GR

ORIGINAL SUBMISSION

000003

RECEIVED
- 7 2005
OFFICE OF
FOOD ADDITIVE SAFETY

FIRST VENTURE
TECHNOLOGIES

July 5, 2005

VIA COURIER

Office of Premarket Approval (HFS-200)
Center for Food Safety and Applied Nutrition
Food and Drug Administration
200 C St. SW.
Washington, DC
20204

Dear Sir / Madam:

Re: GRAS Notice for Genetically Enhanced *Saccharomyces cerevisiae* Yeast

Enclosed please find a GRAS Notice for the use of a genetically modified yeast, which is useful for production of alcoholic beverages.

After careful review of the literature and extensive scientific analysis of this novel yeast strain, we have determined that its use in fermenting foods and beverages is generally recognized as safe (GRAS), and it should therefore be exempt from the statutory premarket approval requirements.

We have enclosed three copies of a notice of a GRAS exemption claim for your review and look forward to your response in due course.

Yours truly,

Howard J. Louie
President & Chief Executive Officer

encl: *GRAS Notice x 3 copies*

1500 – 885 West Georgia Street, Vancouver, B.C. Canada V6C 3E8
604-648-2200 p | 604-648-2201 f
www.firstventuretech.com

GRAS Notification

For the Use of a Modified Yeast for Reduction of Ethyl Carbamate in Fermented Beverages

Prepared by
First Venture Technologies Corporation
Vancouver, British Columbia, CANADA

First Venture Technologies Corporation

Howard J. Louie
President & Director
Dated: July 5, 2005

Table of Contents

1 GRAS Exemption Claim 2

1.1 Name and address of the notifier 2

1.2 Common or usual name of the substance that is the subject of the GRAS exemption claim 2

1.3 Applicable conditions of use of the notified substance 3

1.3.1 Purposes for which the substance is used 3

1.3.2 Foods in which the substance is to be used 3

1.3.3 Levels of use of the ECMo01 active dry yeast in grape must 4

1.4 The basis for the GRAS determination 4

1.5 data and information that are the basis for the notifiers' GRAS determination are available for the Food and Drug Administration review and copying 4

2 Detailed information about the identity of the notified substance 6

2.1 Host microorganism 6

2.1.1 History of use 6

2.1.2 Characteristics 6

2.2 Donor organisms 7

2.2.1 Taxonomy 7

2.2.2 Genetic material from donor organisms 7

2.3 The modified microorganism 8

2.3.1 Final construct used in the integration strategy 8

2.3.1.1 Construction strategy 8

2.3.1.2 Detailed description of the final construct 9

2.3.2 The transformation event 9

2.3.2.1 Genetic material used for the transformation method .. 9

2.3.2.2 Screening method for transformants 9

2.3.3 Genetic characterization of the modified microorganism 10

2.3.3.1 The loss of the antibiotic resistance genes and of other plasmid sequences 10

2.3.3.2 Genetic analysis by Southern blot 10

2.3.3.3 Genetic stability of the ECMo01 yeast strain 11

2.3.4 DNA sequencing of the integrated locus 11

2.3.4.1 Differences found in the *PGK1* promoter, *PGK1* terminator and *URA3* 12

2.3.4.2 Differences in the *DUR1,2* gene 12

2.3.5 Absence of difference between genetic profiles of the transformed and the host strain 13

2.3.6 Method of manufacture of the modified microorganism 13

3 Information on any self-limiting levels of use 14

4 Determination that the use of the wine active dry yeast obtained from the ECMo01 strain in the fermentation of Wine And Other Alcoholic Beverages is exempt from the pre-market approval requirements of the act because such use is GRAS 15

4.1 Safety assessment of the host strain 15

4.2 Safety assessment of the genetic material used to construct the modified organism 16

4.2.1 Foreign genetic material source and product 16

4.2.2 Native genetic material source and product: *DUR1,2* gene, *PGK1* promoter, *PGK1* terminator and *URA3* sequences. 16

4.2.2.1 The *DUR1,2* gene 16

4.2.2.2 *PGK1* promoter, terminator and *URA3* sequences 16

4.2.3 Construction of the modified organism 16

4.2.3.1 Vector used as source of the integration cassette 16

4.2.3.2 Transformation of the host strain with the integration cassette 17

4.3 Safety assessment of the modified organism 17

4.3.1 Characterization of the transformation event 17

4.3.2 Consequence of the genetic modification on gene expression 18

4.3.2.1 Global gene expression analysis 18

4.3.3 Presence of unintended gene products as a result of the transformation event 18

4.3.3.1 Urea Amidolyase 20

4.3.4 Consequence of the genetic modification on the physiology of the ECMo01 strain 20

4.3.4.1 Overview of the function of the urea amidolyase gene 20

000007

4.3.4.2 Manufacture of active dry yeast during fed-batch aerobic cultures 20

4.3.4.3 Growth, ethanol production, and CO_2 production of ECMo01 and Davis 522 strains during laboratory-scale vinifications 20

4.4 Safety assessment of the product derived from the modified organism 21

4.4.1 Changes in the winemaking procedures as a consequence of the ECMo01 fermentation 21

4.4.2 Changes in wine composition as a consequence of the ECMo01 fermentation 21

4.4.2.1 Yeast cells and release of yeast products during winemaking 21

4.4.2.2 Formation of unwanted substances in wine 24

5 References 83

FIGURES

Figure 1: Natural formation of ethyl carbamate by *S. cerevisiae*. 27

Figure 2: Urea degradation by ECMo01..... 28

Figure 3: Flow diagram illustrating the cloning strategy used to construct plasmid pHV10dur$_{11}$..... 29

Figure 4: Plasmid pHVX2, previously described in Volschenk *et al.*, 1997a. 30

Figure 5: Cloning of the *DUR1,2* gene and construction of plasmid pDUR1,2. 31

Figure 6: Plasmid pHVJH1, previously described by Volschenk *et al.*, 2004. 32

Figure 7: Construction of plasmid pHV10dur$_{11}$. 33

Figure 8: Schematic representation of the introduced genetic material. 34

Figure 9: Loss of Phleomycin resistance in the ECMo01 yeast strain. 35

Figure 10: Plasmid pUT332Δ*URA3* map and schematic representation of the probes used in Southern experiments. 36

Figure 11: Southern blot analysis of the Tn5*Ble* gene that encodes for Phleomycin resistance in the ECMo01 strain. 37

Figure 12: Southern blot analysis of the *bla* gene that encodes for Ampicillin resistance in the ECMo01 strain. 38

Figure 13: Schematic representation of probe hybridization to the *DUR1,2* integration cassette. 39

Figure 14: Southern blot analyses of the *URA3* loci in the host and ECMo01 strains. 40

Figure 15: Southern blot analyses of the *DUR1,2* gene in the host and ECMo01 strains. 41

Figure 16: Genetic patterns of the Davis 522 (lane 2) and ECMo01 (lane 3) yeast strains..... 42

Figure 17: Flow diagram describing the manufacture of active dry yeast.... 43

Figure 18: Putative open reading frames (ORFs) analysis. A) Schematic representation of putative ORFs and TATA boxes found within the integration cassette. 44

Figure 19: A) Comparison of growth of strains ECMo01 and Davis 522 in Chardonnay grape juice; B) Formation of ethanol by Davis 522 and ECMo01 when fermenting Chardonnay grape juice. 45

Figure 20: Evolution of carbon dioxide in fermentations of Davis 522 and ECMo01 as measured as weight loss..... 46

Figure 21: White winemaking practices and the use of a standard or ECMo01 yeast strain. 47

Figure 22: Red winemaking practices and the use of a standard or ECMo01 yeast strain. 48

000010

TABLES

Table 1: Primers and linkers used in the construction of the integration cassette 49

Table 2: Plasmids used for the construction and integration of the integration cassette. 50

Table 3: Detailed description of the nucleotides composing the integration cassette. 51

Table 4: Discrepancies found between the genomic sequence of the integrated locus and previously published S288C sequences..... 52

Table 5: Comparison of ethyl carbamate formation in wines formed using ECMo01 strain and Davis 522 strain. 54

000011

APPENDICES

Appendix 1: Integration cassette in Fasta Format 55

Appendix 2: Optimal alignment of DNA sequences: ECMo01 Integration Cassette and Published *URA3*, *DUR1,2*, *PGK1* promoter and terminator sequences 60

Appendix 3: Nucleotide sequence of the *DUR1,2* gene and protein sequence of the corresponding protein in Fasta format..... 71

Appendix 4: Analysis of global gene expression patterns by DNA microarrays 75

Appendix 5: Putative open reading frames as a result of the integration event 78

Appendix 6: Determination of ethyl carbamate in Wine by Solid-Phase Microextraction and Gas Chromatography/Mass Spectrometry ... 81

000012

GRAS NOTICE

1 GRAS EXEMPTION CLAIM

1.1 NAME AND ADDRESS OF THE NOTIFIER

First Venture Technologies Corporation
Suite 1500 - 885 West Georgia Street
Vancouver, British Columbia, V6C 3E8
Canada

Telephone: (604) 648-2200
Facsimile: (604) 648-2201

Scientific Manager: Dr. Hendrik J.J. van Vuuren
hjjvv@interchange.ubc.ca
First Venture Technologies Corporation
Suite 1500 - 885 West Georgia Street
Vancouver, British Columbia, V6C 3E8
Canada

Telephone: (604) 648-2200
Facsimile: (604) 648-2201

The research and development of the urea degrading yeast strain was initiated in Canada by researchers at Brock University, Ontario (PCT publication no. WO 03/040379) and led by Professor Hennie van Vuuren from Brock University (Canada), who has subsequently moved to The University of British Columbia, British Columbia (Canada). The research was completed at The University of British Columbia.

1.2 COMMON OR USUAL NAME OF THE SUBSTANCE THAT IS THE SUBJECT OF THE GRAS EXEMPTION CLAIM

The subject of the notice is a wine yeast obtained from a commercial *Saccharomyces cerevisiae* wine strain. The strain carries a recombinant gene that encodes a urea amidolyase enzyme from *S. cerevisiae*. This novel yeast strain is called ECMo01.

1.3 APPLICABLE CONDITIONS OF USE OF THE NOTIFIED SUBSTANCE

1.3.1 Purposes for which the substance is used

Certain fermented foods and alcoholic beverages contain ethyl carbamate (Urethane; see Zimmerli and Schlatter, 1991 for review), which is a known carcinogen in animals (IARC V.7, 1974; IARC S.4, 1982; IARC S.7, 1987), and a suspected carcinogen in humans (IARC, 1974). Ethyl carbamate forms as the result of the reaction of ethanol, a product of the fermentation process, and another precursor such as cyanate or urea. In grape-based and rice-based products that have been fermented by *S. cerevisiae*, ethyl carbamate is known to be primarily produced by the reaction of ethanol and urea.

Urea is normally metabolized in *S. cerevisiae* by the enzyme, urea amidolyase. Production of this enzyme is determined by the *DUR1,2* (alias *DUR80*) gene. During fermentation, in the presence of other good nitrogen sources the *DUR1,2* gene is not normally active and urea is poorly metabolized. As a result, urea accumulates in the yeast cell. As urea is toxic to yeast, it's released into the wine (or other medium, as the case may be) where it reacts with ethanol, which has also been produced by the yeast, to form ethyl carbamate (see Figure 1). Heating and ageing of the wine accelerates the production of ethyl carbamate, and therefore the levels are much higher in sherries and sake/rice wine (Zimmerli and Schlatter, 1991; Daudt *et al.*, 1992).

As a result of the formation of ethyl carbamate by *S. cerevisiae* as a byproduct of fermentation, we propose the use of a genetically modified strain of *S. cerevisiae* named ECMo01 to reduce ethyl carbamate formation in yeast fermented beverages. The modified strain contains the urea amidolyase gene (*DUR1,2*) under the *S. cerevisiae PGK1* promoter and terminator signals. Increased expression of the *DUR1,2* gene in the recombinant strain, ECMo01, results in more effective degradation of urea during alcoholic fermentation (Figure 2). **As these DNA fragments are all obtained from *S. cerevisiae*, ECMo01 is non-transgenic.**

1.3.2 Foods in which the substance is to be used

The ECMo01 active dry yeast is intended for use in alcoholic beverage production procedures in accordance with Good Manufacturing Practices, as a yeast starter culture for alcoholic beverage fermentation such as grape must and rice fermentations. Like all active dry yeast, the active dry yeast obtained with the ECMo01 strain is used to perform alcoholic fermentation in the production of alcoholic beverages. The following grape-based and rice-based fermented products have been shown to contain ethyl carbamate which has been formed from urea, and therefore may benefit in reduction of ethyl carbamate levels by employing the ECMo01 strain during production:

- Red and white wine
- Champagne
- Sherry
- Sake & other rice wines

- Brandy & cognac

The following products may also benefit from employing the ECMo01 yeast strain in the fermentation process:

- Whiskey
- Rum
- Other distilled beverages

1.3.3 Levels of use of the ECMo01 active dry yeast in grape must

To maximize urea degradation, we recommend that winemakers use approximately 0.2 gram of active dry yeast per litre of wine for grape must fermentation. The Bureau of Alcohol, Tobacco and Firearms (27 CFR 24.176) and the International Oenoligical Codex (2002), however, allow the winemaker to determine the appropriate level of yeast inoculation in wine.

If a mixture of ECMo01 and a non-urea degrading yeast strain is used in fermentation, urea degradation may not be carried out effectively and ethyl carbamate levels may not be as effectively reduced. By choosing the appropriate yeast starter culture for fermentation, the winemaker can decide not to carry out urea degradative fermentation, or to carry out urea degradative fermentation with the ECMo01 yeast strain.

1.4 THE BASIS FOR THE GRAS DETERMINATION

First Venture Technologies Corporation has determined that the yeast ECMo01 is GRAS for use as a starter culture in grape must fermentation, through scientific analysis. Based on the scientific data, the company also concludes that ECMo01 can be considered GRAS for use as a starter culture in any alcoholic beverage fermentation. First Venture also asserts that ECMo01 is substantially equivalent to the host strain from which it was derived, except for the enhancement of the metabolic pathway catalyzing the conversion of urea to ammonia. **No other metabolic difference is evident between the modified and the host strain during anaerobic alcoholic batch fermentations**.

1.5 DATA AND INFORMATION THAT ARE THE BASIS FOR THE NOTIFIERS' GRAS DETERMINATION ARE AVAILABLE FOR THE FOOD AND DRUG ADMINISTRATION REVIEW AND COPYING

First Venture Technologies Corporation undertakes that all data and information used to set up the basis for the GRAS determination for use of the ECMo01 yeast strain in production of alcoholic beverages will be made available for the FDA to review and copy at the following address:

First Venture Technologies Corporation
Attention: Mr. Howard J. Louie, President & Director
Suite 1500 – 885 West Georgia Street
Vancouver, British Columbia, V6C 3E8
Canada

Data will be sent to the FDA upon request.

2 DETAILED INFORMATION ABOUT THE IDENTITY OF THE NOTIFIED SUBSTANCE

2.1 HOST MICROORGANISM

The host yeast strain was an industrial wine strain of *S. cerevisiae* called Davis 522. The Davis 522 active dry yeast wine yeast strain has been commonly used in industry for the production of commercial wines in the US since 1963.

2.1.1 History of use

The yeast *S. cerevisiae* is found on our bodies and in the air that we breathe, and it grows naturally on foods that we eat regularly. It has also been employed for use in food production for over five thousand years. Ancient Egyptians, Hebrews and Greeks used yeasts to produce fermented products such as bakery products, wine and beer. Over 2.5 million tons of yeasts are commercially produced each year worldwide, the largest volume of any microorganism.

Baker's yeast serves as a model organism for scientific research. The reasons for this are numerous. Yeast cells are structurally and genetically similar to human cells. They each contain about 6,000 genes, many of which are identical or similar to human genes. Over the last 20 years, researchers have found that their genes can be manipulated with relative ease and efficacy. As a result, most of what we know about the cell cycle and human cancer was discovered from research on yeast. Throughout time, the classification of *S. cerevisiae* has undergone many changes (an historical survey of the species is found in Lodder and Kreger-van-Rij, 1967; Lodder, 1970; Kregar-van Rij, 1984; Kurtzman and Fell, 1998).

Yeasts are very convenient organisms to work with in small and large volumes. Their growth rates are high and predictable, and their proliferation can be controlled and scaled as desired. Their needs are very basic. But perhaps **the primary reason that yeast has become one of, if not the most widely-studied and understood organism, is that they are known to be generally non-pathogenic and *safe*.**

2.1.2 Characteristics

The Davis 522 wine yeast strain is commonly used in industry to obtain wine active dry yeast. The production of Davis 522 vegetative cells during manufacturing occurs asexually through budding to obtain wine active dry yeast; after rehydration and during fermentation of must, reproduction also occurs asexually.

Davis 522 is a strong fermenter with good ethanol tolerance, and will readily ferment grape musts and fruit juice to dryness. It also has good tolerance to free sulphur dioxide. It is recommended for use in production of full-bodied red and white wines. The manufacturer, Lesaffre Yeast Corporation, does not recommend its use for grapes that have been dusted with sulfur because of a tendency to produce

hydrogen sulfide in the presence of higher concentrations of sulfur compounds. Davis 522 is noted for low volatile acidity, good complex flavors, and intense color.

Davis 522 is normally grown on cane or beet molasses under highly aerobic conditions, which assures better survival of live cells toward the end of the fermentation than yeasts grown in must under anaerobic conditions. It has been adapted to perform well in the presence of normal levels of SO_2.

2.2 DONOR ORGANISMS

2.2.1 Taxonomy

Only *S. cerevisiae* contributed DNA to ECMo01, and therefore ECMo01 is a strain of the yeast, *S. cerevisiae*. The history of use of this species is given in Section 2.1.1.

Three *S. cerevisiae* strains were used for DNA isolation: strain TCY1 (Cunningham and Cooper, 1991), strain AB972 (Olson *et al.*, 1986; ATCC 204511) which traces directly to strain S288C (Mortimer and Johnston, 1986; ATCC 204508), and strain GC210 (Chisholm and Cooper, 1982).

2.2.2 Genetic material from donor organisms

The *DUR1,2* gene

The *DUR1,2* gene was derived from the *S. cerevisiae* laboratory yeast strain TCY1 (chromosome II). The entire sequence of the *DUR1,2* gene from the same strain has previously been described (Genbauffe and Cooper, 1991). The GenBank accession number of this sequence is NC 001134; GenID is 852507. The *DUR1,2* gene was cloned after amplification by polymerase chain reaction of *S. cerevisiae* TCY1 genomic DNA.

The *DUR1,2* gene encodes the enzyme, urea amidolyase. The molecular weight of this protein has been determined to be 201,830 Da. Urea amidolyase has both urea carboxylase and allophanate hydrolase enzymatic activities, and catalyzes the degradation of urea to CO_2 and NH_3. Expression of urea amidolyase in *S. cerevisiae* is sensitive to nitrogen catabolite repression, and is induced by allophanate, an intermediate in allantoin degradation.

The *PGK1* promoter and terminator sequences

The *PGK1* promoter and terminator sequences were isolated from a genomic library of the *S. cerevisiae* strain, AB972 (chromosome III; Olson *et al.*, 1986). The *PGK1* promoter was chosen for its strong promotion of transcription. It contains a complex UAS at -473 to -422, and the TATA box is found at position -145. A pyrimidine-rich cluster that serves as a signal for high-level transcription is found between the TATA element and the start site. Although designated as a "constitutive promoter", it is regulated by glucose concentration of the media and drives gene expression best at high glucose concentrations. During alcoholic fermentation of grape juice, highest level of *PGK1* expression in yeast occurs during the exponential growth phase (Puig and Perez-Ortin, 2000a and 2000b). The *PGK1* promoter used in this study consists of a 1486 bp fragment ranging from position -

1487 to position -2 which contains all the *cis*-regulating elements. The *PGK1* terminator allows for correct termination of the mRNA during gene expression. This 264 bp fragment, ranging from position +2756 to +3019 and the *PGK1* promoter have been used in previous studies involving the heterologous expression of genes of oenological interest (Volschenk *et al.*, 1997a; Volschenk *et al.*, 1997b; Volschenk *et al.*, 2004; Lilly *et al.*, 2000). It has also been used in GRAS Notice No. GRN 000120.

The *URA3* sequence

The *URA3* region (chromosome V) was cloned through polymerase chain reaction amplification of genomic DNA of the *S. cerevisiae* strain GC210 (Chisholm and Cooper, 1982). This DNA sequence was used to target the integration of the transforming material by homologous recombination. A fragment corresponding to the *URA3* locus was used to flank both ends of the DNA to be integrated. The first one is a 925 bp region spanning from nucleotide -505 to nucleotide +419, containing the *URA3* 5'-region as well as part of the *URA3* open reading frame. The second region corresponds to a 933 bp fragment spanning from nucleotide +420 to nucleotide +1352, which contains the second half of the *URA3* open reading frame as well as part of the 3'-region. The *URA3* gene (SGD accession number S000747) encodes for an enzyme of the uracil biosynthetic pathway, the orotidine-5'-phosphate decarboxylase (E.C.4.1.23). However, because the gene sequence was divided into two, it is not functionally expressed.

2.3 THE MODIFIED MICROORGANISM

2.3.1 Final construct used in the integration strategy

2.3.1.1 Construction strategy

The final genetic construct used for the transformation of the host organism (the integration cassette) was isolated from plasmid **pHV10dur$_{11}$**. The construction of plasmid **pHV10dur$_{11}$** is the result of successive cloning events which are detailed in the following sections. A flow diagram is given in Figure 3. Sequences of primers used in amplification reactions, and linkers used in ligation reactions are listed in Table 1. Plasmids used in successive cloning events are listed in Table 2.

The *DUR1,2* gene was produced using the polymerase chain reaction technique. The ends of the PCR fragment were modified with T4 DNA polymerase (in the presence of dTTP and dCTP) in order to clone into **pHVX2** (Figure 4).

The *DUR1,2* gene was cloned into the plasmid **pHVX2** by first cleaving the plasmid with *Eco*RI and *Bgl*II restriction enzymes. Klenow was then added in the presence of dATP and dGTP to produce ends that were complementary to the ends of the *DUR1,2* linear fragment above. Ligation of the *DUR1,2* linear fragment and the open **pHVX2** plasmid yielded plasmid **pDUR1,2** (see Figure 5).

Plasmid **pHVX2** has been previously described by Volschenk *et al.* (1997a); plasmid **pHVJH1** is shown in Figure 6 (Volschenk *et al.*, 2004).

The *DUR1,2* expression cassette containing the *DUR1,2* gene between *PGK1*p and *PGK1*t was isolated from plasmid **pDUR1,2** by PCR amplification. Amplification of the *DUR1,2* expression cassette used primers **J22** and **J23**, both carrying a *Not*I site at their 5' ends.

The PCR product of the *DUR1,2* gene was first cloned into plasmid **pHVX2**, the *DUR1,2* cassette was then isolated by PCR as a *NotI* fragment and subcloned into the *Not*I site of plasmid **pHVJH1**, yielding plasmid **pHV10dur$_{11}$** (Figure 7).

2.3.1.2 Detailed description of the final construct

The final construct, designated as the "integration cassette", was isolated from plasmid **pHV10dur$_{11}$** by digestion using the *Srf*I restriction enzyme. The integration cassette consists of the "*DUR1,2 cassette*", which is the fragment to be inserted, flanked by homologous sequences of the *S. cerevisiae URA3* locus. The *DUR1,2* cassette contains the *S. cerevisiae* gene cloned between the promoter and terminator sequences of the *S. cerevisiae PGK1* gene ("*DUR1,2, expression cassette*"). Small synthetic linkers used in the cloning strategy are also present in this cassette. A schematic representation of the integration cassette is shown in Figure 8. The exact nucleic composition of the integration cassette is detailed in Table 3, and the sequence of the cassette is given in Fasta format in Appendix 1.

2.3.2 The transformation event

2.3.2.1 Genetic material used for the transformation method

The integration cassette was isolated from plasmid **pHV10dur$_{11}$** by digestion with the restriction enzyme *Srf*I. This resulted in the purification of a 9191bp fragment containing the *DUR1,2* expression cassette flanked by the *URA3* sequences (the integration cassette; Figure 8). To decrease the number of clones to analyze after transformation, the host strain cells were transformed with a mixture of the integration cassette and the **pUT332Δ*URA3*** plasmid. The plasmid **pUT332Δ*URA3*** was created by *Bgl*II restriction enzyme digest of **pUT332** (Gatignol *et al.*, 1990), removal of the *URA3* sequences, and re-ligation of the vector sequences to yield **pUT332Δ*URA3***. This plasmid carries a selectable transformation marker, the Tn5*Ble* gene, conferring resistance to Phleomycin to yeast cells. The use of plasmid **pUT332Δ*URA3*** in combination with the integration cassette relies on the hypothesis that cells transformed with plasmid **pUT332Δ*URA3*** are more susceptible of also having integrated the cassette than cells not carrying plasmid **pUT332Δ*URA3***. Hence, a first selection of Phleomycin resistant clones increases the chances of finding an integrative clone.

2.3.2.2 Screening method for transformants

After a first screening round on Phleomycin containing media, a thousand Phleomycin resistant clones were further analyzed for the presence of the integration cassette. This second screening procedure relied on colony PCR which used primers inside and outside of the cassette to confirm integration. Out of the clones tested, one was isolated by the second screening method and validated for the integration of the *DUR1,2* cassette into the *URA3* locus and named ECMo01.

2.3.3 Genetic characterization of the modified microorganism

2.3.3.1 The loss of the antibiotic resistance genes and of other plasmid sequences

To eliminate plasmid **pUT332Δ*URA3*** whose only purpose was to serve in the early steps of the screening method, strain ECMo01 was cultivated in non-selective media (*i.e.* media lacking Phleomycin). After successive culturing for five days, replica plates on non-selective media and on Phleomycin containing media were made in order to select a Phleomycin-sensitive clone. The selected clone was then further verified for Phleomycin resistance to confirm the loss plasmid of **pUT332Δ*URA3*** (Figure 9).

Southern blots were performed using a probe representing the Tn5*Ble* gene of plasmid **pUT332Δ*URA3*** (Figure 10). Results clearly show the absence of TN5*Ble* sequences in the ECMo01 strain (Figure 11).

Plasmid **pUT332Δ*URA3*** also contains the Ampicillin resistance gene, *bla*, used as an *Escherichia coli* transformation marker. Verification that this gene had not been integrated in the ECMo01 yeast genome was performed by Southern blot. Results obtained from Southern blot indicate that the ECMo01 strain does not contain the *bla* gene (see Figure 12).

2.3.3.2 Genetic analysis by Southern blot

Genomic DNA from ECMo01 and the host strain was isolated, digested with restriction enzymes and separated by electrophoresis in a 0.7% agarose gel. Following electrophoresis, the DNA was blotted onto positively charged Nylon membranes and fixed.

First, a Southern blot was performed using *Eco*RI-digested genomic DNA and the Tn5*Ble* probe. The probe was generated by PCR using **pUT332Δ*URA3*** used for co-transformation as template. No signals were detected in either the host strain or the ECMo01 strain for the Tn5*Ble* gene that encodes Phleomycin resistance (Figure 11). This indicates that there is no Tn5*Ble* gene integrated into the genome of ECMo01.

A second Southern blot was performed using also *Eco*RI-digested genomic DNA and the *bla* probe. The probe was generated by PCR using **pUT332Δ*URA3*** for co-transformation as template. No signals were detected in either the host strain or the ECMo01 strain for the *bla* gene that encodes for Ampicillin resistance (Figure 12). This indicates that there is no *bla* gene integrated into the genome of ECMo01.

Southern blot was performed using *Bgl*II-digested genomic DNA and a *URA3* probe. One signal was detected in the lane corresponding to the host strain, representing a 4.455 Kbp DNA fragment from a non-integrated copy of the *URA3* locus (Figure 14). The higher band at 4.455 Kbp indicates that strain ECMo01 still contains at least one non-disrupted *URA3* locus, while the lower signal corresponds to a 3.134 Kbp DNA fragment resulting from the integration event in a *URA3* locus.

To clearly identify the integration event in the ECMo01 strain, a Southern blot was performed using the *DUR1,2* probe generated by PCR using plasmid **pDUR1,2** as template. The *DUR1,2* probe hybridized to *Bgl*II-digested genomic DNA, revealing a single signal representing a 5.085 Kbp DNA fragment in the host strain (Figure 15). Whereas two bands were revealed for ECMo01, the lower bands corresponds to wild type *DUR1,2* (5.085 Kbp) whereas the second higher band of 7.456 Kbp corresponds to the correctly integrated *DUR1,2* gene in the *URA3* locus.

