bioRxiv preprint doi: https://doi.org/10.1101/150722; this version posted August 31, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Identification of a novel interspecific hybrid yeast from a metagenomic spontaneously inoculated beer sample using Hi-C

Authors:
Caiti Smukowski Heil*[a], Joshua N. Burton*[a,b], Ivan Liachko*[a,c], Anne Friedrich[d], Noah A. Hanson[a], Cody L. Morris[e], Joseph Schacherer[d], Jay Shendure[a,f], James H. Thomas[a], Maitreya J. Dunham[a]
*: co-first authors

Affiliations:
[a]Department of Genome Sciences, University of Washington, Seattle, WA
[b]Current address: Adaptive Biotechnologies, Seattle, WA
[c]Current address: Phase Genomics, Seattle WA
[d]Genetics, Genomics, and Microbiology, University of Strasbourg, Strasbourg, France
[e]Current address: Mollusk, Seattle, WA
[f]Howard Hughes Medical Institute

Running Head: Identification of a novel hybrid yeast in spontaneously inoculated beer via Hi-C

bioRxiv preprint doi: https://doi.org/10.1101/150722; this version posted August 31, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Abstract:
Interspecific hybridization is a common mechanism enabling genetic diversification and adaptation; however, the detection of hybrid species has been quite difficult. The identification of microbial hybrids is made even more complicated, as most environmental microbes are resistant to culturing and must be studied in their native mixed communities. We have previously adapted the chromosome conformation capture method Hi-C to the assembly of genomes from mixed populations. Here, we show the method's application in assembling genomes directly from an uncultured, mixed population from a spontaneously inoculated beer sample. Our assembly method has enabled us to de-convolute 4 bacterial and 4 yeast genomes from this sample, including a putative yeast hybrid. Downstream isolation and analysis of this hybrid confirmed its genome to consist of *Pichia membranifaciens* and that of another related, but undescribed yeast. Our work shows that Hi-C-based metagenomic methods can overcome the limitation of traditional sequencing methods in studying complex mixtures of genomes.

Keywords: Hi-C; metagenome assembly; metagenomics; hybrid; hybridization; spontaneous fermentation

COI: JNB, IL, J. Shendure, and MJD have financial interests in Phase Genomics, a company that has commercialized the Hi-C genome assembly technique. IL is a current employee of Phase Genomics.

Nucleic Acid Sequences: Sequences are deposited in NCBI SRA and Assembly under the Project ID PRJNA390460.

bioRxiv preprint doi: https://doi.org/10.1101/150722; this version posted August 31, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Introduction

Once thought to be a rare phenomenon, hybrids are increasingly recognized as common in natural and industrial environments (Abbott et al. 2013). Hybridization can act as an instantaneous, large-scale mutational event, introducing abundant genetic variation and potentially producing changes in gene expression, gene content, and karyotype. New combinations of alleles formed during the initial hybridization event, or transgressive segregation from subsequent recombination, may allow the hybrid to explore greater phenotypic space than either of its parents, and thus to colonize new, uninhabited niches. There is abundant evidence suggesting that because hybrids can explore unoccupied ecological niches, they may be particularly adept at invading novel habitats and even become invasive pests (Bullini 1994; Mallet 2007; Muhlfeld et al. 2014; Pryszcz et al. 2014). In line with these findings, extensive hybridization is often associated with human-disturbed habitats (Dowling and Secor 1997; Vallejo-Marin and Hiscock 2016), and climate change is predicted to increase the number of hybridization events (Kelly et al. 2010; Hoffmann and Sgro 2011).

Similarly, hybrids are often found in commercial settings where stress tolerance or adaptation to particular environmental conditions is important. Most famously, the lager brewing yeast *Saccharomyces pastorianus* is a hybrid between the commensal baking and brewing yeast, *Saccharomyces cerevisiae*, and the cryotolerant species *Saccharomyces eubayanus*. *S. cerevisiae* and *S. eubayanus* diverged approximately 20 million years ago and are 20% divergent at the coding level (Tamai et al. 1998; Yamagishi and Ogata 1999; Dunn and Sherlock 2008; Nakao et al. 2009; Baker et al. 2015; Gibson and Liti 2015). After what is presumed to be an allopolyploid event, *S. pastorianus* has undergone massive loss of chromosomes and chromosomal segments. While *S. pastorianus* genomic content remains fairly balanced in representation of both parental genomes, many hybrids are observed to have imbalanced retention and loss patterns. Hybrids amongst other members of the *Saccharomyces* clade have been discovered in wine fermentation and brewing operations (Gonzalez et al. 2006; Gonzalez et al. 2008; Bellon et al. 2015; Wolfe 2015; Magalhaes et al. 2017), in addition to a number of other fungal species hybrids outside of this well-studied clade. For example, a large proportion of strains of the spoilage yeast *Dekkera* (Brettanomyces) *bruxellensis*, which itself is used in brewing, originated through interspecific hybridization (Hellborg and Piskur 2009; Borneman et al. 2014), and these hybrids have a selective advantage over non-hybrids in wineries (Borneman et al. 2014).

It has been suggested that the previously unrecognized ubiquity of hybrid species is at least in part due to issues of detection and identification (Dowling and Secor 1997), which may be particularly true in microbial taxa (Albertin and Marullo 2012). One of the issues complicating detection of microbial hybrids is the difficulty of dissecting microbial communities in general, which can be made up of many different species, most of which are unculturable (Lane et al. 1985; Pace et al. 1986; Handelsman 2004; Marcy et al. 2007). With the advancement of next generation sequencing, the field of metagenomics has moved from relying on 16S rRNA gene fingerprinting methods (Lane et al. 1985; Pace et al. 1986; Rajendhran and Gunasekaran 2011) to various applications of shotgun sequencing (Venter et al. 2004; Human Microbiome Project 2012; Iverson et al. 2012; Howe et al. 2014), single-cell genome sequencing (Marcy et al. 2007; Stepanauskas 2012; Kamke et al. 2013), and associated computational methods (Saeed et al. 2012; Carr et al. 2013; Hug et al. 2013; Wood and Salzberg 2014).

Previous work from our groups and others demonstrated the use of contact probability maps generated through chromosome conformation capture methods (Lieberman-Aiden et al.

bioRxiv preprint doi: https://doi.org/10.1101/150722; this version posted August 31, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

2009; Dekker et al. 2013) to assemble complex genomes (Burton et al. 2013; Kaplan and Dekker 2013), and in particular, to deconvolute complex metagenomic samples (Beitel et al. 2014; Burton et al. 2014; Marbouty et al. 2014). The basis of this idea takes advantage of the Hi-C protocol, in which crosslinking of proximal chromatin segments occurs prior to cell lysis, thereby ensuring that each Hi-C interaction involves a pair of reads generated from within the same cell. In a diverse microbial population, such data can help delineate which genomic sequences originated from individual species. Burton et al. demonstrated that the method, MetaPhase, can reliably deconvolute and assemble genomes from up to 18 eukaryotic and prokaryotic species from synthetic mixtures in the lab. Importantly, this method should also allow hybrids to be identified in a mixed sample, since crosslinks would be made between the different genomes in contrast to the solely within-genome linkages in a simple mix of the parental species.

Here, we extend this work to explore the microbial community of a spontaneously inoculated beer sample. In the tradition of Belgian lambic, boiled wort is left open to the elements and is fermented solely from autochthonous yeast and bacteria (Vanoevelen et al. 1977; Verachtert and Iserentant 1995; Keersmaecker 1996). Past studies of lambic style fermentation have relied heavily on culturing various species over the course of fermentation; more recently, genetic tools such as terminal restriction fragment length polymorphism, barcoded amplicon sequencing, and quantitative PCR have shed light on the microbial progression, but still are limited by deconvolution issues (Bokulich et al. 2012; Spitaels et al. 2014c). Our method deconvolutes the microbial genomes in the beer sample without culturing or a priori knowledge of species or strain, revealing eight organisms, including novel strains and species of bacteria and yeast, and a novel yeast hybrid. The hybrid appears to be a recent hybridization event between *Pichia membranifaciens* and a previously undescribed species approximately 10-20 million years divergent. We characterize the genome and fermenting capabilities of this novel hybrid, and describe the other species identified. This is one of the first demonstrations of computational metagenomic deconvolution of a non-laboratory sample (see also Marbouty et al. 2017 and Marbouty et al. 2014), and the first method with the ability to detect hybrids in a heterogeneous population.

Materials & Methods

Sample collection

We obtained 20 mL of actively growing culture sampled from the top of a wine barrel containing the spontaneously inoculated beer "Old Warehouse," produced by Epic Ales in Seattle, WA on May 8, 2014.

