Original Submission

GR

LESAFFRE YEAST CORPORATION

Lesaffre Yeast Corporation
433 East Michigan Street
Milwaukee, WI 53202
U.S.A.

January, 2, 2003

RECEIVED
JAN 7 2003
OFFICE OF
FOOD ADDITIVE SAFETY

Office of Food Additive Safety (HFS-200)
Center for Food Safety and Applied Nutrition
Food and Drug Administration
5100 Paint Branch Parkway
College Park
MD 20740-3835
U.S.A.

RE: GRAS notification for the use of a modified yeast in winemaking

Dear Sir or Madam,

Lesaffre Yeast Corporation submits the attached information to the Food and Drug Administration pursuant to the Federal Food, Drug and Cosmetic Act in anticipation to the marketing of a yeast strain modified through biotechnology for winemaking.

Lesaffre Yeast Corporation intends to market this modified yeast strain called ML01 in the form of Wine Active Dry Yeast. This yeast, because of its genetic modification, can be considered as a new substance, but would be exempt from the definition of food additive, and thus from pre-market approval, if its safety was generally recognized by qualified experts.

Lesaffre Yeast Corporation is submitting information regarding the use of the modified yeast ML01 in winemaking in the form of a GRAS notice, following the FDA proposed rule issued in the Federal Register (FR) of April 17, 1997.

The submitted notice includes:

- **A statement that according to Lesaffre Yeast Corporation, called hereafter Lesaffre, the active dry ML01 yeast is GRAS, through scientific procedures, for use as a starter culture in grape must fermentation. Lesaffre's submission is that from data concerning the ML01 strain, this genetically enhanced yeast is substantially equivalent to the host strain it is derived from, except for the introduction of a metabolic pathway catalyzing the conversion of L-malate to L-lactate. No other metabolic difference is detected between the modified and the host strain, whether the studies are carried out during fed batch aerobic cultures or during anaerobic alcoholic fermentations. The submitted statement is based on a safety evaluation of the host microorganism, of the foreign genetic material used to construct the modified yeast strain, of the modified yeast strain ML01 and of the wine obtained from this modified yeast strain.**

- An analysis of the intended use of the substance
- A definition of the consumers of the notified substance
- A statement that Lesaffre Yeast Corporation undertakes that all data and information used to set up the basis for the GRAS determination for the use of the ML01 yeast strain in winemaking will be made available for the FDA to review and copy at the following address:

  Lesaffre International
  Attention Dr. Didier Colavizza
  147 rue Gabriel Péri
  59703 Marcq-en-Baroeul cedex
  France

Alternatively, Lesaffre undertakes to send this data and information to the FDA upon FDA's first request.

The scientific manager with whom FDA can correspond is Dr. Didier COLAVIZZA (djc@lesaffre.fr).

At the best of Lesaffre knowledge, the submitted notice is a representative and balanced notice that includes unfavorable as well as favorable information, known to Lesaffre as pertinent for the evaluation of the safety and functionality of the notified substance. The submitted GRAS notice is signed on the text page (p33) by the Project Director (Robert "Bob" Biwersi) and the Scientific Manager (Didier Colavizza).

The undertakings hereabove copied consist in Lesaffre irrevocable undertakings.

Should you have any questions, please do not hesitate to contact me at 414-615-4085.

Sincerely,

Robert Biwersi
Director, Quality Assurance
414-615-4085

433 East Michigan Street • Milwaukee, WI 53202 • (877) 677-7000

GRAS notification for the use of a modified yeast strain in winemaking

Prepared for

Lesaffre Yeast Corporation,

Milwaukee WI

GRAS notice

1- GRAS exemption claim 1

1.1. Name and address of the notifier 1

1.2. Common or usual name of the substance that is the subject of the GRAS exemption claim 1

1.3. Applicable conditions of use of the notified substance 1

1.3.1. Purpose for which the substance is used *2*

1.3.2. Foods in which the substance is to be used *2*

1.3.3. Levels of use of the ML01 wine active dry yeast in grape must *3*

1.3.4. Population expected to consume the substance *3*

1.3.5. Technical effect *3*

1.4. The basis for the GRAS determination (see attached letter) 4

1.5. Statement that the data and information that are the basis for the notifiers GRAS determination are available for the Food and Drug Administration review and copy (see attached letter) 4

2- Detailed information about the identity of the notified substance 4

2.1. Host microorganism 4

2.1.1. Taxonomy *4*

2.1.2. Characteristics *5*

2.2. Donor organisms 5

2.2.1. Taxomomy *5*

2.2.2. Genetic material from donor organisms *6*

2.3. The modified microorganism 7

2.3.1. Final construct used in the integration strategy *7*

2.3.1.1. Construction strategy 7

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 *9*

2.3.3.1. The loss of the antibiotic resistance genes and of other plasmid sequences 9

2.3.3.2. Genetic analysis by Southern blot 10

2.3.3.3. Genetic analysis by pulsed field gel electrophoresis (PFGE) followed by hybridization. 11

2.3.3.4. Genetic stability of the ML01 yeast strain 11

2.3.4. DNA sequencing of the integrated locus *11*

2.3.5. Absence of difference in 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 13

4- Determination that the use of the wine active dry yeast obtained from the ML01 yeast strain in the fermentation of grape must is exempt from the pre-market approval requirements of the act because such use is GRAS. 13

4.1. Safety assessment of the host strain 13

4.2. Safety assessment of the genetic material used to construct the modified organism 14

4.2.1. Foreign genetic material source and product *14*

000005

4.2.1.1. The *mae1* gene 14
4.2.1.2. The *mleA* gene 16
4.2.2. DNA yeast material : the PGK1 promoter, terminator and URA3 sequences 17
4.2.3. Construction of the modified organism 17
4.2.3.1. Vector used as source of the integration cassette 17
4.2.3.2. Transformation of the host strain with the integration cassette 17
4.3. Safety assessment of the modified organism 18
4.3.1.Characterization of the transformation event 18
4.3.2. Consequence of the genetic modification on gene expression 18
4.3.2.1. *mae1* and *mleA* gene expression 18
4.3.2.2. Global gene expression analysis 19
4.3.3. Presence of unintended gene products as a result of the transformation event 19
4.3.4. Study of yeast proteins by two-dimension protein gels 20
4.3.5. Consequence of the genetic modification on the physiology of the ML01 strain 21
4.3.5.1. Overview of the function of the malolactic gene in lactic acid bacteria 21
4.3.5.2. Manufacture of active dry yeast during fed-batch aerobic cultures 21
4.3.5.3. Growth of ML01 and S92 strains during laboratory-scale vinifications 21
4.3.5.4. Growth and fermentation rates during fermentation of high acid -high malate containing must 21
4.3.5.5. Growth of ML01 and S92 strains during batch laboratory cultures in the presence of malate 22
4.3.5.6. Mixed cultures of the S92 host strain and the modified ML01 strain in synthetic grape juice 22
4.3.5.7. Sporulation and killer character of the modified yeast 23
4.4. Safety assessment of the product derived from the modified organism 23
4.4.1. Changes in the winemaking procedures as a consequence of the ML01 fermentation 23
4.4.2. Changes in the physiology of the modified yeast as consequence of the genetic modification during winemaking 23
4.4.3. Changes in wine composition as a consequence of the ML01 fermentation 24
4.4.3.1. Global characteristics of wine 24
4.4.3.2. Flavor modification 24
4.4.3.3. Yeast cells and release of yeast products during winemaking 25
4.4.3.4. Formation of unwanted substances in wine 28

5- References 30

000006

Figures and Tables

Figure 1 : Flow diagram illustrating the cloning strategy used to construct plasmid pJH2. 34
Table 1 : Primers used in the construction of the integration cassette 35
Table 2 : Plasmids used for the construction and integration of the integration cassette.....36
Figure 2 : Cloning of the *mae1* gene and construction of plasmid pHV3.....37
Figure 3 : Cloning of the *mleA* gene and construction of plasmid pJD4. 38
Figure 4 : Construction of plasmid pHVX2. 39
Figure 5 : Construction of plasmid pJC1. 40
Figure 6 : Construction of plasmid pHVJH1. 42
Figure 7: Construction of plasmid pJH1. 43
Figure 8: Construction of plasmid pJH2. 44
Figure 9 : Schematic representation of the introduced genetic material.....45
Table 3 : Detailed description of the nucleotides composing the integration cassette 46
Figure 10 : Loss of phleomycin resistance. 48
Figure 11: Plasmid pUT332 map and schematic representation of the probes used in Southern experiments. 49
Figure 12 : Absence of the Tn5*Ble* gene in the ML01 strain. 50
Figure 13 : Absence of the *Amp*R gene and of plasmid sequences in the ML01 strain. 51
Figure 14: Schematic representation of probes originating from plasmid pJH2 and used in hybridization experiments. 52
Figure 15 : Southern blots representing hybridization of genomic DNA with the ura3 probe and schematic representation of an integrated and non integrated *URA3* loci. 53
Figure 16: Southern blots representing hybridization of genomic DNA with the mae1 probe and schematic representation of an integrated and non integrated URA3 loci. 54
Figure 17: Southern blots representing hybridization of genomic DNA with the mleA probe and schematic representation of an integrated and non integrated URA3 loci. 55
Figure 18: Southern blot representing hybridization of genomic DNA with the pgk1 probe and schematic representation of a *PGK1* locus, an integrated and non integrated *URA3* loci. 56
Figure 19: Pulsed field gel electrophoresis of genomic DNA of the host strain and the ML01 strain and subsequent hybridization with ura3, pgk1 and mae1 probes. 57
Figure 20: Stability of the integrated genetic material in the ML01 strain. 58
Table 4 : Discrepancies found when comparing the genomic sequence of the integrated locus and previously published sequences 59
Figure 21 : Genetic pattern of the ML01 (1) and the S92 (2) yeast strains based on amplification of genomic DNA regions in between δ elements of the Ty1 retrotransposon. 60
Figure 22: Flow diagram describing the manufacture of active dry yeast. 61
Figure 23 : *mleA* and *mae1* gene expression during synthetic grape must fermentation.....62
Figure 24 : Putative open reading frames (ORFs) analysis. 64
Figure 25: Growth comparison of strains ML01 (blue) and S92 (pink) in synthetic and Chardonnay must. 67
Figure 26: Growth comparison of strains ML01 (blue) and S92 (pink) in high acid (pH 3.18) and high malate (9 grams/liter L-malate) Chardonnay must. 68

000007

Figure 27: Growth comparison of strain ML01 and strain S92 in synthetic medium containing 20 grams/liter L-malate and 5 grams/liter glucose. 69
Figure 28 : Ascospore formation by the ML01 and the S92 strains....70
Figure 29 : White winemaking practices and the use of a standard or ML01 yeast strain....71
Figure 30: Red winemaking practices and the use of a standard or ML01 yeast strain....72
Table 5 : Potential ethyl carbamate in Merlot wines....73

000008

Appendices

Appendix 1 : Integration cassette in Fasta format 74
Appendix 2 : Sequence alignment of the integrated DNA from strain ML01 and previously published sequences 78
Appendix 3 : Nucleotide sequence of the *mleA* and *mae1* genes and protein sequence of the corresponding proteins in Fasta format 91
Appendix 4 : Analysis of global gene expression by DNA arrays carried out by the University of British Columbia 95
Appendix 5 : Putative open reading frames as a result of the integration event 100
Appendix 6 : Synthetic grape must fermentations by mixed cultures of the modified ML01 strain and the host strain S92 104

000009

Gras Notice

1- GRAS EXEMPTION CLAIM

1.1. Name and address of the notifier

Lesaffre Yeast Corporation
433 East Michigan Street
Milwaukee, WI 53202

Project Director

Mr. Bob Biwersi
Director of Quality Assurance
Lesaffre Yeast Corporation
433 East Michigan Street
Milwaukee, WI 53202

Telephone : (414) 615 4085
Facsimile : (414) 615 4000

Scientific Managers

Dr. Didier Colavizza djc@lesaffre.fr
Mr. Jean-Paul Rossi jpr@lesaffre.fr
Dr. Joana Coulon

Lesaffre International
147 rue Gabriel Péri
59703 Marcq-en-Baroeul cedex
France

Telephone : +33 (0) 3 20 81 61 00
Facsimile : +33 (0) 3 20 81 62 52

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

The subject of the notice is an active dry wine yeast obtained from a *Saccharomyces cerevisiae* strain carrying two recombinant genes encoding a malate permease from *Schizosaccharomyces pombe* and a malolactic enzyme from *Oenococcus œni*. This novel yeast strain is called ML01.

1.3. Applicable conditions of use of the notified substance

Research concerning the construction of a malolactic wine yeast strain by cloning the malolactic gene of lactic acid bacteria and expressing it in *Saccharomyces cerevisiae* was started in the United States of America by researchers of the University of Davis, California (Williams *et al.*, 1984 ; US Patent 4472502, Snow *et al.* ; European Patent 0103399, The regent of the University of California) and also in the University of Guelph in Canada (Lautensach and Subden, 1984). However, the use of this malolactic gene in order for yeast strains to perform malolactic fermentation was unsuccessful due to the absence of malate permease enabling malate entry into the cell. The malate permease gene was isolated by Prof. van Vuuren from the University of Stellenbosch (South Africa) and Prof. Subden from the University of Guelph (Canada). The study presented here has been conducted in collaboration with these two Universities. Prof. van Vuuren has now moved to the University of British Columbia, Canada.

Results presented here were obtained in close collaboration with the University of British Columbia and the University of Guelph.

1.3.1. Purpose for which the substance is used

After the alcoholic fermentation, most red wines and some white wines are subjected to malolactic fermentation. This secondary fermentation, usually carried out by lactic acid bacteria, is important for deacidification, flavor modification and microbial stability of the wine. Wine deacidification takes place through the conversion of L malate (dicarboxylic acid) to L-lactate (monocarboxylic acid).

Oenococcus œni (formerly *Leuconostoc œnos*, Dicks *et al.*, 1995) is the major lactic acid bacteria responsible for malolactic fermentation of wines, but most wine lactic acid bacteria possess the malolactic enzyme which catalyses the L-malate to L-lactate bio-conversion. Some lactic acid bacteria strains also show specific metabolic pathways which can lead to the formation of undesirable compounds (Lonvaud-Funel, 1999). Because lactic acid bacteria are on one hand nutritionally very demanding and on the other hand very sensitive to the pH, temperature, sulfur dioxide content and ethanol concentration of the wine, malolactic fermentation is often unpredictable and difficult to achieve, even with the use of commercial starter cultures of lactic acid bacteria.

Because *Saccharomyces cerevisiae* is unable to efficiently degrade L-malate and that the L-malate degrading *Schizosaccharomyces pombe* yeast is unsuitable for the production of wine, we propose the use of a genetically enhanced strain of *Saccharomyces cerevisiae* to ensure a reliable malolactic fermentation. The modified strain contains the *Schizosaccharomyces pombe* malate transporter gene (*mae1*) and the *Oenococcus œni* malolactic enzyme gene (*mleA*) under the *Saccharomyces cerevisiae PGK1* promoter and terminator signals. Hence, all genetic donors are wine microorganisms. Simultaneous expression of both genes triggers complete L-malate utilization by the recombinant *Saccharomyces cerevisiae* strain ML01 during alcoholic fermentation. Therefore, malolactic fermentation and alcoholic fermentation occur simultaneously.

1.3.2. Foods in which the substance is to be used

The ML01 wine active dry yeast is intended to be used in winemaking procedures in accordance with Good Manufacturing Practices, as a yeast starter culture for grape must fermentation. Like all wine active dry yeast, the wine active dry yeast obtained with the ML01 strain is used to perform alcoholic fermentation, thus transforming grape must into wine. Simultaneously the ML01 yeast cells conduct malolactic fermentation, a process which is usually done by lactic acid bacteria after the alcoholic fermentation. The use or not of wine active dry yeast obtained with the ML01 strain depends on the winemaker's choice and relies on whether the malolactic fermentation is required for (i) deacidification, (ii) flavor modification, (iii) microbial stability or a combination of these characteristics.

- If deacidification is the major contribution wanted from malolactic fermentation, the use of the ML01 wine active dry yeast is highly recommended. The bioconversion of a dicarboxylic acid (L-malate) into a monocarboxylic acid (L-lactate) decreases the total acidity of the wine and softens its mouthfeel.
- In some wines, such as red wines with aging potential, malolactic fermentation by bacteria not only contributes to deacidification, but also to flavor modification through bacterial secondary metabolism. In this case, it is often preferable not to use the ML01 wine active dry yeast and to carry out malolactic fermentation with lactic acid bacteria.

- The ML01 wine active dry yeast can especially be used when wine microbial stability is required without the development of buttery flavors (acetoïn, diacetyl) due to lactic acid bacteria secondary metabolism, like it is often the case in white wines.

- The ML01 wine yeast strain can also be used in cases where wine stability is required as soon as possible : wine microbial stability is enhanced once the malolactic fermentation completed by removal of L-malate. Moreover, the sooner the malolactic fermentation is completed, the sooner the wine can be sulphited and bottled with a low risk of bacterial contamination.

1.3.3. Levels of use of the ML01 wine active dry yeast in grape must

Although the Bureau of Alcohol, Tobacco and Firearms (27 CFR 24.176) and the International Œnological Codex (2002) leave the winemaker free of choosing the yeast inoculation level in wine, it is recommended to use in between 0.1 and 0.2 gram/liter of wine active dry yeast for grape must fermentation. Within this inoculation range, alcoholic fermentation will be efficiently conducted, whereas a lower inoculation rate could lead to a stuck fermentation. This inoculation rate is also a guaranty for the winemaker for L-malate degradation, which will be completed within the first days of the fermentation.

On another hand, malate degradation will practically not occur if the ML01 yeast represents 1 % or less of the total yeast inoculum. Given that 1 % represents a high level of contamination, significant malolactic fermentation by yeast cannot occur when must is contaminated by the ML01 strain.

Moreover, in the case of a co-culture of the S92 host yeast cells with ML01 yeast cells, the recombinant strain will not take over the fermentation.

As a consequence, by choosing the appropriate yeast starter culture for wine fermentation, the winemaker can decide not to carry out the malolactic fermentation, or to carry out the malolactic fermentation with either the ML01 yeast strain or lactic acid bacteria.

1.3.4. Population expected to consume the substance

The population expected to consume the substance are wine consumers.

1.3.5. Technical effect

Winemaking trials were performed :

- in laboratory conditions
- in pilot scale conditions in Canada (Okanagan Valley), under the authority of the Wine Research Centre from the University of British Columbia (Vancouver)
- in pilot scale conditions in South Africa, under the authority of the University of Stellenbosch
- in 2002 on a commercial scale in Moldavia : following an assay conducted in 2001, the Department of Agriculture and Food Industry of the Moldavian Republic issued a marketing authorization for the use of wine active dry ML01 yeast in wines.

During these assays, simultaneous malolactic and alcoholic fermentation occurred in wines made with ML01. In these wines, malate was efficiently degraded in the first few days following the beginning of alcoholic fermentation, and the ML01 yeast degraded L-malate into L-lactate in stochiometric amounts. The ML01 and the host yeast (S92) showed similar fermentation kinetics, except in low pH must where malate degradation by ML01 triggered a pH increase and enhanced the fermentation rate. ML01 wines showed no organoleptic default. Expect for parameters influenced by malate degradation (pH, total acidity, richer perception of wines...), the ML01 strain did not change the wine properties. However, ML01 wines conserved better their color than wines where malolactic fermentation was performed by

lactic acid bacteria. Sensory analysis indicated that ML01 wines were considered as superior to control wines, mostly because they were fuller bodied, a characteristic which could be attributed to deacidification through malate degradation by ML01. In wines tested where L-malate degradation occurred through lactic acid bacteria, the delay and temperature required for completion of the malolactic fermentation negatively influenced wine quality.

The ML01 and the parental S92 strains have the same fermentation kinetics. The ML01 strain converts in stochiometric amounts L-malate into L-lactate, indicating that the genetic transformation doesn't disturb the endogenous L-malate metabolism. All results prove that ML01 efficiently performs malolactic fermentation with no other metabolic difference being suspected compared to the host strain.

1.4. The basis for the GRAS determination (see attached letter)

According to Lesaffre, the active dry ML01 yeast is GRAS, through scientific procedures for use as a starter culture in grape must fermentation. Lesaffre's submission is that from data concerning the ML01 strain, this genetically enhanced yeast is substantially equivalent to the host strain it is derived from, except for the introduction of a metabolic pathway catalyzing the conversion of L-malate to L-lactate. No other metabolic difference is detected between the modified and the host strain, whether the studies are carried out during fed batch aerobic cultures or during anaerobic alcoholic fermentations.

This notice includes a safety evaluation of the host microorganism, of the foreign genetic material used to construct the modified yeast strain, of the modified yeast strain and of the product derived from the modified yeast strain.

1.5. Statement that the data and information that are the basis for the notifiers GRAS determination are available for the Food and Drug Administration review and copy (see attached letter)

Lesaffre undertakes that all data and information used to set up the basis for the GRAS determination for the use of the ML01 yeast strain in winemaking will be made available for the FDA to review and copy at the following address :

Lesaffre International
Attention Dr. Didier Colavizza
47 rue Gabriel Péri
59703 Marcq-en-Baroeul cedex
France

Alternatively, data will be sent to the FDA upon FDA's first request.

2- DETAILED INFORMATION ABOUT THE IDENTITY OF THE NOTIFIED SUBSTANCE

2.1. Host microorganism

The host yeast strain is an industrial wine strain called S92 in Lesaffre strain collection.

2.1.1.Taxonomy

S92 is an isolate from the Champagne region in France and belongs to a family of very close or identical commercial strains which are commonly used as wine strains and often designed as "Prise de Mousse" strains. This strain family is considered by oenologists as *Saccharomyces bayanus* or *Saccharomyces cerevisiae var. bayanus* strains. This designation

construction strategy, a flow diagram is given Figure 1. Sequences of primers used in amplification reactions and linkers used in ligation reactions are listed Table 1. All plasmids used in successive cloning events are listed Table 2.

The *mae1* gene was isolated from a genomic library and subcloned as a *BalI-NdeI* fragment into plasmid **YEplac181**. The *mae1* gene was subcloned a second time into plasmid **pHVX2** as an *EcoRI-BglII* fragment, yielding plasmid **pHV3** (Figure 2).

The *mleA* gene was isolated from *Oenococcus œni* Lo 8413 by amplification using primers **J18** and **J19**, containing *EcoRI* and *BglII* sites respectively. The amplified fragment was cleaved with *EcoRI* and *BglII* restriction enzymes and cloned into the *EcoRI* and *BglII* sites of plasmid **pHVX2**, yielding plasmid **pJD4** (Figure 3).

Plasmid **pHVX2** was obtained by subcloning a *HindIII* fragment from plasmid **pJC1** containing the *PGK1* promoter and terminator sequences into the *HindIII* site of **YEplac181** (Figure 4). Construction of plasmid **pJC1** is summarized as follows. The *PGK1* region from *Saccharomyces cerevisiae* was isolated as a *HindIII* fragment from the lambda clone PM-5052 (ATCC 70513; Olson *et al.*, 1986) and cloned into the **pUC8** plasmid (Messing and Viera, 1982). The *PGK1* promoter (*PGK1*p) was subcloned as a *HindIII-EcoRI* DNA fragment into the *HindIII* and *EcoRI* sites of plasmid **pBluescript KS$^+$** (Short *et al.*, 1988). A *BglII* restriction site was introduced at the beginning of the *PGK1* open reading frame through site directed mutagenesis (Figure 5). This was accomplished by altering the DNA sequence with the aid of primer **J17**. The *PGK1* terminator (*PGK1*t) was subcloned as a *BglII-HindIII* fragment into the *BamHI-HindIII* restriction sites of **pUC8**, destroying the *BglII* restriction site in the process. The *PGK1*t region was cleaved at the *ClaI* site downstream from the *PGK1* gene, the overhanging ends filled in by treatment with the Klenow fragment and **linker 1** (containing a *BglII* site) inserted. The *PGK1*p was retrieved as a *HindIII-BglII* DNA fragment from **pBluescript KS$^+$** and the *PGK1*t region as a *BglII-HindIII* DNA fragment from **pUC8**. Both fragments were ligated together at the *BglII* sites (Figure 5). The resulting *PGK1*p-*PGK1*t cassette in **pUC8** was linearized at the *BglII* site and **linker 2** containing the *EcoRI*, *BglII* and *XhoI* restriction sites inserted. The assembly of vector **pJC1** was initiated by linearizing vector **YEp352** at the unique *EcoRI* restriction site, filling in the protruding ends by a Klenow treatment and ligating the resulting blunt ends together. The resulting plasmid with a destroyed *EcoRI* site was named **YEp352'**. The *PGK1*p-*PGK1*t cassette was introduced into the unique *HindIII* restriction site of **YEp352'** yielding plasmid **pJC1** (Figure 5).

The *mae1* and *mleA* expression cassettes containing the *mae1* and *mleA* genes between *PGK1*p and *PGK1*t were isolated from plasmid **pHV3** and plasmid **pJD4** by PCR amplification, respectively. Amplification of the *mae1* expression cassette used primers **J13** and **J14**, both carrying a *KpnI* site at their 5' ends. Amplification of the *mleA* cassette was performed using primers **J15** and **J16**, both carrying a *NotI* site at their 5' ends.

Simultaneously, a plasmid was constructed, plasmid **pHVJH1**, in order to clone both amplification products in between sequences homologous to the *URA3* locus. For construction of plasmid **pHVJH1**, the **YEp352** plasmid (ATCC 37673 ; Hill *et al.*, 1986) was cleaved at the unique *KpnI* restriction site, Klenow-treated and ligated (Figure 6). This resulted in a plasmid in which the *KpnI* site had been destroyed (**YEp352Δ*KpnI***). A 925 bp fragment containing part of the *URA3* open reading frame and the 5' sequence of the same gene was amplified by polymerase chain reaction with primers **J9** and **J10** containing *XbaI* and *SrfI* sites at the 5'-end and a *KpnI* site at the 3'-end, to yield a 944 bp fragment (Figure 6). A 933 bp fragment containing part of the *URA3* open reading frame and the 3' sequence of the same gene was PCR amplified from *Saccharomyces cerevisiae* with primers **J11** and **J12** which contain *KpnI* and *NotI* sites at the 5'-end and *SrfI* and *XbaI* sites at the 3'-end, to yield a 959 bp fragment. The 945 bp and the 959 bp amplified fragments were digested with restriction enzyme *KpnI* and ligated to yield a *URA3* fragment with *XbaI*, *KpnI*, *SrfI* and *NotI* restriction sites. Following PCR amplification with primers **J9** and **J12**, the modified *URA3* fragment was digested with *XbaI* and cloned into plasmid **YEp352Δ*KpnI***, yielding plasmid **pHVJH1** (Figure 6).

000015

relies on early identification tests based on biochemical criteria. Indeed, these strains poorly ferment galactose, which lead oenologists to distinguish them from the other *Saccharomyces cerevisiae* wine yeast strains. As for many other yeast strains, these strains have been incorporated into the *Saccharomyces bayanus* species without this classification being thoroughly checked (Nguyen *et al.*, 2000). It is now admitted that the galactose fermenting capacity is not a significative differentiation criteria between *Saccharomyces cerevisiae* and *Saccharomyces bayanus* (Barnett, 1992 ; Kurtzman and Fell, 1998). Identification methods based on molecular analysis of the yeast genome allowed scientists to classify this family of strains as strictly belonging to the *Saccharomyces cerevisiae* species (Johnston *et al.*, 2000 ; Ribéreau-Gayon *et al.*, 1998a ; Masneuf *et al.*, 1996). Nevertheless, the winemaking industry continues to describe these strains as *Saccharomyces bayanus* or *Saccharomyces cerevisiae var. bayanus*, hereby designating galactose non-fermenting *Saccharomyces cerevisiae* wine strains.

Therefore, the S92 strain belongs to the *Saccharomyces cerevisiae* species.

Since millenaries, *Saccharomyces cerevisiae* is associated with fermented alcoholic beverages, in particular the production of beer and wine. It is also the only species employed in the production of baker's yeast. Throughout the years and as taxonomy rules evolved, the classification of *Saccharomyces cerevisiae* has suffered many changes. The historical survey of the species is given in standard compendia (Lodder and Kreger-van-Rij, 1967 ; Lodder, 1970 ; Kreger-van Rij, 1984 ; Kurtzman and Fell, 1998).

2.1.2. Characteristics

The S92 wine yeast strain is probably diploid, sporulates and is a K2 killer wine yeast. The S92 wine yeast strain is commonly used in industry in order to obtain wine active dry yeast (WADY). The multiplication of S92 vegetative cells during their production in order to obtain WADY and, after rehydration, during fermentation of must, occurs through budding (no sexual multiplication).

2.2. Donor organisms

2.2.1. Taxomomy

The following organisms contributed DNA to the modified microorganism ML01.

Schizosaccharomyces pombe

This yeast species was first isolated from African beer. Since then, it has frequently been cultured from sugar containing products in tropical and sub-tropical countries, and from grape must and cider in moderate climates (Lodder, 1970). Osterwalter (1924) isolated a strain from sulfurized grape juice. Leupold (1950) named it "972h⁻" (ATCC 24843 ; NCYC 1354). It is a heterothallic strain which has been extensively used in genetic research. Many mutants are derived from that strain. The one used in this study as a genetic material donor is an auxotrophic mutant of 972h⁻, carrying a leu1-32 mutation (NCYC 1913).

Oenococcus œni

This bacterial species, formerly designed *Leuconostoc œnos* (Dicks *et al.*, 1995) has only been isolated from wines and related habitats such as wineries and vineyards (Williams *et al.*, 1989). It is a facultative aerobe which grows better in low pH media and is tolerant to high concentrations of ethanol and SO_2. It is the preferred and most common lactic acid bacteria for the malolactic fermentation of wines (Lonvaud-Funel, 1999). The Lo 8413 strain used as a genetic material donor was isolated at the Œnology Faculty of Bordeaux from the University Bordeaux2-Victor Segalen, France.

Saccharomyces cerevisiae

General taxonomy of this species is given elsewhere in this notice (§.2.1.1.). Two different strains were used for DNA isolation. One of them is the strain AB972 (Olson *et al.*, 1986 ; ATCC 204511) which traces directly to strain S288C (Mortimer and Johnston, 1986 ; ATCC 204508). It has been shown that strain S288C has 6 progenitors, 3 of them being baker's yeast strains (Mortimer and Johnston, 1986). The second strain used in this study is the laboratory strain GC210 (Cunningham and Cooper, 1991).

2.2.2. Genetic material from donor organisms

The *mleA* gene

The *mleA* gene is derived from the *Oenococcus œni* wine strain Lo 8413. The entire sequence of the *mleA* gene from the same strain has previously been described (Labarre *et al.*, 1996a). The GenBank accession number of this sequence is X82326. In *Oenococcus œni*, the *mleA* gene is comprised in an operon-like structure which includes the *mleA* gene and the *mleP* gene (encoding the *Oenococcus œni* putative malate permease) (Labarre *et al.*, 1996b). The *mleA* fragment used in order to construct the ML01 strain is devoid of regulatory sequences and only consists of the *mleA* open reading frame. The *mleA* gene was cloned after amplification by polymerase chain reaction of *Oenococcus œni* Lo 8413 genomic DNA.

The *mleA* gene encodes the malolactic enzyme which converts L-malate to L-lactate. This enzyme has previously been purified, but from different strains than the one used in this study (Spettoli *et al.*, 1984 ; Naouri *et al.*, 1990). Studies based on non-denaturing analytical methods attribute to this protein a molecular weight of 130 or 132 kDa (Spettoli *et al.*, 1984 ; Naouri *et al.*, 1990). Expression of the *mleA* gene in *Escherichia coli* revealed that the gene encoded a protein with an apparent molar mass close to 60 kDa, indicating that the malolactic enzyme consists of a dimer comprising two identical subunits (Labarre *et al.*, 1996a). Activity of the malolactic enzyme requires the presence of NAD^+ and Mn^{2+} (Naouri *et al.*, 1990 ; Ribéreau-Gayon *et al.*, 1998a). Several highly conserved regions are present in this protein as well as in the malolactic enzyme sequence of other lactic acid bacteria, the malic enzyme sequence of *Escherichia coli* and of eukaryotic systems. These conserved regions include the malate and NAD^+ binding sites as well as other non-defined regions (Labarre *et al.*, 1996a, Ribéreau-Gayon *et al.*, 1998a).

The *mae1* gene

The GenBank accession number for this gene is U21002. The *mae1* gene was recovered from a *Schizosaccharomyces pombe* (972h⁻ leu1-32 strain) genomic library as a *Hind*III-*Hind*III fragment (Grobler *et al.*, 1995) and was then subcloned as a *Bal*I-*Nde*I fragment (Volschenk *et al.*, 1997a). As a consequence of this cloning strategy, the isolated *mae1* gene used for the construction of the ML01 strain contains the entire open reading frame as well as 232 bp from the 3' non-coding region and 49 bp from the 5' non-coding region.

The *mae1* gene encodes the malate permease which enables the entry of L-malate into the cell. The malate permease is a 49 kDa protein, which hydrophobicity profile shows 10 putative membrane-spanning helices, typical of membrane-transport proteins (Grobler *et al.*, 1995). In addition to L-malate, it has been shown that the malate permease allows transport of other dicarboxylic acids such as malonate and succinate (Grobler *et al.*, 1995). Previous investigations involving *Schizosaccharomyces pombe* have shown this transport to be energy-dependent and associated with a proton symport (Osothsilp and Subden, 1986 ; Sousa *et al.*, 1992). However, when *mae1* is expressed *in Saccharomyces cerevisiae*, the net proton flux accompanying L-malate entry into the cell can no longer be visualized, leading Camarasa and collaborators (2001) to suggest the presence of an active proton counterflow in *Saccharomyces cerevisiae*. In *Schizosaccharomyces pombe*, the L-malate monoanionic form is the preferred substrate for this permease. Camarasa and collaborators (2001) have shown that it is this same

form which is taken up by *Saccharomyces cerevisiae* cells expressing the *Schizosaccharomyces pombe mae1* gene. Whether in *Schizosaccharomyces pombe* or *Saccharomyces cerevisiae*, the undissociated form enters by simple diffusion (Salmon, 1987 ; Sousa *et al.*, 1992). Given that the pKa of the two acid functions of malate are respectively 3.4 and 5.1 and that the pH of must ranges from 3 to 3.5, it is the monoanionic form and the undissociated form which predominate in enological conditions, allowing *Schizosaccharomyces pombe* or the genetically enhanced *Saccharomyces cerevisiae* cells to efficiently transport L-malate present in the media into the cells (Osothsilp and Subden, 1986 ; Camarasa *et al.*, 2001).

The *PGK1* promoter and terminator sequences

The *PGK1* promoter and terminator sequences were isolated from a genomic library of the *Saccharomyces cerevisiae* strain AB972 (chromosome III) (Olson *et al.*, 1986). The *PGK1* promoter has been chosen for the strong expression it drives. It contains a complex UAS at -473 to -422 and the TATA box is found at position -145. Between the TATA element and the start site, is a pyrimidine-rich cluster that serves as a signal for high-level transcription. Although designated as a "constitutive promoter", it is regulated by the glucose concentration of the media and drives gene expression best at high glucose concentrations. Puig and Perez-Ortin (2000a and 2000b) have investigated the expression of the *PGK1* gene during the alcoholic fermentation of grape juice, and found the highest level of expression to be during the exponential growth phase of yeast. 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 described above. The *PGK1* terminator allows for correct termination of the mRNA elongation during gene expression. This 264 bp fragment, ranging from position +2756 to +3019 as well as the previously described *PGK1* promoter have both been used in previous studies involving the heterologous expression of genes of œnological interest (Volschenk *et al.*, 1997a ; Volschenk *et al.*, 1997b ; Lilly *et al.*, 2000).

The *URA3* sequence

The *URA3* region (chromosome V) was cloned after amplification by the polymerase chain reaction of genomic DNA of the *Saccharomyces cerevisiae* strain GC210. This DNA sequence was used to target the integration of the transforming material by homologous recombination. A fragment corresponding to the *URA3* locus was hence used to flank both ends of the DNA to be integrated. Although previous studies had demonstrated that regions of homology as short as 30 to 50 bp could suffice to mediate homologous recombination in yeast, in order to obtain the wished integration event with certainty, we used two larger regions. The first one is a 925 bp region spanning form 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.4.1.1.23). However, because the gene sequence was split in half and used in two different fragments, it should not be 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 **pJH2**. The construction of plasmid **pJH2** is the result of successive cloning events which will be detailed in the following section, and each cloning step will be illustrated. However, in order to have a more global view of the

After digestion with restriction enzyme *Kpn*I, the amplification product of the *mae1* expression cassette was inserted into the *Kpn*I site of plasmid **pHVJH1**, in between the *URA3* sequences, yielding plasmid **pJH1** (Figure 7). The PCR product of the *mleA* expression cassette was first cloned into plasmid **pBluescript KS$^+$**, the *mleA* cassette was then isolated as a *Not*I fragment and subcloned into the *Not*I site of plasmid **pJH1**, yielding plasmid **pJH2** (Figure 8).

2.3.1.2. Detailed description of the final construct

The final construct, designated as the "integration cassette", was isolated from plasmid **pJH2** by digestion using the *Srf*I restriction enzyme. The integration cassette consists of the "malolactic cassette", which is the fragment to be integrated, flanked by homologous sequences of the *Saccharomyces cerevisiae URA3* locus. The malolactic cassette contains the *Oenococcus œni mleA* gene cloned between the promoter and terminator sequences of the *Saccharomyces cerevisiae PGK1* gene ("*mleA* expression cassette"), and the *Schizosaccharomyces pombe mae1* gene cloned between the promoter and terminator sequences of the *Saccharomyces cerevisiae PGK1* gene ("*mae1* expression cassette"). Small synthetic linkers used in the cloning strategy are also present in this cassette. A schematic representation of the integration cassette is given Figure 9.The exact nucleic composition of the integration cassette is detailed Table 3 and the sequence of the cassette is given in Fasta format in the 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 **pJH2** by digestion with the restriction enzyme *Srf*I. This resulted in the purification of a 8683 bp fragment containing the *mae1* and *mleA* expression cassettes flanked by the *URA3* sequences (the integration cassette) (Figure 9). A screening method allowing to detect cells carrying this integration cassette was developed in order to avoid the integration of a positive selection marker. In order 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 **pUT332** plasmid (Gatignol *et al.*, 1990). This plasmid carries a selectable transformation marker, the Tn5*Ble* gene, conferring resistance to phleomycin to yeast cells. The use of plasmid **pUT332** in combination with the integration cassette relies on the hypothesis that cells transformed with plasmid **pUT332** are more susceptible of also having integrated the cassette than cells not carrying plasmid **pUT332**. 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, two thousand phleomycin resistant clones were further analyzed for the presence of the integration cassette. This second screening procedure relied on the detection of L-lactate producing clones. A small amount of cells from each colony was used to inoculate sterile microtiter plates containing sterile synthetic juice. After 5 days, the culture media of each microtiter plate well was analyzed for L-lactate formation using a colorimetric reaction method. Out of the clones tested, four were isolated by the second screening method and one was validated for the integration of the malolactic cassette into the *URA3* locus and named ML01.

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

In order to get rid of plasmid **pUT332** which only purpose was to serve in the early steps of the screening method, strain ML01 was cultivated in non-selective media (i.e. media

lacking phleomycin). After successive culturing for 7 days, replica plates on non-selective media and on phleomycin containing media were done in order to select a phleomycin sensitive clone. The selected clone was then further verified for phleomycin resistance to confirm the loss plasmid of **pUT332** (Figure 10).

To further verify that the phleomycin resistance gene carried by plasmid **pUT332** had not been integrated into the yeast ML01 genome, Southern blots were performed using a probe representing the Tn5*Ble* gene of plasmid **pUT332** (Figure 11). Results clearly show the absence of Tn5*Ble* sequences in the ML01 strain (Figure 12).

Plasmid **pUT332** also contains the ampicillin resistance gene *Amp*R, used as an *Escherichia coli* transformation marker. We also verified that this gene had not been integrated in the ML01 yeast genome by Southern blot. A probe was constructed, representing the sequence of the **pUT332** plasmid lacking all *Saccharomyces cerevisiae* borne sequences. Hence, the probe represents the *AmpR* gene as well as bacterial derived **pUT332** sequences, which trace directly to plasmid pUC19, a well known and well characterized plasmid (ATCC 37254) (Figure 11). Results obtained from low stringency Southern blots (Figure 13) indicate that the ML01 strain contains neither the *Amp*R gene, nor other bacterial derived plasmid sequence.

2.3.3.2. Genetic analysis by Southern blot

Genomic DNA from ML01 and the host strain was isolated, digested with either *Nsi*I, *Pvu*II or *Eco*RI, and separated by electrophoresis in a 1 % agarose gel. Following electrophoresis, the DNA was blotted onto positively charged Nylon membranes and fixed. All probe templates were excised from the **pJH2** plasmid with appropriate restriction enzymes as shown in Figure 14.

First, a Southern blot was performed using *Nsi*I-digested genomic DNA and the ura3 probe. One signal was detected in the lane corresponding to the host strain, representing a 1.7 Kbp DNA fragment from a non-integrated copy of the *URA3* locus (Figure 15). In the lane representing strain ML01, 2 signals can be visualized (Figure 15). The lower band indicates that strain ML01 still contains at least one non-disrupted *URA3* locus, while the higher signal corresponds to a 2.8 Kbp DNA fragment resulting from the integration event in at least one *URA3* locus.

In order to clearly identify the integration event in the ML01 strain, Southern blots were performed using the *mae1* and *mleA* probes. The *mae1* probe was hybridized to *Nsi*I-digested genomic DNA, revealing a single signal representing a 2.8 Kpb DNA fragment (Figure 16). No signal was detected in the lane with DNA derived from the host strain (Figure 16). The Southern blot using the *mleA* probe indicates the presence of a 4.9 Kbp fragment containing part of the *mleA* gene in *Pvu*II-digested DNA of the ML01 strain (Figure 17). In contrast, no signal was detected with the host strain. Hence, these findings confirm the presence of the *mae1* and the *mleA* genes in the ML01 strain. The size of the hybridized fragments are in accordance with the integration of the malolactic cassette at the *URA3* locus.