Results gathered from the Southern blots suggest that the *DUR1,2* cassette has been integrated in a *URA3* locus.

2.3.3.3 Genetic stability of the ECMo01 yeast strain

The ECMo01 strain was cultivated in YPD media for 100 generations. After 100 generations, the yeast were plated on YPD media and 97 random colonies were analyzed by colony PCR for presence of the *DUR1,2* cassette; 97 of the 97 colonies (100%) tested positive for the *DUR1,2* cassette, indicating that **the integrated *DUR1,2* cassette in the *URA3* locus of the ECMo01 strain is genetically stable.**

Genetic analysis of the ECMo01 strain shows that the integration event has occurred. A Southern blot on digested genomic DNA shows that the integrated material contains the *DUR1,2* cassette.

Results from the Southern blots using the *URA3* probe indicate that in strain ECMo01, the *DUR1,2* cassette is integrated into one locus of the *URA3* gene, while the other *URA3* locus is left intact. In addition, all existing data show that the ECMo01 yeast is stable.

2.3.4 DNA sequencing of the integrated locus

Additional analyses were performed to verify the sequence of the integrated DNA at the *URA3* locus. The strategy involved the sequencing of several amplification products corresponding to the inserted genomic region. At least two templates (for each region sequenced) were obtained by PCR using a proof-reading polymerase. The sequences obtained from one template were aligned to the sequences obtained from the second template and analyzed for differences. Sequencing of two sets of templates (each from a specific PCR amplification) minimizes sequencing errors due the amplification procedure. If differences occurred, another round of PCR and sequencing of the region in question was run to determine the correct sequence.

The sequence obtained from the genomic integrated locus was compared to previously published sequences (see Appendix 2). The published sequences were obtained from the *S.* Genome Database for *URA3*, *DUR1,2*, *PGK1*p and *PGK1*t. Table 4 lists the differences between the sequences.

2.3.4.1 Differences found in the *PGK1* promoter, *PGK1* terminator and *URA3*

Four differences were found in the *URA3* sequences (see comments 1, 32, 33, and 34 in Table 4). These sequence variations could be the result of a DNA synthesis error during the amplification steps involved in the construction of the integration cassette, but could also be due to genetic polymorphisms. Strain AB972, from which the sequence for our study was isolated, derives directly from strain S288C, the strain from which the published sequence originates. However, strain AB972 is an ethidium bromide induced mutant, in which ethidium bromide may have induced DNA mutations in the *URA3* locus without any visible phenotype. Alternatively, differences may be attributed to errors in the original published sequence.

Three other differences were found, involving the *PGK1* promoter and terminator sequences (comments 3, 29 and 30 of Table 4). These sequence variations could be the result of a DNA synthesis error during the amplification steps involved in the construction of the integration cassette, but could also be due to genetic polymorphisms. The strains from which the compared sequences derive from are different (S288C for the published sequence, GC210 for the sequence of our study), increasing the possibilities of genetic polymorphisms. Alternatively, differences may be attributed to errors in the original published sequence.

2.3.4.2 Differences in the *DUR1,2* gene

Twenty-three differences were found in the coding sequence of the *DUR1,2* gene. Seventeen of these differences result in silent mutations involving no change in the amino acid sequence. Six differences result in amino acid changes (Table 4). These disparities could originate from an error during the amplification steps involved in the cassette construction, but could also be due to genetic polymorphisms. The strains from which the compared sequences derive from are different (S288C for the published sequence, TCY1 for the sequence of our study), increasing the possibilities of genetic polymorphisms. Alternatively, these differences can be attributed to errors in the published sequences.

The integration cassette was shown to contain only synthetic polylinkers and DNA from *S. cerevisiae*. When comparing the integrated cassette sequences to previously reported sequences, seven differences were found in either non-coding regions (promoter and terminator), or in the non-functional *URA3* open reading frame. These differences may be due to strain genetic polymorphism. However, they may also be the result of DNA amplification errors. Twenty-three changes were found in the *DUR1,2* sequence, seventeen corresponding to silent mutations, the other six involving the exchange of one amino acid for another. These results indicate that the sequences composing the integration cassette of the ECMo01 strain are not significantly different from the original sequences isolated from the donor strains.

2.3.5 Absence of difference between genetic profiles of the transformed and the host strain

The genome of *S. cerevisiae* contains long terminal repeat sequences known as δ elements. These δ elements are the remnants of Ty1 transposon integration events. The number and location of these δ elements are specific to a strain and has been used to fingerprint and differentiate between strains of *S. cerevisiae*. Using PCR and primers J32 and J33 to amplify these δ sequences we have verified the genetic relationship between the host strain and ECMo01. This technique is commonly employed to differentiate between industrial *S. cerevisiae* strains.

We therefore conclude that the recombinant ECMo01 strain is not different than the Davis 522 parental strain, except for the fully-characterized integrated segment. No major DNA reorganization event has occurred upon integration of the *DUR1,2* cassette into the ECMo01 strain.

2.3.6 Method of manufacture of the modified microorganism

The ECMo01 yeast was modified by integration of native DNA into its genome and the integration event was shown to be stable. Thus, no selective media needs to be used for the propagation of the ECMo01 yeast strain. The ECMo01 wine active dry yeast can therefore be manufactured in the exact manner as all active dry yeast used in brewing or winemaking, as described in the literature (*e.g.*, see Reed and Nagodawithana, 1991). A brief description of the manufacturing of such yeast is given in Figure 17.

3 Information on any self-limiting levels of use

First Venture Technologies Corp. sees no reason why the inoculation rate should be different than the one recommended as indicated in Section 1.3.3.

4 DETERMINATION THAT THE USE OF THE WINE ACTIVE DRY YEAST OBTAINED FROM THE ECM001 STRAIN IN THE FERMENTATION OF WINE AND OTHER ALCOHOLIC BEVERAGES IS EXEMPT FROM THE PRE-MARKET APPROVAL REQUIREMENTS OF THE ACT BECAUSE SUCH USE IS GRAS

4.1 SAFETY ASSESSMENT OF THE HOST STRAIN

Davis 522 belongs to a well-known family of wine yeast strains corresponding to the *S. cerevisiae* species. It was derived from the yeast collection of the University of California, Davis. Davis 522 was originally isolated by Pacottet from the Burgundy region in France and was brought to UC Davis by F.T. Boletti (Subden, 1990). The strain belongs to a family of very close or identical commercial strains that are commonly used as wine strains and often designated as "Montrachet" strains (Dunn *et al.*, 2005). It is also commonly referred to as UCD 522.

S. cerevisiae is an organism which has an extensive history of safe use. It has been used for millennia in fermentation processes, such as bread leavening and wine or beer production. It is responsible for spontaneous fermentation of grape juice (Lodder and Kreger-van-Rij, 1967). The Bureau of Alcohol, Tobacco and Firearms rates yeast or yeast cultures grown in juice of the same kind of fruit (here grapes) as permitted material added in the production of natural wines (27 CRF 24.176).

S. cerevisiae is considered GRAS through its use in the brewing, baking and winemaking industry. Its genome has been sequenced, and it has been determined that the yeast is free of known pathogenicity traits. In addition, the Environment Protection Agency (EPA) has included *S. cerevisiae* as a recipient microorganism for exemptions from EPA review and expedited EPA review (40 CFR 725.420). This exemption was made because this species is found to have little adverse effects. They also determined that the introduction of genetic material will not increase the potential for adverse effects, provided that the genetic material is limited in size, well characterized, free of certain sequences and poorly mobilizable.

The Davis 522 yeast strain belongs to the *Saccharomyces cerevisiae* species which has been used for more than 7,000 years by humans in fermented beverages. It can therefore be concluded that species will remain GRAS, even after genetic transformation, as long as no harmful DNA material is added to it.

4.2 SAFETY ASSESSMENT OF THE GENETIC MATERIAL USED TO CONSTRUCT THE MODIFIED ORGANISM

4.2.1 Foreign genetic material source and product

No foreign genetic material has been incorporated into the ECMo01 yeast strain.

4.2.2 Native genetic material source and product: *DUR1,2* gene, *PGK1* promoter, *PGK1* terminator and *URA3* sequences.

4.2.2.1 The *DUR1,2* gene

Source*: *S. cerevisiae

The *DUR1,2* gene was isolated from the laboratory strain of *S. cerevisiae* as described in Section 2.2.2. TCY1 is a 5-FOA^r derivative of GC210, a yeast that is Matα and *lys*2 (Cunningham and Cooper, 1991); GC210 is an isogenic derivative of sigma 1278b (Chisholm and Cooper, 1982) which is Matα wild-type. The parents of sigma 1278b were the baking strains, "Yeast Foam" and 1422-11D (Grenson *et al.*, 1966; Bechet *et al.*, 1970).

Product:* *urea amidolyase

When comparing the sequence of the recombinant *DUR1,2* gene (from TCY1) in the ECMo01 strain and the native gene sequence in *S. cerevisiae* (S288C), 23 changes were found (Table 4). However, 17 of these 23 differences correspond to silent mutations. The remaining six nucleotide differences do correspond to a change in amino acid (Table 4), nevertheless we can infer that these changes do not affect the overall protein since the *PGK1*p-*DUR1,2*-*PGK1*t expression cassette is fully functional.

Therefore, to the best of our knowledge, the *DUR1,2* gene product of the ECMo01 strain is similar to that of the parent *S. cerevisiae* strain.

4.2.2.2 *PGK1* promoter, terminator and *URA3* sequences

When comparing the sequence of the recombinant *DUR1,2* gene in the ECMo01 strain and the native gene sequence in *S. cerevisiae*, 11 changes were found (Table 4). However, the safety assessment of *S. cerevisiae* has already been established in Section 4.1 and since the *PGK1* regulatory sequences do not code for any proteins, and the *URA3* locus is disrupted, no further safety assessment is needed.

4.2.3 Construction of the modified organism

4.2.3.1 Vector used as source of the integration cassette

Plasmid **pHV10dur$_{11}$** was used in the construction of the modified ECMo01 yeast strain as a source of the integration cassette. Plasmid **pHV10dur$_{11}$** is the result of the insertion of the integration cassette in plasmid **YEp352** whose unique *Kpn*I site had been destroyed. Plasmid **YEp352** is a well known and well

characterized plasmid and is deposited in the American Type Culture Collection (ATCC) under accession number 37673. It is a *Escherichia coli*/yeast episomal shuttle vector containing multiple restriction sites for cloning. It contains an Ampicillin resistance gene and the *URA3* gene as selectable markers. Plasmid **YEp352** is itself the result of a construction by filling in and blunt-end ligating plasmid **pUC18** (ATCC 37253; Viera and Messing, 1982) linearized with a fragment of the 2 micron plasmid and a fragment containing *URA3* (Hill *et al.*, 1986).

The final construct (*i.e.*, the integration cassette) was isolated from plasmid **pHV10dur$_{11}$** by digestion with restriction enzyme *Srf*I and subsequent extraction from an agarose gel. Hence, all **pHV10dur$_{11}$** sequences not belonging to the integration cassette have been eliminated from the DNA to be used in the transformation procedure. The integration cassette only contains the *S. cerevisiae PGK1* promoter and terminator, the *S. cerevisiae URA3* sequences, the *S. cerevisiae DUR1,2* gene, and short synthetic polylinkers needed for the cloning strategy.

4.2.3.2 Transformation of the host strain with the integration cassette

The integration cassette was used to transform the host strain Davis 522. As described in Section 2.3.2., host yeast cells were transformed with a mixture of the integration cassette and plasmid **pUT332Δ*URA3*** which was utilized in the first screening round. Once a clone with the integrated *DUR1,2* cassette was isolated, it was cultured in a non-selective medium in order to allow for loss of plasmid **pUT332Δ*URA3***. Different methods were employed in order to verify that sequences from this plasmid had been eliminated from the ECMo01 strain. These methods relied on:

- Verification that the ECMo01 strain has lost its ability to grow on a Phleomycin containing media, indicating the absence of a functional Phleomycin resistance gene (Figure 9);
- Verification by Southern blot of the absence of the Phleomycin resistance gene using a probe corresponding to the **pUT332Δ*URA3*** Tn5*Ble* gene (Figure 11); and
- Verification by Southern blot of the absence of the **pUT332Δ*URA3*** plasmid sequence containing the Ampicillin resistance gene (Figure 12).

The results conclusively show that the pUT332Δ*URA3* antibiotic resistance genes are absent in the ECMo01 genome.

4.3 SAFETY ASSESSMENT OF THE MODIFIED ORGANISM

4.3.1 Characterization of the transformation event

The analysis of the transformation event is described in Section 2.3.3.

Results obtained from hybridization of genomic DNA with probes corresponding to *URA3 or DUR1,2*, have shown that the *DUR1,2* cassette is correctly integrated into a *URA3* locus (Figures 14 to 16).

The genome of *S. cerevisiae* contains long terminal repeat sequences known as δ elements. These δ elements are the remnants of Ty1 transposon integration events. The number and location of these δ elements are specific to a strain and have been used to fingerprint and differentiate between strains of *S. cerevisiae*. Using PCR and primers **J32** and **J33** to amplify these δ sequences, we have verified the genetic relationship between the host strain and ECMo01 (Figure 16).

DNA sequencing has confirmed that only sequences belonging to the *DUR1,2* gene, the *PGK1* promoter and terminator sequences, synthetic linkers used in the cloning strategy and sequences belonging to the *URA3* locus are present at the integration site (Table 3).

The ECMo01 strain has been genetically characterized. The integration site contains no DNA sequences other than those present in the isolated integration cassette. No antibiotic resistance sequences are present in the genetic material of strain ECMo01. Therefore, to the best of our knowledge, the ECMo01 strain is genetically substantially equivalent to the Davis 522 strain, except for the *DUR1,2* cassette present in one *URA3* locus.

4.3.2 Consequence of the genetic modification on gene expression

4.3.2.1 Global gene expression analysis

Global gene expression of the ECMo01 yeast was investigated using the Affymetrix GeneChip® Yeast Genome S98 Array. These gene chips contain probes from all known *S. cerevisiae* S288C strain open reading frames, including putative open reading frames suggested by Serial Analysis of Gene Expression (Velculescu *et al.,* 1997), mitochondrial proteins, Ty proteins, open reading frames from the 2 micron plasmid and a small number of open reading frames from strains other than S288C. The expression pattern of the ECMo01 strain during the fermentation of Chardonnay juice was compared to that of the host strain in identical conditions at one time point (24 hours) in triplicate (see Appendix 4). At 24 hours, 5 genes were either up-regulated or down-regulated. *DUR1,2* was the only gene that was up-regulated. One can conclude from these results that no metabolic pathway is affected by the presence of the expression cassette in strain ECMo01.

Hence, from global gene expression analysis, we have concluded that strain ECMo01 can be regarded as substantially equivalent to the wild type strain Davis 522.

4.3.3 Presence of unintended gene products as a result of the transformation event

S. cerevisiae has been employed by humans in production of food and beverages for millenia. It is known that during this period, genetic changes have occurred, however, this yeast has never been described as having adverse effects on humans.

The sequences composing the integration cassette do not raise any safety concerns. However, in the *DUR1,2* cassette of the ECMo01 strain, the *URA3, PGK1p, PGK1t, DUR1,2* sequences are present in a different genetic environment

than their homologous native sequences. Therefore, we identified new open reading frames (ORFs) putatively expressed as a consequence of the integrated cassette.

ORFs were identified using Omiga software (version 5), which searched for ORFs of more than 100 codons. Using the software, three new open reading frames were found (Figure 18). The nucleotide sequence of the putative ORFs and of the deduced amino acid sequences are given in a Fasta format in Appendix 5.

In order for these ORFs to correspond to putative proteins, they first need to be transcribed into messenger RNA. It is possible that putative ORF1 (5'*ura3* truncated) and putative ORF2 (5'*adp1* truncated) are transcribed as they have the same upstream regulatory sequences as the native *URA3* and *ADP1* genes. However, it is unclear if putative ORF 3 (3'*ura3* truncated) will be transcribed since the upstream sequences are in fact *PGK1* terminator sequences.

If any of these putative ORFs are transcribed, RNA stability would be a limiting factor for protein coding capacities of these putative ORFs. In this regard, the 3' untranslated region of these ORFs is of main importance. If these regions are not adapted or are poorly adapted to enhance mRNA stability, then the RNA corresponding to these ORFs will likely have a very short half life. In this respect, putative ORF1 and ORF2 may not be stable, and may degrade rapidly.

If transcription of any putative ORFs occurs, the corresponding RNAs need to be translated. Efficient translation will only occur if the putative ORF possesses untranslated 5' sequences favoring translation initiation. In *S. cerevisiae*, consensus sequences that participate in translation efficiency exist, but it is difficult to predict whether an ORF will be translated by analyzing its 5' sequence. Nevertheless, because they share the same 5' untranslated sequences as the *URA3* and *ADP1* genes, putative ORF1 and ORF2 could be translated. It is difficult to precisely assess the translation efficiency of putative ORF3, but it is unlikely.

Although the putative ORFs may be translated into proteins, we do not feel that this is a concern. Each of these sequences mainly corresponds to *S. cerevisiae* sequences (putative ORF1: mainly *URA3* sequences plus 16 nucleotides from synthetic sequences plus 11 nucleotides from *PGK1* promoter; putative ORF2: mainly *ADP1* and *URA3* sequences plus 16 nucleotides from synthetic sequences; putative ORF 3 is only comprised of *URA3* sequences). Hence, the large majority of the nucleotide sequences of these putative ORFs can also be found in the parental strain.

In conclusion, given:

- that the Davis 522 strain was never shown to release allergenic or toxic substances, and
- the close sequence similarity between the putative ORFs (ORF1, ORF2 and ORF3) and *S. cerevisiae* parental sequences,

there is no reason to believe that the ECMo01 strain will release toxic or allergenic substances as a result of the presence of unintended gene products.

4.3.3.1 Urea Amidolyase

It has been determined that the *DUR1,2* gene in the ECMo01 strain is up-regulated about 17-fold when compared to the parent strain. It may therefore be assumed that the urea amidolyase protein is also up-regulated by a similar amount. This level of increase in enzyme production can also be assumed by the dramatic reduction in ethyl carbamate formation (see Table 5), which normally forms in wine through the reaction of ethanol with urea (the latter is degraded by urea amidolyase).

It may be assumed, however, that active urea amidolyase is not present in the final wine product at the time of consumption for the following reasons: 1) Residual activity declines sharply with an increase in temperature, and falls below 50% after 7 days at 35°C (Roon and Levenberg, 1972), and 2) During storage on lees, cell proteins normally undergo intracellular enzymatic degradation by yeast intracellular proteases and amino peptidases (see Section 4.4.2.1). It should also be noted that urea amidolyase activity is negligible at a pH less than 6.5 (Roon and Levenberg, 1972); the pH of white and red wines normally ranges between 3 and 4. In addition, bentonite treatment, which is common in white wines, will also remove protein fractions of the wine (see Dawes *et al.*, 1994). **It is therefore reasonable to conclude that use of the ECMo01 strain will not result in additional urea amidolyase presence in wines at the time they are consumed.**

4.3.4 Consequence of the genetic modification on the physiology of the ECMo01 strain

4.3.4.1 Overview of the function of the urea amidolyase gene

The *DUR1,2* gene codes for the enzyme, urea amidolyase. Urea amidolyase, contains both urea carboxylase and allophanate hydrolase activities, degrades urea to CO_2 and NH_3; expression sensitive to nitrogen catabolite repression and induced by allophanate, an intermediate in allantoin degradation. This enzyme requires the co-factor biotin, and ATP.

4.3.4.2 Manufacture of active dry yeast during fed-batch aerobic cultures

Section 2.3.6 gave an overview of the method of manufacture of the ECMo01 strain in the form of wine active dry yeast. This method of manufacture is identical to that used for the current commercial manufacture of wine active dry yeast obtained with the host strain Davis 522.

4.3.4.3 Growth, ethanol production, and CO_2 production of ECMo01 and Davis 522 strains during laboratory-scale vinifications

Growth rates of the ECMo01 and Davis 522 host strain fermenting Chardonnay must were monitored by measuring optical density. Results presented in Figure 19 indicate that the growth curves are similar for both strains.

Vinification trials performed in the laboratory have also shown that the Davis 522 and the ECMo01 strain have similar fermentation rates and ethanol yields (see Figure 19). CO_2 production during fermentation is also virtually identical to that of the parent strain (Figure 20).

Therefore, the presence of additional urea amidolyase does not give the ECMo01 strain any growth advantage, even though the organism is able to degrade urea more effectively than the parent strain.

Vinification trials and results from global gene expression showed that no other difference is found between the ECMo01 strain and the Davis 522 host strain.

Genetic studies have shown that the ECMo01 strain is genetically similar to the Davis 522 host strain, except for the integrated *DUR1,2* cassette. Global gene expression analysis indicates that the introduction of the *DUR1,2* cassette does not change gene expression patterns in the ECMo01 yeast significantly. This leads to the conclusion that strain ECMo01 can be regarded as substantially equivalent to the wild type strain Davis 522, with an enhanced capacity to degrade urea to ammonia. Physiological studies carried out in different culture conditions indicate that no difference between the ECMo01 and Davis 522 strains can be seen, except for urea degradation. The enhanced degradation doesn't confer any significant growth advantage to the ECMo01 strain. Therefore, it can be concluded that the ECMo01 strain is substantially equivalent to the host strain Davis 522 except for the enhanced ability of ECMo01 to degrade urea to ammonia during wine fermentations.

4.4 SAFETY ASSESSMENT OF THE PRODUCT DERIVED FROM THE MODIFIED ORGANISM

4.4.1 Changes in the winemaking procedures as a consequence of the ECMo01 fermentation

The use of the ECMo01 yeast strain leads to no change in winemaking procedures. Figures 21 and 22 show winemaking procedures of white and red wines, respectively, that may be employed using our yeast.

4.4.2 Changes in wine composition as a consequence of the ECMo01 fermentation

4.4.2.1 Yeast cells and release of yeast products during winemaking

As the alcohol content of the wine rises, yeast mortality slowly increases and autolysis occurs with time. To remove the yeast from the wine, it is then either clarified immediately or kept on lees during ageing.

Clarification of wine

During wine clarification, solid particles in the wine are allowed to settle by gravity. Through this type of clarification the majority of yeast cells can be eliminated (see Ribéreau-Gayon *et al.*, 1998b).

Wine may also clarified by centrifugation; whereby more than 99% of the yeast cells can be eliminated (Ribéreau-Gayon *et al.*, 1998b). Although this clarification procedure is obligatory in commercial wine production, it may be delayed by storage of the wine on lees.

Ageing of wine on lees

Keeping wine on lees is an optional winemaking procedure that involves storage of the yeast cells and wine together, unclarified in tanks or in barrels. This procedure is followed to further release yeast constituents that enhance wine properties. During ageing, cell proteins and nucleic material undergo intracellular enzymatic degradation as yeast intracellular proteases, amino peptidases, nucleases and phosphatases are gradually released (Fornairon-Bonnefond *et al.*, 2001; Charpentier and Feuillat, 1993). The yeast cells eventually lyse and release their hydrolyzed constituents into the wine where they can be further degraded by proteases in wine. Wine amino acid and peptide content increases as the yeast cells lyse (Martinez-Rodriguez *et al.*, 2000), however, wine protein content decreases as proteases are released, and larger peptides are cleaved (Ari'lzumi *et al.*, 1994; Martinez-Rodriguez *et al.*, 2000).

Employing this storage practice after ECMo01 fermentation may therefore result in the release of some additional *DUR1,2* intact proteins, however, it is more likely that the hydrolysis products of urea amidolyase, such as peptides and amino acids, will be found in these wines stored on lees. Nucleic material from yeast cells is normally found in wine as single nucleic bases or small nucleic base chains, as autolyzing yeast is rich in nucleases and phosphatases.

Wine stabilization

Clarification is a normal step in the winemaking process. The specific clarification procedure, however, is dependent on whether the wine is white or red.

Bentonite treatment of white wines

During white wine production, a white haze may form after bottling as a result of instable protein accumulation (Ribéreau-Gayon *et al.*, 1998b). This haze can be prevented by removing these proteins, which is commonly done by treating the wine with bentonite, a montmorillonite clay. This treatment will remove all protein fractions, and larger polypeptides in wine, including the protein fractions that cause the haze (Dawes *et al.*, 1994). Yeast proteins and yeast protein products are normally only found in wines that have been stored on lees.

Therefore, although a small amount of additional *DUR1,2* protein and larger hydrolysis products of this protein (and other yeast proteins) could be present in wine, they will be partially removed using bentonite treatment.

Stabilization of red wines

In red wines, tannins associate with the proteins that can cause haze in white wines and therefore, far fewer free proteins are available to form a protein haze (Ribéreau-Gayon *et al.*, 1998a). Gelatins or egg white albumin can be employed to precipitate the protein-tannin complexes and facilitate removal from the wine (Ribéreau-Gayon et al., 1998b).

Therefore, if *DUR1,2* protein is present in red wine, it will interact with the wine tannins and be partially removed upon clarification.

***DUR1,2* protein is present in all wine yeast cells currently used by the industry. The presence of additional *DUR1,2* protein expressed under control of the *PGK1* promoter and terminator signals should be insignificant, unless the wine has been stored on lees. Clarification processes, which are obligatory in commercial winemaking, will remove *DUR1,2* protein, and larger polypeptide fragments of this protein.**

Products of the DUR1,2 protein will also be found in wine as a result of fermentation by unmodified yeast. Therefore, the only significant difference expected between a wine produced by Davis 522 yeast, and that produced employing ECMo01, is the possible presence of minor amounts of additional *DUR1,2* protein.

Filtration

Filtration decreases the turbidity of wine before bottling, and is therefore an important step in the winemaking process. This practice is commonly used, however, it is optional in the production of red wines as some winemakers elect to clarify red wine with gelatin or egg white albumin. There are several ways of filtering wine.

Filtration on diatomaceous earth

Filtration on diatomaceous earth is employed for wines that have not been pre-filtered. This type of filtration can remove a large proportion of wine yeast cells (Ribéreau-Gayon *et al.*, 1998b). After filtration, the wine may be bottled or subjected to an additional filtration procedure that usually involves cellulose filters or membrane filters.

Cellulose filters

Clarification of wines using cellulose filters can reduce yeast cell count to up to 1 cell per 100 milliliters of wine (Ribéreau-Gayon *et al.*, 1998b). Cellulose filtration is primarily employed in wine has been previously clarified by gravity sedimentation, centrifugation, or diatomaceous earth filtration.

Membrane filters

This type of filtration is employed just prior to bottling, and is commonly used to obtain sterile wine. Membrane filter porosity normally ranges from 0.45 micrometers to 10 micrometers, however, a porosity of less than 1.2 micrometers is required for removal of yeast cells.

Standard and modern procedures

Modern and standard winemaking procedures consist of filtering the wine to make it as sterile as possible. Therefore, filters with a porosity less than 1 micrometer (usually 0.45 micrometer for white wines and 0.65 micrometers for red wine) are employed. This filtration normally eliminates all live yeast cells or ghost cells from wine.

Our data show that the only significant difference between the Davis 522 and the ECMo01 yeast is the presence of the integration cassette, leading to the increased production of the urea amidolyase enzyme which allows increased urea degradation by the yeast. Vinification trials showed that ECMo01 wines were not different in their global composition than control wines. Hence, the only suspected difference between an ECMo01 wine and a control wine is the potential presence of the ECMo01 recombinant yeast containing the integration cassette. Depending on the clarification and filtration process, different proportions of ECMo01 yeast cells might be found in wine.

4.4.2.2 Formation of unwanted substances in wine

Ethyl Carbamate

Ethyl carbamate is a suspected carcinogen for humans and is often found in wines as a result of the release of precursors by yeast and lactic acid bacteria metabolism.

Ethyl carbamate precursors in wines produced by ECMo01, or Davis 522 yeast. Chardonnay grape must obtained from Calona Vineyards with a total yeast assimilable nitrogen (YAN) content of 402 mg N/L was used for fermentations. Fermentations were conducted as previously described, in triplicate. Wine was analysed by Gas Chromatography / Mass Spectrometry (GC/MS) and ethyl carbamate quantified using the following protocol (for equipment, chemicals and solutions, see Appendix 6).

A 10 mL sample was pipetted into a 20 mL sample vial. A small magnetic stirring bar and 3 g of NaCl were added and the vial was capped with PTFE/silicone septum. The vial was then placed on a stirrer at 22°C and allowed to equilibrate, with stirring, for 15 min. After equilibration, a CW/DVB SPME fiber was inserted into the headspace for 30 min. Then fiber was removed from the sample vial and inserted into the injection port of the GC/MS for 15 min. Analyses were done in triplicate.