Shotgun, Hi-C libraries

Approximately 5 mL of the sample was pelleted and total DNA was isolated using a standard phenol/chloroform glass bead extraction. Shotgun libraries were prepared using the Nextera Kit (Illumina). Hi-C libraries were prepared as described (Burton et al. 2014). Sequencing was performed on the NextSeq 500 Illumina platform.

De novo assembly, deconvolution, and individual species assembly

The draft metagenome assembly was created using the IDBA-UD assembler (Peng et al. 2012) with the following parameters: '--pre_correction --mink 20 --maxk 60 --step 10'. Hi-C reads were aligned to the draft assembly using BWA (Li and Durbin 2009), following the strategy of

bioRxiv preprint doi: https://doi.org/10.1101/150722; this version posted August 31, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Burton et al. (2014). Clustering of contigs into individual clusters was done using MetaPhase (Burton et al. 2014); https://github.com/shendurelab/MetaPhase). Independently of MetaPhase, in order to determine species identity, contigs were mapped to the BLAST sequence database (https://blast.ncbi.nlm.nih.gov/Blast.cgi; July 2014 database), using blastn with the following parameters: -perc_identity 95 -evalue 1e-30 -word_size 50. Percentage of each cluster assembly aligning to the reference was estimated using AssemblyEvaluator (https://github.com/snayfach/AssemblyEvaluator).

Hybrid assembly and analysis

To confirm the putative hybrid from the metagenome assembly, the hybrid was isolated from a single colony. DNA was extracted and prepared with a Nextera kit, as above. Reads were mapped to the metagenome assembly using BWA (Li and Durbin 2009), and once confirmed as a hybrid, a new draft assembly was created using IDBA-UD with parameters as above (Peng et al. 2012). To split the assembly into species-specific sub-genomes, contigs from this new assembly were compared against the *P. membranifaciens* genome v2.0 (Riley et al. 2016) using blastn with an e-value of 1E-12 (Altschul et al. 1990). All contigs whose single best blastn match to the *P. membranifaciens* genome is >= 97% identical were classified as sub-genome A, whereas all contigs whose single best blastn match to the *P. membranifaciens* genome is >= 77% and < 92% identical were classified as sub-genome B. Contigs of high divergence and contigs with identity between 92-97% were unable to be parsed into sub-genomes. Augustus v3.2.1 was used to create gene predictions for both sub-genomes (Stanke et al. 2004; Stanke and Morgenstern 2005). Gene predictions between sub-genomes were compared using blastn and blastp using the best hit. BLAST was also used to assess coverage differences between the sub-genomes. A rough approximation of divergence between the sub-genomes was estimated by constructing neighbor-joining trees for *P. membranifaciens* predicted genes (g1201, g2530) using ClustalOmega (Sievers et al. 2011) and PHYLIP (Felsenstein 2005), and sequence from each sub-genome, as compared with results using *S. cerevisiae*, *S. paradoxus*, *S. kudriavzevii*, and *S. uvarum*.

Neighbor-joining tree building for S. cerevisiae *strain phylogeny*

The biallelic segregating sites detected among a set of previously sequenced *S. cerevisiae* isolates (Strope et al. 2015; Gallone et al. 2016) and the OW sequence were used to elucidate the phylogenetic relationships among them. Based on these sites, a distance matrix was first computed and submitted to bionj algorithm (R packages ape and SNPrelate) to construct a neighbor-joining tree.

Lab-appropriate wort

Our protocol was designed to make 1.75 liters of a lab-appropriate American style amber beer wort. 320 grams of an amber liquid malt extract are used to meet the Brewers Association's Beer Standard Guidelines for an American style amber/red ale with an original gravity of 11.9 - 14.3 ºPlato. Our wort was within this range at 13.77 ºPlato. The sugar composition was: 44% maltose, 11% maltotriose, 13% glucose, 3% sucrose, 4% fructose, and 25% unfermentable dextrins. To prepare media, 320 grams of liquid malt extract was added to 1.5 liters of 77 °C water using a heat block with a running stir bar. Once in solution, the heat was adjusted such that the wort was allowed to achieve a rolling boil without overflowing for 1 hour. The media was then repeatedly poured through paper filters until achieving a uniform, cloudless consistency. Distilled water was

bioRxiv preprint doi: https://doi.org/10.1101/150722; this version posted August 31, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

added to bring the solution to a total of 1.75 liters before being covered. The media was vacuum filtered through Nalgene 500mL Rapid-Flow Bottle Top Filters with a pore size of 0.45µm, and into a sterile 2L bottle. The bottle was stored at room temperature and away from light exposure. Wort was inoculated with a single colony suspended in 2 mL of water after the temperature reached 22 °C. The wort was aerated via swirling at the point of inoculation, after which no shaker was used and no air was pumped in. It was kept at room temperature for the duration of 5 weeks. After 5 weeks, ethanol volume, pH, and sugar content were analyzed (White Labs, San Diego, CA). No single sugar fermentation studies were conducted at this time.

Data availability

Reads and assemblies for the metagenome are deposited at NCBI under the Project ID PRJNA390460.

Results

Hi-C/MetaPhase on a craft beer sample

We obtained a 20mL sample of active culture growing in wort that had been inoculated in open air and serially pitched in a used wine barrel for several years by Epic Ales in Seattle, Washington. We performed shotgun sequencing and Hi-C on this sample and applied the approach utilized by Burton et al. (2014) to assemble and deconvolute the metagenomic sample. Briefly, we sequenced two shotgun libraries, one with 3.61 M read pairs of 35 bp each, and one with 27.5 M read pairs of 100 bp each, and two Hi-C libraries, one using the HindIII restriction site and one using the NcoI restriction site, with 13.6 M and 14.5 M read pairs respectively. We converted the shotgun reads into a *de novo* assembly using IDBA-UD (Peng et al. 2012), yielding 32,549 contigs with an N50 of 5,285 bp at a minimum contig length of 200 bp and a total size of 67,496,013 bp. We mapped the Hi-C reads to the assembly using BWA (Li and Durbin 2009), and applied the MetaPhase software to deconvolute draft contigs into a progressively decreasing number of clusters while measuring intra-cluster enrichment signal as described in Burton et al. (2014). This signal reached a maximum at N=8 clusters (**Figure 1B**), suggesting the presence of 8 discernable species-level genomes in this assembly (**Figure 1A, Table 1**). 11,189 contigs are included in these eight clusters, representing 73.3% of the initial sequence assembly (49,498,624 bp); the remaining contigs are small (N50 of 1349 bp) and were unable to be confidently clustered. Seven of the eight clusters matched to a unique species of either fungal or bacterial origin (**Figure 1C**), and in each of these clusters, the quantity of sequence served to indicate the completeness of coverage of the species' genome (**Table 1**). In the eighth cluster, no significant quantity of sequence closely matched a known reference genome; however, several sequences in this cluster matched individual genes from *Pediococcus damnosus*, a known spoilage bacterium for which no complete reference genome existed in the BLAST database at the time of analysis (but see (Snauwaert et al. 2015)). This cluster is now confirmed to be *P. damnosus*.

Identification and characterization of a hybrid yeast

One of the more abundant organisms in the sample had sequencing reads from the yeast *Pichia membranifaciens*, a common food spoilage species (Kurtzman et al. 2011; Riley et al. 2016). *P. membranifaciens* has been identified in spontaneously inoculated beer (Spitaels et al. 2014c), wine (Rankine 1966; Stefanini et al. 2016a), and other beverages (Lachance 1995; Silva et al. 2009). The assembly generated here has a genome size roughly twice that of *P.*

bioRxiv preprint doi: https://doi.org/10.1101/150722; this version posted August 31, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

membranifaciens, and upon further inspection, BLAST results were consistent with two separate genomes, one belonging to *P. membranifaciens*, and the other to an unknown, related species. We hypothesized that this could result either from a technical error in MetaPhase collapsing two species into one cluster, or alternatively, a hybridization event. To test this hypothesis, isolates of the putative hybrid were grown and sequenced (**Figure 2A**). Sequencing reads were mapped back to the MetaPhase assembly, and mapped exclusively to the cluster identified as the hybrid or to unclustered reads (which were unable to be assembled), substantiating cluster 3 as a true hybrid. This is further supported by ploidy analysis confirming a genome size of approximately 24 Mb (**Figure 2B**).

