ANALYSIS OF YEAST DNA SEQUENCES TO DETERMINE YEAST STRAINS UNIQUE TO TEXAS THAT ARE OPTIMAL FOR THE PRODUCTION OF CRAFT DISTILLED SPIRITS

by

Elizabeth T. Do

Submitted in partial fulfillment of the

requirements for Departmental Honors in

the Department of Biology

Texas Christian University

Fort Worth, Texas

May 4, 2015

ANALYSIS OF YEAST DNA SEQUENCES TO DETERMINE

YEAST STRAINS UNIQUE TO TEXAS

THAT ARE OPTIMAL FOR

THE PRODUCTION OF CRAFT DISTILLED SPIRITS

Project Approved:

Supervising Professor: Dean Williams, Ph.D.

Department of Biology

Matthew Chumchal, Ph.D.

Department of Biology

Julie Fry, Ph.D.

Department of Chemistry

ABSTRACT

Craft distilling in Texas is a growing business where start-up distilleries are interested in finding yeast strains that will give craft spirits a unique and fine taste. This study genetically characterized yeast strains of *Saccharomyces cerevisiae* that have the potential to distill craft spirits, in order to determine how unique they were compared to other known strains. We extracted DNA from 11 yeast strains originally isolated from local plants at a ranch in North Texas. We then sequenced three different loci: NUP116 gene region, LAS1 gene region and the IntAY intergenic spacer. The DNA sequences were compared to sequences from 124 other yeast strains, obtained from GenBank, that were isolated from three geographic locations: China, France, and lab strains of baker's yeast. We constructed 95% parsimony networks from the DNA sequences to determine how the strains were related to each other. The most promising candidate strain for alcohol production, based on studies conducted by professionals in the craft distillation industry, was unique and most closely related to a yeast strain used to produce wine. This suggests that this strain may be unique to Texas.

TABLE OF CONTENTS

INTRODUCTION……………………………………………………………………...1

METHODS……………………………………………………………………………..3

DNA Sampling and Strain Isolation………………………………………………3

Polymerase Chain Reaction………………………………………………………4

DNA Sequencing…………………………………………………………………6

DNA Manipulation, Subsequent Analysis and Phylogenetic Study……………...6

RESULTS………………………………………………………………………………9

DISCUSSION…………………………………………………………………………21

APPENDIX A: List of sequences obtained from GenBank with respective accession codes ……………………………………………………………………………….25

REFERENCES………………………………………………………………………..27

INTRODUCTION

The species *Saccharomyces cerevisiae*, commonly known as yeast, is one of the most important microorganisms in our society today. Yeast plays a major role in the food industry due to its ability to undergo fermentation – the chemical process of converting sugars into alcohol, a key component in the production of certain foods and beverages such as breads and alcoholic drinks. The use of yeast in food and drink production dates back thousands of years (Duncan and Leu 2009). However, only recently have we gained the scientific knowledge about yeast that helps us optimize its use. From Lavoisier's discovery in the late 1700's that yeasts are in fact living organisms (Barnett 1998) to the biological and chemical characterization of yeast and Pasteur's discovery of alcoholic fermentation in the mid 1800's (Barnett 2000), we now have access to information and techniques that allow us to maximize the use of yeast in our society.

We have developed variants, or "strains," of the *Saccharomyces cerevisiae* yeast species. A strain, defined to be "made up of descendants of a single isolation in pure culture and usually is made up of a succession of cultures ultimately derived from an initial single colony" (Dijshoorn et al. 2000), may be distinguished from other strains but ultimately all strains are of the same species. Particularly, yeast strains have been specialized based on their ability to produce different foods. Today, yeast strains such as baker's yeast and wine yeast are used frequently in the production of their respective products. This suggests that different strains of yeast exhibit different qualities and characteristics that make them better for producing certain foods. Within certain food categories, there are different strains of yeasts. For instance, there are multiple yeast strains used to ferment different kinds of liquors such as wine yeast, beer yeast and sake

yeast. Moreover, the discovery of pure cultures and the development of strain collection techniques have been vital in the microbiology and biotechnology field as we have been able to target these specific desired traits (Roberts et al. 2010). Subsequently, we have domesticated strains through isolation and culturing of respective strains. In this case, with the growing knowledge of biological and chemical traits of yeasts, we have selected for advantageous yeast traits through cross-breeding and production – identifying desired traits for fermentation and allowing individuals with these desired traits to mate (Liti et al. 2009). Consequently, this has resulted in the proliferation of these traits and the aforementioned frequently used commercial strains that optimize the production of food and beverage today.

In addition to intentional cross-breeding of yeast in the lab, human activity has allowed for changes in the genetic composition and differentiation of yeasts by geographic location. This is because strains cross-bred due to human movement of yeasts through trade (Wang et al. 2012). Specifically, for instance, some wine, beer, and distilled spirits were first introduced to the United States from Europe. This created the opportunity for yeast strains from the United States to cross-breed with those from Europe. Therefore, yeast from human-associated fermentations may actually be the source strain of some wild strains found in nature (Fay and Benavides 2005).

Population genetic studies of yeast using genetic markers such as single nucleotide polymorphisms (SNPs) have shown that yeast strains exhibit genetic differentiation by strain and geographic location (Liti et al. 2009). Additionally, past studies have shown that there is great genetic variation amongst liquor-specific strains such as in a yeast strain that produces Bacanora (Álvarez-Ainza et al. 2015). The use of

genetic information to differentiate yeast strains allows head distillers and brewmasters to select for specialized traits and domesticate respective yeast strains that are optimal for producing alcoholic beverages. In fact, craft distilling in Texas is a growing business where start-up distilleries are interested in finding yeast strains that will give craft spirits a unique and fine taste. In this study, we aimed to genetically characterize wild yeast strains that have the potential to create craft spirits. We sequenced three DNA regions in these yeast strains and compared them to sequences from previously characterized strains. We then used parsimony networks to test the following predictions: (1) the most promising candidate craft spirits strain, based on investigations by professionals in the craft distillation industry, is genetically distinct from other known yeast strains (2) the most promising candidate strain is closely related to other alcohol brewing strains and (3) yeast strains exhibit genetic differentiation by geographic location.