The fiber was conditioned at 250°C for 30 min before use. A blank run was performed before sample runs. Quantification was done using external standard method (Whiton and Zoecklein, 2002).

Results shown in Table 5 indicate that the ECMo01 strain produces significantly less ethyl carbamate than the parental Davis 522 strain. This results from the increased efficacy in urea degradation, which is the main precursor of ethyl carbamate in wine.

Hence, the ECMo01 strain produces less ethyl carbamate than the parental strain, and could be used to help reduce the ethyl carbamate content of alcoholic beverages in which urea is a precursor of ethyl carbamate.

Other undesirable substances

The additional *DUR1,2* gene present in the ECMo01 strain as the result of the genetic modification does not code for either toxic or allergenic proteins, nor proteins implicated in the formation of undesirable compounds.

We have carried out risk assessments related to the presence of potential unintended gene products as the result of the genetic modification and found the following:

- Computer analysis shows that the integrated region contains 3 putative open reading frames;
- Our data suggests that they are not subsequently translated into proteins; and
- All of the ORFs correspond to *S. cerevisiae* sequences. Moreover, they share with their parental counterpart the same 5′ untranslated region. This means that if proteins corresponding to these ORFs do exist, then very similar proteins will be found in the parental strain. As the parental strain has never been shown to release unwanted products, neither should the new ECMo01 strain.

Although the more effective degradation of urea will result in more ammonia production by yeast, ammonia is not expected to accumulate in the wine. As a preferred nitrogen source, the yeast is expected to rapidly utilize the additional ammonia.

These considerations lead us to conclude that allergenic or toxic risks related to the presence of these putative ORFs within the new ECMo01 strain are negligible.

Therefore, to the best of our knowledge, the use of the ECMo01 strain in winemaking will not lead to the release or the enhancement of the presence of undesirable compounds in wine. Moreover, the undesirable compound, ethyl carbamate is significantly reduced in wines produced using the ECMo01 strain.

Figures & Tables

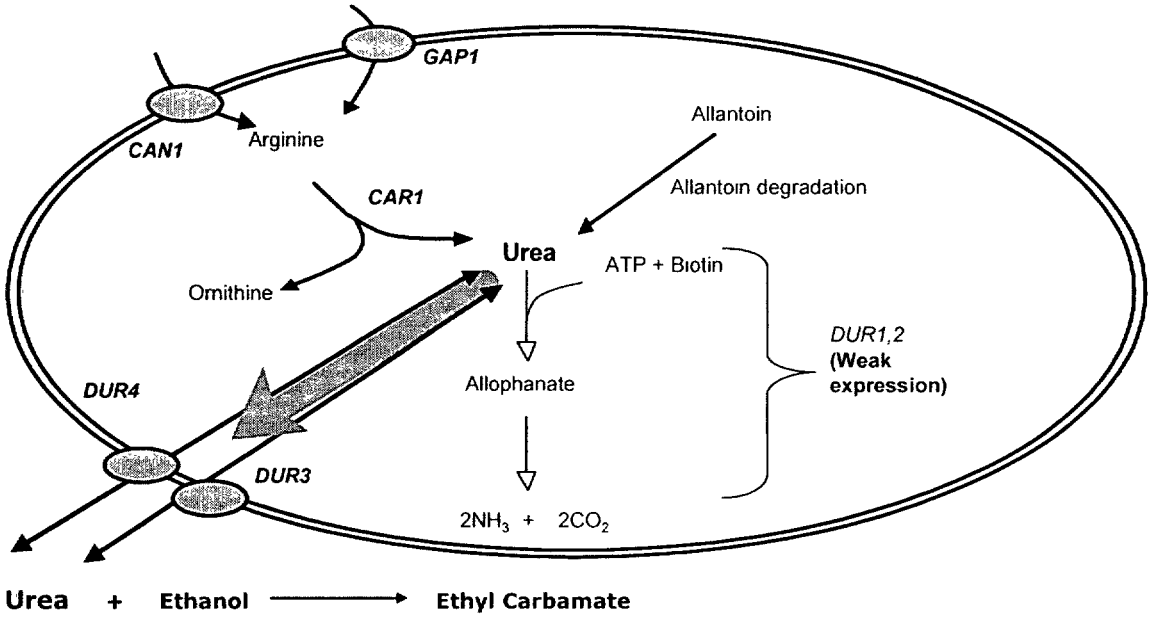


Figure 1: Natural formation of ethyl carbamate by *S. cerevisiae*.

BEST ORIGINAL COPY

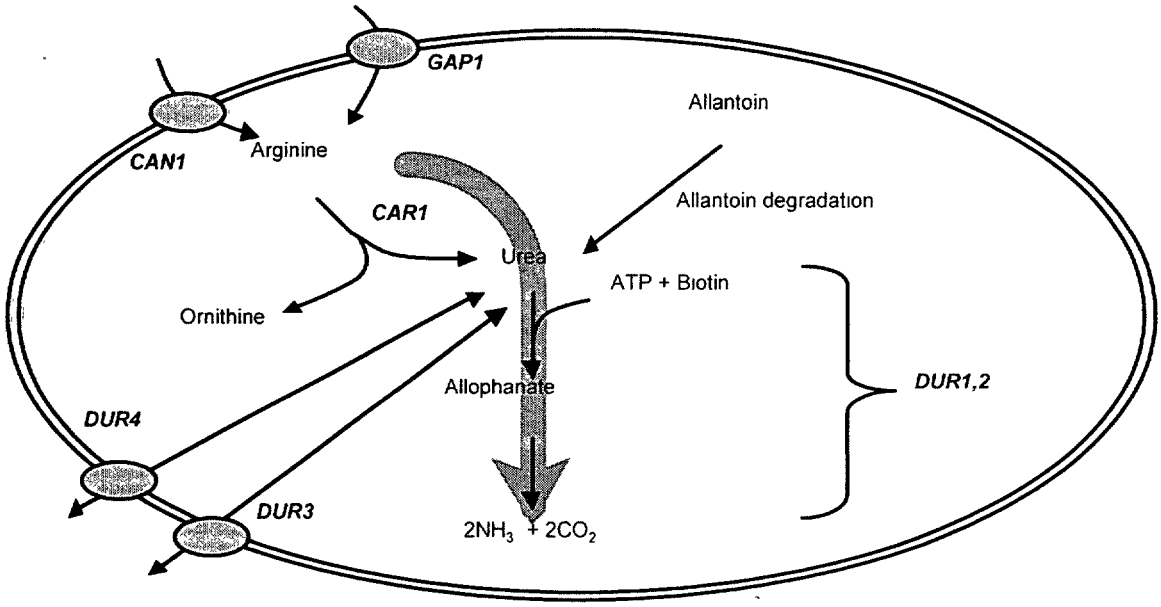


Figure 2: Urea degradation by ECMo01.

BEST ORIGINAL COPY

000040

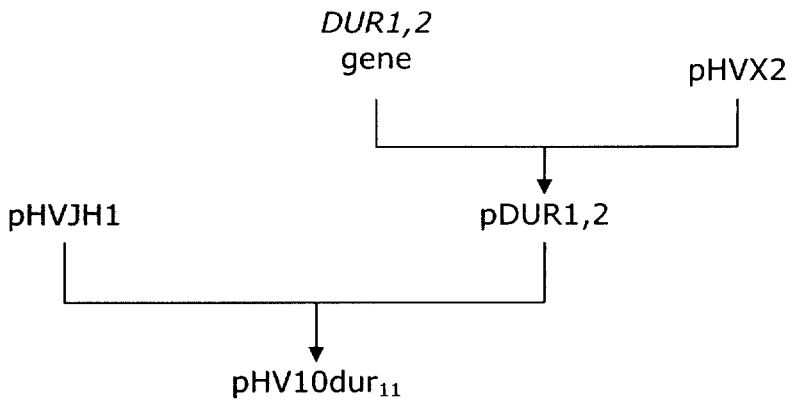


Figure 3: Flow diagram illustrating the cloning strategy used to construct plasmid pHV10dur$_{11}$.

All plasmids are listed in Table 2.

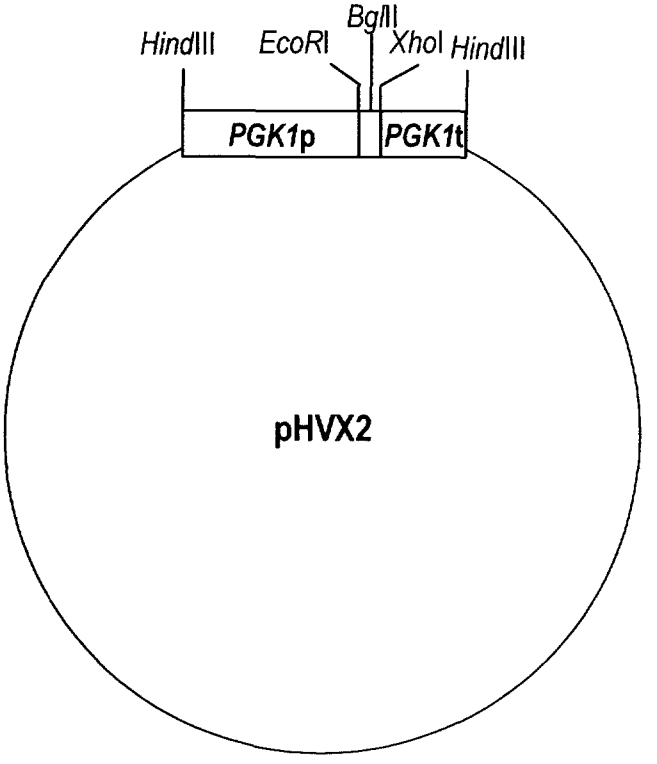


Figure 4: Plasmid pHVX2, previously described in Volschenk *et al.*, 1997a.

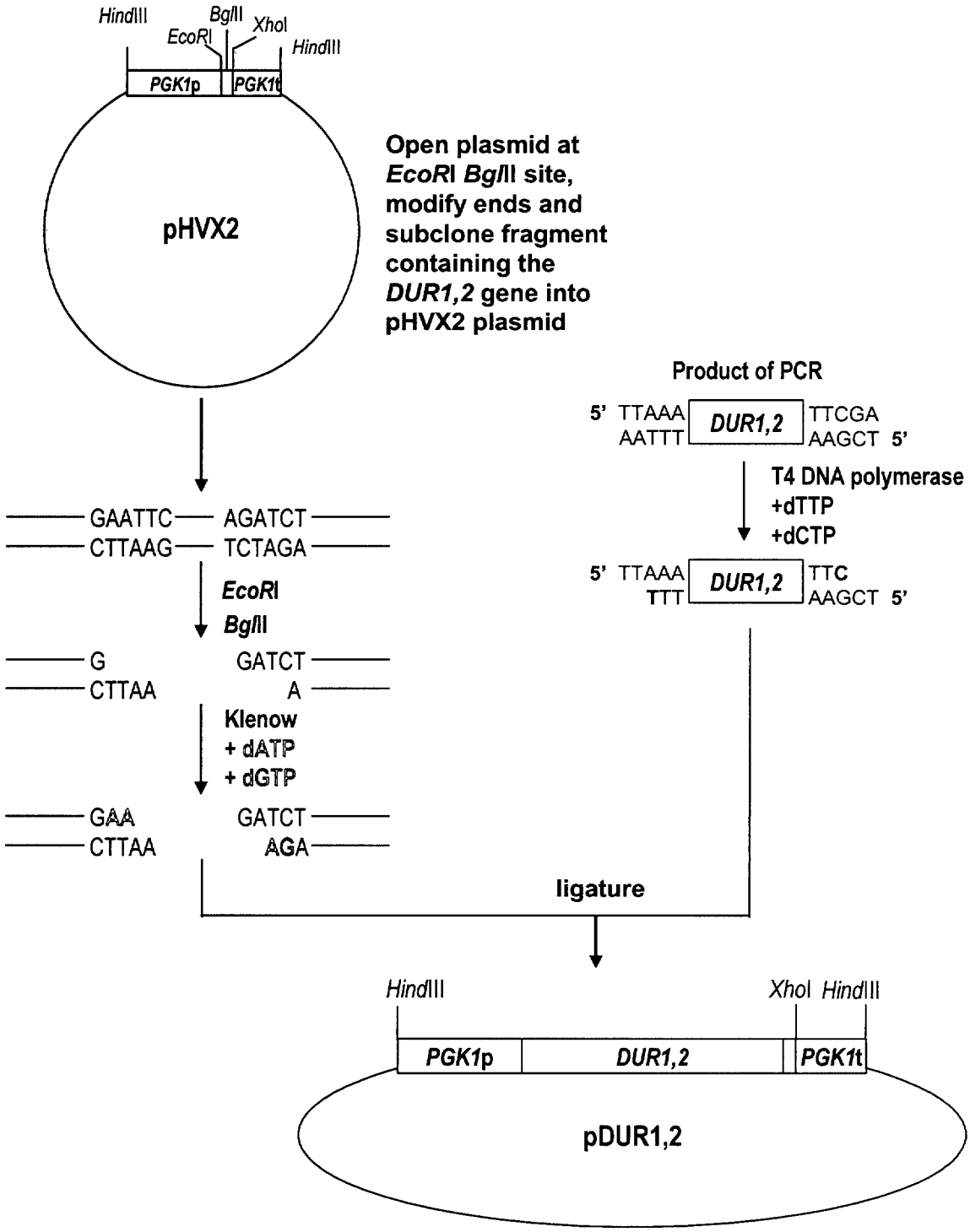


Figure 5: Cloning of the *DUR1,2* gene and construction of plasmid pDUR1,2.

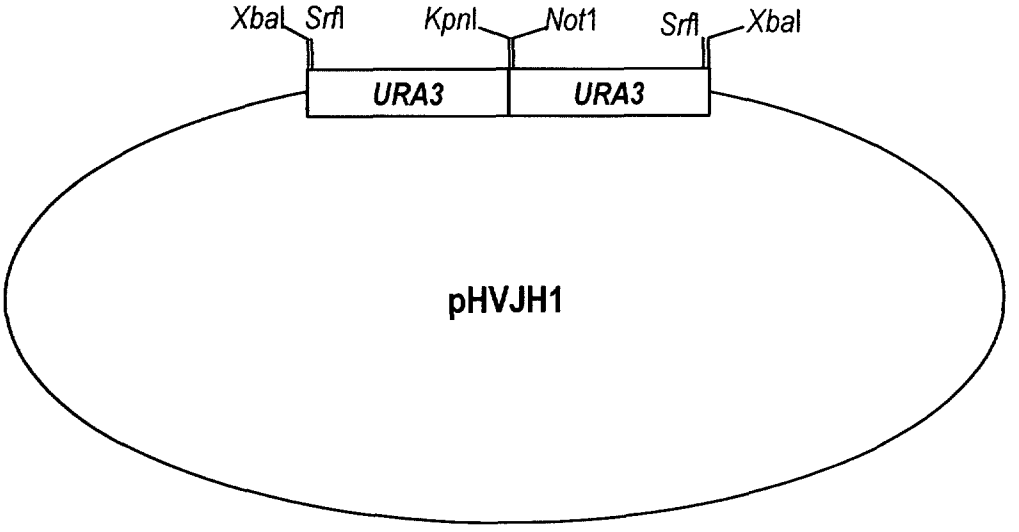


Figure 6: Plasmid pHVJH1, previously described by Volschenk *et al.*, 2004.

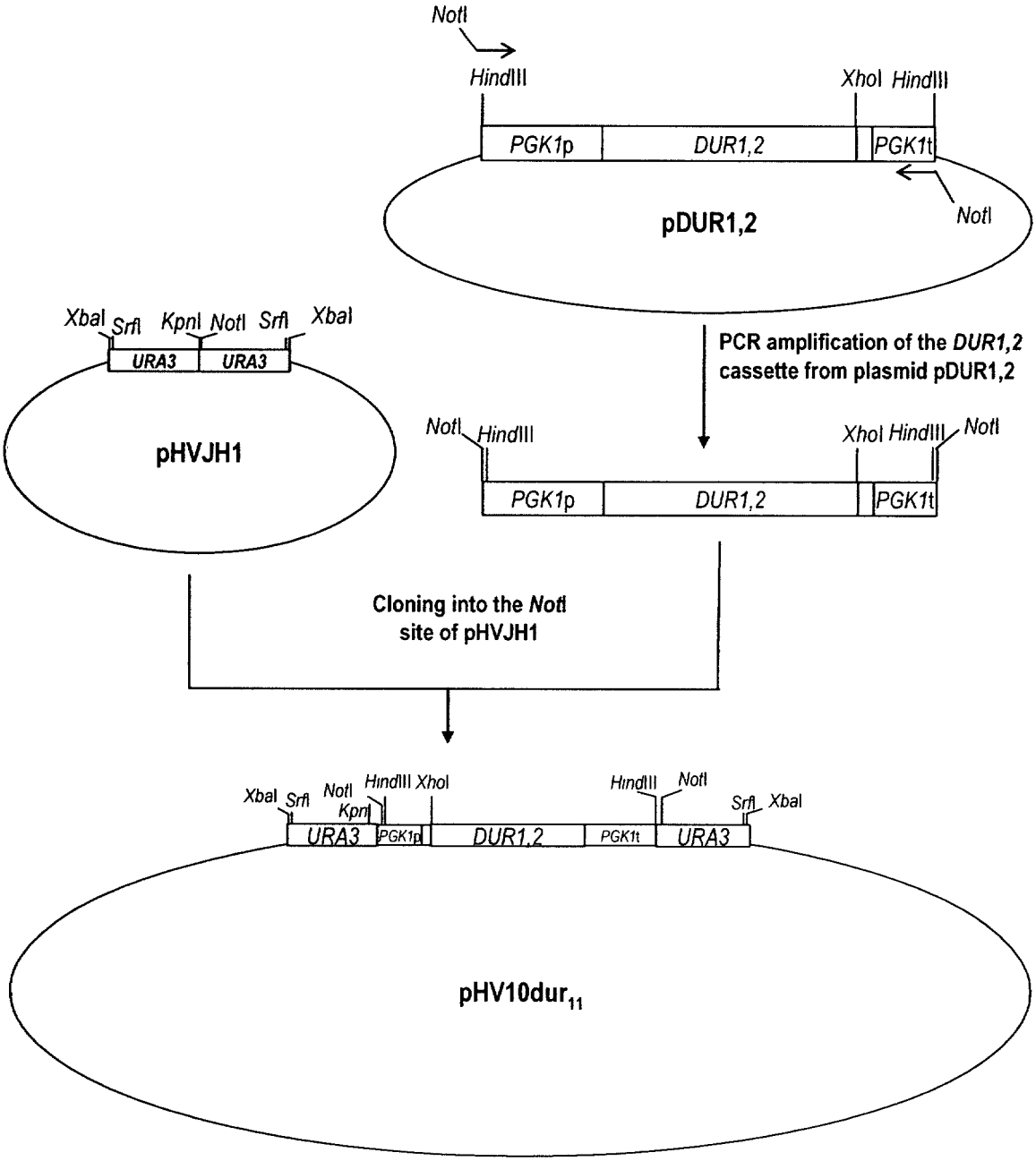


Figure 7: Construction of plasmid pHV10dur$_{11}$.

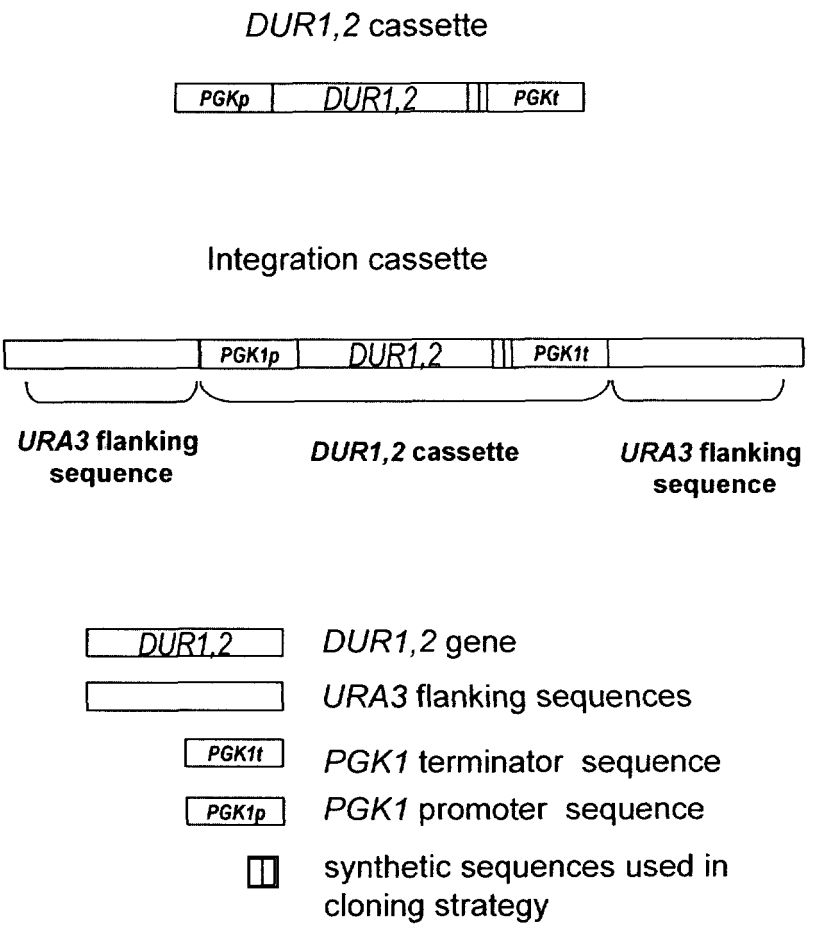


Figure 8: Schematic representation of the introduced genetic material.

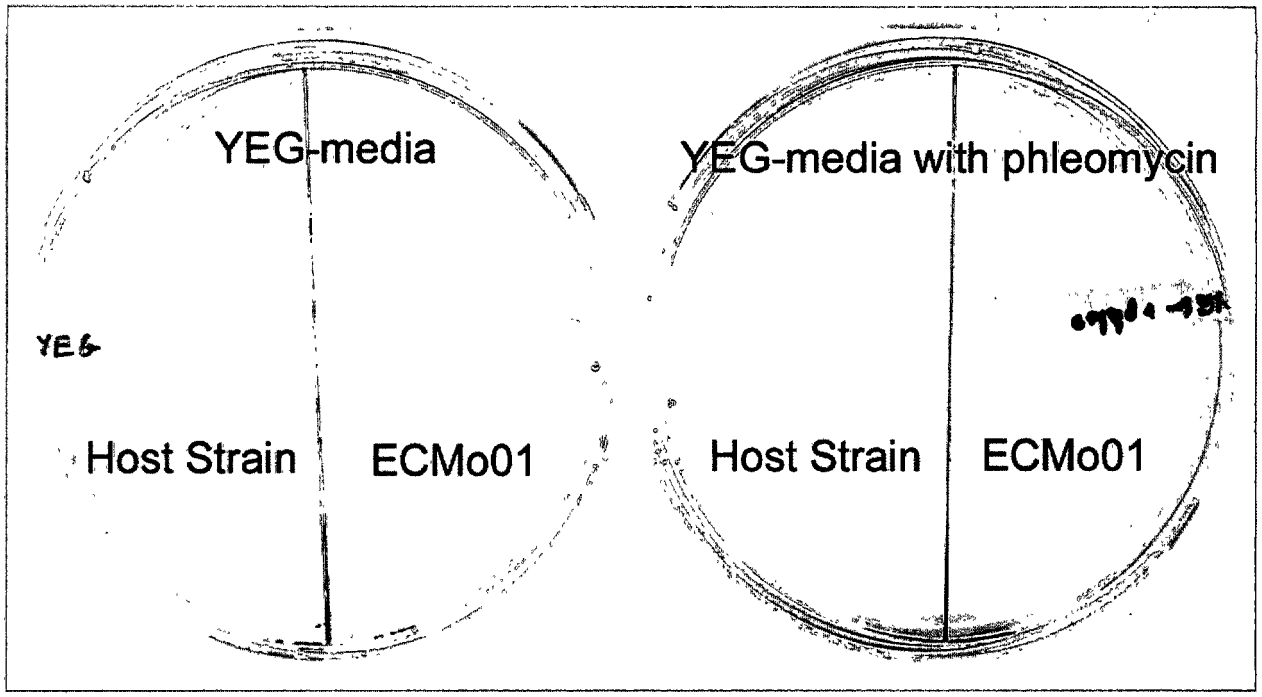


Figure 9: Loss of Phleomycin resistance in the ECMo01 yeast strain.

BEST ORIGINAL COPY

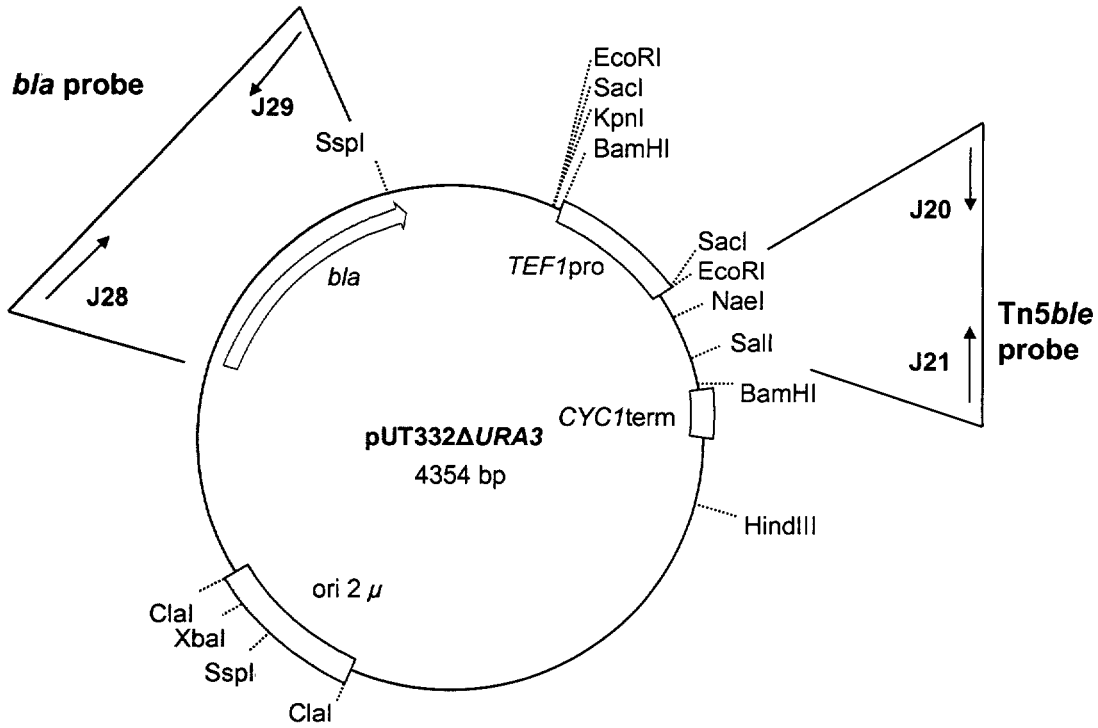


Figure 10: Plasmid pUT332Δ*URA3* map and schematic representation of the probes used in Southern experiments.

The *bla* gene probe was created by PCR amplification of the **pUT332Δ*URA3*** plasmid using primers J28 and J29. The Tn5*Ble* probe was retrieved by PCR amplification of a region of the **pUT332Δ*URA3*** plasmid using primers J20 and J21 (Table 1).

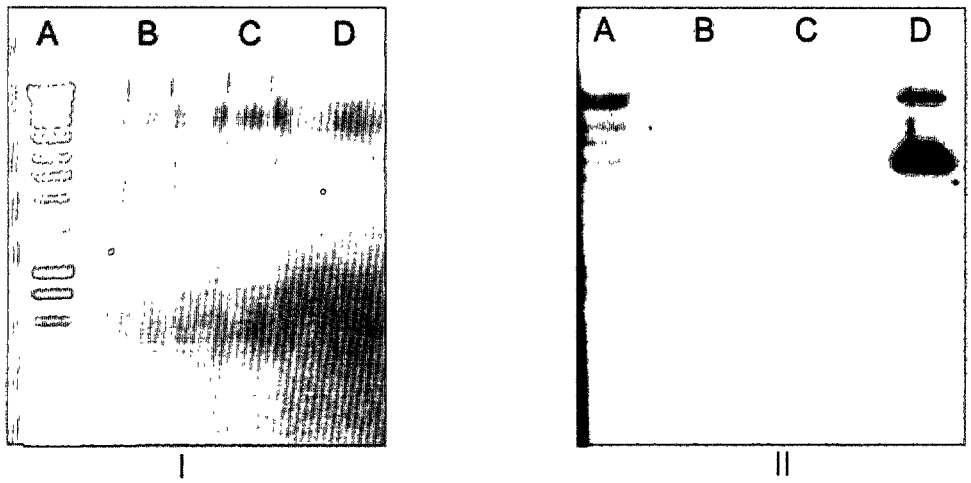


Figure 11: Southern blot analysis of the Tn5*Ble* gene that encodes for Phleomycin resistance in the ECMo01 strain.