Finally, a Southern blot was performed with a pgk1 probe and *Eco*RV-digested genomic DNA in order to verify that no integration had occurred in a *PGK1* genomic locus, despite the presence of the *PGK1* promoter and terminator sequences in the malolactic cassette. The probe used corresponds to the *PGK1* promoter sequence (Figure 14). In the lane representing the host strain, one signal is detected, corresponding to a 6.3 Kbp DNA fragment. As shown in Figure 18, this signal represents the hybridization of the probe with an *Eco*RV fragment containing the native *PGK1* locus. Strain ML01 shows 3 signals (Figure 18), one of them being the 6.3 Kbp fragment previously described which indicates that *PGK1* loci are intact. The other two detected signals are well in accordance with the presence of two fragments from the malolactic cassette, each one of them containing the *PGK1* promoter sequence. *Eco*RV cuts in the malolactic cassette, which yields, if the malolactic cassette is indeed integrated in a *URA3*

locus, one 6.0 Kbp and one 5.1 Kbp fragment, each containing one of the *PGK1* promoters. The 6.0 Kbp fragment is hardly distinguishable from the 6.3 Kbp one, due to their close size.

Results gathered from the Southern blots are all in accordance with the presence of the malolactic cassette integrated in at least one *URA3* locus.

2.3.3.3. Genetic analysis by pulsed field gel electrophoresis (PFGE) followed by hybridization

Genomic DNA from strain ML01 and the host strain was isolated and chromosomes separated by PFGE. Results (Figure 19), show that both patterns are identical, proving that no major DNA reorganization has occurred upon integration of the malolactic cassette.

The pulsed fields obtained were then used for transfer onto a positively charged Nylon membrane and hybridized with different probes. The *mae1* probe hybridizes to at least one chromosome, the same as the one hybridizing the *ura3* probe (Figure 19). Hybridization of the PFGE with the *pgk1* probe shows that while at least one chromosome hybridizes with the pgk1 probe in the host strain, one other chromosome is highlighted in strain ML01 (Figure 19). This chromosome also correspond to the one hybridizing with the *ura3* and *mae1* probes, indicating the presence of the *PGK1* promoter sequence in at least one chromosome carrying the *URA3* locus. **Hence, results gathered from these hybridization experiments are also in accordance with the presence of the malolactic cassette in at least one *URA3* locus.**

2.3.3.4. Genetic stability of the ML01 yeast strain

The ML01 strain was cultivated in YPG media for more than 100 generations. Southern blots were then performed on genomic DNA using the pgk1 probe. Figure 20 shows that the hybridization profile of the strain remains the same after more than 100 generations, indicating that **the integrated locus of the ML01 strain is genetically stable.**

Nothing indicates that the ML01 yeast could be genetically instable given the results found in all assays. Moreover, despite its use for over 20 years as a wine active dry yeast, the host strain has never shown any changes in its fermentation performances. **Therefore, we can affirm that, at the best of our knowledge, the ML01 strain is stable.**

Genetic analysis of the ML01 strain shows that the integration event has occurred. Southern blots on digested genomic DNA provided us with the proof that the integrated material contains at least the *URA3* flanking sequences, the *mae1* and *mleA* genes as well as the *PGK1* promoter. The sizes of the fragments recovered from the different hybridization experiments correspond well to the integration of the entire malolactic cassette. Moreover, Southern blots with the pgk1 probe showed that the integration had not occurred in a *PGK1* locus, despite the presence of the *PGK1* promoter and terminator sequences in the integration cassette. PFGE and subsequent hybridizations indicate that the integration event has occurred in at least one chromosome carrying the *URA3* locus. Results from the Southern blots using the *ura3* probe indicate that in strain ML01, the malolactic cassette is integrated into at least one locus of the *URA3* gene whilst at least one other *URA3* locus is left intact. In addition, all existing data show that the ML01 yeast is stable.

2.3.4. DNA sequencing of the integrated locus

Further analysis was performed in order 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. Two sets of templates spanning the entire integration cassette were obtained by PCR using a proof-reading polymerase. The first set of templates was used for the sequencing of one strand while the other set used for the sequencing of the reverse strand. Sequencing of each strand with different sets of templates minimizes sequencing errors due the amplification procedure. The two complete sequences

were aligned and analyzed for differences. If differences occurred, another round of PCR and sequencing of the region in doubt 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 Saccharomyces Genome Database for *URA3*, *PGK1*p and *PGK1*t, and the National Center for Biotechnology Information for *mae1* and *mleA*. Table 4 lists the differences in between both sequences. Out of 15 differences, 7 result from the presence of additional sequences in the genomic copy due to the cloning strategy (comments 2, 3, 4, 7, 8, 11 and 12 of Table 4).

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

Four differences were found in the *URA3* sequences (comments 1, 13, 14 and 15 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 polymorphism. 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.

Two other differences are found, involving the *PGK1*p sequence (comments 5 and 6 of Table 4). Because the *PGK1* promoter sequence is found twice in the malolactic cassette and that only the one driving the expression of the *mae1* gene shows differences with the published sequence, these differences more than likely originate from polymerase errors during the amplification steps involved in the cloning of the malate permease expression cassette into plasmid pHVJH1 (Figure 7).

Differences in the *mleA* and the *mae1* genes

Two differences were found in the coding sequence of the *mleA* gene. These disparities could originate from a mistake during the amplification steps involved in the cassette construction as both sequences come from the same strain. Alternatively, these differences can be attributed to errors in the published sequence or a mutation within the strain isolate used for the cloning of *mleA*. The first one (comment 9 of Table 4), introduces a change of amino acid from aspartic acid to glutamic acid. It is important to note that both amino acids are acidic amino acids, hence belong to the same family. The second difference (comment 10 of Table 4) results in a silent mutation involving no change in the amino acid sequence.

No difference were found in the *mae1* sequence when compared to the published sequence which comes from the same *Schizosaccharomyces pombe* strain.

In order to fully analyze the integrated locus, we proceeded to the sequencing of the integrated genomic region. The integration cassette was shown to contain only synthetic polylinkers and DNA from donor microorganisms. When comparing the integrated cassette sequences to previously reported sequences, six differences were found in either non-coding regions (promoters, terminators), or in the non-functional (because disrupted) *URA3* open reading frame. These differences may be due to strain genetic polymorphism or polymerase errors during the amplification steps. Two changes were found in the *mleA* sequence, one corresponding to a silent mutation, the other one involving the exchange of one amino acid for another of the same family. The *mae1* gene has the exact same sequence than the original one. These results indicate that the sequences composing the integration cassette of the ML01 strain are not significantly different than the original sequences isolated from the donor strains.

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

The genetic material used to modify the S92 host strain is of restricted size and should not allow to differentiate the recombinant strain from the host strain. We however verified the relationship between both strains by characterizing them genetically using a molecular method. The method used relies on the amplification of Ty1 elements sequences and is commonly used to characterize strains of the *Saccharomyces* genus (Ness *et al.*, 1993). Figure 21 shows the genetic pattern of strain S92 and strain ML01. Both patterns are identical except for an additional small faint band in the S92 pattern. Such band is qualified as a "ghost band" by Ness and collaborators (1993) which presence can vary from one pattern to another of the same strain.

Therefore, it can be concluded that the recombinant ML01 strain is not different from the S92 host strain (except for the integration event) by the polymerase chain reaction targeted to Ty1 sequence elements technique which permits to differentiate between industrial *Saccharomyces cerevisiae* strains.

Furthermore, chromosomal patterns obtained by PFGE (Figure 19) confirmed that both patterns are identical. These results show that no major DNA reorganization event had occurred upon integration of the malolactic cassette.

2.3.6. Method of manufacture of the modified microorganism

The modified ML01 yeast was modified by integration of foreign 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 ML01 yeast strain. Furthermore, the modification concerns the catalytic conversion of L-malate into L-lactate, while malate and equivalent dicarboxylic acids are never present in the medium used to produce industrial yeast such as bakers yeast, wine yeast... Consequently, the ML01 wine active dry yeast can be manufactured in the exact same way as all active dry yeast used in bakery, brewing or winemaking, as it is described in reference literature (Reed and Nagodawithana, 1991). A brief description of the manufacturing of such yeast is given Figure 22.

3-INFORMATION ON ANY SELF-LIMITING LEVELS OF USE

Lesaffre sees no reason why the inoculation rate should be different than the one recommended as indicated in § 1.3.3.

Anyway, if in the grape must the inoculation rate is superior to the recommended level, it is not certain that the quantity of yeast cells present in the wine at the end of the fermentation will be significantly different. This should lead to no change at the wine level.

4- DETERMINATION THAT THE USE OF THE WINE ACTIVE DRY YEAST OBTAINED FROM THE ML01 YEAST STRAIN IN THE FERMENTATION OF GRAPE MUST IS EXEMPT FROM THE PRE-MARKET APPROVAL REQUIREMENTS OF THE ACT BECAUSE SUCH USE IS GRAS.

4.1. Safety assessment of the host strain

Scientific data available through peer-reviewed publications indicate that the host strain S92 belongs to a well known family of wine yeast strains corresponding to the *Saccharomyces cerevisiae* species although it is often referred to by oenologists as *Saccharomyces bayanus*. In the frame of this study, we have only verified that this strain belongs to the species *Saccharomyces cerevisiae* variety *bayanus* as defined by Kreger-van-Rij (1984).

The S92 strain belongs to the family of strains usually called "Prise de Mousse" but which also have other numerous commercial names. This family of strains has been isolated from the Champagne region in France. This wine region is known for having in its cellars a family of very close or very similar strains which predominate in the micro-flora and which perform very well in the fermentation of sparkling wines of the Champagne type. This family of yeast strains is probably one of the most utilized, even the most utilized, family of yeast strains in the world for the production of wine active dry yeast. This family of yeast has been shown to produce the K2 killer factor and to sporulate (Johnston *et al.*, 2000 ; our study). These characteristics are commonly found in wild type strains of *Saccharomyces cerevisiae*.

Saccharomyces 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 regarded as a yeast 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 a permitted material added in the production of natural wines (27 CRF 24.176).**

Saccharomyces cerevisiae is considered as GRAS through its use in the brewing, baking and winemaking industry. Its genome has entirely been sequenced which confirmed that this yeast is free of known pathogenicity traits. Some organizations have evaluated the safety of *Saccharomyces cerevisiae* :

- The Scientific Committee for Human Food of the European Community in its 27th report indicates that this yeast has a safe history of use in food and belongs to a species which is known not to produce toxins.

- The Environmental Protection Agency (EPA) has included *Saccharomyces cerevisiae* as a recipient microorganism for exemptions from EPA review and expedited EPA review (**40 CFR 725.420**). This decision was based on the fact that *Saccharomyces cerevisiae* is found to have little adverse effects and 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 S92 yeast strain belongs to the *Saccharomyces cerevisiae* species which has been used for more than 7000 years by humans in fermented beverages. Given its history of use, it can be affirmed that this species, whatever transformation with no harmful DNA material or mutation it may be subjected to, will stay GRAS and by no means will produce dangerous substances for humans.

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

4.2.1. Foreign genetic material source and product

4.2.1.1. The *mae1* gene

Source* : *Schizosaccharomyces pombe

The *mae1* gene was isolated from a strain of *Schizosaccharomyces pombe*. *Schizosaccharomyces pombe* is a yeast species which has been originally isolated from millet beer from East Africa. An early report by Lindner (1893) emphasizes "that millet beer was drunk with pleasure by the Germans who considered it as a healthy and refreshing beverage".

A bibliographic research carried out by Lesaffre's library department did not find any paper relating the implication of *Schizosaccharomyces pombe* in allergy reactions. The

data bases used were Medline (10.7 billion references since 1966) and Biosis (12.4 billion references since 1969).

Schizosaccharomyces pombe is commonly found on grape berries and in wine and may occasionally participate to spontaneous grape must fermentation.

Schizosaccharomyces pombe has been proposed to conduct alcoholic fermentation (Dharmadhikari and Wilker, 1998 ; Pretorius, 2000) :

- This species is being used for the production of potable alcohol from beet juice and molasse by a high cell density fermentation process called the Alpha Laval process.
- Due to the production of unsavory flavor compounds, this yeast has never been commercially utilized as a yeast starter for grape must fermentation (Delfini and Formica, 2001).

As specified in § 2.2, the strain used as source of the *mae1* gene (972h$^-$, leu1-32) originates from a strain isolated from sulphurized grape juice by Osterwalter in 1924 (Leupold, 1950).

Given its origin, the safe history of the presence of *Schizosaccharomyces pombe* strains in fermented beverages as well as the absence of data concerning the allergenicity of this yeast, it can reasonably be submitted that the *Schizosaccharomyces pombe* strain 972h$^-$ leu1-32 is GRAS.

Product **: malate permease**

The two foreign gene products assessment are performed using data provided by this study and results reported in the literature.

Sequencing of the inserted *mae1* gene of the ML01 strain indicates that no sequence discrepancy could be found in between the recombinant *mae1* gene in ML01 and the native *mae1* gene from *Schizosaccharomyces pombe* (see Appendix 2). Hence, both malate permeases, the one from the recombinant gene and the one from the native gene have the same primary structure.

Volschenk and collaborators (1997b) have analyzed the native and the recombinant malate permease by electrophoresis on a denaturing gel, and showed that both proteins have close molecular weights. Moreover, they are both recognized by rabbit anti-mae1p antibodies, indicating that no major change has occurred that could significantly affect the antigen properties of the mae1 protein when expressed in *Saccharomyces cerevisiae*.

In both cases, the monoanionic form of malate has been shown to be the substrate of the malate permease. Camarasa and collaborators (2001) found the Km of the malate permease to be of 1.6 mM when present in a recombinant strain of *Saccharomyces cerevisiae*. Osothsilp and Subden (1986) found a similar, although not identical value for the native *Schizosaccharomyces pombe* malate permease (3.7 mM). The model used for the Km calculation by Camarasa and collaborators, which took into account the background resulting from malate passive diffusion, could explain the difference found in between both values.

The fact that the ML01 yeast efficiently degrades L-malate shows that the malate permease is functional. This is in itself a proof that the secondary and tertiary structures of the recombinant protein have not suffered major changes compared to the native malate permease.

Hence, expressing the *mae1* gene in *Saccharomyces cerevisiae* has no major consequence on the properties of the malate permease, which can therefore be assessed as safe as the native malate transporter in *Schizosaccharomyces pombe*.

4.2.1.2. The *mleA* gene

Source **: *Oenococcus œni***

The *mleA* gene was isolated from *Oenococcus œni*. This species, formerly called *Leuconostoc œnos*, is widely found in wine and is usually the major lactic acid bacterium responsible for malolactic fermentation (Henick-Kling, 1993, Ribéreau-Gayon *et al.*, 1998a). **The Bureau of Alcohol Tobacco and Firearms rates *Leuconostoc œnos* as an authorized material for treatment of wine and juice (27 CFR 24.246)** and bases this authorization on an FDA advisory opinion dated of 02/25/85. The strain used in this study has specifically been isolated from wine at the Œnology Faculty of Bordeaux from the University Bordeaux2-Victor Segalen, France, and therefore participated in the spontaneous malolactic fermentation of wines in the Bordeaux area (Labarre *et al.*, 1996a). **This strain raises no consumer safety concerns.**

Product **: malolactic enzyme**

When comparing the sequence of the recombinant *mleA* gene in the ML01 strain and the native gene sequence in *Oenococcus œni*, two changes were found (Table 4). However, one difference corresponds to a silent mutation while the other involves the change of an acidic amino acid for another. Hence, despite these two differences in the gene sequence, no significant change has occurred in the primary sequence of the malolactic enzyme.

Working with a *Saccharomyces cerevisiae* strain expressing the *mleS* gene encoding the *Lactococcus lactis* malolactic enzyme, Ansanay and collaborators (1996) showed that the malolactic enzyme had a similar molecular weight than the native *Lactococcus lactis* one. Moreover, both proteins, the native and the recombinant one, were detected after SDS-PAGE by Western blot using anti-mleSp antibodies. Therefore, no major change has occurred that could significantly affect the size and the antigen properties of the mleSp protein when expressed in *Saccharomyces cerevisiae*.

Groisillier and Lonvaud-Funel (1999) have shown by phylogenetic analysis that the *Lactococcus lactis mleS* gene and the *Oenococcus œni mleA* gene are close. If as just previously described, the *mleS* heterologous expression doesn't alter the protein properties, then the expression of the *mleA* gene in the ML01 strain should not change the malolactic enzyme characteristics either.

The ML01 strain shows an efficient L-malate degrading activity indicating the functionality of the malolactic enzyme, hence the conservation of its secondary and tertiary structure.

Therefore, to the best of our knowledge, the *mleA* gene product of the ML01 strain is similar to that of *Oenococcus œni*. Since this species is authorized by the Bureau of Alcohol, Tobacco and Firearms to conduct malolactic fermentation in wine, the presence of the *mleA* gene product in ML01 should raise no safety concern.

4.2.2. DNA yeast material : the *PGK1* promoter, terminator and *URA3* sequences

The *PGK1* promoter and terminator as well as the *URA3* sequences have been isolated from laboratory strains of *Saccharomyces cerevisiae* as described in § 2.2. The safety assessment of *Saccharomyces cerevisiae* has already been established in § 4.1 and since the *PGK1* regulatory sequences do not code for any proteins and that the *URA3* locus is disrupted, no further safety assessment needs to be performed.

The source of microorganisms and the foreign gene products employed (i.e. the malate permease and the malolactic enzyme) do not raise any safety concern.

4.2.3. Construction of the modified organism

4.2.3.1. Vector used as source of the integration cassette

Plasmid **pJH2** was used in the construction of the modified ML01 yeast strain as a source of the integration cassette. Plasmid **pJH2** is the result of the insertion of the integration cassette in plasmid **YEp352** whose unique *Kpn*I site had been destroyed (Figures 1 to 8). 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 **pJH2** by digestion with restriction enzyme *Srf*I and subsequent extraction from an agarose gel. Hence, all **pJH2** 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 *Saccharomyces cerevisiae PGK1* promoter and terminator, the *Saccharomyces cerevisiae URA3* sequences, the *Schizosaccharomyces pombe mae1* gene, the *Oenococcus œni mleA* 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 S92. As described in § 2.3.2., host yeast cells were transformed with a mixture of the integration cassette and plasmid **pUT332**, which was utilized in the first screening round. Once a clone with the integrated malolactic cassette was isolated, it was cultured in a non-selective medium in order to allow for loss of plasmid **pUT332**. Different methods were employed in order to verify that sequences from this plasmid had been eliminated from the ML01 strain. These methods relied on :

- Verification that the ML01 strain has lost its ability to grow on a phleomycin containing media, indicating the absence of a functional phleomycin resistance gene (Figure 10).
- Verification by Southern blot of the absence of the phleomycin resistance gene using a probe corresponding to the **pUT332** Tn5*Ble* gene (Figure 12).
- Verification by Southern blot of the absence of the **pUT332** plasmid sequence containing the ampicillin resistance gene and all other bacterial derived plasmid sequences (Figure 13).

The results conclusively show that the pUT332 sequences are not present in the ML01 strain. Furthermore, emphasis is made on the absence of antibiotic resistance genes in the ML01 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 § 2.3.3.

Results obtained from hybridization of genomic DNA with probes corresponding to *URA3*, *mae1*, *PGK1p*, or *mleA*, have shown that the malolactic cassette is correctly integrated into a least one *URA3* locus (Figures 15 to 18)

The chromosomal patterns of the S92 and the ML01 strain are identical as well as amplification patterns based on sequence elements of the Ty1 transposon (Figures 19 and 21).

DNA sequencing has confirmed that only sequences belonging to the *mae1* and *mleA* genes, 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 ML01 strain has been genetically characterized. The integration site contains no DNA sequences other than those present in the isolated integration cassette. No plasmid sequence, especially antibiotic resistance sequences, are present in the genetic material of strain ML01. Moreover the S92 and ML01 strain have identical chromosomal patterns. It can thus be concluded that at the best of our knowledge, the ML01 strain is genetically substantially equivalent to the S92 host strain, and most probably identical to the S92 host strain, except for the malolactic cassette present in at least one *URA3* locus.

4.3.2. Consequence of the genetic modification on gene expression

4.3.2.1. *mae1* and *mleA* gene expression

Studies on the expression of the *mleA* and *mae1* genes during the fermentation of synthetic must by the ML01 strain were carried out.

One experiment analyzed samples at 48 hours (2 days) and 144 hours (6 days) after inoculation. Results indicated that both transcripts are present at 48 hours and 144 hours, although *mae1* seems to be less expressed at 144 hours than at 48 hours (Figure 23). This later observation can not be taken as a definitive conclusion as RT-PCR constitutes only a semi-quantitative method, but is consistent with our knowledge of the expression of *PGK1p* driven genes (Puig and Perez-Ortin, 2000a and 2000b, § 2.2.2.).

Another experiment further analyzed gene expression during fermentation. Results, only given for *mleA*, indicate that messenger RNAs are still present 12 days after the inoculation (Figure 23).

In both experiments, malate was entirely degraded within 6 days, when cells entered the stationary phase. However, *mleA* continued to be expressed during at least 6 more days. This is in accordance with the literature which shows that *PGK1* promoter mediated expression still occurs during the stationary phase of grape must fermenting yeast cells, although to a lesser extend than during the exponential growth phase (Puig and Perez-Ortin, 2000a and 2000b ; § 2.2.2).

Despite the *mleA* expression throughout the fermentation, we have observed that malolactic fermentation was either always completed before the cells reached the stationary phase or that malolactic fermentation will stop before the cells enter the stationary phase, even if malate has not entirely been degraded. The same conclusion has been drawn by Bony and collaborators (1997). The limiting step in malate degradation, leading to the absence of

malolactic fermentation at the stationary phase, could be, as suggested by the literature, malate permease, through the lack of presence, or of functionality of this protein (Bony *et al.*, 1997). This is reinforced by the presence of a PEST region in the protein sequence, indicating that this protein is rapidly degraded within the yeast cell. This latter hypothesis is the one currently predominating as it has been suggested by Bony and collaborators (1997).

4.3.2.2. Global gene expression analysis

Global gene expression of the ML01 yeast was investigated using the Affymetrix GeneChip® Yeast Genome S98 Array. These gene chips contain probes from all known *Saccharomyces 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 ML01 strain during the fermentation of synthetic must containing 4.5 grams/liter of malate was compared to that of the host strain in identical conditions at two time points (48 hours and 144 hours) in duplicate. This study was carried out at the Wine Research Center of the University of British Columbia. Results were analyzed by experienced scientists and are presented here in the form of an affidavit signed by the head of the research center, Prof. van Vuuren (see Appendix 4). At 144 hours, 15 genes were up or down-regulated more than 2-fold. One can conclude from these results that no metabolic pathway is affected by the presence of the expression cassette in strain ML01.

Hence, from global gene expression analysis, the trained scientists of the Wine Research Centre of the University of British Columbia have concluded that strain ML01 can be regarded as substantially equivalent to the wild type strain S92.

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

Saccharomyces cerevisiae is a yeast that has been used for more than 7000 years by humans. Throughout this period, many genetic changes have occurred but this yeast has never been described as having adverse effects as a result of unintended genetic changes. Moreover, independently from one another, the sequences composing the integration cassette do not raise any safety concern. However, in the malolactic cassette of the ML01 strain the *URA3*, *PGK1p*, *PGK1t*, *mleA* and *mae1* sequences are present in a different genetic environment than their homologous native sequence. Hence, an investigation in order to establish the occurrence of new open reading frames (ORFs) putatively expressed as a consequence of the rearrangement of sequences within the integration cassette was conducted. The sequence of the integrated cassette has been analyzed with the Omiga software (version 5.0) for search of ORFs of more than 90 bp with a TATA box in the 5'-flanking region. Seven new open reading frames were found, four having two possible start codons (Figure 24). The sequence of the ORFs and of the deduced amino acid sequences are given in a Fasta format in Appendix 5.

In order for these open reading frames to correspond to putative proteins, they first need to be transcribed in messenger RNA. We investigated whether or not these open reading frames were transcribed at two time points (48 hours and 144 hours after the beginning of fermentation) during the fermentation of synthetic grape must. Figure 24 shows that all open reading frames are indeed transcribed in these conditions. However, the RT-PCR technique doesn't allow us to quantify the expression levels, neither gives us information on RNA stability.

RNA stability is a limiting factor in the 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 or poorly adapted to enhance mRNA stability, then chances are that the RNA corresponding to these ORFs will have a very short half life. Moreover, once transcribed, the corresponding RNAs need to be translated.

Efficient translation will only occur if the ORF possesses untranslated 5' sequences favoring translation initiation. In *Saccharomyces cerevisiae*, consensus sequences which participate in translation efficiency exist, but it stays difficult to predict whether an open reading frame will be translated or not just by analyzing its 5' sequence. Nevertheless, ORF1, because it shares with the *URA3* gene the same 5' untranslated sequence, could be translated, like the native *URA3* sequence. For the other putative ORFs, it is difficult to precisely assess their translation efficiency but it is unlikely. Sequence size may be an indicator of translation probability as it has been conclusively shown that ORFs of less than 100 codons (300 bp) have a chance below 10 % to actually code for a protein (Mackiewicz *et al.*, 2002). Furthermore, if these proteins exist, they should be very unstable due to their small size. Hence, smaller ORFs like ORF2, ORF3, ORF5 and ORF7 will more than likely not encode for a protein, while larger ones such as ORF4 end ORF6 might have more chances of being translated, albeit much less than ORF1, because their sequence from the 5' region corresponds mainly to the PGK1 promoter which is not a coding sequence in the yeast genome, whether on the Watson or Crick DNA strand.

The fact that ORF1, ORF4 and ORF6 may be translated into a protein is not a problem *per se*. Each of these sequences mainly corresponds to *Saccharomyces cerevisiae* sequences (ORF1 : mainly *URA3* + 6 nucleotides from synthetic sequence + 6 nucleotides from *PGK1t* ; ORF4 : mainly *PGK1p* + 14 nucleotides from synthetic sequences + 32 nucleotides from *PGK1t* ; ORF6 : mainly *PGK1p* + 7 nucleotides from synthetic sequence + 12 nucleotides from *URA3*). Hence, the large majority of the nucleotide sequence of these ORFs can also be found in the parental strain. Moreover, because these ORFs share the same untranslated region than their parental *Saccharomyces cerevisiae* counterpart, the transcription and translation pattern of parental sequences and of the putative ORFs should be similar. Thus, if these ORFs actually encode for proteins, then very similar native proteins should also be found in the parental strain.

In conclusion, given :

- that the S92 strain was never shown to release allergenic or toxic substances
- the very weak chances of smaller putative ORFs like ORF2, ORF3, ORF5 and ORF7 being translated
- the close sequence similarity in between the bigger putative ORFs (ORF1, ORF4 and ORF6) and *Saccharomyces cerevisiae* parental sequences, **there is not reason to think that the ML01 strain will release toxic or allergenic substances as a result of the presence of unintended gene products.**

4.3.4. Study of yeast proteins by two-dimension protein gels

We tried to compare the global protein content of the ML01 strain to that of the S92 host strain by analysis by two-dimension protein gels during synthetic grape must fermentation. The exact same conditions than the ones used for global gene expression were applied. Two dimension gels of the ML01 and S92 soluble proteins were performed at 48 hours and 144 hours after fermentation.

Analysis of the little and poor data obtained showed the presence of the malolactic enzyme at 48 hours in the ML01 yeast. This protein is much less present at 144 hours. This technique because it cannot detect insoluble proteins, did not allow us to visualize malate permease.

Concerning other yeast proteins, no significant difference could be established when comparing the ML01 strain with the parental strain. However, it was impossible to draw any sure conclusion from this study, because the method used did not allow us to visualize many yeast proteins and was not reproducible. Hence, the study of yeast proteins by two-dimension protein gels was abandoned.

4.3.5. Consequence of the genetic modification on the physiology of the ML01 strain

4.3.5.1. Overview of the function of the malolactic gene in lactic acid bacteria

The *mleA* gene encodes for the malolactic enzyme which catalyzes the conversion of L-malate to L-lactate. In order to function, this enzyme requires two co-factors, Mn^{2+} and NAD^+. These co-factors, although they are absolutely necessary for the catalytic activity of the enzyme, are not transformed during the bio-conversion reaction, and no net reduction of NAD^+ to NADH can be seen.

4.3.5.2. Manufacture of active dry yeast during fed-batch aerobic cultures

Section 2.3.6. gave an overview of the method of manufacture of the ML01 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 S92. Both of these strains, during the pilot scale or industrial manufacture of active dry yeast, showed similar multiplication rates and alcohol production kinetics. No growth advantage was found for the ML01 strain. **No difference was noted during the preparation of both strains as active dry yeast.**

4.3.5.3. Growth of ML01 and S92 strains during laboratory-scale vinifications

Growth rates of the ML01 and S92 host strain fermenting synthetic or Chardonnay must were monitored using an automated incubating turbidometric reader. This allowed us to test 18 replicates per strain per test media. Results presented in Figure 25 indicate that the growth curves are similar for both strains, whether fermenting synthetic must containing 4.5 grams/liter L-malate or Chardonnay must containing 4.5 grams/liter L-malate.

Vinification trials performed in the laboratory have also shown that the S92 and the ML01 strain have similar fermentation rates and ethanol yields. These results have also been confirmed by large scale vinification trials.

Hence, the presence of L-malate doesn't give the ML01 strain any growth advantage, even though malate is being degraded. This indicates that the degradation of L-malate by the yeast cell is not of any metabolic use, whether as a carbon or an energy source.

Vinification trials summarized in §1.3.5 and results from global gene expression showed that no other difference than the stochiometric conversion of L-malate to L-lactate is found in between the ML01 strain and the S92 host strain.

Hence, in these winemaking conditions, the ML01 strain has the same growth kinetic, fermentation rate and ethanol yield as the parental strain. All data indicate that the only difference in between the S92 and the ML01 strain is the L-malate degradation capacity of the ML01 strain, without this property conferring any growth advantage to the cell.

4.3.5.4. Growth and fermentation rates during fermentation of high acid -high malate containing must

The growth rates of the ML01 and S92 strains were monitored in a high-acid must containing 9 grams/liter L-malate with an incubating turbidometric reader (18 replicates per strain). These characteristics are found in cool climate wine regions.

Results show that both strains have similar exponential growth curves but at and after the inflection point of the curve, the ML01 grows at a slightly higher rate (Figure 26). This growth difference should be linked to an increase of the pH as a result of malate degradation which occurs during the exponential growth phase.

Similar results were obtained with larger scale assays in Canada. Here, the must also had a low pH (3.18) and a high malate content (9 grams/liter). Completion of the alcoholic fermentation was faster for the ML01 strain than for the S92 strain. Again, the faster fermentation rate of the ML01 strain is linked to the capacity of this strain to higher the must pH by L-malate degradation.

Hence, the ML01 strain has a slight growth advantage in high acid and malate must, due to its capacity to degrade L-malate and higher the must pH. However, this growth advantage will benefit not only to the ML01 strain, but to all yeast cells present in the must as long as there is enough ML01 cells to significantly degrade malate and higher the pH.

4.3.5.5. Growth of ML01 and S92 strains during batch laboratory cultures in the presence of malate

We evaluated whether the genetic modification affected the yeast growth kinetics in a media containing L-malate and a limited glucose concentration sufficient to trigger *mleA* and *mae1 PGK1* promoter expression. A minimal medium containing 5 grams/liter glucose and 20 grams/liter L-malate was used. This study was performed twice :

- Once with ML01 and S92 in separate cultures using microtiter plates wells. This allowed us to evaluate growth rates with an automated incubating turbidometric reader. Eighteen replicates per strain were performed.

- Once using larger scale cultures in 150 milliliter media. These trials were inoculated with either strain (ML01 or S92) or with a mixture containing 50 % of ML01 cells and 50 % of S92 cells. The percentage of ML01 cells throughout the culture was then estimated.

Figure 27 showing growth curves in this media indicates that in all cases, the growth rates for both strains are similar, despite the presence of L-malate in the growth media. Furthermore, the proportion of ML01 cells does not increase during the fermentation compared to that of the total yeast inoculum. Hence, the presence of L-malate doesn't give the ML01 strain any growth advantage.

In this culture condition, no malate is degraded by the ML01 strain after 60 hours of culture.

The presence of L-malate in the yeast natural environment does not confer any competitive advantage to the ML01 strain.

4.3.5.6. Mixed cultures of the S92 host strain and the modified ML01 strain in synthetic grape juice

Fermentations in synthetic grape juice were performed using mixed cultures of the S92 and ML01 yeast strains. These experiments are detailed in the Appendix 6 and results indicate that :

- Malolactic fermentation only occurs during the yeast exponential growth phase. There is practically no malolactic fermentation if the ML01 strain represents about 1 % of the total yeast cell population at the beginning of the alcoholic fermentation. In trials involving mixed inoculations with the ML01 and the S92 strains, the percentage of ML01 yeast cells does not increase during alcoholic fermentation (the percentage of the ML01 yeast cells would maybe even have a slight tendency to decrease).

- The ML01 strain never takes over the fermentation, indicating that the presence of malate doesn't confer any selective advantage to this yeast strain.

4.3.5.7. Sporulation and killer character of the modified yeast

The sporulation capacity on minimum sporulation medium of the S92 and the ML01 strains was investigated. Formation of ascospores containing at least 3 spores was evaluated for both strains. Results summarized in Figure 28 indicated that the S92 and the ML01 strains have close ascospore formation rates.

Preliminary results concerning the study of these spores seem to show that the spores of ML01 which have kept the capacity to carry out the malolactic fermentation are auxotrophic for uracil.

The killer character of both yeast was evaluated. Results confirmed that the S92 and the ML01 strain are both K2 killer strains.

The genetic manipulation of the ML01 strain has not changed the ascospore formation (for ascospores of at least 3 spores) nor the K2 killer character of this yeast.

Genetic studies have shown that the ML01 strain is genetically at least similar to the S92 host strain, except for the integrated malolactic cassette. Global gene expression analysis performed by experienced scientists indicated that the introduction of the malolactic cassette does not change gene expression patterns in the ML01 yeast significantly. This leads to the conclusion that strain ML01 can be regarded as substantially equivalent to the wild type strain S92 with a supplementary property to degrade L-malate into L-lactate. Physiological studies carried out in many different culture conditions indicate that no difference between the ML01 and S92 strains can be seen, except for malate degradation which occurs in specific conditions of exponential growth, before the wine reaches 5 % alcohol. The degradation of malate doesn't confer any significant growth advantage to the ML01 strain. The ML01 strain possesses the same ascospore formation and killer characteristics than the S92 host strain. Therefore, it can be concluded that the ML01 strain is substantially equivalent to the host strain S92 except for the ability of ML01 to decarboxylate L-malate to L-lactate 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 ML01 fermentation

The ML01 strain enables the winemaker to carry out the alcoholic and malolactic fermentation in one single step. Figures 29 and 30 indicate changes occurring in winemaking procedures of white and red wines as a consequence of the use of the ML01 yeast strain. **The use of the ML01 yeast strain leads to no drastic change in the winemaking procedures.** Only at the stage of malolactic fermentation does the procedure vary. Consequently, the winemaking procedure will be shorter. This enables the winemaker to stabilize the wine earlier by addition of sulfur dioxide and filtration. However, the delays used in traditional winemaking procedures may want to be kept. This is the case when aging of the wines on lees or in barrels, or both. The use of the ML01 yeast strain is not incompatible with such procedures.

4.4.2. Changes in the physiology of the modified yeast as consequence of the genetic modification during winemaking

Vinification trials summarized in § 1.3. 5. showed that both strains, S92 and ML01, have similar ethanol yields and fermentation rates. The latter results were confirmed by Bony and collaborators (1997) who worked with a recombinant strain transformed with plasmids (multicopy or monocopy) carrying the *Lactococcus lactis* malolactic enzyme gene and the *Schizosaccharomyces pombe* malate permease gene under the *PGK1* promoter.

These authors also indicate that the growth characteristics of the recombinant strain remained unchanged from the native strain when fermenting synthetic must. We have shown similar results in laboratory scale vinifications using an automated incubating turbidometric reader (Figure 25).

As indicated by wine sensory analysis performed after wine trials, the flavor profile of the wines produced by the ML01 strain is not significantly different from that of control wines, except for the organoleptic characteristics linked to the deacidification process.

During malolactic fermentation by the ML01 yeast, L-malate is converted into L-lactate in stochiometric amounts. This indicates that the ML01 strain only degrades exogenous malate and does not metabolize the endogenous L-malate which could be formed from intermediates of the Krebs cycle, into L-lactate. Therefore, even though a new L-malate degradation pathway has been introduced into the cell, it does not interfere with the endogenous metabolism of malate. The same results were found by Bony and collaborators (1997) when analyzing a recombinant yeast strain carrying the *Lactococcus lactis* malolactic enzyme gene and the *Schizosaccharomyces pombe* malate permease gene under the *PGK1* promoter on monocopy plasmids.

Global gene analysis performed at two time points during laboratory vinification trials, have not shown any significant difference in between the modified ML01 yeast and the S92 host strain.

4.4.3. Changes in wine composition as a consequence of the ML01 fermentation

4.4.3.1. Global characteristics of wine

Vinification assays indicated that many parameters measured in wine such as the ethanol content, residual sugar concentration and volatile acidity stay similar in wines produced from the ML01 strain or a control wine yeast.

This implies that the ML01 wine yeast does not change the global constitution of wine.

4.4.3.2. Flavor modification

Laboratory, pilot scale vinification trials and large scale assays showed that the ML01 strain doesn't lead to any organoleptic deviation of wine.

Descriptive characters of the different ML01 wines did not significantly vary from control wines, except for those related to deacidification.

The use of the ML01 wine yeast strain does not lead to any wine faults. Furthermore, it does not lead to any organoleptic difference compared to wines produced with the S92 host strain, except for the deacidification of wine.

Data provided by our investigations as well as results given in the literature indicate that the growth kinetics, fermentation rates, ethanol yields and flavor profiles of wines are not affected by the use of a modified strain having a new L-malate degradation pathway, during wine fermentation. Moreover, the malolactic metabolism does not interfere with the endogenous malate metabolism of the yeast. Thus, nothing indicates that the ML01 yeast metabolism is different from that of the host strain. This is in accordance with results found during global gene expression studies. The ML01 strain should be regarded as a system substantially equivalent to a system involving independently the S92 host strain and a L-malate to L-lactate conversion pathway in presence of Mn^{2+} and NAD^{+}(which are necessary co-factors although not consumed).

4.4.3.3. Yeast cells and release of yeast products during winemaking

After alcoholic fermentation, the viable yeast population starts to decrease in wine and yeast autolysis occurs. The wine is either immediately clarified or is kept on lees.

Wine clarification

Wine clarification is usually done by letting the solid particles of the wine sediment by gravity and subsequent elimination of the sediments. During this process, a majority of the yeast cells is eliminated and the total yeast population falls to 10^3 or 10^4 cells/milliliter of wine (Ribéreau-Gayon *et al.*, 1998b).

Alternatively, wine can be clarified by centrifugation. This process allows to eliminate more than 99 % of the yeast cells (Ribéreau-Gayon *et al.*, 1998b).

This clarification procedure is an obligatory step in the wine processing. However, it can be delayed by storage of the wine on lees.

Keeping the wines on lees

Some winemaking practices involve storage of the wine on lees. This facultative winemaking practice involves the conservation of the wine and yeast cells that have performed the alcoholic fermentation together, either in tanks or in barrels. The main objective of this step is the release of wine yeast constituents that influence the wine sensory properties. During storage on lees, cell proteins and nucleic material first undergo an intracellular enzymatic degradation due to the liberation of yeast intracellular proteases, amino peptidases, nucleases and phosphatases (Fornairon-Bonnefond *et al.*, 2001 ; Charpentier and Feuillat, 1993). The yeast cell gradually losses its hydrolyzed constituents by breakage of the cell wall. In the wine, the yeast constituents can be further degraded by proteases present in the extracellular media. Autolysis of the yeast cells is thus accompanied by an increase of the wine in amino acids and peptides (Martinez-Rodriguez *et al.*, 2000). Simultaneously, the protein content of the wine decreases as a result of the release of proteases and larger peptides are broken down, giving rise to smaller ones (Ari'Izumi *et al.*, 1994 ; Martinez-Rodriguez *et al.*, 2000).

Storing a wine on ML01 lees will therefore result in the release of very little *mleA* and *mae1* intact proteins, as autolysed yeast cells liberate small amounts of proteins (Martinez-Rodriguez *et al.*, 2000). Rather it will be their hydrolysis products such as peptides and amino acids which will be found in wines stored on lees.

Similarly, proteolysis products of the malolactic enzyme will be present in wines after malolactic fermentation by lactic acid bacteria. Like yeast, lactic acid bacteria loose their intracellular material which will be hydrolyzed by proteases present in the wine.

Proteolysis products of the malate permease protein can also be present in wines where *Schizosaccharomyces pombe* is part of the endogenous yeast flora of the wine. The more this species will participate in the alcoholic fermentation of must, the more *mae1* protein products will be found in the wine after storage on lees. An important participation of *Schizosaccharomyces* in wine alcoholic fermentation has been reported by Delphini and Formica (2001). Where this is not the case, storage of wine on lees of the ML01 strain can differ from that of the S92 strain by the presence of the *mae1* proteolysis products.

However, it must be kept in mind that the malate permease protein contains a PEST region which characterizes proteins with a rapid turn over (Grobler *et al.*, 1995). Hence, at the end of the fermentation, it is likely that very little malate permease proteins remain in the yeast cells due to the rapid turn-over of that protein. Moreover, if the hypothesis that the absence of malolactic fermentation at the stationary phase of cells is due to the absence of

malate permease or absence of *mae1* expression, is verified, this will also contribute to a low presence of the mae1 protein in wine.

Concerning yeast nucleic material, it can be said that the cellular content of an autolyzing yeast is rich in nucleases and phosphatases. The normal result of complete yeast autolysis is therefore that nucleic material will essentially be found in wine as single nucleic bases or small nucleic base chains. More generally, what has been enounced on yeast proteins seems also true for nucleic material.

Wine stabilization

After clarification, the wine must be stabilized. This step involves different procedures depending on whether it is red or white wine.

Bentonite treatment of white wines

Stabilization of white wine requires the removal of instable proteins because these may, upon bottling of the wine, form a protein haze (Ribéreau-Gayon *et al.*, 1998b). Elimination of these proteins is done by treating the wine with bentonite, a montmorillonite clay. Although only a certain part of proteins are responsible for formation of protein haze in wine, bentonite treatment removes equally all protein fractions of the wine (Dawes *et al.*, 1994).The amount of proteins precipitated upon bentonite treatment is function of the amount of bentonite used in the treatment (Dawes *et al.*, 1994).

If the wines are not stored on lees, yeast proteins and yeast protein products will not be found in wines. If the wines are stored on lees, yeast proteins will be liberated, but mostly as hydrolyzed fractions. However, the intact proteins as well as the bigger polypeptides may be removed from wine upon bentonite treatment.