METHODS

DNA Sampling and Strain Isolation

We isolated yeast strains from various fruits, berries, nuts and seeds at from a North Texas ranch. Specific samples include the North Carolina red berry (*Fragaria sp.*), Texas bird of paradise (*Caesalpinia gilliesii*) , live oak acorn (*Quercus virginiana*), pecan nut (*Carya illinoinensis*), red berry/cowberry (*Vaccinium vitis-idaea*), Bradford pear (*Pyrus calleryanaI*), Eastern red cedar (*Juniperus virginiana*), rose hip (*Rosa sp.*), hackberry (*Celtis sp.*), blueberry juniper (*Juniperus ashei*), wooly bucket (*Bumelia lanuginosa*), Texas prickly pear (*Opuntia engelmannii*), Turk’s cap (*Malvaviscus arboreus*), apricot (*Prunus sp.*) and peach (*Prunus persica*). We screened one hundred

strains for the presence of *Saccharomyces cerevisiae* by amplifying a small genomic region of yeast with *Saccharomyces cerevisiae* specific primers. If a sample amplified, it was identified as *Saccharomyces cerevisiae.* Upon screening, we narrowed down the field of candidates from 100 yeast strains to 20 yeast strains. Eleven of these strains were specifically isolated from the pecan nut, and professionals in the craft distillation industry have deemed at least one of these strains potentially suitable for the production of craft spirits. We extracted the DNA from these 11 strains by first growing the yeast in media, pelleting the cells, pouring off the media, and replacing with 400 µL lysis buffer (75 mM Tris pH 8.0, 25 mM EDTA, 0.5% sodium dodecyl sulfate). The yeast was then disrupted using glass beads in an MP BioPrep. One and a half volumes 7.5 M ammonium acetate was added and the tube put on ice for 10 min to precipitate proteins, which were then pelleted by centrifugation for 10 min. We precipitated the DNA by adding 0.7 volumes isopropanol to the supernatant and centrifuging at 13 g for 15 min to pellet the DNA. The DNA pellet was washed with 400 µL 70% ethanol and then air-dried before resuspending in 100 µl 10 mM Tris–HCl pH 8.5.

Polymerase Chain Reaction (PCR)

I amplified each strain at three different regions of the yeast genome: NUP116 (gene for a nucleoporin component of the central core of the nuclear pore complex), LAS1 (gene for an essential nuclear protein required for pre-rRNA processing at both ends of the region ITS2) and IntAY (an intergenic spacer between genes APP1 and YPT53). Sequences for primers used are as follows: NUP116 forward (5'-AAG CAA CTG TCA CCA ACA CG-3'), NUP116 reverse (5'-CTT CCC CAT CGT TCT TTG AG-

3'), LAS1 forward (5'-GTT GGG CCG CCA AT AAG-3'), LAS1 reverse (5'-AAG ACG CCA AAT CGC CTA G-3'), IntAY forward (5'-TCG CAG AAT TAG GGA GAA GT-3') and IntAY reverse (5'-ACT TCC CGA CAG CAG ATT C-3'). I set up each 10µL PCR with 1.0µL deoxyribose nucleoside triphosphates (dNTPs 2mM), 1.0µL 10X PCR buffer with 25mM $MgCl_2$, 0.04µL Taq polymerase, and 6 µL H_2O, 1µL of the respective primers (5µm) and 1µL of the yeast DNA sample (~50 ng). The samples were placed in an ABI 2720 Thermal Cycler to run the PCR reaction. To amplify the NUP116 and IntAYF regions, I used the following cycling parameters: 2 min at 95°C, 30 cycles of 30 s at 94°C, 30 s at 60°C and 1 min at 72°C with a final 5 min of elongation at 72°C. To amplify the LAS1 region, we used the following cycling parameters: 2 min of denaturation at 95°C, 30 cycles of 30 s of denaturation at 94°C, 30 s of annealing at 55°C and 1 min of elongation at 72°C with a final 5 min of elongation at 72°C.

After the PCR reaction, I ran 5µL of each PCR product on a 1% agarose gel to check for successful amplification. Loading dye with GelRed ® (Biotium) (2 µL) was added to 5µL of each PCR product and then placed into the wells of a 1% agarose gel. Electrophoresis ran at 250 volts for 15 minutes. After electrophoresis, we checked the gel under UV light.

I cleaned each PCR product by running an ExoSap reaction to remove unused dNTPs and primers with Exonuclease (ExoI) and Shrimp Alkaline Phosphatase (rSAP) enzymes (0.01µL of ExoI, 0.2µL rSAP, 0.7 1X NEB 1 buffer, 1.19µL H_2O and 5.0µL of the PCR product). The ExoSap reactions were run in the ABI 2720 Thermal Cycler and the cycling parameters were set as follows: 15 min at 37°C followed by 15 min at 80°C.

DNA Sequencing

I sequenced each sample at each of the three regions, each once with the forward primer and once with the reverse primer. I carried out each reaction with 1µL of the ExoSap PCR product (DNA sample), 3.75µL H_2O, 4.0µL 5X sequencing buffer, 1.0µL of the respective primer, and 0.25µL BigDye v3.1 (containing dideoxynucleotide triphosphates (ddNTP), dNTPs, Taq polymerase and a buffer). The reactions were run in the ABI 2720 thermal cycler set at the following cycling parameters: 25 cycles of 10 s at 94°C, 5 s at 50°C and 4 min at 60°C.

After sequencing, I conducted post-sequencing cleanup for each reaction to remove unincorporated fluorescently labeled ddNTPs: we added 20µL BET (Compel beads (Bangs Laboratories), ethanol, tetraethylene glycol) to each reaction and then set them on a magnet to magnetize the beads and then removed the liquid. I rinsed the beads in the tubes twice with 100µL of 70% ethanol. I then took the tubes off the magnet and added 40µL of 0.1 mM EDTA pH 8.0 to each reaction. I placed the tubes on the magnet again to magnetize the beads. I removed the liquid from the tubes and placed into a semi-skirted plate. The plate was then put on the ABI 3130XL Genetic Analyzer and sequences were read.

Data Manipulation, Subsequent Analysis and Phylogenetic Study

I trimmed the sequences and assembled the reverse and forward sequences of each respective sample as contigs using Sequencher v5.0 (Gene Codes USA). I ran each sequence through a nucleotide BLAST analysis in GenBank

(http://blast.ncbi.nlm.nih.gov/Blast.cgi) to check that the sequences belonged to *Saccharomyces cerevisiae.*

I obtained sequences of other known strains from GenBank for comparative phylogenetic analysis of samples (Appendix 1). A total of 124 additional sequences were obtained from GenBank: 99 were wild strains from China, which are strains collected from natural habitats, 24 were lab strains, which are strains cultured and contained in the lab that are commonly used in scientific study and 1 was a common strain used to produce wine, isolated in France. I aligned the sequences of all samples generated in this study and GenBank sequences in MEGA 6.0 (Tamura et al. 2013) using Muscle (Edgar 2004). A total of 135 sequences were included in this study: 116 were studied under NUP116, 117 under LAS1 and 127 under IntAY (appendix I). I also concatenated the three separate regions in the order of IntAY, LAS1 and NUP116. I saved files as aligned FASTA files and converted them into TCS files using FaBox (Villesen 2007). I loaded the converted TCS files onto TCS Analyzer (Clement et al. 2000) to generate 95% parsimony networks for each DNA region separately and for the concatenated sequences. 95% parsimony networks group haplotypes based on their genetic similarity (Templeton et al. 1992). Haplotypes are sections of DNA that are typically inherited together and different haplotypes in this case are simply different alleles of a particular gene or intergenic region.

