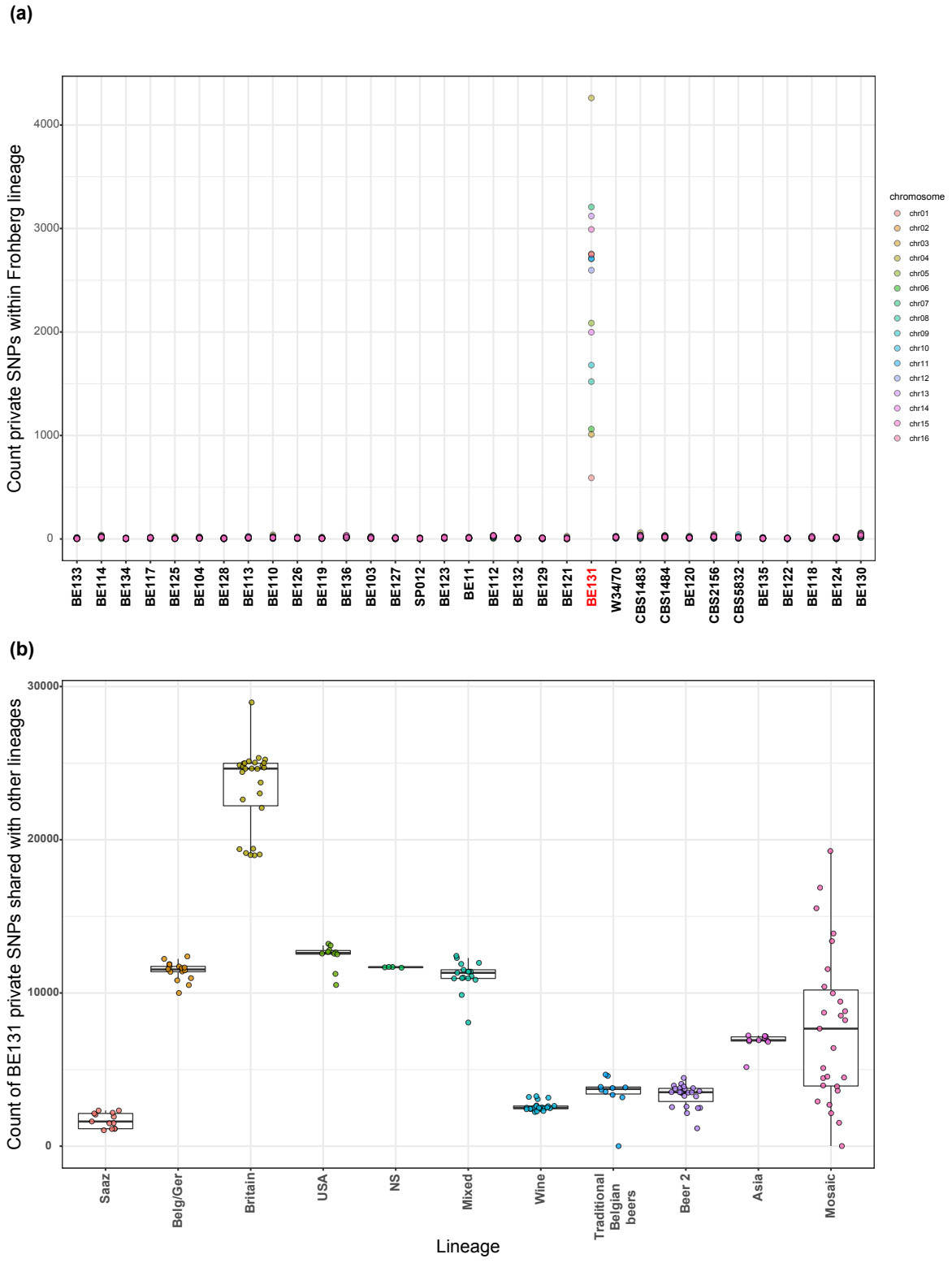


Figure 8.21: Potential introgressive hybridization in a *S. cer* x *S. eub* strain. (a) Strain BE131 (red font) harbors a significant amount of private SNPs compared to closely related strains within the same lineage (Frohberg) in its *S. cerevisiae* subgenome **(b)** Private SNPs from strain BE131 are shared by strains belonging to other *Saccharomyces* lineages and most notably to strain in the "Britain" clade, suggesting a potential introgressive hybridization with a strain belonging to this lineage. SNP count is dissected per chromosome in both panels. NS: Not specified.

8.6 Methods

8.6.1 gDNA Extraction

Genomic DNA was prepared using the GENTRA PUREGENE Yeast KIT (Qiagen, Germany) with some modifications to the recommended protocol. The main modification involves a 2 hours treatment of the overnight cell culture with zymolyase to efficiently digest yeast cell wall. Final DNA concentrations were measured using Qubit (Thermo Fisher Scientific, USA), 260/230 and 260/280 ratios with Nanodrop (Thermo Fisher Scientific, USA).

8.6.2 Library prep and Whole Genome Sequencing

Paired-end sequencing libraries (2 x 300bp, v3 reagent kit, 600 cycles) with a mean insert size of 300bp were prepared and run according to the manufacturer's instructions on an Illumina MiSeq at the Nucleomics Core facility in Leuven (http://www.nucleomics.be/).

8.6.3 Reference-based alignments and variant calling

Reads were pre-processed by filtering low-quality and ambiguous reads, adapters and PhiX contaminations, using Trimmomatic (v0.30). Clean reads were mapped to 7 *Saccharomyces* species concatenated into a single a fasta file: *Saccharomyces cerevisiae* reference genome S288c (R64-1-1, EF4-Ensemble Release 74), *Saccharomyces paradoxus* (strain YP138, [81]), *Saccharomyces mikatae* (strain IFO 1815, [81]), *Saccharomyces kudriavzevii* (IFO 1802, [81]), *Saccharomyces arboricola* (type strain H-6^{T}; CBS 10644^{TT}, [82]), *Saccharomyces eubayanus* (strain FM1318, [56]), *Saccharomyces uvarum* (CBS 7001, [81]) with the Burrows-Wheeler Aligner (BWA-aln) using default parameters except for -q 10. Non-primary alignments were filtered out and duplicate reads were marked using Picard Tools (v1.56) (http://picard.sourceforge.net). Coverage was estimated based on read depth in non-overlapping 1kb windows (reported mean coverage per window). using bedtools [83].

8.6.4 De novo assembly

For each library, low quality and ambiguous reads were trimmed using Trimmomatic (v0.30). Reads were error-corrected with SPAdes (v3.10.1) and subsequently used for the de novo assembly with SPAdes (v3.10.1). Next the redundans pipeline (v1.2) [84] was used to detect and remove redundant contigs and perform scaffolding and gap closing based on paired reads information. In order to determine the coordinates of contigs from each newly assembled strain relative to the reference strain and to obtain pseudo chromosomes, whole genome alignments were performed against the species identified within each hybrid strain using Ragout (v1.2) [85]. To confirm the correct order and placement of contigs and to detect introgressions, pseudo-chromosomes and

unplaced contigs were blasted against a database including the 7 sensu stricto species using blast (v2.2.30+).

8.6.5 **Phylogenomic analyses**

In order to infer the origin of the hybrid *Saccharomyces* genomes and their genetic relationship across species and strains within the *Saccharomyces* species complex, single-copy genes across all strains and species have been identified. The starting set of genes included 4722 1:1:1:1 orthologs among *S. cerevisiae* s288c, *S. paradoxus* YP138, *S. kudriavzevii* IFO1802 and *S. uvarum* CBS7001 identified by Scannell and co-autohors [81]. The starting set of genes was reduced to 4125 genes after including *S. eubayanus* FM1318 [56] and the outgroup species, *Saccharomyces castellii* [86]. Next, the presence of these genes and their single-copy status was tested within a collection of 423 *Saccharomyces* isolates, in order to identify the set of genes present in a single copy in all isolates. The collection of strains investigated included 286 *S. cerevisiae* isolates, 2 (check) *S. kudriavzevii* isolates, 45 *S. uvarum* isolates, 20 *S. eubayanus* isolates, 10 *S. cerevisiae* x *S. kudriavzevii* hybrids, 47 *S. pastorianus* hybrids and 13 *S. uvarum* x *S. eubayanus hybrids* **(8.1, p. 195)**. The sub-genomes within the hybrid isolates were considered as distinct species: for instance, for a *S. uvarum x S. eubayanus* hybrid, two sets of orthologs were identified, one for each species respectively. A local BLAST database was set up for all the genomes assembled (see earlier) complemented with publicly available genomes **(8.1, p. 195)** and BLASTN searches were performed (1E-04 E-value cut-off, >98% similarity and >85% coverage blast v2.5.0+) for each pure species using the set of genes identified in the previous step. For hybrid genomes, an additional BLASTN step was implemented to compare the set of genes identified for the distinct parental species and exclude genes with high similarity between species that cannot be unequivocally assigned to one or the other species. Five set of genes were obtained: 1) 1389 genes across the *S. cerevisiae* genomes and subgenomes, dubbed "*S. cer*" set; 2) 1571 genes across the *S. eubayanus* genomes and subgenomes, dubbed "*Seub*" set; 3) 1364 genes across *S. uvarum* genomes and subgenomes, dubbed "*S. uva*" set; 4) 1750 genes across the *S. kudriavzevii* genomes and subgenomes, dubbed "*S. kud*" set. Considering the high level of species-specific subgenome loss and fragmentation observed in some *Saccharomyces* hybrids, and in order to maximize the number of isolates included, 3% of missing data per gene was allowed. Extreme cases with >50% missing genes per strain were excluded from the analysis. Nucleotide-level multiple sequence alignments (MSAs) for each gene in each set identified above were generated using MAFFT (v7.187), with default settings and 1,000 refinement iterations. The MSAs for each set were concatenated into a supermatrix using FASconCAT (v1.0) (https://www.zfmk.de/en/research/research-centres-and-groups/fasconcat). Quality checks and format conversions were performed using trimAl (version 1.2) [87]. The final *S. cer* supermatrix included 337 taxa and 1,556,065 positions, 97.345% nucleotides, 2.655% gaps and 0% ambiguities. The final *S. eub* supermatrix included 81 taxa and 2,323,546 positions, 88.141%

nucleotides, 11.859% gaps and 0% ambiguities. The final *S. uva* supermatrix included 55 taxa and 1,453,393 positions, 91.772% nucleotides, 8.228% gaps and 0% ambiguities. The final *S. kud* supermatrix included 14 taxa and 2,635,158 positions, 90.741% nucleotides, 9.259% gaps and 0% ambiguities. Within each supermatrix, each gene was considered as separate data partition. Twenty-five random starting trees for the *S. cer* supermatrix and 20 random starting trees for the *S. eub*, *S. uva*, *S. kud* supermatrices were obtained using Randomized Axelerated Maximum Likelihood (RAxML, v 8.2.8;[88]). Maximum-likelihood (ML) tree searches were performed on each random starting tree under the GTRGAMMA model (4 discrete rate categories) using Exascale Maximum Likelihood (ExaML v 3.0.17;[89]) and the rapid hill-climbing algorithm (-f d). During the ML search, the alpha parameter of the model of rate heterogeneity and the rates of the GTR model of nucleotide substitution were optimized independently for each partition. The branch lengths were optimized jointly across all partitions. For each starting tree, the best tree was selected based on the highest log-likelihood score. The tree with the best overall log-likelihood score of all tree inferences was considered the final ML tree. For a given supermatrix, non-parametric bootstrap analysis was performed for each final tree using RaxML on the corresponding supermatrix. The a posteriori boot-stopping criterion [90] (MR bootstrapping convergence criterion) was applied to define the number of replicates. After every 50 replicates, the set of bootstrapped trees generated so far is repeatedly (1000 permutations) split in two equal subsets, and the Weighted Robinson-Foulds (WRF) distance is calculated between the extended-majority-rule consensus trees of both subsets (for each permutation). Low WRF distances ($< 3\%$) for $>= 99\%$ of permutations were used to indicate bootstrapping convergence. Convergence was reached after 200 replicates for the *S. cer* phylogeny: average WRF = 2.33%, percentage of permutations in which the WRF was <=3.00 = 997; 300 replicates for the *S. eub* phylogeny: average WRF = 1.85%, percentage of permutations in which the WRF was <=3.00 = 998; 600 replicates for the *S. uva* phylogeny: average WRF = 1.36%, percentage of permutations in which the WRF was <=3.00 = 992; 50 replicates for the *S. kud* phylogeny: average WRF = 0.26%, percentage of permutations in which the WRF was <=3.00 = 1000. The final trees were visualized and rooted in R (ggtree package) using *S. paradoxus* as outgroup for the *S. cer* tree and *S. castellii* for the other trees.