Contigs from the assembly were parsed into two genomes using percent identity compared to *P. membranifaciens* (Riley et al. 2016), with contig identity above 97% classified as sub-genome A (*P. membranifaciens*), and contig identity between 77-92% classified as sub-genome B (unknown species). Sub-genome A has 5792 predicted genes and sub-genome B has 5742 predicted genes, similar to the 6604 ORFs in *S. cerevisiae* (of which 5159 have been verified). Of 4780 unique protein matches between the two sub-genomes, 92% of proteins share 70% amino acid similarity or higher (4410 proteins), and of these, 56% have greater than 90% amino acid identity. Phylogenetic analysis suggests approximately 10-20 million years divergence between the hybrid parentals, which appears consistent with the amino acid divergence. However, we note that this could be an underestimate since we only considered genes with matches between sub-genomes.

Many hybrids experience gene loss, copy number changes, and aneuploidy following a hybridization event (Abbott et al. 2013). In a crude sense, the amount of loss can date a hybridization event, with more genomic homogenization indicating a more ancient hybridization event. To estimate the "completeness" of the hybrid genome, hybrid contigs were compared to *P. membranifaciens* to obtain coverage estimates. About 2.5% of the *P. membranifaciens* genome has no coverage from the hybrid, mostly in telomeric regions. Of the rest of the genome, 88.8% is covered exactly twice, and 89.6% is covered two or more times. Most of the regions with greater than 2X coverage are likely artifacts of the blast method and overlapping assembly contigs, although we have evidence for potential copy number variants in gene regions in both sub-genomes.

Sub-genome B contig 60_487 is duplicated relative to the rest of the genome and contains three predicted genes: g1333 (methionine permease, *MUP1*), g1334 (hypothetical protein in *P. membranifaciens*, similarity to methionine permease in *P. kudriavzevii*), and g1335 (hypothetical protein in *P. membranifaciens*, similar to allantoate permease in *P. kudriavzevii*). All three of these genes are present in single copy in sub-genome A. Sub-genome A contig 60_428 is also duplicated and contains a single predicted gene, g1144, an oligopeptide transporter (*OPT2*). It is not present in sub-genome B, either representing a loss of heterozygosity event, or incomplete assembly. Breakpoints of both duplication events are not identifiable as the copy number increase spans the whole contig, suggesting that the breakpoints are in unassembled regions which may contain additional genes of interest.

Approximately 10.4% of the *P. membranifaciens* genome is covered only once, but the large majority of that is in short fragments distributed across the genome (average length: 202.19 bp, median length: 106 bp), most likely an artifact of incomplete hybrid assembly or unusually divergent regions of subgenome B. Still, there are 120 segments of 1 kb or more that are covered once, representing 2.7% of the genome (total length of 317 kb) and might represent missing

bioRxiv preprint doi: https://doi.org/10.1101/150722; this version posted August 31, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

genes in one or the other sub-genomes. Overall, these observations are consistent with a recent hybridization event.

Pichia hybrid does not play a significant role in fermentation

To characterize the fermenting capabilities of the *Pichia* hybrid, the isolate was used to inoculate a wort media at room temperature for five weeks. During this time, aggregates accumulated and could be seen rising to the wort's surface and attaching to walls of the fermentation vessel (**Figure 2C**). When viewed under a microscope, cells appeared to bend at the points of cell division, resulting in a palisade arrangement with stacking patterns (**Figure 2D**). No pellicles formed on the surface of the wort, in contrast to both *P. membranifaciens* and *P. kudriavzevii*, indicating that the hybrid may have lost this phenotype (Matthew Bochman, pers. comm.).

Filtrates of the starting wort media and the culture after five weeks were analyzed. The pH of the starting wort and the finished product reflect standard brewing measurements at 4.57 and 4.53 pH, respectively. These pH values are consistent with the avoidance of flavor and contamination problems (for example, a higher pH would signal food safety issues), indicating that the *Pichia* hybrid isn't typical of a spoilage species in this respect.

Extract analyses were completed to determine the amount of non-fermentable carbohydrates remaining in the product after culturing. According to the Real Extract test, the original wort was valued at 13.75 ºPlato and after 5 weeks at 13.61 ºPlato. The Apparent Extract/Final Gravity in an American style amber/red ale should be 2.5 - 4.6 ºPlato (Papazian 2017). The alcohol by weight in this style of beer should be 3.50% - 4.80%, whereas the *Pichia* hybrid product was 0.02%. The final calorie content per a 12 oz. serving of beer typically drops to approximately 160 kcal, whereas the hybrid fermentation product was valued at 202.68 kcal. In addition to this, percentages based on the entire contents of the media showed that over five weeks, glucose levels dropped from an initial measure of 1.62% to a final measure of 1.22%. The hybrid may incrementally breakdown maltotriose and fructose, dropping from 1.46% to 1.05% and 0.57% to 0.22% respectively, but did not appear to be able to reduce the levels of maltose. However, this test would not distinguish sugars created as metabolic byproducts from those present initially. Single sugar fermentation studies could be done in the future to better define the metabolic activities of the hybrid. These results do indicate that the *Pichia* hybrid did not significantly metabolize much of the available carbohydrates into alcohol within this wort environment, and may suggest that it's dependent on byproducts or enzymes from other organisms in a mixed population. These results may also reflect a lack of aeration, as we did not employ any aeration methods during the five week fermentation, and a recent study noted *P. kudriavzevii* is not able to grow in the absence of oxygen and therefore can only grow at early stages of fermentation (van Rijswijck et al. 2017).

Metagenomic deconvolution reveals novel strains and species of bacteria, yeast

In addition to this novel hybrid, we additionally identified and assembled whole genome sequences for three other species of yeast and four species of bacteria. These included the yeasts *Saccharomyces cerevisiae*, *Dekkera bruxellensis*, and *Pichia kudriavzevii*. *D. bruxellensis* and *P. kudriavzevii* are both generally associated with spoilage in beer and wine, though *D. bruxellensis* is strongly associated with spontaneous fermentation beers and is prevalent in these cultures after 4 to 8 months of fermentation (Vanoevelen et al. 1977; Verachtert and Iserentant 1995; Spitaels et al. 2014c; Steensels et al. 2015). *D. bruxellensis* is known to have considerable strain variation in chromosome number and ploidy with genome size estimates ranging dramatically; here, the

bioRxiv preprint doi: https://doi.org/10.1101/150722; this version posted August 31, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

size estimate is comparable to haploid *S. cerevisiae*, considerably smaller than the 20-30 Mb previously estimated (Hellborg and Piskur 2009), although this could partially be explained by incomplete assembly.

The *S. cerevisiae* sequence appears to be closely related to the Canadian beer strain BE018 and the domesticated European wine cluster (Strope et al. 2015; Gallone et al. 2016) (**Figure 3**), though analysis of allele frequencies indicated multiple strains may be present, or that the strain varies in ploidy. Although the original wort inoculation was spontaneous, at various points over the course of the Old Warehouse production the barrel was topped off with other beers which were brewed using commercial ale strains. At this point it is unclear if the *S. cerevisiae* isolated from this sample was initially present, or derived from these later additions.

Consistent with previous studies of spontaneous fermentation beers, we observe acetic acid bacteria (AAB) *Acetobacter pomorum* and a previously unidentified species of Acetobacter (OWB1), and lactic acid bacteria (LAB) *Pediococcus damnosus* and a previously unidentified species of Lactobacillus (OWB2), similar to *L. brevis*. Spitaels *et al.* recently identified two new species of AAB in a traditional Belgian lambic fermentation: *Acetobacter lambici* and *Glunobacter cerevisiae* (Spitaels et al. 2014a; Spitaels et al. 2014b), although we are unable to determine if OWB1 is related to *A. lambici* at this time. Both OWB1 and OWB2 have beer spoilage genes as identified by Fujii *et al.* (Fujii et al. 2005), who identified a locus specific to a wide variety of beer-spoilage strains including *P. damnosus*, *L. collinoides*, *L. coryniformis*, and *L. brevis*.

Discussion

Here, we leverage Hi-C to characterize a metagenomic sample isolated from a spontaneously inoculated beer from Epic Ales in Seattle, Washington. The method identified the yeasts *D. bruxellensis*, *P. kudriavzevii*, *S. cerevisiae*, and a novel *Pichia* hybrid composed of *P. membranifaciens* and an unidentified, related species. We also observed the AAB *Acetobacter pomorum* and a novel Acetobacter species, and the LAB *Pediococcus damnosus* and a novel Lactobacillus species. We present genome assemblies for each, without culturing or any a priori knowledge of community makeup. This represents one of the first studies to deconvolute an unknown mixed sample without culturing (see also (Marbouty et al. 2014; Marbouty et al. 2017)), and demonstrates the ability of this method to both identify novel species, and of particular interest, identify novel hybrids.