I color coded the strains in the networks according to the geographic location where they were isolated. These color codes and respective location are shown in the legend in Fig 1. Specifically, the most promising yeast strain for the production of craft

spirits is labeled as 4 (labelled as NUP116_4, LAS1_4, IntAY_4 and 4 respective to each DNA region studied).

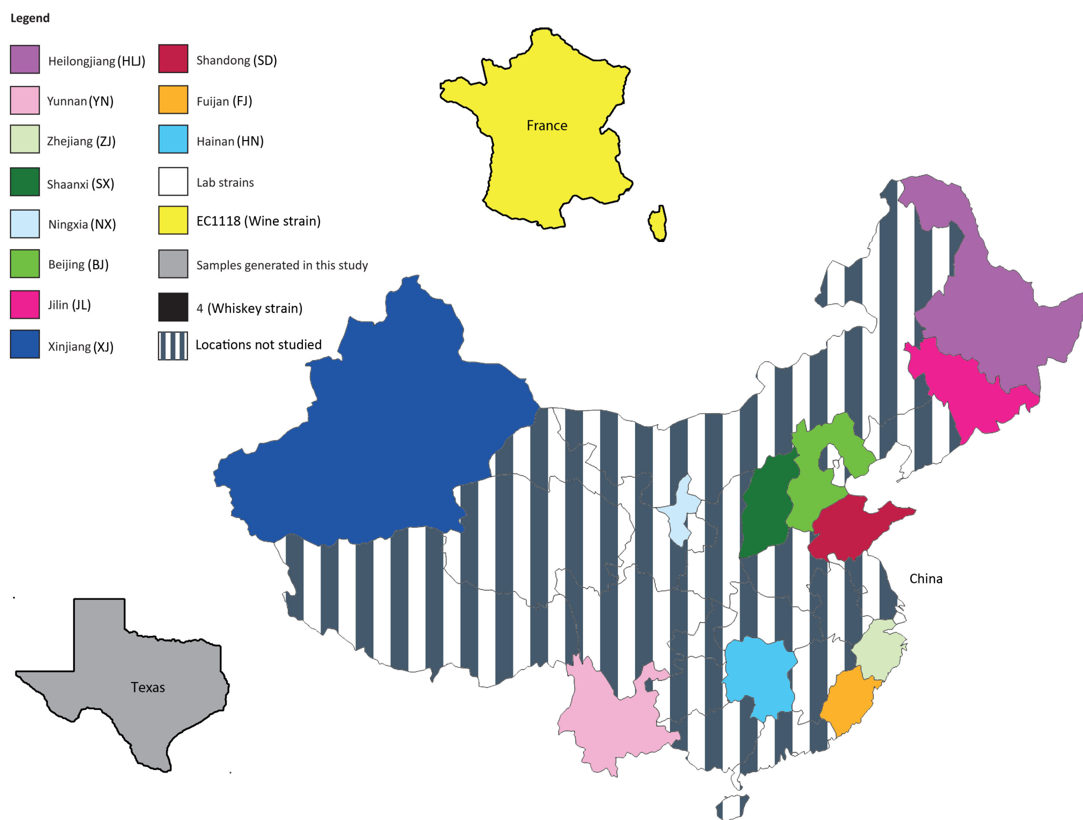

Fig 1. Map indicating geographic locations of samples studied

RESULTS

NUP116

There were a total of seventeen different haplotypes for NUP116 (Fig 2). The most common haplotype tHN was predicted to be the ancestral haplotype from which all other haplotypes branched and included most of the samples from Texas (NUP116_1, NUP116_2, NUP116_3, NUP116_6, NUP116_7, NUP116_8, NUP116_10, and NUP116_11) as well as thirty-seven other samples (Table 1). Of these other samples that share their haplotype with tHN, thirty-three were wild strains isolated in areas across China, covering a large amount of area, ranging from the Zhejiang, Yunnan, Xinjiang, Shaanxi Shandong, Jilin, Hainan, Heilongjiang, Fuijan and Beijin provinces. The other strains included in the haplotype tHN were lab strains. The second most common haplotype was JL5 with twenty samples (Table 1). Many of the samples within the haplotype JL5 were strains isolated from the Northeastern parts of China – Heilongjiang and Beijing provinces. The Texas craft spirits strain, NUP116_4, formed a haplotype with another Texas sample, NUP116_5 that is different by two-nucleotides from JL5 and is one nucleotide different from EC118, a haplotype that includes a strain that is used for making wine and seven other samples.

LAS1

There were a total of thirty-three different haplotypes for LAS1 (Fig. 3). The most common haplotype for LAS1 includes both EC1118 and the Texas samples LAS1_4 and LAS1_5 as well as ten other samples (Table 2). The ten other samples in the most common haplotype were a lab strain and wild strains isolated from areas across China,

ranging from the Xinjiang, Shandong, Ningxia and Beijing provinces. The other Texas samples were all part of haplotypes LAS1_1 and LAS1_10 (Table 2).

IntAY

There are a total of 53 different haplotypes for IntAY (Fig 4). The most common haplotype for IntAY was IntAY_1 and included most of the samples collected in Texas (IntAY_1, IntAY_2, IntAY_3, IntAY_6, IntAY_8, IntAY_9, IntAY_10, and IntAY_11) and four other samples (Table 3). The other four samples were wild strains isolated from areas across China, ranging from the Zhejiang, Xinjiang and Beijing provinces. Another sample collected in Texas, IntAY_7 was one nucleotide different from this predicted ancestral haplotype. The Texas craft spirits strain, IntAY_4 and another Texas sample (IntAYF_5) formed another haplotype that was also one nucleotide different from the IntAY_1 haplotype. The Texas craft spirits strain haplotype was seven nucleotides away from EC1118.

Concatenated NUP116, LAS1 and IntAY

There were a total of eighty different haplotypes for the concatenated series. The most common haplotype (1) contained six samples collected in Texas (1, 2, 3, 6, 8, 11). Several other Texas samples (7, 9, 10) were closely related to this haplotype. The Texas craft spirits strain (4) formed a haplotype with Texas sample 5 and was five nucleotides different from the EC1118 strain, which formed a unique haplotype (Fig 4). The network reveals that the Texas craft spirits strain haplotype is very distinct from the other Texas samples (1, 2, 3, 6, 7, 8, 9, 10, 11). A majority of the other samples formed independent haplotypes (Table 4). Of the wild strains isolated in China, the Texas craft spirits strain

was closest to the haplotype XJ5, differing by eight nucleotides. XJ5 was isolated from the Xinjiang province.