8.6.6 **Divergence-time estimation**

Diverge-time inference was performed for the "Beer 1" clade on the *S. cer* phylogenetic tree using the best scoring ML tree obtained for a supermatrix of 1389 protein coding genes as described in section 8.6.5. To provide this tree with divergence time estimates, we assessed the performance of several molecular clock models using BEAST 1.10 [77] in combination with BEAGLE 3 [91]. Specifically, we analysed the data using an uncorrelated relaxed clock with an underlying lognormal distribution [92], a random local clock [93] and a fixed local clock with predefined clades (Lager-Frohberg, Lager-Saaz, Ale-British, Ale-Belgium/Germany, Ale-United States; [94]). A pure-birth

Yule speciation prior and a random starting tree were used for the Bayesian inference analyses. Additionally, to date our phylogeny we used a calibration prior on the split between US and British beer yeasts, using a normal prior with a 99% confidence interval falling between 1607 and 1637 AD, as based on historical events [33]. Each analysis was run until ESS values of at least 100 could be obtained for all relevant parameters, as computed by Tracer 1.7 [95]. Of these models, the random local clock provided a significantly better fit to the data than its two competing model, as estimated using generalized stepping-stone sampling [96]. The uncorrelated relaxed clock model performed the worst among the models tested, indicating that a single rate distribution is not able to account for shifts in evolutionary rates having occurred. Assuming a strict clock rate within each predefined clade is equally restrictive, which explains why the random local clock is able to significantly outperform the fixed local clock model. Given these findings, we present results for the random local clock model, by summarizing its divergence time estimates in a maximum clade credibility tree as obtained through TreeAnnotator, which is part of the BEAST software package [77].

8.6.7 **Estimation of ploidy and identification of chimeric regions across species**

The starting files for this analysis were the alignment of paired-end reads on the multi-species reference genome, obtained in section 8.6.3. A new alignment file was obtained for each strain by retaining paired reads with high mapping quality (> 20), for which the two mates are mapping on homeologous/non-homeologous chromosomes of the different subgenomes. E.g. for read pair A, one read is mapped on *S. cerevisiae* chr I and its mate on *S. eubayanus* chr 1. We dubbed these reads "discordant reads". The absolute number of discordant reads over the total amount of reads was calculated in non-overlapping 1k windows over the full multi-species reference genome excluding unplaced contigs and mitochondrial contigs. In order to identify windows in the genome supporting the presence of a chimeric event, we selected windows with at least 15 reads and a minimum of 3 discordant reads. This very conservative threshold allows the identification of potential chimeric events across areas of the genome with varying coverage levels within the same hybrid-type and across hybrids with different genome size and hence varying coverage levels. Windows with a minimum of 3 discordant reads were defined "breakpoint windows"; breakpoint windows preceded and/or followed by a breakpoint window were categorized as "major breakpoint" windows, because of the presence of consecutive windows supporting a chimeric event. Since breakpoints windows often coincided with changes in ploidy, we calculated simultaneously absolute ploidy level and occurrence of discordant reads in each 1k window along the genome. First, the mean raw coverage per 1k non-overlapping windows as calculated in section 8.6.3 is smoothed using a running median function on 11 consecutive 1k windows (CaTools package v1.6 implemented in R). Second, the density of the smoothed coverage is plotted per subgenome and for the full hybrid genome. In order to map the raw coverage values representing major shifts in ploidy levels, a Gaussian mixture model

is fitted to the coverage density distribution and mean and standard deviation are calculated for each peak (peaks are proxies for ploidy levels detected in the hybrid genome). Third, the smoothed raw coverage values are divided by the mean value of the haploid peak to obtain estimated ploidy values for each pre-computed window in the genome. Due to the presence of noise and potential copy number heterogeneity in the population of cells sequenced, the absolute copy values do not always fall within the expected intervals and "in-between" ploidy levels were detected **(Figure 8.18, p. 182)**. In order to define ploidy intervals on a per-strain basis, a second Gaussian mixture model was fitted on the distribution ploidy values. This allowed the identification of mean and standard deviation of each ploidy peak and the definition of "ploidy shift" intervals based on the intersection point between two different ploidy density distributions in the mixture model. The identification of breakpoint windows and their ploidy context was then followed by the identification of the exact mapping location of the reads within the breakpoint window across the two subgenome for each hybrid strain based on the following steps: i) extract the reads from the selected breakpoint windows; ii) retain reads that are still in pair after the identification and selection of breakpoint windows; iii) calculate for each pair their summed edit distance from the corresponding reference sequence and normalise it by the summed length of the two reads; iv) select reads with a percentage identity >=95% against the corresponding reference location; iv) intersect the position of the reads with annotated features from the respective reference sequence **(Table 8.2, p. 195)**. In order to calculate nucleotide identity (%) between subgenomes for breakpoint windows and non-breakpoint windows, we obtained pairwise whole-genome alignments between *Saccharomyces* species corresponding to the species combinations identified in our set of hybrids using Mugsy (v1.2.3) [97]. Next, we mapped the 1kb intervals on the whole-genome alignments and we calculated percentage identity as number of matches on the total amount of bases in the window excluding gaps. We repeated the calculations for 500, 100 and 50bp windows in order to identify microhomology regions within the initial 1kb windows. Then, we compared the identity distribution across breakpoint windows and non-breakpoint windows **(Figure 8.20, p. 184)**.

8.6.8 Flavor Production and Flocculation in Fermentation Conditions

To assess the production of aroma-active compounds, lab-scale fermentation experiments were performed. These fermentations were performed in rich growth medium (YPGlu 10%; peptone 2% w v^{-1}, yeast extract 1% w v^{-1}, glucose 10% w v^{-1})and beer wort (13 °P, Brouwland malt extract). Yeast precultures were shaken overnight at 30 °C in test tubes containing 5mL of yeast extract (1% w v^{-1}), peptone (2% w v^{-1}) and glucose (4% w v^{-1}) medium (YPGlu 4%). After 16 hr of growth, 0.5 mL of the preculture was used to inoculate 50 mL of YPGlu 4% medium in 250 mL Erlenmeyer flasks, and this second preculture was shaken at 30 °C for 16 hr. This second preculture was used for inoculation of the fermentation medium (YPGlu 10%) at an initial optical density (at 600nm; OD600) of 0.5, roughly equivalent to 107 cells mL^{-1}. The fermentations, performed

in 250 mL Schott bottles with a water lock placed on each bottle, were incubated statically for 7 days at 20 °C (YPD) or 16 °C (wort). Weight loss was measured daily to estimate fermentation progress. After 7 days, the fermentations were stopped, filtered (0.15 mm paper filter) and samples for chromatographic analysis, density and ethanol measurements were taken. Headspace gas chromatography coupled with flame ionization detection (HS-GC-FID) (Agilent Technologies, USA) was used for the quantification of yeast aroma production. The GC was calibrated for 16 important aroma compounds, including esters (ethyl acetate, isobutyl acetate, propyl acetate, isoamyl acetate, phenyl ethyl acetate, ethyl propionate, ethyl butyrate, ethyl hexanoate, ethyl octanoate, ethyl decanoate), higher alcohols (isoamyl alcohol, isobutanol, butanol, phenyl ethanol), acetaldehyde and 4-VG. The GC was equipped with a headspace autosampler (PAL system, CTC analytics, Switzerland) and contained a DB-WAXETER column (length, 30 m; internal diameter, 0.25 mm; layer thickness, 0.5 μm, Agilent Technologies, USA) and N2 was used as the carrier gas. Samples were heated for 25 min at 70 °C in the autosampler. The injector block and FID temperatures were both kept constant at 250 °C. Samples of 5 mL filtered fermentation medium were collected in 15 ml glass tubes containing 1.75 g of sodium chloride each. These tubes were immediately closed and cooled to minimize evaporation of volatile compounds. The oven temperature was held at 50 °C for 5 min, after which it increased to 80 °C at 4 °C min^{-1}. Next, it increased to 200 °C at 5°C min^{-1} and was held at 200 °C for 3 min. Results were analyzed with the Agilent Chemstation software (Agilent Technologies, USA). Additionally, after fermentation, the flocculation character of each strain was scored visually from 1 (not flocculent) to 6 (extremely flocculent, big flocs).

8.6.9 **Ethanol Accumulation Capacity**

To assess the maximal ethanol accumulation capacity of all strains, fermentation tests were performed in rich medium containing 35% w v^{-1} glucose. Yeast precultures were shaken overnight at 30 °C in test tubes containing 5 mL YPGlu 4%. After 16 hr of growth, 0.5 mL of the preculture was used to inoculate 50 mL of YPGlu 4% medium in 250 mL Erlenmeyer flasks, and this second preculture was shaken at 30 °C for 16 hr. This preculture was used for inoculation of the fermentation medium (peptone 2% w v^{-1}, yeast extract 1% w v^{-1}, glucose 35% w v^{-1}; YPGlu 35%) at an initial OD600 of 0.5, roughly equivalent to 107 cells mL^{-1}. The fermentations, performed in 250 ml Schott bottles with a water lock placed on each bottle, were incubated statically for 14 days at 30 °C. Weight loss was measured daily to estimate fermentation progress. After 14 days, the fermentations were stopped, filtered (0.15 mm paper filter) and samples for ethanol measurements [performed with the Alcolyzer Beer DMA 4500M (Anton Paar, Austria)] were taken.

8.6.10 **Screening for Environmental and Nutrient Stress Tolerance**

All strains were tested in robot-assisted, high-throughput spotting assays in several conditions. All isolates were evaluated on YPGlu 2% agar (Yeast Extract 1% w v^{-1}, Peptone 2% w v^{-1},

Glucose 2% w v^{-1}, agar 2% w v^{-1}) for (i) temperature tolerance (4 °C, 12 °C, 16 °C, 30 °C, 37 °C, 39 °C), (ii) sugar- and/or osmotolerance using increasing concentrations of glucose (final osmolyte concentration of 44 - 46 - 48 w v^{-1}), (iii) acid tolerance using increasing concentrations of acetic acid (12.5 - 25 - 50 - 75 mM), (iv) ethanol tolerance using increasing concentrations of ethanol (5 - 7 - 9 - 10 - 11 - % v v^{-1}), and (v) copper tolerance using 0.050-0.075-0.100 mM copper. Stressor concentrations were selected based on previous pilot experiments with a broader concentration range (data not shown). For each of these experiments, growth on YPGlu 2% agar on 20 °C was used as a control condition. Prior to the experiment, the 96-well microtiter plates containing the isolates (stored at -80 °C) were thawed and spotted on YPGlu 2% agar and incubated at 30 °C for 48 hr. Next, 96-well plates containing 150 μl of YPGlu 2% in each well were inoculated with the isolates and incubated overnight at 30 °C on a microtiter plate shaking platform (Heidolph Instruments, Germany) at 600rpm, allowing the cells to reach stationary phase. Then, the OD600 of all wells was measured using a microtiter plate reader (Molecular Devices, Sunnyvale, USA). Subsequently, the cell density was manually adjusted to OD600 ~0.1 in a second 96-well microtiter plate using sterile deionized water in order to standardize the starting cell density for all isolates. This plate was used as the source plate for spotting the test media. In order to maximize throughput and reproducibility, a high-density array robot (Singer Instruments, UK) was used for all spotting or replication steps. After spotting, all plates were sealed using plastic paraffin film and all plates (except for plates used in the thermotolerance assays) were incubated at 20 °C. After 4-14 days of incubation (depending on the experiment), all plates were scanned using a high-definition scanner (Seiko Epson, Japan). Scanned images were processed using ImageJ combined with the ScreenMill macro [98]. Data were processed by calculating relative growth compared to the control condition **(Table 8.3, p. 195)** . Heat maps were obtained using the R function heatmap.2 from gplots package (v3.0.1 ref). Strains were hierarchically clustered based on phenotypic behavior (z-scores - normalized values) using Ward's method on Euclidean distances, **(Figure 8.9, p. 168)**.

8.6.11 Investigation of the Yeast's Sexual Life Cycle

Sporulation was induced on minimal sporulation medium [1% (w v^{-1}) KAc, 0.05% (w v^{-1}) amino acids, 2% (w v^{-1}) agar] at 23 °C after pre-growth in YPGlu 2%. Tetrad dissection of 4 tetrads of each strains was carried out using a Singer micromanipulator (Singer Instruments, UK), and mating-type determination of all germinated spores was carried out by mating-type PCR.

8.6.12 Yeast survival in beer

To assess survival of all strains, the viability of cultures aging in Duvel Green [blond ale, 7% (v v^{-1}) ethanol] was tracked over a one month period. Yeast precultures were shaken for 48 hours at 30 °C in test tubes containing 5 mL YPGlu 2%. This sample was then transferred to 15 mL

falcons and centrifuged for 3 minutes at 3000 rpm. After discarding the supernatant, the samples were resuspended in sterile water to reach an initial cell count of 4 x 107 cells mL^{-1}. 0,5 mL of this sample was used to inoculate a sterile GC vial containing 10 mL of filter-sterilized Duvel Green supplemented with 0.2% (w w^{-1}) glucose to reach a final concentration of 2 x 106 cells mL^{-1}. The headspace of the samples was flushed with CO^2 before capping the vials. The vials were incubated statically at 30 °C. After 30 days, samples were taken, and viability was assessed using a methylene blue [0.1% (w v^{-1}), Sigma-Aldrich] staining. Automated cell counting was performed with the TC20 automated cell counter (Biorad, USA).