Spontaneous inoculation brewing

The traditional Belgian lambic style of fermentation dates back hundreds of years (Keersmaecker 1996), and has recently gained popularity in the American craft brewery community. Phenotypic identification and culturing, and more recently various genetic methods, have illustrated the microbial progression of these beers, which typically lasts 1-3 years. The first phase lasts for 30-40 days and is dominated by Enterobacteriaceae (Vanoevelen et al. 1977; Martens et al. 1991; Martens et al. 1997; Bokulich et al. 2012; Spitaels et al. 2014c); this is followed by the main fermentation, characterized by *Saccharomyces* yeasts (including *S. cerevisiae*, *S. pastorianus*, and *S. uvarum*) and lasts 3-4 months. The acidification phase succeeds fermentation and is characterized by the increasing isolation of LAB, and *Dekkera* species after 4 to 8 months of fermentation. Around 10 months, LAB start to decrease as the wort is gradually attenuated. AAB are isolated throughout the brewing process.

bioRxiv preprint doi: https://doi.org/10.1101/150722; this version posted August 31, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

In comparison to previous spontaneous inoculation studies, which typically sample over the course of 1-3 years, we only sampled one batch at a single time-point. The sample was obtained after new wort was added to a used wine barrel, from which several prior batches of the "Old Warehouse" beer had already been bottled. Thus, the mixed population we extracted had likely been established for several years in various configurations. We recapitulate some of the previous findings, and also identify several novel strains of bacteria. The diversity of AAB, LAB, and other yeasts found in cultures at various timepoints in other samples can be explained by multiple possibilities. It's possible that some species are present at undetectable levels in our sample, or that more species are present in the slurry of the barrel, or are embedded in the barrel itself, that were not captured with our sampling method, which only extracted actively growing culture from the aqueous phase. Furthermore, many researchers have found that microbiota are specific to a given environment (Knight et al. 2015; Jara et al. 2016), so differences between different breweries are to be expected. For example, traditional lambics are brewed in breweries in Belgium that produce no other style of beer, whereas in the United States, a different community of yeast and bacterial species are established in breweries that may be experimenting with lambic style fermentation (Bokulich et al. 2012).

While *S. cerevisiae* appears chiefly responsible for fermentation, mixed fermentations with non-*Saccharomyces* yeasts in wine and cider have been shown to produce desirable aroma and flavor complexity (Rojas et al. 2001; Clemente-Jimenez et al. 2005; Viana et al. 2008; Ye et al. 2014; van Rijswijck et al. 2017). The yeast *D. bruxellensis* helps characterize the "funky" flavor of spontaneous inoculation beers; it produces phenol compounds, described as clove-y or medicinal, and smaller compounds often described as barnyard or horse blanket. *D. bruxellensis* and *S. cerevisiae* actually share many important brewing traits, such as the ability to produce high ethanol levels, high tolerance towards ethanol, and the ability to grow without oxygen and in acidic environments (Rozpedowska et al. 2011), but they diverged approximately 200 million years ago, and are thought to have independently evolved these phenotypes (Woolfit et al. 2007). While *S. cerevisiae* outcompetes *D. bruxellensis* when sugar is abundant, the roles are changed at the end of fermentation, when sugar is depleted and only low amounts of nitrogen are available (Nardi et al. 2010).

D. bruxellensis is more closely related to the *Pichia* species than it is to the *Saccharomyces* yeasts (Woolfit et al. 2007; Piskur et al. 2012). Prior to reclassification, the *Pichia* genus included more than 100 species, but has been refined to 20, including the type species of the genus *P. membranifaciens*, and the distantly related *P. kudriavzevii* (Kurtzman et al. 2008; Kurtzman 2011). *P. membranifaciens* and *P. kudriavzevii* are most commonly associated with food and beverage spoilage (Kurtzman et al. 2011; Chan et al. 2012), but seem to play a role in various fermentations (Lachance 1995; Daniel et al. 2009; Silva et al. 2009; N'Guessan K et al. 2011; Del Monaco et al. 2014; Spitaels et al. 2015; Stefanini et al. 2016a; Varela and Borneman 2017). A study of wine deacidification showed that *P. kudriavzevii* increased a wine's fruity aroma and could degrade L-malic acid, increase the pH, and produce low levels of ethanol, important levels of glycerol, and acceptable amounts of acetic acid (Del Monaco et al. 2014). *P. membranifaciens* is repeatedly isolated from lambic and gueuze beers (Spitaels et al. 2014c; Spitaels et al. 2015). Attempts to ferment with the *Pichia* hybrid discovered in this sample suggest that it does not significantly contribute to ethanol production, nor did it change the acidity of the wort; however, it might contribute flavor compounds and/or act in concert with other members of the community. Though recent work suggests sensory

bioRxiv preprint doi: https://doi.org/10.1101/150722; this version posted August 31, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

profiles of beers fermented with *Pichia* yeasts are undesirable (Osburn et al. 2017), we did not detect an obviously objectionable profile.

Conclusions

The discovery of a new interspecific hybrid from a brewing environment adds to an increasing number of similarly documented cases, raising several interesting questions. Are there particular conditions in brewing environments that select for hybridization events? For example, a recent study found that chemical-physical conditions in the gut of wasps selected for hybridization events between *S. cerevisiae* and *S. paradoxus* (Stefanini et al. 2016b). Brewing environments could also simply be enriched in both population size and diversity of species, thus presenting more opportunity for hybridization than their natural environments. Perhaps previously established hybrids have a selective advantage in brewing environments, which there is some evidence for with *Dekkera* hybrids in wine environments (Borneman et al. 2014). Alternatively, the explanation could simply be sampling bias: yeast hybrids happen in many places but breweries and wineries are heavily sampled. These hypotheses are not mutually exclusive, and present a rich field for further study. With new detection tools, like the method presented here, we have the ability to identify new hybrids from a variety of taxa and environments, thus providing a critical tool to understanding hybrid biology and ecology.

Acknowledgements

We thank Benjamin Vernot for initiating the collaboration with Epic Ales. We appreciate the feedback we received from two anonymous reviewers, and from many brewers, particularly Matthew L. Bochman, Ron Extract, and Tom Schmidlin. This work was supported by National Science Foundation grant 1516330 (M.J.D.), National Institute of Health (NIH)/National Human Genome Research Institute (NHGRI) grant T32HG000035 (J.N.B.), NIH/NHGRI grant HG006283 (J.S.), NIH/National Institute of General Medical Sciences grant P41 GM103533 (M.J.D.), DOE-LBL-JGI grant 7074345/DE-AC02-05CH11231 (J.S.), and NIAID grant 1R43AI122654 (I.L.). I.L. was also supported by the UW Commercialization Gap Fund and Commercialization Fellows Program. M.J.D. is a Fellow in the Genetic Networks program at the Canadian Institute for Advanced Research and is supported in part by a Faculty Scholar grant from the Howard Hughes Medical Institute.

bioRxiv preprint doi: https://doi.org/10.1101/150722; this version posted August 31, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

References

Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJ, Bierne N, Boughman J, Brelsford A, Buerkle CA, Buggs R et al. 2013. Hybridization and speciation. *J Evol Biol* **26**(2): 229-246.

Albertin W, Marullo P. 2012. Polyploidy in fungi: evolution after whole-genome duplication. *P Roy Soc B-Biol Sci* **279**(1738): 2497-2509.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic Local Alignment Search Tool. *Journal of Molecular Biology* **215**(3): 403-410.

Baker E, Wang B, Bellora N, Peris D, Hulfachor AB, Koshalek JA, Adams M, Libkind D, Hittinger CT. 2015. The genome sequence of *Saccharomyces eubayanus* and the domestication of lager-brewing yeasts. *Molecular Biology and Evolution* **32**(11): 2818-2831.

Beitel CW, Froenicke L, Lang JM, Korf IF, Michelmore RW, Eisen JA, Darling AE. 2014. Strain- and plasmid-level deconvolution of a synthetic metagenome by sequencing proximity ligation products. *PeerJ* **2**: e415.

Bellon JR, Yang F, Day MP, Inglis DL, Chambers PJ. 2015. Designing and creating *Saccharomyces* interspecific hybrids for improved, industry relevant, phenotypes. *Applied Microbiology and Biotechnology* **99**(20): 8597-8609.

Bokulich NA, Bamforth CW, Mills DA. 2012. Brewhouse-resident microbiota are responsible for multi-stage fermentation of American coolship ale. *PLoS One* **7**(4): e35507.

Borneman AR, Zeppel R, Chambers PJ, Curtin CD. 2014. Insights into the *Dekkera bruxellensis* genomic landscape: Comparative genomics reveals variations in ploidy and nutrient utilisation potential amongst wine isolates. *PLoS Genetics* **10**(2): e1004161.