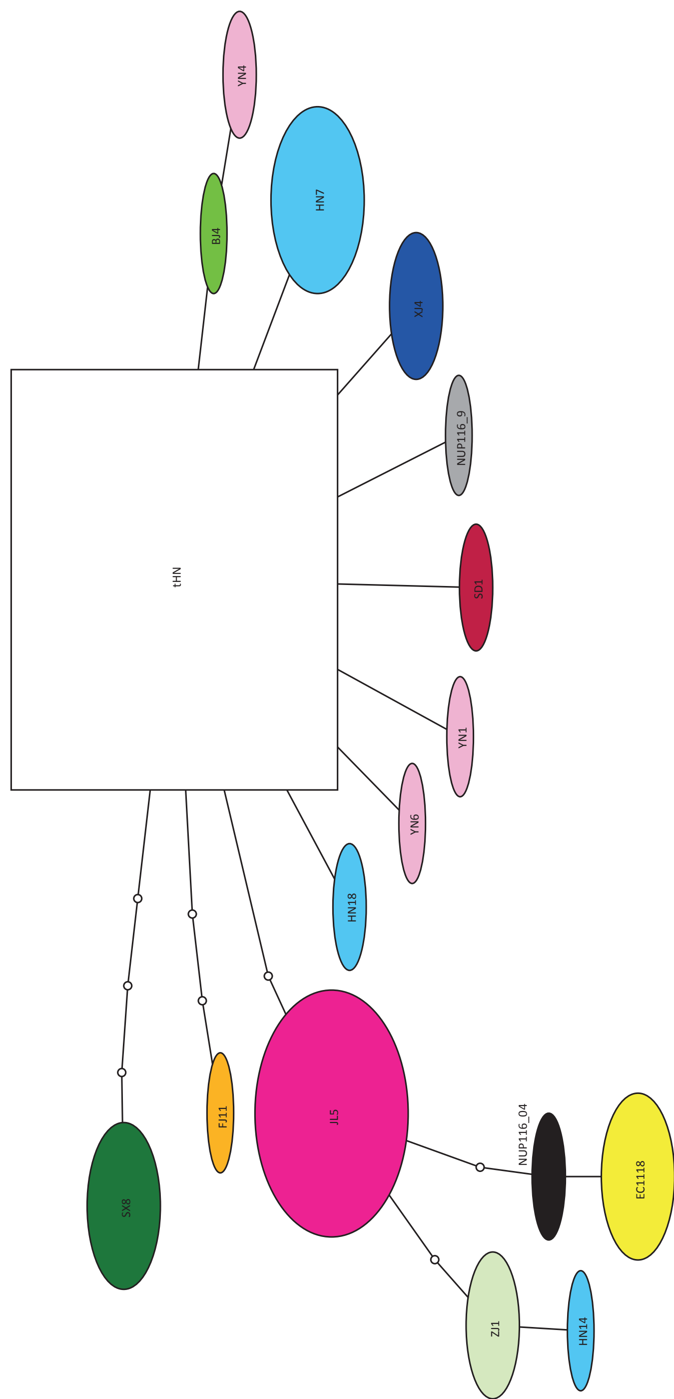

Fig 2. 95% parsimony network of NUP116 gene region for *Saccharomyces cerevisiae*. Ovals are haplotypes, square is predicted ancestral haplotype, each dotted line is one nucleotide difference, regardless of length and small circles are inferred intermediate haplotypes that were not sampled. Colors correspond to colored regions in Fig 1.

Table 1. Shared haplotypes of NUP116. Total number of haplotypes: 17.

Haplotype	Strains sharing haplotype
EC1118	YJM993, XJ5, XJ3, NX2, NX1, JL3, JL1
tHN	cosmid9532, S288c, XN, S288c_mRNA, NUP116_1, NUP116_2, NUP116_3 NUP116_6, NUP116_7, NUP116_8, NUP116_10, NUP116_11, ZJ2, YN5, YN3, YN2, XJ7, XJ6, XJ2, XJ1, SX11, SX10, SX9, SD3, SD2, JL4, JL2, HN19, HN16, HN11, HN10, HN9, HN8, HN4, HLJ3, HLJ2, FJ12, FJ6, FJ3, FJ2, FJ1, BJ28, BJ13, BJ9, BJ5
NUP116_4	NUP116_5
NUP116_9	
ZJ1	HN15, HN13, HN12, FJ4
YN6	
YN4	BJ3
YN1	
XJ4	SD4, NX3, BJ17, BJ7
SX8	SX7, SX6, SX5, SX4, SX3, SX2, SX1
SD1	BJ1
JL5	HLJ1, BJ27, BJ26, BJ25, BJ24, BJ23, BJ22, BJ21, BJ20, BJ19, BJ18, BJ16, BJ15, BJ14, BJ12, BJ10, BJ8, BJ6, BJ2
HN18	HN17
HN14	
HN7	HN6, HN5, HN3, HN2, HN1, FJ10, FJ9, FJ8, FJ7, FJ5
FJ11	
BJ4	