8.6.13 **Desiccation tolerance**

Desiccation tolerance was measured by a modified version of the assay described in Calahan et al. 2011 [99]. All strains were streaked on YPGlu 2% agar. Strains were pregrown in YPGlu 2%, supplemented with 0.5% v v^{-1} Tween80 and 20 μg mL^{-1} ergosterol in deep-well plates at 16 °C for 3 days. Cells were harvested by centrifugation (3000 rpm, 3 minutes) and supernatant was removed. 108 cells were resuspended in 1 mL assay buffer (x8 dilution of phosphate buffered saline: 17.1 mM NaCl, 0.338 mM KCl,, 1.25 mM Na2HPO4, 0.220 mM KH2PO4, pH 7.4), after which viability was checked using plating. 100 μL of this cell culture was transferred to a 96-well microtiter plate. The lid of the 96-well microtiter plate was lifted 1.7 cm (using 4 small pieces of cardboard) during the experiment to allow sufficient air flow. Plates were incubated for 7 days at 8 °C. Afterwards, cells were resuspended in 200 μL assay buffer and viability was assayed using plating. Next, the ratio between the viability after desiccation and the viability before desiccation was calculated. These data was subsequently binned to score desiccation tolerance (viability <1% = '1', viability 1-20% = '2', viability 20-50% = '3', viability >50% = '4'). All strains were tested in biological duplicates.

8.6.14 **Biomass production in cold wort**

To assess biomass production in beer wort at very low temperatures strains were pregrown in 100 μL of YPGlu+Mal 2% (Yeast Extract 1% w v^{-1}, Peptone 2% w v^{-1}, Glucose 1% w v^{-1}, maltose 1% w v^{-1}) at 16°C for 2 days. Next, 5 μL of these cultures were transferred to 95 μL wort (8 °P, 8 EBC Brewferm, Belgium) in a 96-well microtiter plate and incubated at 8 °P for 7 days on a microtiter plate shaking platform (Heidolph Instruments, Germany) at 600 rpm. Optical density of the strains was assessed after 0, 3 and 7 days. Experiments were performed in duplicate.

8.6.15 **Development and phenotypic evaluation of artificial hybrids**

Hybridisation was induced by placing single spores from both parental strains together with a micromanipulator (Singer instruments MSM, UK) on YPGlu 2% agar, followed by visual inspection of zygote formation after 6-8 hours of incubation at room temperature. Candidate interspecific

hybrids were purified by streaking on wort agar medium [12% w v^{-1} malt extract (8 EBC Brewferm, Belgium) and 1.5% w v^{-1} agar]. Hybrids were confirmed through a species PCR [78, 100, 101]. PCR-confirmed interspecific hybrids were streaked another three consecutive times on 12°P wort agar prior to long term storage at -80 °C to ensure strain purity. Three independent hybrids were developed from the same two parents (BE014 and WL021). These hybrids were tested for biomass during wort fermentation at cold temperature as described in the previous section. The maltotriose that was still present after 7 days of fermentation was measured using the Dionex system (ICS 5000+, Thermo Fisher Scientific, USA). Fermentation samples were diluted in dH2O, filtered (0.2 μm), and 10 μL were injected into the system. A Carbopac PA20 column, kept at 30 °C, was used, with a column flow of 0.3 μl min^{-1}. 250mM NaOH (Eluent 1) and 500 mM CH3COONa + 100 mM NaOH (Eluent 2) were used as eluents.

8.7 Supplementary tables

Table 8.1: Additional Information on the *Saccharomyces* strains investigated in this chapter.

This table is provided at:
https://www.dropbox.com/s/vjjjkyjbmv1q31d/Table%208.1.xlsx?dl=0

Table 8.2: Detailed overview of chimeric breakpoints identified in each hybrid. Sheet1: *S.cer* x *S.kud* hybrids; sheet2: *S.uva* x *S.eub* hybrids; sheet3: *S.cer* x *S.eub* hybrids. Related to **Figures 8.6, p. 163, 8.7, p. 164, 8.8, p. 165** and **Figures 8.19, p. 183, 8.20, p. 184**.

This table is provided at:
https://www.dropbox.com/s/dcgpm3z4bm4mbsu/Table%208.2.xlsx?dl=0

Table 8.3: Phenotypic variation across *Saccharomyces* hybrids and *Saccharomyces* pure species for 76 traits. Related to **Figure 8.9, p. 168**.

This table is provided at:
https://www.dropbox.com/s/9q48okepa6gcseb/Table%208.3.xlsx?dl=0

8.8 References

[1] J. Mallet, N. Besansky, and M. W. Hahn, "How reticulated are species?," *BioEssays*, vol. 38, no. 2, pp. 140–149, 2016.

[2] B. J. Shapiro, J.-B. Leducq, and J. Mallet, "What Is Speciation?," *PLOS Genetics*, vol. 12, no. 3, p. e1005860, 2016.

[3] R. Abbott, D. Albach, S. Ansell, J. W. Arntzen, S. J. Baird, N. Bierne, J. Boughman, A. Brelsford, C. A. Buerkle, R. Buggs, R. K. Butlin, U. Dieckmann, F. Eroukhmanoff, A. Grill, S. H. Cahan, J. S. Hermansen, G. Hewitt, A. G. Hudson, C. Jiggins, J. Jones, B. Keller, T. Marczewski, J. Mallet, P. Martinez-Rodriguez, M. Möst, S. Mullen, R. Nichols, A. W. Nolte, C. Parisod, K. Pfennig, A. M. Rice, M. G. Ritchie, B. Seifert, C. M. Smadja, R. Stelkens, J. M. Szymura, R. Väinölä, J. B. Wolf, and D. Zinner, "Hybridization and speciation," *Journal of Evolutionary Biology*, vol. 26, no. 2, pp. 229–246, 2013.

[4] E. J. Baack and L. H. Rieseberg, "A genomic view of introgression and hybrid speciation.," *Current opinion in genetics & development*, vol. 17, no. 6, pp. 513–8, 2007.

[5] A. W. Nolte and D. Tautz, "Understanding the onset of hybrid speciation," *Trends Genet.*, vol. 26, no. 2, pp. 54–58, 2010.

[6] L. H. Rieseberg, M. A. Archer, and R. K. Wayne, "Transgressive segregation, adaptation and speciation," *Heredity (Edinb)*, vol. 83 (Pt 4), pp. 363–372, 1999.

[7] S. Lamichhaney, F. Han, M. T. Webster, L. Andersson, B. R. Grant, and P. R. Grant, "Rapid hybrid speciation in Darwin's finches.," *Science (New York, N.Y.)*, vol. 359, no. 6372, pp. 224–228, 2018.

[8] L. H. Rieseberg, "Humoploid Reticulate Evolution in *Helianthus* (*Asteraceae*): Evidence from Ribosomal genes," *American Journal of Botany*, vol. 78, no. 9, pp. 1218–1237, 1991.

[9] L. H. Rieseberg, S.-C. Kim, R. A. Randell, K. D. Whitney, B. L. Gross, C. Lexer, and K. Clay, "Hybridization and the colonization of novel habitats by annual sunflowers.," *Genetica*, vol. 129, no. 2, pp. 149–65, 2007.

[10] A. E. Schwarzbach and L. H. Rieseberg, "Likely multiple origins of a diploid hybrid sunflower species.," *Molecular ecology*, vol. 11, no. 9, pp. 1703–15, 2002.

[11] L. P. Pryszcz, T. Németh, E. Saus, E. Ksiezopolska, E. Hegedűsová, J. Nosek, K. H. Wolfe, A. Gacser, and T. Gabaldón, "The Genomic Aftermath of Hybridization in the Opportunistic Pathogen *Candida metapsilosis*," *PLOS Genetics*, vol. 11, no. 10, p. e1005626, 2015.

[12] M. Kellis, N. Patterson, M. Endrizzi, B. Birren, and E. S. Lander, "Sequencing and comparison of yeast species to identify genes and regulatory elements," *Nature*, vol. 423, no. 6937, pp. 241–254, 2003.

[13] D. Greig, "Reproductive isolation in *Saccharomyces*," *Heredity*, vol. 102, no. 1, pp. 39–44, 2009.

[14] G. Liti, D. B. H. Barton, and E. J. Louis, "Sequence diversity, reproductive isolation and species concepts in *Saccharomyces*.," *Genetics*, vol. 174, no. 2, pp. 839–50, 2006.

[15] J. Hou, A. Friedrich, J. De Montigny, and J. Schacherer, "Chromosomal rearrangements as a major mechanism in the onset of reproductive isolation in *Saccharomyces cerevisiae*," *Current Biology*, 2014.

[16] L. Morales and B. Dujon, "Evolutionary role of interspecies hybridization and genetic exchanges in yeasts," *Microbiology and molecular biology reviews : MMBR*, vol. 76, no. 4, pp. 721–39, 2012.

[17] G. I. Naumov, E. S. Naumova, I. Masneuf, M. Aigle, V. I. Kondratieva, and D. Dubourdieu, "Natural polyploidization of some cultured yeast *Saccharomyces* sensu stricto: Auto- and allotetraploidy," *Systematic and Applied Microbiology*, 2000.

[18] I. Stefanini, L. Dapporto, L. Berná, M. Polsinelli, S. Turillazzi, and D. Cavalieri, "Social wasps are a *Saccharomyces* mating nest.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 113, no. 8, pp. 2247–51, 2016.

[19] H. Zhang, A. Skelton, R. C. Gardner, and M. R. Goddard, "*Saccharomyces paradoxus* and *Saccharomyces cerevisiae* reside on oak trees in New Zealand: evidence for migration from Europe and interspecies hybrids," *FEMS Yeast Research*, vol. 10, no. 7, pp. 941–947, 2010.

[20] J. Bing, P.-J. Han, W.-Q. Liu, Q.-M. Wang, and F.-Y. Bai, "Evidence for a Far East Asian origin of lager beer yeast.," *Current biology : CB*, vol. 24, no. 10, pp. R380–1, 2014.

[21] B. Dunn and G. Sherlock, "Reconstruction of the genome origins and evolution of the hybrid lager yeast *Saccharomyces pastorianus*," *Genome research*, vol. 18, no. 10, pp. 1610–23, 2008.

[22] Y. Nakao, T. Kanamori, T. Itoh, Y. Kodama, S. Rainieri, N. Nakamura, T. Shimonaga, M. Hattori, and T. Ashikari, "Genome sequence of the lager brewing yeast, an interspecies hybrid.," *DNA research : an international journal for rapid publication of reports on genes and genomes*, vol. 16, no. 2, pp. 115–29, 2009.

[23] A. Walther, A. Hesselbart, and J. Wendland, "Genome sequence of *Saccharomyces carlsbergensis*, the world's first pure culture lager yeast.," *G3 (Bethesda, Md.)*, vol. 4, no. 5, pp. 783–93, 2014.

[24] D. Cavalieri, P. E. McGovern, D. L. Hartl, R. Mortimer, and M. Polsinelli, "Evidence for *S. cerevisiae* fermentation in ancient wine.," *Journal of molecular evolution*, vol. 57 Suppl 1, pp. S226–32, 2003.

[25] R. Michel and P. McGovern, "Chemical evidence for ancient beer," *Nature*, pp. 145–150, 1992.

[26] P. E. McGovern, J. Zhang, J. Tang, Z. Zhang, G. R. Hall, R. A. Moreau, A. Nunez, E. D. Butrym, M. P. Richards, C. S. Wang, G. Cheng, Z. Zhao, and C. Wang, "Fermented beverages of pre- and proto-historic China," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 101, pp. 17593–17598, Dec 2004.

[27] R. Barbosa, A. Pontes, R. O. Santos, G. G. Montandon, C. M. de Ponzzes-Gomes, P. B. Morais, P. Goncalves, C. A. Rosa, and J. P. Sampaio, "Multiple Rounds of Artificial Selection Promote Microbe Secondary Domestication-The Case of Cachaça Yeasts," *Genome Biol Evol*, vol. 10, no. 8, pp. 1939–1955, 2018.

[28] A. R. Borneman, B. A. Desany, D. Riches, J. P. Affourtit, A. H. Forgan, I. S. Pretorius, M. Egholm, and P. J. Chambers, "Whole-genome comparison reveals novel genetic elements that characterize the genome of industrial strains of *Saccharomyces cerevisiae*," *PLoS genetics*, vol. 7, no. 2, p. e1001287, 2011.

[29] C. L. Ludlow, G. A. Cromie, C. Garmendia-Torres, A. Sirr, M. Hays, C. Field, E. W. Jeffery, J. C. Fay, and A. M. Dudley, "Independent Origins of Yeast Associated with Coffee and Cacao Fermentation," *Curr. Biol.*, vol. 26, no. 7, pp. 965–971, 2016.

[30] G. A. Cromie, K. E. Hyma, C. L. Ludlow, C. Garmendia-Torres, T. L. Gilbert, P. May, A. a. Huang, A. M. Dudley, and J. C. Fay, "Genomic sequence diversity and population structure of *Saccharomyces cerevisiae* assessed by RAD-seq," *G3 (Bethesda, Md.)*, vol. 3, no. 12, pp. 2163–71, 2013.