Hence, although a small proportion of the *mleA* and *mae1* intact proteins should be present in wine, they will be partially removed like all wine proteins by a bentonite treatment, as well as they larger hydrolysis products.

Red wine stabilization

Red wines are not subject to protein haze like white wines, because proteins associate with tannins and little remain freely in the wine (Ribéreau-Gayon *et al.*, 1998). The protein-tannins complexes are partially eliminated after clarification with gelatins or egg white albumin which both possess the ability to precipitate colloidal structures (Ribéreau-Gayon *et al.*, 1998b). **If the *mleA* and *mae1* proteins are present in red wine, they will interact with the wine tannins and be partially removed upon clarification.**

The presence of the recombinant proteins mleA and mae1 should only occur significantly if the wine has been stored on lees after alcoholic fermentation. During storage on lees, mainly hydrolysis products of the mleA and mae1 proteins will be released in wine. Clarification processes, which are obligatory winemaking practices, will remove part of the mae1 and mleA proteins as well as larger polypeptides fragments of these proteins. Hence, only the smaller polypeptides and amino acids of these proteins will remain in the wine.

Products of the mleA protein can also be found in wine as a result of bacterial malolactic fermentation. Therefore, the only significant difference expected in between a control wine and a ML01 wine is the presence of the mae1 protein products. However, the rapid turn over of the malate permease as a consequence of the presence of a PEST region in its amino acid sequence should render this difference minimal.

Filtration

Filtration is an important step in wine clarification as it decreases the turbidity of the wine before bottling. This practice is not always used for red wines and some winemakers rely simply on clarification of the wine with gelatins or egg white albumin. However, the later procedure is rarely used and most wines are submitted to a filtration before being bottled.

Filtration of the wine can be done in several ways, but a perfect wine clarification will require several filtration steps.

Filtration on diatomaceous earth

This type of filtration is used for wines that have not been previously pre-filtered and is usually the first step in wine filtration. Different types of diatomaceous earth exist, some being more permeable than others. Filtration can remove a substantial proportion of wine yeast cells, the wine viable yeast cell count upon filtration being in the range of 10^3cells/100 milliliters of wine (Ribéreau-Gayon *et al.*, 1998b). After this type of filtration, the wine is either bottled immediately or submitted to another filtration step, usually involving cellulose filters or membrane filters.

Cellulose filters

This type of filtration is largely utilized just before bottling of the wine when the wine has been previously clarified either by gravity sedimentation, centrifugation, or diatomaceous earth filtration. Two types of filters exist, their use depend on whether a complete sterilization of the wine is wanted or just a fine clarification is needed. Clarification of wine using cellulose filters will allow the yeast cell count to drop to less than 50 cells/100 milliliters of wine (Ribéreau-Gayon *et al.*, 1998b). More efficient filters will decrease this count to less than one cell/100 milliliters of wine. A prerequisite for the efficiency of the filtration for microbial stability of the wine is the sterilization of the filter before use.

Membrane filters

Just before bottling, a membrane filtration of the wine can be performed. This type of filtration is particularly used when the wine is required to be sterile. Membrane filters can have different porosity, spanning from 0.45 micrometers to 10 micrometers. For yeast cells to be removed from wine, a porosity of less than 1.2 micrometers is needed. Hence, when a complete removal of yeast cells is required, these are the types of filters that should be used.

Standard and modern procedures

The modern and standard procedure consists in filtering wine in order to obtain a limpid solution having a very low NTU by turbidimetry, especially in the case of white wines and in order to obtain a wine as sterile as possible. Consequently, the modern wine procedure consists in filtering the wine with a porosity inferior to 1 micrometer, generally 0.45 micrometers for white wines, and 0.65 micrometers for red wine. The filtration eliminates all live yeast cells or ghost cells.

Previous studies have shown that the only significant difference in between the S92 and the ML01 wine yeast is the presence of the integration cassette, leading to the presence of the malate permease and malolactic enzyme which allow L-malate degradation by the yeast. Analysis of the wines showed that ML01 wines were not different in their global composition and organoleptic properties than control wines. Hence, the only suspected difference between a ML01 wine and a control wine, would be the possible presence of the ML01 recombinant yeast containing the malate permease and malolactic enzyme. These proteins, and most of all their hydrolysis products will essentially be found upon storage of the wines on lees. Moreover, the intact proteins and their larger polypeptides will be removed by obligatory clarification practices. Hence only the smaller peptides and amino acids will remain. Products of the mleA protein are also found in wines in which bacterial malolactic fermentation occurred. As for the mae1p products, one can note that malate permease is suspected for not further being present at the end of the alcoholic fermentation in reason of its probable quick turnover due to a PEST region, and the absence of malate degradation since the end of the exponential growth phase.

Depending on the clarification and the filtration process, different proportions of ML01 yeast cells might be found in wine. The most common practice is to filter the wine before bottling. Different filtering devices exist, each with different retention thresholds. Depending on the filtration unit utilized, yeast count can range from 10^3 cells/100 milliliters to less than one per bottle. The standard filtration procedure is the use of filter units containing membranes of porosity less than 1 micrometer allowing the complete removal of the yeast cells of wine.

4.4.3.4. Formation of unwanted substances in wine

Ethyl carbamate

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

We measured ethyl carbamate precursors in wines produced by ML01, S92 alone or S92 and where lactic acid bacteria performed malolactic fermentation. Results shown in Table 5 indicated that :

- the ML01 strain does not produce more ethyl carbamate precursors than the parental S92 strain. This shows that the genetic modification does not interfere with ethyl carbamate precursors metabolism. In yeast the main ethyl carbamate precursor is urea, which in wine is involved in arginine metabolism.

- wine in which malolactic fermentation was conducted by lactic acid bacteria contained more ethyl carbamate precursors than the wine where malolactic fermentation was carried out by the ML01 strain, showing that lactic acid bacteria played a non negligible role in the ethyl carbamate formation of the wine.

Hence, not only does the ML01 strain produces no more ethyl carbamate precursors than the parental strain, but the use of this modified yeast could help to reduce the ethyl carbamate content of wines where malolactic fermentation needs to be carried out by avoiding the formation of bacterial precursors during malolactic fermentation.

Biogenic amines

The presence biogenic amines in wines can be of major concern for consumers as they have been shown, and particularly histamine, to be the causative agent of allergenic symptoms such as head aches, vomiting, diarrhea, rashes, palpitations (Lonvaud-Funel, 2001). The presence of biogenic amines in wines is the result of bacterial activity, mainly lactic acid bacteria, which often perform the malolactic fermentation of wines. Metabolic pathways

leading to the formation of biogenic amines in wine are well known. Lactic acid bacteria histidine decarboxylase, tyrosine decarboxylase and ornithine decarboxylase activities all lead to the formation of biogenic amines from amino acids in wines.

Although the ML01 strain has been transformed with a lactic acid bacteria gene encoding for the malolactic enzyme, it is clear and well known that this enzymatic activity doesn't interfere with biogenic formation in wines.

Hence, because *Saccharomyces cerevisiae* has never been shown to participate to biogenic amine synthesis in wine and that the ML01 strain transformation does not involve genes of the biogenic amines formation pathways, there is no reason to think that the ML01 strain will produce more biogenic amines than the parental S92 strain.

Moreover, given the role of lactic acid bacteria in biogenic amine formation, one can reasonably assume that wines in which malolactic fermentation was performed by the ML01 yeast will contain less biogenic amines than wines in which it was conducted by lactic acid bacteria.

Other undesirable substances

The *mleA* and *mae1* genes present in the ML01 strain as the result of the genetic modification do not code for neither toxic or allergenic proteins, nor proteins implicated in the formation of undesirable compounds

We have carried out risks assessments related to the presence of potential unintended gene products as the result of the genetic modification :

- Computer analysis showed that the integrated region contained 7 new putative open reading frames

- Despite the fact that we showed that these ORFs were transcribed into mRNA, nothing indicates that they are subsequently translated into proteins. Moreover, among these 7 putative ORFs, only three have sequences of more than 100 codons (300 nucleotides), thus have reasonable chances of being translated into proteins

- These bigger ORFs mainly correspond to *Saccharomyces cerevisiae* sequences. Only a few nucleotides in each of these ORFs have an exogenous origin. 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 ML01 strain.

These considerations lead us to conclude that allergenic or toxic risks related to the presence of these putative ORFs within the new ML01 yeast strain were negligible.

Therefore, to the best of our knowledge, the use of the ML01 yeast strain in winemaking will not lead to the release or the enhancement of the presence of undesirable compounds in wine. Moreover, the undesirable compounds ethyl carbamate and biogenic amines may show reduced concentrations in wines produced with the ML01 strain compared to wines where the malolactic fermentation was carried out by lactic acid bacteria.

5- REFERENCES

Ansanay V., Dequin S., Camarasa C., Schaeffer V., Grivet P., Blondin B., Salmon J.M., Barre P. Malolactic fermentation by engineered *Saccharomyces cerevisiae* as compared with engineered *Schizosaccharomyces pombe*. Yeast. 1996. 12: 215-225

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

Barnett J.A. The taxonomy of the genus *Saccharomyces* Meyen *ex* Reess : a short review for non-taxonomists. Yeast. 1992. 8:1-23

Bony M., Bidard F., Camarasa C., Ansanay V., Dulau L., Barre P., Dequin . Metabolic analysis of *S. cerevisiae* strains engineered for malolactic fermentation. FEBS Lett. 1997. 410:452-456

Camarasa C., Bidard F., Bony M., Barre P., Dequin S. Characterization of *Schizosaccharomyces pombe* malate permease by expression in *Saccharomyces cerevisiae*. Appl Environ Microbiol. 2001. 67:4144-4151

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

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

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

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

Delfini C., Formica J.V. The use of selected yeast strains. In : *Wine Microbiology. Science and Technology*. New York : Marcel Dekker. 2001. 291-326

Dharmadhikari M.R., Wilker K.L. Deacidification of high malate must with *Schizosaccharomyces pombe*. Am J Enol Vitic. 1998. 49 (4):408-412

Dicks L.M., Dellagio F., Collins M.G. Proposal to reclassify Leuconostoc oenos as Oenococcus oeni [corrig.]. Int J Syst Bacteriol. 1995. 45:395-397

Flick G.J., Oria M.P., Douglas L. Potential hazards in cold-smoked fish : biogenic amines. J Food Sci. 2001. 66: S 1088-S1090

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

Gerbaux V., Briffox C. Influence de l'ensemencement en bactéries lactiques sur l'évolution de la couleur des vins de Pinot noir pendant l'élevage. Revue des Œnologues. 2002. 130:19-23

Gietz R.D., Sugino A. New yeast *Escherichia coli* shuttle vectors constructed with in vitro mutagenized yeast genes lacking six-base pair restriction sites. Gene.1988.74:527-534

Grobler J., Bauer F., Subden R.E., van Vuuren H.J.J. The *mae1* gene of *Schizosaccharomyces pombe* encodes a permease for malate and other C4 dicarboxylic acids. Yeast. 1995. 11:1485-1491

Groisiller A., Lonvaud-Funel A. Comparision of partial malolactic enzyme gene sequences for phylogenetic analysis of some lactic acid bacteria species and relationships with the malic enzyme. Int J Syst Bacteriol. 1999. 49:1417-1428

Henick-Kling T. Malolactic fermentation. In : *Wine Microbiology and Biotechnology.* Fleet G.H. (ed). Reading : Hardwood Academic. 1993. 289-326

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

Johnston J.R., Baccari C., Mortimer R.K. Genotypic characterization of strains of commercial wine yeasts by tetrad analysis. Res Microbiol. 2000. 151:583-590

Kreger-van-Rij N.J.W. (Ed). The yeasts : a taxonomy study. 3rd ed. Amsterdam : Elsevier Science Publishers B.V. 1984.1082p

Kurtzman C.P., Fell J.W. (Eds). The yeasts. A taxonomic study. 4th edition. Amsterdam : Elsevier Science B.V. 1998.1055p

Labarre C., Diviès C., Guzzo J . Genetic organization of the *mle* locus and identification of a *mleR*-like gene from *Leuconostoc oenos.* Appl Environ Microbiol. 1996b. 62 (12):4493-4498

Labarre C., Guzzo J., Cavin J.F., Diviès C. Cloning and characterization of the genes encoding the malolactic enzyme and the malate permease of *Leuconostoc Oenos.* Appl Environ Microbiol . 1996a. 62 (4): 1274-1282

Lautensach A., Subden R.E. Cloning of malic acid assimilating activity from *Leuconostoc oenos* in *E. coli.* Microbios. 1984. 39:29-39

Leupold U. Die vererbung von homothallic und heterothallic bei *Schizosaccharomyces pombe.* C R Trav Lab Carlsberg, Ser Physiol. 1950. 24:381-480

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

Lindner P. *Schizosaccharomyces pombe* n.sp. neuer Gärungserreger. Wochenschr Brau. 1893. 10:1298-1300

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

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

Lonvaud-Funel A. Lactic acid bacteria in the quality improvement and depreciation of wine. Antonie van Leeuw. 1999. 76:317-331

Lonvaud-Funel A. Biogenic amines in wines : role of lactic acid bacteria. FEMS Microbiol Lett. 2001. 199 (1):9-13

Machiewicz P. Kowalczuk M., Machiewicz D., Nowicka A., Dudkiewicz M., Laszkiewicz A., Dudek M., Cebrat S. How many protein-coding genes are there in the *Saccharomyces cerevisiae* genome ? Yeast. 2002. 19:619-629.

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

Masneuf I., Aigle M., Dubourdieu D. Development of a polymerase chain reaction/restriction fragment lengh polymorphism method for *Saccharomyces cerevisiae* and *Saccharomyces bayanus* identification in enology. FEMS Microbiol. 1996. 138:239-244

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

Naouri P., Chagnaud P., Arnaud A., Galzy P. Purification and properties of a malolactic enzyme from *Leuconostoc oenos* ATCC23278. J Basic Microbiol. 1990. 30 (8):577-585

Ness F., Lavallée F., Dubourdieu D., Aigle M., Dulau L. Identification of yeast strains using the polymerase chain reaction. J Sci Food Agric. 1993. 62:89-94

000041

Nguyen H-V., Lepingle A., Gaillardin C. Molecular typing demonstrates homogeneity of *Saccharomyces uvarum* strains and reveals the existance of hybrids between *S. uvarum* and *S. cerevisiae* including the *S. bayanus* type strain CBS 380. Syst Appl Microbiol. 2000. 23:71-85

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

Osothsilp C., Subden R.E. Malate transport in *Schizosaccharomyces pombe.* J Bacteriol. 1986. 168 (3):1439-1443

Osterwalder A. *Schizosaccharomyces liquefaciens* n.sp., eine gegen freie schweflige Säure widerstandsfähige Gärhefe. Mitt Geb Lebensmitt Hyg. 1924. 15:5-28

Pretorius I.S. Tailoring wine yeast for the new millennium : novel approches to the ancient art of winemaking. Yeast. 2000.16:675-729

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

Radler F. Metabolism of organic acids. In : *Wine microbiology and Biotechnology.* Fleet G.H. (ed). Reading : Hardwood Academic. 1993.165-182

Reed R., Nagodawithana T.W. Yeast Technology. 2nd edition. New York : Van Nostrand Reinhold. 1991.454p

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

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

Salmon J-M. L-malic acid permeation in resting cells of anaerobically grown *Saccharomyces cerevisiae.* Bioch Biophys Acta. 1987. 901:30-34

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

Snow S.R., Kunkee R.E., Hodges R.A., Williams S.A. Malolactic gene. US4472502. Sep. 18, 1984.

Sousa M.J., Mota M., Leao C. Transport of malic acid in the yeast *Schizosaccharomyces pombe* : evidence for a proton-dicarboxylate transport. Yeast. 1992. 8:1025-1031

Spetolli P., Nuti M.P., Zamorani A. Properties of malolactic activity purified from *Leuconostoc oenos* ML34 by affinity chromatography. Appl Environ Microbiol. 1984. 48 (4):900-901

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.* Unpublished.

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

Vivas N., Bellemere., Lonvaud-Funel A., Glories Y., Augustin M. Etudes sur la fermentation malolactique des vins rouges en barriques et en cuves. R F Œnologie. 1994. 149:37-42

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

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

Williams S.A., Hodges R.A., Strikes T.L., Snow R., Kunkee R.E. Cloning of the gene for the malolactic fermentation of wine from *Lactobacillus delbruekii* in *Escherichia coli* and yeasts. Appl and Environ Microbiol. 1984. 47:288-293

Williams S.T., Sharpe M.E., Holt J.C. (Eds). Bergey's manual of systematic bacteriology. Volume 4. Baltimore, USA : Williams & Wilkins. 1989.2299-264

Dr Didier COLAVIZZA
Scientific Manager
Lesaffre International

Signature

M. Bob BIWERSI
Project Director
Lesaffre Yeast Corporation

Signature

000043

Figures and Tables

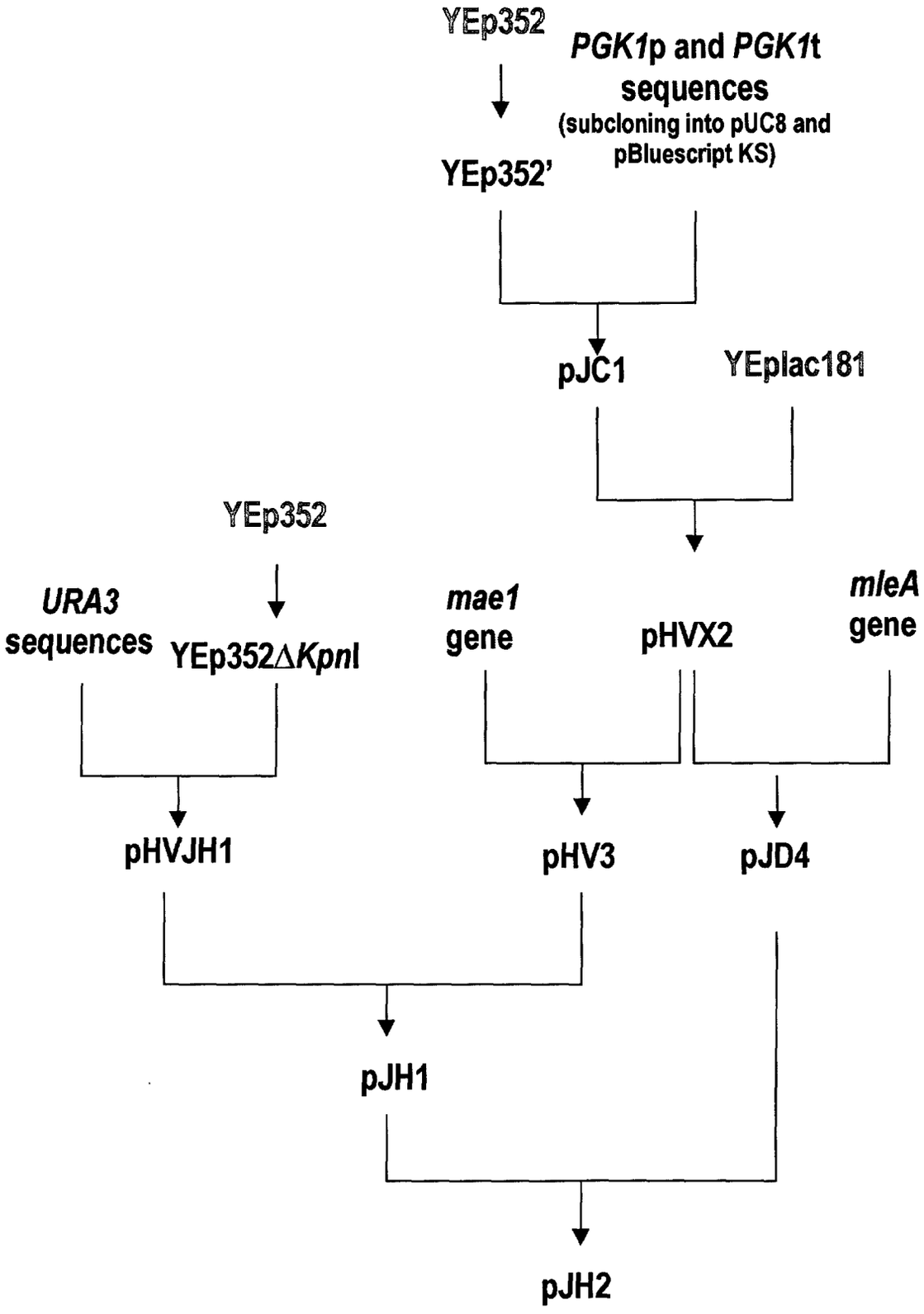


Figure 1 : Flow diagram illustrating the cloning strategy used to construct plasmid pJH2.

Standard plasmids from which all other plasmids used in this study are derived from are in red.Genomic sequences originating from donor organisms are shown in blue. All plasmids are listed in Table 2.

000045

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

Designation	DNA sequence	Description
J9	5'GATC**TCTAGA**GCCCGGGCAACGGTTCATCATCTCATGGATC TGC3'	5'-primer for *URA3* upstream flanking sequence with ***Xba*I** and *SrfI* sites
J10	5'GATC**GGTACC**TACTTCTTCCGCCGCCTGCAAACCGCCT3'	3'- primer for *URA3* upstream flanking sequence with ***Kpn*I** site
J11	5'GATC**GGTACC**GCGGCCGCACAAAGGAACCTAGAGGCCTTT TGATGTTAG3'	5'- primer for *URA3* downstream flanking sequence with ***Kpn*I** and *NotI* sites
J12	5'GATC**TCTAGA**GCCCGGGCTACACCAGAGATACATAATTAGA TAT3'	3'- primer for *URA3* downstream flanking sequence with ***Xba*I** and *SrfI* sites
J13	5'AACCAAAAAT**GGTACC**AAGCTTTCTAACTGATCTATCCAAA ACTGA3'	5'-primer used to amplify the *mae1* expression cassette from pHV3, building in a ***Kpn*I** site
J14	5'AAGGAAAAAA**GGTACC**AAGCTTTAACGCAGAATTTCGAGT T3'	3'-primer used to amplify the *mae1* expression cassette from pHV3, building in a ***Kpn*I** site
J15	5'ATGCTTTTTT**GCGGCCGC**AAGCTTTCTAACTGATCTATCCA AAACTGAAA3'	5'-primer used to amplify the *mleA* expression cassette from pJD4, building in a ***Not*I** site
J16	5'ATGCATATAA**GCGGCCGC**AAGTTTAACGAACGAACGCAGA ATTTTCGAGTTAT3'	3'-primer used to amplify the *mleA* expression cassette from pJD4, building in a ***Not*I** site
J17	5'GAAGATAAA**GATCT**TG3'	Used for site directed mutagenesis, introducing a ***BglI*I** site at the end of PGK1p
J18	5'GAGGA**GAATTC**ATGACAGATCCAG3'	5'-primer used to amplify the *mleA* gene from *Oenococcus œni* genomic DNA, building in a ***Eco*RI** site
J19	5'ACG**AGATCT**CATTAGTATTTCGGATCC3'	3'-primer used to amplify the *mleA* gene from *Oenococcus œni* genomic DNA, building in a ***BglII*** site
J20	5'AATGACCGACCAAGCGACG3'	5'-primer used to amplify the PhleoR probe from plasmid pUT332
J21	5'ATCCTGGGTGGTGAGCAG3'	3'-primer used to amplify the PhleoR probe from plasmid pUT332
Linker 1	5'GAAGATAA**AGATCT**TG3'	Linker used in construction strategy of plasmid pJC1, containing a ***BglII*** site
Linker 2	5'**CCTCGAGG**AGATCTGGAATTCC3'	Linker used in construction strategy of plasmid pC1, containing a ***Xho*I** site, a *BglII* site and an *EcoRI* site

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

Designation	Description	Source
pBluescript II KS	High copy number phagemid	Short *et al.*, 1988
YEp352	*Escherichia coli*/yeast episomal shuttle vector	Hill *et al.*, 1986
YEp352'	YEp352 in which the *EcoR*I site has been destroyed	Crous *et al.*, 1995
YEp352Δ*Kpn*I	YEp352 in which the *Kpn*I site has been destroyed	This study
YEplac181	*Escherichia coli*/yeast episomal shuttle vector containing the Tn5*Ble* dominant marker for resistance to phleomycin	Gietz and Sugino, 1988
pUC8	*Escherichia coli* expression vector encoding the beta-galactosidase alpha peptide (lacZ'), permitting blue-white visual detection of recombinants	Messing and Viera, 1982
pUT332	*Escherichia coli*/yeast episomal shuttle vector	Gatignol *et al* ., 1990
pJC1	YEp352 derived plasmid containing the *PGK1* promoter and terminator sequences	Crous et al., 1995
pHVX2	YEplac181 derived plasmid containing the *PGK1* promoter and terminator sequences	Volschenk *et al.*, 1997a
pHV3	pHVX2 containing the *mae1* gene subcloned between the *PGK1* promoter and terminator sequences	Volschenk *et al.*, 1997a
pJD4	pHVX2 containing the *mleA* open reading frame subcloned between the *PGK1* promoter and terminator sequences	Volschenk *et al.*, unpublished
pHVJH1	YEp352(Δ*Kpn*I) containing the *URA3* flanking sequences	This study
pJH1	pHVJH1 containing the *mae1* cassette cloned between the *URA3* flanking sequences	This study
pJH2	pHVJH1 containing the *mae1* and the mleA cassettes cloned between the *URA3* flanking sequences (subcloned from pJH1)	This study

000047

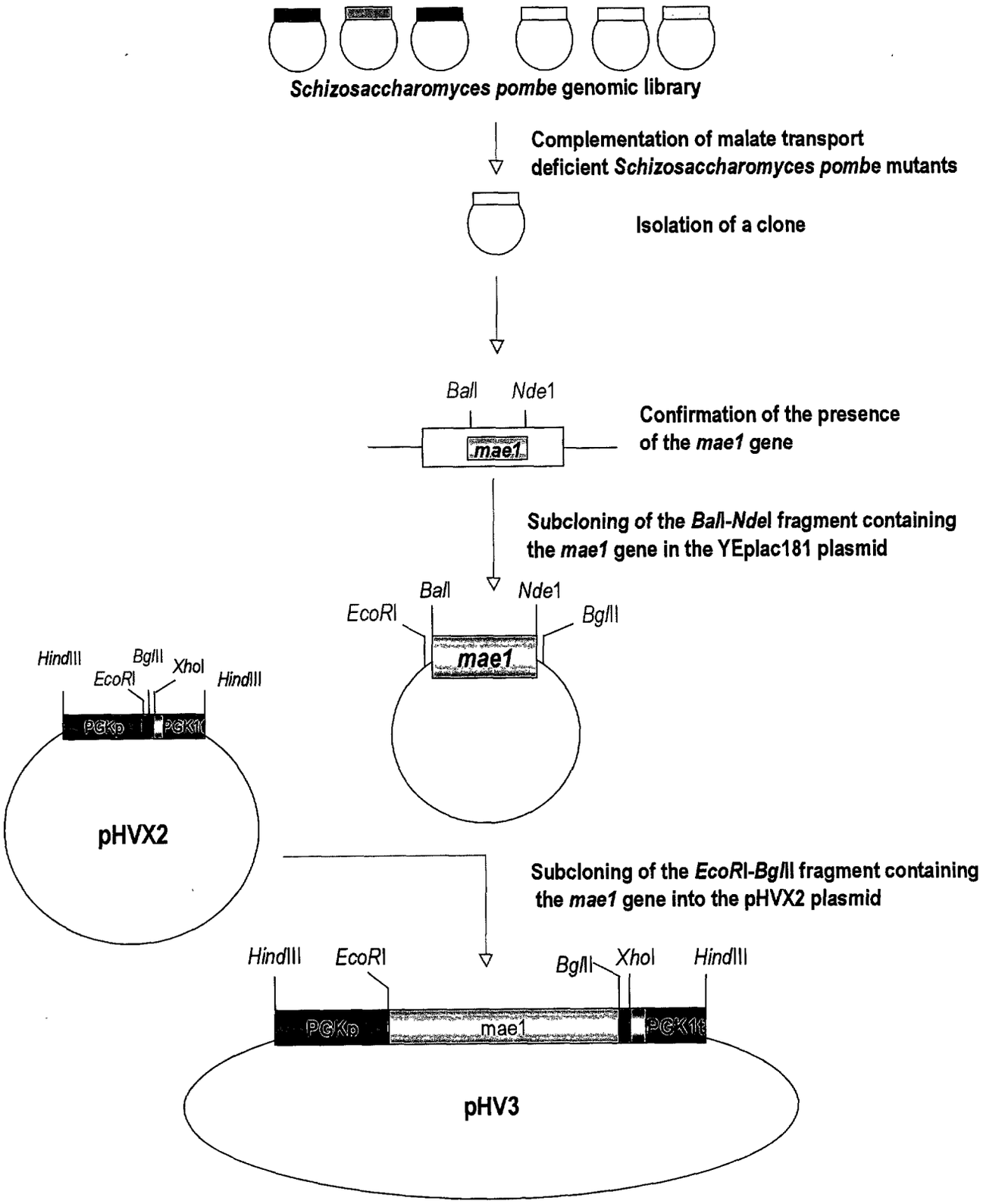


Figure 2 : Cloning of the *mae1* gene and construction of plasmid pHV3.

000048

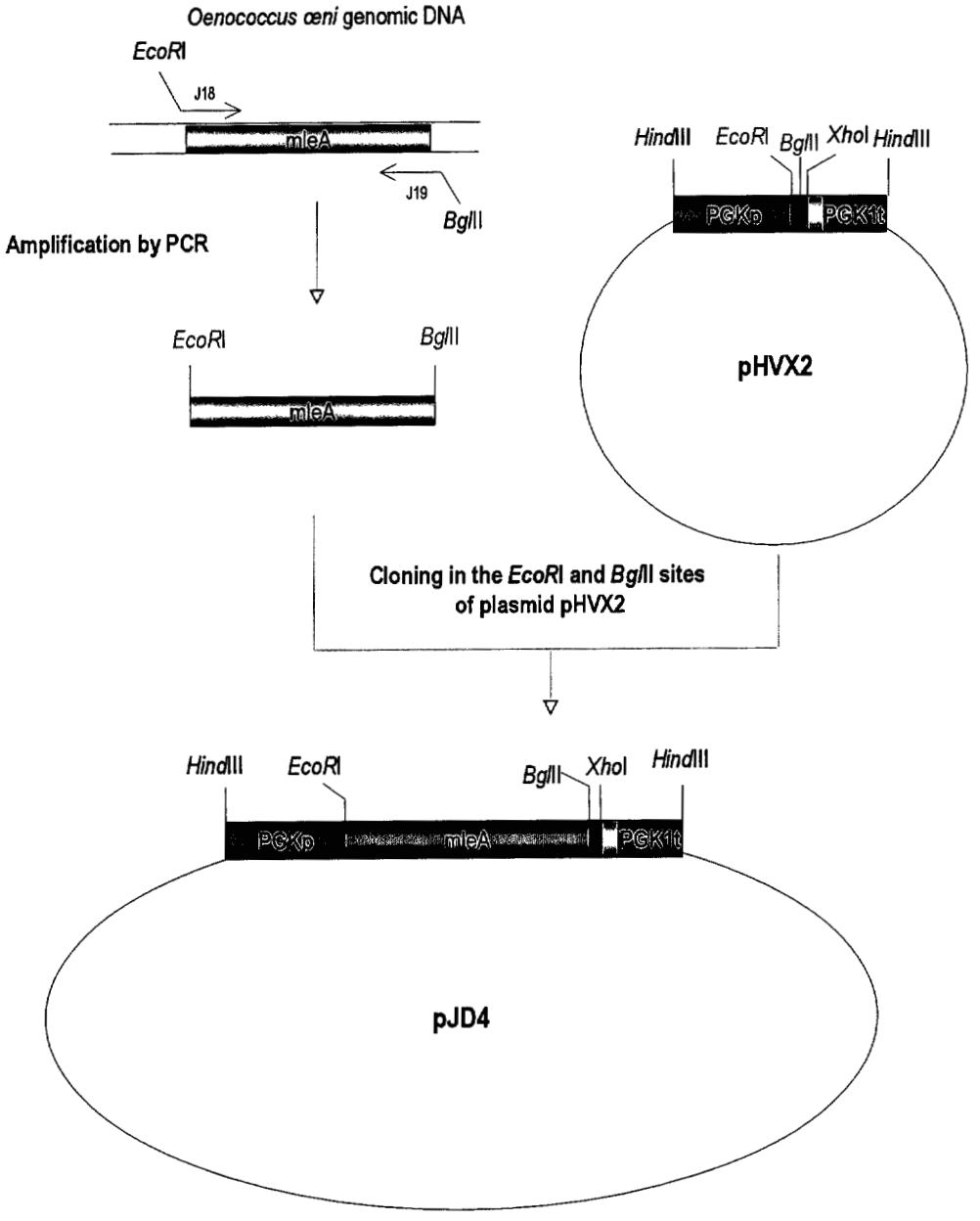


Figure 3 : Cloning of the *mleA* gene and construction of plasmid pJD4.

Primers are described Table 1.

000049

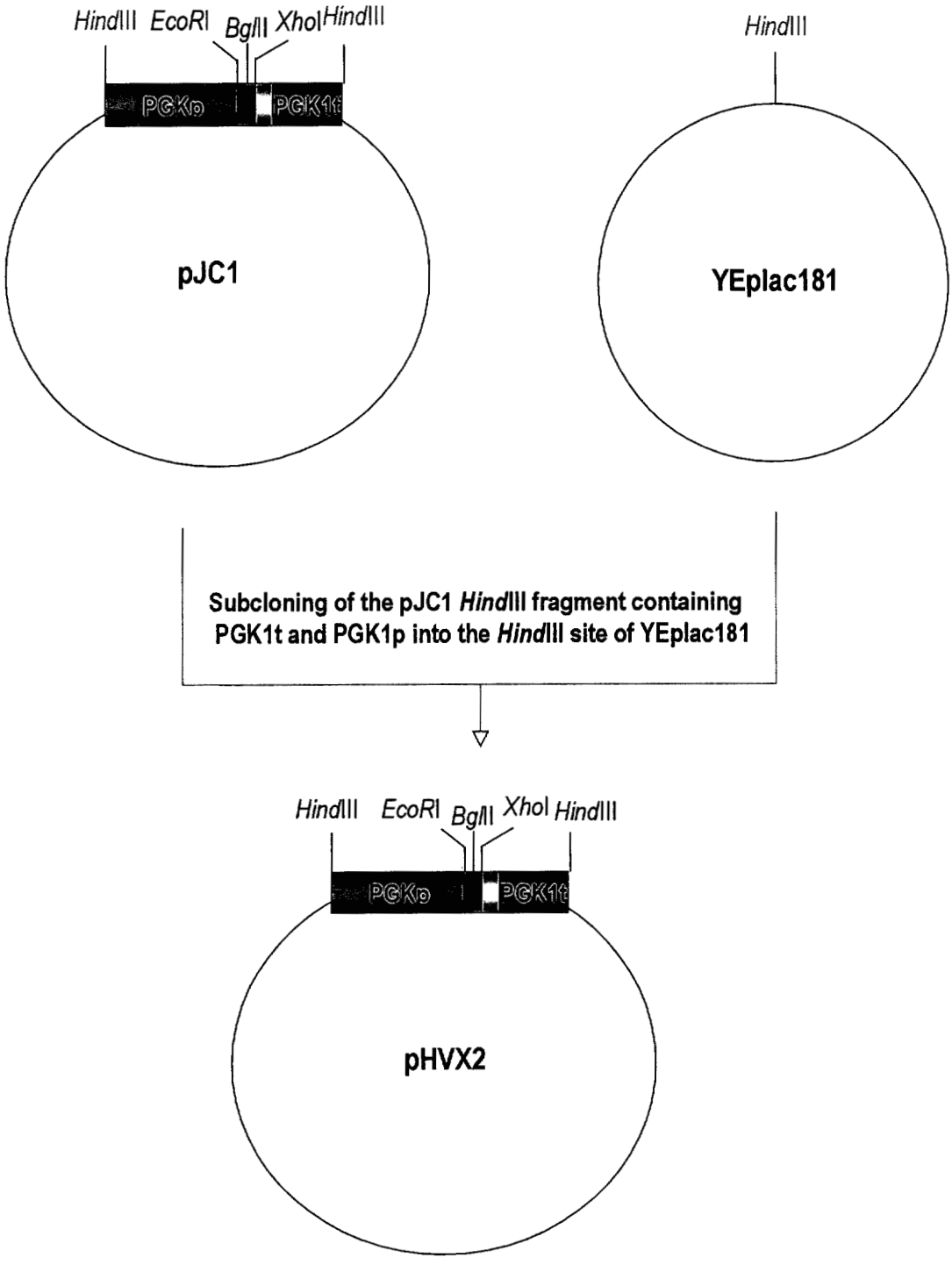


Figure 4 : Construction of plasmid pHVX2.

000050

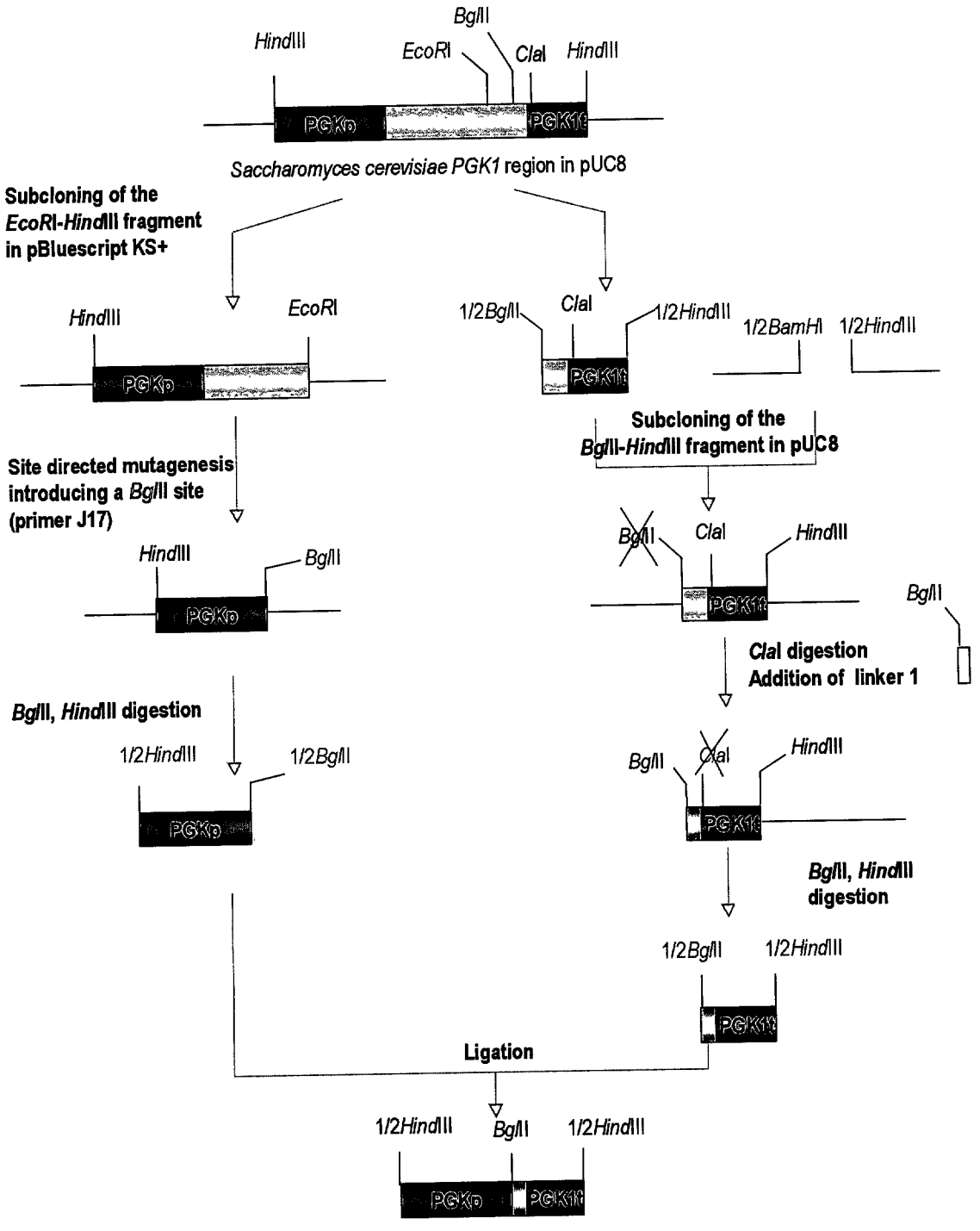


Figure 5 : Construction of plasmid pJC1.

000051

Primers and linkers are described Table 1.