Bullini L. 1994. Origin and evolution of animal hybrid species. *Trends in Ecology & Evolution* **9**(11): 422-426.

Burton JN, Adey A, Patwardhan RP, Qiu R, Kitzman JO, Shendure J. 2013. Chromosome-scale scaffolding of *de novo* genome assemblies based on chromatin interactions. *Nature Biotechnology* **31**(12): 1119-1125.

Burton JN, Liachko I, Dunham MJ, Shendure J. 2014. Species-level deconvolution of metagenome assemblies with Hi-C-based contact probability maps. *G3* **4**(7): 1339-1346.

Carr R, Shen-Orr SS, Borenstein E. 2013. Reconstructing the genomic content of microbiome taxa through shotgun metagenomic deconvolution. *PLoS Comput Biol* **9**(10): e1003292.

Chan GF, Gan HM, Ling HL, Rashid NA. 2012. Genome sequence of *Pichia kudriavzevii* M12, a potential producer of bioethanol and phytase. *Eukaryotic Cell* **11**(10): 1300-1301.

Clemente-Jimenez JM, Mingorance-Cazorla L, Martinez-Rodriguez S, Las Heras-Vazquez FJ, Rodriguez-Vico F. 2005. Influence of sequential yeast mixtures on wine fermentation. *Int J Food Microbiol* **98**(3): 301-308.

Daniel HM, Vrancken G, Takrama JF, Camu N, De Vos P, De Vuyst L. 2009. Yeast diversity of Ghanaian cocoa bean heap fermentations. *FEMS Yeast Res* **9**(5): 774-783.

Dekker J, Marti-Renom MA, Mirny LA. 2013. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nature Reviews Genetics* **14**(6): 390-403.

Del Monaco SM, Barda NB, Rubio NC, Caballero AC. 2014. Selection and characterization of a Patagonian *Pichia kudriavzevii* for wine deacidification. *Journal of Applied Microbiology* **117**(2): 451-464.

bioRxiv preprint doi: https://doi.org/10.1101/150722; this version posted August 31, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Dowling TE, Secor CL. 1997. The role of hybridization and introgression in the diversification of animals. *Annu Rev Ecol Syst* **28**: 593-619.

Dunn B, Sherlock G. 2008. Reconstruction of the genome origins and evolution of the hybrid lager yeast *Saccharomyces pastorianus*. *Genome Res* **18**(10): 1610-1623.

Felsenstein J. 2005. PHYLIP (Phylogeny Inference Package).

Fujii T, Nakashima K, Hayashi N. 2005. Random amplified polymorphic DNA-PCR based cloning of markers to identify the beer-spoilage strains of *Lactobacillus brevis*, *Pediococcus damnosus*, *Lactobacillus collinoides* and *Lactobacillus coryniformis*. *Journal of Applied Microbiology* **98**(5): 1209-1220.

Gallone B, Steensels J, Prahl T, Soriaga L, Saels V, Herrera-Malaver B, Merlevede A, Roncoroni M, Voordeckers K, Miraglia L et al. 2016. Domestication and divergence of *Saccharomyces cerevisiae* beer yeasts. *Cell* **166**(6): 1397-1410 e1316.

Gibson B, Liti G. 2015. *Saccharomyces pastorianus*: genomic insights inspiring innovation for industry. *Yeast* **32**(1): 17-27.

Gonzalez SS, Barrio E, Gafner J, Querol A. 2006. Natural hybrids from *Saccharomyces cerevisiae*, *Saccharomyces bayanus* and *Saccharomyces kudriavzevii* in wine fermentations. *FEMS Yeast Res* **6**(8): 1221-1234.

Gonzalez SS, Barrio E, Querol A. 2008. Molecular characterization of new natural hybrids of *Saccharomyces cerevisiae* and *S. kudriavzevii* in brewing. *Appl Environ Microb* **74**(8): 2314-2320.

Handelsman J. 2004. Metagenomics: Application of genomics to uncultured microorganisms. *Microbiol Mol Biol R* **68**(4): 669-685.

Hellborg L, Piskur J. 2009. Complex nature of the genome in a wine spoilage yeast, *Dekkera bruxellensis*. *Eukaryotic Cell* **8**(11): 1739-1749.

Hoffmann AA, Sgro CM. 2011. Climate change and evolutionary adaptation. *Nature* **470**(7335): 479-485.

Howe AC, Jansson JK, Malfatti SA, Tringe SG, Tiedje JM, Brown CT. 2014. Tackling soil diversity with the assembly of large, complex metagenomes. *Proceedings of the National Academy of Sciences of the United States of America* **111**(13): 4904-4909.

Hug LA, Castelle CJ, Wrighton KC, Thomas BC, Sharon I, Frischkorn KR, Williams KH, Tringe SG, Banfield JF. 2013. Community genomic analyses constrain the distribution of metabolic traits across the Chloroflexi phylum and indicate roles in sediment carbon cycling. *Microbiome* **1**(1): 22.

Human Microbiome Project C. 2012. Structure, function and diversity of the healthy human microbiome. *Nature* **486**(7402): 207-214.

Iverson V, Morris RM, Frazar CD, Berthiaume CT, Morales RL, Armbrust EV. 2012. Untangling genomes from metagenomes: revealing an uncultured class of marine Euryarchaeota. *Science* **335**(6068): 587-590.

Jara C, Laurie VF, Mas A, Romero J. 2016. Microbial terroir in Chilean valleys: Diversity of non-conventional yeast. *Frontiers in Microbiology* **7**: 663.

Kamke J, Sczyrba A, Ivanova N, Schwientek P, Rinke C, Mavromatis K, Woyke T, Hentschel U. 2013. Single-cell genomics reveals complex carbohydrate degradation patterns in poribacterial symbionts of marine sponges. *The ISME Journal* **7**(12): 2287-2300.

Kaplan N, Dekker J. 2013. High-throughput genome scaffolding from *in vivo* DNA interaction frequency. *Nature Biotechnology* **31**(12): 1143-1147.

bioRxiv preprint doi: https://doi.org/10.1101/150722; this version posted August 31, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Keersmaecker JD. 1996. The mystery of lambic beer. In *Scientific American*, Vol August, pp. 74-80.

Kelly B, Whiteley A, Tallmon D. 2010. The Arctic melting pot. *Nature* **468**(7326): 891-891.

Knight S, Klaere S, Fedrizzi B, Goddard MR. 2015. Regional microbial signatures positively correlate with differential wine phenotypes: evidence for a microbial aspect to terroir. *Scientific Reports* **5**: 14233.

Kurtzman CP. 2011. Phylogeny of the ascomycetous yeasts and the renaming of *Pichia anomala* to *Wickerhamomyces anomalus*. *Antonie van Leeuwenhoek* **99**(1): 13-23.

Kurtzman CP, Fell JW, Boekhout T. 2011. *The Yeasts, a Taxonomic Study. Vol 1*. Elsevier.

Kurtzman CP, Robnett CJ, Basehoar-Powers E. 2008. Phylogenetic relationships among species of *Pichia*, *Issatchenkia* and *Williopsis* determined from multigene sequence analysis, and the proposal of *Barnettozyma* gen. nov., *Lindnera* gen. nov. and *Wickerhamomyces* gen. nov. *FEMS Yeast Res* **8**(6): 939-954.

Lachance MA. 1995. Yeast communities in a natural tequila fermentation. *Antonie van Leeuwenhoek* **68**(2): 151-160.

Lane DJ, Pace B, Olsen GJ, Stahl DA, Sogin ML, Pace NR. 1985. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proceedings of the National Academy of Sciences of the United States of America* **82**(20): 6955-6959.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**(14): 1754-1760.

Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO et al. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**(5950): 289-293.

Magalhaes F, Krogerus K, Vidgren V, Sandell M, Gibson B. 2017. Improved cider fermentation performance and quality with newly generated *Saccharomyces cerevisiae* x *Saccharomyces eubayanus* hybrids. *Journal of Industrial Microbiology & Biotechnology*. **44**(8): 1203-1213.

Mallet J. 2007. Hybrid speciation. *Nature* **446**(7133): 279-283.

Marbouty M, Baudry L, Cournac A, Koszul R. 2017. Scaffolding bacterial genomes and probing host-virus interactions in gut microbiome by proximity ligation (chromosome capture) assay. *Science Advances* **3**(2): e1602105.

Marbouty M, Cournac A, Flot JF, Marie-Nelly H, Mozziconacci J, Koszul R. 2014. Metagenomic chromosome conformation capture (meta3C) unveils the diversity of chromosome organization in microorganisms. *eLife* **3**: e03318.