[31] S.-F. Duan, P.-J. Han, Q.-M. Wang, W.-Q. Liu, J.-Y. Shi, K. Li, X.-L. Zhang, and F.-Y. Bai, "The origin and adaptive evolution of domesticated populations of yeast from Far East Asia," *Nature Communications*, vol. 9, no. 1, p. 2690, 2018.

[32] B. Dunn, C. Richter, D. J. Kvitek, T. Pugh, and G. Sherlock, "Analysis of the *Saccharomyces cerevisiae* pan-genome reveals a pool of copy number variants distributed in diverse yeast strains from differing industrial environments," *Genome research*, vol. 22, no. 5, pp. 908–24, 2012.

[33] B. Gallone, J. Steensels, T. Prahl, L. Soriaga, V. Saels, B. Herrera-Malaver, A. Merlevede, M. Roncoroni, K. Voordeckers, L. Miraglia, C. Teiling, B. Steffy, M. Taylor, A. Schwartz, T. Richardson, C. White, G. Baele, S. Maere, and K. Verstrepen, "Domestication and Divergence of *Saccharomyces cerevisiae* Beer Yeasts," *Cell*, vol. 166, no. 6, pp. 1397–1410.e16, 2016.

[34] B. Gallone, S. Mertens, J. L. Gordon, S. Maere, K. J. Verstrepen, and J. Steensels, "Origins, evolution, domestication and diversity of *Saccharomyces* beer yeasts," *Current Opinion in Biotechnology*, vol. 49, pp. 148–155, 2018.

[35] M. Gonçalves, A. Pontes, P. Almeida, R. Barbosa, M. Serra, D. Libkind, M. Hutzler, P. Gonçalves, J. Sampaio, W. Heijne, and et Al., "Distinct Domestication Trajectories in Top-Fermenting Beer Yeasts and Wine Yeasts," *Current Biology*, vol. 26, no. 20, pp. 2750–2761, 2016.

[36] J.-L. Legras, V. Galeote, F. Bigey, C. Camarasa, S. Marsit, T. Nidelet, I. Sanchez, A. Couloux, J. Guy, R. Franco-Duarte, M. Marcet-Houben, T. Gabaldon, D. Schuller, J. P. Sampaio, and S. Dequin, "Adaptation of *S. cerevisiae* to Fermented Food Environments Reveals Remarkable Genome Plasticity and the Footprints of Domestication," *Molecular Biology and Evolution*, vol. 35, no. 7, pp. 1712–1727, 2018.

[37] J. Peter, M. De Chiara, A. Friedrich, J.-X. Yue, D. Pflieger, A. Bergström, A. Sigwalt, B. Barre, K. Freel, A. Llored, C. Cruaud, K. Labadie, J.-M. Aury, B. Istace, K. Lebrigand, P. Barbry, S. Engelen, A. Lemainque, P. Wincker, G. Liti, and J. Schacherer, "Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates," *Nature*, vol. 556, no. 7701, pp. 339–344, 2018.

[38] Q.-M. Wang, W.-Q. Liu, G. Liti, S.-A. Wang, and F.-Y. Bai, "Surprisingly diverged populations of *Saccharomyces cerevisiae* in natural environments remote from human activity.," *Molecular ecology*, vol. 21, no. 22, pp. 5404–17, 2012.

[39] J.-X. Yue, J. Li, L. Aigrain, J. Hallin, K. Persson, K. Oliver, A. Bergström, P. Coupland, J. Warringer, M. C. Lagomarsino, G. Fischer, R. Durbin, and G. Liti, "Contrasting evolutionary genome dynamics between domesticated and wild yeasts," *Nature Genetics*, 2017.

[40] J. I. Eizaguirre, D. Peris, M. E. Rodriguez, C. A. Lopes, P. De Los Rios, C. T. Hittinger, and D. Libkind, "Phylogeography of the wild Lager-brewing ancestor (*Saccharomyces eubayanus*) in Patagonia," *Environ. Microbiol.*, vol. 20, no. 10, pp. 3732–3743, 2018.

[41] D. Libkind, C. T. Hittinger, E. Valério, C. Gonçalves, J. Dover, M. Johnston, P. Gonçalves, and J. P. Sampaio, "Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 35, pp. 14539–44, 2011.

[42] D. Peris, K. Sylvester, D. Libkind, P. Gonçalves, J. P. Sampaio, W. G. Alexander, and C. T. Hittinger, "Population structure and reticulate evolution of *Saccharomyces eubayanus* and its lager-brewing hybrids," *Molecular ecology*, vol. 23, no. 8, pp. 2031–45, 2014.

[43] D. Peris, Q. K. Langdon, R. V. Moriarty, K. Sylvester, M. Bontrager, G. Charron, J.-B. Leducq, C. R. Landry, D. Libkind, and C. T. Hittinger, "Complex Ancestries of Lager-Brewing Hybrids Were Shaped by Standing Variation in the Wild Yeast *Saccharomyces eubayanus*," *PLOS Genetics*, vol. 12, no. 7, p. e1006155, 2016.

[44] B. R. Gibson, E. Storgårds, K. Krogerus, and V. Vidgren, "Comparative physiology and fermentation performance of Saaz and Frohberg lager yeast strains and the parental species *Saccharomyces eubayanus*," *Yeast*, vol. 30, no. 7, pp. 255–266, 2013.

[45] S. S. González, E. Barrio, and A. Querol, "Molecular characterization of new natural hybrids of *Saccharomyces cerevisiae* and *S. kudriavzevii* in brewing," *Applied and environmental microbiology*, vol. 74, no. 8, pp. 2314–20, 2008.

[46] S. S. González, E. Barrio, J. Gafner, and A. Querol, "Natural hybrids from *Saccharomyces cerevisiae*, *Saccharomyces bayanus* and *Saccharomyces kudriavzevii* in wine fermentations," *FEMS Yeast Research*, vol. 6, no. 8, pp. 1221–1234, 2006.

[47] C. Le Jeune, M. Lollier, C. Demuyter, C. Erny, J. L. Legras, M. Aigle, and I. Masneuf-Pomarede, "Characterization of natural hybrids of Saccharomyces cerevisiae and Saccharomyces bayanus var. uvarum," *FEMS Yeast Res.*, vol. 7, no. 4, pp. 540–549, 2007.

[48] C. A. Lopes, E. Barrio, and A. Querol, "Natural hybrids of *S. cerevisiae* × *S. kudriavzevii* share alleles with European wild populations of *Saccharomyces kudriavzevii*," *FEMS Yeast Research*, vol. 10, no. 4, pp. 412–421, 2010.

[49] I. Masneuf, J. Hansen, C. Groth, J. Piskur, and D. Dubourdieu, "New hybrids between *Saccharomyces* sensu stricto yeast species found among wine and cider production strains.," *Applied and environmental microbiology*, vol. 64, no. 10, pp. 3887–92, 1998.

[50] D. Peris, R. Perez-Torrado, C. T. Hittinger, E. Barrio, and A. Querol, "On the origins and industrial applications of *Saccharomyces cerevisiae* × *Saccharomyces kudriavzevii* hybrids," *Yeast*, vol. 35, no. 1, pp. 51–69, 2018.

[51] D. Peris, C. A. Lopes, C. Belloch, A. Querol, and E. Barrio, "Comparative genomics among *Saccharomyces cerevisiae* × *Saccharomyces kudriavzevii* natural hybrid strains isolated from wine and beer reveals different origins," *BMC Genomics*, vol. 13, p. 407, 2012.

[52] B. Gibson and G. Liti, "*Saccharomyces pastorianus*: genomic insights inspiring innovation for industry," *Yeast*, vol. 32, no. 1, pp. 17–27, 2015.

[53] J. Wendland, "Lager yeast comes of age," *Eukaryotic Cell*, vol. 13, no. 10, pp. 1256–1265, 2014.

[54] F. Dai, E. Nevo, D. Wu, J. Comadran, M. Zhou, L. Qiu, Z. Chen, A. Beiles, G. Chen, and G. Zhang, "Tibet is one of the centers of domestication of cultivated barley," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 109, no. 42, pp. 16969–16973, 2012.

[55] F. Meußdoerffer and M. Zarnkow, *Das Bier: Eine Geschichte von Hopfen und Malz*. Beck C. H., 2014.

[56] E. Baker, B. Wang, N. Bellora, D. Peris, A. B. Hulfachor, J. A. Koshalek, M. Adams, D. Libkind, and C. T. Hittinger, "The Genome Sequence of *Saccharomyces eubayanus* and the Domestication of Lager-Brewing Yeasts," *Molecular Biology and Evolution*, vol. 32, no. 11, pp. 2818–2831, 2015.

[57] C. Belloch, R. Perez-Torrado, S. S. Gonzalez, J. E. Perez-Ortin, J. Garcia-Martinez, A. Querol, and E. Barrio, "Chimeric genomes of natural hybrids of *Saccharomyces cerevisiae* and *Saccharomyces kudriavzevii*," *Appl. Environ. Microbiol.*, vol. 75, no. 8, pp. 2534–2544, 2009.

[58] S. K. Hewitt, I. J. Donaldson, S. C. Lovell, and D. Delneri, "Sequencing and characterisation of rearrangements in three *S. pastorianus* strains reveals the presence of chimeric genes and gives evidence of breakpoint reuse," *PLoS ONE*, vol. 9, no. 3, p. e92203, 2014.

[59] M. Okuno, R. Kajitani, R. Ryusui, H. Morimoto, Y. Kodama, and T. Itoh, "Next-generation sequencing analysis of lager brewing yeast strains reveals the evolutionary history of interspecies hybridization," *DNA Research*, vol. 23, no. 1, p. dsv037, 2016.

[60] C. Monerawela and U. Bond, "Recombination sites on hybrid chromosomes in Saccharomyces pastorianus share common sequence motifs and define a complex evolutionary relationship between group I and II lager yeasts," *FEMS yeast research*, vol. 17, no. 5, 2017.

[61] P. Almeida, C. Gonçalves, S. Teixeira, D. Libkind, M. Bontrager, I. Masneuf-Pomarède, W. Albertin, P. Durrens, D. J. Sherman, P. Marullo, C. T. Hittinger, P. Gonçalves, and J. P. Sampaio, "A Gondwanan imprint on global diversity and domestication of wine and cider yeast *Saccharomyces uvarum*," *Nature communications*, vol. 5, no. May, p. 4044, 2014.

[62] M. van den Broek, I. Bolat, J. F. Nijkamp, E. Ramos, M. A. H. Luttik, F. Koopman, J. M. Geertman, D. de Ridder, J. T. Pronk, and J.-M. Daran, "Chromosomal Copy Number Variation in *Saccharomyces pastorianus* Is Evidence for Extensive Genome Dynamics in Industrial Lager Brewing Strains," *Applied and Environmental Microbiology*, vol. 81, no. 18, pp. 6253–6267, 2015.

[63] K. Krogerus, M. Arvas, M. De Chiara, F. Magalhães, L. Mattinen, M. Oja, V. Vidgren, J.-X. Yue, G. Liti, and B. Gibson, "Ploidy influences the functional attributes of de novo lager yeast hybrids," *Applied Microbiology and Biotechnology*, vol. 100, no. 16, pp. 7203–7222, 2016.

[64] K. J. Fisher, S. W. Buskirk, R. C. Vignogna, D. A. Marad, and G. I. Lang, "Adaptive genome duplication affects patterns of molecular evolution in *Saccharomyces cerevisiae*," *PLoS Genet.*, vol. 14, no. 5, p. e1007396, 2018.

[65] A. M. Selmecki, Y. E. Maruvka, P. A. Richmond, M. Guillet, N. Shoresh, A. L. Sorenson, S. De, R. Kishony, F. Michor, R. Dowell, and D. Pellman, "Polyploidy can drive rapid adaptation in yeast," *Nature*, vol. 519, no. 7543, pp. 349–352, 2015.

[66] R. T. Todd, A. Forche, and A. Selmecki, "Ploidy Variation in Fungi: Polyploidy, Aneuploidy, and Genome Evolution," *Microbiology Spectrum*, vol. 5, no. 4, 2017.

[67] J. Steensels and K. J. Verstrepen, "Taming Wild Yeast: Potential of Conventional and Nonconventional Yeasts in Industrial Fermentations," *Annual Review of Microbiology*, vol. 68, no. 1, pp. 61–80, 2014.

[68] J. De Roos and L. De Vuyst, "Acetic acid bacteria in fermented foods and beverages," *Curr. Opin. Biotechnol.*, vol. 49, pp. 115–119, 2018.

[69] M. C. Dzialo, R. Park, J. Steensels, B. Lievens, and K. J. Verstrepen, "Physiology, ecology and industrial applications of aroma formation in yeast," *FEMS Microbiology Reviews*, vol. 41, no. Supp_1, pp. S95–S128, 2017.

[70] N. Mukai, K. Masaki, T. Fujii, M. Kawamukai, and H. Iefuji, "PAD1 and FDC1 are essential for the decarboxylation of phenylacrylic acids in Saccharomyces cerevisiae.," *Journal of bioscience and bioengineering*, vol. 109, no. 6, pp. 564–9, 2010.