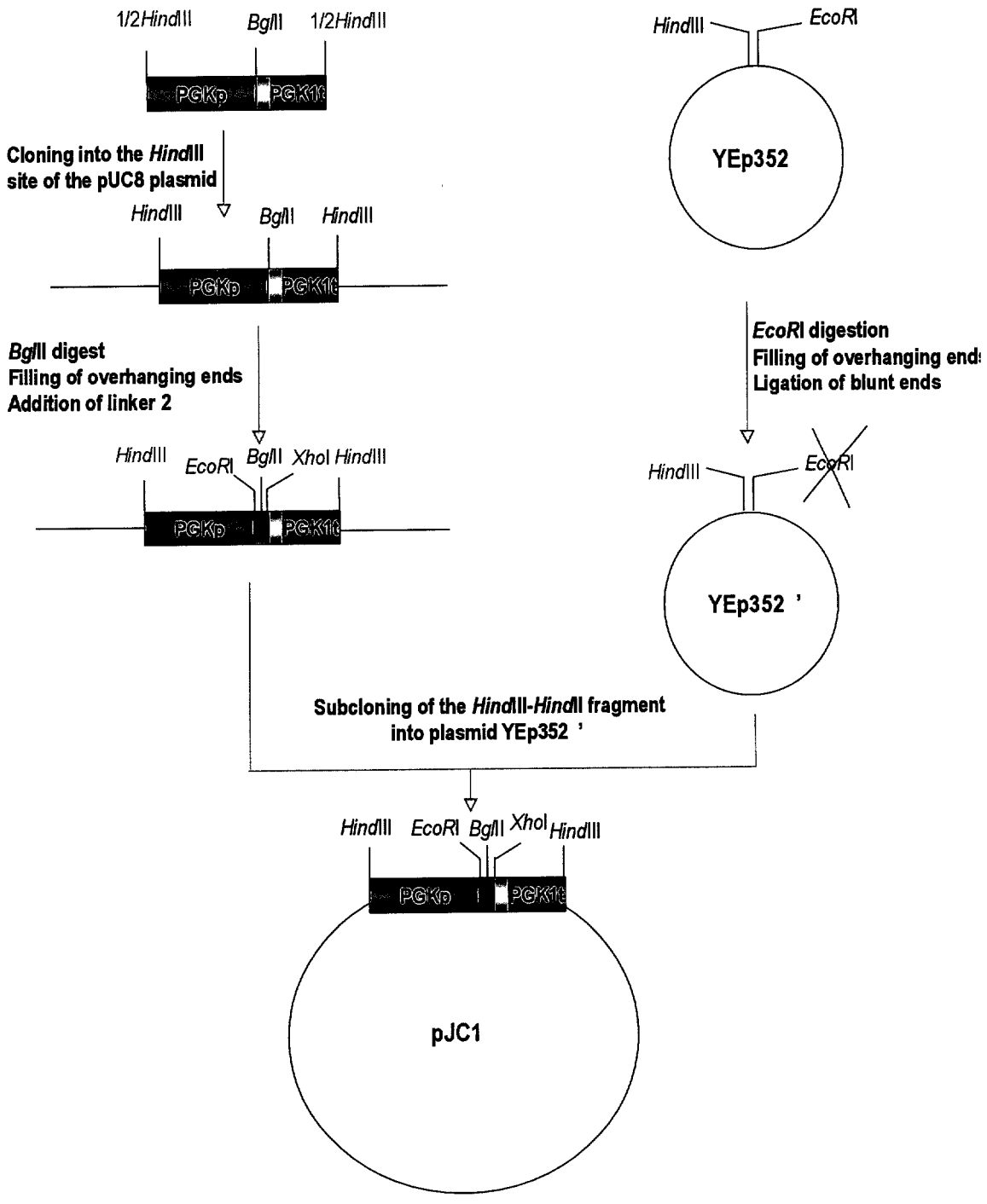
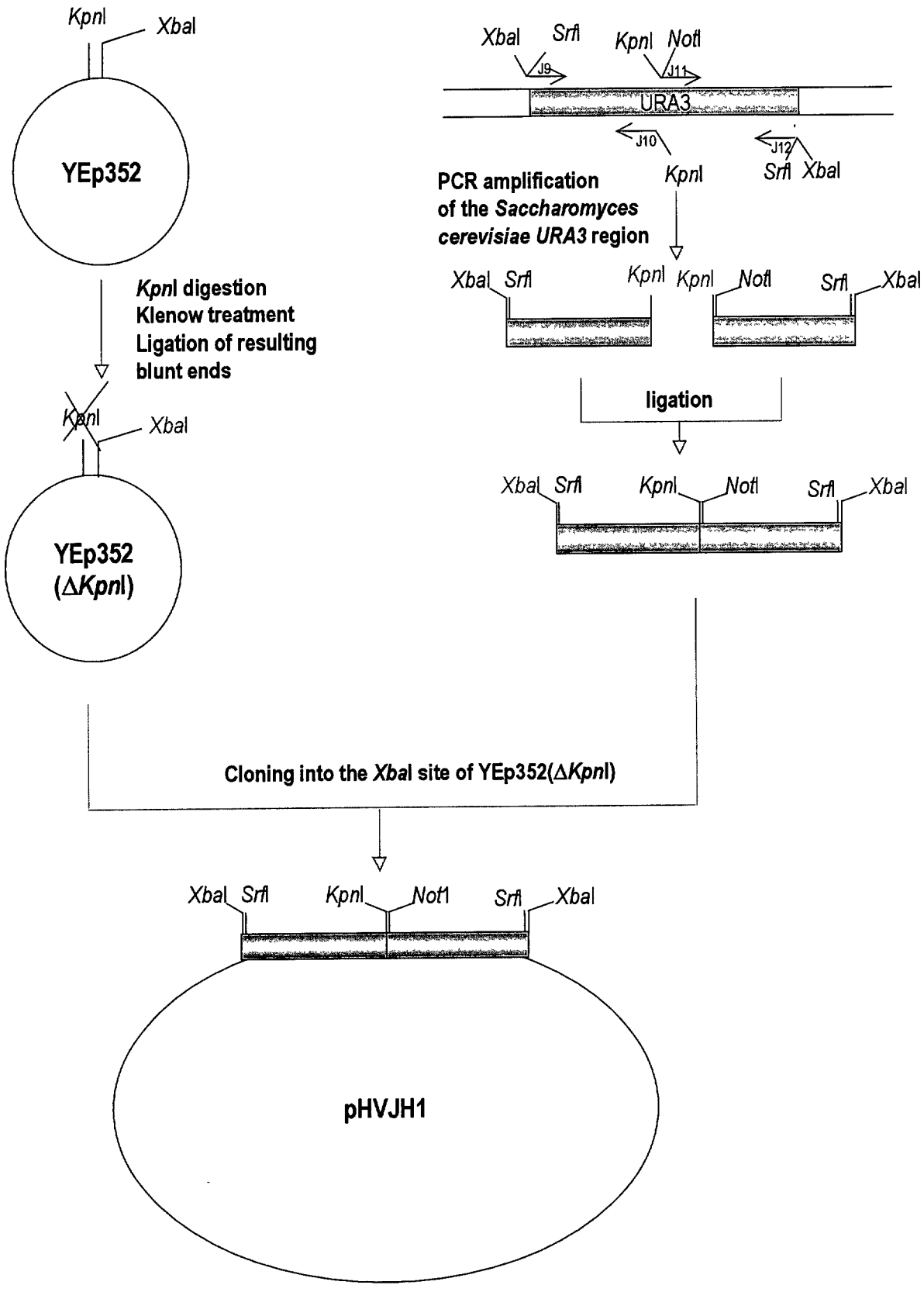


Figure 5 (continued) : construction of plasmid pJC1.

Linkers are described Table 1.

000052



000053

Figure 6 : Construction of plasmid pHVJH1.

Primers and linkers are described Table 1

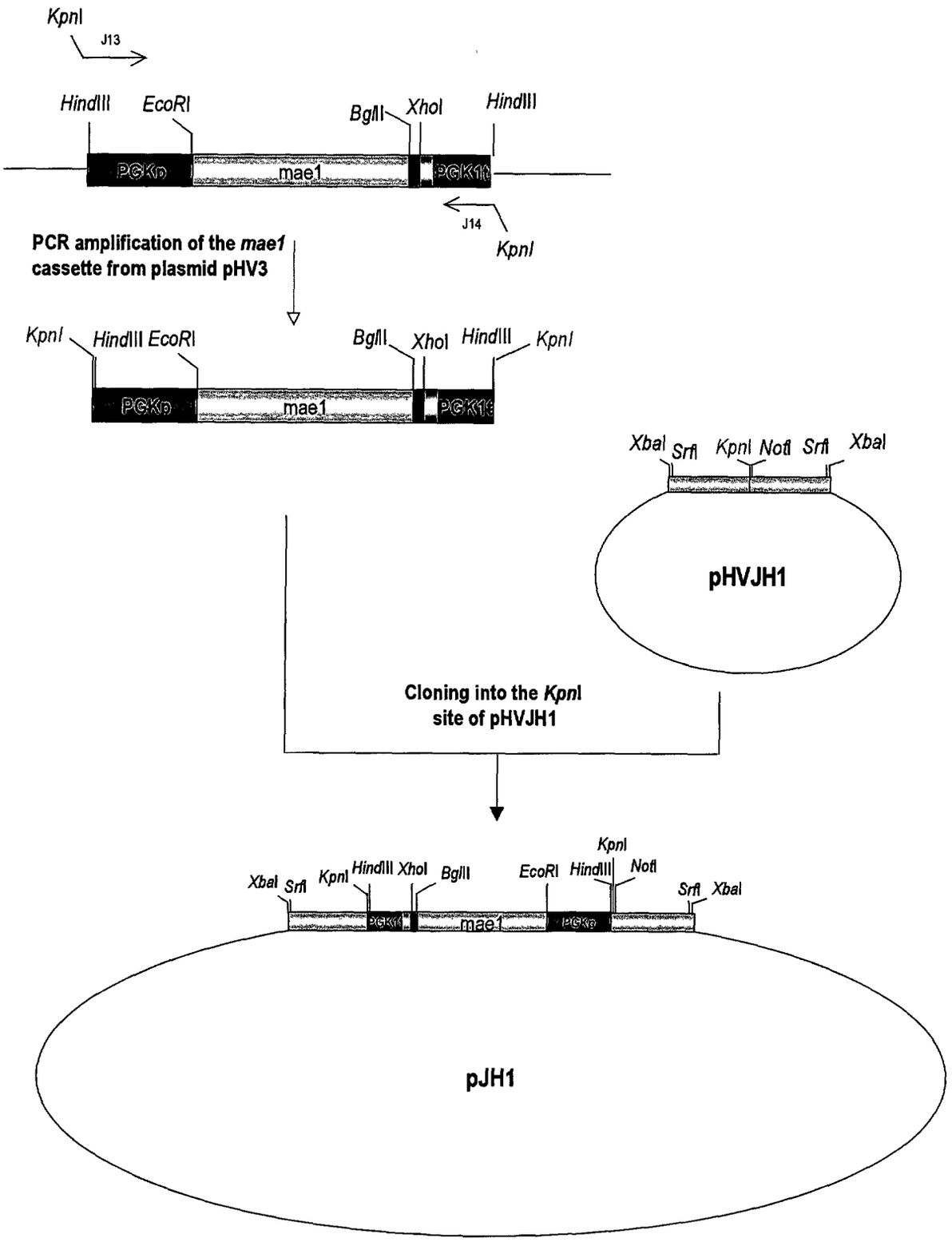


Figure 7: Construction of plasmid pJH1.

Primers are described Table 1.

000054

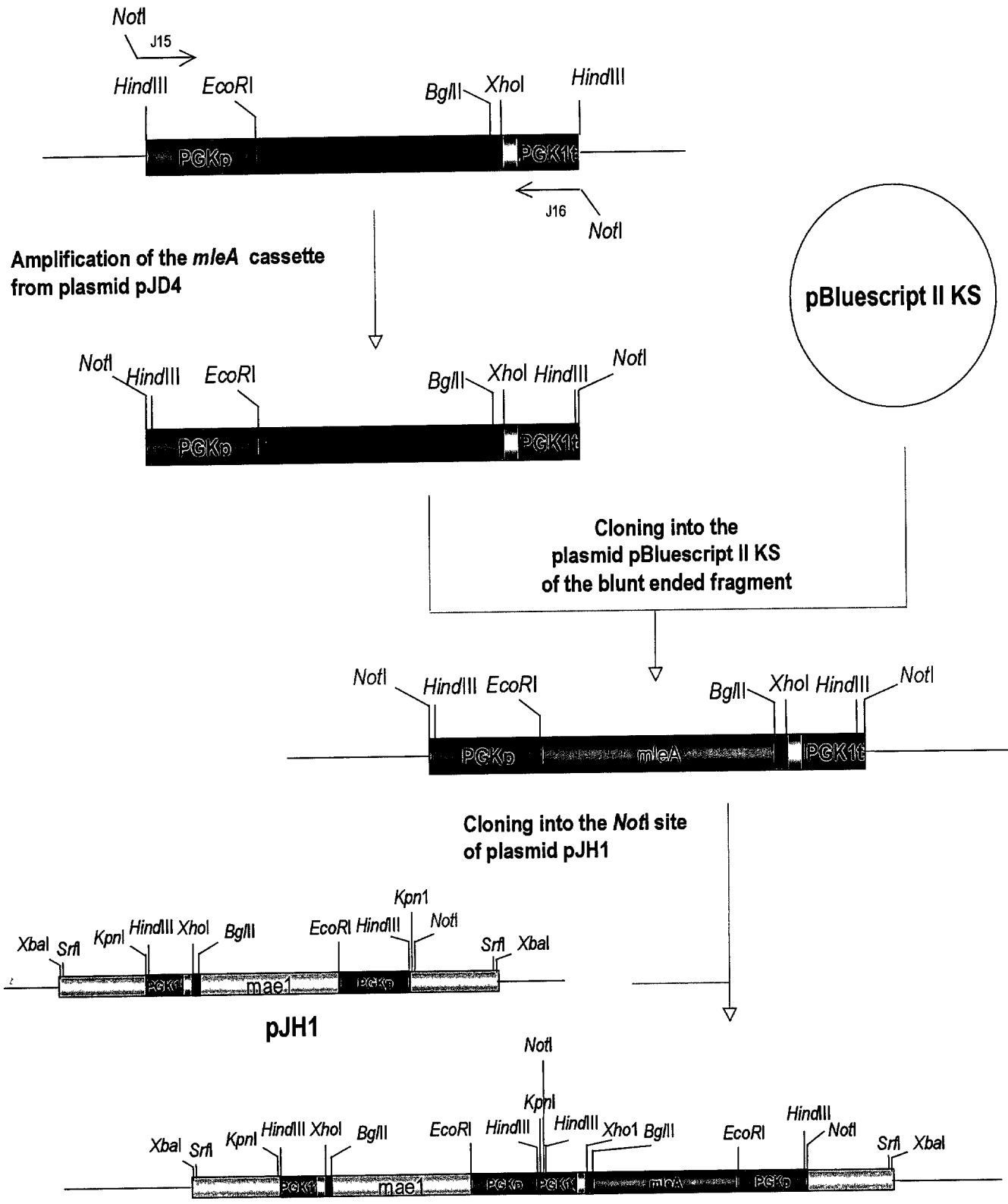


Figure 8: Construction of plasmid pJH2.

Primers are described Table 1.

000055

Malolactic cassette

mae1 expression cassette mleA expression cassette

Integration cassette

URA3 flanking sequence Malolactic cassette URA3 flanking sequence

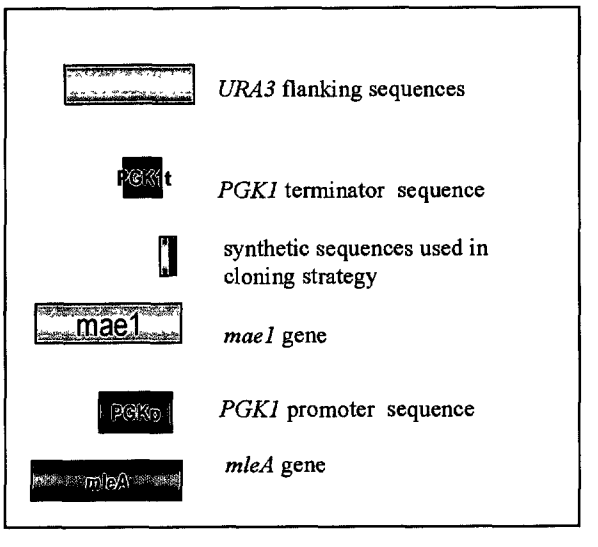


Figure 9 : Schematic representation of the introduced genetic material.

000056

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

Nucleotide position	Designation	Reference for cloning details	Source
1- 3	***Srfl* half cloning site**	**This study**	**Synthetic**
4-928	***URA3* sequence**	This study	***Saccharomyces cerevisiae* GC210**
4-508	5' non coding sequence		
509-928	Part of open reading frame		
929-934	***KpnI* cloning site**	**This study**	**Synthetic**
935-1198	*PGK1* terminator	**Crous *et al.*, 1995**	***Saccharomyces cerevisiae* AB972**
935-940	*Hind*III site		
941-1194	Rest of *PGK*1t		
1195-1198	Remaining of *Cla*I site		
1199-1218	**Part of linker used in cloning strategy**	**Crous *et al.*, 1995**	**Synthetic**
1199	C residue from linker		
1200-1204	remaining of *Bgl*II cloning site		
1205	C residue		
1206-1211	*Xho*I cloning site		
1212	G residue		
1213-1218	*Bgl*II cloning site		
1219-2818	***mae1* gene**	Grobler *et al.*, 1995 **Volschenk *et al.*, 1997a and 1997b**	***Schizosaccharomyces pombe* 972 h⁻ (leu1-32)**
1219-1224	*Nde*I site		
1225-1456	3' non coding region		
1457-1459	STOP codon		
1460-2770	Coding region		
2771-2773	START codon		
2774-2812	5' non coding region		
2813-2818	*Bal*I site		
2819-2830	Part of linker used in cloning strategy	**Crous *et al.*, 1995**	**Synthetic**
2819-2824	*EcoR*I cloning site		
2825	C residue		
2826-2830	remaining of *Bgl*II cloning site		
2831-4316	*PGK1* promoter	**Crous *et al.*, 1995**	***Saccharomyces cerevisiae* AB972**

000057

2831-4310	Part of *PGK1*p		
4311-4316	*Hind*III cloning site		
4317-4322	***Kpn*I cloning site**	**This study**	**Synthetic**
4323-4330	***Not*I cloning site**	**This study**	**Synthetic**
4331-4594	***PGK1* terminator**	**Crous *et al.*, 1995**	***Saccharomyces cerevisiae* AB972**
4331-4336	*Hind*III site		
4337-4590	Rest of PGK1t		
4591-4594	Remaining of *Cla*I site		
4595-4600	Part of linker used in cloning strategy	**Crous *et al.*, 1995**	**Synthetic**
4595	C residue		
4596-4600	remaining of *Bgl*II cloning site		
4601	C residue		
4602-4607	*Xho*I cloning site		
4608	G residue		
4609-4614	*Bgl*II cloning site		
4615-4616	**CA residues left from oligonucleotide used for *mleA* amplification**	**Volschenk, unpublished**	**Synthetic**
4617-6242	***mleA* gene**	Volschenk, unpublished	***Oenococcus œni* Lo 8413**
4617-4619	STOP codon		
4620-6239	*mleA* open reading frame		
6240-6242	START codon		
6243-6253	Part of linker used in cloning strategy	**Crous *et al.*, 1995**	**Synthetic**
6243-6248	*EcoR*I cloning site		
6249	C residue		
6250-6253	remaining of *Bgl*II cloning site		
6254-7740	***PGK1* promoter**	**Crous *et al.*, 1995**	***Saccharomyces cerevisiae* AB972**
6254-7734	Part of *PGK1*p		
7735-7740	*Hind*III cloning site		
7741-7747	***Not*I cloning site**	**This study**	**Synthetic**
7748-8680	*URA3* sequence	This study	***Saccharomyces cerevisiae* GC210**
7748-8132	Part of open reading frame		
8133-8680	3' non coding region		
8681-8683	***Srf*I half cloning site**	**This study**	**Synthetic**

000058

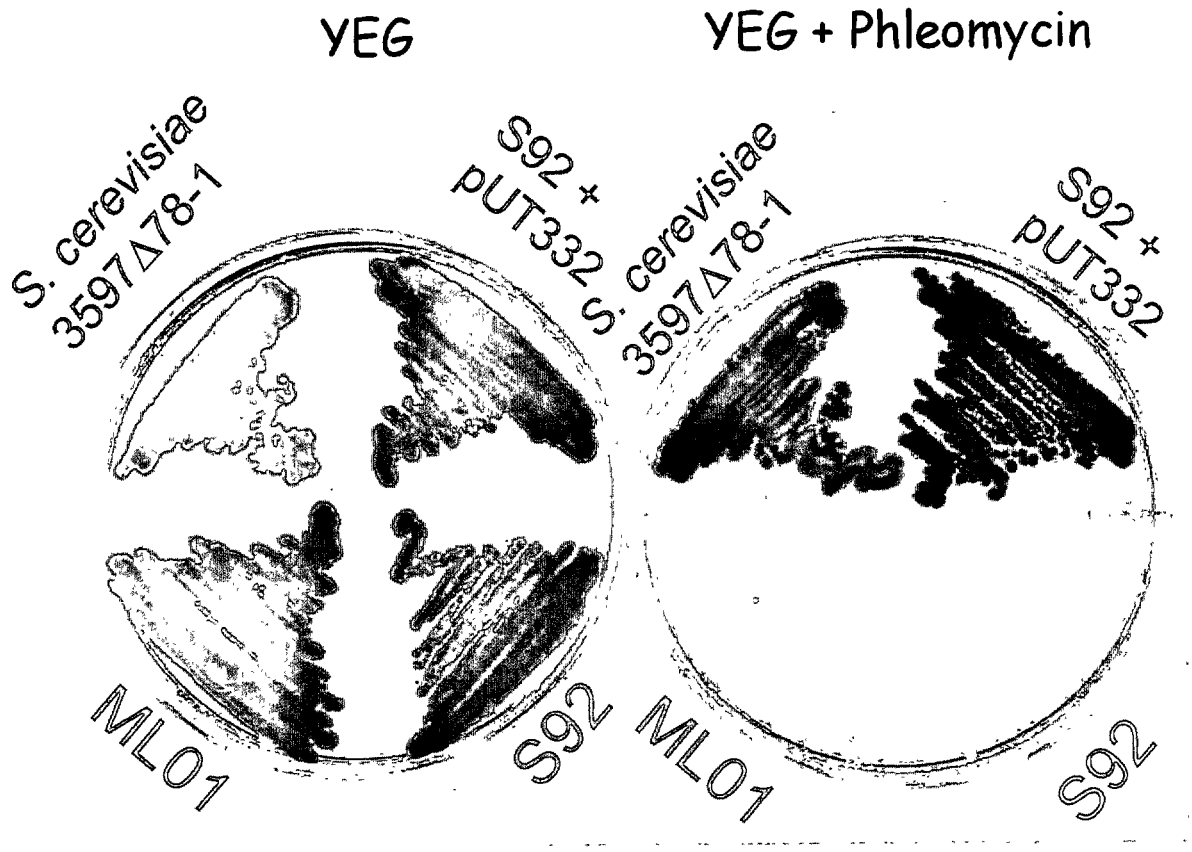


Figure 10 : Loss of phleomycin resistance.

000059

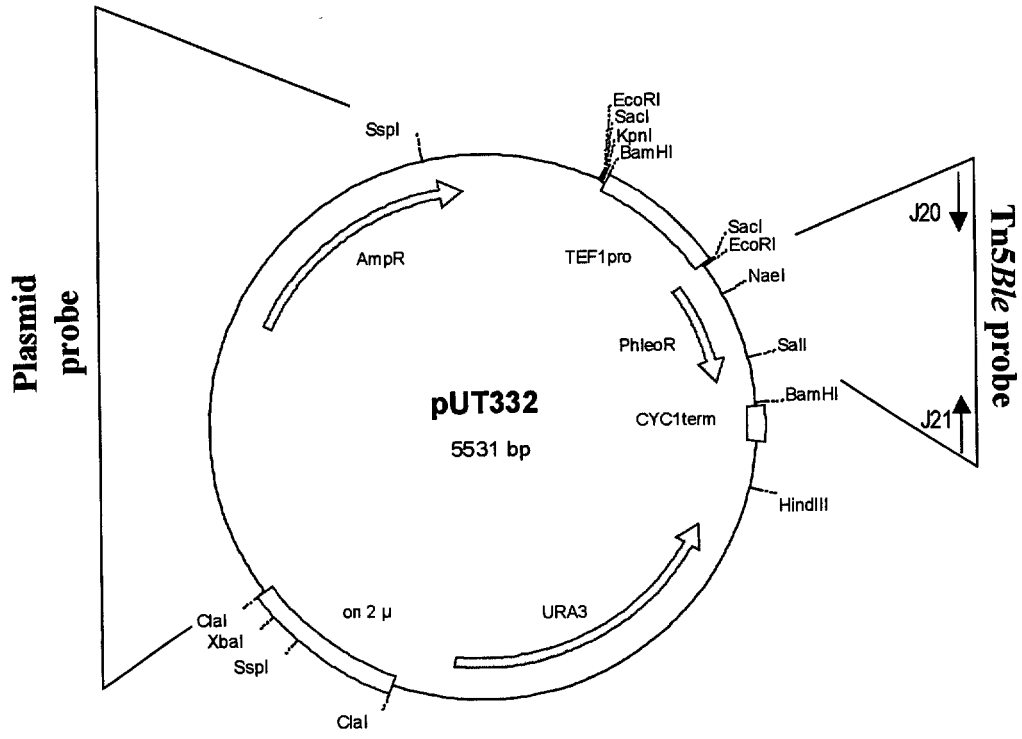


Figure 11: Plasmid pUT332 map and schematic representation of the probes used in Southern experiments.

The plasmid probe was retrieved as a *Cla*I-*Ssp*I fragment after restriction digest of the pUT332 plasmid. This probe represents all bacterial borne sequences of plasmid pUT332 which traces directly to plasmid pUC19 (Gatignol *et al.*, 1990). The Tn5*Ble* probe was retrieved by PCR amplification of a region of the pUT332 plasmid using primers J20 and J21 (Table 1).

000060

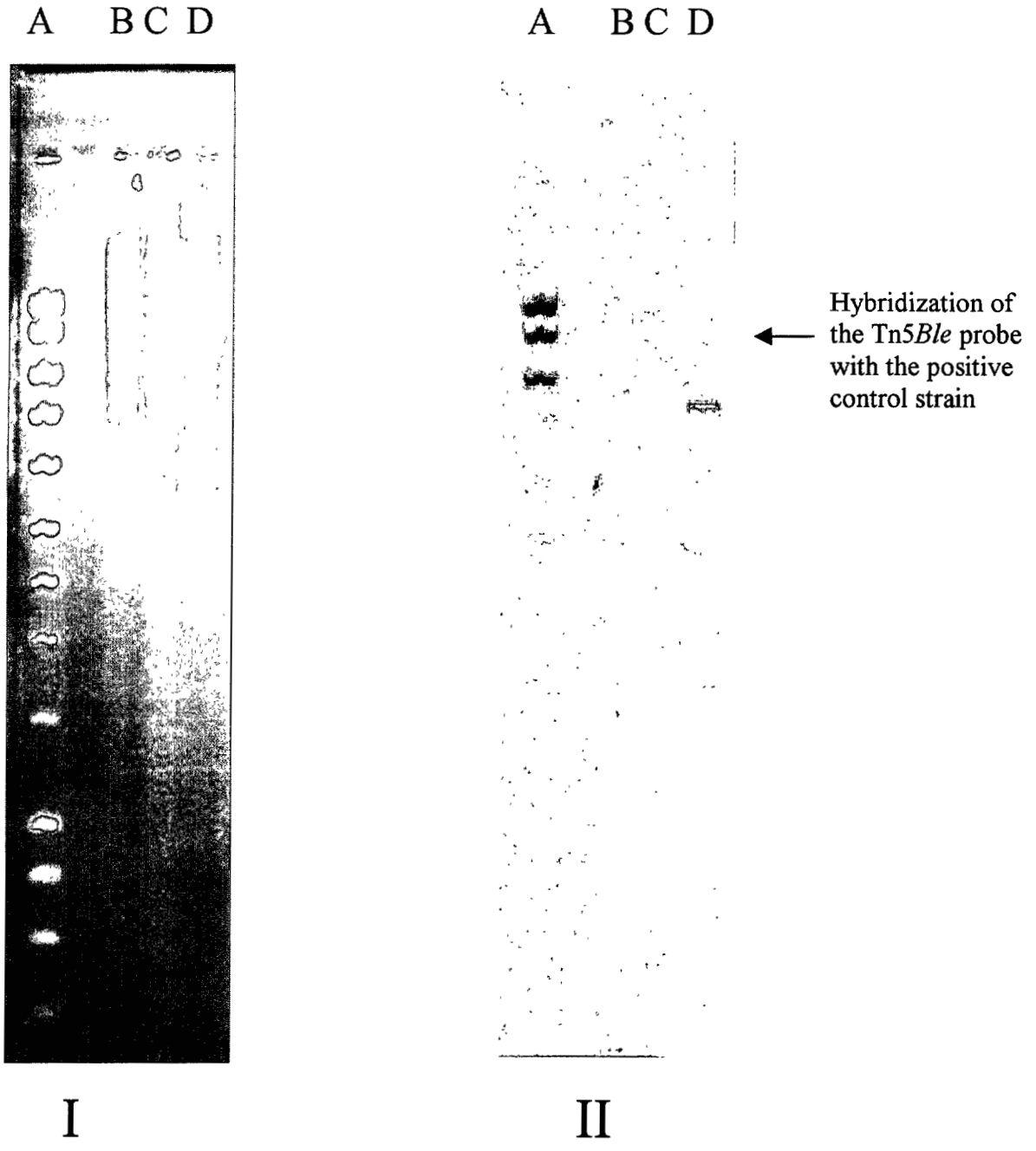


Figure 12 : Absence of the Tn5*Ble* gene in the ML01 strain.

Genomic DNA was digested with the *Nco*I restriction enzyme and separated onto an agarose gel. The host strain S92 was used as a negative control and a strain carrying an integrated Tn5*Ble* gene (strain 3597 Δ78-1) was used as a positive control.

Panel I: Genomic digest of strains ML01 (lane B), S92 (lane C) and *Saccharomyces cerevisiae* 3597 Δ78-1 (lane D). Lane A, DNA Ladder.

Panel II: Hybridization of genomic DNA from strains ML01 (lane B), S92 (lane C) and *Saccharomyces cerevisiae* 3597 Δ78-1 (lane D) with the Tn5*Ble* probe. Lane A, DNA Ladder. Probe labeling, hybridization and detection were done using the ECL direct nucleic acid labeling and detection system (Amersham Pharmacia). A description of the probe used is given Figure 11.

In our assay conditions the hybridization experiment allows to detect an integrated copy of the Tn5*Ble* gene in the yeast genome of a positive control strain and not in the genome of strain ML01.

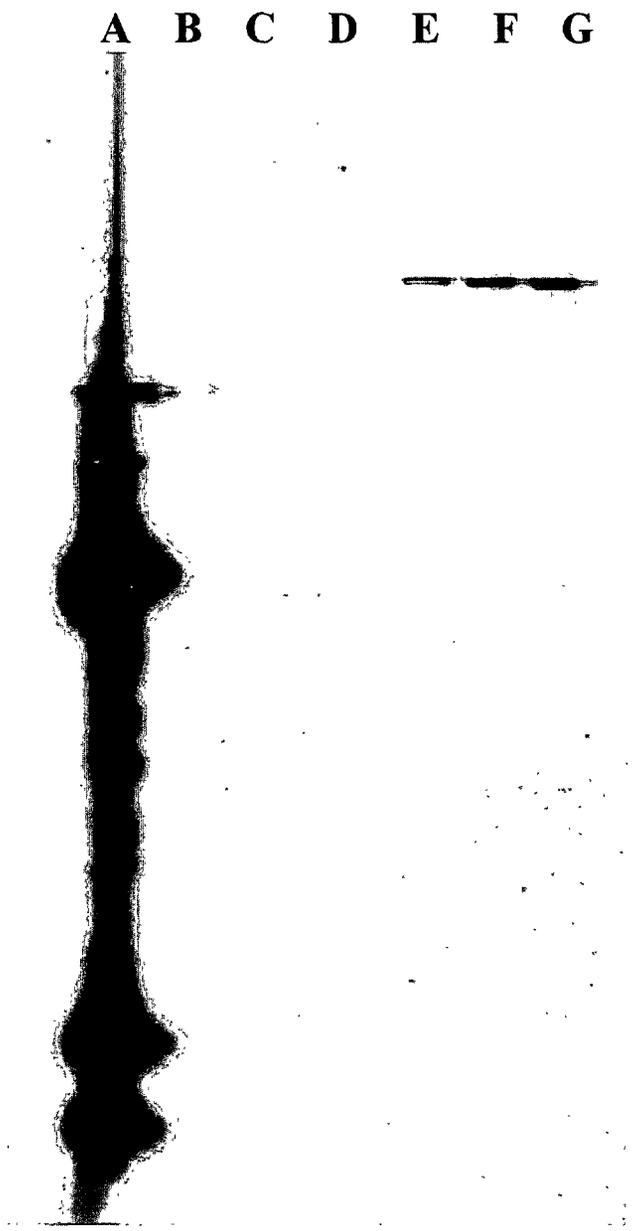


Figure 13 : Absence of the *Amp*R gene and of plasmid sequences in the ML01 strain.

DNA was digested and separated onto an agarose gel. Lane A represents a DNA Ladder, lane B genomic DNA of the host strain S92, lanes C and D genomic DNA of the ML01 strain (two different active dry yeast pilot preparations), lanes E to G mixtures of genomic DNA of the strain ML01 with increasing concentrations of pUT332. Probe labeling, hybridization and detection were performed using the ECL direct nucleic acid labeling hybridization and detection kit (Amersham Pharmacia). A description of the probe used is given Figure 11. Our assay conditions represent low stringency hybridization.

Plasmid pJH2

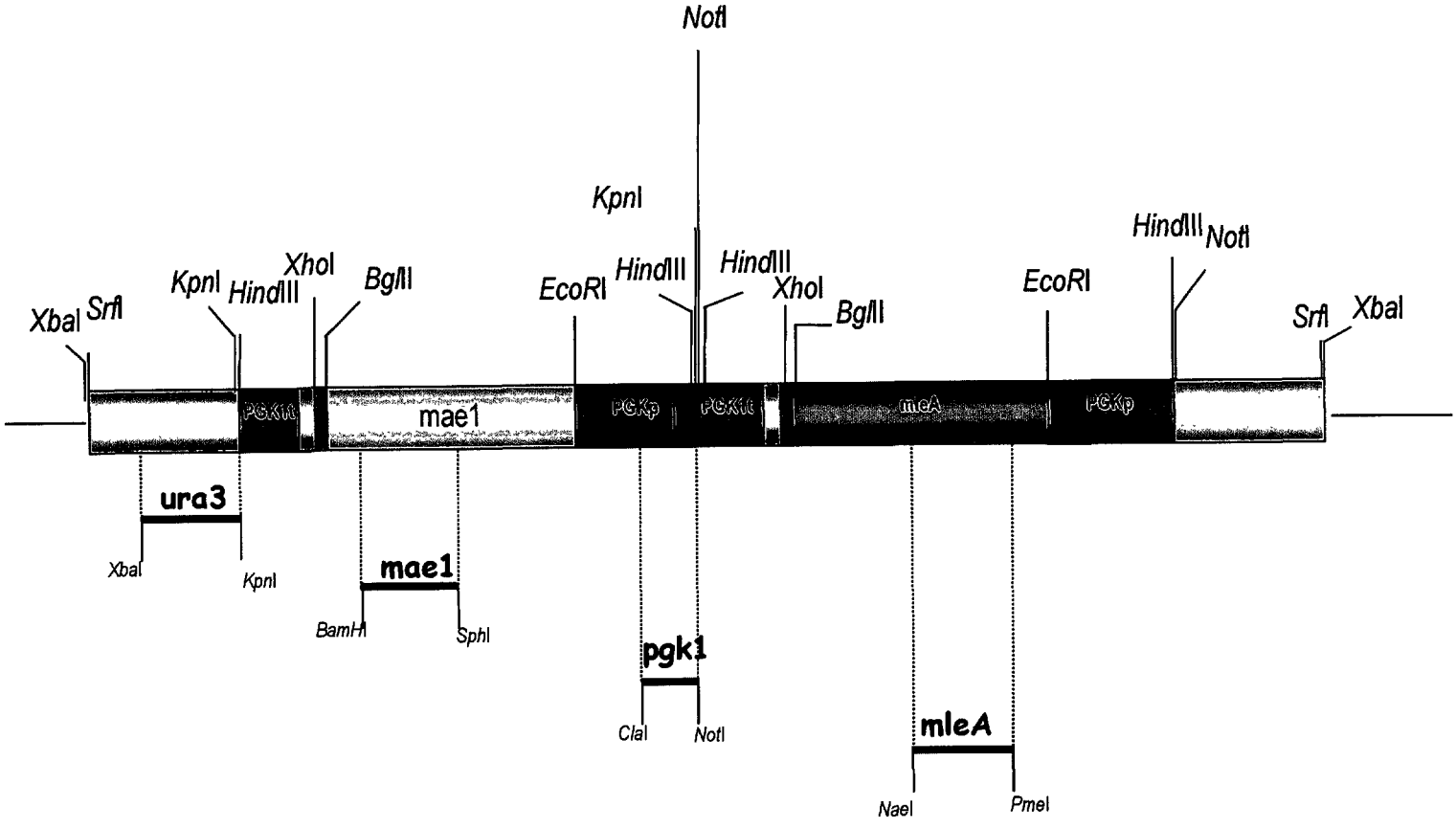


Figure 14: Schematic representation of probes originating from plasmid pJH2 and used in hybridization experiments.

All probes templates were isolated from plasmid **pJH2** by digestion with appropriate restriction enzymes. Template of the ura3 probe was retrieved as a *Xba*I-*Kpn*I fragment, template of the mae1 probe as a *BamH*I-*Sph*I fragment, template of the pgk1 probe as a *Cla*I-*Not*I fragment, and template of the mleA probe as a *Nae*I-*Pme*I fragment.

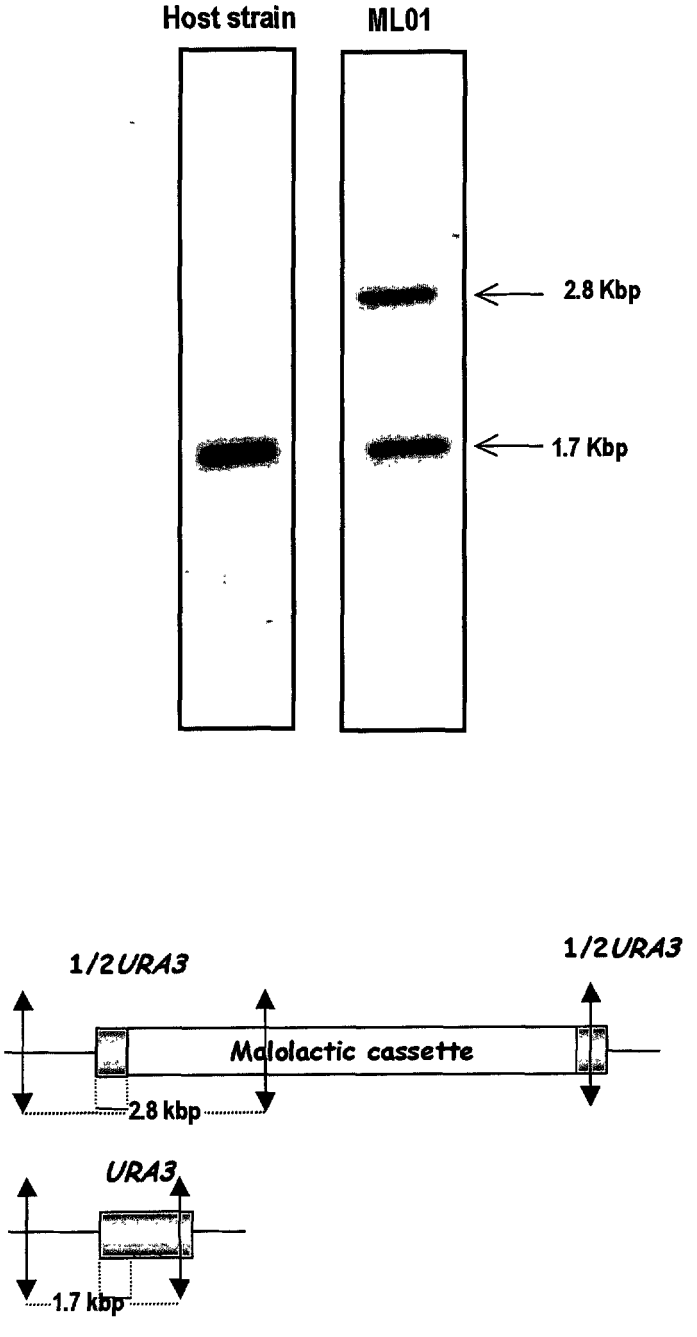


Figure 15 : Southern blots representing hybridization of genomic DNA with the ura3 probe and schematic representation of an integrated and non integrated *URA3* loci.

Genomic DNA from either the host strain or the ML01 strain was totally digested with restriction enzyme *Nsi*I. *Nsi*I restriction sites are represented by double arrows. Ura3 probe hybridization sites are schematized as full red lines and restriction fragments detected by the probe are represented as dotted lines. Probe labeling, hybridization and detection were done by the ECL direct nucleic acid labeling and detection system (Amersham Pharmacia).

000064

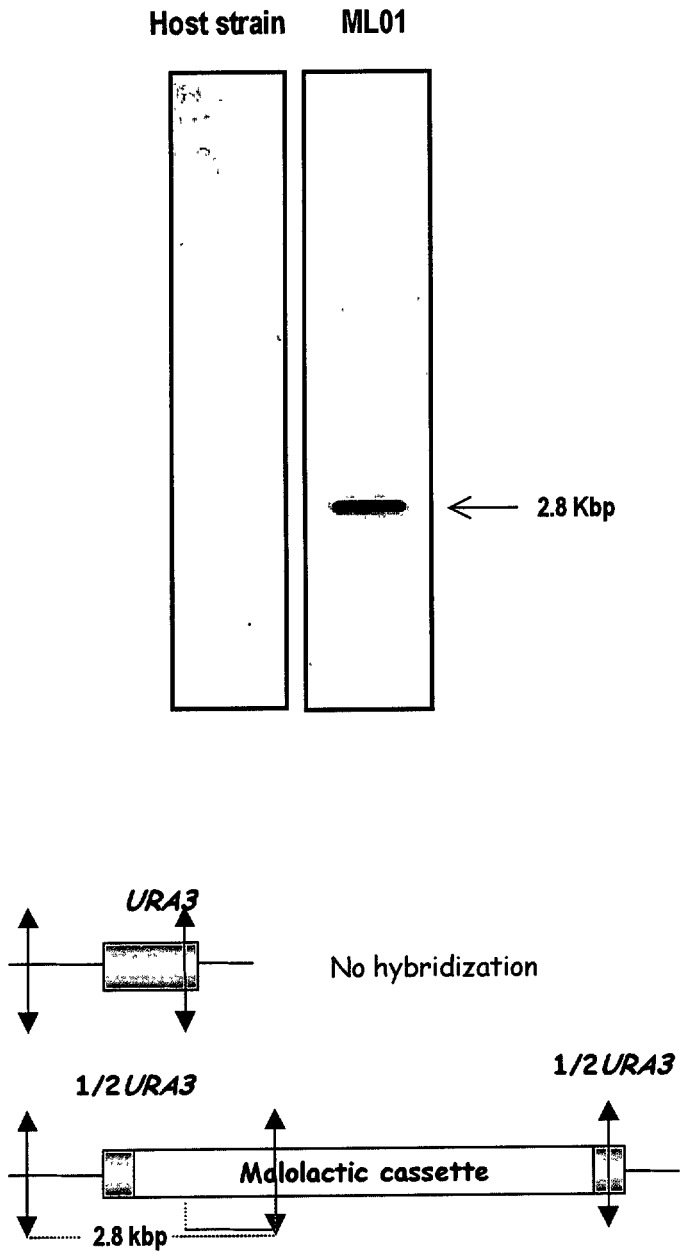


Figure 16: Southern blots representing hybridization of genomic DNA with the mae1 probe and schematic representation of an integrated and non integrated URA3 loci.