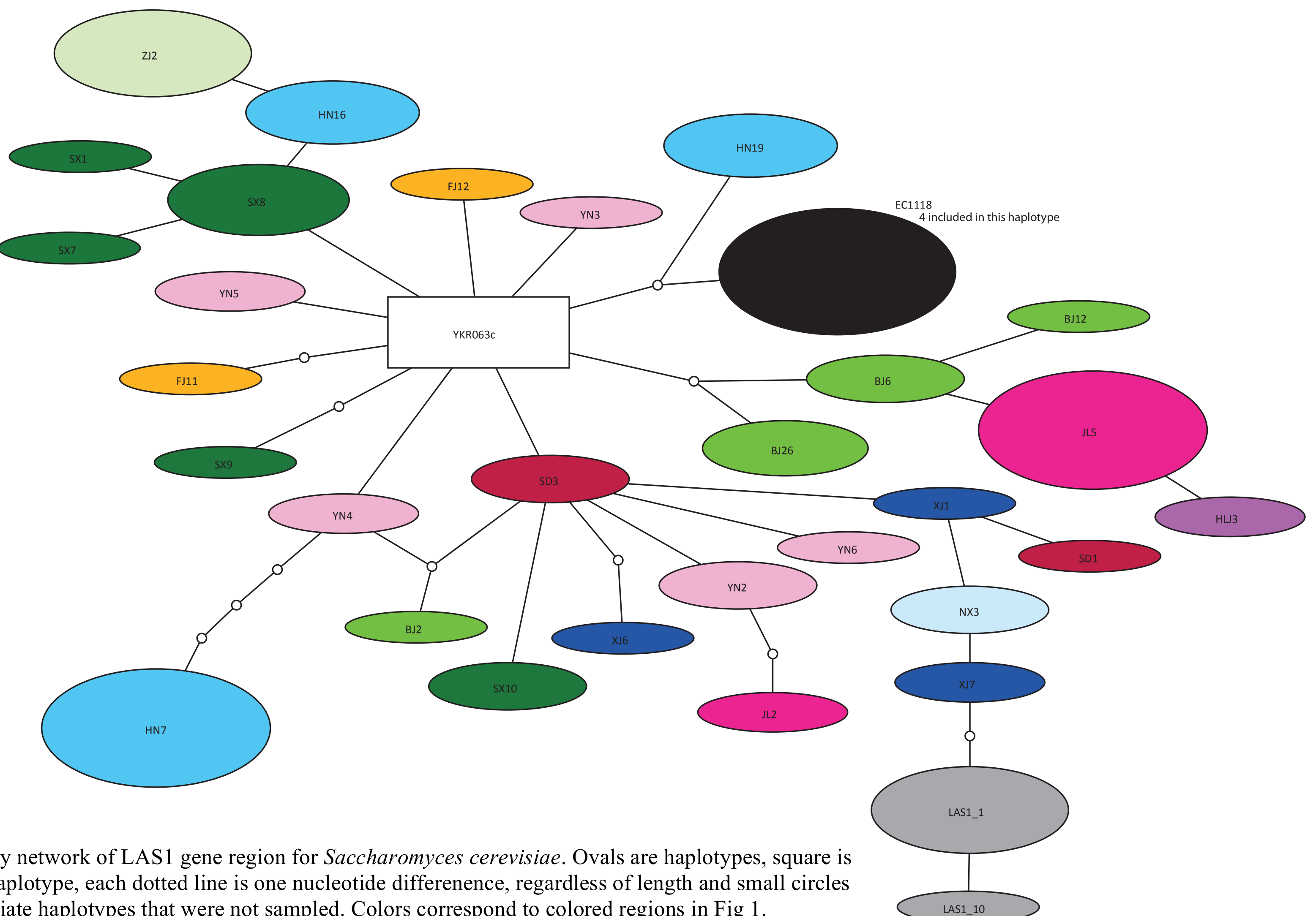

Fig 3. 95% parsimony network of LAS1 gene region for *Saccharomyces cerevisiae*. Ovals are haplotypes, square is predicted ancestral haplotype, each dotted line is one nucleotide difference, regardless of length and small circles are inferred intermediate haplotypes that were not sampled. Colors correspond to colored regions in Fig 1.

Table 2. Shared haplotypes of LAS1. Total number of haplotypes: 33

Haplotype	Strains sharing haplotype
EC1118	YJM993, LAS1_4, LAS1_5, XJ5, SD4, NX2, NX1, BJ28, BJ13, BJ10, BJ9, BJ5
YKR063c	YC, grfat8, S288c_mRNA, S288c, XJ4
LAS1_1	LAS1_2, LAS1_3, LAS1_6, LAS1_7, LAS1_8, LAS1_9, LAS1_11
LAS1_10	
ZJ2	ZJ1, HN18, HN17, HN14, HN13, HN12, FJ6
YN6	
YN5	YN1
YN4	SX11
YN3	
YN2	XJ2, BJ7
XJ7	XJ3
XJ6	
XJ1	
SX10	SD2, BJ4
SX9	
SX8	SX6, SX5, SX4, SX3, SX2
SX7	
SX1	
SD3	JL3, JL1
SD1	
NX3	FJ1, BJ8
JL5	JL4, HJL1, BJ27, BJ24, BJ22, BJ21. BJ19, BJ18, BJ17, BJ15, BJ14
JL2	BJ1
HN19	HN11, HN10, HN9, HN8
HN16	HN15, FJ4, FJ3, FJ2
HN7	HN6, HN5, HN4, HN3, HN2, HN1, FJ10, FJ9, FJ8, FJ7, FJ5
HLJ3	HLJ2
FJ12	
FJ11	
BJ6	BJ16, BJ3
BJ26	BJ25, BJ23, BJ20
BJ12	
BJ2	

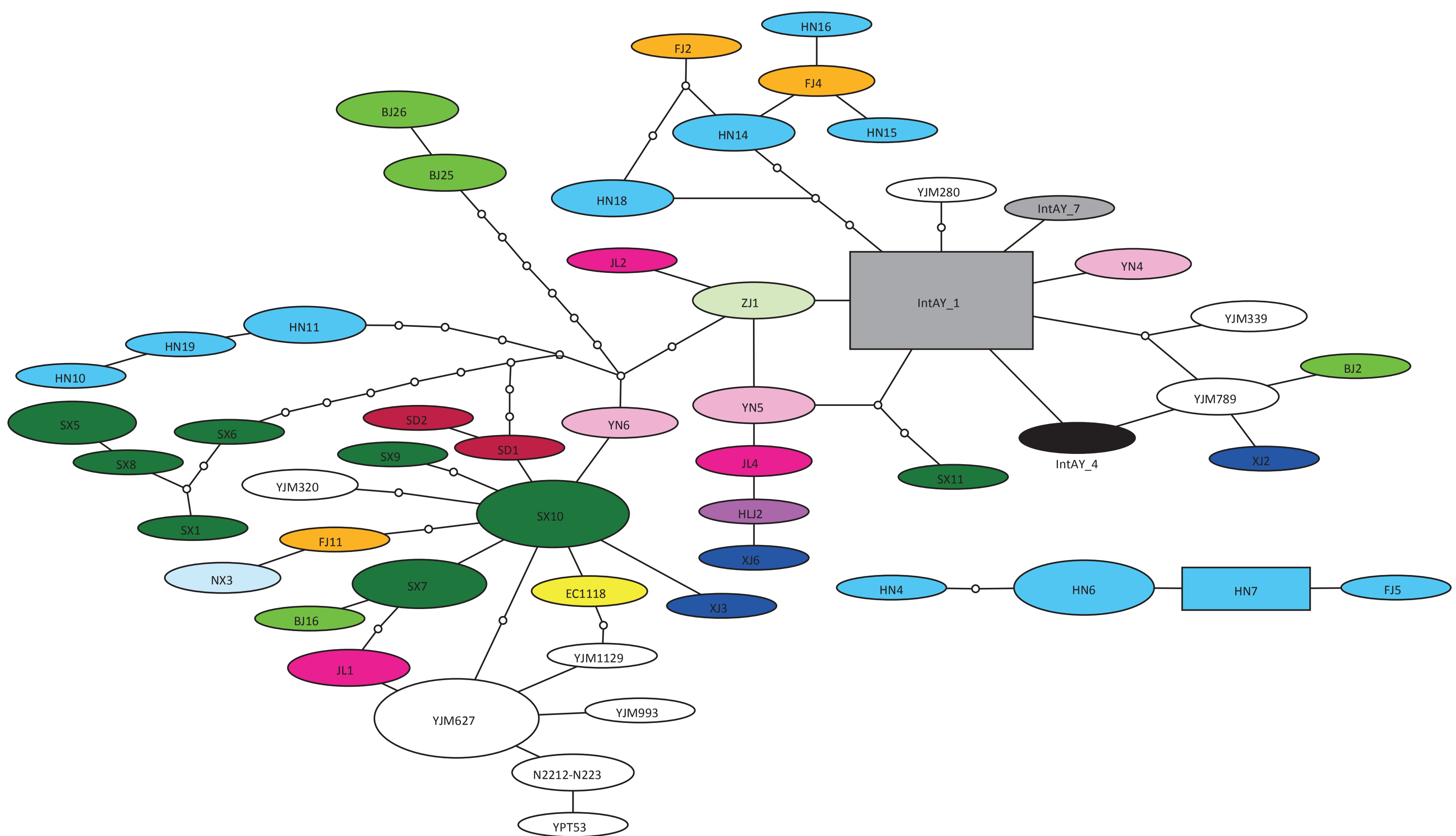

Fig 4. 95% parsimony network of IntAY intergenic spacer for *Saccharomyces cerevisiae*. Ovals are haplotypes, square is predicted ancestral haplotype, each dotted line is one nucleotide differenence, regardless of length and small circles are inferred intermediate haplotypes that were not sampled. Colors correspond to colored regions in Fig 1.