Panel I. Agarose gel of genomic DNA digested with *Eco*RI restriction enzyme. (A) DNA size marker (B) genomic DNA of the host strain (C) genomic DNA of ECMo01 (D) plasmid DNA of **pUT332Δ*URA3*** containing the Tn5*Ble* gene. Panel II. Southern blot analysis probing for the Tn5*Ble* gene in the genome of ECMo01.

BEST ORIGINAL COPY

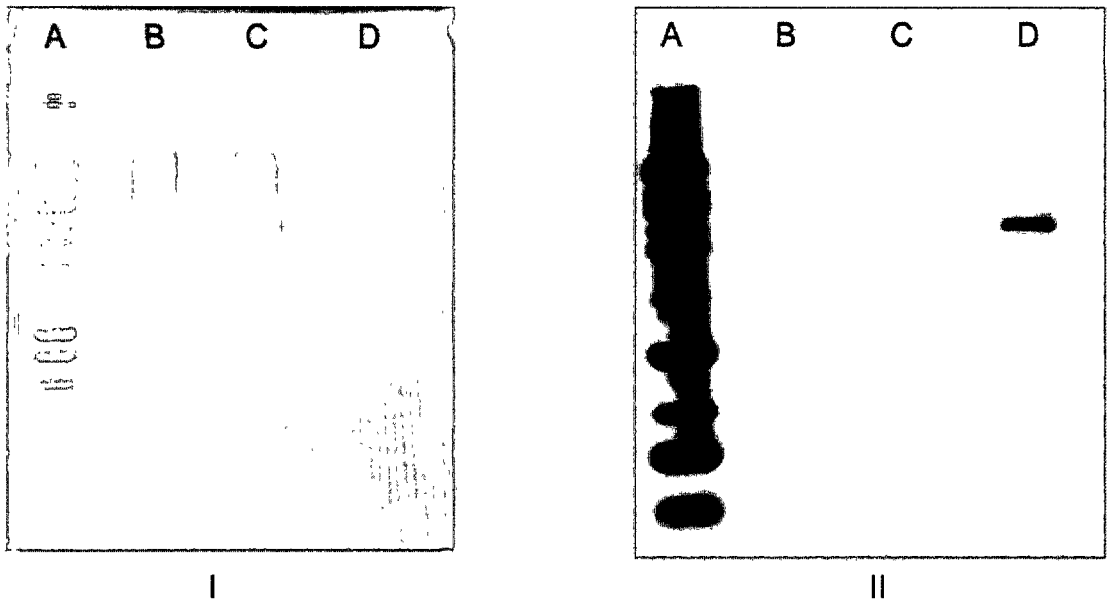


Figure 12: Southern blot analysis of the *bla* gene that encodes for Ampicillin resistance in the ECMo01 strain.

Panel I. Agarose gel of fractionated genomic DNA digested with *Eco*RI restriction enzyme. (A) DNA size marker (B) genomic DNA of the host strain (C) genomic DNA of ECMo01 (D) plasmid DNA of **pUT332Δ*URA3*** containing the *bla* gene. Panel II Southern blot analysis probing for the *bla* gene in the genome of ECMo01.

BEST ORIGINAL COPY

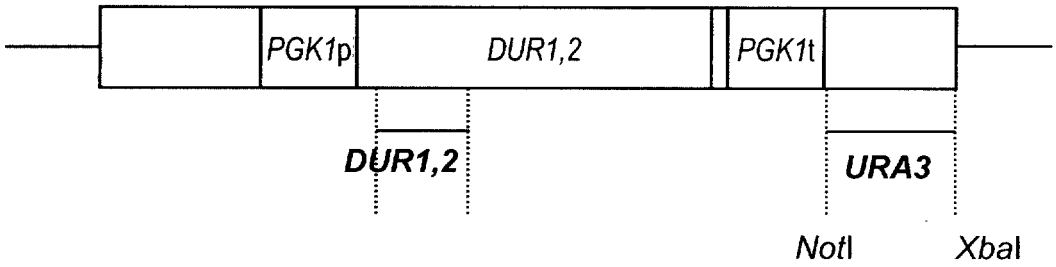


Figure 13: Schematic representation of probe hybridization to the *DUR1,2* integration cassette.

The 736 bp *DUR1,2* probe was generated by PCR using plasmid **pDUR1,2** and primers J30 and J31. A 927 bp *ura3* fragment spanning nt 430 to nt 1357 downstream from the *URA3* ATG codon was used as a probe.

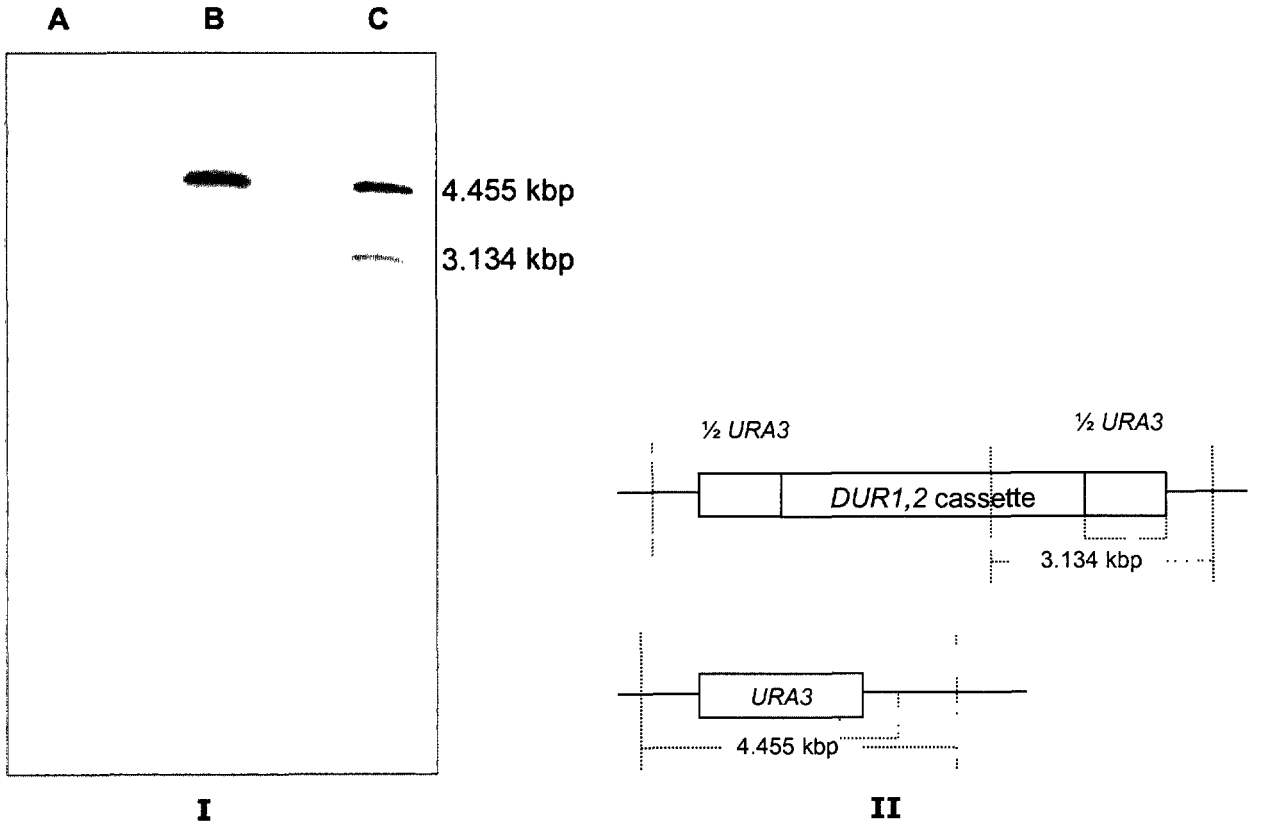


Figure 14: Southern blot analyses of the *URA3* loci in the host and ECMo01 strains.

Panel I. Southern blot analysis of genomic DNA digested with *Bgl*II restriction enzyme (A) DNA size marker (B) genomic DNA of the host strain (C) genomic DNA of ECMo01. Panel II. Schematic representation of the integrated and wild type *URA3* loci. Vertical, dotted lines indicate *Bgl*II restriction enzyme sites. Red lines indicated areas of probe hybridization to genomic DNA.

BEST ORIGINAL COPY

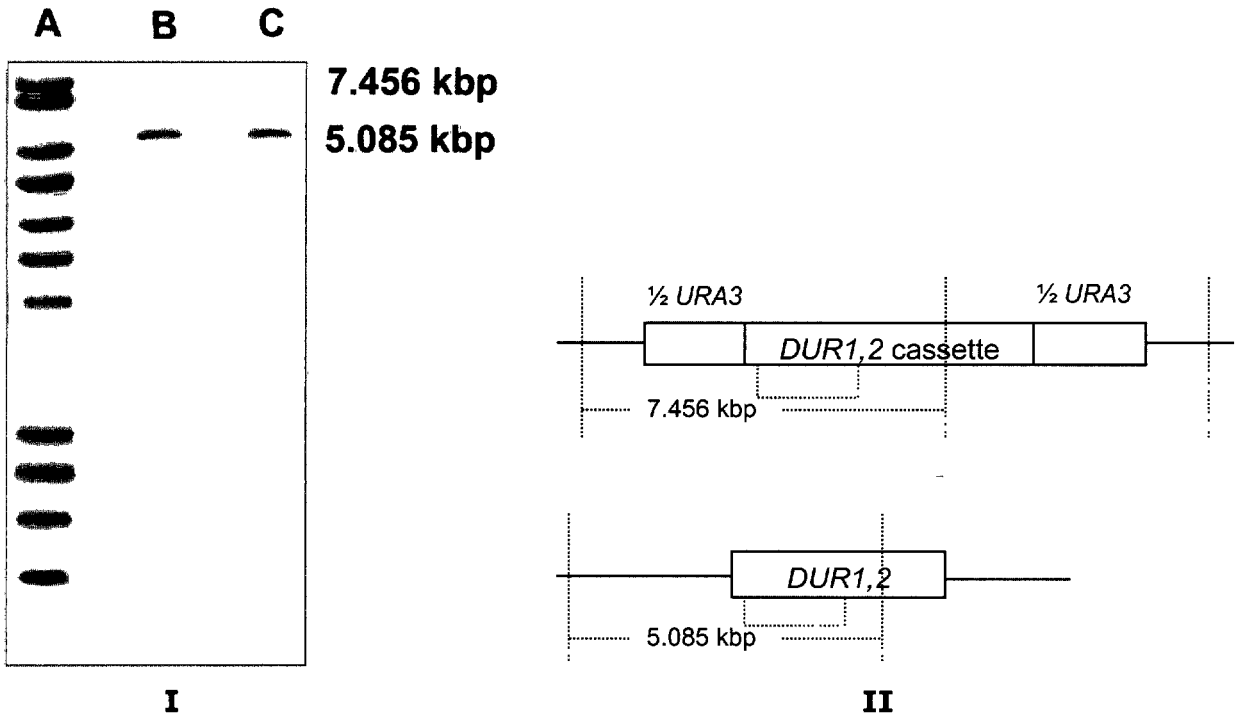


Figure 15: Southern blot analyses of the *DUR1,2* gene in the host and ECMo01 strains.

Panel I. Southern blot analysis of genomic DNA digested with *Bgl*II restriction enzyme (A) DNA size marker (B) genomic DNA of the host strain (C) genomic DNA of ECMo01. Panel II. Schematic representation of the integrated and wild type *DUR1,2* genes. Vertical, dotted lines indicate *Bgl*II restriction enzyme sites. Red lines indicated areas of probe hybridization to genomic DNA.

BEST ORIGINAL COPY

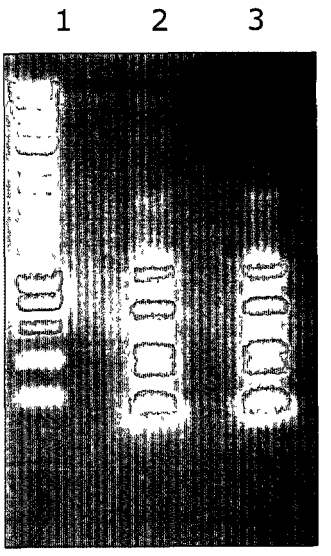


Figure 16: Genetic patterns of the Davis 522 (lane 2) and ECMo01 (lane 3) yeast strains.

PCR was performed using primers J32 and J33 (see Table 1) on genomic DNA isolated from Davis 522 and ECMo01 strains. This technique is described by Schuller *et al.* (2004).

BEST ORIGINAL COPY

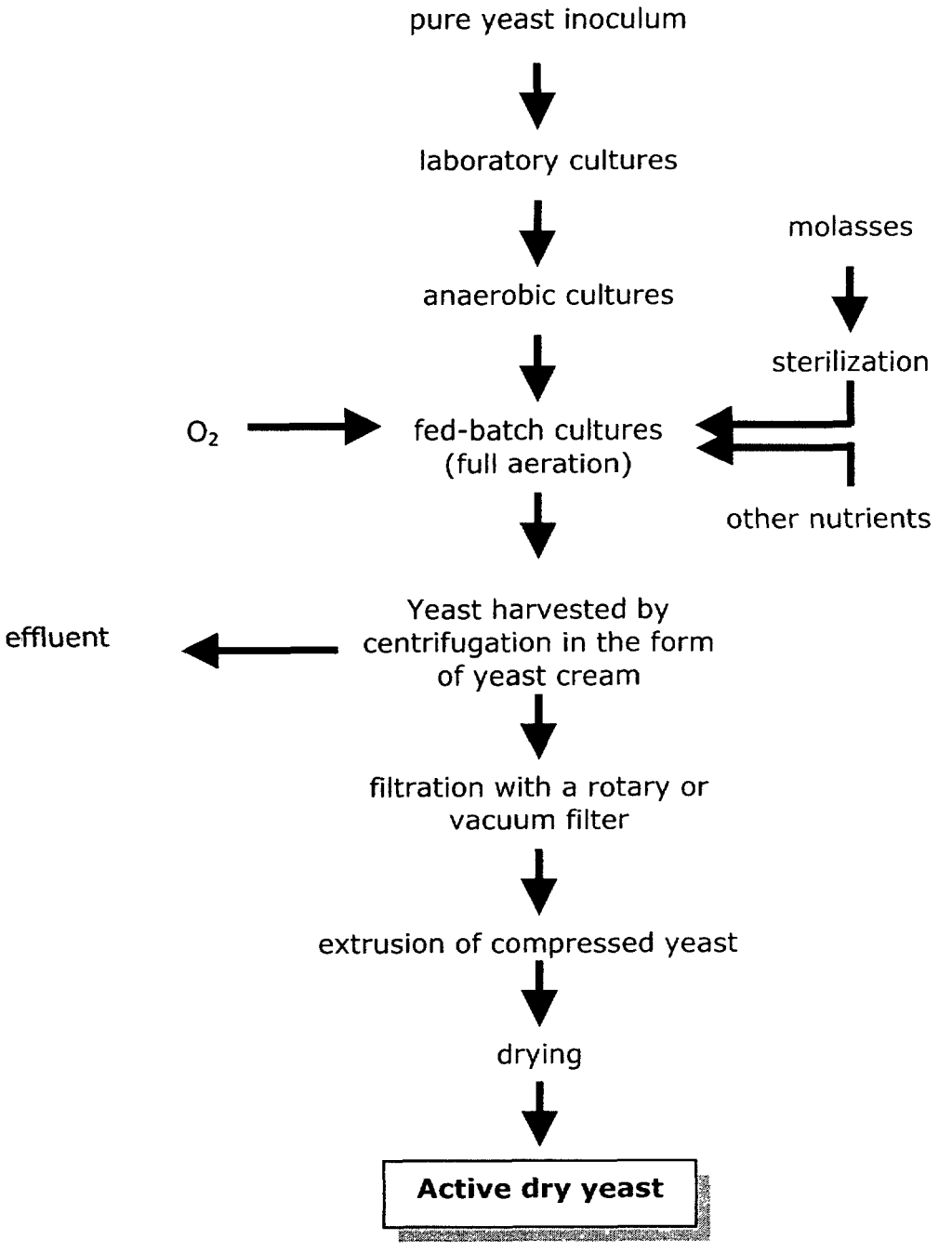


Figure 17: Flow diagram describing the manufacture of active dry yeast.

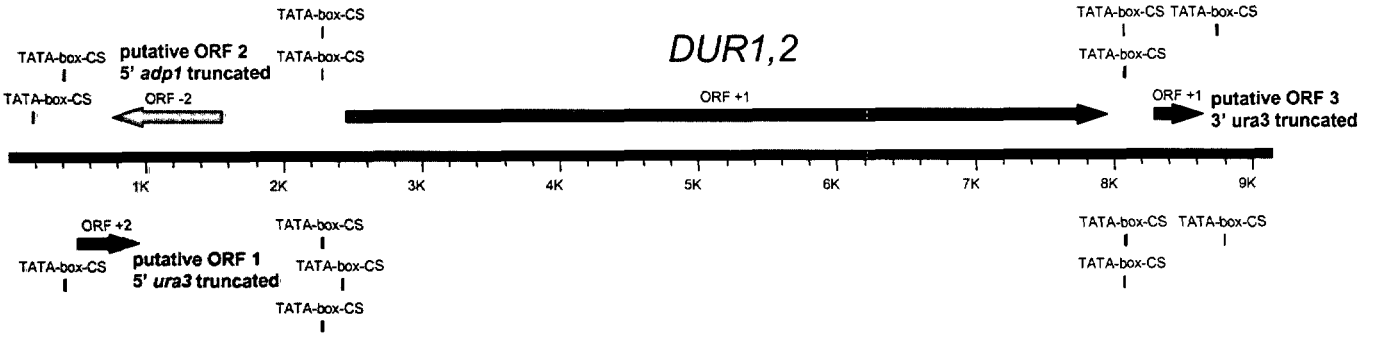
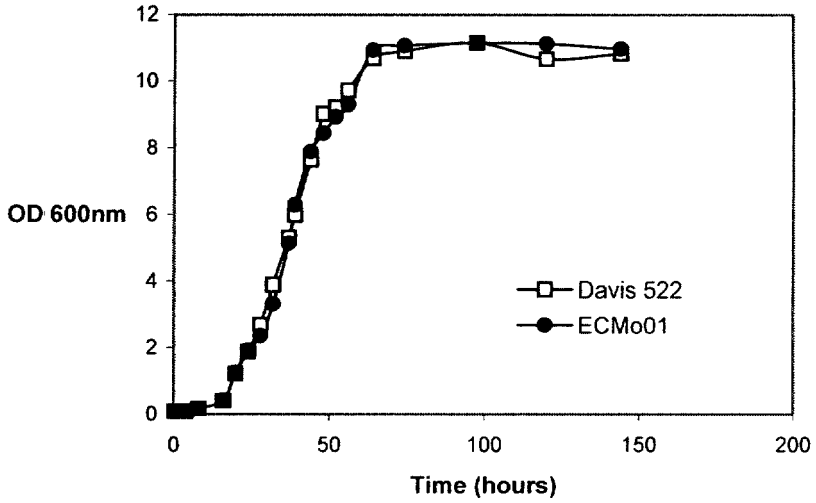


Figure 18: Putative open reading frames (ORFs) analysis. A) Schematic representation of putative ORFs and TATA boxes found within the integration cassette.

A



B

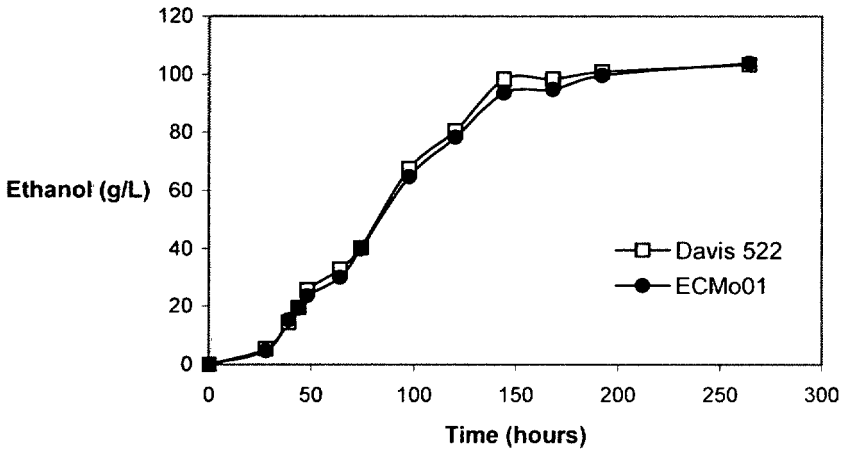


Figure 19: A) Comparison of growth of strains ECMo01 and Davis 522 in Chardonnay grape juice; B) Formation of ethanol by Davis 522 and ECMo01 when fermenting Chardonnay grape juice.

Sterile flasks capped with vapor locks containing 200 mL sterile filtered Chardonnay grape juice were inoculated to $1x10^6$ cells/mL with either ECMo01 or Davis 522. Fermentations were conducted at 20°C. Fermentations were slowly stirred before samples were taken for growth and ethanol determinations.

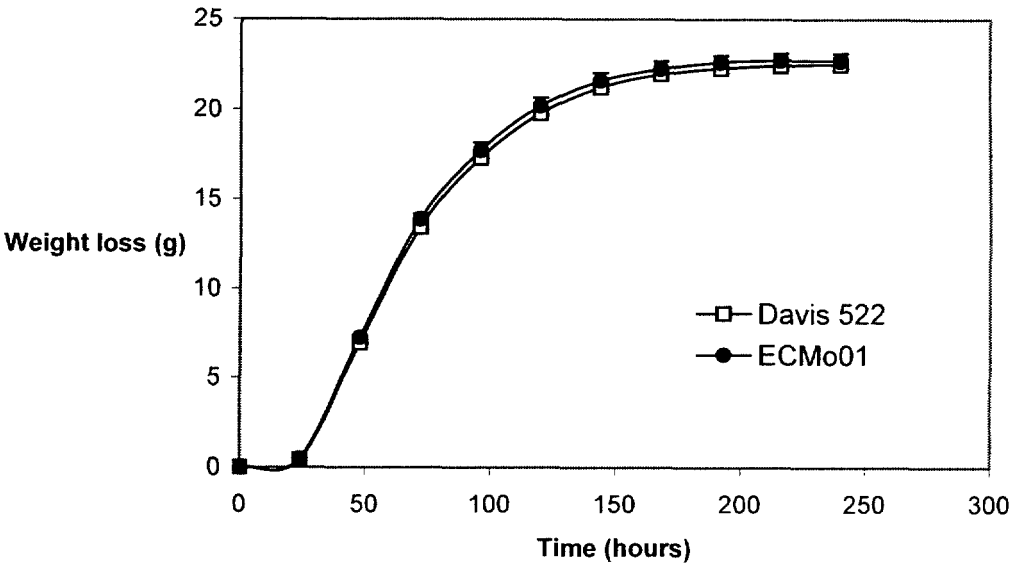


Figure 20: Evolution of carbon dioxide in fermentations of Davis 522 and ECMo01 as measured as weight loss.

Sterile flasks capped with vapor locks containing 500 mL unfiltered Chardonnay grape juice were inoculated to $2x10^6$ cells/ mL with either ECMo01 or Davis 522. Fermentations were conducted at 20 °C. Fermentations were monitored daily for CO_2 evolution.

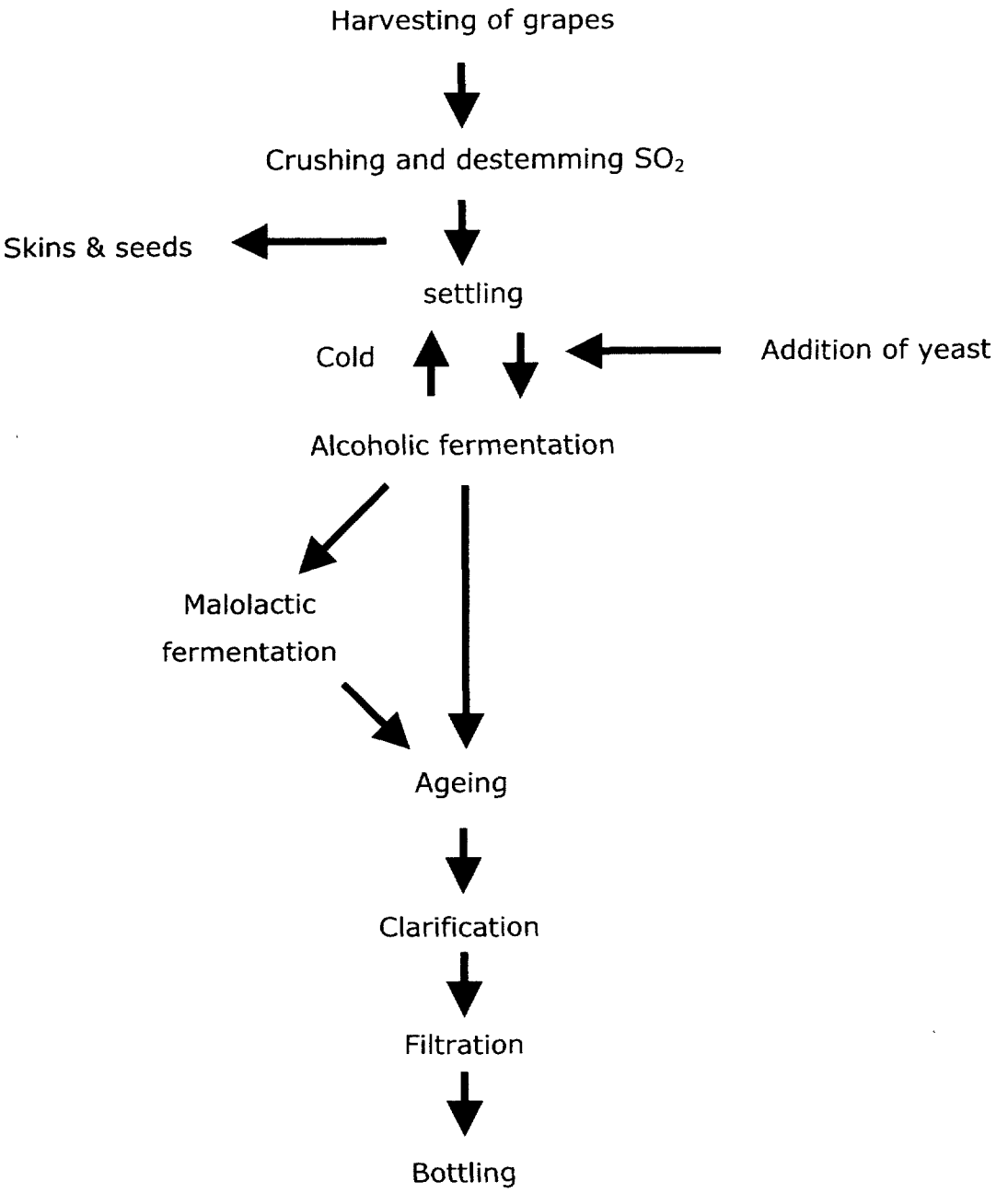


Figure 21: White winemaking practices and the use of a standard or ECMo01 yeast strain.