[71] N. Mukai, K. Masaki, T. Fujii, and H. Iefuji, "Single nucleotide polymorphisms of *PAD1* and *FDC1* show a positive relationship with ferulic acid decarboxylation ability among industrial yeasts used in alcoholic beverage production," *Journal of bioscience and bioengineering*, vol. 118, no. 1, pp. 50–5, 2014.

[72] Dale P. Van Wieren, *American Breweries II.* Eastern Coast Breweriana Association, 1995.

[73] I. S. Hornsey, *A History of Beer and Brewing.* RSC Paperbacks, The Royal Society of Chemistry, 2003.

[74] C. Monerawela, T. C. James, K. H. Wolfe, and U. Bond, "Loss of lager specific genes and subtelomeric regions define two different *Saccharomyces cerevisiae* lineages for *Saccharomyces pastorianus* Group I and II strains," *FEMS Yeast Research*, vol. 15, no. 2, 2015.

[75] L. Pasteur, "Etudes sur la bière ses maladies, causes qui les provoquent, procédé pour la rendre inaltérable, avec une théorie nouvelle de la fermentation," 1876.

[76] E. Hansen, "Recherches sur la physiologie et la morphologie des ferments alcooliques V. Methodes pour obtenir des cultures pures de *Saccharomyces* et de mikroorganismes analogues.," *C. R. Trav. Lab. Carlsberg*, vol. 2, pp. 92–105, 1883.

[77] M. A. Suchard, P. Lemey, G. Baele, D. L. Ayres, A. J. Drummond, and A. Rambaut, "Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10," *Virus Evolution*, vol. 4, no. 1, 2018.

[78] S. Mertens, J. Steensels, V. Saels, G. De Rouck, G. Aerts, and K. J. Verstrepen, "A Large Set of Newly Created Interspecific *Saccharomyces* Hybrids Increases Aromatic Diversity in Lager Beers," *Applied and Environmental Microbiology*, vol. 81, no. 23, pp. 8202–8214, 2015.

[79] K. Krogerus, F. Magalhães, V. Vidgren, and B. Gibson, "Novel brewing yeast hybrids: creation and application," *Applied microbiology and biotechnology*, vol. 101, no. 1, pp. 65–78, 2017.

[80] D. Peris, R. V. Moriarty, W. G. Alexander, E. Baker, K. Sylvester, M. Sardi, Q. K. Langdon, D. Libkind, Q. M. Wang, F. Y. Bai, J. B. Leducq, G. Charron, C. R. Landry, J. P. Sampaio, P. Goncalves, K. E. Hyma, J. C. Fay, T. K. Sato, and C. T. Hittinger, "Hybridization and adaptive evolution of diverse *Saccharomyces* species for cellulosic biofuel production," *Biotechnol Biofuels*, vol. 10, p. 78, 2017.

[81] D. R. Scannell, O. a. Zill, A. Rokas, C. Payen, M. J. Dunham, M. B. Eisen, J. Rine, M. Johnston, and C. T. Hittinger, "The Awesome Power of Yeast Evolutionary Genetics: New Genome Sequences and Strain Resources for the *Saccharomyces* sensu stricto Genus," *G3 (Bethesda, Md.)*, vol. 1, no. 1, pp. 11–25, 2011.

[82] G. Liti, A. N. N. Ba, M. Blythe, C. A. Müller, A. Bergström, F. A. Cubillos, F. Dafhnis-Calas, S. Khoshraftar, S. Malla, N. Mehta, C. C. Siow, J. Warringer, A. M. Moses, E. J. Louis, and C. A. Nieduszynski, "High quality de novo sequencing and assembly of the *Saccharomyces arboricolus* genome," *BMC Genomics*, vol. 14, no. 1, p. 69, 2013.

[83] A. R. Quinlan and I. M. Hall, "BEDTools: a flexible suite of utilities for comparing genomic features," *Bioinformatics*, vol. 26, no. 6, pp. 841–842, 2010.

[84] L. P. Pryszcz and T. Gabaldon, "Redundans: an assembly pipeline for highly heterozygous genomes," *Nucleic Acids Res.*, vol. 44, no. 12, p. e113, 2016.

[85] M. Kolmogorov, B. Raney, B. Paten, and S. Pham, "Ragout a reference-assisted assembly tool for bacterial genomes," *Bioinformatics*, vol. 30, no. 12, pp. i302–309, 2014.

[86] I. Wapinski, A. Pfeffer, N. Friedman, and A. Regev, "Natural history and evolutionary principles of gene duplication in fungi.," *Nature*, vol. 449, no. 7158, pp. 54–61, 2007.

[87] S. Capella-Gutiérrez, J. M. Silla-Martínez, and T. Gabaldón, "trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses," *Bioinformatics (Oxford, England)*, vol. 25, no. 15, pp. 1972–3, 2009.

[88] A. Stamatakis, "RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies," *Bioinformatics (Oxford, England)*, vol. 30, no. 9, pp. 1312–3, 2014.

[89] A. M. Kozlov, A. J. Aberer, and A. Stamatakis, "ExaML version 3: a tool for phylogenomic analyses on supercomputers," *Bioinformatics (Oxford, England)*, vol. 31, no. 15, pp. 2577–9, 2015.

[90] N. D. Pattengale, M. Alipour, O. R. Bininda-Emonds, B. M. Moret, and A. Stamatakis, "How many bootstrap replicates are necessary?," *J. Comput. Biol.*, vol. 17, no. 3, pp. 337–354, 2010.

[91] D. L. Ayres, A. Darling, D. J. Zwickl, P. Beerli, M. T. Holder, P. O. Lewis, J. P. Huelsenbeck, F. Ronquist, D. L. Swofford, M. P. Cummings, A. Rambaut, and M. A. Suchard, "BEAGLE: An Application Programming Interface and High-Performance Computing Library for Statistical Phylogenetics," *Systematic Biology*, vol. 61, no. 1, pp. 170–173, 2012.

[92] A. J. Drummond and A. Rambaut, "BEAST: Bayesian evolutionary analysis by sampling trees.," *BMC evolutionary biology*, vol. 7, no. 1, p. 214, 2007.

[93] A. J. Drummond and M. A. Suchard, "Bayesian random local clocks, or one rate to rule them all.," *BMC biology*, vol. 8, no. 1, p. 114, 2010.

[94] M. Worobey, G. Z. Han, and A. Rambaut, "A synchronized global sweep of the internal genes of modern avian influenza virus," *Nature*, vol. 508, no. 7495, pp. 254–257, 2014.

[95] A. Rambaut, A. J. Drummond, D. Xie, G. Baele, and M. A. Suchard, "Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7," *Syst. Biol.*, vol. 67, no. 5, pp. 901–904, 2018.

[96] G. Baele, P. Lemey, and M. A. Suchard, "Genealogical Working Distributions for Bayesian Model Testing with Phylogenetic Uncertainty," *Syst. Biol.*, vol. 65, no. 2, pp. 250–264, 2016.

[97] S. V. Angiuoli and S. L. Salzberg, "Mugsy: fast multiple alignment of closely related whole genomes," *Bioinformatics*, vol. 27, no. 3, pp. 334–342, 2011.

[98] J. C. Dittmar, R. J. Reid, and R. Rothstein, "ScreenMill: a freely available software suite for growth measurement, analysis and visualization of high-throughput screen data," *BMC Bioinformatics*, vol. 11, p. 353, 2010.

[99] D. Calahan, M. Dunham, C. DeSevo, and D. E. Koshland, "Genetic analysis of desiccation tolerance in *Sachharomyces cerevisiae*," *Genetics*, vol. 189, no. 2, pp. 507–519, 2011.

[100] A. Muir, E. Harrison, and A. Wheals, "A multiplex set of species-specific primers for rapid identification of members of the genus *Saccharomyces*," *FEMS Yeast Res.*, vol. 11, no. 7, pp. 552–563, 2011.

[101] R. J. Pengelly and A. E. Wheals, "Rapid identification of *Saccharomyces eubayanus and its hybrids*," *FEMS Yeast Res.*, vol. 13, no. 2, pp. 156–161, 2013.

9

Conclusions and future prospects

9.1 General conclusions

In this project we investigated the origin and diversity of industrial *Saccharomyces* yeasts by undertaking a whole-genome sequencing approach combined with high-throughput phenotyping. While *Saccharomyces cerevisiae* is the main workhorse in many of today's industrial fermentations, a large part of yeasts associated with man-made environments are interspecific hybrids between various *Saccharomyces* species. Our results demonstrated that industrial *S. cerevisiae* yeasts are genetically and phenotypically separated from wild stocks due to adaptation to industrial niches and human selection. They stem from only few ancestral strains that made their way into fermentations and subsequently spread across different industrial sites (e.g. breweries) and evolved into separate lineages. In particular, we identified two distinct lineages of *S. cerevisiae* beer strains that show signs of convergent evolution towards key traits related to the brewing environment (e.g. maltotriose utilization, alchol tolerance, 4-VG production). One of the two lineages, Beer 1, shows a strong geographical stratification, with clear separation of strains originating from Belgium-Germany, UK and US, indicating allopatric divergence after the separation from their most recent common ancestor. We did not identify strong subclustering of beer strains into beer styles, probably due to the brewers' practice of reusing one strain or a few strains for a wide range of beers.

Industrial *S. cerevisiae* yeasts show strong signatures of domestication both at the genetic and phenotypic level. Interestingly, domestication is most clear in beer strains, where hundreds of

years of unwilling selection led to impaired sexual reproduction, genome decay, general loss of stress resistance and selection for traits favorable in the brewing environment. The strong domestication signatures detected in beer yeasts may be a direct consequence of the nature of the brewing process, which differs substantially from other industries. Beer is produced throughout the year, which means yeasts are continuously growing in nutrient-rich medium. In addition "back-slopping" is a common practice in beer production: at the end of the fermentation process yeast is harvested and used for the inoculation of the next fermentation batch for multiple cycles. This practice not only allows yeast to continuously grow within the same environment but also confines yeast to this specific niche, promoting allopatric divergence that mirrors human trafficking and colonization, as observed in the phylogenetic analyses. This is in sharp contrast with wine yeasts that never completely lost contact with their natural environment: wine yeasts can only grow in wine must for a short period every year, spending the rest of their lives in and around the vineyards or in the guts of insects. This highlights the presence of multiple domestication "trajectories" that define and characterize ecological diversification among industrial yeasts and across different industries.

Moreover, the brewing environment holds a surprisingly rich biodiversity of *Saccharomyces* interspecific hybrids. We identified three distinct types of hybrids in the isolates we sequenced: *S. cerevisiae* x *S. eubayanus* hybrids in lager beers, *S. cerevisiae* x *S. kudriavzevii* hybrids in Trappist and Lambic beers and *S. uvarum* x *S. eubayanus* in Lambic beers and as contaminant of the brewing environment. Hybrid subgenomes tend to occupy specific lineages within the global population of the corresponding pure species, indicating the presence of relatively few targeted hybridization events or multiple events involving very similar strains. Hybrid lineages are most closely related to industrial lineages and their niche specificity suggests that they originated and were put to use early on in the industrialization of beer brewing. The establishment of new strains after the industrialization phase appears to have been largely due to spread and allopatric diversification of existing strains rather than the selection of new hybrids.

Our results indicate that beer yeast domestication started before the discovery of microbes and the isolation of the first pure yeast cultures by Emil Hansen at Carlsberg brewery in 1883 AD, but well after the invention of brewing. We estimated that the most recent common ancestor of today's commercial beer yeasts, both ale and lager, dates to the mid-16th century. Since then, multiple waves of human selection, driven by dramatic alterations and improvements of brewing practices and technology, shaped the current beer yeasts diversity.

Nowadays, evolution of most of commercial beer yeasts can be considered halted: most breweries rely on frozen yeast cultures that allow consistency of the product and efficiency of the process. However, the process has now moved to specialized laboratories, where the expanding experimental toolbox and the wealth of 'omics' data available for *Saccharomyces* yeasts opens new avenues to generate novel and superior industrial variants. New yeast variants can be obtained by exploring and screening the natural yeast biodiversity (both *Saccharomyces* and non-*Saccharomyces*) and subsequently by generating artificial combinations of phenotypic traits not (yet) encountered in

nature. In the last decade several new techniques have been developed that allow the creation of artificial diversity highly targeted to specific industrial processes. These techniques include non-genetically modified (GM) strategies (such as hybridization and marker-assisted breeding) as well as approaches that yield GM yeasts (such as CRISPR genome editing) (for a review see [1–5]). For the beer industry in particular, there is a great demand for product diversification and successful improved yeast variants were developed in order to improve fermentation performance, stress tolerance (e.g. temperature) and aroma formation [6–9]. For example, the discovery in 2011 by Libkind and coworkers of S. eubayanus as one of the parental species in the lager hybrid *S. pastorianus* [10] promoted a wave of effective attempts to recreate the ancestral hybridization event. Our improved knowledge of the genetic and phenotypic diversity, in particular of *S. cerevisiae*, allows the meticulous selection of the parental species for the artificial crosses in order to maximize the phenotypic diversity of the resulting hybrids compared to commercial hybrids. Thus, several centuries after the dawn of beer yeast domestication by commercial-scale brewing, a second revolution, sparked by biotechnology, is now driving a new era of beer yeast evolution and domestication.