Table 3. Shared haplotypes of IntAY. Total number of haplotypes: 53

Haplotype	**Strains sharing haplotype**
N2212-N223	S96, S288c
YPT53	
EC1118	POL1
YJM280	
YJM320	YJM326
YJM339	YJM421
YJM627	XJ5, XJ4, SD4, NX2, NX1, JL3, BJ28, BJ13, BJ10
YJM789	YN3, YN2
YJM1129	
YJM993	
IntAY_1	IntAY_2, IntAY_3, IntAY_6, IntAY_8, IntAY_9, IntAY_11, IntAY_10, ZJ2, XJ7, BJ27, BJ24, BJ22
IntAY_4	IntAY_5
IntAY_7	
ZJ1	YN1, BJ1
YN6	FJ12
YN5	SD3, BJ7
YN4	XJ1
XJ6	
XJ3	
XJ2	
SX11	
SX10	JL5, HLJ1, BJ21, BJ15, BJ14, BJ8, BJ4
SX9	
SX8	
SX7	BJ19, BJ18, BJ17, BJ12
SX6	
SX5	SX4, SX3, SX2
SX1	
SD2	
SD1	
NX3	FJ1
JL4	HLJ3
JL2	
JL1	BJ9, BJ5
HN19	

Table 3 continued on page 17

Table 3. Shared haplotypes of IntAY (continued)

Haplotype	**Strains sharing haplotype**
HN18	HN17, HN12
HN16	
HN15	
HN14	HN13, FJ6
HN11	HN9, HN8
HN10	
HN7	HN5, HN3, HN2
HN6	HN1, FJ10, FJ9, FJ8, FJ7
HN4	
HLJ2	
FJ11	
FJ5	
FJ4	FJ3
FJ2	
BJ26	BJ23, BJ20
BJ25	BJ6, BJ3
BJ16	
BJ2	

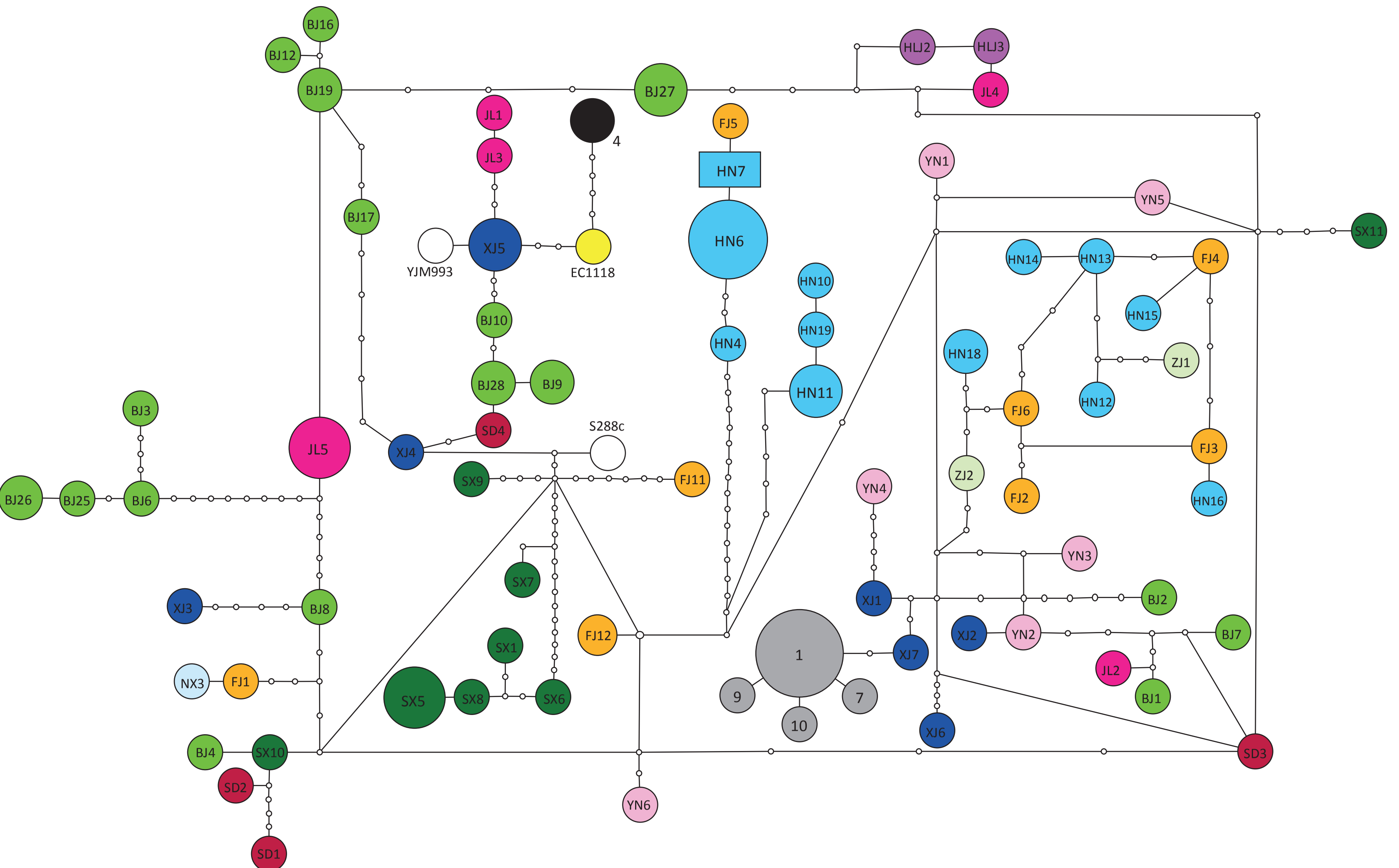

Fig 5. 95% parsimony network of concatenated sequence in the order of NUP116, LAS1 and IntAY for *Saccharomyces cerevisiae*. Ovals are haplotypes, square is predicted ancestral haplotype, each dotted line is one nucleotide differenence, regardless of length and small circles are inferred intermediate haplotypes that were not sampled. Colors correspond to colored regions in Fig 1.