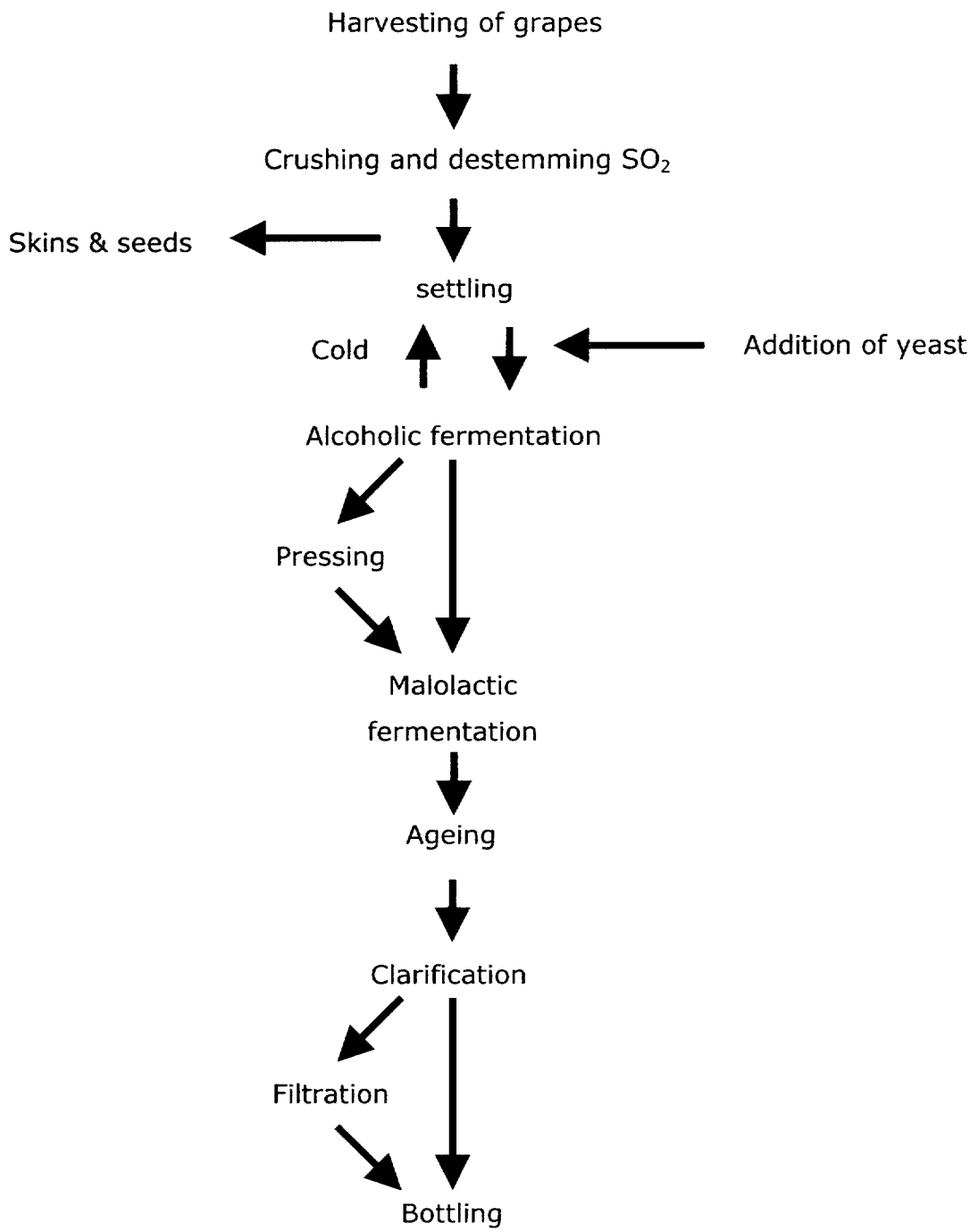


Figure 22: Red winemaking practices and the use of a standard or ECMo01 yeast strain.

Table 1: Primers and linkers used in the construction of the integration cassette

Designation	DNA sequences	Description
J9	5'GATC**TCTAGA**GCCCGGGCAACGGTTCATCATCTCATGGATCTGC3'	5'-primer for *URA3* upstream flanking sequence with ***XbaI*** and *SrfI* sites
J10	5'GATC**GGTACC**TACTTCTTCCGCCGCCTGCAAACCGCCT3'	3'-primer for *URA3* upstream flanking sequence with ***KpnI*** site
J11	5'GATC**GGTACC**GCGGCCGCACAAAGGAACCTAGAGGCCTTTTGATGTTAG3'	5'-primer for *URA3* downstream flanking sequence with ***KpnI*** and *NotI* sites
J12	5'GATC**TCTAGA**GCCCGGGCTACACCAGAGATACATAATTAGATAT3'	3'-primer for *URA3* downstream flanking sequence with ***XbaI*** and *SrfI* sites
J17	5' GAAGATAA**AGATCT**TG 3'	Used for site directed mutagenesis , introducing a ***BglII*** site at the end of *PGK1p*
J20	5' AATGACCGACCAAGCGACG 3'	Primer used to amplify the Tn5*ble* probe (Phleomycin resistance) from plasmid pUT332Δ*URA3*
J21	5' ATCCTGGGTGGTGAGCAG 3'	Primer used to amplify the Tn5*ble* probe (Phleomycin resistance) from plasmid pUT332Δ*URA3*
J22	5' AAGGAAAAAA**GCGGCCGC**AAAAGCTTTCTAACTGATCTAT 3'	5'-end primer used to amplify the *DUR1,2* expression cassette from pHV3, building in a ***NotI*** site.
J23	5' AAGGAAAAAA**GCGGCCGC**AAAAGCTTTAACGACGAGAATT 3'	3'-end primer used to amplify the *DUR1,2* expression cassette from pHV3, building in a ***NotI*** site.
J24	5' TTAAAAAA**ATG**ACAGTTAGTTCCGATACA 3'	*DUR1,2* gene – start codon
J25	5' TCGAAAAAGGTATT**TCA**TGCCAATGTTATGAC 3'	*DUR1,2* gene – stop codon
J28	5' TTGCGGCATTTTGCCTTCCT 3'	Primer used to amplify the *bla* probe (Ampicillin resistance) from plasmid pUT332Δ*URA3*
J29	5' GTTGCCGGGAAGCTAGAGTA 3'	Primer used to amplify the *bla* probe (Ampicillin resistance) from plasmid pUT332Δ*URA3*
J30	5'-TTAGACTGCGTCTCCATCTTTG-3'	Primer used to amplify *DUR1,2* probe for Southern blot analysis from plasmid pDUR1,2
J31	5'-TGCTGGCTTTACTGAAGAAGAG-3'	Primer used to amplify *DUR1,2* probe for Southern blot analysis from plasmid pDUR1,2
J32	5'-GTGGATTTTTATTCCAAC-3'	Primer used to amplify δ elements of Ty1 retro-transposons for δ sequence typing
J33	5'-TCAACAATGGAATCCCAAC-3'	Primer used to amplify δ elements of Ty1 retro-transposons for δ sequence typing

Table 2: Plasmids used for the construction and integration of the integration cassette.

Designation	Description	Source
pHVX2	YEplac181 derived plasmid containing the *PGK1* promoter and terminator sequences	Volschenk *et al.*, 1997a
pHVJH1	YEp352(Δ*Kpn*I) containing the *URA3* flanking sequences	Volschenk *et al.*, 2004
pDUR1,2	pHVX2 containing the *DUR1,2* open reading frame subcloned between the *PGK1* promoter and terminator sequences	Coulon, 2001
pHV10dur$_{11}$	pHVJH1 containing the *DUR1,2* cassette cloned between the URA3 flanking sequences	Coulon, 2001
pUT332	*Escherichia coli*, *Saccharomyces cerevisiae*. Plasmid that confers resistance to Phleomycin.	Cayla S.A.S.
pUT332Δ*URA3*	pUT332 with the *URA3* removed by *Bgl*II digestion and ligation of pUT332 backbone.	This study.

Table 3: Detailed description of the nucleotides composing the integration cassette.

Nucleotide position	Designation	Reference for cloning details	Source
1- 4	***SrfI* half cloning site**	**GRAS Notice No. 000120**	***S. cerevisiae* GC210**
1-928 4-508 509-928	***URA3* sequence** 5′ non coding sequence Part of open reading frame	**GRAS Notice No. 000120**	***S. cerevisiae* GC210**
929-934	***Kpn*I cloning site**	**GRAS Notice No. 000120**	**Synthetic**
935-942	***Not*I cloning site**	**GRAS Notice No. 000120**	**Synthetic**
943-944	**AA residues introduced with *DUR1,2* expression cassette primer**	**This study**	**Synthetic**
945-2428	***PGK1* promoter**	**Crous *et al.*, 1995**	***S. cerevisiae* AB972**
2429-2445 2429 2430-2434 2435 2436-2445	**Part of linker used in cloning strategy** A residue Remaining *Bgl*II cloning site G residue Sequences used in cloning *DUR1,2* into pHVX2	**Crous *et al.*, 1995, and this study** Crous *et al.*, 1995 Crous *et al.*, 1995 Crous *et al.*, 1995 **This study**	**Synthetic**
2446-7953 2446-2448 2449-7950 7951-7953	***DUR1,2* gene** START codon Coding region STOP codon	**This study**	***S. cerevisae* TCY1**
7954-7984	**Part of linker used in cloning strategy**	**Crous *et al.*, 1995 and this study**	**Synthetic**
7985-7970	**Sequences used in cloning *DUR1,2* into pHVX2**	**This study**	**Synthetic**
7971	**C residue**	**Crous *et al.*, 1995**	**Synthetic**
7972-7977	***XhoI* cloning site**	**Crous *et al.*, 1995**	**Synthetic**
7978	**G residue**	**Crous *et al.*, 1995**	**Synthetic**
7979-7983	**Remaining *Bgl*II cloning site**	**Crous *et al.*, 1995**	**Synthetic**
7984	**G residue**	**Crous *et al.*, 1995**	**Synthetic**
7985-8246	***PGK1* terminator**	**Crous *et al.*, 1995**	***S. cerevisiae* AB972**
8247-8248	**TT residues introduced with *DUR1,2* expression cassette primer**	**This study**	**Synthetic**
8249-8256	***NotI* cloning site**	**GRAS Notice No. 000120**	**Synthetic**
8257-9191 8257-8640 8641-9191	***URA3* sequence** Part of open reading frame 3′ non coding region	**GRAS Notice No. 000120**	***S. cerevisiae* GC210**
9188-9191	***Srf*I half cloning site**	**GRAS Notice No. 000120**	***S. cerevisiae* GC210**

Table 4: Discrepancies found between the genomic sequence of the integrated locus and previously published S288C sequences.

Comments and nucleotide positions refer to the sequence composition detailed in Table 3.

Comment	Nucleotide position	Description
1	821	Difference in the 5' region of the *URA3* open reading frame
2	929-944	Additional sequence resulting from cloning strategy
3	1623/1624	Difference in the *PGK1* promoter sequence. One base deletion between 1623 and 1624 bp.
4	2429-2446	Additional sequence resulting from cloning strategy
5	2495	Difference in the *DUR1,2* open reading frame. This difference corresponds to a change of amino acids from isoleucine (in the published sequence) to threonine (in the genomic sequence).
6	2607	Difference in the *DUR1,2* open reading frame. This difference corresponds to **no change** of the amino acid sequence.
7	2732	Difference in the *DUR1,2* open reading frame. This difference corresponds to a change of amino acids from proline (in the published sequence) to arginine (in the genomic sequence).
8 and 9	2757 and 2814	Differences in the *DUR1,2* open reading frame. These differences correspond to **no change** of the amino acid sequence.
10	3219	Difference in the *DUR1,2* open reading frame. This difference corresponds to a change of amino acids from lysine (in the published sequence) to asparagine (in the genomic sequence).
11 and 12	3441 and 3585	Differences in the *DUR1,2* open reading frame. These differences correspond to **no change** of the amino acid sequence.
13	3722	Difference in the *DUR1,2* open reading frame. This difference corresponds to a change of amino acids from isoleucine (in the published sequence) to methionine (in the genomic sequence).
14 - 24	3981, 4383, 4797, 4824, 4863, 4911, 4956, 4965, 5250, 5418, and 6408	Differences in the *DUR1,2* open reading frame. These differences correspond to **no change** of the amino acid sequence.
25	7005	Difference in the *DUR1,2* open reading frame. This difference corresponds to a change of amino acids from lysine (in the published sequence) to asparagine (in the genomic sequence).
26	7234	Difference in the *DUR1,2* open reading frame. This difference corresponds to a change of amino acids from alanine (in the published sequence) to threonine (in the genomic sequence).
27	7788	Difference in the *DUR1,2* open reading frame. This difference corresponds to **no change** of the amino acid sequence.
28	7954-7984	Additional sequence resulting from cloning strategy
29	8232/8233	Difference in the *PGK1* terminator sequence between. One base deletion between 8232 and 8233 bp.

30	8234/8235	Difference in the *PGK1* terminator sequence between. One base deletion between 8234 and 8235 bp.
31	8249-8256	Additional sequence resulting from cloning strategy
32	8259	Difference in the 3′ region of the *URA3* open reading frame
33	8742	Difference in the 3′ region of the *URA3* non coding sequence
34	8804	Difference in the 3′ region of the *URA3* non coding sequence

Table 5: Comparison of ethyl carbamate formation in wines formed using ECMo01 strain and Davis 522 strain.

Values are presented as means ± 1 S.D.

	ECMo01	Davis 522
Ethyl carbamate (µg/L)	**9.61±0.36** (N=3)	**87.85±5.82** (N=3)
Reduction of ethyl carbamate relative to Davis 522 (%)	89.0±0.36 (N=3)	

Appendix 1: Integration cassette in Fasta Format

>
GGGCAACGGTTCATCATCTCATGGATCTGCACATGAACAAACACCAGAGT
CAAACGACGTTGAAATTGAGGCTACTGCGCCAATTGATGACAATACAGAC
GATGATAACAAACCGAAGTTATCTGATGTAGAAAAGGATTAAAGATGCTA
AGAGATAGTGATGATATTTCATAAATAATGTAATTCTATATATGTTAATT
ACCTTTTTTGCGAGGCATATTTATGGTGAAGGATAAGTTTTGACCATCAA
AGAAGGTTAATGTGGCTGTGGTTTCAGGGTCCATAAAGCTTTTCAATTCA
TCTTTTTTTTTTTTGTTCTTTTTTTTGATTCCGGTTTCTTTGAAATTTTT
TTGATTCGGTAATCTCCGAGCAGAAGGAAGAACGAAGGAAGGAGCACAGA
CTTAGATTGGTATATATACGCATATGTGGTGTTGAAGAAACATGAAATTG
CCCAGTATTCTTAACCCAACTGCACAGAACAAAAACCTGCAGGAAACGAA
GATAAATCATGTCGAAAGCTACATATAAGGAACGTGCTGCTACTCATCCT
AGTCCTGTTGCTGCCAAGCTATTTAATATCATGCACGAAAAGCAAACAAA
CTTGTGTGCTTCATTGGATGTTCGTACCACCAAGGAATTACTGGAGTTAG
TTGAAGCATTAGGTCCCAAAATTTGTTTACTAAAAACACATGTGGATATC
TTGACTGATTTTTCCATGGAGGGCACAGTTAAGCCGCTAAAGGCATTATC
CGCCAAGTACAATTTTTTACTCTTCGAAGACAGAAAATTTGCTGACATTG
GTAATACAGTCAAATTGCAGCACTCTGCGGGTGTATACAGAATAGCAGAA
TGGGCAGACATTACGAATGCACACGGTGTGGTGGGCCCAGGTATTGTTAG
CGGTTTGAAGCAGGCGGCGGAAGAAGTAGGTACCGCGGCCGCAAAAGCTT
TCTAACTGATCTATCCAAAACTGAAAATTACATTCTTGATTAGGTTTATC
ACAGGCAAATGTAATTTGTGGTATTTTGCCGTTCAAAATCTGTAGAATTT
TCTCATTGGTCACATTACAACCTGAAAATACTTTATCTACAATCATACCA
TTCTTATAACATGTCCCCTTAATACTAGGATCAGGCATGAACGCATCACA
GACAAAATCTTCTTGACAAACGTCACAATTGATCCCTCCCCATCCGTTAT
CACAATGACAGGTGTCATTTTGTGCTCTTATGGGACGATCCTTATTACCG
CTTTCATCCGGTGATAGACCGCCACAGAGGGGCAGAGAGCAATCATCACC
TGCAAACCCTTCTATACACTCACATCTACCAGTGTACGAATTGCATTCAG
AAAACTGTTTGCATTCAAAAATAGGTAGCATACAATTAAAACATGGCGGG
CATGTATCATTGCCCTTATCTTGTGCAGTTAGACGCGAATTTTTCGAAGA
AGTACCTTCAAAGAATGGGGTCTTATCTTGTTTTGCAAGTACCACTGAGC
AGGATAATAATAGAAATGATAATATACTATAGTAGAGATAACGTCGATGA
CTTCCCATACTGTAATTGCTTTTAGTTGTGTATTTTTAGTGTGCAAGTTT
CTGTAAATCGATTAATTTTTTTTCTTTCCTCTTTTTATTAACCTTAATTT
TTATTTTAGATTCCTGACTTCAACTCAAGACGCACAGATATTATAACATC
TGCATAATAGGCATTTGCAAGAATTACTCGTGAGTAAGGAAAGAGTGAGG
AACTATCGCATACCTGCATTTAAAGATGCCGATTTGGGCGCGAATCCTTT
ATTTTGGCTTCACCCTCATACTATTATCAGGGCCAGAAAAAGGAAGTGTT
TCCCTCCTTCTTGAATTGATGTTACCCTCATAAAGCACGTGGCCTCTTAT
CGAGAAAGAAATTACCGTCGCTCGTGATTTGTTTGCAAAAAGAACAAAAC
TGAAAAAACCCAGACACGCTCGACTTCCTGTCTTCCTATTGATTGCAGCT
TCCAATTTCGTCACACAACAAGGTCCTAGCGACGGCTCACAGGTTTTGTA
ACAAGCAATCGAAGGTTCTGGAATGGCGGGAAAGGGTTTAGTACCACATG
CTATGATGCCCACTGTGATCTCCAGAGCAAAGTTCGTTCGATCGTACTGT
TACTCTCTCTCTTTCAAACAGAATTGTCCGAATCGTGTGACAACAACAGC
CTGTTCTCACACACTCTTTTCTTCTAACCAAGGGGGTGGTTTAGTTTAGT
AGAACCTCGTGAAACTTACATTTACATATATATAAACTTGCATAAATTGG
TCAATGCAAGAAATACATATTTGGTCTTTTCTAATTCGTAGTTTTTCAAG
TTCTTAGATGCTTTCTTTTTCTCTTTTTTACAGATCATCAAGGAAGTAAT
TATCTACTTTTTACAACAAATATAAAACAAGATCGGAATTAAAAAATGAC
AGTTAGTTCCGATACAACTGCTGAAATATCGTTAGGTTGGTCAACCCAAG
ACTGGATTGATTTCCACAAGTCATCAAGCTCCCAGGCTTCACTAAGGCTT
CTTGAATCACTACTAGACTCTCAAAATGTTGCGCCAGTCGATAATGCGTG
GATATCACTAATTTCAAAGGAAAATTTACTGCACCAATTCCAAATTTTAA
AGAGCAGAGAAAATAAAGAAACTCTACCTCTCTACGGTGTCCCTATTGCT
GTTAAGGACAACATCGACGTTAGAGGTCTACGCACCACCGCTGCATGTCC
ATCCTTCGCATATGAGCCTTCCAAAGACTCTAAAGTAGTAGAACTACTAA

GAAATGCAGGTGCAATAATCGTGGGTAAGACAAACTTGGACCAATTTGCC
ACAGGATTAGTCGGCACACGGTCTCCATATGGGAAAACACCTTGCGCTTT
TAGCAAAGAGCATGTATCTGGTGGTTCCTCCGCTGGGTCAGCATCGGTGG
TCGCCAGAGGTATCGTACCAATTGCATTGGGTACTGATACAGCAGGTTCT
GGTAGAGTCCCAGCCGCCTTGAACAACCTGATTGGCCTAAAGCCAACAAA
GGGCGTCTTTTCCTGTCAAGGTGTAGTTCCCGCTTGTAAATCTTTAGACT
GCGTCTCCATCTTTGCATTAAACCTAAGTGATGCTGAACGCTGCTTCCGC
ATCATGTGCCAGCCAGATCCTGATAATGATGAATATTCTAGACCCTATGT
TTCCAACCCTTTGAAAAATTTTTCAAGCAATGTAACGATTGCTATTCCTA
AAAATATCCCATGGTATGGTGAAACCAAGAATCCTGTACTGTTTTCCAAT
GCTGTCGAAAATCTATCAAGAACGGGCGCTAACGTCATAGAAATTGATTT
TGAGCCTCTTTTAGAGTTAGCTCGCTGTTTATACGAAGGTACTTGGGTGG
CCGAGCGTTATCAAGCTATTCAATCGTTTTTGGACAGTAAGCCACCAAAG
GAATCTTTGGACCCTACTGTTATTTCAATTATAGAAGGGGCCAAGAAATA
CAGTGCAGTAGACTGCTTCAGTTTTGAATACAAAAGACAAGGCATCTTGC
AAAAAGTGAGACGACTTCTCGAATCAGTCGATGTCTTGTGTGTGCCCACA
TGTCCTTTAAATCCTACTATGCAACAAGTTGCGGATGAACCAGTCCTAGT
CAATTCAAGACAAGGCACATGGACTAATTTTGTCAACTTGGCAGATTTGG
CAGCCCTTGCTGTTCCCGCAGGGTTCCGAGACGATGGTTTGCCAAATGGT
ATTACTTTAATCGGTAAAAAATTCACAGATTACGCACTATTAGAGTTGGC
TAACCGCTATTTCCAAAATATGTTCCCCAACGGTTCCAGAACATACGGTA
CTTTTACCTCTTCTTCAGTAAAGCCAGCAAACGATCAATTAGTGGGACCA
GACTATGACCCATCTACGTCCATAAAATTGGCTGTTGTCGGTGCACATCT
TAAGGGTCTGCCTCTACATTGGCAATTGGAGAAGGTCAATGCAACATATT
TATGTACAACAAAAACATCAAAAGCTTACCAGCTTTTTGCTTTGCCCAAA
AATGGACCAGTTTTAAAACCTGGTTTGAGAAGAGTTCAAGATAGCAATGG
CTCTCAAATCGAATTAGAAGTGTACAGTGTTCCAAAAGAACTGTTCGGTG
CTTTTATTTCCATGGTTCCTGAACCATTAGGAATAGGTTCAGTGGAGTTA
GAATCTGGTGAATGGATCAAATCCTTTATTTGTGAAGAATCTGGTTACAA
AGCCAAAGGTACAGTTGATATCACAAAGTATGGTGGATTTAGAGCATATT
TTGAAATGTTGAAGAAAAAAGAGTCCCAAAAGAAGAAGTTATTTGATACC
GTGTTAATTGCCAATAGAGGTGAAATTGCCGTCCGTATTATCAAGACATT
AAAAAAATTGGGTATTAGATCAGTTGCAGTTTATTCCGACCCTGATAAAT
ATTCTCAACACGTTACTGATGCAGATGTTTCTGTACCCCTTCATGGCACA
ACCGCAGCCCAAACTTATTTAGACATGAATAAGATCATAGATGCCGCTAA
GCAAACTAATGCACAGGCCATTATTCCTGGTTATGGTTTCTTGTCGGAAA
ATGCGGATTTTTCTGATGCGTGCACCAGTGCTGGCATTACCTTTGTTGGT
CCTTCGGGAGATATTATCAGAGGTTTAGGGTTAAAACATTCTGCTAGACA
GATTGCACAGAAGGCTGGCGTTCCTCTAGTGCCAGGCTCTTTGCTTATCA
CATCAGTTGAAGAGGCTAAGAAAGTCGCAGCGGAATTGGAATACCCGGTT
ATGGTGAAGTCAACTGCTGGTGGTGGTGGTATTGGTTTGCAGAAAGTCGA
TTCTGAAGAGGATATCGAGCATATTTTTGAGACTGTGAAACATCAAGGTG
AAACATTTTTTGGTGACGCTGGTGTATTTCTGGAACGGTTTATCGAAAAT
GCCAGACATGTTGAGGTCCAACTTATGGGAGATGGTTTTGGTAAGGCCAT
TGCTTTGGGCGAACGTGATTGTTCTTTACAGCGTCGTAACCAAAAAGTTA
TCGAAGAAACTCCTGCACCAAATTTGCCAGAAAAGACGAGGTTGGCGTTA
AGAAAGGCAGCTGAAAGTTTGGGATCTTTATTGAATTACAAGTGTGCTGG
TACGGTTGAATTTATTTACGATGAGAAAAAGGACGAGTTTTACTTTTTAG
AAGTTAATACAAGATTACAAGTTGAACATCCAATAACAGAAATGGTTACG
GGGTTAGACTTGGTCGAGTGGATGATCAGGATTGCCGCTAATGATGCACC
TGATTTTGATTCTACAAAGGTAGAAGTCAATGGGGTTTCAATGGAGGCAC
GTTTATATGCTGAAAATCCATTGAAAAATTTCAGACCTTCTCCAGGTTTA
CTTGTCGATGTGAAATTCCCTGATTGGGCAAGAGTGGATACTTGGGTTAA
GAAAGGTACTAATATTTCTCCCGAATATGATCCAACATTGGCCAAAATTA
TCGTTCATGGGAAAGACCGTGATGATGCAATTTCCAAGTTAAATCAAGCG
TTAGAAGAAACAAAAGTTTACGGATGTATTACTAACATTGACTACCTGAA
GTCTATCATTACCAGTGATTTCTTTGCTAAAGCAAAAGTTTCTACAAACA

TTTTGAACTCTTATCAATATGAGCCTACCGCCATCGAAATTACTTTGCCC
GGTGCACACACTAGTATTCAGGATTACCCCGGTAGAGTTGGGTACTGGAG
AATTGGTGTTCCGCCCTCTGGTCCAATGGACGCATATTCGTTTAGATTGG
CGAACAGAATTGTTGGTAATGACTACAGGACTCCTGCCATTGAAGTAACG
TTGACTGGTCCATCCATCGTTTTCCATTGTGAAACTGTCATTGCCATTAC
TGGTGGTACCGCTCTATGTACATTAGACGGCCAAGAAATTCCCCAACACA
AACCGGTCGAAGTTAAGAGGGGATCTACTTTATCCATTGGCAAGTTGACA
AGCGGCTGTAGAGCATACTTAGGTATCAGGGGTGGCATTGATGTGCCTAA
ATACTTGGGCTCTTATTCTACTTTCACTCTAGGAAATGTCGGTGGATACA
ATGGAAGGGTGCTAAAACTTGGAGACGTACTATTCTTACCAAGCAATGAA
GAAAATAAATCAGTTGAGTGCCTTCCACAGAATATTCCTCAATCATTAAT
TCCTCAAATTTCCGAAACTAAGGAATGGAGAATTGGTGTAACATGTGGTC
CCCATGGGTCTCCAGATTTTTTTAAACCTGAGTCCATCGAAGAATTTTTC
AGTGAGAAGTGGAAGGTTCATTACAACTCCAATAGATTTGGTGTCCGTTT
GATTGGACCTAAACCTAAGTGGGCAAGAAGTAATGGTGGTGAAGGTGGTA
TGCATCCCTCAAACACTCACGATTACGTTTATTCTCTGGGTGCAATTAAT
TTCACGGGTGATGAGCCAGTTATTATTACTTGCGATGGTCCTTCCTTAGG
TGGTTTTGTGTGTCAAGCTGTTGTCCCAGAAGCAGAACTGTGGAAGGTTG
GACAGGTTAAACCCGGTGATTCCATTCAGTTTGTGCCACTTTCTTACGAA
AGCTCGAGATCCTTAAAGGAATCTCAGGATGTTGCAATTAAATCATTGGA
TGGTACTAAGTTAAGGCGCTTAGACTCTGTTTCAATTTTACCATCATTCG
AAACGCCTATTCTTGCACAAATGGAAAAAGTGAATGAGCTTTCACCAAAG
GTTGTATACAGACAAGCAGGTGATCGTTATGTTTTGGTGGAATACGGTGA
TAATGAAATGAATTTTAATATTTCCTATAGAATTGAATGCCTGATCTCCC
TTGTGAAAAAGAATAAGACTATTGGTATTGTTGAAATGTCCCAAGGTGTT
AGATCTGTATTGATAGAATTTGATGGTTACAAAGTCACTCAAAAAGAATT
GCTTAAAGTATTGGTGGCATATGAAACAGAAATCCAGTTTGATGAAAATT
GGAACATAACTTCTAATATAATAAGATTACCGATGGCTTTCGAAGACTCG
AAGACTTTGGCATGTGTTCAAAGGTATCAAGAAACAATTCGTTCGTCTGC
TCCATGGTTGCCAAATAACGTTGATTTCATTGCCAATGTAAATGGAATTT
CAAGGAATGAAGTTTATGATATGTTGTATTCTGCCAGATTTATGGTTTTA
GGTTTAGGTGATGTCTTCCTAGGGTCGCCTTGTACTGTTCCATTAGATCC
TCGTCACAGATTTTTGGGAAGCAAGTACAACCCAAGTAGAACATATACAG
AAAGAGGTGCAGTCGGTATTGGCGGTATGTATATGTGCATATATGCTGCT
AACAGTCCTGGTGGGTACCAATTAGTGGGTAGAACAATACCAATTTGGGA
CAAACTATGTCTGGCCGCATCTTCTGAGGTTCCGTGGTTGATGAACCCAT
TTGACCAAGTCGAATTTTACCCAGTTTCTGAAGAAGATTTGGATAAAATG
ACTGAAGATTGTGATAATGGTGTTTATAAAGTCAATATCGAAAAGAGTGT
TTTTGATCATCAAGAATACTTGAGATGGATCAACGCAAACAAAGATTCCA
TCACAGCATTCCAGGAGGGCCAGCTTGGTGAAAGAGCAGAGGAATTTGCC
AAATTGATTCAAAATGCAAACTCTGAACTAAAAGAAAGTGTCACAGTCAA
ACCTGACGAGGAAGAAGACTTCCCAGAAGGTGCAGAAATTGTATATTCTG
AGTATTCTGGGCGTTTTTGGAAATCCATAGCATCTGTCGGAGATGTTATT
GAAGCAGGTCAAGGGCTACTAATTATTGAAGCCATGAAAGCGGAAATGAT
TATATCCGCTCCTAAATCGGGTAAGATTATCAAGATTTGCCATGGCAATG
GTGATATGGTTGATTCTGGTGACATAGTGGCCGTCATAGAGACATTGGCA
TGAAATACCTTTTTCGATCTCCTCGAGGGATCTGCGATAGATCAATTTTT
TTCTTTTCTCTTTCCCCATCCTTTACGCTAAAATAATAGTTTATTTTATT
TTTTGAATATTTTTTATTTATATACGTATATATAGACTATTATTTATCTT
TTAATGATTATTAAGATTTTTATTAAAAAAAAATTCGCTCCTCTTTTAAT
GCCTTTATGCAGTTTTTTTTTCCCATTCGATATTTCTATGTTCGGGTTCA
GCGTATTTTAAGTTTAATAACTCGAAAATTCTCGTCGTTAAAGCTTTTGC
GGCCGCACGAAGGAACCTAGAGGCCTTTTGATGTTAGCAGAATTGTCATG
CAAGGGCTCCCTAGCTACTGGAGAATATACTAAGGGTACTGTTGACATTG
CGAAGAGCGACAAAGATTTTGTTATCGGCTTTATTGCTCAAAGAGACATG
GGTGGAAGAGATGAAGGTTACGATTGGTTGATTATGACACCCGGTGTGGG
TTTAGATGACAAGGGAGACGCATTGGGTCAACAGTATAGAACCGTGGATG