Marcy Y, Ouverney C, Bik EM, Losekann T, Ivanova N, Martin HG, Szeto E, Platt D, Hugenholtz P, Relman DA et al. 2007. Dissecting biological "dark matter" with single-cell genetic analysis of rare and uncultivated TM7 microbes from the human mouth. *Proceedings of the National Academy of Sciences of the United States of America* **104**(29): 11889-11894.

Martens H, Dawoud E, Verachtert H. 1991. Wort Enterobacteria and other microbial populations involved during the first month of lambic fermentation. *J I Brewing* **97**(6): 435-439.

Martens H, Iserentant D, Verachtert H. 1997. Microbiological aspects of a mixed yeast-bacterial fermentation in the production of a special Belgian acidic ale. *J I Brewing* **103**(2): 85-91.

bioRxiv preprint doi: https://doi.org/10.1101/150722; this version posted August 31, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Muhlfeld CC, Kovach RP, Jones LA, Al-Chokhachy R, Boyer MC, Leary RF, Lowe WH, Luikart G, Allendorf FW. 2014. Invasive hybridization in a threatened species is accelerated by climate change. *Nat Clim Change* **4**(7): 620-624.

N'Guessan K F, Brou K, Jacques N, Casaregola S, Dje KM. 2011. Identification of yeasts during alcoholic fermentation of tchapalo, a traditional sorghum beer from Cote d'Ivoire. *Antonie van Leeuwenhoek* **99**(4): 855-864.

Nakao Y, Kanamori T, Itoh T, Kodama Y, Rainieri S, Nakamura N, Shimonaga T, Hattori M, Ashikari T. 2009. Genome sequence of the lager brewing yeast, an interspecies hybrid. *DNA Res* **16**(2): 115-129.

Nardi T, Remize F, Alexandre H. 2010. Adaptation of yeasts *Saccharomyces cerevisiae* and *Brettanomyces bruxellensis* to winemaking conditions: a comparative study of stress genes expression. *Applied Microbiology and Biotechnology* **88**(4): 925-937.

Osburn K, Amaral J, Metcalf SR, Nickens DM, Rogers CM, Sausen C, Caputo R, Miller J, Li H, Tennessen JM et al. 2017. Primary souring: a novel bacteria-free method for sour beer production. *bioRxiv*.

Pace NR, Stahl DA, Lane DJ, Olsen GJ. 1986. The Analysis of Natural Microbial-Populations by Ribosomal RNA Sequences. *Adv Microb Ecol* **9**: 1-55.

Papazian C. 2017. Brewers Association Beer Style Guidelines. Brewers Association.

Peng Y, Leung HC, Yiu SM, Chin FY. 2012. IDBA-UD: a *de novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **28**(11): 1420-1428.

Piskur J, Ling Z, Marcet-Houben M, Ishchuk OP, Aerts A, LaButti K, Copeland A, Lindquist E, Barry K, Compagno C et al. 2012. The genome of wine yeast *Dekkera bruxellensis* provides a tool to explore its food-related properties. *Int J Food Microbiol* **157**(2): 202-209.

Pryszcz LP, Nemeth T, Gacser A, Gabaldon T. 2014. Genome comparison of *Candida orthopsilosis* clinical strains reveals the existence of hybrids between two distinct subspecies. *Genome Biology and Evolution* **6**(5): 1069-1078.

Rajendhran J, Gunasekaran P. 2011. Microbial phylogeny and diversity: Small subunit ribosomal RNA sequence analysis and beyond. *Microbiol Res* **166**(2): 99-110.

Rankine BC. 1966. *Pichia membranaefaciens*, a yeast causing film formation and off-flavor in table wine. *American Journal of Enology and Viticulture* **17**(2): 82-86.

Riley R, Haridas S, Wolfe KH, Lopes MR, Hittinger CT, Goker M, Salamov AA, Wisecaver JH, Long TM, Calvey CH et al. 2016. Comparative genomics of biotechnologically important yeasts. *Proceedings of the National Academy of Sciences of the United States of America* **113**(35): 9882-9887.

Rojas V, Gil JV, Pinaga F, Manzanares P. 2001. Studies on acetate ester production by non-*Saccharomyces* wine yeasts. *Int J Food Microbiol* **70**(3): 283-289.

Rozpedowska E, Hellborg L, Ishchuk OP, Orhan F, Galafassi S, Merico A, Woolfit M, Compagno C, Piskur J. 2011. Parallel evolution of the make-accumulate-consume strategy in *Saccharomyces* and *Dekkera* yeasts. *Nature Communications* **2**: 302.

Saeed I, Tang SL, Halgamuge SK. 2012. Unsupervised discovery of microbial population structure within metagenomes using nucleotide base composition. *Nucleic Acids Res* **40**(5): e34.

bioRxiv preprint doi: https://doi.org/10.1101/150722; this version posted August 31, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li WZ, Lopez R, McWilliam H, Remmert M, Soding J et al. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**: 539.

Silva CL, Vianna CR, Cadete RM, Santos RO, Gomes FC, Oliveira ES, Rosa CA. 2009. Selection, growth, and chemo-sensory evaluation of flocculent starter culture strains of *Saccharomyces cerevisiae* in the large-scale production of traditional Brazilian cachaca. *International Journal of Food Microbiology* **131**(2-3): 203-210.

Snauwaert I, Stragier P, De Vuyst L, Vandamme P. 2015. Comparative genome analysis of *Pediococcus damnosus* LMG 28219, a strain well-adapted to the beer environment. *BMC Genomics* **16**: 267.

Spitaels F, Li L, Wieme A, Balzarini T, Cleenwerck I, Van Landschoot A, De Vuyst L, Vandamme P. 2014a. *Acetobacter lambici* sp. nov., isolated from fermenting lambic beer. *International Journal of Systematic and Evolutionary Microbiology* **64**(Pt 4): 1083-1089.

Spitaels F, Van Kerrebroeck S, Wieme AD, Snauwaert I, Aerts M, Van Landschoot A, De Vuyst L, Vandamme P. 2015. Microbiota and metabolites of aged bottled gueuze beers converge to the same composition. *Food Microbiology* **47**: 1-11.

Spitaels F, Wieme A, Balzarini T, Cleenwerck I, Van Landschoot A, De Vuyst L, Vandamme P. 2014b. *Gluconobacter cerevisiae* sp. nov., isolated from the brewery environment. *International Journal of Systematic and Evolutionary Microbiology* **64**(Pt 4): 1134-1141.

Spitaels F, Wieme AD, Janssens M, Aerts M, Daniel HM, Van Landschoot A, De Vuyst L, Vandamme P. 2014c. The microbial diversity of traditional spontaneously fermented lambic beer. *PLoS One* **9**(4): e95384.

Stanke M, Morgenstern B. 2005. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res* **33**(Web Server issue): W465-467.

Stanke M, Steinkamp R, Waack S, Morgenstern B. 2004. AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Res* **32**(Web Server issue): W309-312.

Steensels J, Daenen L, Malcorps P, Derdelinckx G, Verachtert H, Verstrepen KJ. 2015. *Brettanomyces* yeasts–From spoilage organisms to valuable contributors to industrial fermentations. *Int J Food Microbiol* **206**: 24-38.

Stefanini I, Albanese D, Cavazza A, Franciosi E, De Filippo C, Donati C, Cavalieri D. 2016a. Dynamic changes in microbiota and mycobiota during spontaneous 'Vino Santo Trentino' fermentation. *Microbial Biotechnology* **9**(2): 195-208.

Stefanini I, Dapporto L, Berna L, Polsinelli M, Turillazzi S, Cavalieri D. 2016b. Social wasps are a *Saccharomyces* mating nest. *Proceedings of the National Academy of Sciences of the United States of America* **113**(8): 2247-2251.

Stepanauskas R. 2012. Single cell genomics: an individual look at microbes. *Current Opinion in Microbiology* **15**(5): 613-620.

Strope PK, Skelly DA, Kozmin SG, Mahadevan G, Stone EA, Magwene PM, Dietrich FS, McCusker JH. 2015. The 100-genomes strains, an *S. cerevisiae* resource that illuminates its natural phenotypic and genotypic variation and emergence as an opportunistic pathogen. *Genome Res* **25**(5): 762-774.

Tamai Y, Momma T, Yoshimoto H, Kaneko Y. 1998. Co-existence of two types of chromosome in the bottom fermenting yeast, *Saccharomyces pastorianus*. *Yeast* **14**(10): 923-933.