Table 4. Haplotype groups of concatenated sequences. Total number of haplotypes: 80

Haplotype	**Strains in haplotype group**
1	2, 3, 6, 8, 11
4	5
7	
9	
10	
ZJ2	
ZJ1	
YN6	
YN5	
YN4	
YN3	
YN2	
YN1	
XJ7	
XJ6	
XJ5	NX2, NX1
XJ4	
XJ3	
XJ2	
XJ1	
SX11	
SX10	
SX9	
SX8	
SX7	
SX6	
SX5	SX4, SX3, SX2
SX1	
SD4	
SD3	
SD2	
SD1	
NX3	
JL5	HLJ1, BJ21, BJ15, BJ14
JL4	
JL3	

Table 4 continued on page 20.

Table 4. Haplotype groups of concatenated sequences. (continued)

Haplotype	**Strains sharing haplotype**
JL2	
JL1	
HN19	
HN18	HN17
HN16	
HN15	
HN14	
HN13	
HN12	
HN11	HN9, HN8
HN10	
HN7	HN5, HN3, HN2
HN6	HN1, FJ10, FJ9, FJ8, FJ7
HN4	
HLJ3	
HLJ2	
FJ12	
FJ11	
FJ6	
FJ5	
FJ4	
FJ3	
FJ2	
FJ1	
BJ28	BJ13
BJ27	BJ24, BJ22
BJ26	BJ23, BJ20
BJ25	
BJ19	BJ18
BJ17	
BJ16	
BJ12	
BJ10	
BJ9	BJ5
BJ8	
BJ7	

Table 4 continued on page 21.

Table 4. Haplotype groups of concatenated sequences. (continued)

Haplotype	Strains sharing haplotype
BJ6	
BJ4	
BJ3	
BJ2	
BJ1	
YJM993	
S288c	
EC1118	

DISCUSSION

In general, Strain 4, the most promising candidate craft spirits strain, based on investigations by professionals in the craft distillation industry, was genetically unique relative to the other strains in this study. This was apparent with respect to all the three gene regions that were analyzed – NUP116, LAS1, IntAY – and the concatenated series.

The gene NUP116 is an essential gene in that it codes for a nucleoporin component of the central core of the nuclear pore complex. As a result, the DNA sequence is highly conserved. This is apparent as the NUP116 network exhibited the least genetic diversity as shown by the lesser amount of branching in the network for NUP116. However, despite being conserved, strains 4 (NUP116_4) and 5 (NUP116_5) form their own unique haplotype. Moreover, we find that the EC1118 strain, used to ferment wine, is closely related to strains 4 and 5, as the EC1118 haplotype is only one nucleotide difference from that of strains 4 and 5.

The gene LAS1 is another gene region that is conserved due to its function in coding for another nuclear protein involved in pre-rRNA processing. However, there is more genetic diversity amongst the strains for the LAS1 gene region compared to

NUP116, as shown by the greater amount of branching in the network for LAS1. This suggests that the DNA sequence for LAS1 is not as highly conserved. For this gene, strain 4 (LAS1_4) shares a haplotype with the EC1118 strain, further suggesting a close genetic relatedness between the two strains.

As an intergenic spacer, IntAY is not highly conserved over evolutionary time because it is does not code for a protein that is important for the functioning of yeast. Hence there will be greater genetic variation in this region as it will accumulate mutations over time. This is why there are many more haplotypes for this region. In the case of IntAY, we see that strain 4 (IntAY_4) still branches out in its own haplotype. However, different from NUP116 and LAS1, IntAY_4 does not share a haplotype with or branch off from EC1118. This suggests a case of genetic difference between the strains.

When the sequences are concatenated, strain 4 branches out in its own haplotype. Again, strain 4 shares a haplotype with strain 5. Moreover, as seen in LAS1 and NUP116, there is a close relation between strain 4 and the EC1118 wine strain, as the haplotype of strain 4 branches off from that of EC118. In this case, there is only a 5 nucelotide difference between the respective haploytypes.

The close relationship between strain 4 and the EC1118 wine strain is not surprising, as both are involved in producing alcohol. This suggests that there may be certain alleles that increase yeast's ability to successfully produce fine-tasting alcohol. And it is consistent with the fact that yeasts are typically specialized for alcoholic beverages (Ingledew et al. 2009). In addition to EC1118, however, strain 4 typically showed some connection to other commercial lab strains. For instance, for LAS1, the haplotype of strain 4 differed only by two nucleotides from that of YKR063c, a haplotype

shared with five other lab strains. For IntAY, the haplotype of strain 4 (IntAY_4) was only one nucleotide different from another lab strain, YJM789. This is consistent with the idea that yeast strains used commericially, such as baker's yeast and other alcohol fermentors, are isolated and selected for their desirable traits that make them better for food and alcohol production.

Furthermore, there is no consistent pattern in the relationships between strains that share a geographic location. For instance, although some of the strains are located in the same areas of the network of the concatenated series suggesting that they are genetically similar, there were a few strains isolated from the same geographic location that were quite different. For example, the strains from Beijing, BJ16, BJ12, BJ19 and BJ17, which had similar haplotypes, differing by one to three nucleotides, were more genetically different from other strains from Beijing, such as BJ7 and BJ1 that were more closely related to strains from other provinces. This is also apparent in the case of the Texas strains that we isolated for this study. Particularly, the strain of interest (strain 4) was genetically different from the other Texas strains (1, 2, 3, 6, 7, 8, 9, 10, 11). As a microorganism, it is not surprising that yeast strains are not genetically similar by geographic location. This is because microbes are ubiquitous in nature and have the ability to move great distances (Duncan and Leu 2009). Since yeast is a microbe useful to humans, humans have also facilitated the movement of yeast. Furthermore, yeast reproduce rapidly, giving more opportunity for strains to accumulate mutations.

In conclusion, yeast did not show strong genetic differentiation by geographic location in this study. Strains 4 and 5, which we isolated in Texas, appear to be the same strain as they shared haplotypes in all four networks. However, overall, with respect to all

strains isolated in Texas and those included in this study, the Texas craft spirits strain is unique. Further study of strain 4 in relation to other commercial strains used to produce alcohol would be insightful in further characterizing strain 4 as an alcohol-producing strain.

APPENDIX A

Appendix A. List of sequences obtained from GenBank with respective accession codes

Strain Name	Accession Code
BJ1	JQ284751.1
BJ2	JQ284752.1
BJ3	JQ284753.1
BJ4	JQ284754.1
BJ5	JQ284755.1
BJ6	JQ284756.1
BJ7	JQ284757.1
BJ8	JQ284758.1
BJ9	JQ284759.1
BJ10	JQ284760.1
BJ12	JQ284761.1
BJ13	JQ284762.1
BJ14	JQ284763.1
BJ15	JQ284764.1
BJ16	JQ284765.1
BJ17	JQ284766.1
BJ18	JQ284767.1
BJ19	JQ284768.1
BJ20	JQ284769.1
BJ21	JQ284770.1
BJ22	JQ284771.1
BJ23	JQ284772.1
BJ24	JQ284773.1
BJ25	JQ284774.1
BJ26	JQ284775.1
BJ27	JQ284776.1
BJ28	JQ284777.1
FJ1	JQ284778.1
FJ2	JQ284779.1

Appendix A continues on page 25.