9.2 Finding meaning in complexity

Complex genomes result in suboptimal assemblies. The most challenging yet fascinating aspect of the yeasts investigated in this project is their complex genome structure. Sequencing the isolates at their natural ploidy revealed an immense diversity in genome size and configuration: for industrial yeasts, diploid isolates are certainly the exception rather than the rule. The majority of beer strains, both pure species and interspecific hybrids, are polyploid and/or highly aneuploid. Abnormal numbers of chromosomes often occur together with high heterozygosity levels in pure *Saccharomyces* strains and within and between subgenomes in *Saccharomyces* hybrids. This complexity presents a challenge for current sequencing technologies and algorithms to assemble these sequences into whole genomes. The assemblies produced in the context of this project using next generation short-read sequencing **(Table 6.3, p. 127)**, represent a relatively fragmented, haploid consensus sequence of the true genome structure. During the course of my PhD I've witnessed an extraordinary and fast shift from next generation short-read sequencing to long-read sequencing (third-generation sequencing). Pacific Biosciences (PacBio) and Oxford Nanopore MinION have been successfully applied to generate chromosomal end-to-end assemblies of multiple yeasts [11–18]. Long-read sequencing also allows the study of large-scale structural variants (SVs) that constitute a substantial source of genetic variation within *Saccharomyces* yeasts but a difficult to resolve with short-read sequencing [18].

Despite the enormous progress in the field, the nature of industrial *Saccharomyces* genomes in general, and the poly-aneuploidy of beer yeasts in particular, still constitute a big challenge in the era of third generation sequencing. Based on the knowledge acquired within this project, we are currently resequencing part of the yeast collection using PacBio long-read sequencing,

in order to obtain chromosome-level assemblies (Gordon, J.L. et al., in preparation). Although we achieved higher contig continuity, genome completeness and better resolution of structural variants, the resulting assemblies are still collapsed (i.e. each contig represents a variable number of copies of the homologous region). Therefore, long-range haplotype information is lacking and heterozygosity (in terms of single nucleotide variants and short InDels, as well as larger structural variants such as deletions, duplications, or inversions) is often lost and/or not completely resolved even in simple diploid genomes. Phasing information is fundamental to understand the context and potential function of variants and to study evolutionary and population genetic processes. In the last few years, together with third generation sequencing technologies, new long-range scaffolding methodologies became available such as Bionano Genomics' next-generation mapping (NGM) [19], Chromosome Conformation Capture (Hi-C) and 10X Genomics. Although fundamentally different, they are all based on detecting physical linkage of elements within the genome (e.g., physical location of restriction enzymes, chromatin crosslinks, barcoded linked reads), thus improving the scaffolding process. The coupling of short- and long-read sequencing with long-range scaffolding has been successfully applied to obtain relatively long contiguous phased blocks of genomic variants in diploid human germline and cancer genomes [20–24], and diploid plants [25, 26]. In the near future the challenge will be to integrate multiple approaches and develop new bioinformatics tools in order to obtain phase-aware assemblies and overcome the limitations of studying genome evolution on haploid consensus sequences.

Polyploid genomes are still a challenge for population genomics. Our basic understanding of evolution of polyploid genomes is relatively poor compared to haploid and diploid genomes. In addition, poly-aneuploid genomes constitute an extra level of complexity due to the presence of variable copy number levels along the genome. The limitations in the resolution of the assembly discussed in the previous paragraph have important implications in the study of patterns of natural variations across individuals and populations. The difficulties in inferring the correct allelic configuration, dosage and segregation patterns for a specific locus hamper the application of the majority of standard population genomics analyses (e.g. demographic, gene flow, migration pattern analyses) and have impact on phylogenetic reconstruction [27–31]. This is particularly important in an intraspecific context where admixture is more likely to happen in the evolutionary history of each subpopulation. Heterozygosity might reflect (alternative) common origin(s) but also secondary contact and the two scenarios are not exclusive. Thus, haplotype information is crucial in order to depict the complete spectrum of genetic diversity across natural populations. The phylogenetic analyses reported in my thesis, for example, are merely the best statistical hypothesis of the true relationships across the yeasts investigated, based on the chosen approach. We detected and reported reticulation at different levels (population and species) but we only partially addressed the extent and directionality of the process due to analytical challenges. Furthermore, each *S. cerevisiae* lineage identified in this study has a unique life history in terms of ecology (wild

vs. domesticated), mating behavior (sexual versus obligate asexual), selection regime, ploidy profiles, population dynamics and demographic history. Therefore, within an intra-species context, these subpopulations represent genotypic clusters with different ecological specialization as a result of the distinct evolutionary trajectories followed. Our understanding of the genetic mechanisms (genome-wide vs. gene-specific) driving ecological diversification and sweep within *Saccharomyces* remains limited. Further research is needed to develop a comprehensive framework that takes into account the complexity of *Saccharomyces* species, in particular, and of eukaryotic microbial populations, in general, in the context of populations genomics studies.

From genomes to function: unraveling the molecular basis of adaptation during domestication. Nowadays, with more than 1000 *S. cerevisiae* strains and hundreds of other *Saccharomyces* strains sequenced, it is clear that variation in the *Saccharomyces* complex, both small and large-scale, is pervasive. The ultimate of goal of the work presented in this thesis is to unravel the genetic determinants of the natural phenotypic variability detected and to understand the nature of molecular changes that enable adaptations to man-made environments.

In this thesis we investigated the origin and the evolution of the phenotypic diversity observed in 4-vinyl guaiacol (4-VG) production, the main compound responsible for phenolic off-flavors (POF). The spicy, medicinal, clove-like aroma that this compound produces is tolerated in some specialty beers but it is strictly undesirable in the majority of ale beers and in lager beers. Two genes are responsible for the production of 4-VG, phenylacrylic acid decarboxylase (*PAD1*) and ferulic acid decarboxylase (*FDC1*). We showed that the majority of beer strains lost the ability to produce 4-VG due to the accumulation of loss-of-function mutations in the two genes or deletion of the region carrying the *PAD1-FDC1* gene cluster. In addition we described that very early in the domestication of beer strains within the Beer 1 clade, a 4-VG negative variant arose from a 4-VG positive variant and spread. While we suggest that human selection favored 4-VG negative variants during the domestication process of beer strains, it is still unclear if the absence of the *PAD1-FDC1* cluster gives a selective advantage within the beer environment or if it is a byproduct of other selective processes. Future experiments are needed to establish the precise role of the *PAD1-FDC1* gene cluster within the beer environment and clarify whether the presence of many 4-VG negative beer strains was actively, but unknowingly pursued by early brewers or whether it is simply the consequence of the artificial environmental conditions that beer strains experienced.

Establishing the adaptive nature of the diversity observed across populations during domestication, represents one of the most promising and challenging research avenues in the field. Most of the evidence of adaptive changes in domesticated lineages might be contingent on multiple factors affecting the mutational spectrum such as sampling, demographic processes, mode of reproduction and environmental conditions [32]. One way to test the potential adaptive role of the molecular change identified might be mimicking the domestication process in the laboratory by experimental evolution of an ancestral wild strain under the selective pressures of the specific man-made envi-

ronments and examine if changes arise as result of the artificial selection or of drift and stochastic mutational dynamics and test if the pattern observed can be replicated. Another way would be to engineer the ancestral allele in the genetic background of the domesticated strain and setup a competition assay between the natural and the engineered strains in the relevant environment. This experiment would test the presence and the degree of the fitness advantage conferred by the putative adaptive change in the conditions considered. However, the genetics of several adaptations does not follow simple modes of inheritance: for instance, 4-VG production is a trait with a known, and relatively simple genetic basis but several industrially relevant traits, likely targeted during the domestication process, are complex quantitative traits with a polygenic nature. These traits might be controlled by tens to hundreds different genes with a wide range of individual and concerted contributions to the phenotypic trait. This implies that the classical view of adaptation that involves selective sweeps at key loci should be extended in order to consider the importance of modest changes in allele frequencies at many loci, the so called "polygenic adaptation" [33, 34]. The strength of the selection and the heritability of trait will eventually determine the speed of adaptation of a population in a new environment. Identification of the loci that control variation in such complex traits represents the first step and one of the major challenges in the omics era. Several studies using both quantitative trait loci (QTL) mapping and genome-wide association mapping (GWAS) approaches confirmed that many industrially relevant traits in *S. cerevisiae* are controlled by hundreds of QTLs such as in volatile compounds production [35], in low-temperature tolerance [36], in high-temperature and ethanol tolerance [37–41] and in tolerance to toxins that arise during biofuel production [42]. The collection of strains sequenced and phenotype in this study represent an ideal starting point for designing QTL studies because it allows a fine-tuned selection of the parental candidates based on their phenotypic behaviour, genetic distance, ability to sporulate, viability of the spores and ploidy [43]. A key challenge for future research will be to connect quantitative genetics and population genetics approaches in order to understand the role of polygenic adaptation in the evolution of domesticated microbes. Furthermore we will need to go beyond the simplistic view of a DNA molecular change that affects the phenotypes and integrate RNA, protein and metabolites data in order to achieve a high-resolution view of the genotype-phenotype map within *Saccharomyces* [44–47].

9.3 References

[1] A. Dank, E. J. Smid, and R. A. Notebaart, "CRISPR-Cas genome engineering of esterase activity in *Saccharomyces cerevisiae* steers aroma formation," *BMC Res Notes*, vol. 11, no. 1, p. 682, 2018.

[2] B. Gibson, J.-M. A. Geertman, C. T. Hittinger, K. Krogerus, D. Libkind, E. J. Louis, F. Magalhães, and J. P. Sampaio, "New yeasts—new brews: modern approaches to brewing yeast design and development," *FEMS Yeast Research*, vol. 17, jun 2017.

[3] A. R. Gorter de Vries, P. A. de Groot, M. van den Broek, and J. G. Daran, "CRISPR-Cas9 mediated gene deletions in lager yeast *Saccharomyces pastorianus*," *Microb. Cell Fact.*, vol. 16, no. 1, p. 222, 2017.

[4] J. Steensels and K. J. Verstrepen, "Taming Wild Yeast: Potential of Conventional and Nonconventional Yeasts in Industrial Fermentations," *Annual Review of Microbiology*, vol. 68, pp. 61–80, sep 2014.

[5] J. Steensels, T. Snoek, E. Meersman, M. P. Nicolino, K. Voordeckers, and K. J. Verstrepen, "Improving industrial yeast strains: exploiting natural and artificial diversity.," *FEMS microbiology reviews*, vol. 38, pp. 947–95, sep 2014.

[6] J. R. Bellon, F. Schmid, D. L. Capone, B. L. Dunn, and P. J. Chambers, "Introducing a new breed of wine yeast: interspecific hybridisation between a commercial Saccharomyces cerevisiae wine yeast and Saccharomyces mikatae," *PLoS ONE*, vol. 8, no. 4, p. e62053, 2013.

[7] S. Mertens, J. Steensels, V. Saels, G. De Rouck, G. Aerts, and K. J. Verstrepen, "A Large Set of Newly Created Interspecific Saccharomyces Hybrids Increases Aromatic Diversity in Lager Beers," *Applied and Environmental Microbiology*, vol. 81, pp. 8202–8214, dec 2015.

[8] K. Krogerus, F. Magalhães, V. Vidgren, and B. Gibson, "Novel brewing yeast hybrids: creation and application.," *Applied microbiology and biotechnology*, vol. 101, pp. 65–78, jan 2017.

[9] K. Krogerus, F. Magalhaes, V. Vidgren, and B. Gibson, "New lager yeast strains generated by interspecific hybridization," *J. Ind. Microbiol. Biotechnol.*, vol. 42, no. 5, pp. 769–778, 2015.

[10] D. Libkind, C. T. Hittinger, E. Valério, C. Gonçalves, J. Dover, M. Johnston, P. Gonçalves, and J. P. Sampaio, "Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, pp. 14539–44, aug 2011.

[11] A. P. Douglass, B. Offei, S. Braun-Galleani, A. Y. Coughlan, A. A. R. Martos, R. A. Ortiz-Merino, K. P. Byrne, and K. H. Wolfe, "Population genomics shows no distinction between pathogenic *Candida krusei* and environmental *Pichia kudriavzevii*: One species, four names," *PLOS Pathogens*, vol. 14, no. 7, p. e1007138, 2018.

[12] T. Fournier, J.-S. Gounot, K. Freel, C. Cruaud, A. Lemainque, J.-M. Aury, P. Wincker, J. Schacherer, and A. Friedrich, "High-Quality de Novo Genome Assembly of the *Dekkera bruxellensis* Yeast Using Nanopore MinION Sequencing," *G3 (Bethesda, Md.)*, vol. 7, no. 10, pp. 3243–3250, 2017.