ATGTGGTCTCTACAGGATCTGACATTATTATTGTTGGAAGAGGACTATTT
GCAAAGGGAAGGGATGCTAAGGTAGAGGGTGAACGTTACAGAAAAGCAGG
CTGGGAAGCATATTTGAGAAGATGCGGCCAGCAAAACTAAAAAACTGTAT
TATAAGTAAATGCATGTATACTAAACTCACAAATTAGAGCTTCAATTTAA
TTATATCAGTTATTACCCGGGAATCTCGGTCGTAATGATTTTTATAATGA
CGAAAAAAAAAAAAATTGGAAAGAAAAAGCTTCATGGCCTTTATAAAAAGG
AACCATCCAATACCTCGCCAGAACCAAGTAACAGTATTTTACGGGGCACA
AATCAAGAACAATAAGACAGGACTGTAAAGATGGACGCATTGAACTCCAA
AGAACAACAAGAGTTCCAAAAAGTAGTGGAACAAAAGCAAATGAAGGATT
TCATGCGTTTGTACTCTAATCTGGTAGAAAGATGTTTCACAGACTGTGTC
AATGACTTCACAACATCAAAGCTAACCAATAAGGAACAAACATGCATCAT
GAAGTGCTCAGAAAAGTTCTTGAAGCATAGCGAACGTGTAGGGCAGCGTT
TCCAAGAACAAAACGCTGCCTTGGGACAAGGCTTGGGCCGATAAGGTGTA
CTGGCGTATATATATCTAATTATGTATCTCTGGTGTAGCCC

Appendix 2: Optimal alignment of DNA sequences: ECMo01 Integration Cassette and Published *URA3*, *DUR1,2*, *PGK1* promoter and terminator sequences

Gap_Open_Penalty=10.0 Gap_Extend_Penalty=5.0

Upper line: ECMo01 Integration Cassette, from 1 to 9191
Lower line: Published sequences, from 1 to 9120

ECMo01 Integration Cassette: Published sequences identity= 99%

```
1     GGGCAACGGTTCATCATCTCATGGATCTGCACATGAACAAACACCAGAGTCAAACGACGT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1     GGGCAACGGTTCATCATCTCATGGATCTGCACATGAACAAACACCAGAGTCAAACGACGT

61    TGAAATTGAGGCTACTGCGCCAATTGATGACAATACAGACGATGATAACAAACCGAAGTT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
61    TGAAATTGAGGCTACTGCGCCAATTGATGACAATACAGACGATGATAACAAACCGAAGTT

121   ATCTGATGTAGAAAAGGATTAAAGATGCTAAGAGATAGTGATGATATTTCATAAATAATG
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
121   ATCTGATGTAGAAAAGGATTAAAGATGCTAAGAGATAGTGATGATATTTCATAAATAATG

181   TAATTCTATATATGTTAATTACCTTTTTTGCGAGGCATATTTATGGTGAAGGATAAGTTT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
181   TAATTCTATATATGTTAATTACCTTTTTTGCGAGGCATATTTATGGTGAAGGATAAGTTT

241   TGACCATCAAAGAAGGTTAATGTGGCTGTGGTTTCAGGGTCCATAAAGCTTTTCAATTCA
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
241   TGACCATCAAAGAAGGTTAATGTGGCTGTGGTTTCAGGGTCCATAAAGCTTTTCAATTCA

301   TCTTTTTTTTTTTTGTTCTTTTTTTTGATTCCGGTTTCTTTGAAATTTTTTTGATTCGGT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
301   TCTTTTTTTTTTTTGTTCTTTTTTTTGATTCCGGTTTCTTTGAAATTTTTTTGATTCGGT

361   AATCTCCGAGCAGAAGGAAGAACGAAGGAAGGAGCACAGACTTAGATTGGTATATATACG
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
361   AATCTCCGAGCAGAAGGAAGAACGAAGGAAGGAGCACAGACTTAGATTGGTATATATACG

421   CATATGTGGTGTTGAAGAAACATGAAATTGCCCAGTATTCTTAACCCAACTGCACAGAAC
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
421   CATATGTGGTGTTGAAGAAACATGAAATTGCCCAGTATTCTTAACCCAACTGCACAGAAC

481   AAAAACCTGCAGGAAACGAAGATAAATCATGTCGAAAGCTACATATAAGGAACGTGCTGC
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
481   AAAAACCTGCAGGAAACGAAGATAAATCATGTCGAAAGCTACATATAAGGAACGTGCTGC

541   TACTCATCCTAGTCCTGTTGCTGCCAAGCTATTTAATATCATGCACGAAAAGCAAACAAA
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
541   TACTCATCCTAGTCCTGTTGCTGCCAAGCTATTTAATATCATGCACGAAAAGCAAACAAA

601   CTTGTGTGCTTCATTGGATGTTCGTACCACCAAGGAATTACTGGAGTTAGTTGAAGCATT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
601   CTTGTGTGCTTCATTGGATGTTCGTACCACCAAGGAATTACTGGAGTTAGTTGAAGCATT

661   AGGTCCCAAAATTTGTTTACTAAAAACACATGTGGATATCTTGACTGATTTTTCCATGGA
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
661   AGGTCCCAAAATTTGTTTACTAAAAACACATGTGGATATCTTGACTGATTTTTCCATGGA

721   GGGCACAGTTAAGCCGCTAAAGGCATTATCCGCCAAGTACAATTTTTTACTCTTCGAAGA
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
721   GGGCACAGTTAAGCCGCTAAAGGCATTATCCGCCAAGTACAATTTTTTACTCTTCGAAGA
                                          Comment 1
781   CAGAAAATTTGCTGACATTGGTAATACAGTCAAATTGCAGCACTCTGCGGGTGTATACAG
      ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
781   CAGAAAATTTGCTGACATTGGTAATACAGTCAAATTGCAGTACTCTGCGGGTGTATACAG

841   AATAGCAGAATGGGCAGACATTACGAATGCACACGGTGTGGTGGGCCCAGGTATTGTTAG
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
841   AATAGCAGAATGGGCAGACATTACGAATGCACACGGTGTGGTGGGCCCAGGTATTGTTAG
                                      Comment 2
901   CGGTTTGAAGCAGGCGGCGGAAGAAGTAGGTACCGCGGCCGCAAAAGCTTTCTAACTGAT
      ||||||||||||||||||||||||||||                  ||||||||||||||||
901   CGGTTTGAAGCAGGCGGCGGAAGAAGTA.....ccccccccccc.AAGCTTTCTAACTGAT

961   CTATCCAAAACTGAAAATTACATTCTTGATTAGGTTTATCACAGGCAAATGTAATTTGTG
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
945   CTATCCAAAACTGAAAATTACATTCTTGATTAGGTTTATCACAGGCAAATGTAATTTGTG

1021  GTATTTTGCCGTTCAAAATCTGTAGAATTTTCTCATTGGTCACATTACAACCTGAAAATA
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1005  GTATTTTGCCGTTCAAAATCTGTAGAATTTTCTCATTGGTCACATTACAACCTGAAAATA
```

```
1081  CTTTATCTACAATCATACCATTCTTATAACATGTCCCCTTAATACTAGGATCAGGCATGA
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1065  CTTTATCTACAATCATACCATTCTTATAACATGTCCCCTTAATACTAGGATCAGGCATGA

1141  ACGCATCACAGACAAAATCTTCTTGACAAACGTCACAATTGATCCCTCCCCATCCGTTAT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1125  ACGCATCACAGACAAAATCTTCTTGACAAACGTCACAATTGATCCCTCCCCATCCGTTAT

1201  CACAATGACAGGTGTCATTTTGTGCTCTTATGGGACGATCCTTATTACCGCTTTCATCCG
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1185  CACAATGACAGGTGTCATTTTGTGCTCTTATGGGACGATCCTTATTACCGCTTTCATCCG

1261  GTGATAGACCGCCACAGAGGGGCAGAGAGCAATCATCACCTGCAAACCCTTCTATACACT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1245  GTGATAGACCGCCACAGAGGGGCAGAGAGCAATCATCACCTGCAAACCCTTCTATACACT

1321  CACATCTACCAGTGTACGAATTGCATTCAGAAAACTGTTTGCATTCAAAAATAGGTAGCA
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1305  CACATCTACCAGTGTACGAATTGCATTCAGAAAACTGTTTGCATTCAAAAATAGGTAGCA

1381  TACAATTAAAACATGGCGGGCATGTATCATTGCCCTTATCTTGTGCAGTTAGACGCGAAT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1365  TACAATTAAAACATGGCGGGCATGTATCATTGCCCTTATCTTGTGCAGTTAGACGCGAAT

1441  TTTTCGAAGAAGTACCTTCAAAGAATGGGGTCTTATCTTGTTTTGCAAGTACCACTGAGC
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1425  TTTTCGAAGAAGTACCTTCAAAGAATGGGGTCTTATCTTGTTTTGCAAGTACCACTGAGC

1501  AGGATAATAATAGAAATGATAATATACTATAGTAGAGATAACGTCGATGACTTCCCATAC
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1485  AGGATAATAATAGAAATGATAATATACTATAGTAGAGATAACGTCGATGACTTCCCATAC

1561  TGTAATTGCTTTTAGTTGTGTATTTTTAGTGTGCAAGTTTCTGTAAATCGATTAATTTTT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1545  TGTAATTGCTTTTAGTTGTGTATTTTTAGTGTGCAAGTTTCTGTAAATCGATTAATTTTT
```

Comment 3

```
1621  TTTCTTTCCTCTTTTTATTAACCTTAATTTTTATTTTAGATTCCTGACTTCAACTCAAG
      ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1605  TTTTCTTTCCTCTTTTTATTAACCTTAATTTTTATTTTAGATTCCTGACTTCAACTCAAG

1680  ACGCACAGATATTATAACATCTGCATAATAGGCATTTGCAAGAATTACTCGTGAGTAAGG
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1665  ACGCACAGATATTATAACATCTGCATAATAGGCATTTGCAAGAATTACTCGTGAGTAAGG

1740  AAAGAGTGAGGAACTATCGCATACCTGCATTTAAAGATGCCGATTTGGGCGCGAATCCTT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1725  AAAGAGTGAGGAACTATCGCATACCTGCATTTAAAGATGCCGATTTGGGCGCGAATCCTT

1800  TATTTTGGCTTCACCCTCATACTATTATCAGGGCCAGAAAAAGGAAGTGTTTCCCTCCTT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1785  TATTTTGGCTTCACCCTCATACTATTATCAGGGCCAGAAAAAGGAAGTGTTTCCCTCCTT

1860  CTTGAATTGATGTTACCCTCATAAAGCACGTGGCCTCTTATCGAGAAAGAAATTACCGTC
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1845  CTTGAATTGATGTTACCCTCATAAAGCACGTGGCCTCTTATCGAGAAAGAAATTACCGTC

1920  GCTCGTGATTTGTTTGCAAAAAGAACAAAACTGAAAAAACCCAGACACGCTCGACTTCCT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1905  GCTCGTGATTTGTTTGCAAAAAGAACAAAACTGAAAAAACCCAGACACGCTCGACTTCCT

1980  GTCTTCCTATTGATTGCAGCTTCCAATTTCGTCACACAACAAGGTCCTAGCGACGGCTCA
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1965  GTCTTCCTATTGATTGCAGCTTCCAATTTCGTCACACAACAAGGTCCTAGCGACGGCTCA

2040  CAGGTTTTGTAACAAGCAATCGAAGGTTCTGGAATGGCGGGAAAGGGTTTAGTACCACAT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
2025  CAGGTTTTGTAACAAGCAATCGAAGGTTCTGGAATGGCGGGAAAGGGTTTAGTACCACAT

2100  GCTATGATGCCCACTGTGATCTCCAGAGCAAAGTTCGTTCGATCGTACTGTTACTCTCTC
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
2085  GCTATGATGCCCACTGTGATCTCCAGAGCAAAGTTCGTTCGATCGTACTGTTACTCTCTC

2160  TCTTTCAAACAGAATTGTCCGAATCGTGTGACAACAACAGCCTGTTCTCACACACTCTTT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
2145  TCTTTCAAACAGAATTGTCCGAATCGTGTGACAACAACAGCCTGTTCTCACACACTCTTT

2220  TCTTCTAACCAAGGGGGTGGTTTAGTTTAGTAGAACCTCGTGAAACTTACATTTACATAT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
2205  TCTTCTAACCAAGGGGGTGGTTTAGTTTAGTAGAACCTCGTGAAACTTACATTTACATAT

2280  ATATAAACTTGCATAAATTGGTCAATGCAAGAAATACATATTTGGTCTTTTCTAATTCGT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
2265  ATATAAACTTGCATAAATTGGTCAATGCAAGAAATACATATTTGGTCTTTTCTAATTCGT

2340  AGTTTTTCAAGTTCTTAGATGCTTTCTTTTTCTCTTTTTTACAGATCATCAAGGAAGTAA
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
2325  AGTTTTTCAAGTTCTTAGATGCTTTCTTTTTCTCTTTTTTACAGATCATCAAGGAAGTAA
```

```
                                        Comment 4
2400  TTATCTACTTTTTACAACAAATATAAAACAAGATCGGAATTAAAAAATGACAGTTAGTTC
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
2385  TTATCTACTTTTTACAACAAATATAAAACA................TGACAGTTAGTTC
                                        Comment 5
2460  CGATACAACTGCTGAAATATCGTTAGGTTGGTCAACCCAAGACTGGATTGATTTCCACAA
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
2428  CGATACAACTGCTGAAATATCGTTAGGTTGGTCAATCCAAGACTGGATTGATTTCCACAA

2520  GTCATCAAGCTCCCAGGCTTCACTAAGGCTTCTTGAATCACTACTAGACTCTCAAAATGT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
2488  GTCATCAAGCTCCCAGGCTTCACTAAGGCTTCTTGAATCACTACTAGACTCTCAAAATGT
                                        Comment 6
2580  TGCGCCAGTCGATAATGCGTGGATATCGCTAATTTCAAAGGAAAATTTACTGCACCAATT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
2548  TGCGCCAGTCGATAATGCGTGGATATCGCTAATTTCAAAGGAAAATTTACTGCACCAATT

2640  CCAAATTTTAAAGAGCAGAGAAAATAAAGAAACTCTACCTCTCTACGGTGTCCCTATTGC
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
2608  CCAAATTTTAAAGAGCAGAGAAAATAAAGAAACTCTACCTCTCTACGGTGTCCCTATTGC
                                        Comment 7          Comment 8
2700  TGTTAAGGACAACATCGACGTTAGAGGTCTACGCACCACCGCTGCATGTCCATCCTTCGC
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
2668  TGTTAAGGACAACATCGACGTTAGAGGTCTACCCACCACCGCTGCATGTCCATCCTTTGC
                                                           Comment 9
2760  ATATGAGCCTTCCAAAGACTCTAAAGTAGTAGAACTACTAAGAAATGCAGGTGCAATAAT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
2728  ATATGAGCCTTCCAAAGACTCTAAAGTAGTAGAACTACTAAGAAATGCAGGTGCGATAAT

2820  CGTGGGTAAGACAAACTTGGACCAATTTGCCACAGGATTAGTCGGCACACGGTCTCCATA
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
2788  CGTGGGTAAGACAAACTTGGACCAATTTGCCACAGGATTAGTCGGCACACGGTCTCCATA

2880  TGGGAAAACACCTTGCGCTTTTAGCAAAGAGCATGTATCTGGTGGTTCCTCCGCTGGGTC
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
2848  TGGGAAAACACCTTGCGCTTTTAGCAAAGAGCATGTATCTGGTGGTTCCTCCGCTGGGTC

2940  AGCATCGGTGGTCGCCAGAGGTATCGTACCAATTGCATTGGGTACTGATACAGCAGGTTC
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
2908  AGCATCGGTGGTCGCCAGAGGTATCGTACCAATTGCATTGGGTACTGATACAGCAGGTTC

3000  TGGTAGAGTCCCAGCCGCCTTGAACAACCTGATTGGCCTAAAGCCAACAAAGGGCGTCTT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
2968  TGGTAGAGTCCCAGCCGCCTTGAACAACCTGATTGGCCTAAAGCCAACAAAGGGCGTCTT

3060  TTCCTGTCAAGGTGTAGTTCCCGCTTGTAAATCTTTAGACTGCGTCTCCATCTTTGCATT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3028  TTCCTGTCAAGGTGTAGTTCCCGCTTGTAAATCTTTAGACTGCGTCTCCATCTTTGCATT

3120  AAACCTAAGTGATGCTGAACGCTGCTTCCGCATCATGTGCCAGCCAGATCCTGATAATGA
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3088  AAACCTAAGTGATGCTGAACGCTGCTTCCGCATCATGTGCCAGCCAGATCCTGATAATGA
                                        Comment 10
3180  TGAATATTCTAGACCCTATGTTTCCAACCCTTTGAAAAATTTTTCAAGCAATGTAACGAT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3148  TGAATATTCTAGACCCTATGTTTCCAACCCTTTGAAAAAATTTTCAAGCAATGTAACGAT

3240  TGCTATTCCTAAAAATATCCCATGGTATGGTGAAACCAAGAATCCTGTACTGTTTTCCAA
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3208  TGCTATTCCTAAAAATATCCCATGGTATGGTGAAACCAAGAATCCTGTACTGTTTTCCAA

3300  TGCTGTCGAAAATCTATCAAGAACGGGCGCTAACGTCATAGAAATTGATTTTGAGCCTCT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3268  TGCTGTCGAAAATCTATCAAGAACGGGCGCTAACGTCATAGAAATTGATTTTGAGCCTCT

3360  TTTAGAGTTAGCTCGCTGTTTATACGAAGGTACTTGGGTGGCCGAGCGTTATCAAGCTAT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3328  TTTAGAGTTAGCTCGCTGTTTATACGAAGGTACTTGGGTGGCCGAGCGTTATCAAGCTAT
                Comment 11
3420  TCAATCGTTTTTGGACAGTAAGCCACCAAAGGAATCTTTGGACCCTACTGTTATTTCAAT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3388  TCAATCGTTTTTGGACAGTAAACCACCAAAGGAATCTTTGGACCCTACTGTTATTTCAAT

3480  TATAGAAGGGGCCAAGAAATACAGTGCAGTAGACTGCTTCAGTTTTGAATACAAAAGACA
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3448  TATAGAAGGGGCCAAGAAATACAGTGCAGTAGACTGCTTCAGTTTTGAATACAAAAGACA
                                                Comment 12
3540  AGGCATCTTGCAAAAAGTGAGACGACTTCTCGAATCAGTCGATGTCTTGTGTGTGCCCAC
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3508  AGGCATCTTGCAAAAAGTGAGACGACTTCTCGAATCAGTCGATGTATTGTGTGTGCCCAC

3600  ATGTCCTTTAAATCCTACTATGCAACAAGTTGCGGATGAACCAGTCCTAGTCAATTCAAG
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3568  ATGTCCTTTAAATCCTACTATGCAACAAGTTGCGGATGAACCAGTCCTAGTCAATTCAAG
```

```
3660  ACAAGGCACATGGACTAATTTTGTCAACTTGGCAGATTTGGCAGCCCTTGCTGTTCCCGC
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3628  ACAAGGCACATGGACTAATTTTGTCAACTTGGCAGATTTGGCAGCCCTTGCTGTTCCCGC

3720  AGGGTTCCGAGACGATGGTTTGCCAAATGGTATTACTTTAATCGGTAAAAAATTCACAGA
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3688  AGGGTTCCGAGACGATGGTTTGCCAAATGGTATTACTTTAATCGGTAAAAAATTCACAGA
                                                Comment 13
3780  TTACGCACTATTAGAGTTGGCTAACCGCTATTTCCAAAATATGTTCCCCAACGGTTCCAG
      ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
3748  TTACGCACTATTAGAGTTGGCTAACCGCTATTTCCAAAATATATTCCCCAACGGTTCCAG

3840  AACATACGGTACTTTTACCTCTTCTTCAGTAAAGCCAGCAAACGATCAATTAGTGGGACC
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3808  AACATACGGTACTTTTACCTCTTCTTCAGTAAAGCCAGCAAACGATCAATTAGTGGGACC

3900  AGACTATGACCCATCTACGTCCATAAAATTGGCTGTTGTCGGTGCACATCTTAAGGGTCT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3868  AGACTATGACCCATCTACGTCCATAAAATTGGCTGTTGTCGGTGCACATCTTAAGGGTCT
                  Comment 14
3960  GCCTCTACATTGGCAATTGGAGAAGGTCAATGCAACATATTTATGTACAACAAAAACATC
      ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
3928  GCCTCTACATTGGCAATTGGAAAAGGTCAATGCAACATATTTATGTACAACAAAAACATC

4020  AAAAGCTTACCAGCTTTTTGCTTTGCCCAAAAATGGACCAGTTTTAAAACCTGGTTTGAG
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3988  AAAAGCTTACCAGCTTTTTGCTTTGCCCAAAAATGGACCAGTTTTAAAACCTGGTTTGAG

4080  AAGAGTTCAAGATAGCAATGGCTCTCAAATCGAATTAGAAGTGTACAGTGTTCCAAAAGA
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
4048  AAGAGTTCAAGATAGCAATGGCTCTCAAATCGAATTAGAAGTGTACAGTGTTCCAAAAGA

4140  ACTGTTCGGTGCTTTTATTTCCATGGTTCCTGAACCATTAGGAATAGGTTCAGTGGAGTT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
4108  ACTGTTCGGTGCTTTTATTTCCATGGTTCCTGAACCATTAGGAATAGGTTCAGTGGAGTT

4200  AGAATCTGGTGAATGGATCAAATCCTTTATTTGTGAAGAATCTGGTTACAAAGCCAAAGG
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
4168  AGAATCTGGTGAATGGATCAAATCCTTTATTTGTGAAGAATCTGGTTACAAAGCCAAAGG

4260  TACAGTTGATATCACAAAGTATGGTGGATTTAGAGCATATTTTGAAATGTTGAAGAAAAA
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
4228  TACAGTTGATATCACAAAGTATGGTGGATTTAGAGCATATTTTGAAATGTTGAAGAAAAA

4320  AGAGTCCCAAAAGAAGAAGTTATTTGATACCGTGTTAATTGCCAATAGAGGTGAAATTGC
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
4288  AGAGTCCCAAAAGAAGAAGTTATTTGATACCGTGTTAATTGCCAATAGAGGTGAAATTGC
Comment 15
4380  CGTCCGTATTATCAAGACATTAAAAAAATTGGGTATTAGATCAGTTGCAGTTTATTCCGA
      ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
4348  CGTTCGTATTATCAAGACATTAAAAAAATTGGGTATTAGATCAGTTGCAGTTTATTCCGA

4440  CCCTGATAAATATTCTCAACACGTTACTGATGCAGATGTTTCTGTACCCCTTCATGGCAC
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
4408  CCCTGATAAATATTCTCAACACGTTACTGATGCAGATGTTTCTGTACCCCTTCATGGCAC

4500  AACCGCAGCCCAAACTTATTTAGACATGAATAAGATCATAGATGCCGCTAAGCAAACTAA
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
4468  AACCGCAGCCCAAACTTATTTAGACATGAATAAGATCATAGATGCCGCTAAGCAAACTAA

4560  TGCACAGGCCATTATTCCTGGTTATGGTTTCTTGTCGGAAAATGCGGATTTTTCTGATGC
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
4528  TGCACAGGCCATTATTCCTGGTTATGGTTTCTTGTCGGAAAATGCGGATTTTTCTGATGC

4620  GTGCACCAGTGCTGGCATTACCTTTGTTGGTCCTTCGGGAGATATTATCAGAGGTTTAGG
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
4588  GTGCACCAGTGCTGGCATTACCTTTGTTGGTCCTTCGGGAGATATTATCAGAGGTTTAGG

4680  GTTAAAACATTCTGCTAGACAGATTGCACAGAAGGCTGGCGTTCCTCTAGTGCCAGGCTC
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
4648  GTTAAAACATTCTGCTAGACAGATTGCACAGAAGGCTGGCGTTCCTCTAGTGCCAGGCTC
                                                Comment 16
4740  TTTGCTTATCACATCAGTTGAAGAGGCTAAGAAAGTCGCAGCGGAATTGGAATACCCGGT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
4708  TTTGCTTATCACATCAGTTGAAGAGGCTAAGAAAGTCGCAGCGGAATTGGAATACCCAGT
                  Comment 17
4800  TATGGTGAAGTCAACTGCTGGTGGTGGTGGTATTGGTTTGCAGAAAGTCGATTCTGAAGA
      ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
4768  TATGGTGAAGTCAACTGCTGGTGGCGGTGGTATTGGTTTGCAGAAAGTCGATTCTGAAGA
```

```
                Comment 18                                  Comment 19
4860  GGAGATCGAGCATATTTTTGAGACTGTGAAACATCAAGGTGAAACATTTTTTGGTGACGC
      ||| |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
4828  GGACATCGAGCATATTTTTGAGACTGTGAAACATCAAGGTGAAACATTTTTCGGTGACGC
                                   Comment 20  Comment 21
4920  TGGTGTATTTCTGGAACGGTTTATCGAAAATGCCAGGCATGTTGAGGTCCAACTTATGGG
      |||||||||||||||||||||||||||||||||||||||| |||||||| |||||||||||||
4888  TGGTGTATTTCTGGAACGGTTTATCGAAAATGCCAGGCATGTTGAAGTCCAACTTATGGG

4980  AGATGGTTTTGGTAAGGCCATTGCTTTGGGCGAACGTGATTGTTCTTTACAGCGTCGTAA
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
4948  AGATGGTTTTGGTAAGGCCATTGCTTTGGGCGAACGTGATTGTTCTTTACAGCGTCGTAA

5040  CCAAAAAGTTATCGAAGAAACTCCTGCACCAAATTTGCCAGAAAAGACGAGGTTGGCGTT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
5008  CCAAAAAGTTATCGAAGAAACTCCTGCACCAAATTTGCCAGAAAAGACGAGGTTGGCGTT

5100  AAGAAAGGCAGCTGAAAGTTTGGGATCTTTATTGAATTACAAGTGTGCTGGTACGGTTGA
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
5068  AAGAAAGGCAGCTGAAAGTTTGGGATCTTTATTGAATTACAAGTGTGCTGGTACGGTTGA

5160  ATTTATTTACGATGAGAAAAAGGACGAGTTTTACTTTTTAGAAGTTAATACAAGATTACA
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
5128  ATTTATTTACGATGAGAAAAAGGACGAGTTTTACTTTTTAGAAGTTAATACAAGATTACA
                                Comment 22
5220  AGTTGAACATCCAATAACAGAAATGGTTACGGGGTTAGACTTGGTCGAGTGGATGATCAG
      ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
5188  AGTTGAACATCCAATAACAGAAATGGTTACAGGGTTAGACTTGGTCGAGTGGATGATCAG

5280  GATTGCCGCTAATGATGCACCTGATTTTGATTCTACAAAGGTAGAAGTCAATGGGGTTTC
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
5248  GATTGCCGCTAATGATGCACCTGATTTTGATTCTACAAAGGTAGAAGTCAATGGGGTTTC

5340  AATGGAGGCACGTTTATATGCTGAAAATCCATTGAAAAATTTCAGACCTTCTCCAGGTTT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
5308  AATGGAGGCACGTTTATATGCTGAAAATCCATTGAAAAATTTCAGACCTTCTCCAGGTTT
                  Comment 23
5400  ACTTGTCGATGTGAAATTCCCTGATTGGGCAAGAGTGGATACTTGGGTTAAGAAAGGTAC
      |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
5368  ACTTGTCGATGTGAAATTTCCTGATTGGGCAAGAGTGGATACTTGGGTTAAGAAAGGTAC

5460  TAATATTTCTCCCGAATATGATCCAACATTGGCCAAAATTATCGTTCATGGGAAAGACCG
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
5428  TAATATTTCTCCCGAATATGATCCAACATTGGCCAAAATTATCGTTCATGGGAAAGACCG

5520  TGATGATGCAATTTCCAAGTTAAATCAAGCGTTAGAAGAAACAAAAGTTTACGGATGTAT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
5488  TGATGATGCAATTTCCAAGTTAAATCAAGCGTTAGAAGAAACAAAAGTTTACGGATGTAT

5580  TACTAACATTGACTACCTGAAGTCTATCATTACCAGTGATTTCTTTGCTAAAGCAAAAGT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
5548  TACTAACATTGACTACCTGAAGTCTATCATTACCAGTGATTTCTTTGCTAAAGCAAAAGT

5640  TTCTACAAACATTTTGAACTCTTATCAATATGAGCCTACCGCCATCGAAATTACTTTGCC
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
5608  TTCTACAAACATTTTGAACTCTTATCAATATGAGCCTACCGCCATCGAAATTACTTTGCC

5700  CGGTGCACACACTAGTATTCAGGATTACCCCGGTAGAGTTGGGTACTGGAGAATTGGTGT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
5668  CGGTGCACACACTAGTATTCAGGATTACCCCGGTAGAGTTGGGTACTGGAGAATTGGTGT

5760  TCCGCCCTCTGGTCCAATGGACGCATATTCGTTTAGATTGGCGAACAGAATTGTTGGTAA
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
5728  TCCGCCCTCTGGTCCAATGGACGCATATTCGTTTAGATTGGCGAACAGAATTGTTGGTAA

5820  TGACTACAGGACTCCTGCCATTGAAGTAACGTTGACTGGTCCATCCATCGTTTTCCATTG
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
5788  TGACTACAGGACTCCTGCCATTGAAGTAACGTTGACTGGTCCATCCATCGTTTTCCATTG

5880  TGAAACTGTCATTGCCATTACTGGTGGTACCGCTCTATGTACATTAGACGGCCAAGAAAT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
5848  TGAAACTGTCATTGCCATTACTGGTGGTACCGCTCTATGTACATTAGACGGCCAAGAAAT

5940  TCCCCAACACAAACCGGTCGAAGTTAAGAGGGGATCTACTTTATCCATTGGCAAGTTGAC
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
5908  TCCCCAACACAAACCGGTCGAAGTTAAGAGGGGATCTACTTTATCCATTGGCAAGTTGAC

6000  AAGCGGCTGTAGAGCATACTTAGGTATCAGGGGTGGCATTGATGTGCCTAAATACTTGGG
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
5968  AAGCGGCTGTAGAGCATACTTAGGTATCAGGGGTGGCATTGATGTGCCTAAATACTTGGG

6060  CTCTTATTCTACTTTCACTCTAGGAAATGTCGGTGGATACAATGGAAGGGTGCTAAAACT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
6028  CTCTTATTCTACTTTCACTCTAGGAAATGTCGGTGGATACAATGGAAGGGTGCTAAAACT
```

```
6120  TGGAGACGTACTATTCTTACCAAGCAATGAAGAAAATAAATCAGTTGAGTGCCTTCCACA
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
6088  TGGAGACGTACTATTCTTACCAAGCAATGAAGAAAATAAATCAGTTGAGTGCCTTCCACA

6180  GAATATTCCTCAATCATTAATTCCTCAAATTTCCGAAACTAAGGAATGGAGAATTGGTGT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
6148  GAATATTCCTCAATCATTAATTCCTCAAATTTCCGAAACTAAGGAATGGAGAATTGGTGT

6240  AACATGTGGTCCCCATGGGTCTCCAGATTTTTTTAAACCTGAGTCCATCGAAGAATTTTT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
6208  AACATGTGGTCCCCATGGGTCTCCAGATTTTTTTAAACCTGAGTCCATCGAAGAATTTTT

6300  CAGTGAGAAGTGGAAGGTTCATTACAACTCCAATAGATTTGGTGTCCGTTTGATTGGACC
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
6268  CAGTGAGAAGTGGAAGGTTCATTACAACTCCAATAGATTTGGTGTCCGTTTGATTGGACC
                                                Comment 24
6360  TAAACCTAAGTGGGCAAGAAGTAATGGTGGTGAAGGTGGTATGCATCCCTCAAACACTCA
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
6328  TAAACCTAAGTGGGCAAGAAGTAATGGTGGTGAAGGTGGTATGCATCCTTCAAACACTCA

6420  CGATTACGTTTATTCTCTGGGTGCAATTAATTTCACGGGTGATGAGCCAGTTATTATTAC
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
6388  CGATTACGTTTATTCTCTGGGTGCAATTAATTTCACGGGTGATGAGCCAGTTATTATTAC

6480  TTGCGATGGTCCTTCCTTAGGTGGTTTTGTGTGTCAAGCTGTTGTCCCAGAAGCAGAACT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
6448  TTGCGATGGTCCTTCCTTAGGTGGTTTTGTGTGTCAAGCTGTTGTCCCAGAAGCAGAACT

6540  GTGGAAGGTTGGACAGGTTAAACCCGGTGATTCCATTCAGTTTGTGCCACTTTCTTACGA
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
6508  GTGGAAGGTTGGACAGGTTAAACCCGGTGATTCCATTCAGTTTGTGCCACTTTCTTACGA

6600  AAGCTCGAGATCCTTAAAGGAATCTCAGGATGTTGCAATTAAATCATTGGATGGTACTAA
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
6568  AAGCTCGAGATCCTTAAAGGAATCTCAGGATGTTGCAATTAAATCATTGGATGGTACTAA

6660  GTTAAGGCGCTTAGACTCTGTTTCAATTTTACCATCATTCGAAACGCCTATTCTTGCACA
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
6628  GTTAAGGCGCTTAGACTCTGTTTCAATTTTACCATCATTCGAAACGCCTATTCTTGCACA

6720  AATGGAAAAAGTGAATGAGCTTTCACCAAAGGTTGTATACAGACAAGCAGGTGATCGTTA
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
6688  AATGGAAAAAGTGAATGAGCTTTCACCAAAGGTTGTATACAGACAAGCAGGTGATCGTTA

6780  TGTTTTGGTGGAATACGGTGATAATGAAATGAATTTTAATATTTCCTATAGAATTGAATG
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
6748  TGTTTTGGTGGAATACGGTGATAATGAAATGAATTTTAATATTTCCTATAGAATTGAATG

6840  CCTGATCTCCCTTGTGAAAAAGAATAAGACTATTGGTATTGTTGAAATGTCCCAAGGTGT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
6808  CCTGATCTCCCTTGTGAAAAAGAATAAGACTATTGGTATTGTTGAAATGTCCCAAGGTGT

6900  TAGATCTGTATTGATAGAATTTGATGGTTACAAAGTCACTCAAAAAGAATTGCTTAAAGT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
6868  TAGATCTGTATTGATAGAATTTGATGGTTACAAAGTCACTCAAAAAGAATTGCTTAAAGT
                                                Comment 25
6960  ATTGGTGGCATATGAAACAGAAATCCAGTTTGATGAAAATTGGAACATAACTTCTAATAT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
6928  ATTGGTGGCATATGAAACAGAAATCCAGTTTGATGAAAATTGGAAGATAACTTCTAATAT

7020  AATAAGATTACCGATGGCTTTCGAAGACTCGAAGACTTTGGCATGTGTTCAAAGGTATCA
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
6988  AATAAGATTACCGATGGCTTTCGAAGACTCGAAGACTTTGGCATGTGTTCAAAGGTATCA

7080  AGAAACAATTCGTTCGTCTGCTCCATGGTTGCCAAATAACGTTGATTTCATTGCCAATGT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
7048  AGAAACAATTCGTTCGTCTGCTCCATGGTTGCCAAATAACGTTGATTTCATTGCCAATGT

7140  AAATGGAATTTCAAGGAATGAAGTTTATGATATGTTGTATTCTGCCAGATTTATGGTTTT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
7108  AAATGGAATTTCAAGGAATGAAGTTTATGATATGTTGTATTCTGCCAGATTTATGGTTTT
                                    Comment 26
7200  AGGTTTAGGTGATGTCTTCCTAGGGTCGCCTTGTACTGTTCCATTAGATCCTCGTCACAG
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
7168  AGGTTTAGGTGATGTCTTCCTAGGGTCGCCTTGTGCTGTTCCATTAGATCCTCGTCACAG

7260  ATTTTTGGGAAGCAAGTACAACCCAAGTAGAACATATACAGAAAGAGGTGCAGTCGGTAT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
7228  ATTTTTGGGAAGCAAGTACAACCCAAGTAGAACATATACAGAAAGAGGTGCAGTCGGTAT

7320  TGGCGGTATGTATATGTGCATATATGCTGCTAACAGTCCTGGTGGGTACCAATTAGTGGG
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
7288  TGGCGGTATGTATATGTGCATATATGCTGCTAACAGTCCTGGTGGGTACCAATTAGTGGG
```