Appendix A (continued)

Strain Name	Accession Code
FJ3	JQ284780.1
FJ4	JQ284781.1
FJ5	JQ284782.1
FJ6	JQ284783.1
FJ7	JQ284784.1
FJ8	JQ284785.1
FJ9	JQ284786.1
FJ10	JQ284787.1
FJ11	JQ284788.1
FJ12	JQ284789.1
HLJ1	JQ284790.1
HLJ2	JQ284791.1
HLJ3	JQ284792.1
HN1	JQ284793.1
HN2	JQ284794.1
HN3	JQ284795.1
HN4	JQ284796.1
HN5	JQ284797.1
HN6	JQ284798.1
HN7	JQ284799.1
HN8	JQ284800.1
HN9	JQ284801.1
HN10	JQ284802.1
HN11	JQ284803.1
HN12	JQ284804.1
HN13	JQ284805.1
HN14	JQ284806.1
HN15	JQ284807.1
HN16	JQ284808.1
HN17	JQ284809.1
HN18	JQ284810.1
HN19	JQ284811.1
JL1	JQ284812.1
JL2	JQ284813.1
JL3	JQ284814.1
JL4	JQ284815.1
JL5	JQ284816.1
NX1	JQ284817.1
NX2	JQ284818.1
NX3	JQ284819.1
SD1	JQ284820.1

Appendix A continues on page 26.

Appendix A (continued)

Strain Name	Accession Code
SD2	JQ284821.1
SD3	JQ284822.1
SD4	JQ284823.1
SX1	JQ284824.1
SX2	JQ284825.1
SX3	JQ284826.1
SX4	JQ284827.1
SX5	JQ284828.1
SX6	JQ284829.1
SX7	JQ284830.1
SX8	JQ284831.1
SX9	JQ284832.1
SX10	JQ284833.1
SX11	JQ284834.1
XJ1	JQ284835.1
XJ2	JQ284836.1
XJ3	JQ284837.1
XJ4	JQ284838.1
XJ5	JQ284839.1
XJ6	JQ284840.1
XJ7	JQ284841.1
YN1	JQ284842.1
YN2	JQ284843.1
YN3	JQ284844.1
YN4	JQ284845.1
YN5	JQ284846.1
YN6	JQ284847.1
ZJ1	JQ284848.1
ZJ2	JQ284849.1
YNL093w	Z71369.1
YNL094w	Z71370.1
N2212-N223	X85811.1
YPT53	X76175.1
EC1118	FN393086.1
POL1	DQ115393.1
S96	AF458969.1
YJM280	AF458970.1
YJM320	AF458971.1
YJM326	AF458972.1
YJM339	AF458973.1
YJM421	AF458974.1
YJM627	AF458981.1

Appendix A continues on page 27.

Appendix A (continued)

Strain name	**Accession code**
YJM789	AF458975.1
YJM1129	AF458978.1
YJM993	CP005526.1
S288c	BK006947.3
YKR063c	CAA82142.1
YC	AAU09759.1
grfat8	AAC49111.1
S288c mRNA	NP_012989.3
tHN	CAA78754.1
cosmid9532	CAA88413.1
XN	CAA48228.1

REFERENCES

M.L. Álvarez-Ainza, K.A. Zamora-Quiñonez, G.M. Moreno-Ibarra and E.Acedo-Félix. 2015. Genomic diversity of *Saccharomyces cerevisiae* yeasts associated with alcoholic fermentation of Bacanora produced by artisanal methods. Applied Biochemistry and Biotechnology 175: 2668-2676

James A. Barnett.1998. A history of research on yeasts 1: work by chemists and biologists 1789-1850. Yeast 14: 1439-1451

James A. Barnett. 2000. A history of research on yeasts 2: Louis Pasteur and his contemporaries, 1850-1880. Yeast 16: 755-771

M. Clement, D. Posada and K. Crandall. 2000. TCS: a computer program to estimate gene genealogies. Molecular Ecology 9(10): 1657-1660

L. Dijkshoorn, B.M. Ursing and J.B. Ursing. 2000. Strain, clone and species: comments on three basic concepts of bacteriology. Microbiology 49: 397-401

Robert C. Edgar. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32: 1792-1797

Justin C. Fay and Joseph A. Benavides. 2005. Evidence for domesticated and wild populations of *Saccharomyces cerevisiae*. PLOS Genetics 1: 66-71

Greig Duncan and Jun-Yi Leu. 2009. Natural History of Budding Yeast. Current Biology 19: 886-890

W.M Ingledew, D.R. Kelsall, G.D. Austin and C. Kluhspies. 2009. The Alcohol Textbook. Nottingham University Press, Notingham, UK.

Gianni Liti, David M. Carter, Alan M. Moses, Jonas Warringer, Leopold Parts, Stephen A. James, Robert P. Davey, Ian N. Roberts, Austin Burt, Vassiliki Koufopanou, Isheng J. Tsai, Casey M. Bergman, Douda Bensasson, Michael J. T. O'Kelly, Alexander van Oudenaarden, David B. H. Barton, Elizabeth Bailes, Alex N. Nguyen, Matthew Jones, Michael A. Quail, Ian Goodhead, Sarah Sims, Frances Smith, Anders Blomberg, Richard Durbin and Edward J. Louis. 2009. Population genomics of domestic and wild yeasts. Nature 485: 337-341

Ian N. Roberts and Stephen G. Oliver. 2010. The yin and yang of yeast: biodiversity research and systems biology as complementary forces driving innovation in biotechnology. Biotechnology letters 33: 477-487

Koichiro Tamura, Glen Stecher, Daniel Peterson, Alan Filipski, and Sudhir Kumar (2013) MEGA6: Molecular Evolutionary Genetics Analysisversion 6.0. Molecular Biology and Evolution:30 2725-2729.

Alan R. Templeton, Keith A. Crandall and Charles F. Sing. 1992. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence date. III. Cladogram estimation. Genetics Society of America 132: 619-633

Sequencher® version 5.3 sequence analysis software, Gene Codes

Villesen, P (2007), *FaBox: an online fasta sequence toolbox*, http://www.birc.au.dk/software/fabox

Qi-Ming Wang, Wan-Qiu Liu, Gianni Liti, Shi-An Wang and Fend-Yen Bai. 2012. Surprisingly diverged populations of *Saccharomyces cerevisiae* in natural environments remote from human activity. Molecular Ecology 21: 5404-54417