Genomic DNA from either the host strain or the ML01 strain was totally digested with restriction enzyme *Nsi*I. *Nsi*I restriction sites are represented by double arrows.
Mae1 probe hybridization sites are schematized as full red lines and restriction fragments detected by the probe are represented as dotted lines. Probe labeling, hybridization and detection were done by the ECL direct nucleic acid labeling and detection system (Amersham Pharmacia).

000065

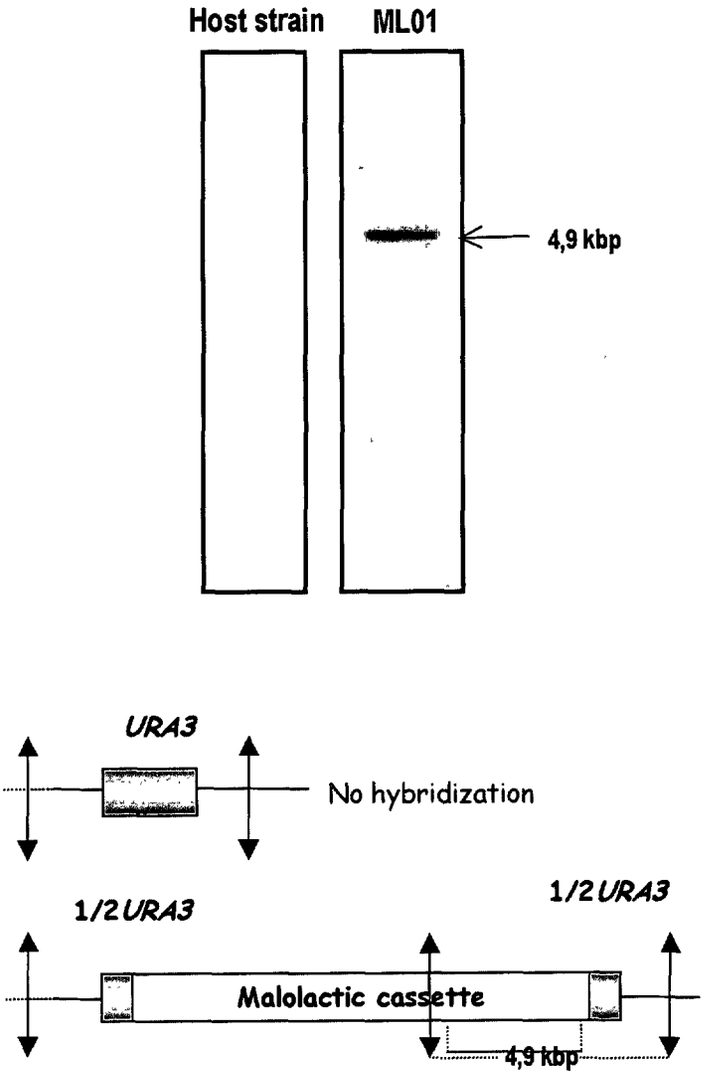


Figure 17: Southern blots representing hybridization of genomic DNA with the mleA probe and schematic representation of an integrated and non integrated URA3 loci.

Genomic DNA from either the host strain or the ML01 strain was totally digested with restriction enzyme *Pvu*II. *Pvu*II restriction sites are represented by double arrows.
MleA probe hybridization sites are schematized as full red lines and restriction fragments detected by the probe are represented as dotted lines. Probe labeling, hybridization and detection were done by the ECL direct nucleic acid labeling and detection system (Amersham Pharmacia).

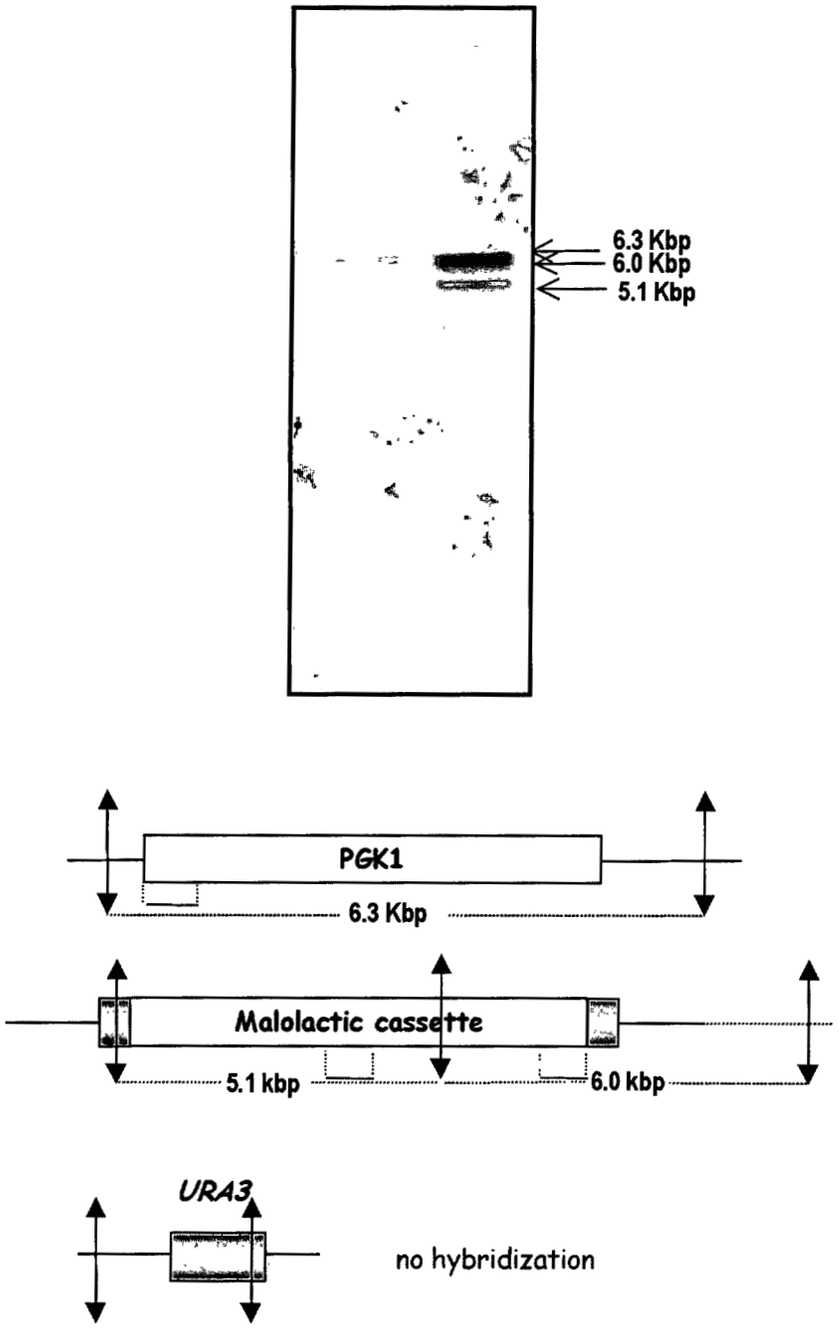


Figure 18: Southern blot representing hybridization of genomic DNA with the pgk1 probe and schematic representation of a *PGK1* locus, an integrated and non integrated *URA3* loci.

Genomic DNA from either the host strain or the ML01 strain was totally digested with restriction enzyme *Eco*RV. *Eco*RV restriction sites are represented by double arrows.
Pgk1 probe hybridization sites are schematized as full red lines and restriction fragments detected by the probe are represented as dotted lines. Probe labeling, hybridization and detection were done by the ECL direct nucleic acid labeling and detection system (Amersham Pharmacia).

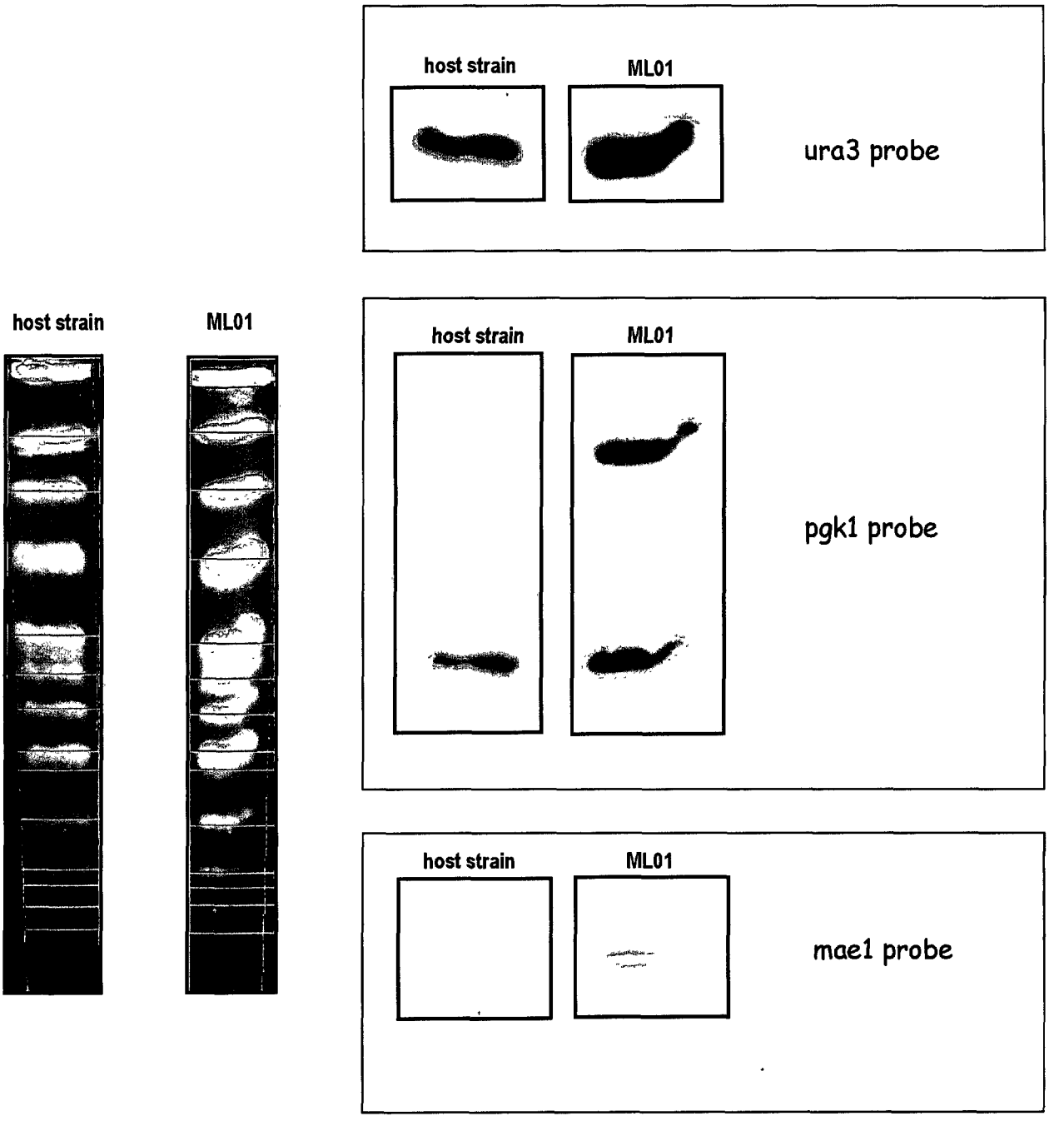


Figure 19: Pulsed field gel electrophoresis of genomic DNA of the host strain and the ML01 strain and subsequent hybridization with ura3, pgk1 and mae1 probes.

The same membrane was used in all hybridization experiments by stripping and re-probing.
Probe labeling, hybridization and detection were done by the ECL direct nucleic acid labeling and detection system (Amersham Pharmacia).

000068

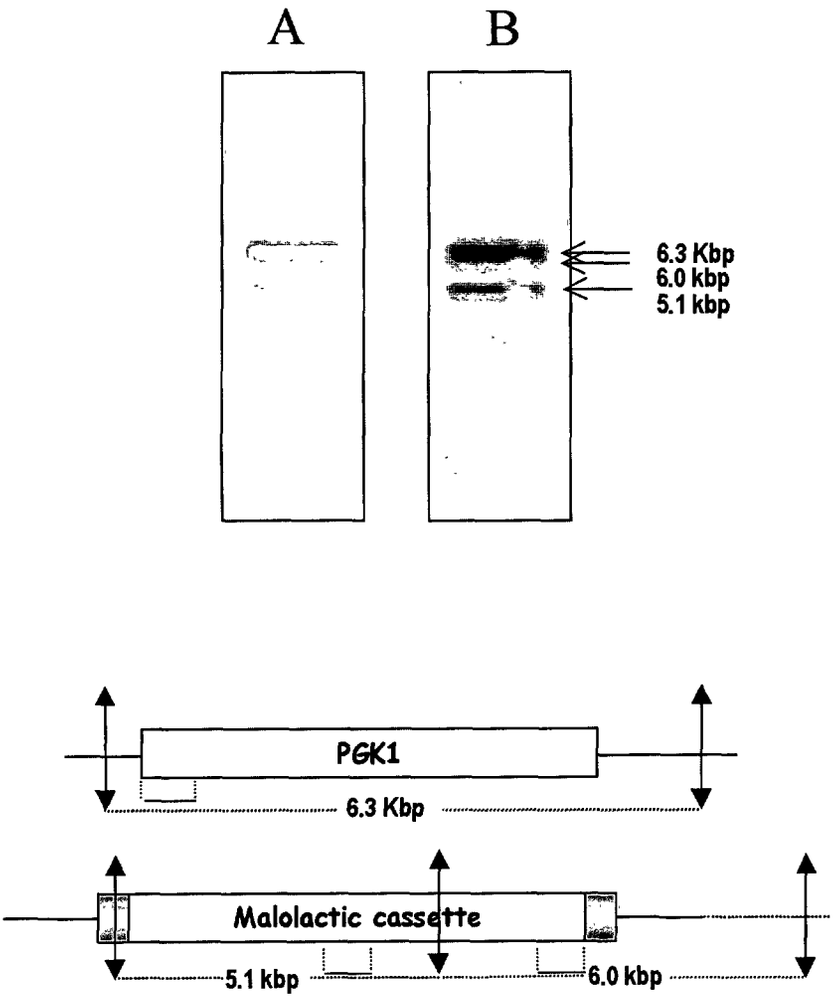


Figure 20: Stability of the integrated genetic material in the ML01 strain.

Southern blots of genomic DNA digested with *Eco*RI and hybridized with the pgk1 probe. Lane A : ML01 strain; Lane B : ML01 strain after more than 100 generations.
Probe labeling, hybridization and detection were done by the ECL direct nucleic acid labeling and detection system (Amersham Pharmacia).

000069

Table 4 : Discrepancies found when comparing the genomic sequence of the integrated locus and previously published sequences

Comments and nucleotide positions reefer 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	929-934	Additional sequence resulting from cloning strategy
3	1199-1217	Additional sequence resulting from cloning strategy
4	2819-2829	Additional sequence resulting from cloning strategy
5	2896	Difference in the *PGK1*p sequence
6	4298	Difference in the *PGK1*p sequence
7	4317-4330	Additional sequence resulting from cloning strategy
8	4595-4613	Additional sequence resulting from cloning strategy
9	4629	Difference in the *mleA* open reading frame. This difference corresponds to a change of amino acids from aspartic acid (in the published sequence) to glutamic acid (in the genomic sequence)
10	5247	Difference in the *mleA* open reading frame. This difference corresponds to **no change** of the amino acid sequence
11	6243-6253	Additional sequence resulting from cloning strategy
12	7741-7748	Additional sequence resulting from cloning strategy
13	7751	Difference in the 3' region of the *URA3* open reading frame
14	8234	Difference in the 3' region of the *URA3* non coding sequence
15	8543	Difference in the 3' region of the *URA3* non coding sequence

000070

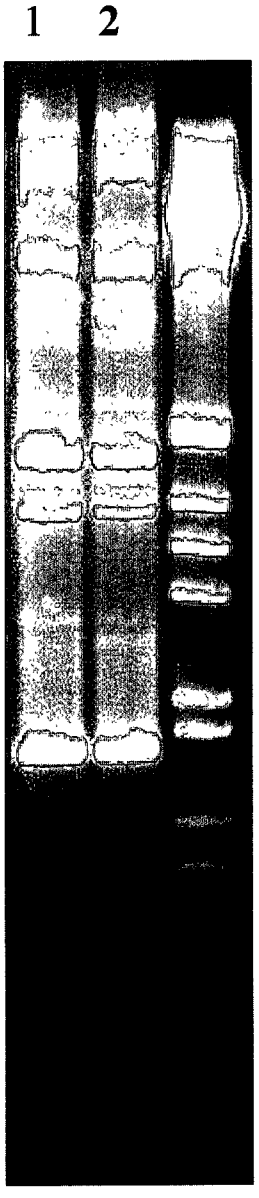


Figure 21 : Genetic pattern of the ML01 (1) and the S92 (2) yeast strains based on amplification of genomic DNA regions in between δ elements of the Ty1 retrotransposon.

Lane 3 corresponds to molecular weight markers. A detailed method of the PCR amplification procedure is given by Ness *et al.*, 1993.

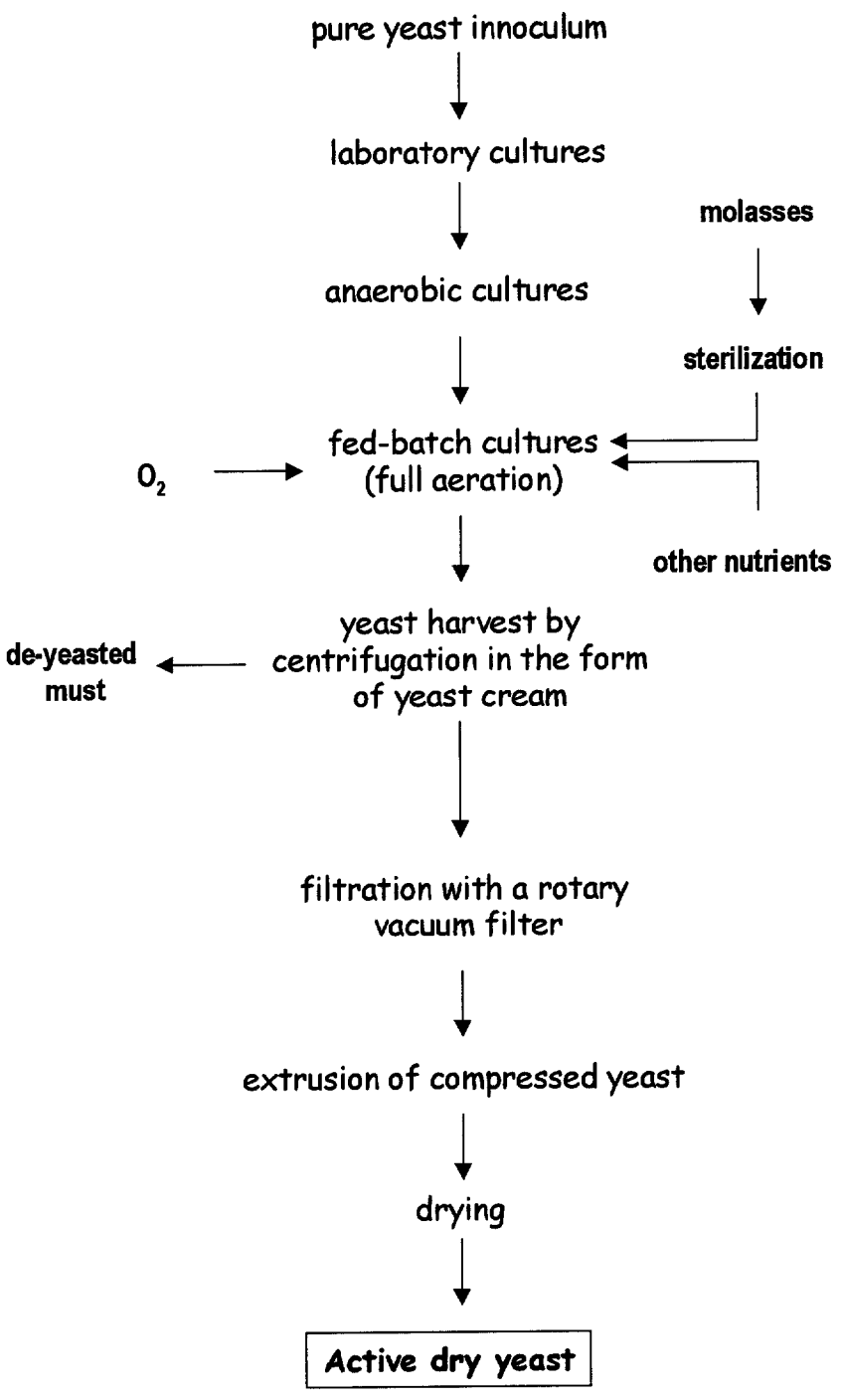


Figure 22: Flow diagram describing the manufacture of active dry yeast.

A

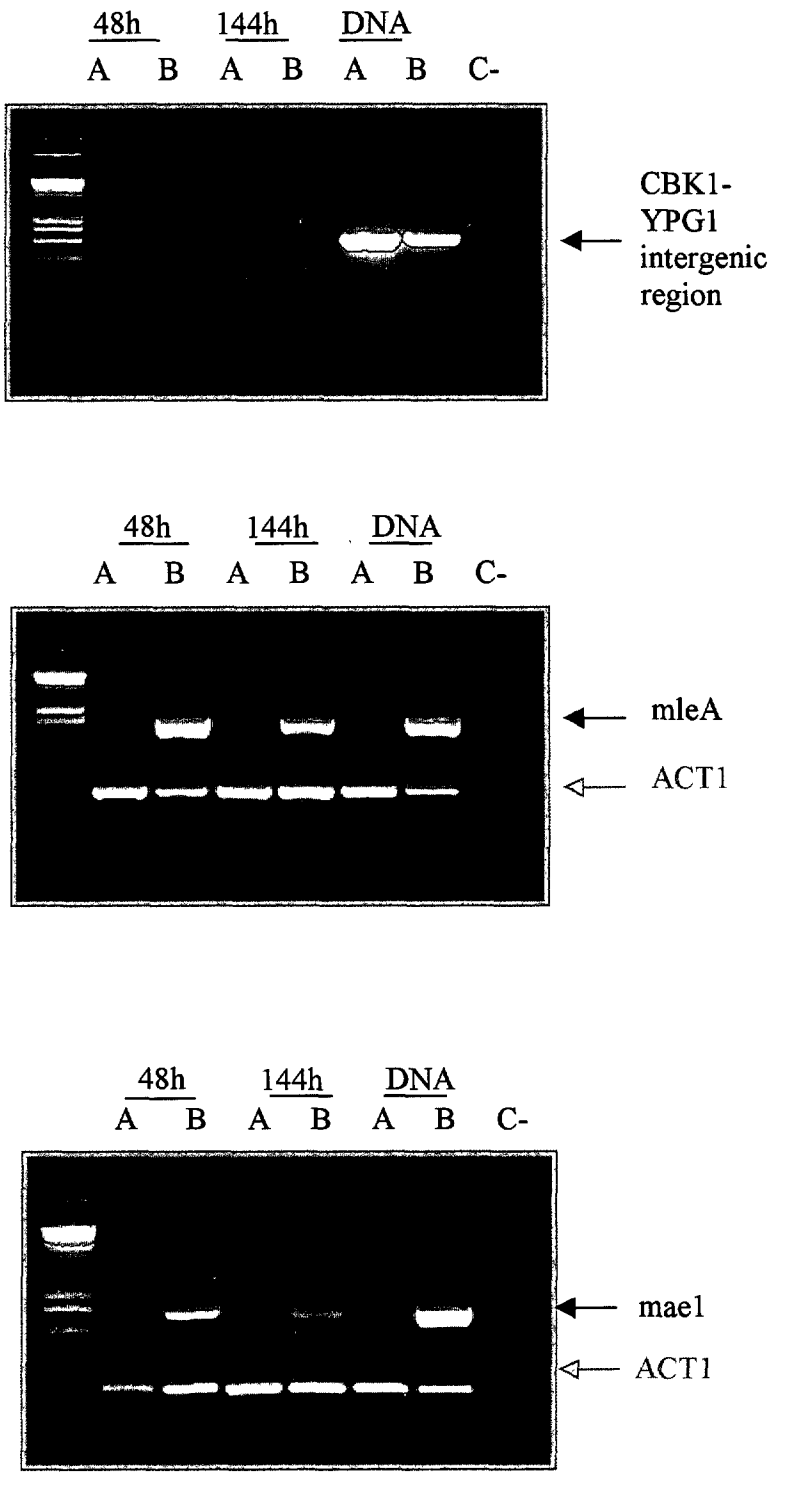


Figure 23 : *mleA* and *mae1* gene expression during synthetic grape must fermentation

A- *mleA* and *mae1* genes expression of the ML01 strain (B) or the S92 strain (A) during the fermentation of synthetic grape juice at two time points , 48 hours (48h) and 144 hours (144h) after the beginning of fermentation were analyzed by RT-PCR. *ACT1* gene expression was used as an internal control. A control was carried out by using genomic DNA (DNA) and a negative control was carried out by adding no template to the reaction mixture (C-).

B

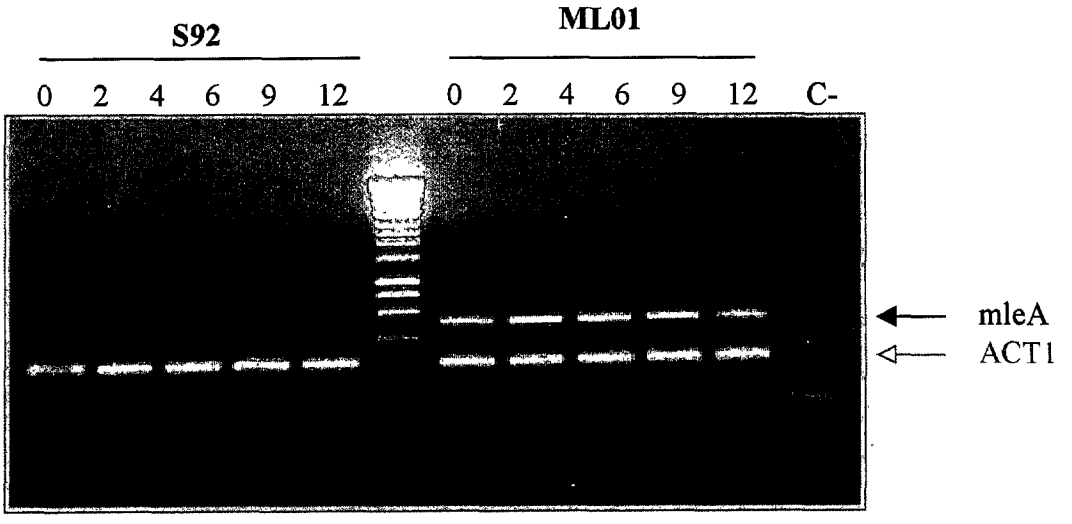
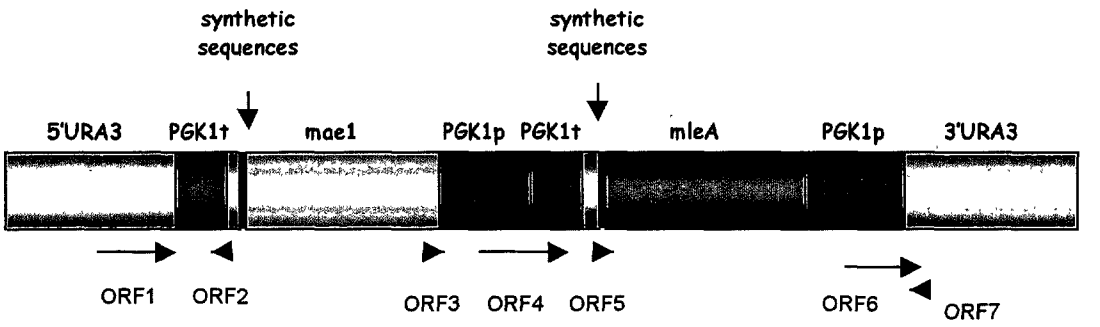


Figure 23 (continued) : *mleA* and *mae1* gene expression during synthetic grape must fermentation.

B- *mleA* gene expression analyzed by RT-PCR during synthetic grape must fermentation at 5 time points after the beginning of fermentation : 0, 2, 4, 6, 9 and 12 days. A negative control was carried out by adding no template to the amplification reaction (C-). *ACT1* gene expression was used as an internal standard.

000074

A



ORF1	ORF2	ORF3	ORF4	ORF5	ORF6	ORF7	
432	120	135	660	183	633	108	**ORF size (bp)**
509	1269	2675	3703	4560	7127	7840	**Start position***
941	1149	2810	4363	4743	7760	7732	**Stop position***

* these positions reefer to those described in Table 3.

Figure 24 : Putative open reading frames (ORFs) analysis.

A- schematic representation of putative open reading frames (ORF) of more than 90 bp and presenting a TATA box found within the integration cassette.

B

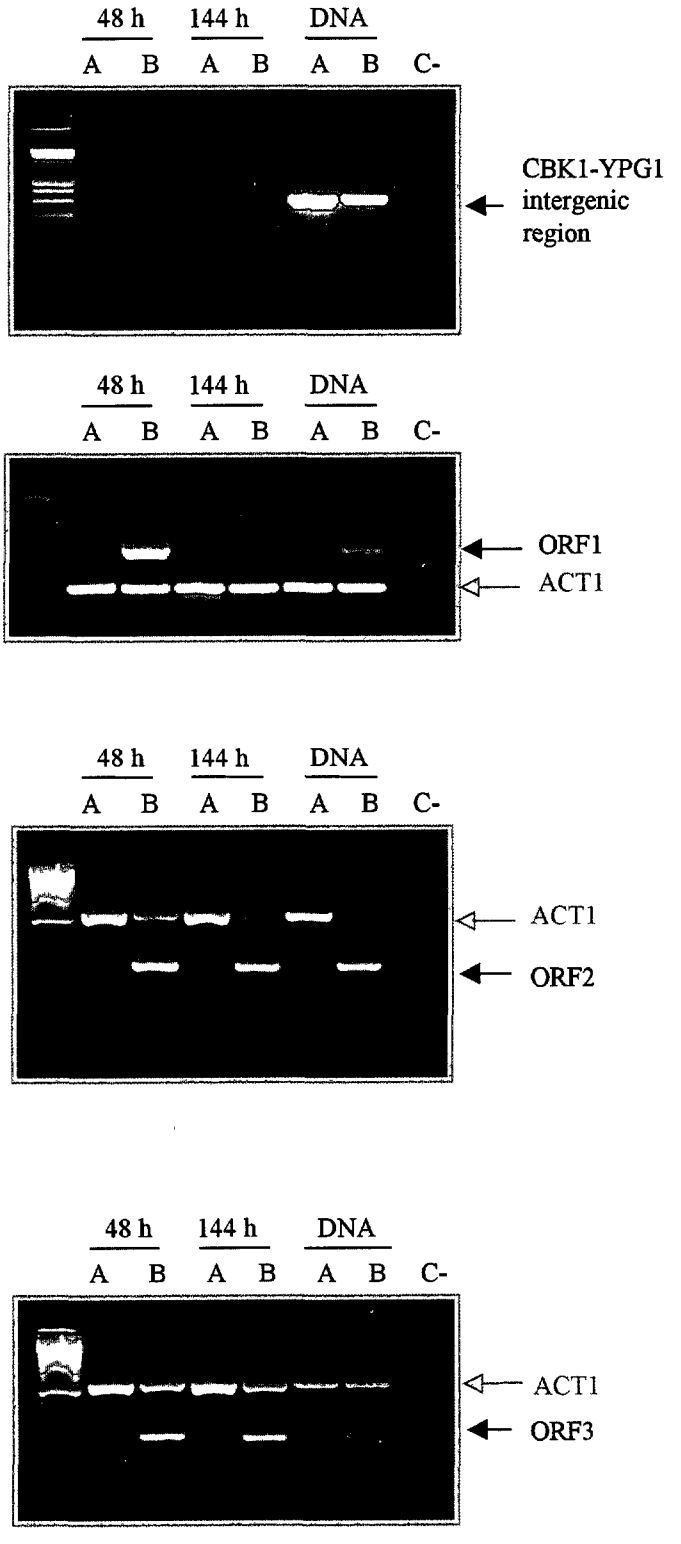


Figure 24 (continued) : Putative open reading frames (ORFs) analysis
B- Putative ORFs expression analysis by RT-PCR.
Oligonucleotides targeted to an intergenic region (CBK1-YGP1) were used to verify that RNA samples were not contaminated with genomic DNA. Controls (DNA) were performed for all amplifications by using genomic DNA from either S92 (A) or ML01 (B). A negative control (C-) was carried out by adding no template to the amplification mixture. RNA samples from yeast cells of S92 (A) or ML01 (B) harvested 48 hours (48h) or 144hours (144h) after the beginning of the fermentation of synthetic grape must, were used as templates for each RT-PCR reaction. Amplification of the *ACT1* transcript was used as a reaction control.

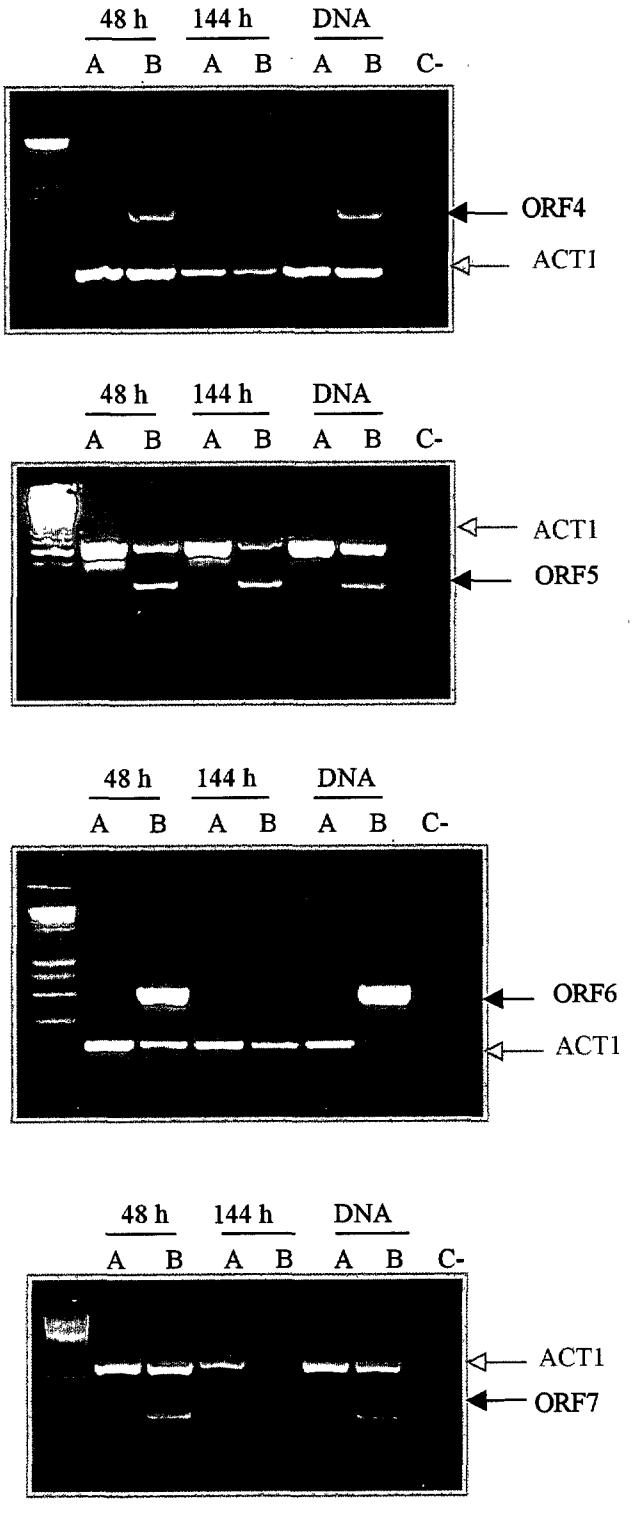
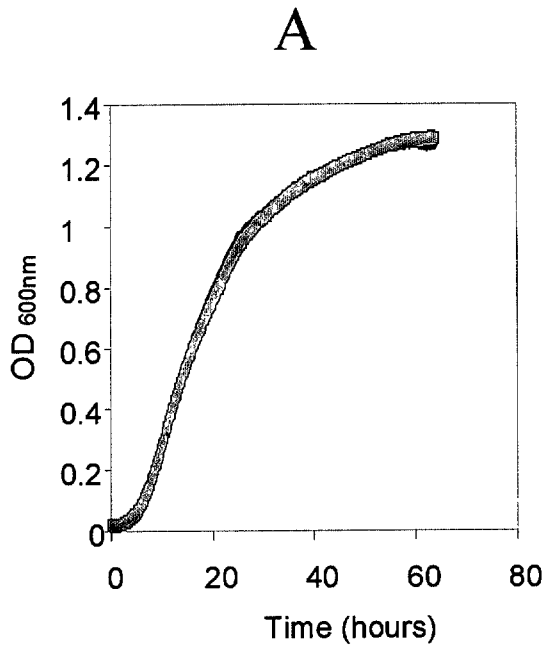


Figure 24 (continued) : Putative open reading frames (ORFs) analysis
B- Putative ORFs expression analysis by RT-PCR.
Oligonucleotides targeted to an intergenic region (CBK1-YGP1) were used to verify that RNA samples were not contaminated with genomic DNA. Controls (DNA) were performed for all amplifications by using genomic DNA from either S92 (A) or ML01 (B). A negative control (C-) was carried out by adding no template to the amplification mixture. RNA samples from yeast cells of S92 (A) or ML01 (B) harvested 48 hours (48h) or 144hours (144h) after the beginning of the fermentation of synthetic grape must, were used as templates for each RT-PCR reaction. Amplification of the *ACT1* transcript was used as a reaction control.



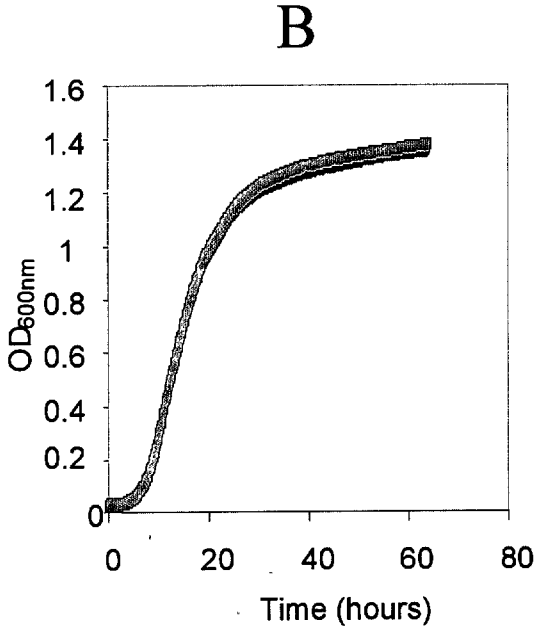


Figure 25: Growth comparison of strains ML01 (blue) and S92 (pink) in synthetic and Chardonnay must.

Panel A: Synthetic grape must ; panel B : Chardonnay must containing 4.5 grams/liter L-malate.

000078

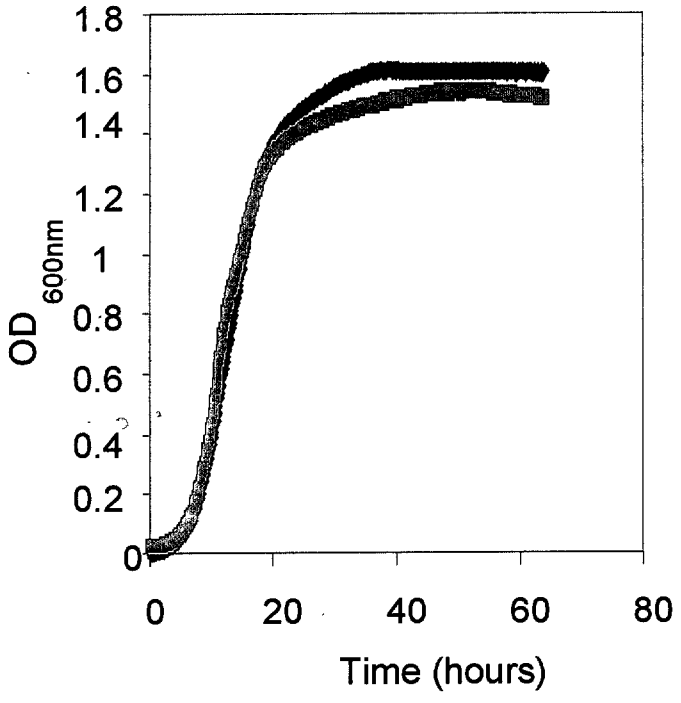
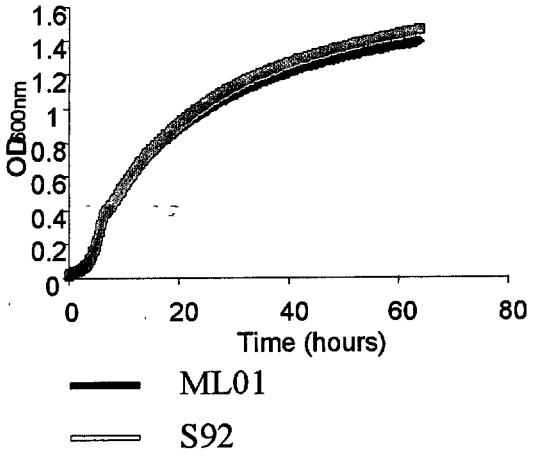


Figure 26: Growth comparison of strains ML01 (blue) and S92 (pink) in high acid (pH 3.18) and high malate (9 grams/liter L-malate) Chardonnay must.