BEST ORIGINAL COPY

```
7380  TAGAACAATACCAATTTGGGACAAACTATGTCTGGCCGCATCTTCTGAGGTTCCGTGGTT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
7348  TAGAACAATACCAATTTGGGACAAACTATGTCTGGCCGCATCTTCTGAGGTTCCGTGGTT

7440  GATGAACCCATTTGACCAAGTCGAATTTTACCCAGTTTCTGAAGAAGATTTGGATAAAAT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
7408  GATGAACCCATTTGACCAAGTCGAATTTTACCCAGTTTCTGAAGAAGATTTGGATAAAAT

7500  GACTGAAGATTGTGATAATGGTGTTTATAAAGTCAATATCGAAAAGAGTGTTTTTGATCA
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
7468  GACTGAAGATTGTGATAATGGTGTTTATAAAGTCAATATCGAAAAGAGTGTTTTTGATCA

7560  TCAAGAATACTTGAGATGGATCAACGCAAACAAAGATTCCATCACAGCATTCCAGGAGGG
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
7528  TCAAGAATACTTGAGATGGATCAACGCAAACAAAGATTCCATCACAGCATTCCAGGAGGG

7620  CCAGCTTGGTGAAAGAGCAGAGGAATTTGCCAAATTGATTCAAAATGCAAACTCTGAACT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
7588  CCAGCTTGGTGAAAGAGCAGAGGAATTTGCCAAATTGATTCAAAATGCAAACTCTGAACT

7680  AAAAGAAAGTGTCACAGTCAAACCTGACGAGGAAGAAGACTTCCCAGAAGGTGCAGAAAT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
7648  AAAAGAAAGTGTCACAGTCAAACCTGACGAGGAAGAAGACTTCCCAGAAGGTGCAGAAAT
                                                   Comment 27
7740  TGTATATTCTGAGTATTCTGGGCGTTTTTGGAAATCCATAGCATCTGTCGGAGATGTTAT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
7708  TGTATATTCTGAGTATTCTGGGCGTTTTTGGAAATCCATAGCATCTGTTGGAGATGTTAT

7800  TGAAGCAGGTCAAGGGCTACTAATTATTGAAGCCATGAAAGCGGAAATGATTATATCCGC
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
7768  TGAAGCAGGTCAAGGGCTACTAATTATTGAAGCCATGAAAGCGGAAATGATTATATCCGC

7860  TCCTAAATCGGGTAAGATTATCAAGATTTGCCATGGCAATGGTGATATGGTTGATTCTGG
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
7828  TCCTAAATCGGGTAAGATTATCAAGATTTGCCATGGCAATGGTGATATGGTTGATTCTGG
                                                   Comment 28
7920  TGACATAGTGGCCGTCATAGAGACATTGGCATGAAATACCTTTTTCGATCTCCTCGAGGG
      ||||||||||||||||||||||||||||||||||
7888  TGACATAGTGGCCGTCATAGAGACATTGGCATGA.........................

7980  AUCTGCGATAGATCAATTTTTTTCTTTTCTCTTTCCCCATCCTTTACGCTAAAATAATAG
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
7922  .....CGATAGATCAATTTTTTTCTTTTCTCTTTCCCCATCCTTTACGCTAAAATAATAG

8040  TTTATTTTATTTTTTGAATATTTTTTATTTATATACGTATATATAGACTATTATTTATCT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
7977  TTTATTTTATTTTTTGAATATTTTTTATTTATATACGTATATATAGACTATTATTTATCT

8100  TTTAATGATTATTAAGATTTTTATTAAAAAAAAAATTCGCTCCTCTTTTAATGCCTTTATG
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
8037  TTTAATGATTATTAAGATTTTTATTAAAAAAAAAATTCGCTCCTCTTTTAATGCCTTTATG

8160  CAGTTTTTTTTTCCCATTCGATATTTCTATGTTCGGGTTCAGCGTATTTTAAGTTTAATA
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
8097  CAGTTTTTTTTTCCCATTCGATATTTCTATGTTCGGGTTCAGCGTATTTTAAGTTTAATA
    Comment 29 Comment 30 Comment 31 Comment 32
8220  ACTCGAAAATTCTCCGTCGTTAAAGCTTTTGCGGCGGCACGAAGGAACCTAGAGGCCTT
      ||||||||||||| || |||||||||||          || ||||||||||||||||||
8157  ACTCGAAAATTCTGCGTTCGTTAAAGCTT..........ACAAAGGAACCTAGAGGCCTT

8278  TTGATGTTAGCAGAATTGTCATGCAAGGGCTCCCTAGCTACTGGAGAATATACTAAGGGT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
8207  TTGATGTTAGCAGAATTGTCATGCAAGGGCTCCCTAGCTACTGGAGAATATACTAAGGGT

8338  ACTGTTGACATTGCGAAGAGCGACAAAGATTTTGTTATCGGCTTTATTGCTCAAAGAGAC
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
8267  ACTGTTGACATTGCGAAGAGCGACAAAGATTTTGTTATCGGCTTTATTGCTCAAAGAGAC

8398  ATGGGTGGAAGAGATGAAGGTTACGATTGGTTGATTATGACACCCGGTGTGGGTTTAGAT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
8327  ATGGGTGGAAGAGATGAAGGTTACGATTGGTTGATTATGACACCCGGTGTGGGTTTAGAT

8458  GACAAGGGAGACGCATTGGGTCAACAGTATAGAACCGTGGATGATGTGGTCTCTACAGGA
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
8387  GACAAGGGAGACGCATTGGGTCAACAGTATAGAACCGTGGATGATGTGGTCTCTACAGGA

8518  TCTGACATTATTATTGTTGGAAGAGGACTATTTGCAAAGGGAAGGGATGCTAAGGTAGAG
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
8447  TCTGACATTATTATTGTTGGAAGAGGACTATTTGCAAAGGGAAGGGATGCTAAGGTAGAG

8578  GGTGAACGTTACAGAAAAGCAGGCTGGGAAGCATATTTGAGAAGATGCGGCCAGCAAAAC
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
8507  GGTGAACGTTACAGAAAAGCAGGCTGGGAAGCATATTTGAGAAGATGCGGCCAGCAAAAC
```

BEST ORIGINAL COPY

000079

```
8638  TAAAAAACTGTATTATAAGTAAATGCATGTATACTAAACTCACAAATTAGAGCTTCAATT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
8567  TAAAAAACTGTATTATAAGTAAATGCATGTATACTAAACTCACAAATTAGAGCTTCAATT
```

Comment 33

```
8698  TAATTATATCAGTTATTACCCGGGAATCTCGGTCGTAATGATTTTTATAATGACGAAAAA
      ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
8627  TAATTATATCAGTTATTACCCGGGAATCTCGGTCGTAATGATTTCTATAATGACGAAAAA
```

Comment 34

```
8758  AAAAAAATTGGAAAGAAAAAGCTTCATGGCCTTTATAAAAAGGAACCATCCAATACCTCG
      ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
8687  AAAAAAATTGGAAAGAAAAAGCTTCATGGCCTTTATAAAAAGGAACTATCCAATACCTCG

8818  CCAGAACCAAGTAACAGTATTTTACGGGGCACAAATCAAGAACAATAAGACAGGACTGTA
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
8747  CCAGAACCAAGTAACAGTATTTTACGGGGCACAAATCAAGAACAATAAGACAGGACTGTA

8878  AAGATGGACGCATTGAACTCCAAAGAACAACAAGAGTTCCAAAAAGTAGTGGAACAAAAG
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
8807  AAGATGGACGCATTGAACTCCAAAGAACAACAAGAGTTCCAAAAAGTAGTGGAACAAAAG

8938  CAAATGAAGGATTTCATGCGTTTGTACTCTAATCTGGTAGAAAGATGTTTCACAGACTGT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
8867  CAAATGAAGGATTTCATGCGTTTGTACTCTAATCTGGTAGAAAGATGTTTCACAGACTGT

8998  GTCAATGACTTCACAACATCAAAGCTAACCAATAAGGAACAAACATGCATCATGAAGTGC
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
8927  GTCAATGACTTCACAACATCAAAGCTAACCAATAAGGAACAAACATGCATCATGAAGTGC

9058  TCAGAAAAGTTCTTGAAGCATAGCGAACGTGTAGGGCAGCGTTTCCAAGAACAAAACGCT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
8987  TCAGAAAAGTTCTTGAAGCATAGCGAACGTGTAGGGCAGCGTTTCCAAGAACAAAACGCT

9118  GCCTTGGGACAAGGCTTGGGCCGATAAGGTGTACTGGCGTATATATATCTAATTATGTAT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
9047  GCCTTGGGACAAGGCTTGGGCCGATAAGGTGTACTGGCGTATATATATCTAATTATGTAT

9178  CTCTGGTGTAGCCC
      ||||||||||||||
9107  CTCTGGTGTAGCCC
```

BEST ORIGINAL COPY

Discrepancies found between the genomic sequence of the integrated locus and previously published S288C sequences (comments and nucleotide positions refer to the sequence composition detailed in Table 3).

Comments and nucleotide positions refer to the sequence composition detailed in Table 3.

Comment	Nucleotide position	Description
1	821	Difference in the 5' region of the *URA3* open reading frame
2	929-944	Additional sequence resulting from cloning strategy
3	1623/1624	Difference in the *PGK1* promoter sequence. One base deletion between 1623 and 1624 bp.
4	2429-2446	Additional sequence resulting from cloning strategy
5	2495	Difference in the *DUR1,2* open reading frame. This difference corresponds to a change of amino acids from isoleucine (in the published sequence) to threonine (in the genomic sequence).
6	2607	Difference in the *DUR1,2* open reading frame. This difference corresponds to **no change** of the amino acid sequence.
7	2732	Difference in the *DUR1,2* open reading frame. This difference corresponds to a change of amino acids from proline (in the published sequence) to arginine (in the genomic sequence).
8 and 9	2757 and 2814	Differences in the *DUR1,2* open reading frame. These differences correspond to **no change** of the amino acid sequence.
10	3219	Difference in the *DUR1,2* open reading frame. This difference corresponds to a change of amino acids from lysine (in the published sequence) to asparagine (in the genomic sequence).
11 and 12	3441 and 3585	Differences in the *DUR1,2* open reading frame. These differences correspond to **no change** of the amino acid sequence.
13	3722	Difference in the *DUR1,2* open reading frame. This difference corresponds to a change of amino acids from isoleucine (in the published sequence) to methionine (in the genomic sequence).
14 - 24	3981, 4383, 4797, 4824, 4863, 4911, 4956, 4965, 5250, 5418, and 6408	Differences in the *DUR1,2* open reading frame. These differences correspond to **no change** of the amino acid sequence.
25	7005	Difference in the *DUR1,2* open reading frame. This difference corresponds to a change of amino acids from lysine (in the published sequence) to asparagine (in the genomic sequence).
26	7234	Difference in the *DUR1,2* open reading frame. This difference corresponds to a change of amino acids from alanine (in the published sequence) to threonine (in the genomic sequence).
27	7788	Difference in the *DUR1,2* open reading frame. This difference corresponds to **no change** of the amino acid sequence.
28	7954-7984	Additional sequence resulting from cloning strategy
29	8232/8233	Difference in the *PGK1* terminator sequence between. One base deletion between 8232 and 8233 bp.

30	8234/8235	Difference in the *PGK1* terminator sequence between. One base deletion between 8234 and 8235 bp.
31	8249-8256	Additional sequence resulting from cloning strategy
32	8259	Difference in the 3′ region of the *URA3* open reading frame
33	8742	Difference in the 3′ region of the *URA3* non coding sequence
34	8804	Difference in the 3′ region of the *URA3* non coding sequence

Appendix 3: Nucleotide sequence of the *DUR1,2* gene and protein sequence of the corresponding protein in Fasta format

Nucleotide sequence integrated into ECMo01

>
ATGACAGTTAGTTCCGATACAACTGCTGAAATATCGTTAGGTTGGTCAAC
CCAAGACTGGATTGATTTCCACAAGTCATCAAGCTCCCAGGCTTCACTAA
GGCTTCTTGAATCACTACTAGACTCTCAAAATGTTGCGCCAGTCGATAAT
GCGTGGATATCACTAATTTCAAAGGAAAATTTACTGCACCAATTCCAAAT
TTTAAAGAGCAGAGAAAATAAAGAAACTCTACCTCTCTACGGTGTCCCTA
TTGCTGTTAAGGACAACATCGACGTTAGAGGTCTACGCACCACCGCTGCA
TGTCCATCCTTCGCATATGAGCCTTCCAAAGACTCTAAAGTAGTAGAACT
ACTAAGAAATGCAGGTGCAATAATCGTGGGTAAGACAAACTTGGACCAAT
TTGCCACAGGATTAGTCGGCACACGGTCTCCATATGGGAAAACACCTTGC
GCTTTTAGCAAAGAGCATGTATCTGGTGGTTCCTCCGCTGGGTCAGCATC
GGTGGTCGCCAGAGGTATCGTACCAATTGCATTGGGTACTGATACAGCAG
GTTCTGGTAGAGTCCCAGCCGCCTTGAACAACCTGATTGGCCTAAAGCCA
ACAAAGGGCGTCTTTTCCTGTCAAGGTGTAGTTCCCGCTTGTAAATCTTT
AGACTGCGTCTCCATCTTTGCATTAAACCTAAGTGATGCTGAACGCTGCT
TCCGCATCATGTGCCAGCCAGATCCTGATAATGATGAATATTCTAGACCC
TATGTTTCCAACCCTTTGAAAAATTTTTCAAGCAATGTAACGATTGCTAT
TCCTAAAAATATCCCATGGTATGGTGAAACCAAGAATCCTGTACTGTTTT
CCAATGCTGTCGAAAATCTATCAAGAACGGGCGCTAACGTCATAGAAATT
GATTTTGAGCCTCTTTTAGAGTTAGCTCGCTGTTTATACGAAGGTACTTG
GGTGGCCGAGCGTTATCAAGCTATTCAATCGTTTTTGGACAGTAAGCCAC
CAAAGGAATCTTTGGACCCTACTGTTATTTCAATTATAGAAGGGGCCAAG
AAATACAGTGCAGTAGACTGCTTCAGTTTTGAATACAAAAGACAAGGCAT
CTTGCAAAAAGTGAGACGACTTCTCGAATCAGTCGATGTCTTGTGTGTGC
CCACATGTCCTTTAAATCCTACTATGCAACAAGTTGCGGATGAACCAGTC
CTAGTCAATTCAAGACAAGGCACATGGACTAATTTTGTCAACTTGGCAGA
TTTGGCAGCCCTTGCTGTTCCCGCAGGGTTCCGAGACGATGGTTTGCCAA
ATGGTATTACTTTAATCGGTAAAAAATTCACAGATTACGCACTATTAGAG
TTGGCTAACCGCTATTTCCAAAATATGTTCCCCAACGGTTCCAGAACATA
CGGTACTTTTACCTCTTCTTCAGTAAAGCCAGCAAACGATCAATTAGTGG
GACCAGACTATGACCCATCTACGTCCATAAAATTGGCTGTTGTCGGTGCA
CATCTTAAGGGTCTGCCTCTACATTGGCAATTGGAGAAGGTCAATGCAAC
ATATTTATGTACAACAAAAACATCAAAAGCTTACCAGCTTTTTGCTTTGC
CCAAAAATGGACCAGTTTTAAAACCTGGTTTGAGAAGAGTTCAAGATAGC
AATGGCTCTCAAATCGAATTAGAAGTGTACAGTGTTCCAAAAGAACTGTT
CGGTGCTTTTATTTCCATGGTTCCTGAACCATTAGGAATAGGTTCAGTGG
AGTTAGAATCTGGTGAATGGATCAAATCCTTTATTTGTGAAGAATCTGGT
TACAAAGCCAAAGGTACAGTTGATATCACAAAGTATGGTGGATTTAGAGC
ATATTTTGAAATGTTGAAGAAAAAAGAGTCCCAAAAGAAGAAGTTATTTG
ATACCGTGTTAATTGCCAATAGAGGTGAAATTGCCGTCCGTATTATCAAG
ACATTAAAAAAATTGGGTATTAGATCAGTTGCAGTTTATTCCGACCCTGA
TAAATATTCTCAACACGTTACTGATGCAGATGTTTCTGTACCCCTTCATG
GCACAACCGCAGCCCAAACTTATTTAGACATGAATAAGATCATAGATGCC
GCTAAGCAAACTAATGCACAGGCCATTATTCCTGGTTATGGTTTCTTGTC
GGAAAATGCGGATTTTTCTGATGCGTGCACCAGTGCTGGCATTACCTTTG
TTGGTCCTTCGGGAGATATTATCAGAGGTTTAGGGTTAAAACATTCTGCT
AGACAGATTGCACAGAAGGCTGGCGTTCCTCTAGTGCCAGGCTCTTTGCT
TATCACATCAGTTGAAGAGGCTAAGAAAGTCGCAGCGGAATTGGAATACC
CGGTTATGGTGAAGTCAACTGCTGGTGGTGGTGGTATTGGTTTGCAGAAA
GTCGATTCTGAAGAGGATATCGAGCATATTTTTGAGACTGTGAAACATCA
AGGTGAAACATTTTTTGGTGACGCTGGTGTATTTCTGGAACGGTTTATCG
AAAATGCCAGACATGTTGAGGTCCAACTTATGGGAGATGGTTTTGGTAAG
GCCATTGCTTTGGGCGAACGTGATTGTTCTTTACAGCGTCGTAACCAAAA
AGTTATCGAAGAAACTCCTGCACCAAATTTGCCAGAAAAGACGAGGTTGG
CGTTAAGAAAGGCAGCTGAAAGTTTGGGATCTTTATTGAATTACAAGTGT