[13] F. Giordano, L. Aigrain, M. A. Quail, P. Coupland, J. K. Bonfield, R. M. Davies, G. Tischler, D. K. Jackson, T. M. Keane, J. Li, J.-X. Yue, G. Liti, R. Durbin, and Z. Ning, "De novo yeast genome assemblies from MinION, PacBio and MiSeq platforms," *Scientific Reports*, vol. 7, no. 1, p. 3935, 2017.

[14] P. Jenjaroenpun, T. Wongsurawat, R. Pereira, P. Patumcharoenpol, D. W. Ussery, J. Nielsen, and I. Nookaew, "Complete genomic and transcriptional landscape analysis using third-generation sequencing: a case study of *Saccharomyces cerevisiae* CEN.PK113-7D," *Nucleic Acids Research*, vol. 46, no. 7, pp. e38–e38, 2018.

[15] A. N. Salazar, A. R. Gorter de Vries, M. van den Broek, M. Wijsman, P. de la Torre Cortés, A. Brickwedde, N. Brouwers, J.-M. G. Daran, and T. Abeel, "Nanopore sequencing enables near-complete de novo assembly of *Saccharomyces cerevisiae* reference strain CEN.PK113-7D," *FEMS Yeast Research*, vol. 17, no. 7, 2017.

[16] F. Tondini, V. Jiranek, P. R. Grbin, and C. A. Onetto, "Genome Sequence of Australian Indigenous Wine Yeast *Torulaspora delbrueckii* COFT1 Using Nanopore Sequencing.," *Genome announcements*, vol. 6, no. 17, pp. e00321–18, 2018.

[17] J.-X. Yue and G. Liti, "Long-read sequencing data analysis for yeasts," *Nature Protocols*, vol. 13, no. 6, pp. 1213–1231, 2018.

[18] J.-X. Yue, J. Li, L. Aigrain, J. Hallin, K. Persson, K. Oliver, A. Bergström, P. Coupland, J. Warringer, M. C. Lagomarsino, G. Fischer, R. Durbin, and G. Liti, "Contrasting evolutionary genome dynamics between domesticated and wild yeasts," *Nature Genetics*, 2017.

[19] A. R. Hastie, E. T. Lam, A. W. C. Pang, X. Zhang, W. Andrews, J. Lee, T. Y. Liang, J. Wang, X. Zhou, Z. Zhu, T. Anantharaman, Z. Zdzakula, S. Bocklandt, U. Surti, M. Saghbini, M. Austin, M. Borodkin, R. E. Holmlin, and H. Cao, "Rapid Automated Large Structural Variation Detection in a Diploid Genome by NanoChannel Based Next-Generation Mapping," *bioRxiv*, p. 102764, 2017.

[20] C.-S. Chin, P. Peluso, F. J. Sedlazeck, M. Nattestad, G. T. Concepcion, A. Clum, C. Dunn, R. O'Malley, R. Figueroa-Balderas, A. Morales-Cruz, G. R. Cramer, M. Delledonne, C. Luo, J. R. Ecker, D. Cantu, D. R. Rank, and M. C. Schatz, "Phased diploid genome assembly with single-molecule real-time sequencing," *Nature Methods*, vol. 13, no. 12, pp. 1050–1054, 2016.

[21] D. Porubsky, S. Garg, A. D. Sanders, J. O. Korbel, V. Guryev, P. M. Lansdorp, and T. Marschall, "Dense and accurate whole-chromosome haplotyping of individual genomes," *Nature Communications*, vol. 8, no. 1, p. 1293, 2017.

[22] N. I. Weisenfeld, V. Kumar, P. Shah, D. M. Church, and D. B. Jaffe, "Direct determination of diploid genome sequences.," *Genome research*, vol. 27, no. 5, pp. 757–767, 2017.

[23] K. H. Y. Wong, M. Levy-Sakin, and P.-Y. Kwok, "De novo human genome assemblies reveal spectrum of alternative haplotypes in diverse populations," *Nature Communications*, vol. 9, no. 1, p. 3040, 2018.

[24] G. X. Y. Zheng, B. T. Lau, M. Schnall-Levin, M. Jarosz, J. M. Bell, C. M. Hindson, S. Kyriazopoulou-Panagiotopoulou, D. A. Masquelier, L. Merrill, J. M. Terry, P. A. Mudivarti, P. W. Wyatt, R. Bharadwaj, A. J. Makarewicz, Y. Li, P. Belgrader, A. D. Price, A. J. Lowe, P. Marks, G. M. Vurens, P. Hardenbol, L. Montesclaros, M. Luo, L. Greenfield, A. Wong, D. E. Birch, S. W. Short, K. P. Bjornson, P. Patel, E. S. Hopmans, C. Wood, S. Kaur, G. K. Lockwood, D. Stafford, J. P. Delaney, I. Wu, H. S. Ordonez, S. M. Grimes, S. Greer, J. Y. Lee, K. Belhocine, K. M. Giorda, W. H. Heaton, G. P. McDermott, Z. W. Bent, F. Meschi, N. O. Kondov, R. Wilson, J. A. Bernate, S. Gauby, A. Kindwall, C. Bermejo, A. N. Fehr, A. Chan, S. Saxonov, K. D. Ness, B. J. Hindson, and H. P. Ji, "Haplotyping germline and cancer genomes with high-throughput linked-read sequencing," *Nature Biotechnology*, vol. 34, no. 3, pp. 303–311, 2016.

[25] A. M. Hulse-Kemp, S. Maheshwari, K. Stoffel, T. A. Hill, D. Jaffe, S. R. Williams, N. Weisenfeld, S. Ramakrishnan, V. Kumar, P. Shah, M. C. Schatz, D. M. Church, and A. Van Deynze, "Reference quality assembly of the 3.5-Gb genome of *Capsicum annuum* from a single linked-read library," *Horticulture Research*, vol. 5, no. 1, p. 4, 2018.

[26] A. Minio, J. Lin, B. S. Gaut, and D. Cantu, "How Single Molecule Real-Time Sequencing and Haplotype Phasing Have Enabled Reference-Grade Diploid Genome Assembly of Wine Grapes," *Frontiers in plant science*, vol. 8, p. 826, 2017.

[27] A. K. Brysting, C. Mathiesen, and T. Marcussen, "Challenges in polyploid phylogenetic reconstruction: A case story from the arctic-alpine Cerastium alpinum complex," *Taxon*, vol. 60, no. 2, pp. 333–347, 2011.

[28] F. Dufresne, M. Stift, R. Vergilino, and B. K. Mable, "Recent progress and challenges in population genetics of polyploid organisms: An overview of current state-of-the-art molecular and statistical tools," 2014.

[29] H. R. Kates, M. G. Johnson, E. M. Gardner, N. J. C. Zerega, and N. J. Wickett, "Allele phasing has minimal impact on phylogenetic reconstruction from targeted nuclear gene sequences in a case study of *Artocarpus*," *American Journal of Botany*, vol. 105, no. 3, pp. 404–416, 2018.

[30] H. E. Lischer, L. Excoffier, and G. Heckel, "Ignoring Heterozygous Sites Biases Phylogenomic Estimates of Divergence Times: Implications for the Evolutionary History of *Microtus Voles*," *Molecular Biology and Evolution*, vol. 31, no. 4, pp. 817–831, 2014.

[31] A. J. Potts, T. A. Hedderson, and G. W. Grimm, "Constructing Phylogenies in the Presence Of Intra-Individual Site Polymorphisms (2ISPs) with a Focus on the Nuclear Ribosomal Cistron," *Systematic Biology*, vol. 63, no. 1, pp. 1–16, 2014.

[32] S. Marsit, J. B. Leducq, E. Durand, A. Marchant, M. Filteau, and C. R. Landry, "Evolutionary biology through the lens of budding yeast comparative genomics," *Nat. Rev. Genet.*, vol. 18, no. 10, pp. 581–598, 2017.

[33] N. H. Barton and P. D. Keightley, "Understanding quantitative genetic variation," *Nat. Rev. Genet.*, vol. 3, no. 1, pp. 11–21, 2002.

[34] J. K. Pritchard and A. Di Rienzo, "Adaptation - not by sweeps alone," *Nat. Rev. Genet.*, vol. 11, no. 10, pp. 665–667, 2010.

[35] M. Eder, I. Sanchez, C. Brice, C. Camarasa, J.-L. Legras, and S. Dequin, "QTL mapping of volatile compound production in *Saccharomyces cerevisiae* during alcoholic fermentation," *BMC Genomics*, vol. 19, no. 1, p. 166, 2018.

[36] E. García-Ríos, M. Morard, L. Parts, G. Liti, and J. M. Guillamón, "The genetic architecture of low-temperature adaptation in the wine yeast *Saccharomyces cerevisiae*," *BMC genomics*, vol. 18, no. 1, p. 159, 2017.

[37] X. H. Hu, M. H. Wang, T. Tan, J. R. Li, H. Yang, L. Leach, R. M. Zhang, and Z. W. Luo, "Genetic dissection of ethanol tolerance in the budding yeast *Saccharomyces cerevisiae*," *Genetics*, vol. 175, pp. 1479–1487, mar 2007.

[38] L. Riles and J. C. Fay, "Genetic basis of variation in heat and ethanol tolerance in *Saccharomyces cerevisiae*," *bioRxiv*, p. 318212, 2018.

[39] S. Swinnen, K. Schaerlaekens, T. Pais, J. Claesen, G. Hubmann, Y. Yang, M. Demeke, M. R. Foulquié-Moreno, A. Goovaerts, K. Souvereyns, L. Clement, F. Dumortier, and J. M. Thevelein, "Identification of novel causative genes determining the complex trait of high ethanol tolerance in yeast using pooled-segregant whole-genome sequence analysis," *Genome research*, vol. 22, no. 5, pp. 975–84, 2012.

[40] S. Wilkening, G. Lin, E. S. Fritsch, M. M. Tekkedil, S. Anders, R. Kuehn, M. Nguyen, R. S. Aiyar, M. Proctor, N. A. Sakhanenko, D. J. Galas, J. Gagneur, A. Deutschbauer, and L. M. Steinmetz, "An evaluation of high-throughput approaches to QTL mapping in *Saccharomyces cerevisiae*," *Genetics*, vol. 196, no. 3, pp. 853–65, 2014.

[41] C. V. Weiss, J. I. Roop, R. K. Hackley, J. N. Chuong, I. V. Grigoriev, A. P. Arkin, J. M. Skerker, and R. B. Brem, "Genetic dissection of interspecific differences in yeast thermotolerance," *Nat. Genet.*, vol. 50, no. 11, pp. 1501–1504, 2018.

[42] M. Sardi, V. Paithane, M. Place, E. Robinson, J. Hose, D. J. Wohlbach, and A. P. Gasch, "Genome-wide association across *Saccharomyces cerevisiae* strains reveals substantial variation in underlying gene requirements for toxin tolerance," *PLoS Genet.*, vol. 14, no. 2, p. e1007217, 2018.

[43] G. Liti, J. Warringer, and A. Blomberg, "Mapping quantitative trait loci in yeast," *Cold Spring Harbor Protocols*, vol. 2017, no. 8, pp. 631–635, 2017.

[44] A. Acharjee, B. Kloosterman, R. G. F. Visser, and C. Maliepaard, "Integration of multi-omics data for prediction of phenotypic traits using random forest," *BMC bioinformatics*, vol. 17 Suppl 5, no. Suppl 5, p. 180, 2016.

[45] J. M. Buescher and E. M. Driggers, "Integration of omics: more than the sum of its parts.," *Cancer & metabolism*, vol. 4, p. 4, 2016.

[46] H. Li, X. Wang, D. Rukina, Q. Huang, T. Lin, V. Sorrentino, H. Zhang, M. Bou Sleiman, D. Arends, A. McDaid, P. Luan, N. Ziari, L. A. Velázquez-Villegas, K. Gariani, Z. Kutalik, K. Schoonjans, R. A. Radcliffe, P. Prins, S. Morgenthaler, R. W. Williams, and J. Auwerx, "An Integrated Systems Genetics and Omics Toolkit to Probe Gene Function," *Cell Systems*, vol. 6, no. 1, pp. 90–102.e4, 2018.

[47] C. Manzoni, D. A. Kia, J. Vandrovcova, J. Hardy, N. W. Wood, P. A. Lewis, and R. Ferrari, "Genome, transcriptome and proteome: the rise of omics data and their integration in biomedical sciences," *Briefings in Bioinformatics*, vol. 19, no. 2, pp. 286–302, 2018.

10

Funding, Author Contributions and Conflict of Interest

10.1 Funding

Research at the laboratory of Kevin J. Verstrepen is supported by VIB, AB-InBev-Baillet Latour Fund, KULeuven, FWO, VLAIO, European Research Council (ERC) Consolidator Grant CoG682009 and Human Frontier Science Program (HFSP) grant 246 RGP0050/2013. Research at the laboratory of Steven Maere is supported by VIB and Ghent University. Brigida Gallone acknowledges funding from the VIB International PhD Program in Life Sciences.

10.2 Contribution

In this section I will briefly summarize my contribution to the research chapters of this thesis.