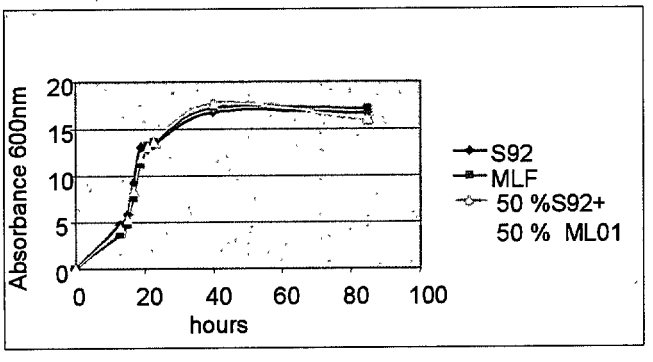
I-Measurements using an automated tubidometric reader

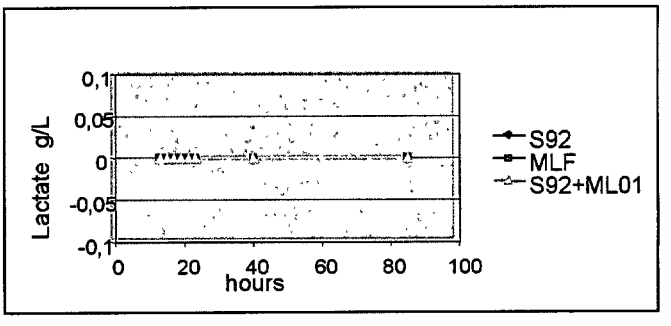


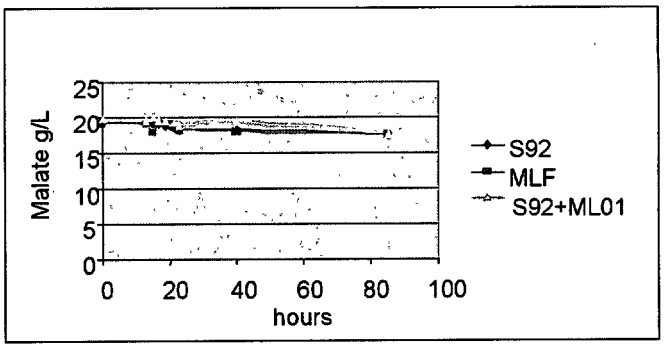
	ML01	S92
L-malate gram/liter	19.3	18.99
L-lactate gram/liter	0.036	0.002

Malate and lactate measurements after 60 hours culture

II- Larger scale cultures







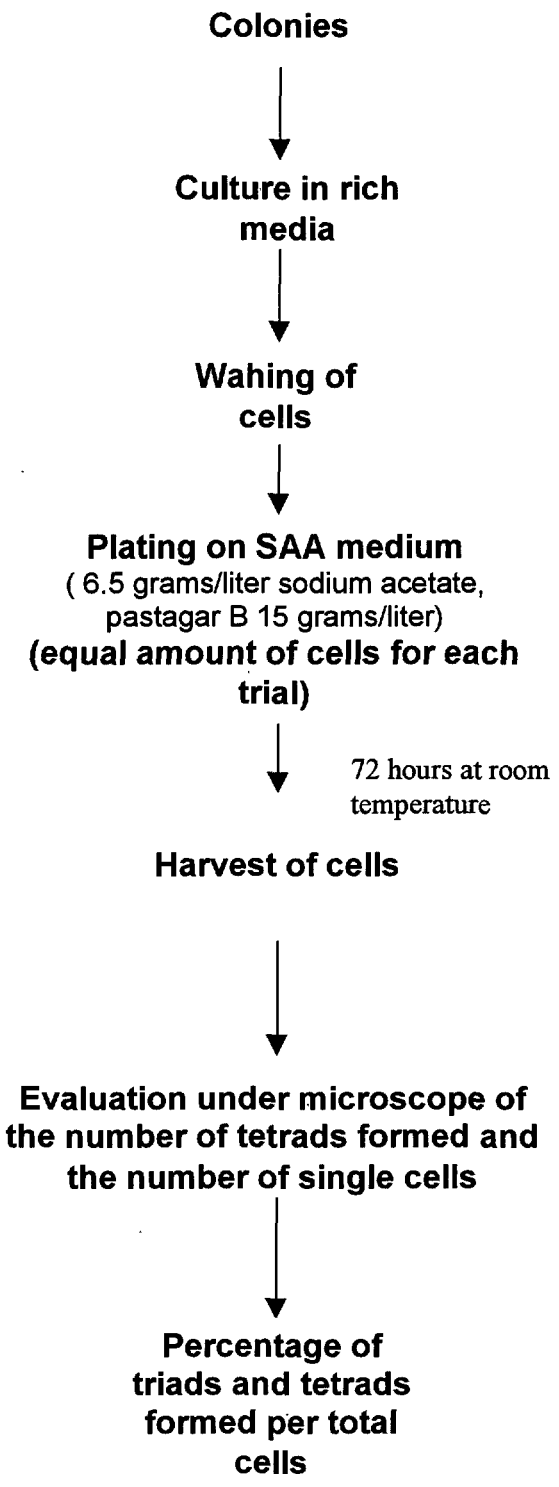
Hours	0	15	23
% S92	48 %	65 %	53 %
% ML01	52 %	35 %	47 %

Percent of S92 and ML01 cells during the fermentation

Figure 27: Growth comparison of strain ML01 and strain S92 in synthetic medium containing 20 grams/liter L-malate and 5 grams/liter glucose.

The culture media was prepared by mixing 10 grams/liter Yeast Extract (Difco), 20 grams/liter peptone (Difco), 5 grams/liter glucose and 20 grams/liter L-malate. The pH of the solution was adjusted to 6.5 with KOH and filter sterilized.

000080



	ML01	S92
Total counted	782	612
Percentage of ascospores (containing 3 or 4 spores)	11 %	8%

Figure 28 : Ascospore formation by the ML01 and the S92 strains

000081

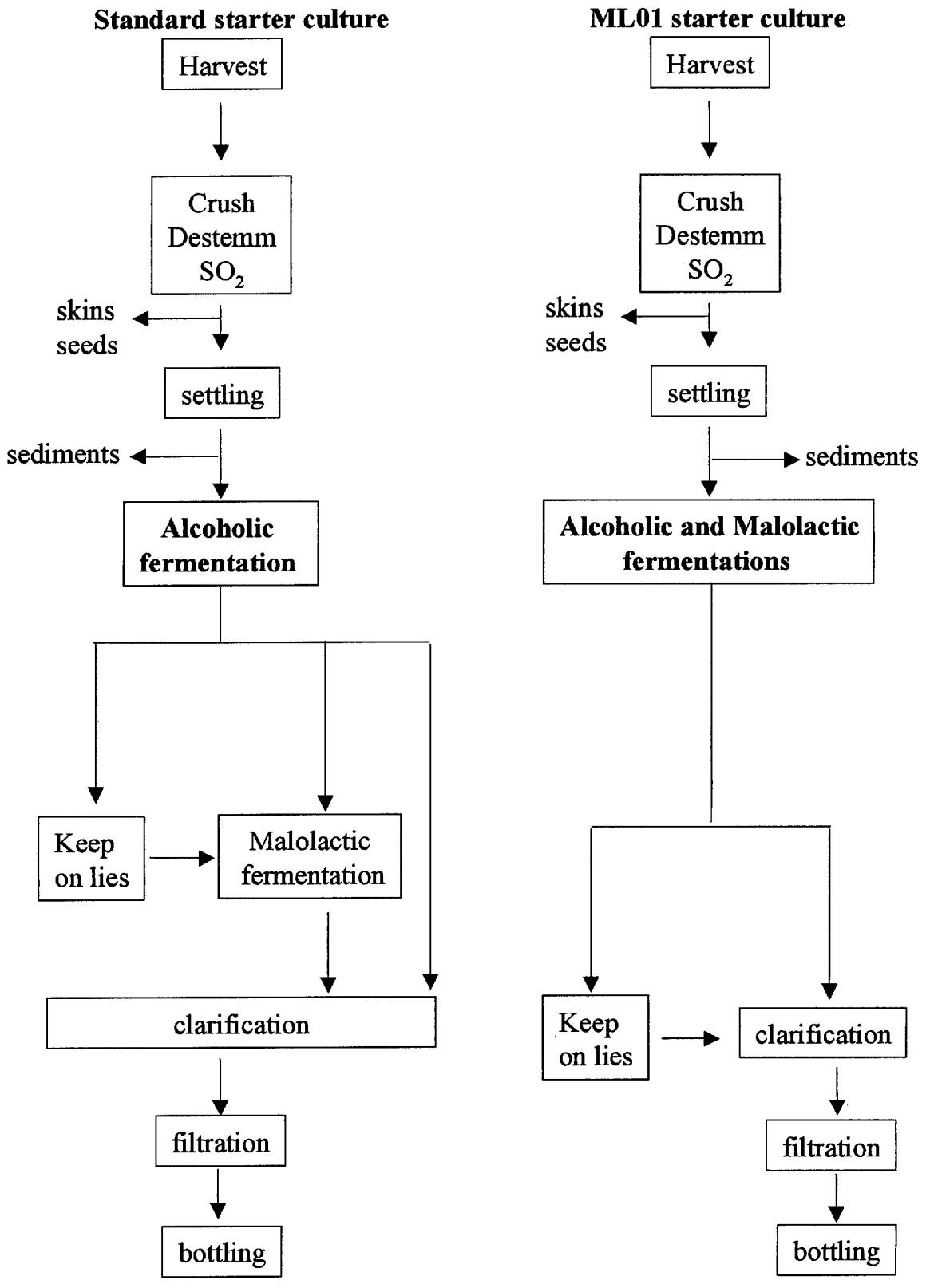


Figure 29 : White winemaking practices and the use of a standard or ML01 yeast strain.

000082

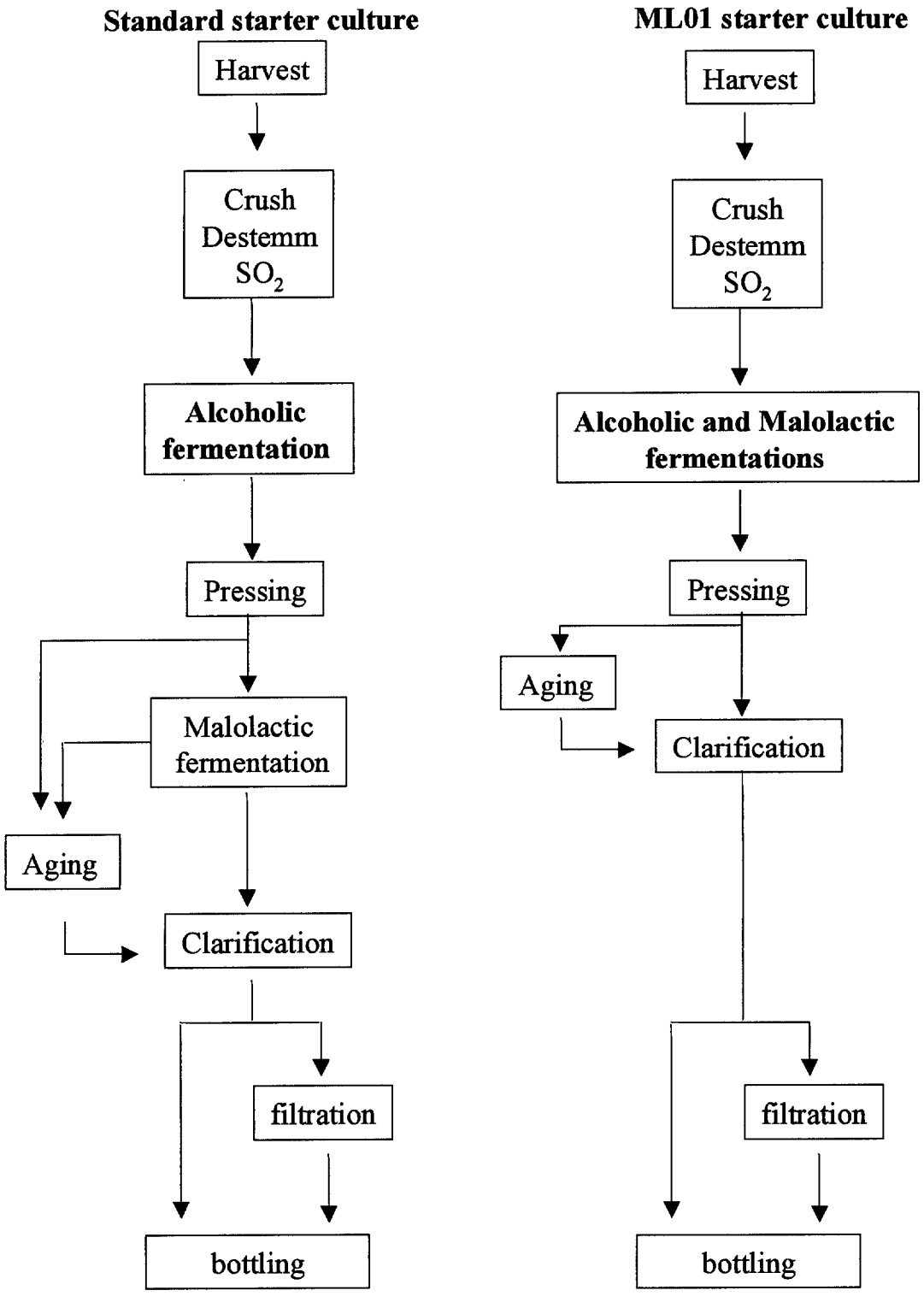


Figure 30: Red winemaking practices and the use of a standard or ML01 yeast strain

000083

Table 5 : Potential ethyl carbamate in Merlot wines

	Ethyl carbamate (ppb)	Standard deviation (ppb)
ML01	12.0	0
S92	13.0	1
S92 + Oenococcus œni	20.3	0.6

Ethyl carbamate was measured after heating the samples 48 hours at 72 °C, thus accelerating the ethyl carbamate formation. This represents potential ethyl carbamate which could be formed ultimately in the wine. The wine with which these dosages were done is a pilot-scale assay of Merlot wine carried out at the Pacific Agri Research Centre (PARC) in the Okanagan Valley, under the supervision of the University of British Columbia. Ethyl carbamate measurements were done in triplicate by the Liquor Control Board of Ontario, by gas chromatography coupled to mass spectrometry.

Appendices

Appendix 1 : Integration cassette in Fasta format

000086

>readseq.input(1), 8683 bases, 3976C049 checksum.
gggcaacggttcatcatctcatggatctgcacatgaacaaacaccagagt
caaacgacgttgaaattgaggctactgcgccaattgatgacaatacagac
gatgataacaaaccgaagttatctgatgtagaaaaggattaaagatgcta
agagatagtgatgatatttcataaataatgtaattctatatatgttaatt
acctttttgcgaggcatatttatggtgaaggataagttttgaccatcaa
agaaggttaatgtggctgtggtttcagggtccataaagcttttcaattca
tctttttttttttgttctttttttgattccggtttctttgaaattttt
ttgattcggtaatctccgagcagaaggaagaacgaaggaaggagcacaga
cttagattggtatatatacgcatatgtggtgttgaagaaacatgaaattg
cccagtattcttaacccaactgcacagaacaaaaacctgcaggaaacgaa
gataaatcatgtcgaaagctacatataaggaacgtgctgctactcatcct
agtcctgttgctgccaagctatttaatatcatgcacgaaaagcaaacaaa
cttgtgtgcttcattggatgttcgtaccaccaaggaattactggagttag
ttgaagcattaggtcccaaaatttgtttactaaaaacacatgtggatatc
ttgactgatttttccatggagggcacagttaagccgctaaaggcattatc
cgccaagtacaatttttactcttcgaagacagaaaatttgctgacattg
gtaatacagtcaaattgcagcactctgcgggtgtatacagaatagcagaa
tgggcagacattacgaatgcacacggtgtggtgggcccaggtattgttag
cggtttgaagcaggcggcggaagaagtaggtaccaagctttaacgaacgc
agaattttcgagttattaaacttaaaatacgctgaacccgaacatagaaa
tatcgaatgggaaaaaaaaactgcataaaggcattaaaagaggagcgaat
tttttttaataaaaatcttaataatcattaaaagataaataatagtcta
tatatacgtatataaataaaaaatattcaaaaaataaaataaactattat
tttagcgtaaaggatggggaaagagaaaagaaaaaaattgatctatcgca
gatccctcgaggagatctcatatgcaagacgcatacatatttaacactac
gacaaccaaaaattaccattattcaaagtccaaaattttggtatgtataa
tttaaacaaaaagaagaagaaggagcaaaaatgtctaagcaaaattagta
ccagtagtacaataaagatacagaatgattcaaaatgattaatctataat
ggtattaagagcagggcacacagatttatagaaaaattaaggaaaagcat
acaagcttaaacgctttcatgttcactactaggaggatccgattcaccac
cagtagatgtgacatgtgtatcgactttttccaaagatgcaggtgctttt
tcaggtgggaaggtagggttaaggacacctgtatttggttttggtggagg
atgggcatcttcgtctttgccaggatagcaaagatcattgacgagaaacg
cacggaccattaaatacattaggaggatccactgaatacaaagaatgacc
ccaatgatatgtccaaacatttggaaagctttggaatctatcattttacc
tatctcaatggtacaattaacaaaacccacgttggggaaaatgaatgcaa
accatccacaagcaaacttgagaggggctcgagtgaaaaagcccgctaaa
aagctaaccatggcgagacagtaacaccaagcagcaagaccccaaataaa
aatagccataaaggtagaaacaaaaccaagatactcggatgagttggcgc
caacaaaaatataagggcgactgcccatagcaccacgcgcaatattaatt
aaggccaaacctgagaaagctggtggaccgacaaacataaacataccagg
tcgatcttggggttttgccaggcctacagtaaaaaaccgtaagacattga
cggcaaacagtaaaagataaacccaaaaaccaagtccttgaaagaggata
ccaaagataaccatatttttttaattgatgagcgggttgtgtagaattgac
ggcgccagcaatgacaccacaaatcataggagggaaaataggaagaatcc
aagcaggagatgcggtttcaatggtatatacatggttgttgaaaattgta
aaaaaagccattacgcagtatataaaggatactgcaacgtaaatgtaata
aaggattcgaatgacccacaccatccactcgccggtatcaggataggcgt
atatggcaagcatgtcgatgaacgtggatattgaaagaagacaagtagca
atgaaaagcttttccaaatgatggttccaggaatccttgatagttgaagg
atatttaataaagcgaaaaagcatgcatgatccaaagagagaaaacaaaa
agatttgaagaatataaacaattttgccaattgtattaagaccataaaat
cgaaaggggaaagaaccaataatcaaaccaacaccaccagttgccatagt
acatgcaaaccaagaccatgtaaaatgcttcagtcgttgactgagaggga
catgaggggctttgacattccagtcaagcaactcatgatacctctgtttc
aagatttccttgagttcacccatgtcgaatcgagagataaaggggaatta
aaaaaaaaatagtggccagaattccgatcttgttttatatttgttgtaaa
aagtagataattacttccttgatgatctgtaaaaaagagaaaaaggaagc
atctaagaacttgaaaaactacgaattagaaaagaccaaatatgtatttc
ttgcattgaccaatttatgcaagtttatatatatgtaaatgtaagtttca
cgaggttctactaaactaaaccaccccccttggttagaagaaaagagtgtg
tgagaacaggctgttgttgtcacacgattcggacaattctgtttgaaaga
gagagagtaacagtacgatcgaacgaactttgctctggagatcacagtgg
gcatcatagcatgtggtactaaacccttcccgccattccagaaccttcg
attgcttgttacaaaacctgtgagccgtcgctaggaccttgttgtgtgac
gaaattggaagctgcaatcaataggaagacaggaagtcgagcgtgtctgg
gtttttcagttttgttctttttgcaaacaaatcacgagcgacggtaatt
tctttctcgataagaggccacgtgctttatgagggtaacatcaattcaag
aaggagggaaacacttccttttctggccctgataatagtatgagggtga
agccaaaataaaggattcgcgcccaaatcggcatctttaaatgcaggtat
gcgatagttcctcactctttccttactcacgagtaattcttgcaaatgcc

000087

tattatgcagatgttataatatctgtgcgtcttgagttgaagtcaggaat
ctaaaataaaaattaaggttaataaaaagaggaaagaaaaaaaaattaat
cgatttacagaaacttgcacactaaaaatacacaactaaaagcaattaca
gtatgggaagtcatcgacgttatctctactatagtatattatcatttcta
ttattatcctgctcagtggtacttgcaaaacaagataagaccccattctt
tgaaggtacttcttcgaaaaattcgcgtctaactgcacaagataagggca
atgatacatgcccgccatgttttaattgtatgctacctatttttgaatgc
aaacagttttctgaatgcaattcgtacactggtagatgtgagtgtataga
agggtttgcaggtgatgattgctctctgcccctctgtggcggtctatcac
cggatgaaagcggtaataaggatcgtcccataagagcacaaaatgacacc
tgtcattgtgataacggatggggagggatcaattgtgacgtttgtcaaga
agattttgtctgtgatgcgttcatgcctgatcctagtattaaggggacat
gttataagaatggtatgattgtagataaagtattttcaggttgtaatgtg
accaatgagaaaattctacagattttgaacggcaaaataccacaaattac
atttgcctgtgataaacctaatcaagaatgtaattttcagttttggacag
atcagttagaaagcttggtaccgcggccgcaagctttaacgaacgcagaa
ttttcgagttattaaacttaaaatacgctgaacccgaacatagaaatatc
gaatgggaaaaaaaaactgcataaaggcattaaaagaggagcgaattttt
ttttaataaaaatcttaataatcattaaaagataaataatagtctatata
tacgtatataaataaaaaatattcaaaaaataaaataaactattatttta
gcgtaaaggatggggaaagagaaaagaaaaaaattgatctatcgcagatc
cctcgaggagatctcattagtatttcggctcccacttcaaatcgtcgacg
gccttttcaacatcatcaatcggctggcgattaagaccttgttcaacagc
ttttttagcgacagcgacagcgactgtacgcgaaaagtcggctaattttg
aaactggtggcaaaacagcagcaccaacttttgttgtatcaacgattcct
ccaagactatgggcagcagcagaaatcatttcaggcgtaagcagctttga
ttgagcggcaatggcaccaaagccaagacctggatagatcaaggcattgt
tggcctgaccgatttggtaagcgttgccctcatattcaatatcgtcaact
ggaacaccagtaccgatcaaggcttttccattagaccatttcaaaacatc
ttcggcttttgcttcggctaattttgttggattggaaattggaaaaatga
ttggtctttcagtataaccagacatatctttaacaatttcttcagtaaag
gaatttggatgtgtcgaggttccgaccaaaatggtcgggtggacagcttc
aacagctgcttggagattggtcaattgattagcatttttgaaatcacttc
gtttagcagcgaaaggcttttgctctggagttaaatccggatcatcatca
aacaagaggccttgtttgtcaacaaggaagaaatgctttttagcctcttc
gtcggaaagaccctgttcaaccatttcttcatgcaactgtttaacaattc
ccattccagcagttccggcaccgaagctcatgtaagtttgatcagttaat
ttctgaccggaaatcttcaacgctccaagaacgccggcaagaacgacgat
tccagttccttgaatatcatcattaaaagtagcaattttatctttatagc
tgtttaagatattagaagcattcgaacggccaaaatcttcccaatgcaaa
tataaattaggaaaaagcgattcggcatgattaacaaatttgtcgataaa
atcatagtacttatcgccacgaacacgattaaatttatttcccaaataca
taggatctttcaaaagcttttcgttatttgttccagcatcaataacaact
gcaagaactgttgatggatcgattccggccgcaactgtataaaccatcag
tttgccgacagcaatatcaacaccctggacaccccagtctccaatcccaa
gaataccttcggcatcagaaacgaccagcagcttgatatcgcggccatta
gcagcatttttcagagtcgattgaatgttttccggatgattaatatccaa
aaaagcggcaccttgcggttcaacaaataattccgaataattttcaattg
tatcagcaattgtcgggtcataaacaattggcataaattcaacaacatgt
tgagaaaaaagcttataaaacaacacgtgattcgtattgaatatttccat
taaaaacaatcgtttttcgagatttgaaaccttgctttgaaattgagcat
aagtctgatcaacttgctcttgtaaagcctgaaccttggccggtaataaa
ccgtttaaaccaagctcctctctctccgcttccgtaaaagcagttccttt
gttaataaaaggatcatttaaaatacttactggatctgtcatgaattccg
atcttgttttatatttgttgtaaaaagtagataattacttccttgatgat
ctgtaaaaaagagaaaaagaaagcatctaagaacttgaaaaactacgaat
tagaaaagaccaaatatgtatttcttgcattgaccaatttatgcaagttt
atatatatgtaaatgtaagtttcacgaggttctactaaactaaaccaccc
ccttggttagaagaaaagagtgtgtgagaacaggctgttgttgtcacacg
attcggacaattctgtttgaaagagagagagtaacagtacgatcgaacga
actttgctctggagatcacagtgggcatcatagcatgtggtactaaaccc
tttcccgccattccagaaccttcgattgcttgttacaaaacctgtgagcc
gtcgctaggaccttgttgtgtgacgaaattggaagctgcaatcaatagga
agacaggaagtcgagcgtgtctgggtttttcagttttgttcttttttgca
aacaaatcacgagcgacggtaatttctttctcgataagaggccacgtgct
ttatgagggtaacatcaattcaagaaggagggaaacacttccttttctg
gccctgataatagtatgagggtgaagccaaaataaaggattcgcgcccaa
atcggcatctttaaatgcaggtatgcgatagttcctcactctttccttac
tcacgagtaattcttgcaaatgcctattatgcagatgttataatatctgt
gcgtcttgagttgaagtcaggaatctaaaataaaaattaaggttaataaa
aagaggaaagaaaaaaaaattaatcgatttacagaaacttgcacactaaa
aatacacaactaaaagcaattacagtatgggaagtcatcgacgttatctc

000088

tactatagtatattatcatttctattattatcctgctcagtggtacttgc
aaaacaagataagaccccattctttgaaggtacttcttcgaaaaattcgc
gtctaactgcacaagataagggcaatgatacatgcccgccatgttttaat
tgtatgctacctatttttgaatgcaaacagttttctgaatgcaattcgta
cactggtagatgtgagtgtatagaagggtttgcaggtgatgattgctctc
tgcccctctgtggcggtctatcaccggatgaaagcggtaataaggatcgt
cccataagagcacaaaatgacacctgtcattgtgataacggatggggagg
gatcaattgtgacgtttgtcaagaagattttgtctgtgatgcgttcatgc
ctgatcctagtattaaggggacatgttataagaatggtatgattgtagat
aaagtattttcaggttgtaatgtgaccaatgagaaaattctacagatttt
gaacggcaaaataccacaaattacatttgcctgtgataaacctaatcaag
aatgtaattttcagttttggatagatcagttagaaagcttgcggccgcac
gaaggaacctagaggccttttgatgttagcagaattgtcatgcaagggct
ccctagctactggagaatatactaagggtactgttgacattgcgaagagc
gacaaagattttgttatcggctttattgctcaaagagacatgggtggaag
agatgaaggttacgattggttgattatgacacccggtgtgggtttagatg
acaagggagacgcattgggtcaacagtatagaaccgtggatgatgtggtc
tctacaggatctgacattattattgttggaagaggactatttgcaaaggg
aagggatgctaaggtagagggtgaacgttacagaaaagcaggctgggaag
catatttgagaagatgcggccagcaaaactaaaaaactgtattataagta
aatgcatgtatactaaactcacaaattagagcttcaatttaattatatca
gttattacccgggaatctcggtcgtaatgatttttataatgacgaaaaaa
aaaaaattggaaagaaaaagcttcatggcctttataaaaaggaactatcc
aatacctcgccagaaccaagtaacagtattttacggggcacaaatcaaga
acaataagacaggactgtaaagatggacgcattgaactccaaagaacaac
aagagttccaaaaagtagtggaacaaaagcaaatgaaggatttcatgcgt
ttgtactctaatctggtagaaagatgtttcacagactgtgtcaatgactt
cacaacatcaaagctaaccaataaggaacaaacatgcatcataaagtgct
cagaaaagttcttgaagcatagcgaacgtgtagggcagcgtttccaagaa
caaaacgctgccttgggacaaggcttgggccgataaggtgtactggcgta
tatatatctaattatgtatctctggtgtagccc

000089

Appendix 2 : Sequence alignment of the integrated DNA from strain ML01 and previously published sequences

000090

half *Srf*I site which starts the integration locus

```
Genomic Sequence   ..........GGGCAACGGTTCATCATCTCATGGATCTGC     30
Published Sequence ..........GGGCAACGGTTCATCATCTCATGGATCTGC

Genomic Sequence   ACATGAACAAACACCAGAGTCAAACGACGTTGAAATTGAG     70
Published Sequence ACATGAACAAACACCAGAGTCAAACGACGTTGAAATTGAG

Genomic Sequence   GCTACTGCGCCAATTGATGACAATACAGACGATGATAACA    110
Published Sequence GCTACTGCGCCAATTGATGACAATACAGACGATGATAACA

Genomic Sequence   AACCGAAGTTATCTGATGTAGAAAAGGATTAAAGATGCTA    150
Published Sequence AACCGAAGTTATCTGATGTAGAAAAGGATTAAAGATGCTA

Genomic Sequence   AGAGATAGTGATGATATTTCATAAATAATGTAATTCTATA    190
Published Sequence AGAGATAGTGATGATATTTCATAAATAATGTAATTCTATA

Genomic Sequence   TATGTTAATTACCTTTTTTGCGAGGCATATTTATGGTGAA    230
Published Sequence TATGTTAATTACCTTTTTTGCGAGGCATATTTATGGTGAA

Genomic Sequence   GGATAAGTTTTGACCATCAAAGAAGGTTAATGTGGCTGTG    270
Published Sequence GGATAAGTTTTGACCATCAAAGAAGGTTAATGTGGCTGTG

Genomic Sequence   GTTTCAGGGTCCATAAAGCTTTTCAATTCATCTTTTTTTT    310
Published Sequence GTTTCAGGGTCCATAAAGCTTTTCAATTCATCTTTTTTTT

Genomic Sequence   TTTTGTTCTTTTTTTTGATTCCGGTTTCTTTGAAATTTTT    350
Published Sequence TTTTGTTCTTTTTTTTGATTCCGGTTTCTTTGAAATTTTT

Genomic Sequence   TTGATTCGGTAATCTCCGAGCAGAAGGAAGAACGAAGGAA    390
Published Sequence TTGATTCGGTAATCTCCGAGCAGAAGGAAGAACGAAGGAA

Genomic Sequence   GGAGCACAGACTTAGATTGGTATATATACGCATATGTGGT    430
Published Sequence GGAGCACAGACTTAGATTGGTATATATACGCATATGTGGT

Genomic Sequence   GTTGAAGAAACATGAAATTGCCCAGTATTCTTAACCCAAC    470
Published Sequence GTTGAAGAAACATGAAATTGCCCAGTATTCTTAACCCAAC

Genomic Sequence   TGCACAGAACAAAAACCTGCAGGAAACGAAGATAAATCAT    510
Published Sequence TGCACAGAACAAAAACCTGCAGGAAACGAAGATAAATCAT

Genomic Sequence   GTCGAAAGCTACATATAAGGAACGTGCTGCTACTCATCCT    550
Published Sequence GTCGAAAGCTACATATAAGGAACGTGCTGCTACTCATCCT

Genomic Sequence   AGTCCTGTTGCTGCCAAGCTATTTAATATCATGCACGAAA    590
Published Sequence AGTCCTGTTGCTGCCAAGCTATTTAATATCATGCACGAAA

Genomic Sequence   AGCAAACAAACTTGTGTGCTTCATTGGATGTTCGTACCAC    630
Published Sequence AGCAAACAAACTTGTGTGCTTCATTGGATGTTCGTACCAC

Genomic Sequence   CAAGGAATTACTGGAGTTAGTTGAAGCATTAGGTCCCAAA    670
Published Sequence CAAGGAATTACTGGAGTTAGTTGAAGCATTAGGTCCCAAA

Genomic Sequence   ATTTGTTTACTAAAAACACATGTGGATATCTTGACTGATT    710
Published Sequence ATTTGTTTACTAAAAACACATGTGGATATCTTGACTGATT

Genomic Sequence   TTTCCATGGAGGGCACAGTTAAGCCGCTAAAGGCATTATC    750
Published Sequence TTTCCATGGAGGGCACAGTTAAGCCGCTAAAGGCATTATC

Genomic Sequence   CGCCAAGTACAATTTTTTACTCTTCGAAGACAGAAAATTT    790
Published Sequence CGCCAAGTACAATTTTTTACTCTTCGAAGACAGAAAATTT
                                          Comment #1
Genomic Sequence   GCTGACATTGGTAATACAGTCAAATTGCAGCACTCTGCGG    830
Published Sequence GCTGACATTGGTAATACAGTCAAATTGCAGtACTCTGCGG
```

```
Genomic Sequence   GTGTATACAGAATAGCAGAATGGGCAGACATTACGAATGC    870
Published Sequence GTGTATACAGAATAGCAGAATGGGCAGACATTACGAATGC

Genomic Sequence   ACACGGTGTGGTGGGCCCAGGTATTGTTAGCGGTTTGAAG    910
Published Sequence ACACGGTGTGGTGGGCCCAGGTATTGTTAGCGGTTTGAAG
                                   Comment #2
Genomic Sequence   CAGGCGGCGGAAGAAGTAGGTACCAAGCTTTAACGAACGC    950
Published Sequence CAGGCGGCGGAAGAAGTA......AAGCTTTAACGAACGC

Genomic Sequence   AGAATTTTCGAGTTATTAAACTTAAAATACGCTGAACCCG    990
Published Sequence AGAATTTTCGAGTTATTAAACTTAAAATACGCTGAACCCG

Genomic Sequence   AACATAGAAATATCGAATGGGAAAAAAAAACTGCATAAAG   1030
Published Sequence AACATAGAAATATCGAATGGGAAAAAAAAACTGCATAAAG

Genomic Sequence   GCATTAAAAGAGGAGCGAATTTTTTTTTAATAAAAATCTT   1070
Published Sequence GCATTAAAAGAGGAGCGAATTTTTTTTTAATAAAAATCTT

Genomic Sequence   AATAATCATTAAAAGATAAATAATAGTCTATATATACGTA   1110
Published Sequence AATAATCATTAAAAGATAAATAATAGTCTATATATACGTA

Genomic Sequence   TATAAATAAAAAATATTCAAAAAATAAAATAAACTATTAT   1150
Published Sequence TATAAATAAAAAATATTCAAAAAATAAAATAAACTATTAT

Genomic Sequence   TTTAGCGTAAAGGATGGGGAAAGAGAAAAGAAAAAAATTG   1190
Published Sequence TTTAGCGTAAAGGATGGGGAAAGAGAAAAGAAAAAAATTG
                                   Comment #3
Genomic Sequence   ATCTATCGCAGATCCCTCGAGGAGATCTCATATGCAAGAC   1230
Published Sequence ATCTATCG...................TCATATGCAAGAC

Genomic Sequence   GCATACATATTTAACACTACGACAACCAAAAATTACCATT   1270
Published Sequence GCATACATATTTAACACTACGACAACCAAAAATTACCATT

Genomic Sequence   ATTCAAAGTCCAAAATTTTGGTATGTATAATTTAAACAAA   1310
Published Sequence ATTCAAAGTCCAAAATTTTGGTATGTATAATTTAAACAAA

Genomic Sequence   AAGAAGAAGAAGGAGCAAAAATGTCTAAGCAAAATTAGTA   1350
Published Sequence AAGAAGAAGAAGGAGCAAAAATGTCTAAGCAAAATTAGTA

Genomic Sequence   CCAGTAGTACAATAAAGATACAGAATGATTCAAAATGATT   1390
Published Sequence CCAGTAGTACAATAAAGATACAGAATGATTCAAAATGATT

Genomic Sequence   AATCTATAATGGTATTAAGAGCAGGGCACACAGATTTATA   1430
Published Sequence AATCTATAATGGTATTAAGAGCAGGGCACACAGATTTATA

Genomic Sequence   GAAAAATTAAGGAAAAGCATACAAGCTTAAACGCTTTCAT   1470
Published Sequence GAAAAATTAAGGAAAAGCATACAAGCTTAAACGCTTTCAT

Genomic Sequence   GTTCACTACTAGGAGGATCCGATTCACCACCAGTAGATGT   1510
Published Sequence GTTCACTACTAGGAGGATCCGATTCACCACCAGTAGATGT

Genomic Sequence   GACATGTGTATCGACTTTTTCCAAAGATGCAGGTGCTTTT   1550
Published Sequence GACATGTGTATCGACTTTTTCCAAAGATGCAGGTGCTTTT

Genomic Sequence   TCAGGTGGGAAGGTAGGGTTAAGGACACCTGTATTTGGTT   1590
Published Sequence TCAGGTGGGAAGGTAGGGTTAAGGACACCTGTATTTGGTT

Genomic Sequence   TTGGTGGAGGATGGGCATCTTCGTCTTTGCCAGGATAGCA   1630
Published Sequence TTGGTGGAGGATGGGCATCTTCGTCTTTGCCAGGATAGCA

Genomic Sequence   AAGATCATTGACGAGAAACGCACGGACCATTAAATACATT   1670
Published Sequence AAGATCATTGACGAGAAACGCACGGACCATTAAATACATT

Genomic Sequence   AGGAGGATCCACTGAATACAAAGAATGACCCCAATGATAT   1710
Published Sequence AGGAGGATCCACTGAATACAAAGAATGACCCCAATGATAT
```

```
Genomic Sequence   GTCCAAACATTTGGAAAGCTTTGGAATCTATCATTTTACC   1750
Published Sequence GTCCAAACATTTGGAAAGCTTTGGAATCTATCATTTTACC

Genomic Sequence   TATCTCAATGGTACAATTAACAAAACCCACGTTGGGGAAA   1790
Published Sequence TATCTCAATGGTACAATTAACAAAACCCACGTTGGGGAAA

Genomic Sequence   ATGAATGCAAACCATCCACAAGCAAACTTGAGAGGGGCTC   1830
Published Sequence ATGAATGCAAACCATCCACAAGCAAACTTGAGAGGGGCTC

Genomic Sequence   GAGTGAAAAAGCCCGCTAAAAAGCTAACCATGGCGAGACA   1870
Published Sequence GAGTGAAAAAGCCCGCTAAAAAGCTAACCATGGCGAGACA

Genomic Sequence   GTAACACCAAGCAGCAAGACCCCAAATAAAAATAGCCATA   1910
Published Sequence GTAACACCAAGCAGCAAGACCCCAAATAAAAATAGCCATA

Genomic Sequence   AAGGTAGAAACAAAACCAAGATACTCGGATGAGTTGGCGC   1950
Published Sequence AAGGTAGAAACAAAACCAAGATACTCGGATGAGTTGGCGC

Genomic Sequence   CAACAAAAATATAAGGGCGACTGCCCATAGCACCACGCGC   1990
Published Sequence CAACAAAAATATAAGGGCGACTGCCCATAGCACCACGCGC

Genomic Sequence   AATATTAATTAAGGCCAAACCTGAGAAAGCTGGTGGACCG   2030
Published Sequence AATATTAATTAAGGCCAAACCTGAGAAAGCTGGTGGACCG

Genomic Sequence   ACAAACATAAACATACCAGGTCGATCTTGGGGTTTTGCCA   2070
Published Sequence ACAAACATAAACATACCAGGTCGATCTTGGGGTTTTGCCA

Genomic Sequence   GGCCTACAGTAAAAAACCGTAAGACATTGACGGCAAACAG   2110
Published Sequence GGCCTACAGTAAAAAACCGTAAGACATTGACGGCAAACAG

Genomic Sequence   TAAAAGATAAACCCAAAAACCAAGTCCTTGAAAGAGGATA   2150
Published Sequence TAAAAGATAAACCCAAAAACCAAGTCCTTGAAAGAGGATA

Genomic Sequence   CCAAAGATAACCATATTTTTTAATTGATGAGCGGGTTGTG   2190
Published Sequence CCAAAGATAACCATATTTTTTAATTGATGAGCGGGTTGTG

Genomic Sequence   TAGAATTGACGGCGCCAGCAATGACACCACAAATCATAGG   2230
Published Sequence TAGAATTGACGGCGCCAGCAATGACACCACAAATCATAGG

Genomic Sequence   AGGGAAAATAGGAAGAATCCAAGCAGGAGATGCGGTTTCA   2270
Published Sequence AGGGAAAATAGGAAGAATCCAAGCAGGAGATGCGGTTTCA

Genomic Sequence   ATGGTATATACATGGTTGTTGAAAATTGTAAAAAAAGCCA   2310
Published Sequence ATGGTATATACATGGTTGTTGAAAATTGTAAAAAAAGCCA

Genomic Sequence   TTACGCAGTATATAAAGGATACTGCAACGTAAATGTAATA   2350
Published Sequence TTACGCAGTATATAAAGGATACTGCAACGTAAATGTAATA

Genomic Sequence   AAGGATTCGAATGACCCACACCATCCACTCGCCGGTATCA   2390
Published Sequence AAGGATTCGAATGACCCACACCATCCACTCGCCGGTATCA

Genomic Sequence   GGATAGGCGTATATGGCAAGCATGTCGATGAACGTGGATA   2430
Published Sequence GGATAGGCGTATATGGCAAGCATGTCGATGAACGTGGATA

Genomic Sequence   TTGAAAGAAGACAAGTAGCAATGAAAAGCTTTTCCAAATG   2470
Published Sequence TTGAAAGAAGACAAGTAGCAATGAAAAGCTTTTCCAAATG

Genomic Sequence   ATGGTTCCAGGAATCCTTGATAGTTGAAGGATATTTAATA   2510
Published Sequence ATGGTTCCAGGAATCCTTGATAGTTGAAGGATATTTAATA

Genomic Sequence   AAGCGAAAAAGCATGCATGATCCAAAGAGAGAAAACAAAA   2550
Published Sequence AAGCGAAAAAGCATGCATGATCCAAAGAGAGAAAACAAAA
```

```
Genomic Sequence   AGATTTGAAGAATATAAACAATTTTGCCAATTGTATTAAG   2590
Published Sequence AGATTTGAAGAATATAAACAATTTTGCCAATTGTATTAAG

Genomic Sequence   ACCATAAAATCGAAAGGGGAAAGAACCAATAATCAAACCA   2630
Published Sequence ACCATAAAATCGAAAGGGGAAAGAACCAATAATCAAACCA

Genomic Sequence   ACACCACCAGTTGCCATAGTACATGCAAACCAAGACCATG   2670
Published Sequence ACACCACCAGTTGCCATAGTACATGCAAACCAAGACCATG

Genomic Sequence   TAAAATGCTTCAGTCGTTGACTGAGAGGGACATGAGGGGC   2710
Published Sequence TAAAATGCTTCAGTCGTTGACTGAGAGGGACATGAGGGGC

Genomic Sequence   TTTGACATTCCAGTCAAGCAACTCATGATACCTCTGTTTC   2750
Published Sequence TTTGACATTCCAGTCAAGCAACTCATGATACCTCTGTTTC

Genomic Sequence   AAGATTTCCTTGAGTTCACCCATGTCGAATCGAGAGATAA   2790
Published Sequence AAGATTTCCTTGAGTTCACCCATGTCGAATCGAGAGATAA
                                                Comment #4
Genomic Sequence   AGGGGAATTAAAAAAAAAATAGTGGCCAGAATTCCGATCT   2830
Published Sequence AGGGGAATTAAAAAAAAAATAGTGGCCA...........T

Genomic Sequence   TGTTTTATATTTGTTGTAAAAAGTAGATAATTACTTCCTT   2870
Published Sequence TGTTTTATATTTGTTGTAAAAAGTAGATAATTACTTCCTT
                                            Comment #5
Genomic Sequence   GATGATCTGTAAAAAAGAGAAAAAGGAAGCATCTAAGAAC   2910
Published Sequence GATGATCTGTAAAAAAGAGAAAAAGaAAGCATCTAAGAAC

Genomic Sequence   TTGAAAAACTACGAATTAGAAAAGACCAAATATGTATTTC   2950
Published Sequence TTGAAAAACTACGAATTAGAAAAGACCAAATATGTATTTC

Genomic Sequence   TTGCATTGACCAATTTATGCAAGTTTATATATATGTAAAT   2990
Published Sequence TTGCATTGACCAATTTATGCAAGTTTATATATATGTAAAT

Genomic Sequence   GTAAGTTTCACGAGGTTCTACTAAACTAAACCACCCCCTT   3030
Published Sequence GTAAGTTTCACGAGGTTCTACTAAACTAAACCACCCCCTT

Genomic Sequence   GGTTAGAAGAAAAGAGTGTGTGAGAACAGGCTGTTGTTGT   3070
Published Sequence GGTTAGAAGAAAAGAGTGTGTGAGAACAGGCTGTTGTTGT

Genomic Sequence   CACACGATTCGGACAATTCTGTTTGAAAGAGAGAGAGTAA   3110
Published Sequence CACACGATTCGGACAATTCTGTTTGAAAGAGAGAGAGTAA

Genomic Sequence   CAGTACGATCGAACGAACTTTGCTCTGGAGATCACAGTGG   3150
Published Sequence CAGTACGATCGAACGAACTTTGCTCTGGAGATCACAGTGG

Genomic Sequence   GCATCATAGCATGTGGTACTAAACCCTTTCCCGCCATTCC   3190
Published Sequence GCATCATAGCATGTGGTACTAAACCCTTTCCCGCCATTCC

Genomic Sequence   AGAACCTTCGATTGCTTGTTACAAAACCTGTGAGCCGTCG   3230
Published Sequence AGAACCTTCGATTGCTTGTTACAAAACCTGTGAGCCGTCG

Genomic Sequence   CTAGGACCTTGTTGTGTGACGAAATTGGAAGCTGCAATCA   3270
Published Sequence CTAGGACCTTGTTGTGTGACGAAATTGGAAGCTGCAATCA

Genomic Sequence   ATAGGAAGACAGGAAGTCGAGCGTGTCTGGGTTTTTTCAG   3310
Published Sequence ATAGGAAGACAGGAAGTCGAGCGTGTCTGGGTTTTTTCAG

Genomic Sequence   TTTTGTTCTTTTTGCAAACAAATCACGAGCGACGGTAATT   3350
Published Sequence TTTTGTTCTTTTTGCAAACAAATCACGAGCGACGGTAATT

Genomic Sequence   TCTTTCTCGATAAGAGGCCACGTGCTTTATGAGGGTAACA   3390
Published Sequence TCTTTCTCGATAAGAGGCCACGTGCTTTATGAGGGTAACA
```