000084

BEST ORIGINAL COPY

GCTGGTACGGTTGAATTTATTTACGATGAGAAAAAGGACGAGTTTTACTT
TTTAGAAGTTAATACAAGATTACAAGTTGAACATCCAATAACAGAAATGG
TTACGGGGTTAGACTTGGTCGAGTGGATGATCAGGATTGCCGCTAATGAT
GCACCTGATTTTGATTCTACAAAGGTAGAAGTCAATGGGGTTTCAATGGA
GGCACGTTTATATGCTGAAAATCCATTGAAAAATTTCAGACCTTCTCCAG
GTTTACTTGTCGATGTGAAATTCCCTGATTGGGCAAGAGTGGATACTTGG
GTTAAGAAAGGTACTAATATTTCTCCCGAATATGATCCAACATTGGCCAA
AATTATCGTTCATGGGAAAGACCGTGATGATGCAATTTCCAAGTTAAATC
AAGCGTTAGAAGAAACAAAAGTTTACGGATGTATTACTAACATTGACTAC
CTGAAGTCTATCATTACCAGTGATTTCTTTGCTAAAGCAAAAGTTTCTAC
AAACATTTTGAACTCTTATCAATATGAGCCTACCGCCATCGAAATTACTT
TGCCCGGTGCACACACTAGTATTCAGGATTACCCCGGTAGAGTTGGGTAC
TGGAGAATTGGTGTTCCGCCCTCTGGTCCAATGGACGCATATTCGTTTAG
ATTGGCGAACAGAATTGTTGGTAATGACTACAGGACTCCTGCCATTGAAG
TAACGTTGACTGGTCCATCCATCGTTTTCCATTGTGAAACTGTCATTGCC
ATTACTGGTGGTACCGCTCTATGTACATTAGACGGCCAAGAAATTCCCCA
ACACAAACCGGTCGAAGTTAAGAGGGGATCTACTTTATCCATTGGCAAGT
TGACAAGCGGCTGTAGAGCATACTTAGGTATCAGGGGTGGCATTGATGTG
CCTAAATACTTGGGCTCTTATTCTACTTTCACTCTAGGAAATGTCGGTGG
ATACAATGGAAGGGTGCTAAAACTTGGAGACGTACTATTCTTACCAAGCA
ATGAAGAAAATAAATCAGTTGAGTGCCTTCCACAGAATATTCCTCAATCA
TTAATTCCTCAAATTTCCGAAACTAAGGAATGGAGAATTGGTGTAACATG
TGGTCCCCATGGGTCTCCAGATTTTTTTAAACCTGAGTCCATCGAAGAAT
TTTTCAGTGAGAAGTGGAAGGTTCATTACAACTCCAATAGATTTGGTGTC
CGTTTGATTGGACCTAAACCTAAGTGGGCAAGAAGTAATGGTGGTGAAGG
TGGTATGCATCCCTCAAACACTCACGATTACGTTTATTCTCTGGGTGCAA
TTAATTTCACGGGTGATGAGCCAGTTATTATTACTTGCGATGGTCCTTCC
TTAGGTGGTTTTGTGTGTCAAGCTGTTGTCCCAGAAGCAGAACTGTGGAA
GGTTGGACAGGTTAAACCCGGTGATTCCATTCAGTTTGTGCCACTTTCTT
ACGAAAGCTCGAGATCCTTAAAGGAATCTCAGGATGTTGCAATTAAATCA
TTGGATGGTACTAAGTTAAGGCGCTTAGACTCTGTTTCAATTTTACCATC
ATTCGAAACGCCTATTCTTGCACAAATGGAAAAAGTGAATGAGCTTTCAC
CAAAGGTTGTATACAGACAAGCAGGTGATCGTTATGTTTTGGTGGAATAC
GGTGATAATGAAATGAATTTTAATATTTCCTATAGAATTGAATGCCTGAT
CTCCCTTGTGAAAAAGAATAAGACTATTGGTATTGTTGAAATGTCCCAAG
GTGTTAGATCTGTATTGATAGAATTTGATGGTTACAAAGTCACTCAAAAA
GAATTGCTTAAAGTATTGGTGGCATATGAAACAGAAATCCAGTTTGATGA
AAATTGGAACATAACTTCTAATATAATAAGATTACCGATGGCTTTCGAAG
ACTCGAAGACTTTGGCATGTGTTCAAAGGTATCAAGAAACAATTCGTTCG
TCTGCTCCATGGTTGCCAAATAACGTTGATTTCATTGCCAATGTAAATGG
AATTTCAAGGAATGAAGTTTATGATATGTTGTATTCTGCCAGATTTATGG
TTTTAGGTTTAGGTGATGTCTTCCTAGGGTCGCCTTGTACTGTTCCATTA
GATCCTCGTCACAGATTTTTGGGAAGCAAGTACAACCCAAGTAGAACATA
TACAGAAAGAGGTGCAGTCGGTATTGGCGGTATGTATATGTGCATATATG
CTGCTAACAGTCCTGGTGGGTACCAATTAGTGGGTAGAACAATACCAATT
TGGGACAAACTATGTCTGGCCGCATCTTCTGAGGTTCCGTGGTTGATGAA
CCCATTTGACCAAGTCGAATTTTACCCAGTTTCTGAAGAAGATTTGGATA
AAATGACTGAAGATTGTGATAATGGTGTTTATAAAGTCAATATCGAAAAG
AGTGTTTTTGATCATCAAGAATACTTGAGATGGATCAACGCAAACAAAGA
TTCCATCACAGCATTCCAGGAGGGCCAGCTTGGTGAAAGAGCAGAGGAAT
TTGCCAAATTGATTCAAAATGCAAACTCTGAACTAAAAGAAAGTGTCACA
GTCAAACCTGACGAGGAAGAAGACTTCCCAGAAGGTGCAGAAATTGTATA
TTCTGAGTATTCTGGGCGTTTTTGGAAATCCATAGCATCTGTCGGAGATG
TTATTGAAGCAGGTCAAGGGCTACTAATTATTGAAGCCATGAAAGCGGAA
ATGATTATATCCGCTCCTAAATCGGGTAAGATTATCAAGATTTGCCATGG
CAATGGTGATATGGTTGATTCTGGTGACATAGTGGCCGTCATAGAGACAT
TGGCATGA

Protein sequence

MTVSSDTTAEISLGWSTQDWIDFHKSSSSQASLRLLESLLDSQNVAPVDN
AWISLISKENLLHQFQILKSRENKETLPLYGVPIAVKDNIDVRGLRTTAA
CPSFAYEPSKDSKVVELLRNAGAIIVGKTNLDQFATGLVGTRSPYGKTPC
AFSKEHVSGGSSAGSASVVARGIVPIALGTDTAGSGRVPAALNNLIGLKP
TKGVFSCQGVVPACKSLDCVSIFALNLSDAERCFRIMCQPDPDNDEYSRP
YVSNPLKNFSSNVTIAIPKNIPWYGETKNPVLFSNAVENLSRTGANVIEI
DFEPLLELARCLYEGTWVAERYQAIQSFLDSKPPKESLDPTVISIIEGAK
KYSAVDCFSFEYKRQGILQKVRRLLESVDVLCVPTCPLNPTMQQVADEPV
LVNSRQGTWTNFVNLADLAALAVPAGFRDDGLPNGITLIGKKFTDYALLE
LANRYFQNMFPNGSRTYGTFTSSSVKPANDQLVGPDYDPSTSIKLAVVGA
HLKGLPLHWQLEKVNATYLCTTKTSKAYQLFALPKNGPVLKPGLRRVQDS
NGSQIELEVYSVPKELFGAFISMVPEPLGIGSVELESGEWIKSFICEESG
YKAKGTVDITKYGGFRAYFEMLKKKESQKKKLFDTVLIANRGEIAVRIIK
TLKKLGIRSVAVYSDPDKYSQHVTDADVSVPLHGTTAAQTYLDMNKIIDA
AKQTNAQAIIPGYGFLSENADFSDACTSAGITFVGPSGDIIRGLGLKHSA
RQIAQKAGVPLVPGSLLITSVEEAKKVAAELEYPVMVKSTAGGGGIGLQK
VDSEEDIEHIFETVKHQGETFFGDAGVFLERFIENARHVEVQLMGDGFGK
AIALGERDCSLQRRNQKVIEETPAPNLPEKTRLALRKAAESLGSLLNYKC
AGTVEFIYDEKKDEFYFLEVNTRLQVEHPITEMVTGLDLVEWMIRIAAND
APDFDSTKVEVNGVSMEARLYAENPLKNFRPSPGLLVDVKFPDWARVDTW
VKKGTNISPEYDPTLAKIIVHGKDRDDAISKLNQALEETKVYGCITNIDY
LKSIITSDFFAKAKVSTNILNSYQYEPTAIEITLPGAHTSIQDYPGRVGY
WRIGVPPSGPMDAYSFRLANRIVGNDYRTPAIEVTLTGPSIVFHCETVIA
ITGGTALCTLDGQEIPQHKPVEVKRGSTLSIGKLTSGCRAYLGIRGGIDV
PKYLGSYSTFTLGNVGGYNGRVLKLGDVLFLPSNEENKSVECLPQNIPQS
LIPQISETKEWRIGVTCGPHGSPDFFKPESIEEFFSEKWKVHYNSNRFGV
RLIGPKPKWARSNGGEGGMHPSNTHDYVYSLGAINFTGDEPVIITCDGPS
LGGFVCQAVVPEAELWKVGQVKPGDSIQFVPLSYESSRSLKESQDVAIKS
LDGTKLRRLDSVSILPSFETPILAQMEKVNELSPKVVYRQAGDRYVLVEY
GDNEMNFNISYRIECLISLVKKNKTIGIVEMSQGVRSVLIEFDGYKVTQK
ELLKVLVAYETEIQFDENWNITSNIIRLPMAFEDSKTLACVQRYQETIRS
SAPWLPNNVDFIANVNGISRNEVYDMLYSARFMVLGLGDVFLGSPCTVPL
DPRHRFLGSKYNPSRTYTERGAVGIGGMYMCIYAANSPGGYQLVGRTIPI
WDKLCLAASSEVPWLMNPFDQVEFYPVSEEDLDKMTEDCDNGVYKVNIEK
SVFDHQEYLRWINANKDSITAFQEGQLGERAEEFAKLIQNANSELKESVT
VKPDEEEDFPEGAEIVYSEYSGRFWKSIASVGDVIEAGQGLLIIEAMKAE
MIISAPKSGKIIKICHGNGDMVDSGDIVAVIETLA*

Appendix 4: Analysis of global gene expression patterns by DNA microarrays

Comparison of ECMo01 to its parental strain (Davis 522) 24 hours after inoculation in Chardonnay grape must

Analysis of DNA array data

DNA micro-arrays were done in triplicate with independently grown cells using S98 GeneChips (Affymetrix, Santa Clara, CA). Data were analyzed using MASv5.0 and DMT (Affymetrix, Santa Clara, CA). All tenable parameters were set to default values (Affymetrix Statistical Algorithm Reference Guide, Affymetrix, Santa Clara, CA). Only protein coding ORF included in the analysis were linked to their gene ontology (GO) annotations using the "orf_geneontology.tab" table (http://www.yeastgenome.org/gene_list.shtml). Nine comparative analyses were performed using Davis 522 as baseline. Genes with change p-values of <0.003 (genes with an increased call) or >0.997 (genes with a decreased call), were considered reproducible and statistically significant. The average of the Signal Log (base 2) Ratio (SLR) values were used to calculate the fold change.

Results and Discussion

When comparing the transcriptome of ECMo01 to its parental strain only *DUR1,2* (17.2-fold) is up-regulated more than two-fold after 24 hours (Table A1). Whereas 4 genes are down-regulated more than two-fold in ECMo01 at the same time point (Table A1). These four genes are involved in different biological processes.

Table A1. Genes that were expressed at least two-fold differently when comparing the transcriptome of ECMo01 to the transcriptome of Davis 522.

Gene	ORF	Biological process	Ave SLR[a]	Fold Change[b]
DUR1,2	YBR208W	Urea metabolism	4.1	+17.2
THI4	YGR144W	Thiamine biosynthesis	-2.6	-6.0
PET18	YCR020C	Mitochondrion organization and biogenesis	-2.1	-4.2
RPL9B	YNL067W	Protein biosynthesis	-1.5	-2.9
URA3	YEL021W	Pyrimidine base biosynthesis	-0.7	-1.6

a SLR - Signal Log Ratio
b Fold Change - Calculated from SLR

The *THI4* gene encodes for a protein involved in thiamine biosynthesis. *PET18* encodes for a protein required for respiratory growth and stability of the mitochondrial genome. Whereas the *RPL9B* gene encodes for a component of the large (60S) ribosomal subunit. The down regulation of *URA3* is expected since the *DUR1,2* cassette was integrated into the *URA3* locus. This resulted in one fewer

copy of *URA3* in the genome of ECMo01. The fact that only four genes were down-regulated suggests that there is no specific response, at transcriptional level, to the genetic modification in ECMo01. We believe that the micro-array data confirms the substantial equivalence of ECMo01 to Davis 522.

Appendix 5: Putative open reading frames as a result of the integration event

ORF1 (5'*ura3* truncated)

Nucleotide sequence

ATGTCGAAAGCTACATATAAGGAACGTGCTGCTACTCATCCTAGTCCTGT
TGCTGCCAAGCTATTTAATATCATGCACGAAAAGCAAACAAACTTGTGTG
CTTCATTGGATGTTCGTACCACCAAGGAATTACTGGAGTTAGTTGAAGCA
TTAGGTCCCAAAATTTGTTTACTAAAAACACATGTGGATATCTTGACTGA
TTTTTCCATGGAGGGCACAGTTAAGCCGCTAAAGGCATTATCCGCCAAGT
ACAATTTTTTACTCTTCGAAGACAGAAAATTTGCTGACATTGGTAATACA
GTCAAATTGCAGCACTCTGCGGGTGTATACAGAATAGCAGAATGGGCAGA
CATTACGAATGCACACGGTGTGGTGGGCCCAGGTATTGTTAGCGGTTTGA
AGCAGGCGGCGGAAGAAGTAGGTACCGCGGCCGCAAAAGCTTTCTAA

Putative protein sequence

MSKATYKERAATHPSPVAAKLFNIMHEKQTNLCASLDVRTTKELLELVEA
LGPKICLLKTHVDILTDFSMEGTVKPLKALSAKYNFLLFEDRKFADIGNT
VKLQHSAGVYRIAEWADITNAHGVVGPGIVSGLKQAAEEVGTAAAKAF*

ORF2 (5'*adp1* truncated)

Nucleotide sequence

ATGGGAAGTCATCGACGTTATCTCTACTATAGTATATTATCATTTCTATT
ATTATCCTGCTCAGTGGTACTTGCAAAACAAGATAAGACCCCATTCTTTG
AAGGTACTTCTTCGAAAAATTCGCGTCTAACTGCACAAGATAAGGGCAAT
GATACATGCCCGCCATGTTTTAATTGTATGCTACCTATTTTTGAATGCAA
ACAGTTTTCTGAATGCAATTCGTACACTGGTAGATGTGAGTGTATAGAAG
GGTTTGCAGGTGATGATTGCTCTCTGCCCCTCTGTGGCGGTCTATCACCG
GATGAAAGCGGTAATAAGGATCGTCCCATAAGAGCACAAAATGACACCTG
TCATTGTGATAACGGATGGGGAGGGATCAATTGTGACGTTTGTCAAGAAG
ATTTTGTCTGTGATGCGTTCATGCCTGATCCTAGTATTAAGGGGACATGT
TATAAGAATGGTATGATTGTAGATAAAGTATTTTCAGGTTGTAATGTGAC
CAATGAGAAAATTCTACAGATTTTGAACGGCAAAATACCACAAATTACAT
TTGCCTGTGATAAACCTAATCAAGAATGTAATTTTCAGTTTTGGATAGAT
CAGTTAGAAAGCTTTTGCGGCCGCGGTACCTACTTCTTCCGCCGCCTGCT
TCAAACCGCTAACAATACCTGGGCCCACCACACCGTGTGCATTCGTAATG
TCTGCCCATTCTGCTATTCTGTATACACCCGCAGAGTGCTGCAATTTGAC
TGTATTACCAATGTCAGCAAATTTTCTGTCTTCGAAGAGTAAA

BEST ORIGINAL COPY

Putative protein sequence

MGSHRRYLYYSILSFLLLSCSVVLAKQDKTPFFEGTSSKNSRLTAQDKGN
DTCPPCFNCMLPIFECKQFSECNSYTGRCECIEGFAGDDCSLPLCGGLSP
DESGNKDRPIRAQNDTCHCDNGWGGINCDVCQEDFVCDAFMPDPSIKGTC
YKNGMIVDKVFSGCNVTNEKILQILNGKIPQITFACDKPNQECNFQFWID
QLESFCGRGTYFFRRLLQTANNTWAHHTVCIRNVCPFCYSVYTRRVLQFD
CITNVSKFSVFEE*

000091

ORF3 (3'*ura3* truncated)

Nucleotide sequence

ATGTTAGCAGAATTGTCATGCAAGGGCTCCCTAGCTACTGGAGAATATAC
TAAGGGTACTGTTGACATTGCGAAGAGCGACAAAGATTTTGTTATCGGCT
TTATTGCTCAAAGAGACATGGGTGGAAGAGATGAAGGTTACGATTGGTTG
ATTATGACACCCGGTGTGGGTTTAGATGACAAGGGAGACGCATTGGGTCA
ACAGTATAGAACCGTGGATGATGTGGTCTCTACAGGATCTGACATTATTA
TTGTTGGAAGAGGACTATTTGCAAAGGGAAGGGATGCTAAGGTAGAGGGT
GAACGTTACAGAAAAGCAGGCTGGGAAGCATATTTGAGAAGATGCGGCCA
GCAAAACTAA

Putative protein sequence

MLAELSCKGSLATGEYTKGTVDIAKSDKDFVIGFIAQRDMGGRDEGYDWL
IMTPGVGLDDKGDALGQQYRTVDDVVSTGSDIIIVGRGLFAKGRDAKVEG
ERYRKAGWEAYLRRCGQQN*

Appendix 6: Determination of ethyl carbamate in Wine by Solid-Phase Microextraction and Gas Chromatography/Mass Spectrometry

Equipment

Solid Phase Microextraction (SPME) fibers and fiber holder were purchased from Supelco (Bellefonte, PA, USA).

Fiber type: 65um Carbowax/Divinylbenzene (CW/DVB)

Gas Chromatograph: Agilent 6890N GC interfaced to a 5973N Mass Selective Detector

Column: 60m x 0.25mm ID, 0.25um thickness DBWAX fused silica open tubular column (J&W Scientific, Folsom, CA, USA).

Carrier Gas: Ultra High Purity Helium @ 36cm/s, constant flow.

Injector temp: 250C

Transfer line temp: 250C

Oven temp: 70C(2min) -> 180C (3min) @ 8C/min -> 220C (15min) @ 20C/min, for a total run time of 35.75 min

Injection mode: splitless for 5min (purge flow: 5mL/min, purge time: 5min)

MS: Operated in Selected Ion Monitoring (SIM) mode with electron impact ionization.

MS quad temp: 150C

MS source temp: 230C

Delay: 8min

Ions monitored: 44, 62, 74, 89 (dwell time 100msec). Mass 62 was used for quantification and the others for confirmation.

Chemicals & Solutions:

Ethyl carbamate was purchased from Sigma-Aldrich (Milwaukee, WI, USA) and was used as standard for calibration curve.

Synthetic wine solution: distilled H2O containing 12% (v/v) ethanol and 1mM tartaric acid at a pH 3.1

Ethyl carbamate Std. stock solution was prepared at 0.1mg/mL in synthetic wine solution.

The calibration standards were prepared with ethyl carbamate concentration of 5, 10, 20, 40, 90 ug/L in synthetic wine solution. The solutions were stored in the refrigerator.

5 References

Ari'Izumi K., Suzuki Y., Kato I., Yagi Y., Otsuka K., Sato M. Winemaking from Koshu variety by the *sur lie* method : change in the content of nitrogen compounds. *Am J Enol Vitic.* 1994. **45 (3)**:312-318.

Bechet, J., M. Grenson and J.M. Waime. Mutations affecting the repressibility of arginine biosynthetic enzymes in *Saccharomyces cerevisiae*. *Eur. J. Biochem*. 1970. **12**: 31-39.

Charpentier C., Feuillat M. Yeast autolysis. In : *Wine Microbiology and Biotechnology.* Fleet G.H. (ed). Reading : Hardwood Academic. 1993. 225-242.

Chisholm G, Cooper TG. Isolation and characterization of mutants that produce the allantoin-degrading enzymes constitutively in *Saccharomyces cerevisiae*. *Mol Cell Biol*. 1982. **Sep; 2(9):**1088-95.

Commission des Communautés Européennes. Rapports du comité scientifique de l'alimentation humaine (vingt-septième série). EUR **14181** FR. p 15-22.

Coulon, J. Ph.D. Thesis. 2001. University of Bordeaux, France.

Crous J.M., Pretorius I.S., van Zyl W.H. Cloning and expression of an *Aspergillus kawachii* endo-1,4-□-xylanase gene in *Saccharomyces cerevisiae. Curr Genet*. 1995. **28**: 467-473.

Cunningham T.S., Cooper T.G. Expression of the *DAL80* gene, whose product is homologous to the GATA factors and is a negative regulator of multiple nitrogen catabolic genes in *Saccharomyces cerevisiae*, is sensitive to nitrogen catabolic repression. *Mol Cell Biol*. 1991. **11 (12)**: 6205-6215.

Daudt, C.E., Ough, C.S., Stevens, D. and Herraiz, T. Investigations into ethyl carbamate n-propyl carbamate, and urea in fortified wines. *Am. J. Enol. Vitic.* 1992. **43(4)**: 318-322.

Dawes H., Boyes S., Keene J., Heatherbell D. Protein instability of wines : influence of protein isoelectric point. *Am J Enol Vitic*. 1994. **45 (3)**:319-325.

Dunn, B., R.P. Levine, G. Sherlock. Microarray karyotyping of commercial wine yeast strains reveals shared, as well as unique, genomic signatures. *BMC Genomics*. 2005. Apr 16; **6 (1)**:53.

Fornairon-Bonnefond C., Camarasa C., Moutonnet M., Salmon J-M. New trends on yeast autolysis and wine ageing on lees : a bibliographic review. *Int J Wine Sci*. 2001. **36 (2)**:49-69.

Gatignol A., Dassain M., Tiraby G. Cloning of *Saccharomyces cerevisiae* promoters using a probe vector based on Phleomycin resistance. *Gene*. 1990. **91 (1)**:35-41.

Genbauffe, F.S. and T.G. Cooper. The urea amidolyase (*DUR1,2*) gene of *Saccharomyces cerevesiae*. *DNA sequence.* 1991. **2**: 19-32.

Grenson, M., M. Mousset, J.M. Wiame and J. Bechet. Multiplicity of the amino acid permeases in *Saccharomyces cerevisiae*. I. Evidence for a speicfic arginine-transporting system. *Biochim. Biophys. Acta.* 1966. **127**: 325-338.

Hill J.E., Myers A.M., Koerner T.J., Tzagoloff A. Yeast/*E.coli* shuttle vectors with multiple unique restriction sites. *Yeast*. 1986. **2**:163-197.

International Agency for Research on Cancer. 1974, 1982, 1987. www.iarc.fr.

Kreger-van-Rij N.J.W. (Ed). *The yeasts : a taxonomy study*. 3rd ed. Amsterdam : Elsevier Science Publishers B.V. 1984. 1082pp.

Kurtzman C.P., Fell J.W. (Eds). *The yeasts. A taxonomic study*. 4th edition. Amsterdam : Elsevier Science B.V. 1998. 1055 pp.

Lilly M., Lambrechts M.G., Pretorius I.S. Effect of increased yeast alcohol acetyltransferase activity on flavor profiles of wines and distillates. *Appl Environ Microbiol*. 2000. **66 (2)**:744-753.

Lodder J. (Ed). *The yeasts : a taxonomic study*. 2nd ed. Amsterdam : North Holland Publishing Company. 1970. 1385 pp.

Lodder J., Kreger-van-Rij N.J.W. (Eds). *The yeasts : a taxonomic study*. 2nd ed. Amsterdam : North Holland Publishing Company. 1967. 713 pp.

Martinez-Rodriguez A.J., Polo M. Characterization of the nitrogen compounds released during yeast autolysis in a model wine system. *Am Chem Soc*. 2000. **48 (4)**: 1081-1085.

Messing J., Vieira J. A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. *Gene*. 1982. **19 (3)**:269-276.

Mortimer R.K., Johnston J.R. Genealogy of principal strains of the yeast genetic stock center. *Genetics*. 1986. **113**: 35-43.

Office International de la vigne et du vin. Codex Œnologique International, Edition 2002. Paris, OIV.

Olson M.V., Dutchik J.E., Graham M.Y., Brodeur G.M., Helms C., Frank M., MacCollin M., Scheinman R., Frank T. Random-clone strategy for genomic restriction mapping in yeast. *Proc Natl Acad Sci USA*. 1986. **83**: 7826-7830.

Puig S., Perez-Ortin J.E. Expression levels and patterns of glycolytic yeast during wine fermentation. *System Appl Microbiol*. 2000a. **23**: 300-303.

Puig S., Perez-Ortin J.E. Stress response and expression patterns in wine fermentations of yeast genes induced at the diauxic shift. *Yeast*. 2000b. **16**: 139-148.

Reed R., Nagodawithana T.W. *Yeast Technology*. 2nd edition. New York : Van Nostrand Reinhold. 1991. 454 pp.

Ribéreau-Gayon P., Dubourdieu D., Donèche B., Lonvaud A. Traité d'Œnologie 1. Microbiologie du vin. Vinifications. Paris : Dunod. 1998a. 617 pp.

Ribéreau-Gayon P., Glories Y., Maujean A., Dubourdieu D. Traité d'Œnologie 2. Chimie du vin. Stabilisation et traitements. Paris : Dunod. 1998b. 519pp.

Roon, R.J. and Levenburg, B. Urea Amidolyase. *J. Biol. Chem*. 1972. **247(13)**: 4107-4113.

Schuller D, Valero E, Dequin S, Casal M. Survey of molecular methods for the typing of wine yeast strains. *FEMS Microbiol Lett*. 2004. **Feb 9; 231(1)**:19-26.

Short J.M., Fernandez J.M., Sorge J.A., Huse W.D. Lamba ZAP : a bacteriophage lamba expression vector with *in vivo* excision properties. *Nucleic Acids Res*. 1988. **16:** 7583-7600.

Subden, R.E. Wine yeast: Selection and modification. In: Panchal C.J. (Ed.), Yeast strain selection, Marcel Dekker Inc., NY, 1990. pp. 113-137.

The Regents of the University of California. Malolactic gene. EP0103399. March 21, 1984.

U.S. Bureau of Alcohol, Tobacco and Firearms. Code of Federal Regulations, Title 27, Parts 1 to 199.

U.S. Environmental Protection Agency. 40 CFR Parts 700, 720, 721, 723, and 725. Microbial products of biotechnology; Final regulation under the Toxic Substances Control Act, Final Rule. *Federal Register*. 1997. **62 (70):** 17909-17958.

U.S. Environmental Protection Agency. Final decision document: TSCA section 5 (H) (4) exemption for *Saccharomyces cerevisiae*. 1996.

Velculescu V.E., Zhang L., Zhou W., Vogelstein J., Basrai M.A., Bassett D.E., Hieter P., Vogelstein B., Kinzler K.W. Characterisation of the yeast transcriptrome. *Cell*. 1997. **88 (2):** 243-251.

Viera J. , Messing J. The pUC plasmids, an M13mp7- derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene*. 1982. **19**: 259-268.

Volschenk H., Viljoen M., Grobler J., Bauer F., Lonvaud-Funel A., Denayrolles M., Subden R.E., van Vuuren H.J.J. Malolactic fermentation in grape must by a genetically engineered strain of *Saccharomyces cerevisiae. Am J Enol Vitic*. 1997a. **48 (2):** 193-197.

Volschenk H., Viljoen M., Grobler J., Petzold B., Bauer F., Subden R.E., Young R.A., Lonvaud A., Denayrolles M. and van Vuuren H.J.J. Engineering pathways for malate degradation in *Saccharomyces cerevisiae. Nature Biotech*. 1997b. **15:** 253-257.

Volschenk, H., M. Viljoen-Bloom, J. Van Staden, J. Husnik and H.J.J. van Vuuren. Genetic engineering of an industrial strain of Saccharomyces bayanus for l malic acid degradation via an efficient malo ethanolic pathway. *South African Journal for Enology and Viticulture*. 2004. **25**:63-73.

Whiton, R.S. and Zoecklein, B.W. Determination of Ethyl Carbamate in Wine by Solid-Phase Microextraction and Gas Chromatography/Mass Spectrometry. *American Journal of Enology and Viticulture*. 2002. **53(1)**: 60-63.

Zimmerli, B and Schlatter, J. Ethyl carbamate: analytical methodology, occurrence, formation, biological activity and risk assessment. *Mut. Res.* 1991. **259**: 325-350.

Submission End

000098