Vallejo-Marin M, Hiscock SJ. 2016. Hybridization and hybrid speciation under global change. *New Phytol* **211**(4): 1170-1187.

bioRxiv preprint doi: https://doi.org/10.1101/150722; this version posted August 31, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

van Rijswijck IMH, Wolkers-Rooijackers JCM, Abee T, Smid EJ. 2017. Performance of non-conventional yeasts in co-culture with brewers' yeast for steering ethanol and aroma production. *Microbial Biotechnology*.

Vanoevelen D, Spaepen M, Timmermans P, Verachtert H. 1977. Microbiological aspects of spontaneous wort fermentation in production of lambic and gueuze. *J I Brewing* **83**(6): 356-360.

Varela C, Borneman AR. 2017. Yeasts found in vineyards and wineries. *Yeast* **34**(3): 111-128.

Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, Wu D, Paulsen I, Nelson KE, Nelson W et al. 2004. Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**(5667): 66-74.

Verachtert H, Iserentant D. 1995. Properties of Belgian acid beers and their microflora. Part I. The production of gueuze and related refreshing acid beers. *Cerevisia, Belgian Journal of Brewing and Biotechnology* **20**: 37-41.

Viana F, Gil JV, Genoves S, Valles S, Manzanares P. 2008. Rational selection of non-*Saccharomyces* wine yeasts for mixed starters based on ester formation and enological traits. *Food Microbiology* **25**(6): 778-785.

Wolfe KH. 2015. Origin of the yeast whole-genome duplication. *PLoS Biol* **13**(8): e1002221.

Wood DE, Salzberg SL. 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biology* **15**(3).

Woolfit M, Rozpedowska E, Piskur J, Wolfe KH. 2007. Genome survey sequencing of the wine spoilage yeast *Dekkera* (*Brettanomyces*) *bruxellensis*. *Eukaryotic Cell* **6**(4): 721-733.

Yamagishi H, Ogata T. 1999. Chromosomal structures of bottom fermenting yeasts. *Syst Appl Microbiol* **22**(3): 341-353.

Ye M, Yue T, Yuan Y. 2014. Effects of sequential mixed cultures of *Wickerhamomyces anomalus* and *Saccharomyces cerevisiae* on apple cider fermentation. *FEMS Yeast Res* **14**(6): 873-882.

bioRxiv preprint doi: https://doi.org/10.1101/150722; this version posted August 31, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure Legends

Figure 1: MetaPhase deconvolutes species from a spontaneously inoculated beer sample
A. Clustering of the Old Warehouse assembly. Each contig is shown as a dot, with size indicating contig length, colored by species. Edge widths represent the densities of Hi-C links between the contigs shown. **B.** Determining the approximate number of species in the metagenome. We ran the hierarchical agglomerative clustering algorithm on the draft assembly. In this algorithm, the number of clusters starts high and gradually decreases as clusters are merged together; to generate this data, we continued clustering all the way down to N = 1 cluster. Shown is the metric *E*, or intra-cluster link enrichment, at each value of *N*. It is assumed that the maximum value of *E*(*N*) occurs when *N* is roughly equal to the true number of distinct species present in the draft assembly. For this assembly, this occurs at *N* = 8 clusters. **C.** Validation. This heatmap indicates what fraction of the sequence in each MetaPhase cluster maps uniquely to each of the reference genomes of the predicted species. Note that not all sequence is expected to map uniquely to one species. x-axis: the 7 species. y-axis: the MetaPhase clusters. Note, this excludes the *P. damnosus*-like cluster because there was no public *P. damnosus* reference genome available.

Figure 2: A novel interspecific *Pichia* hybrid
A. Isolated colonies of the Pichia hybrid show distinct colony morphology. **B.** Flow cytometry indicates the hybrid genome (blue) size is approximately 24 Mb, similar to G2 haploid *S. cerevisiae*. **C.** Growth in wort shows aggregates on the flask side. **D.** Growth in wort shows palisades pattern.

Figure 3: Phylogeny of *S. cerevisiae*
Neighbor-joining tree of *S. cerevisiae* strains representative of the species diversity (Strope et al. 2015; Gallone et al. 2016) and OW (for Old Warehouse) *S. cerevisiae* strain (red).

Table 1: Species identified in spontaneously inoculated beer

Cluster	N contigs	Total length (bp)	N50 (bp)	Percent of reference covered by assembly	Abundance	Plurality reference	%eukaryotic	Plurality taxonomy
1	1716	9251667	8278	69.03%	9.25%	#0 (SC-FY)	100.00%	*Saccharomyces cerevisiae*
2	3367	8294670	4247	55.32%	16.16%	#1 (DB)	99.69%	*Dekkera bruxellensis*
3	3822	16898196	6415	64.01%	19.86%	#2 (PM)	96.43%	*Pichia membranifaciens*
4	1068	9135821	12637	80.40%	15.16%	#3 (PK)	100.00%	*Pichia kudriavzevii*
5	285	286331	69866	NA	0.08%	#4 (OWB1)	0.00%	Acetobacter; *Gluconacetobacter xylinus*
6	212	2618979	12000	47.39%	1.36%	#5 (Ap)	0.00%	Bacteria; a-proteobacteria; *Acetobacter pasteurianus*
7	501	941088	2554	NA	0.43%	#6 (OWB2)	0.00%	Bacteria; firmicutes; *Lactobacillus brevis*
8	218	2071872	15372	84.53%	11.68%	#6 (OWB2)	0.00%	Bacteria; firmicutes; *Pediococcus damnosus*

bioRxiv preprint doi: https://doi.org/10.1101/150722; this version posted August 31, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

bioRxiv preprint doi: https://doi.org/10.1101/150722; this version posted August 31, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure 1