000094

```
Genomic Sequence   TCAATTCAAGAAGGAGGGAAACACTTCCTTTTTCTGGCCC   3430
Published Sequence TCAATTCAAGAAGGAGGGAAACACTTCCTTTTTCTGGCCC

Genomic Sequence   TGATAATAGTATGAGGGTGAAGCCAAAATAAAGGATTCGC   3470
Published Sequence TGATAATAGTATGAGGGTGAAGCCAAAATAAAGGATTCGC

Genomic Sequence   GCCCAAATCGGCATCTTTAAATGCAGGTATGCGATAGTTC   3510
Published Sequence GCCCAAATCGGCATCTTTAAATGCAGGTATGCGATAGTTC

Genomic Sequence   CTCACTCTTTCCTTACTCACGAGTAATTCTTGCAAATGCC   3550
Published Sequence CTCACTCTTTCCTTACTCACGAGTAATTCTTGCAAATGCC

Genomic Sequence   TATTATGCAGATGTTATAATATCTGTGCGTCTTGAGTTGA   3590
Published Sequence TATTATGCAGATGTTATAATATCTGTGCGTCTTGAGTTGA

Genomic Sequence   AGTCAGGAATCTAAAATAAAAATTAAGGTTAATAAAAAGA   3630
Published Sequence AGTCAGGAATCTAAAATAAAAATTAAGGTTAATAAAAAGA

Genomic Sequence   GGAAAGAAAAAAAAATTAATCGATTTACAGAAACTTGCAC   3670
Published Sequence GGAAAGAAAAAAAAATTAATCGATTTACAGAAACTTGCAC

Genomic Sequence   ACTAAAAATACACAACTAAAAGCAATTACAGTATGGGAAG   3710
Published Sequence ACTAAAAATACACAACTAAAAGCAATTACAGTATGGGAAG

Genomic Sequence   TCATCGACGTTATCTCTACTATAGTATATTATCATTTCTA   3750
Published Sequence TCATCGACGTTATCTCTACTATAGTATATTATCATTTCTA

Genomic Sequence   TTATTATCCTGCTCAGTGGTACTTGCAAAACAAGATAAGA   3790
Published Sequence TTATTATCCTGCTCAGTGGTACTTGCAAAACAAGATAAGA

Genomic Sequence   CCCCATTCTTTGAAGGTACTTCTTCGAAAAATTCGCGTCT   3830
Published Sequence CCCCATTCTTTGAAGGTACTTCTTCGAAAAATTCGCGTCT

Genomic Sequence   AACTGCACAAGATAAGGGCAATGATACATGCCCGCCATGT   3870
Published Sequence AACTGCACAAGATAAGGGCAATGATACATGCCCGCCATGT

Genomic Sequence   TTTAATTGTATGCTACCTATTTTTGAATGCAAACAGTTTT   3910
Published Sequence TTTAATTGTATGCTACCTATTTTTGAATGCAAACAGTTTT

Genomic Sequence   CTGAATGCAATTCGTACACTGGTAGATGTGAGTGTATAGA   3950
Published Sequence CTGAATGCAATTCGTACACTGGTAGATGTGAGTGTATAGA

Genomic Sequence   AGGGTTTGCAGGTGATGATTGCTCTCTGCCCCTCTGTGGC   3990
Published Sequence AGGGTTTGCAGGTGATGATTGCTCTCTGCCCCTCTGTGGC

Genomic Sequence   GGTCTATCACCGGATGAAAGCGGTAATAAGGATCGTCCCA   4030
Published Sequence GGTCTATCACCGGATGAAAGCGGTAATAAGGATCGTCCCA

Genomic Sequence   TAAGAGCACAAAATGACACCTGTCATTGTGATAACGGATG   4070
Published Sequence TAAGAGCACAAAATGACACCTGTCATTGTGATAACGGATG

Genomic Sequence   GGGAGGGATCAATTGTGACGTTTGTCAAGAAGATTTTGTC   4110
Published Sequence GGGAGGGATCAATTGTGACGTTTGTCAAGAAGATTTTGTC

Genomic Sequence   TGTGATGCGTTCATGCCTGATCCTAGTATTAAGGGGACAT   4150
Published Sequence TGTGATGCGTTCATGCCTGATCCTAGTATTAAGGGGACAT

Genomic Sequence   GTTATAAGAATGGTATGATTGTAGATAAAGTATTTTCAGG   4190
Published Sequence GTTATAAGAATGGTATGATTGTAGATAAAGTATTTTCAGG

Genomic Sequence   TTGTAATGTGACCAATGAGAAAATTCTACAGATTTTGAAC   4230
Published Sequence TTGTAATGTGACCAATGAGAAAATTCTACAGATTTTGAAC

Genomic Sequence   GGCAAAATACCACAAATTACATTTGCCTGTGATAAACCTA   4270
Published Sequence GGCAAAATACCACAAATTACATTTGCCTGTGATAAACCTA
```

000095

```
                                          Comment #6
Genomic Sequence   ATCAAGAATGTAATTTTCAGTTTTGGACAGATCAGTTAGA   4310
Published Sequence ATCAAGAATGTAATTTTCAGTTTTGGAtAGATCAGTTAGA
                              Comment #7
Genomic Sequence   AAGCTTGGTACCGCGGCCGCAAGCTTTAACGAACGCAGAA   4350
Published Sequence AAGCTT..............AAGCTTTAACGAACGCAGAA

Genomic Sequence   TTTTCGAGTTATTAAACTTAAAATACGCTGAACCCGAACA   4390
Published Sequence TTTTCGAGTTATTAAACTTAAAATACGCTGAACCCGAACA

Genomic Sequence   TAGAAATATCGAATGGGAAAAAAAAACTGCATAAAGGCAT   4430
Published Sequence TAGAAATATCGAATGGGAAAAAAAAACTGCATAAAGGCAT

Genomic Sequence   TAAAAGAGGAGCGAATTTTTTTTTAATAAAAATCTTAATA   4470
Published Sequence TAAAAGAGGAGCGAATTTTTTTTTAATAAAAATCTTAATA

Genomic Sequence   ATCATTAAAAGATAAATAATAGTCTATATATACGTATATA   4510
Published Sequence ATCATTAAAAGATAAATAATAGTCTATATATACGTATATA

Genomic Sequence   AATAAAAAATATTCAAAAAATAAAATAAACTATTATTTTA   4550
Published Sequence AATAAAAAATATTCAAAAAATAAAATAAACTATTATTTTA
Genomic Sequence   GCGTAAAGGATGGGGAAAGAGAAAAGAAAAAAATTGATCT   4590
Published Sequence GCGTAAAGGATGGGGAAAGAGAAAAGAAAAAAATTGATCT
                              Comment #8          Comment #9
Genomic Sequence   ATCGCAGATCCCTCGAGGAGATCTCATTAGTATTTCGGCT   4630
Published Sequence ATCG...................TCATTAGTATTTCGGaT

Genomic Sequence   CCCACTTCAAATCGTCGACGGCCTTTTCAACATCATCAAT   4670
Published Sequence CCCACTTCAAATCGTCGACGGCCTTTTCAACATCATCAAT

Genomic Sequence   CGGCTGGCGATTAAGACCTTGTTCAACAGCTTTTTTAGCG   4710
Published Sequence CGGCTGGCGATTAAGACCTTGTTCAACAGCTTTTTTAGCG

Genomic Sequence   ACAGCGACAGCGACTGTACGCGAAAAGTCGGCTAATTTTG   4750
Published Sequence ACAGCGACAGCGACTGTACGCGAAAAGTCGGCTAATTTTG

Genomic Sequence   AAACTGGTGGCAAAACAGCAGCACCAACTTTTGTTGTATC   4790
Published Sequence AAACTGGTGGCAAAACAGCAGCACCAACTTTTGTTGTATC

Genomic Sequence   AACGATTCCTCCAAGACTATGGGCAGCAGCAGAAATCATT   4830
Published Sequence AACGATTCCTCCAAGACTATGGGCAGCAGCAGAAATCATT

Genomic Sequence   TCAGGCGTAAGCAGCTTTGATTGAGCGGCAATGGCACCAA   4870
Published Sequence TCAGGCGTAAGCAGCTTTGATTGAGCGGCAATGGCACCAA

Genomic Sequence   AGCCAAGACCTGGATAGATCAAGGCATTGTTGGCCTGACC   4910
Published Sequence AGCCAAGACCTGGATAGATCAAGGCATTGTTGGCCTGACC

Genomic Sequence   GATTTGGTAAGCGTTGCCCTCATATTCAATATCGTCAACT   4950
Published Sequence GATTTGGTAAGCGTTGCCCTCATATTCAATATCGTCAACT

Genomic Sequence   GGAACACCAGTACCGATCAAGGCTTTTCCATTAGACCATT   4990
Published Sequence GGAACACCAGTACCGATCAAGGCTTTTCCATTAGACCATT

Genomic Sequence   TCAAAACATCTTCGGCTTTTGCTTCGGCTAATTTTGTTGG   5030
Published Sequence TCAAAACATCTTCGGCTTTTGCTTCGGCTAATTTTGTTGG

Genomic Sequence   ATTGGAAATTGGAAAAATGATTGGTCTTTCAGTATAACCA   5070
Published Sequence ATTGGAAATTGGAAAAATGATTGGTCTTTCAGTATAACCA

Genomic Sequence   GACATATCTTTAACAATTTCTTCAGTAAAGGAATTTGGAT   5110
Published Sequence GACATATCTTTAACAATTTCTTCAGTAAAGGAATTTGGAT

Genomic Sequence   GTGTCGAGGTTCCGACCAAAATGGTCGGGTGGACAGCTTC   5150
Published Sequence GTGTCGAGGTTCCGACCAAAATGGTCGGGTGGACAGCTTC
```

000096

```
Genomic Sequence   AACAGCTGCTTGGAGATTGGTCAATTGATTAGCATTTTTG   5190
Published Sequence AACAGCTGCTTGGAGATTGGTCAATTGATTAGCATTTTTG

Genomic Sequence   AAATCACTTCGTTTAGCAGCGAAAGGCTTTTGCTCTGGAG   5230
Published Sequence AAATCACTTCGTTTAGCAGCGAAAGGCTTTTGCTCTGGAG
                              Comment #10
Genomic Sequence   TTAAATCCGGATCATCATCAAACAAGAGGCCTTGTTTGTC   5270
Published Sequence TTAAATCCGGATCATCgTCAAACAAGAGGCCTTGTTTGTC

Genomic Sequence   AACAAGGAAGAAATGCTTTTTAGCCTCTTCGTCGGAAAGA   5310
Published Sequence AACAAGGAAGAAATGCTTTTTAGCCTCTTCGTCGGAAAGA

Genomic Sequence   CCCTGTTCAACCATTTCTTCATGCAACTGTTTAACAATTC   5350
Published Sequence CCCTGTTCAACCATTTCTTCATGCAACTGTTTAACAATTC

Genomic Sequence   CCATTCCAGCAGTTCCGGCACCGAAGCTCATGTAAGTTTG   5390
Published Sequence CCATTCCAGCAGTTCCGGCACCGAAGCTCATGTAAGTTTG

Genomic Sequence   ATCAGTTAATTTCTGACCGGAAATCTTCAACGCTCCAAGA   5430
Published Sequence ATCAGTTAATTTCTGACCGGAAATCTTCAACGCTCCAAGA

Genomic Sequence   ACGCCGGCAAGAACGACGATTCCAGTTCCTTGAATATCAT   5470
Published Sequence ACGCCGGCAAGAACGACGATTCCAGTTCCTTGAATATCAT
Genomic Sequence   CATTAAAAGTAGCAATTTTATCTTTATAGCTGTTTAAGAT   5510
Published Sequence CATTAAAAGTAGCAATTTTATCTTTATAGCTGTTTAAGAT

Genomic Sequence   ATTAGAAGCATTCGAACGGCCAAAATCTTCCCAATGCAAA   5550
Published Sequence ATTAGAAGCATTCGAACGGCCAAAATCTTCCCAATGCAAA

Genomic Sequence   TATAAATTAGGAAAAAGCGATTCGGCATGATTAACAAATT   5590
Published Sequence TATAAATTAGGAAAAAGCGATTCGGCATGATTAACAAATT

Genomic Sequence   TGTCGATAAAATCATAGTACTTATCGCCACGAACACGATT   5630
Published Sequence TGTCGATAAAATCATAGTACTTATCGCCACGAACACGATT

Genomic Sequence   AAATTTATTTCCCAAATACATAGGATCTTTCAAAAGCTTT   5670
Published Sequence AAATTTATTTCCCAAATACATAGGATCTTTCAAAAGCTTT


Genomic Sequence   TCGTTATTTGTTCCAGCATCAATAACAACTGCAAGAACTG   5710
Published Sequence TCGTTATTTGTTCCAGCATCAATAACAACTGCAAGAACTG

Genomic Sequence   TTGATGGATCGATTCCGGCCGCAACTGTATAAACCATCAG   5750
Published Sequence TTGATGGATCGATTCCGGCCGCAACTGTATAAACCATCAG

Genomic Sequence   TTTGCCGACAGCAATATCAACACCCTGGACACCCCAGTCT   5790
Published Sequence TTTGCCGACAGCAATATCAACACCCTGGACACCCCAGTCT

Genomic Sequence   CCAATCCCAAGAATACCTTCGGCATCAGAAACGACCAGCA   5830
Published Sequence CCAATCCCAAGAATACCTTCGGCATCAGAAACGACCAGCA

Genomic Sequence   GCTTGATATCGCGGCCATTAGCAGCATTTTTCAGAGTCGA   5870
Published Sequence GCTTGATATCGCGGCCATTAGCAGCATTTTTCAGAGTCGA

Genomic Sequence   TTGAATGTTTTCCGGATGATTAATATCCAAAAAAGCGGCA   5910
Published Sequence TTGAATGTTTTCCGGATGATTAATATCCAAAAAAGCGGCA

Genomic Sequence   CCTTGCGGTTCAACAAATAATTCCGAATAATTTTCAATTG   5950
Published Sequence CCTTGCGGTTCAACAAATAATTCCGAATAATTTTCAATTG

Genomic Sequence   TATCAGCAATTGTCGGGTCATAAACAATTGGCATAAATTC   5990
Published Sequence TATCAGCAATTGTCGGGTCATAAACAATTGGCATAAATTC
```

000097

```
Genomic Sequence   AACAACATGTTGAGAAAAAAGCTTATAAAACAACACGTGA   6030
Published Sequence AACAACATGTTGAGAAAAAAGCTTATAAAACAACACGTGA

Genomic Sequence   TTCGTATTGAATATTTCCATTAAAAACAATCGTTTTTCGA   6070
Published Sequence TTCGTATTGAATATTTCCATTAAAAACAATCGTTTTTCGA

Genomic Sequence   GATTTGAAACCTTGCTTTGAAATTGAGCATAAGTCTGATC   6110
Published Sequence GATTTGAAACCTTGCTTTGAAATTGAGCATAAGTCTGATC

Genomic Sequence   AACTTGCTCTTGTAAAGCCTGAACCTTGGCCGGTAATAAA   6150
Published Sequence AACTTGCTCTTGTAAAGCCTGAACCTTGGCCGGTAATAAA

Genomic Sequence   CCGTTTAAACCAAGCTCCTCTCTCTCCGCTTCCGTAAAAG   6190
Published Sequence CCGTTTAAACCAAGCTCCTCTCTCTCCGCTTCCGTAAAAG

Genomic Sequence   CAGTTCCTTTGTTAATAAAAGGATCATTTAAAATACTTAC   6230
Published Sequence CAGTTCCTTTGTTAATAAAAGGATCATTTAAAATACTTAC
                              Comment #11
Genomic Sequence   TGGATCTGTCATGAATTCCGATCTTGTTTTATATTTGTTG   6270
Published Sequence TGGATCTGTCAT...........TTGTTTTATATTTGTTG

Genomic Sequence   TAAAAAGTAGATAATTACTTCCTTGATGATCTGTAAAAAA   6310
Published Sequence TAAAAAGTAGATAATTACTTCCTTGATGATCTGTAAAAAA

Genomic Sequence   GAGAAAAAGAAAGCATCTAAGAACTTGAAAAACTACGAAT   6350
Published Sequence GAGAAAAAGAAAGCATCTAAGAACTTGAAAAACTACGAAT

Genomic Sequence   TAGAAAAGACCAAATATGTATTTCTTGCATTGACCAATTT   6390
Published Sequence TAGAAAAGACCAAATATGTATTTCTTGCATTGACCAATTT

Genomic Sequence   ATGCAAGTTTATATATATGTAAATGTAAGTTTCACGAGGT   6430
Published Sequence ATGCAAGTTTATATATATGTAAATGTAAGTTTCACGAGGT

Genomic Sequence   TCTACTAAACTAAACCACCCCCTTGGTTAGAAGAAAAGAG   6470
Published Sequence TCTACTAAACTAAACCACCCCCTTGGTTAGAAGAAAAGAG

Genomic Sequence   TGTGTGAGAACAGGCTGTTGTTGTCACACGATTCGGACAA   6510
Published Sequence TGTGTGAGAACAGGCTGTTGTTGTCACACGATTCGGACAA

Genomic Sequence   TTCTGTTTGAAAGAGAGAGAGTAACAGTACGATCGAACGA   6550
Published Sequence TTCTGTTTGAAAGAGAGAGAGTAACAGTACGATCGAACGA

Genomic Sequence   ACTTTGCTCTGGAGATCACAGTGGGCATCATAGCATGTGG   6590
Published Sequence ACTTTGCTCTGGAGATCACAGTGGGCATCATAGCATGTGG

Genomic Sequence   TACTAAACCCTTTCCCGCCATTCCAGAACCTTCGATTGCT   6630
Published Sequence TACTAAACCCTTTCCCGCCATTCCAGAACCTTCGATTGCT

Genomic Sequence   TGTTACAAAACCTGTGAGCCGTCGCTAGGACCTTGTTGTG   6670
Published Sequence TGTTACAAAACCTGTGAGCCGTCGCTAGGACCTTGTTGTG

Genomic Sequence   TGACGAAATTGGAAGCTGCAATCAATAGGAAGACAGGAAG   6710
Published Sequence TGACGAAATTGGAAGCTGCAATCAATAGGAAGACAGGAAG

Genomic Sequence   TCGAGCGTGTCTGGGTTTTTTCAGTTTTGTTCTTTTTGCA   6750
Published Sequence TCGAGCGTGTCTGGGTTTTTTCAGTTTTGTTCTTTTTGCA

Genomic Sequence   AACAAATCACGAGCGACGGTAATTTCTTTCTCGATAAGAG   6790
Published Sequence AACAAATCACGAGCGACGGTAATTTCTTTCTCGATAAGAG

Genomic Sequence   GCCACGTGCTTTATGAGGGTAACATCAATTCAAGAAGGAG   6830
Published Sequence GCCACGTGCTTTATGAGGGTAACATCAATTCAAGAAGGAG
```

000098

```
Genomic Sequence   GGAAACACTTCCTTTTTCTGGCCCTGATAATAGTATGAGG   6870
Published Sequence GGAAACACTTCCTTTTTCTGGCCCTGATAATAGTATGAGG

Genomic Sequence   GTGAAGCCAAAATAAAGGATTCGCGCCCAAATCGGCATCT   6910
Published Sequence GTGAAGCCAAAATAAAGGATTCGCGCCCAAATCGGCATCT

Genomic Sequence   TTAAATGCAGGTATGCGATAGTTCCTCACTCTTTCCTTAC   6950
Published Sequence TTAAATGCAGGTATGCGATAGTTCCTCACTCTTTCCTTAC

Genomic Sequence   TCACGAGTAATTCTTGCAAATGCCTATTATGCAGATGTTA   6990
Published Sequence TCACGAGTAATTCTTGCAAATGCCTATTATGCAGATGTTA

Genomic Sequence   TAATATCTGTGCGTCTTGAGTTGAAGTCAGGAATCTAAAA   7030
Published Sequence TAATATCTGTGCGTCTTGAGTTGAAGTCAGGAATCTAAAA

Genomic Sequence   TAAAAATTAAGGTTAATAAAAAGAGGAAAGAAAAAAAAAT   7070
Published Sequence TAAAAATTAAGGTTAATAAAAAGAGGAAAGAAAAAAAAAT

Genomic Sequence   TAATCGATTTACAGAAACTTGCACACTAAAAATACACAAC   7110
Published Sequence TAATCGATTTACAGAAACTTGCACACTAAAAATACACAAC

Genomic Sequence   TAAAAGCAATTACAGTATGGGAAGTCATCGACGTTATCTC   7150
Published Sequence TAAAAGCAATTACAGTATGGGAAGTCATCGACGTTATCTC

Genomic Sequence   TACTATAGTATATTATCATTTCTATTATTATCCTGCTCAG   7190
Published Sequence TACTATAGTATATTATCATTTCTATTATTATCCTGCTCAG

Genomic Sequence   TGGTACTTGCAAAACAAGATAAGACCCCATTCTTTGAAGG   7230
Published Sequence TGGTACTTGCAAAACAAGATAAGACCCCATTCTTTGAAGG

Genomic Sequence   TACTTCTTCGAAAAATTCGCGTCTAACTGCACAAGATAAG   7270
Published Sequence TACTTCTTCGAAAAATTCGCGTCTAACTGCACAAGATAAG

Genomic Sequence   GGCAATGATACATGCCCGCCATGTTTTAATTGTATGCTAC   7310
Published Sequence GGCAATGATACATGCCCGCCATGTTTTAATTGTATGCTAC

Genomic Sequence   CTATTTTTGAATGCAAACAGTTTTCTGAATGCAATTCGTA   7350
Published Sequence CTATTTTTGAATGCAAACAGTTTTCTGAATGCAATTCGTA

Genomic Sequence   CACTGGTAGATGTGAGTGTATAGAAGGGTTTGCAGGTGAT   7390
Published Sequence CACTGGTAGATGTGAGTGTATAGAAGGGTTTGCAGGTGAT

Genomic Sequence   GATTGCTCTCTGCCCCTCTGTGGCGGTCTATCACCGGATG   7430
Published Sequence GATTGCTCTCTGCCCCTCTGTGGCGGTCTATCACCGGATG

Genomic Sequence   AAAGCGGTAATAAGGATCGTCCCATAAGAGCACAAAATGA   7470
Published Sequence AAAGCGGTAATAAGGATCGTCCCATAAGAGCACAAAATGA

Genomic Sequence   CACCTGTCATTGTGATAACGGATGGGGAGGGATCAATTGT   7510
Published Sequence CACCTGTCATTGTGATAACGGATGGGGAGGGATCAATTGT

Genomic Sequence   GACGTTTGTCAAGAAGATTTTGTCTGTGATGCGTTCATGC   7550
Published Sequence GACGTTTGTCAAGAAGATTTTGTCTGTGATGCGTTCATGC

Genomic Sequence   CTGATCCTAGTATTAAGGGGACATGTTATAAGAATGGTAT   7590
Published Sequence CTGATCCTAGTATTAAGGGGACATGTTATAAGAATGGTAT

Genomic Sequence   GATTGTAGATAAAGTATTTTCAGGTTGTAATGTGACCAAT   7630
Published Sequence GATTGTAGATAAAGTATTTTCAGGTTGTAATGTGACCAAT

Genomic Sequence   GAGAAAATTCTACAGATTTTGAACGGCAAAATACCACAAA   7670
Published Sequence GAGAAAATTCTACAGATTTTGAACGGCAAAATACCACAAA

Genomic Sequence   TTACATTTGCCTGTGATAAACCTAATCAAGAATGTAATTT   7710
Published Sequence TTACATTTGCCTGTGATAAACCTAATCAAGAATGTAATTT
```

000099

```
                                                  Comment #12
Genomic Sequence    TCAGTTTTGGATAGATCAGTTAGAAAGCTTGCGGCCGCAC    7750
Published Sequence  TCAGTTTTGGATAGATCAGTTAGAAAGCTT........AC
                    Comment #13
Genomic Sequence    GAAGGAACCTAGAGGCCTTTTGATGTTAGCAGAATTGTCA    7790
Published Sequence  aAAGGAACCTAGAGGCCTTTTGATGTTAGCAGAATTGTCA

Genomic Sequence    TGCAAGGGCTCCCTAGCTACTGGAGAATATACTAAGGGTA    7830
Published Sequence  TGCAAGGGCTCCCTAGCTACTGGAGAATATACTAAGGGTA

Genomic Sequence    CTGTTGACATTGCGAAGAGCGACAAAGATTTTGTTATCGG    7870
Published Sequence  CTGTTGACATTGCGAAGAGCGACAAAGATTTTGTTATCGG

Genomic Sequence    CTTTATTGCTCAAAGAGACATGGGTGGAAGAGATGAAGGT    7910
Published Sequence  CTTTATTGCTCAAAGAGACATGGGTGGAAGAGATGAAGGT

Genomic Sequence    TACGATTGGTTGATTATGACACCCGGTGTGGGTTTAGATG    7950
Published Sequence  TACGATTGGTTGATTATGACACCCGGTGTGGGTTTAGATG

Genomic Sequence    ACAAGGGAGACGCATTGGGTCAACAGTATAGAACCGTGGA    7990
Published Sequence  ACAAGGGAGACGCATTGGGTCAACAGTATAGAACCGTGGA

Genomic Sequence    TGATGTGGTCTCTACAGGATCTGACATTATTATTGTTGGA    8030
Published Sequence  TGATGTGGTCTCTACAGGATCTGACATTATTATTGTTGGA

Genomic Sequence    AGAGGACTATTTGCAAAGGGAAGGGATGCTAAGGTAGAGG    8070
Published Sequence  AGAGGACTATTTGCAAAGGGAAGGGATGCTAAGGTAGAGG

Genomic Sequence    GTGAACGTTACAGAAAAGCAGGCTGGGAAGCATATTTGAG    8110
Published Sequence  GTGAACGTTACAGAAAAGCAGGCTGGGAAGCATATTTGAG


Genomic Sequence    AAGATGCGGCCAGCAAAACTAAAAAACTGTATTATAAGTA    8150
Published Sequence  AAGATGCGGCCAGCAAAACTAAAAAACTGTATTATAAGTA

Genomic Sequence    AATGCATGTATACTAAACTCACAAATTAGAGCTTCAATTT    8190
Published Sequence  AATGCATGTATACTAAACTCACAAATTAGAGCTTCAATTT

Genomic Sequence    AATTATATCAGTTATTACCCGGGAATCTCGGTCGTAATGA    8230
Published Sequence  AATTATATCAGTTATTACCCGGGAATCTCGGTCGTAATGA
            Comment #14
Genomic Sequence    TTTTTATAATGACGAAAAAAAAAAAAATTGGAAAGAAAAAG   8270
Published Sequence  TTTcTATAATGACGAAAAAAAAAAAAATTGGAAAGAAAAAG

Genomic Sequence    CTTCATGGCCTTTATAAAAAGGAACTATCCAATACCTCGC    8310
Published Sequence  CTTCATGGCCTTTATAAAAAGGAACTATCCAATACCTCGC

Genomic Sequence    CAGAACCAAGTAACAGTATTTTACGGGGCACAAATCAAGA    8350
Published Sequence  CAGAACCAAGTAACAGTATTTTACGGGGCACAAATCAAGA

Genomic Sequence    ACAATAAGACAGGACTGTAAAGATGGACGCATTGAACTCC    8390
Published Sequence  ACAATAAGACAGGACTGTAAAGATGGACGCATTGAACTCC

Genomic Sequence    AAAGAACAACAAGAGTTCCAAAAAGTAGTGGAACAAAAGC    8430
Published Sequence  AAAGAACAACAAGAGTTCCAAAAAGTAGTGGAACAAAAGC

Genomic Sequence    AAATGAAGGATTTCATGCGTTTGTACTCTAATCTGGTAGA    8470
Published Sequence  AAATGAAGGATTTCATGCGTTTGTACTCTAATCTGGTAGA

Genomic Sequence    AAGATGTTTCACAGACTGTGTCAATGACTTCACAACATCA    8510
Published Sequence  AAGATGTTTCACAGACTGTGTCAATGACTTCACAACATCA
                                                  Comment #15
Genomic Sequence    AAGCTAACCAATAAGGAACAAACATGCATCATAAAGTGCT    8550
Published Sequence  AAGCTAACCAATAAGGAACAAACATGCATCATgAAGTGCT
```

000100

```
Genomic Sequence   CAGAAAAGTTCTTGAAGCATAGCGAACGTGTAGGGCAGCG   8590
Published Sequence CAGAAAAGTTCTTGAAGCATAGCGAACGTGTAGGGCAGCG

Genomic Sequence   TTTCCAAGAACAAAACGCTGCCTTGGGACAAGGCTTGGGC   8630
Published Sequence TTTCCAAGAACAAAACGCTGCCTTGGGACAAGGCTTGGGC

Genomic Sequence   CGATAAGGTGTACTGGCGTATATATATCTAATTATGTATC   8670
Published Sequence CGATAAGGTGTACTGGCGTATATATATCTAATTATGTATC

Genomic Sequence   TCTGGTGTAGCCC..........                    8683
Published Sequence TCTGGTGTAGCCC..........
```

half *Srf*I site which stops the integration locus

000101

Discrepancies found when comparing the genomic sequence of the integrated locus and previously published sequences.
Comments and nucleotide positions reefer 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	929-934	Additional sequence resulting from cloning strategy
3	1199-1217	Additional sequence resulting from cloning strategy
4	2819-2829	Additional sequence resulting from cloning strategy
5	2896	Difference in the *PGK1*p sequence
6	4298	Difference in the *PGK1*p sequence
7	4317-4330	Additional sequence resulting from cloning strategy
8	4595-4613	Additional sequence resulting from cloning strategy
9	4629	Difference in the *mleA* open reading frame. This difference corresponds to a change of amino acids from aspartic acid (in the published sequence) to glutamic acid (in the genomic sequence)
10	5247	Difference in the *mleA* open reading frame. This difference corresponds to **no change** of the amino acid sequence
11	6243-6253	Additional sequence resulting from cloning strategy
12	7741-7748	Additional sequence resulting from cloning strategy
13	7751	Difference in the 3' region of the *URA3* open reading frame
14	8234	Difference in the 3' region of the *URA3* non coding sequence
15	8543	Difference in the 3' region of the *URA3* non coding sequence

Appendix 3 : Nucleotide sequence of the *mleA* and *mae1* genes and protein sequence of the corresponding proteins in Fasta format

000103

Mae1

Nucleotide sequence integrated in ML01

```
>readseq.input(1), 1317 bases, 2B384AC checksum.
ttaaacgctttcatgttcactactaggaggatccgattcaccaccagtag
atgtgacatgtgtatcgactttttccaaagatgcaggtgctttttcaggt
gggaaggtagggttaaggacacctgtatttggttttggtggaggatgggc
atcttcgtctttgccaggatagcaaagatcattgacgagaaacgcacgga
ccattaaatacattaggaggatccactgaatacaaagaatgaccccaatg
atatgtccaaacatttggaaagctttggaatctatcattttacctatctc
aatggtacaattaacaaaacccacgttggggaaaatgaatgcaaaccatc
cacaagcaaacttgagaggggctcgagtgaaaaagcccgctaaaaagcta
accatggcgagacagtaacaccaagcagcaagaccccaaataaaaatagc
cataaaggtagaaacaaaaccaagatactcggatgagttggcgccaacaa
aaatataagggcgactgcccatagcaccacgcgcaatattaattaaggcc
aaacctgagaaagctggtggaccgacaaacataaacataccaggtcgatc
ttggggttttgccaggcctacagtaaaaaaccgtaagacattgacggcaa
acagtaaaagataaacccaaaaaccaagtccttgaaagaggataccaaag
ataaccatatttttaattgatgagcgggttgtgtagaattgacggcgcc
agcaatgacaccacaaatcataggagggaaaataggaagaatccaagcag
gagatgcggtttcaatggtatatacatggttgttgaaaattgtaaaaaaa
gccattacgcagtatataaaggatactgcaacgtaaatgtaataaaggat
tcgaatgacccacaccatccactcgccggtatcaggataggcgtatatgg
caagcatgtcgatgaacgtggatattgaaagaagacaagtagcaatgaaa
agcttttccaaatgatggttccaggaatccttgatagttgaaggatattt
aataaagcgaaaaagcatgcatgatccaaagagagaaaacaaaaagattt
gaagaatataaacaattttgccaattgtattaagaccataaaatcgaaag
gggaaagaaccaataatcaaaccaacaccaccagttgccatagtacatgc
aaaccaagaccatgtaaaatgcttcagtcgttgactgagagggacatgag
gggctttgacattccagtcaagcaactcatgatacctctgtttcaagatt
tccttgagttcacccat
```

Reverse complement of the integrated nucleotide sequence (coding sequence)

```
>READSEQ.INPUT(1), 1317 BASES, 2B384AC CHECKSUM. : Sens inverse-complémentaire
de la séquence (1 - 1317)
atgggtgaactcaaggaaatcttgaaacagaggtatcatgagttgcttgactggaatgtc
aaagcccctcatgtccctctcagtcaacgactgaagcattttacatggtcttggtttgca
tgtactatggcaactggtggtgttggtttgattattggttctttccccttcgattttat
ggtcttaatacaattggcaaaattgtttatattcttcaaatctttttgttttctctcttt
ggatcatgcatgctttttcgctttattaaatatccttcaactatcaaggattcctggaac
catcatttggaaaagcttttcattgctacttgtcttctttcaatatccacgttcatcgac
atgcttgccatatacgcctatcctgataccggcgagtggatggtgtgggtcattcgaatc
ctttattacatttacgttgcagtatcctttatatactgcgtaatggctttttttacaatt
ttcaacaaccatgtatataccattgaaaccgcatctcctgcttggattcttcctattttc
cctcctatgatttgtggtgtcattgctggcgccgtcaattctacacaacccgctcatcaa
ttaaaaaatatggttatctttggtatcctctttcaaggacttggtttttgggtttatctt
ttactgtttgccgtcaatgtcttacggtttttttactgtaggcctggcaaaaccccaagat
cgacctggtatgtttatgtttgtcggtccaccagctttctcaggtttggccttaattaat
attgcgcgtggtgctatgggcagtcgcccttatatttttgttggcgccaactcatccgag
tatcttggttttgtttctacctttatggctatttttatttggggtcttgctgcttggtgt
tactgtctcgccatggttagctttttagcgggctttttcactcgagcccctctcaagttt
gcttgtggatggtttgcattcattttccccaacgtgggttttgttaattgtaccattgag
ataggtaaaatgatagattccaaagctttccaaatgtttggacatatcattggggtcatt
ctttgtattcagtggatcctcctaatgtatttaatggtccgtgcgtttctcgtcaatgat
ctttgctatcctggcaaagacgaagatgcccatcctccaccaaaaccaaatacaggtgtc
cttaaccctaccttcccacctgaaaaagcacctgcatctttggaaaaagtcgatacacat
gtcacatctactggtggtgaatcggatcctcctagtagtgaacatgaaagcgtttaa
```

000104

Protein sequence

>readseq.input(1), 439 bases, 6473FEF checksum.
MGELKEILKQRYHELLDWNVKAPHVPLSQRLKHFTWSWFACTMATGGVGL
IIGSFPFRFYGLNTIGKIVYILQIFLFSLFGSCMLFRFIKYPSTIKDSWN
HHLEKLFIATCLLSISTFIDMLAIYAYPDTGEWMVWVIRILYYIYVAVSF
IYCVMAFFTIFNNHVYTIETASPAWILPIFPPMICGVIAGAVNSTQPAHQ
LKNMVIFGILFQGLGFWVYLLLFAVNVLRFFTVGLAKPQDRPGMFMFVGP
PAFSGLALINIARGAMGSRPYIFVGANSSEYLGFVSTFMAIFIWGLAAWC
YCLAMVSFLAGFFTRAPLKFACGWFAFIFPNVGFVNCTIEIGKMIDSKAF
QMFGHIIGVILCIQWILLMYLMVRAFLVNDLCYPGKDEDAHPPPKPNTGV
LNPTFPPEKAPASLEKVDTHVTSTGGESDPPSSEHESV*

MleA

Nucleotide sequence integrated in ML01

>readseq.input(1), 1626 bases, 8F69675E checksum.
ttagtatttcggctcccacttcaaatcgtcgacggccttttcaacatcat
caatcggctggcgattaagaccttgttcaacagctttttagcgacagcg
acagcgactgtacgcgaaaagtcggctaattttgaaactggtggcaaaac
agcagcaccaacttttgttgtatcaacgattcctccaagactatgggcag
cagcagaaatcatttcaggcgtaagcagctttgattgagcggcaatggca
ccaaagccaagacctggatagatcaaggcattgttggcctgaccgatttg
gtaagcgttgccctcatattcaatatcgtcaactggaacaccagtaccga
tcaaggcttttccattagaccatttcaaaacatcttcggcttttgcttcg
gctaattttgttggattggaaattggaaaaatgattggtctttcagtata
accagacatatctttaacaatttcttcagtaaaggaatttggatgtgtcg
aggttccgaccaaaatggtcgggtggacagcttcaacagctgcttggaga
ttggtcaattgattagcatttttgaaatcacttcgtttagcagcgaaagg
cttttgctctggagttaaatccggatcatcatcaaacaagaggccttgtt
tgtcaacaaggaagaaatgctttttagcctcttcgtcggaaagaccctgt
tcaaccatttcttcatgcaactgtttaacaattcccattccagcagttcc
ggcaccgaagctcatgtaagtttgatcagttaatttctgaccggaaatct
tcaacgctccaagaacgccggcaagaacgacgattccagttccttgaata
tcatcattaaaagtagcaattttatctttatagctgtttaagatattaga
agcattcgaacggccaaaatcttcccaatgcaaatataaattaggaaaaa
gcgattcggcatgattaacaaatttgtcgataaaatcatagtacttatcg
ccacgaacacgattaaatttatttcccaaatacataggatctttcaaaag
cttttcgttatttgttccagcatcaataacaactgcaagaactgttgatg
gatcgattccggccgcaactgtataaaccatcagtttgccgacagcaata
tcaacaccctggacaccccagtctccaatcccaagaataccttcggcatc
agaaacgaccagcagcttgatatcgcggccattagcagcatttttcagag
tcgattgaatgttttccggatgattaatatccaaaaaagcggcaccttgc
ggttcaacaaataattccgaataattttcaattgtatcagcaattgtcgg
gtcataaacaattggcataaattcaacaacatgttgagaaaaaagcttat
aaaacaacacgtgattcgtattgaatatttccattaaaaacaatcgtttt
tcgagatttgaaaccttgctttgaaattgagcataagtctgatcaacttg
ctcttgtaaagcctgaaccttggccggtaataaaccgtttaaaccaagct
cctctctctccgcttccgtaaaagcagttcctttgttaataaaaggatca
tttaaaatacttactggatctgtcat

000105

Reverse complement of the integrated nucleotide sequence (coding sequence)

>READSEQ.INPUT(1), 1626 BASES, 8F69675E CHECKSUM. : Sens inverse-complémentaire de la séquence (1 - 1626)

```
atgacagatccagtaagtattttaaatgatccttttattaacaaaggaactgcttttacg
gaagcggagagagaggagcttggtttaaacggtttattaccggccaaggttcaggcttta
caagagcaagttgatcagacttatgctcaatttcaaagcaaggtttcaaatctcgaaaaa
cgattgtttttaatggaaatattcaatacgaatcacgtgttgttttataagctttttct
caacatgttgttgaatttatgccaattgtttatgacccgacaattgctgatacaattgaa
aattattcggaattatttgttgaaccgcaaggtgccgctttttggatattaatcatccg
gaaaacattcaatcgactctgaaaaatgctgctaatggccgcgatatcaagctgctggtc
gtttctgatgccgaaggtattcttgggattggagactggggtgtccagggtgttgatatt
gctgtcggcaaactgatggtttatacagttgcggccggaatcgatccatcaacagttctt
gcagttgttattgatgctggaacaaataacgaaaagcttttgaaagatcctatgtatttg
ggaaataaatttaatcgtgttcgtggcgataagtactatgattttatcgacaaatttgtt
aatcatgccgaatcgctttttcctaatttatatttgcattgggaagattttggccgttcg
aatgcttctaatatcttaaacagctataaagataaaattgctacttttaatgatgatatt
caaggaactggaatcgtcgttcttgccggcgttcttggagcgttgaagatttccggtcag
aaattaactgatcaaacttacatgagcttcggtgccggaactgctggaatgggaattgtt
aaacagttgcatgaagaaatggttgaacagggtctttccgacgaagaggctaaaaagcat
ttcttccttgttgacaaacaaggcctcttgtttgatgatgatccggatttaactccagag
caaaagcctttcgctgctaaacgaagtgatttcaaaaatgctaatcaattgaccaatctc
caagcagctgttgaagctgtccacccgaccattttggtcggaacctcgacacatccaaat
tcctttactgaagaaattgttaaagatatgtctggttatactgaaagaccaatcattttt
ccaatttccaatccaacaaaattagccgaagcaaaagccgaagatgttttgaaatggtct
aatggaaaagccttgatcggtactggtgttccagttgacgatattgaatatgagggcaac
gcttaccaaatcggtcaggccaacaatgccttgatctatccaggtcttggctttggtgcc
attgccgctcaatcaaagctgcttacgcctgaaatgatttctgctgctgcccatagtctt
ggaggaatcgttgatacaacaaaagttggtgctgctgttttgccaccagtttcaaaatta
gccgacttttcgcgtacagtcgctgtcgctgtcgctaaaaaagctgttgaacaaggtctt
aatcgccagccgattgatgatgttgaaaaggccgtcgacgatttgaagtgggagccgaaa
tactaa
```

Protein sequence

>readseq.input(1), 542 bases, E5EA13EB checksum.

```
MTDPVSILNDPFINKGTAFTEAEREELGLNGLLPAKVQALQEQVDQTYAQ
FQSKVSNLEKRLFLMEIFNTNHVLFYKLFSQHVVEFMPIVYDPTIADTIE
NYSELFVEPQGAAFLDINHPENIQSTLKNAANGRDIKLLVVSDAEGILGI
GDWGVQGVDIAVGKLMVYTVAAGIDPSTVLAVVIDAGTNNEKLLKDPMYL
GNKFNRVRGDKYYDFIDKFVNHAESLFPNLYLHWEDFGRSNASNILNSYK
DKIATFNDDIQGTGIVVLAGVLGALKISGQKLTDQTYMSFGAGTAGMGIV
KQLHEEMVEQGLSDEEAKKHFFLVDKQGLLFDDDPDLTPEQKPFAAKRSD
FKNANQLTNLQAAVEAVHPTILVGTSTHPNSFTEEIVKDMSGYTERPIIF
PISNPTKLAEAKAEDVLKWSNGKALIGTGVPVDDIEYEGNAYQIGQANNA
LIYPGLGFGAIAAQSKLLTPEMISAAAHSLGGIVDTTKVGAAVLPPVSKL
ADFSRTVAVAVAKKAVEQGLNRQPIDDVEKAVDDLKWEPKY*
```

000106

Appendix 4 : Analysis of global gene expression by DNA arrays carried out by the University of British Columbia

000107

ANALYSIS OF GLOBAL GENE EXPRESSION DATA

Comparison of *S. bayanus* strains MLO1 and S92 after 48 and 144 hours

Analysis of DNA array data

Results were analyzed using Affymetrix MASv5.0 software. All detection and comparison tunable parameters were set to default values (Affymetrix Statistical Algorithm Reference Guide, Affymetrix, Santa Clara, CA). Absolute analysis was completed on duplicate data for *S. bayanus* S92 and ML01 at each time point. Probe sets were included if the detection call (Absent or Present) was the same in the duplicate chips (P/P or A/A). The A/P calls are made by an algorithm that is based on a one-sided Wilcoxon signed rank-test. If present, the probe sets with a detection p-value of ≤ 0.01 were selected. Four comparative analyses were generated using *S. bayanus* S92 as baseline. Signal Log Ratios of either ≥ 1.0 or ≤ -1.0 in all four comparisons were included (corresponding to at least a two-fold change). Only probe sets that had a change call of I (Increase) and change p-value of ≤ 0.002 across all four comparisons were called 'increasers', and probe sets with D (Decrease) and p-value ≥ 0.998 were designated 'decreasers'. The change calls made by the software algorithm is based on a two-sided Wilcoxon signed rank-test. In the increasing data set, the *S. bayanus* ML01 probe sets were required to be present and the S92 probe sets were required to be either absent or present. In the decreasing data set, the *S. bayanus* S92 probe sets were required to be present and the ML01 probe sets were required to be either absent or present. In both groups, any probes with absent calls in both the baseline and experiment were discarded. Probe sets containing less than eight probe pairs were also discarded (based on the default probe pair threshold).