Chapter 6: Domestication and Divergence of *Saccharomyces cerevisiae* Beer yeasts. I performed the data analyses of the genomic and phenotypic data, the screen to test production of 4-VG, the data visualization and I drafted the manuscript together with Jan Steensels. Phenotypic data were collected by Jan Steensels, Veerle Saels and Beatriz Herrera-Malaver. I supervised Adriaan Merlevede during his Master internship while he developed the "Splint" algorithm to correct the read depth bias detected in part of our whole-genome sequencing data. Thank you to all co-authors for editing the manuscript and for the fruitful discussions about the results.

Chapter 7: Origins, evolution, Domestication and Diversity of *Saccharomyces* Beer Yeasts. I performed the data analysis and visualization. All authors contributed to the drafting and editing of the manuscript.

Chapter 8: Interspecific Hybridization Facilitates Niche Adaptation in Beer Yeast. I performed the data analyses of the genomic and phenotypic data, the data visualization and I drafted the manuscript together with Jan Steensels and Maria C. Dzialo. Guy Baele performed the time divergence analysis with feedback from me, Jan Steensels, Steven Maere, Mathias Hutzler and Franz Meussdoerffer. Thank you to all co-authors for editing the manuscript and for the fruitful discussions about the results.

10.3 Conflict of interest

Chapter 8 (Interspecific Hybridization Facilitates Niche Adaptation in Beer Yeasts, p. 151) contains unpublished data and the authors declare the right for future publication in peer-reviewed journals.

Funding: AB InBev partially funded the experiments performed in this study. AB InBev had no role in conceptualization, design, data collection, analysis, decision to publish or preparation of the manuscript. **Employment**: Several authors are employed by AB InBev. BS, PM, LD.

11

Thanks

Thank you to my supervisors, Kevin and Steven. Thank you for your time, help, advice and support. You inspired and encouraged me to achieve goals that I've never thought possible. I had a wonderful time in your labs and I discovered what it means to be a researcher and what kind of questions drive my passion and curiosity most.

Thank you to my thesis committee members for all of their guidance through this process; your discussion, ideas, and feedback have been absolutely invaluable.

Thank you to a multitude of people, fellow graduate students, research technicians, collaborators and interns who contributed to this research. Jan, in particular, I had a great time working with you and I am so happy and grateful that I will keep doing this.

Thank you to all my past and present lab fellows. You are like family to me. I, undoubtedly, could not have done this without you. You contributed immensely to my personal and professional time during this PhD.

Thank you to the Van der Kantjes, Dinia, Niek, Merel and Bjorn. It's not always easy to be so far away from home but since I met you I feel like having a piece of family here and that means the world to me. Thank you for the fun weekends, the family trips and all the cosy moments we shared in the last three years.

Rob, you brought so much happiness in my life halfway through my PhD. More than anybody else you supported me through this crazy roller coaster called PhD. Thank you for all the chats, the walks, the trips and the beers we shared; thank you for the beautiful layout of this thesis and thank you for being so understanding with the stressed version of myself. I can't wait to see where

life takes us next.

Un grazie speciale va alla mia famiglia per il loro supporto, ogni singolo giorno della mia vita. Ai miei genitori che non hanno mai posto limiti ai miei sogni.

12

Resume

12.1 Contact Information

Brigida Gallone
Date and place of Birth: 22-12-1986, Cisternino, Italy
Nationality: Italian
Email: brgal@psb.vib-ugent.be, brigida.gallone@gmail.com
Phone: +32 (0) 483386745
Skype: gallone.brigida
Evolutionary Systems Biology Laboratory - VIB Center for Plant Systems Biology and Department of Plant Biotechnology and Bioinformatics - University of Gent (Belgium)
Laboratory for Systems Biology - VIB Center for Microbiology - University of Leuven (Belgium)
Leuven Institute for Beer Research - LIBR
Address 1: UGent-VIB Research Building FSVM, Technologiepark 927, 9052 GENT, Belgium
Address 2: Bio-Incubator 1, Gaston Geenslaan 1, B-3001, Leuven, Belgium

12.2 Research Interests

Evolutionary and Comparative Genomics; Conservation Genomics; Population Genomics;Migration Processes; Speciation; Time Divergence; Phylogenomics; Domestication & Artificial selection; Ancient DNA; Big data analysis; Next Generation Sequencing (NGS), single molecule real time

(SMRT) sequencing (PacBio), MinION nanopore sequencing.

12.3 Education

Ph.D., Bioinformatics - Bioscience Engineering 10/2012 - Present
VIB - University of Gent - University of Leuven (Gent-Leuven, Belgium)
Master of Science, Biotechnology and Molecular Biology 10/2009 - 3/2012
Universita' degli Studi di Bari "Aldo Moro" (Bari, Italy)
Final grade 110/110 cum Laude
Bachelor of Science, Biotechnology and Molecular Biology 10/2005 - 7/2009
Universita' degli Studi di Bari "Aldo Moro" (Bari, Italy)
Final grade 107/110

12.4 Research Experience

Graduate Student Researcher - VIB Ph.D. in Life Sciences 10/2012 - Present
Evolutionary Systems Biology Laboratory / Laboratory for Systems Biology

- Supervisors: Prof. Steven Maere / Prof. Kevin J. Verstrepen
- Project: "Comparative genomics of industrial yeasts"

Undergraduate Student Researcher 5/2011 - 10/2011
Department of Cell and Molecular Sciences - The James Hutton Institute - University of Saint Andrews (Dundee, Scotland)

- Supervisors: Prof. Lesley Torrance/Dr. Graham Cowan
- Project: "Unravel the genetic basis of Extreme Resistance to Potyviruses in *Solanum tuberosum* group *Phureja*"

Undergraduate Student Researcher 10/2008-3/2009
Department of Plant Pathology and Applied Microbiology - Universita' degli studi di Bari "Aldo Moro" (Bari, Italy)

- Supervisors: Prof. Donato Gallitelli / Dr. Fabrizio Cillo
- Project: "Development of *Tobacco rattle virus* as viral vector for the expression of Hc-Pro in plants"

12.5 Conference Presentations

"31st VHYC Yeast Conference" - Leuven (Belgium) - Poster	04/2018
"Recent computational advances in microbial genomics and metagenomics" - Sint-Katelijne Waver (Belgium) - Talk	11/2016
"Genome Evolution Conference" - Tel Aviv (Israel) - Poster	11/2016
"EMBO Conference on Experimental Approaches to Evolution and Ecology" - Heidelberg (Germany) - Poster	10/2016
"SMBE Symposium 2016" - Queensland (Australia)- Poster	7/2016
"Belgian Brewing Conference - Chair J. De Clerck XV 2015 - Leuven (Belgium) - Talk	8/2015
"Complex Traits: From Models to Diseases" - Nice (France) - Poster	5/2014

12.6 Honors & Awards

- VIB International PhD Program in Life Sciences
  Call 2012 - 8 selected students
- Royal Union of old brewing students KU Leuven: Best brewing article KU Leuven 2016
- First prize "Data visualisation" at KU Leuven Datathon (February 2018)

12.7 Editorial Work

Reviewer for the following peer-reviewed, open access journals: Molecular Biology and Evolution (MBE), eLife, Cell, Genome Biology, Yeast, Nature Communications.

12.8 Teaching & Mentoring

- Daily supervisor of master thesis student Adriaan Merlevede 8/2014 - 06/2015
  (MSc. Bioinformatics - University of Leuven, Leuven, Belgium)
  Dissertation: "Detecting copy number variations in industrial yeast: a novel algorithm that corrects bias in read depth"
- Supervision of three students for the "Integrated Bioinformatics Project" 10/2014 - 1/2015
  (MSc. Bioinformatics, University of Leuven, Leuven, Belgium)
  Project: "Life history choices and genome evolution in *Saccharomyces cerevisiae*: an exploratory analysis"

12.9 References

Professor Kevin J. Verstrepen, PhD (e-mail: kevin.verstrepen@kuleuven.vib.be; Telephone: +32 (0)16 75 1390 - assistant or +32 (0)16 75 1393 - direct)

- Professor Faculty of Bioscience Engineering, University of Leuven
- Director of the VIB Center of Microbiology
- Professor Verstrepen is my PhD advisor

The Verstrepen Lab

Professor Steven Maere, PhD (e-mail: steven.maere@psb.vib-ugent.be; Telephone: +32 (0) 9 3313805)

- Professor of Computational Biology, University of Gent
- Professor Maere is my PhD advisor

The Maere Lab

Dr Jonathan Gordon, PhD (e-mail: jonathan.gordon@kuleuven.vib.be)

- Senior Postdoc in genetics, evolutionary biology and bioinformatics at the Verstrepen Lab
- Jonathan and I worked in close collaboration in the last two years

12.10 List of Publications

Brigida Gallone[*], Jan Steensels[*], Troels Prahl, Leah Soriaga, Veerle Saels, Beatriz Herrera-Malaver, Adriaan Merlevede, Miguel Roncoroni, Karin Voordeckers, Loren Miraglia, Clotilde Teiling, Brian Steffy, Maryann Taylor, Ariel Schwartz, Toby Richardson, Christopher White, Guy Baele, Steven Maere[*], Kevin J Verstrepen[*]. Domestication and divergence of *Saccharomyces cerevisiae* beer yeasts. *Cell*, 166(6):1397-1410, September 2016.
Press coverage:

- Article in Nature highlighting this study
- Article in The Economist highlighting this study
- Article in The New York Times highlighting this study

Brigida Gallone[*], Jan Steensels[*], Stijn Mertens, Maria C. Dzialo, Jonathan L. Gordon, Ruben Wauters, Florian A. Thesseling, Francesca Bellinazzo, Veerle Saels, Beatriz Herrera-Malaver, Mathias Hutzler, Franz Meussdoerffer, Philippe Malcorps, Ben Souffriau, Luk Daenen, Guy Baele, Steven Maere and Kevin J. Verstrepen. Interspecific Hybridization Facilitates Niche Adaptation in Beer Yeasts. 2018, *submitted.*

Brigida Gallone[*], Stijn Mertens, Sam Crauwels, Bart Lievens, Kevin J Verstrepen, Jan Steensels. Genomics and Evolution of Beer Yeasts. *Brewing Microbiology: Current Research, Omics and Microbial Ecology*, 145-177, January 2017.

Jonathan L Gordon, **Brigida Gallone**, Steven Maere, Kevin J Verstrepen. Evolutionary context improves regulatory network predictions. *Cell systems*, 4(5):478-479, May 2017.

Stijn Mertens, Jan Steensels, **Brigida Gallone**, Ben Souffriau, Philippe Malcorps, Kevin J Verstrepen. Rapid screening method for phenolic off-flavor (POF) production in yeast. *Journal of the American Society of Brewing Chemists*, 75(4):318-323, September 2017.

Brigida Gallone[*], Stijn Mertens, Jonathan L Gordon, Steven Maere, Kevin J Verstrepen, Jan Steensels. Origins, evolution, domestication and diversity of *Saccharomyces* beer yeasts. *Current opinion in biotechnology*, 49:148-155, February 2018.

Bram Cerulus, Abbas Jariani, Gemma Perez-Samper, Lieselotte Vermeersch, Julian Pietsch, Matthew Crane, **Brigida Gallone**, Miguel Roncoroni, Maria Dzialo, Sander Govers, Jhana Herdrickx, Eva Galle, Maarten Coomans, Pieter Berden, Sara Verbandt, Peter Swain, Kevin J Verstrepen. Transition between fermentation and respiration determines history-dependent behavior in fluctuating

carbon sources. *Elife*. doi: 10.7554/eLife.39234, October 2018

Gemma Perez-Samper, Bram Cerulus, Abbas Jariani, Lieselotte Vermeersch, Nuria Barraj n Simancas, Markus M M Bisschops, Joost van den Brink, Daniel Solis-Escalante, **Brigida Gallone**, Dries De Maeyer, Elise van Bael, Tom Wenseleers, Jan Michiels, Kathleen Marchal, Pascale Daran-Lapujade, Kevin J Verstrepen. The Crabtree effect shapes the *Saccharomyces cerevisiae* lag phase. *mBio*. doi: 10.1128/mBio.01331-18, October 2018

Alexander J. Kruis*, **Brigida Gallone***, Timo Jonker, Astrid E. Mars, Irma van Rijswijck, Judith C.M Wolkers-Rooijackers, Eddy J. Smid, Jan Steensels, Kevin J. Verstrepen, Servé W.M. Kenger, John van der Oost and Ruud Weusthuis. Role of Eat1 and other alcohol acetyltransferases to ester production in *Saccharomyces cerevisiae*. *Frontiers in Microbiology*, *accepted*

Stijn Mertens, **Brigida Gallone**, Jan Steensels, Beatriz Herrera-Malaver, Jeroen Cortebeek, Robbe Nolmans, Veerle Saels, Valmik K. Vyas and Kevin J. Verstrepen. Reducing phenolic off-flavors through CRISPR-based gene editing of the *FDC1* gene in *Saccharomyces cerevisiae* x *Saccharomyces eubayanus* hybrid lager beer yeasts. *Plos One*, *accepted*