Smukowski Heil

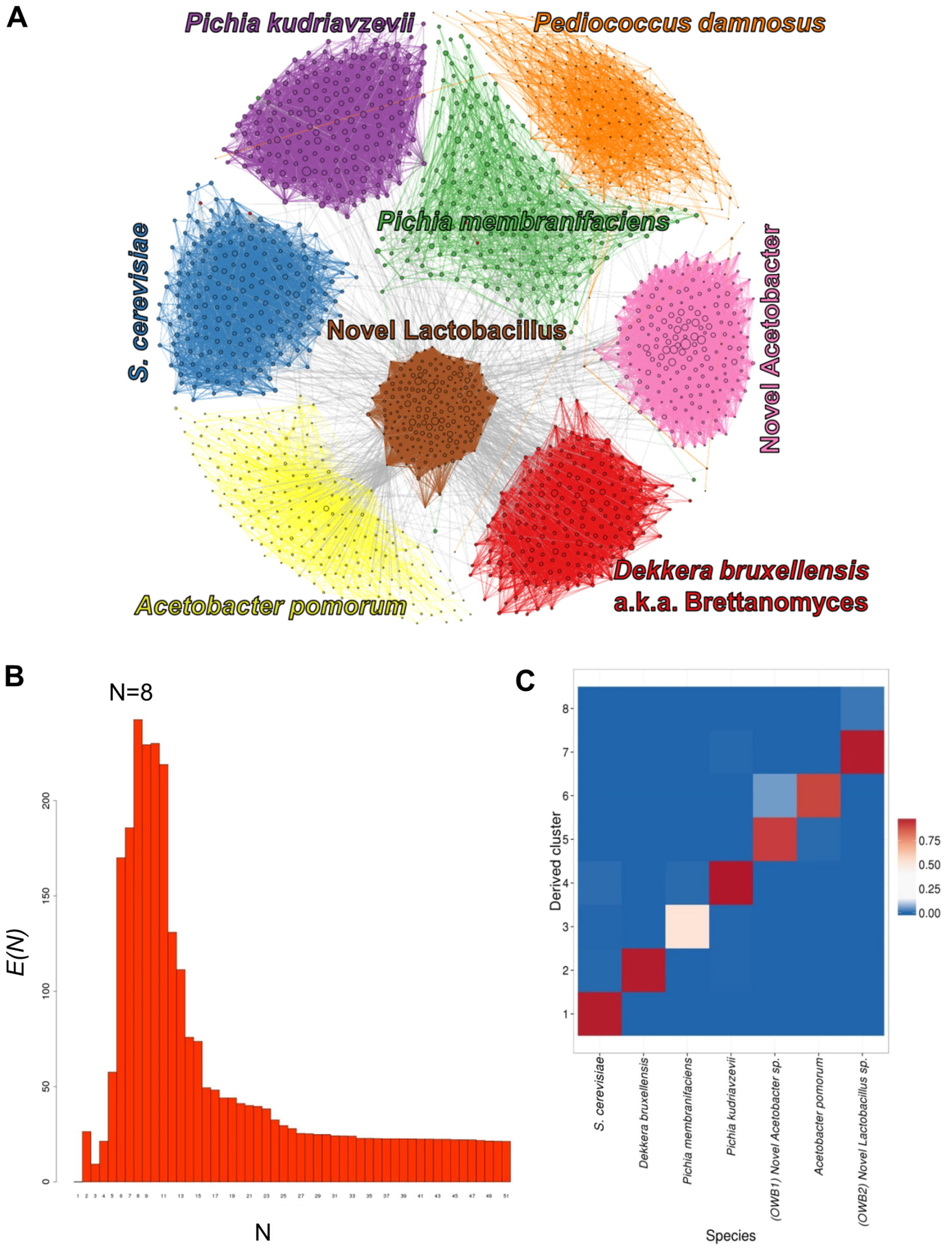
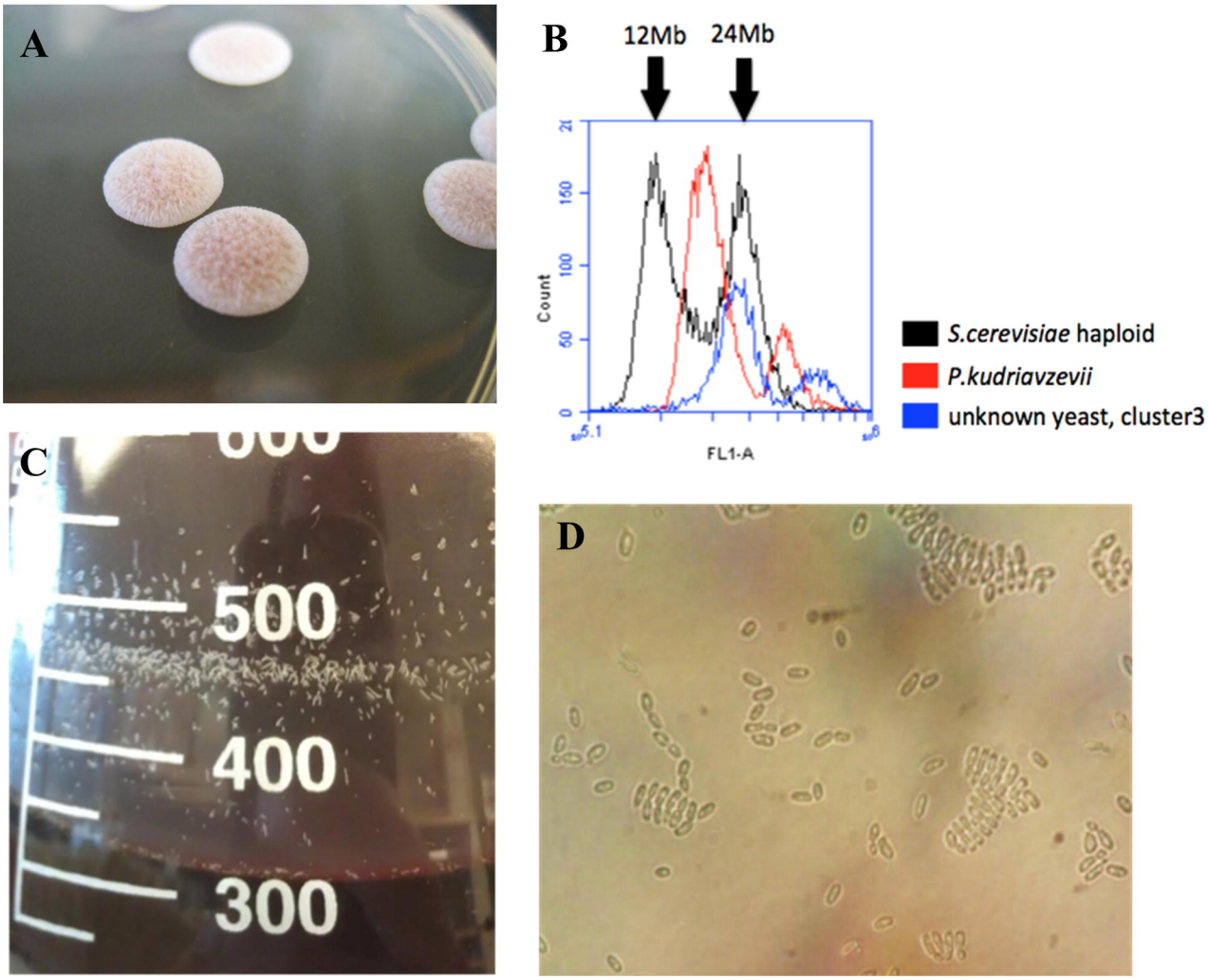
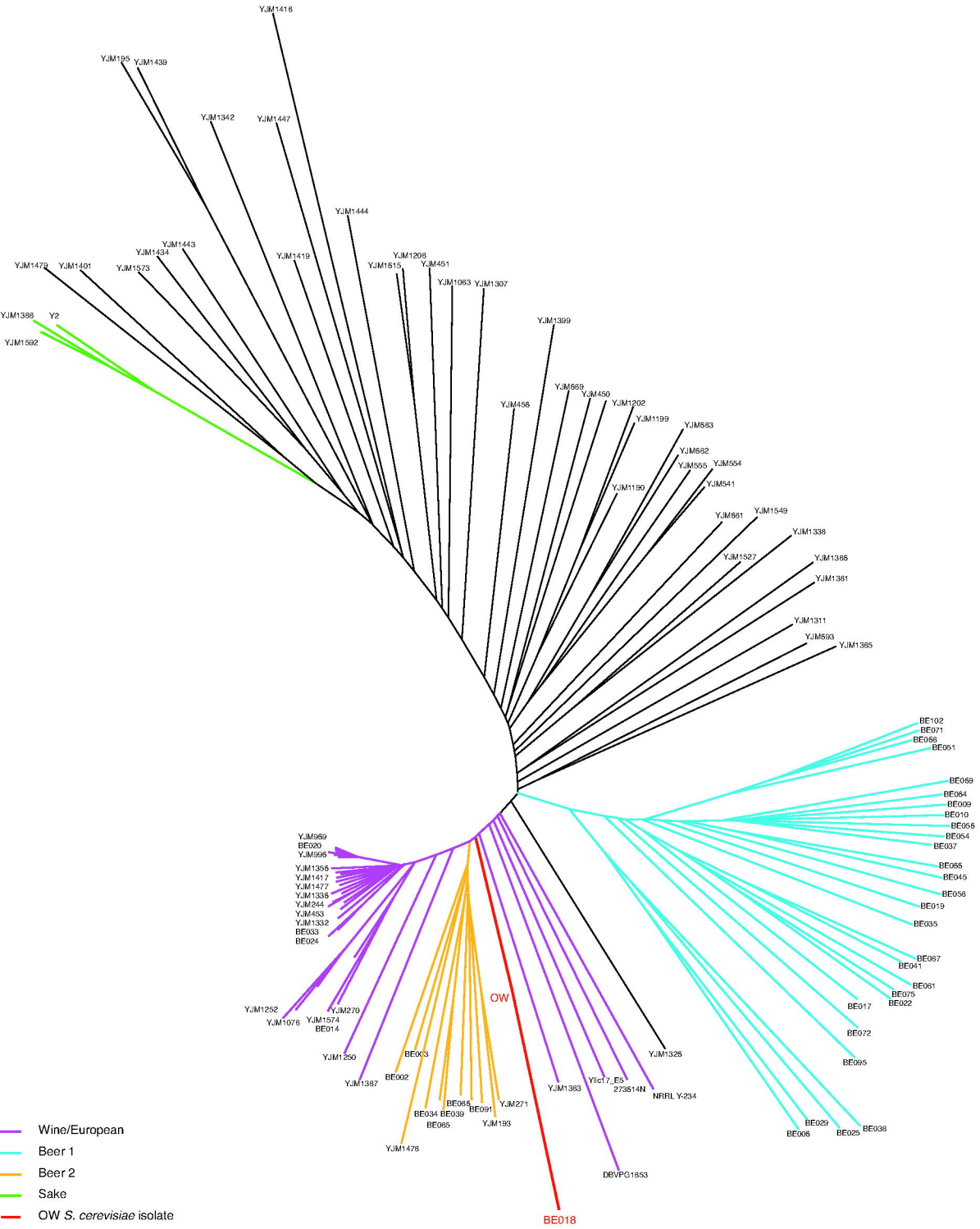


Figure 2

Smukowski Heil



bioRxiv preprint doi: https://doi.org/10.1101/150722; this version posted August 31, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.