000108

Probe sets that corresponded to transposons, non-annotated SAGE orfs, and probe sets that corresponded to hypothetical orfs without an assigned name, were deleted from the data set. Finally, gene names and descriptions from SGD were linked to corresponding public accession numbers (Cherry *et al.*, 1997).

Faculty of Agricultural Sciences, University of British Columbia
Suite 230, 2205 East Mall, Vancouver, BC, V6T 1Z4 Canada
tel: 604.822.0418, fax: 604.822.5143, www.agsci.ubc.ca/wine

Results and Discussion

Transcription of only one gene (*HSP26*) was down-regulated in the malolactic yeast ML01 at 48 hours (Table 1). This gene encodes for a heat shock protein. No genes were up-regulated in *S. bayanus* ML01 after 48 hrs.

Table 1: Gene Down Regulated in *S. bayanus* ML01 at 48 Hours (≥ 2-fold change, High Stringency Analysis)

Gene Name (SGD)	Fold Change	Description
HSP26	-2.589	YBR072W heat shock protein 26

The transcription of 14 genes was affected in the genetically engineered yeast after 144 hrs of fermentation; six genes were up-regulated (Table 2) and eight genes were down-regulated (Table 3).

Table 2: Genes that were up-regulated in *S. bayanus* ML01 at 144 hrs (≥ 2-fold change, High Stringency Analysis)

Gene Name (SGD)	Fold Change	Description
MRT4	5.766	YKL009W mRNA turnover 4
ZPR1	4.438	YGR211W zinc finger protein
BTN2	4.287	YGR142W similarity to hypothetical protein YPR158w
NIP7	2.848	YPL211W Nip7p is required for 60S ribosome subunit biogenesis
RPS9B	2.727	YBR189W Ribosomal protein S9B (S13) (rp21) (YS11)
SSE1	2.329	YPL106C HSP70 family member, highly homologous to Ssa1p and Sse2p

The *MTR4* gene is involved with mRNA turnover. Transcription of this gene was substantially up-regulated (5.7-fold). At this stage we can only speculate why this gene was up-regulated. The *ZPR1* gene encodes a zinc finger transcription factor (Pol1). Loss of *ZPR1* causes disruption of nucleolar function. *BTN2* encodes for a hypothetical protein possibly involved in regulation of the intracellular pH. Deletion of *BTN1* in *S. cerevisiae* diminishes pH buffering capacity and causes poor growth at low pH. The up-regulation of this gene might well compensate for a change in the intracellular pH of the malolactic yeast due to the intracellular presence of malate and/or lactate in the malolactic yeast. *NIP7* and *RPS9B* encode proteins involved with ribosome biogenesis. We have noted in all of our arrays done to date (more than

50 chips) that expression of the RPS genes is extremely sensitive to changes in the medium, i.e. osmotic stress and addition of nutrients. The presence of malate or lactate in the yeast cell, or the accompanying change in pH due the decarboxylation of malate to lactate by the malolactic yeast, might be responsible for a change in the expression level of these two genes. The *SSE1* gene is a member of the heat shock family proteins. This protein is a cochaperone and is involved with protein folding. In our array analyses to date, we have also noted that genes which encode heat shock proteins are sensitive to changes in the medium.

Table 3: Genes that were down-regulated in *S. bayanus* ML01 after 144 hours (≥ 2-fold change, High Stringency Analysis)

Gene Name (SGD)	Fold Change	Description
TIS11	-3.393	YLR137W hypothetical protein
SPS100	-2.819	YHR139C sporulation-specific wall maturation protein
RNR3	-2.621	YIL066C Ribonucleotide reductase (ribonucleoside-diphosphate reductase) large subunit
GRE1	-2.594	YPL223C Induced by osmotic stress
ALD3	-2.527	YMR169C Aldehyde Dehydrogenase (NAD(P)+)
ADE17	-2.479	YMR120C 5-aminoimidazole-4-carboxamide ribonucleotide (AICAR) transformylase/IMP cyclohydrolase
SNR11	-2.329	SNR11 snRNA
PFK2	-2.038	YMR205C phosphofructokinase beta subunit

Transcription of *TIS11* was down-regulated 3.4-fold in the malolactic yeast. This gene encodes for a protein of unknown function. The null mutant is viable but alters metabolism that is reflected by a pH change on YPD plates. Once again, pH seems to be an important factor. The *SPS100* gene is similar to the *YGP1* gene that encodes for a glycoprotein that is synthesised in response to nutrient limitation. We cannot at this stage offer any idea as to why transcription of this gene would be affected. The *RNR3* gene encodes a ribonucleoside-diphosphate reductase involved in DNA replication. Growth of the malolactic yeast was not affected compared to the control yeast. The *GRE1* gene encodes a protein with unknown function. It has previously been established that transcription of this gene is induced by osmotic stress. The *ALD3* gene codes for aldehyde dehydrogenase; expression of this gene is induced by heat shock, oxidative and osmotic stress. The *ADE17* gene encodes for phosphoribosylaminoimidazole-carboxamide formyltransferase that catalyzes the penultimate step in the *de novo* biosynthesis of IMP. The

SNR11 and *PFK2* genes encode for an enzyme involved in rRNA modification and 6-phosfofructokinase, respectively. Apart from the *TIS11* gene that might be involved with pH regulation in the yeast cell, down-regulation of the *SPS100, RNR3, GRE1, ALD3, ADE17, SNR11* and *PFK2* genes in the malolactic yeast cannot be explained at this stage.

Conclusions

It is clear that the introduction of the malolactic gene cassette into *S. cerevisiae* S92 had a minimal effect on the transcription of the more than 6200 genes in the yeast cell. The transcription of a single gene was affected after 48 hours of fermentation. After 144 hrs of fermentation, the transcription of only 14 genes was affected. A few genes encoding enzymes involved with ribosome biogenesis were affected by the introduction of the malolactic cassette. However, transcription of these genes has been shown to be sensitive to changes in the medium. In the presence of malate and a fermentable carbon source, the malolactic yeast will transport malate into the cell and decarboxylate malate to lactate. This acummulation of acids in the yeast cell might be reponsible for up-regulation of the *BTN1* gene in *S. cerevisiae* which regulates pH buffering capacity of the cell.

From the array data we can conclude that the introduction of the malolactic gene cassette did not alter gene expression patterns in the malolactic yeast ML01 significantly. *S. bayanus* ML01 can be regarded as substantially equivalent to the wild type strain *S. bayanus* S92.

Reference

Cherry JM, Ball C, Weng S, Juvik G, Schmidt R, Adler C, Dunn B, Dwight S, Riles L, Mortimer RK, Botstein D. (1997). Genetic and physical maps of *Saccharomyces cerevisiae.* Nature 387(6632 Suppl):67-73.

Hennie J.J. van Vuuren (Ph.D.)
Professor and Eagles Chair in Food Biotechnology
Director

June 13, 2002

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

000112

Nucleic sequences given here contain 3 bp more than the ones described in Figure 24 as they include the stop codon which was not the case in Figure 24.

ORF1

Nucleotide sequence

```
>readseq.input(1), 435 bases, BDF920E1 checksum.
atgtcgaaagctacatataaggaacgtgctgctactcatcctagtcctgt
tgctgccaagctatttaatatcatgcacgaaaagcaaacaaacttgtgtg
cttcattggatgttcgtaccaccaaggaattactggagttagttgaagca
ttaggtcccaaaatttgtttactaaaaacacatgtggatatcttgactga
tttttccatggagggcacagttaagccgctaaaggcattatccgccaagt
acaattttttactcttcgaagacagaaaatttgctgacattggtaataca
gtcaaattgcagcactctgcgggtgtatacagaatagcagaatgggcaga
cattacgaatgcacacggtgtggtgggcccaggtattgttagcggtttga
agcaggcggcggaagaagtaggtaccaagctttaa
```

Putative protein sequence

```
>readseq.input(1), 145 bases, A0D537AF checksum.
MSKATYKERAATHPSPVAAKLFNIMHEKQTNLCASLDVRTTKELLELVEA
LGPKICLLKTHVDILTDFSMEGTVKPLKALSAKYNFLLFEDRKFADIGNT
VKLQHSAGVYRIAEWADITNAHGVVGPGIVSGLKQAAEEVGTKL*
```

ORF2

Nucleotide sequence

```
>readseq.input(1), 123 bases, A340F1C6 checksum.
ttattttagcgtaaaggatggggaaagagaaaagaaaaaaaattgatctat
cgcagatccctcgaggagatctcatatgcaagacgcatacatatttaaca
ctacgacaaccaaaaattaccat
```

Putative protein sequence

```
>readseq.input(1), 41 bases, 1AE15B08 checksum.
MVIFGCRSVKYVCVLHMRSPRGICDRSIFFFSLSPSFTLK*
```

ORF3

Nucleotide sequence

```
>readseq.input(1), 138 bases, 4EB2EB4A checksum.
atgcttcagtcgttgactgagagggacatgaggggctttgacattccagt
caagcaactcatgatacctctgtttcaagatttccttgagttcacccatg
tcgaatcgagagataaaggggaattaaaaaaaaaatag
```

000113

Putative protein sequence

>readseq.input(1), 46 bases, B9403F35 checksum.
MLQSLTERDMRGFDIPVKQLMIPLFQDFLEFTHVESRDKGELKKK*

ORF4

Nucleotide sequence

>readseq.input(1), 663 bases, EE97C39A checksum.
atgggaagtcatcgacgttatctctactatagtatattatcatttctatt
attatcctgctcagtggtacttgcaaaacaagataagaccccattctttg
aaggtacttcttcgaaaaattcgcgtctaactgcacaagataagggcaat
gatac**atg**cccgccatgttttaattgtatgctacctatttttgaatgcaa
acagttttctgaatgcaattcgtacactggtagatgtgagtgtatagaag
ggtttgcaggtgatgattgctctctgcccctctgtggcggtctatcaccg
gatgaaagcggtaataaggatcgtcccataagagcacaaaatgacacctg
tcattgtgataacggatggggagggatcaattgtgacgtttgtcaagaag
attttgtctgtgatgcgttcatgcctgatcctagtattaaggggacatgt
tataagaatggtatgattgtagataaagtattttcaggttgtaatgtgac
caatgagaaaattctacagattttgaacggcaaaataccacaaattacat
ttgcctgtgataaacctaatcaagaatgtaattttcagttttggacagat
cagttagaaagcttggtaccgcggccgcaagctttaacgaacgcagaatt
ttcgagttattaa

Putative protein sequence

>readseq.input(1), 221 bases, 221AC14B checksum.
MGSHRRYLYYSILSFLLLSCSVVLAKQDKTPFFEGTSSKNSRLTAQDKGN
DTCPPCFNCMLPIFECKQFSECNSYTGRCECIEGFAGDDCSLPLCGGLSP
DESGNKDRPIRAQNDTCHCDNGWGGINCDVCQEDFVCDAFMPDPSIKGTC
YKNGMIVDKVFSGCNVTNEKILQILNGKIPQITFACDKPNQECNFQFWTD
QLESLVPRPQALTNAEFSSY*

ORF5

Nucleotide sequence

>readseq.input(1), 186 bases, 1ED040F8 checksum.
atggggaaagagaaaagaaaaaaattgatctatcgcagatccctcgagga
gatctcattagtatttcggctcccacttcaaatcgtcgacggccttttca
acatcatcaatcggctggcgattaagaccttgttcaacagctttttagc
gacagcgacagcgactgtacgcgaaaagtcggctaa

Putative protein sequence

>readseq.input(1), 62 bases, 97E557D4 checksum.
MGKEKRKKLIYRRSLEEISLVFRLPLQIVDGLFNIINRLAIKTLFNSFFS
DSDSDCTRKVG*

000114

ORF6

Nucleotide sequence

```
>readseq.input(1), 636 bases, D7829915 checksum.
atgggaagtcatcgacgttatctctactatagtatattatcatttctatt
attatcctgctcagtggtacttgcaaaacaagataagaccccattctttg
aaggtacttcttcgaaaaattcgcgtctaactgcacaagataagggcaat
gatacatgcccgcc atgttttaattgtatgctacctatttttgaatgcaa
acagttttctgaatgcaattcgtacactggtagatgtgagtgtatagaag
ggtttgcaggtgatgattgctctctgcccctctgtggcggtctatcaccg
gatgaaagcggtaataaggatcgtcccataagagcacaaaatgacacctg
tcattgtgataacggatggggagggatcaattgtgacgtttgtcaagaag
attttgtctgtgatgcgttcatgcctgatcctagtattaaggggacatgt
tataagaatggtatgattgtagataaagtattttcaggttgtaatgtgac
caatgagaaaattctacagattttgaacggcaaaataccacaaattacat
ttgcctgtgataaacctaatcaagaatgtaattttcagttttggatagat
cagttagaaagcttgcggccgcacgaaggaacctag
```

Putative protein sequence

```
>readseq.input(1), 212 bases, FDF54FB2 checksum.
MGSHRRYLYYSILSFLLLSCSVVLAKQDKTPFFEGTSSKNSRLTAQDKGN
DTCPPCFNCMLPIFECKQFSECNSYTGRCECIEGFAGDDCSLPLCGGLSP
DESGNKDRPIRAQNDTCHCDNGWGGINCDVCQEDFVCDAFMPDPSIKGTC
YKNGMIVDKVFSGCNVTNEKILQILNGKIPQITFACDKPNQECNFQFWID
QLESLRPHEGT*
```

ORF7

Nucleotide sequence

```
>readseq.input(1), 111 bases, D0609A9D checksum.
ttagaaagcttgcggccgcacgaaggaacctagaggccttttgatgttag
cagaattgtcatgcaagggctccctagctactggagaatatactaagggt
actgttgacat
```

Putative protein sequence

```
>readseq.input(1), 37 bases, 568F21F0 checksum.
MSTVPLVYSPVAREPLHDNSANIKRPLGSFVRPQAF*
```

000115

Appendix 6 : Synthetic grape must fermentations by mixed cultures of the modified ML01 strain and the host strain S92

000116

First set of experiments : I- Materials and methods

Active dry yeast preparation : Active dry yeast was rehydrated in sterile water at room temperature.

Media : Sterile flask fitted with vapour locks and containing 250 milliliters of filter-sterilized synthetic juice (100 grams/liter D-glucose, 100 grams/liter D-Fructose, 6.0 grams/liter tartaric acid, 4.5 grams/liter L-malic acid, 0.3 gram/liter citric acid, 2.0 grams/liter $(NH_4)_2SO_4$, 1.7 grams/liter YNB without amino acids or ammonium sulphate, 1.0 milliliter/liter Tween 80, 5.0 milligrams/liter oleic acid, pH adjusted to 3.5 with KOH) were prepared.

Inoculation : each fermentation was inoculated with rehydrated active dry yeast strains S92, ML01 or a combination of both. Test fermentations were inoculated as follows :

75 milligrams/liter S92 and 75 milligrams/liter ML01 (50 % ML01 per total inoculum)

135 milligrams/liter S92 and 15 milligrams/liter ML01 (10 % ML01 per total inoculum)

148.5 milligrams/liter S92 and 1.5 milligrams/liter ML01 (1 % ML01 per total inoculum)

149.85 milligrams/liter S92 and 0.15 milligrams/liter ML01 (0.1 % ML01 per total inoculum)

149.985 milligrams/liter S92 and 0.015 milligrams/liter ML01 (0.01 % ML01 per total inoculum)

150 milligrams/liter S92 or ML01 (control fermentations).

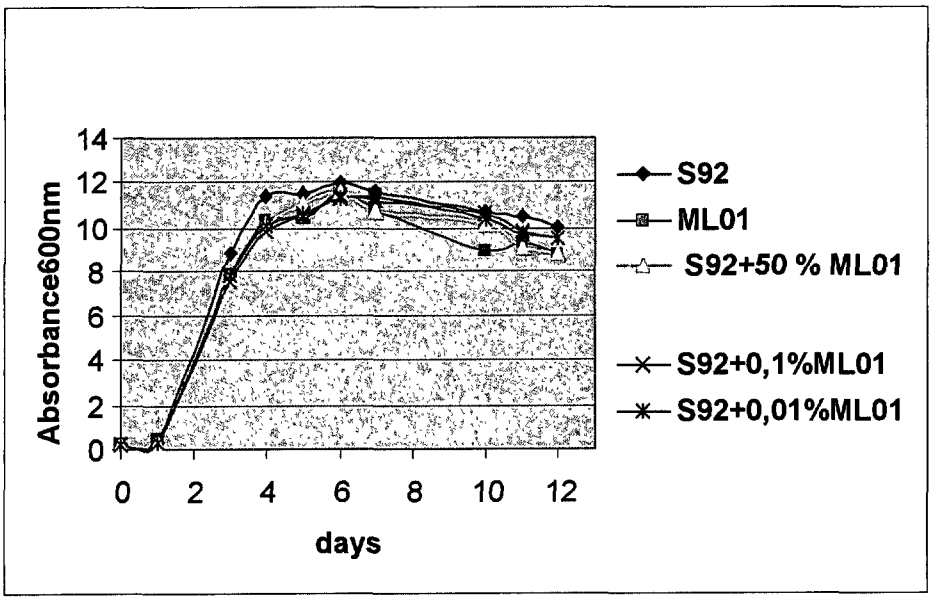
Fermentation conditions : the flasks were incubated at 21 °C and gently shaken at 100 rpm.

Monitoring : Growth was monitored by measuring Absorbance 600nm at the time points indicated. At 0, 6 and 12 or 13 days after inoculation, samples from each flask were taken, diluted and spread onto rich media plates and incubated at 30 °C until colonies appeared.

Chemical analyses : L-malate and L-lactate concentrations were estimated by enzymatic analysis.

Colony analysis : Colonies obtained from each fermentation condition were randomly picked and inoculated into microwell plates containing 200 microliters of synthetic juice and incubated at 30°C for 7 days. After incubation, analysis for the presence of lactate were performed.

First set of experiments : II- Fermentations with 50 % or 0.1 % or 0.01 % ML01 per total inoculum

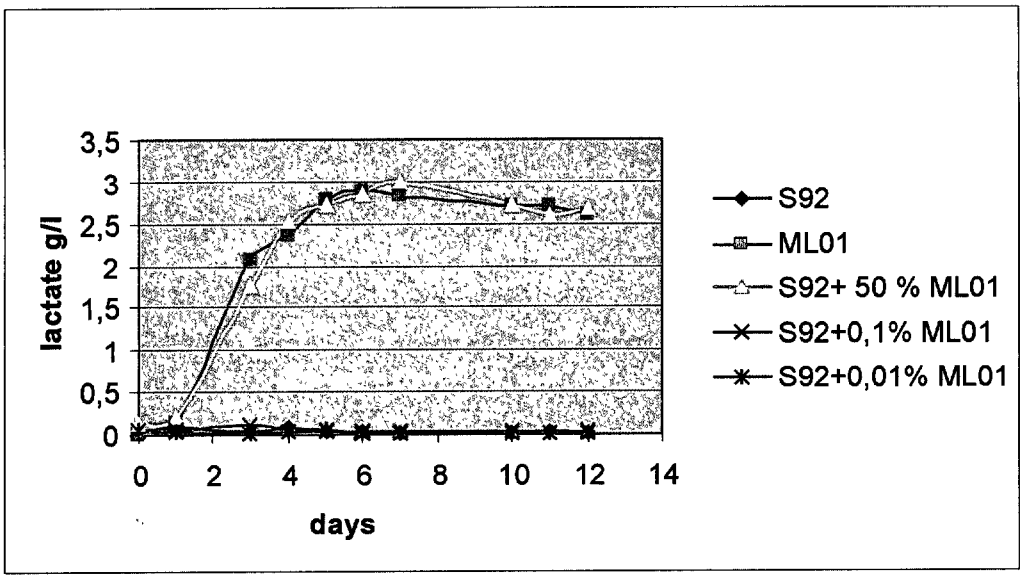


	% of ML01 colonies		
% of ML01 inoculation per total yeast cells	**Day 0**	**Day 6**	**Day 12**
50% ML01	48 %	53 %	31 %
0.1 % ML01	-	-	0 %*
0.01 % ML01	-	-	0 %*

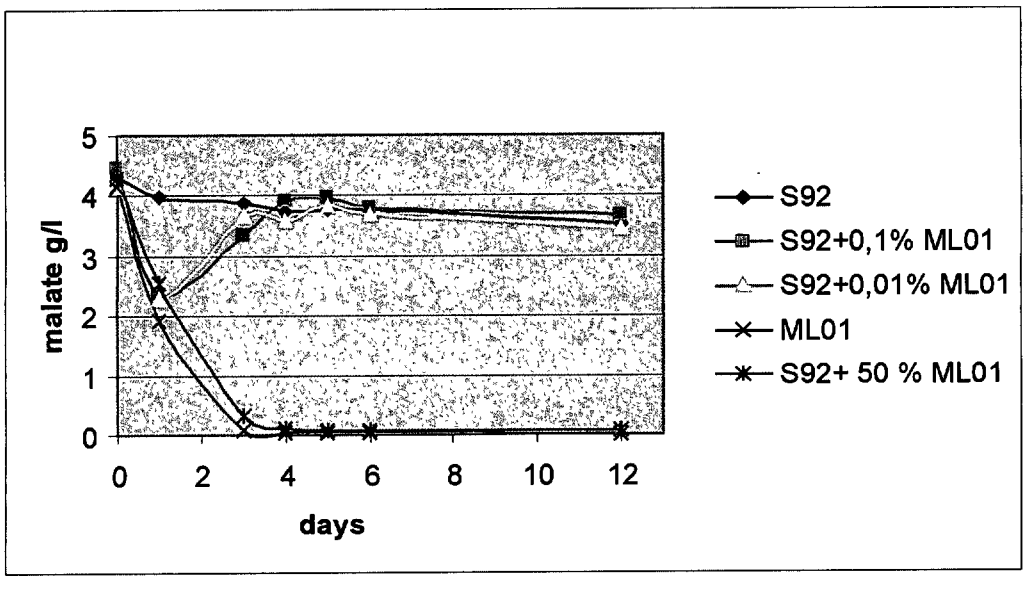
* The number of colonies analyzed does not permit the estimation of ML01 colonies.

000118

First set of experiments : II- Fermentations with 50 % or 0.1 % or 0.01 % ML01 per total inoculum (continued)



There is no lactate production with S92, S92 + 0.1 % ML01, and S92 + 0.01 % ML01

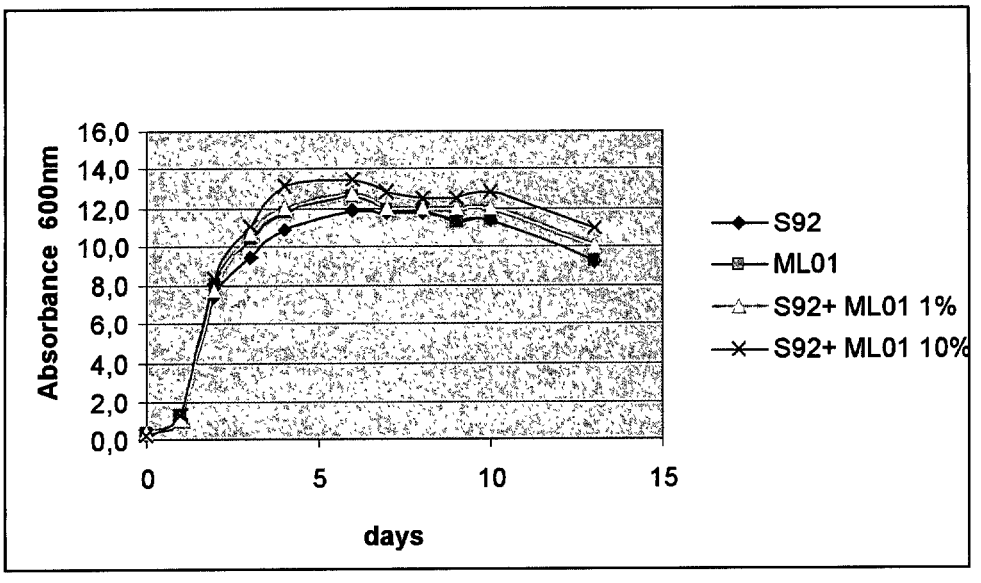


Warning : graphic representation codes have been changed for this graphic.

Note : the graphical points at one day after the beginning of fermentation for S92 + 0.1 % ML01 and S92 + 0.01 % ML01 are aberrant.

000119

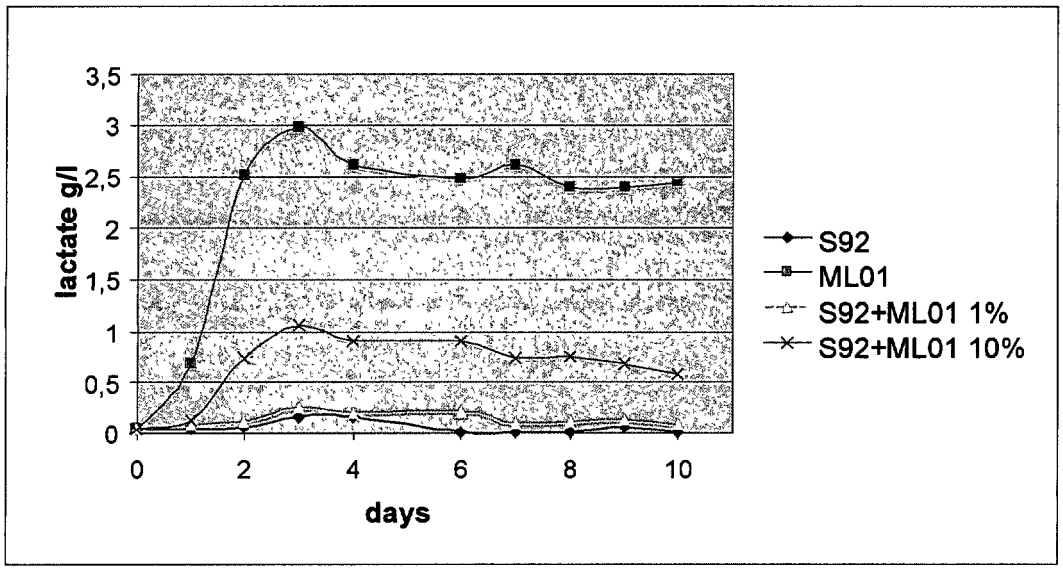
First set of experiments : III- Fermentations with 1 % or 10 % ML01 per total inoculum



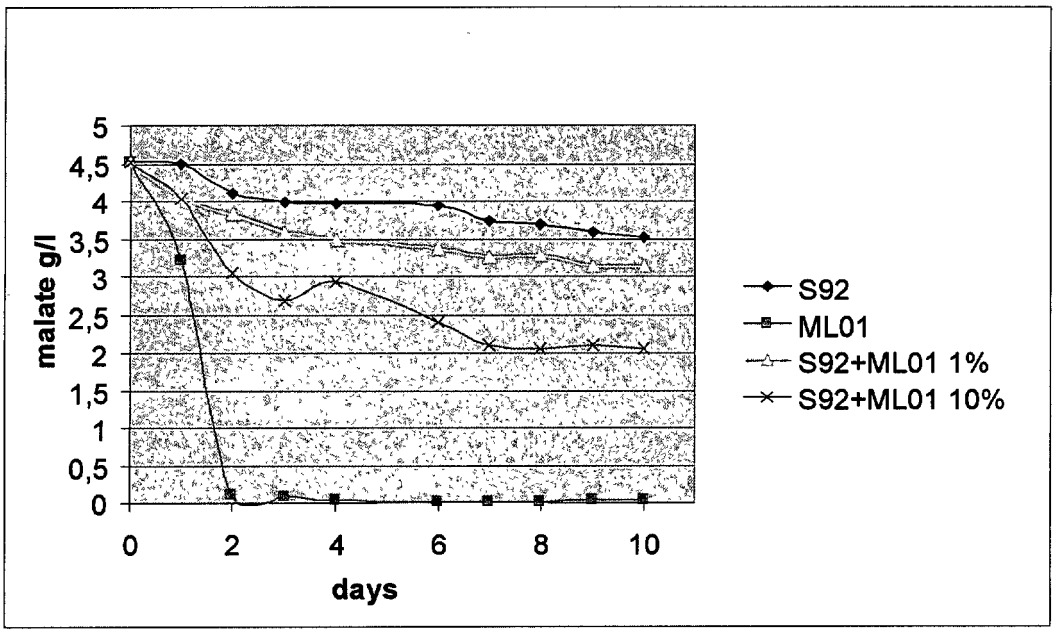
	% of ML01 colonies		
% of ML01 inoculation per total yeast cells	Day 0	Day 6	Day 13
1 % ML01	0%	0 %	1 %
10 % ML01	24 %	9 %	4 %

000120

First set of experiments : III- Fermentations with 1 % or 10 % ML01 per total inoculum (continued)



There is practically no production of lactate with S92 and S92 + ML01 1 %
With S92 + ML01 10 %, the production of lactate stops after 3 days, the first 4 days corresponding to the exponential growth phase.



There is no significant difference between S92 and S92 + ML01 1 % concerning malate degradation.
The curve of malate degradation with S92 + ML01 10 % is difficult to interpret because the production of lactate stops after 3 days. In any case, malate degradation remains very incomplete.

000121

Second set of experiments :I- Materials and Methods

Active Dry Yeast Preparation: One gram of active dry yeast of the host strain strain S92 and one gram of active dry yeast of strain ML01 were each rehydrated in 20 milliliters of sterile distilled water at 37 °C for 15 minutes.

Media: Twenty-two sterile flasks of 500 milliliters fitted with vapour locks and containing 400 milliliters of filter-sterilised synthetic juice (100 grams/liter D-glucose, 100 grams/lilter D-Fructose, 6.0 grams/liter tartaric acid, 4.5 grams/liter L-malic acid, 0.3 gram/liter citric acid, 2.0 grams/liter $(NH_4)_2SO_4$, 1.7 grams/liter YNB without amino acids or ammonium sulphate, 1.0 milliliter/liter Tween 80, 5.0 milligrams/liter oleic acid, pH adjusted to 3.5 with KOH) were prepared.

Inoculation: Each fermentation flask was inoculated with rehydrated ADY strains: S92, ML01, or a combination of S92 and ML01. Test fermentations were inoculated as follows in duplicate:

50 milligrams/liter S92 and 50 milligrams/liter ML01 (50 % ML01 per total inoculum).

50 milligrams/liter S92 and 5 milligrams/liter ML01 (10 % ML01 per total inoculum).

50 milligrams/liter S92 and 0.5 milligrams/liter ML01 (1 % ML01 per total inoculum).

50 milligrams/liter S92 and 0.05 milligrams/liter ML01 (0.1 % ML01 per total inoculum).

50 milligrams/liter S92 and 0.005 milligrams/liter ML01 (0.01 % ML01 per total inoculum).

100 milligrams/liter or 55 milligrams/liter, or 50 milligrams/liter of strain S92 or strain ML01 (control fermentations).

Fermentation Conditions: The flasks were incubated at 20°C (no shaking).

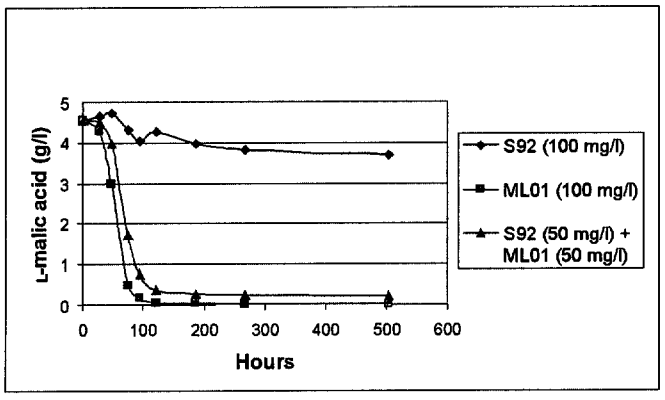
Monitoring: Growth was monitored by measuring Absorbance 600nm at the time points indicated. Two samples from each flask per time point were measured and averaged. After optical density measurements, individual samples were centrifuged, and the supernatant frozen in clean sterile tubes.
At 2 hours (Day 0), 187 hours (Day 8) and 504 hours (Day 21) samples from each flask were taken, serially diluted and spread onto rich medium plates and incubated at 30°C until colonies appeared.

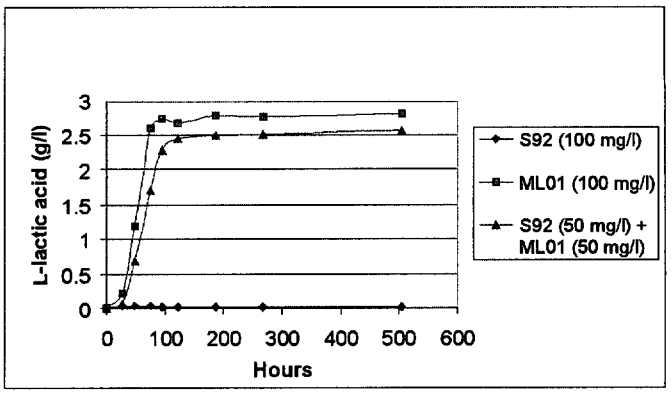
Chemical analyses: L-malic and L-lactic acid concentrations were obtained by enzymatic analysis.

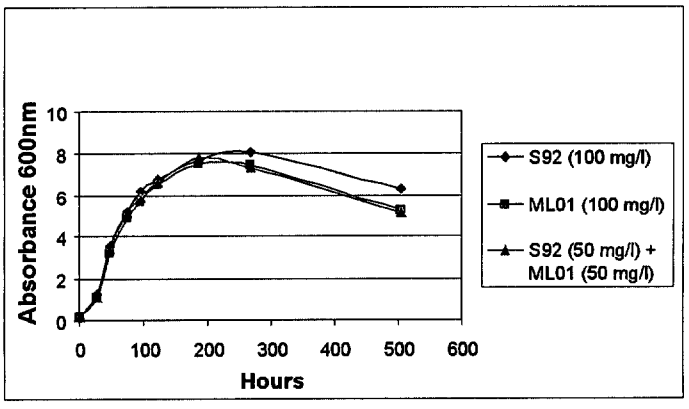
Colony analysis: Colonies obtained from each flask were randomly picked and inoculated into microwell plates containing 200 microliters of synthetic juice and incubated at 30°C for 7 days. Two negative (S92) and 2 positive controls (ML01) were also inoculated. After incubation, analysis for the presence of lactate was performed.

<u>Second set of experiments</u> : II- Fermentations performed with 50 % of ML01 yeast cells per total inoculum

Results represent the mean of two experiments.





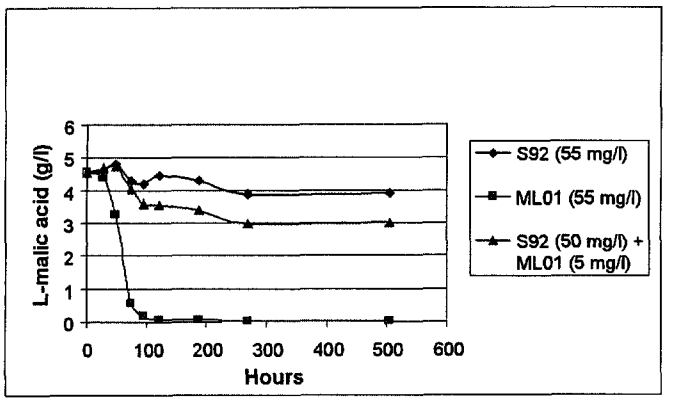


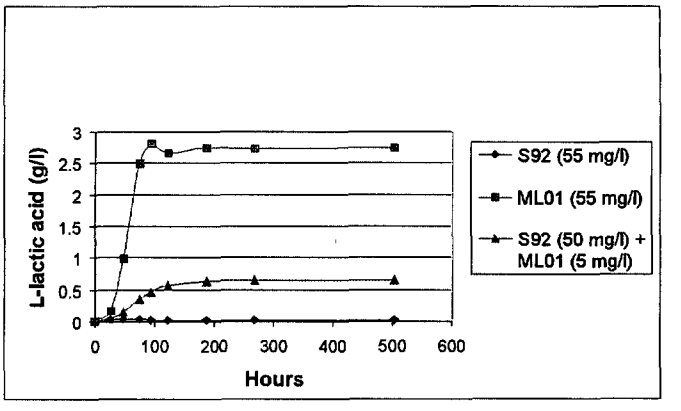
	% of ML01 colonies		
Yeast strain	**Day 0**	**Day 8**	**Day 21**
S92	0	-	0
ML01	98.44	-	100
50 % S92 50 % ML01	54.35	44.47	38.40

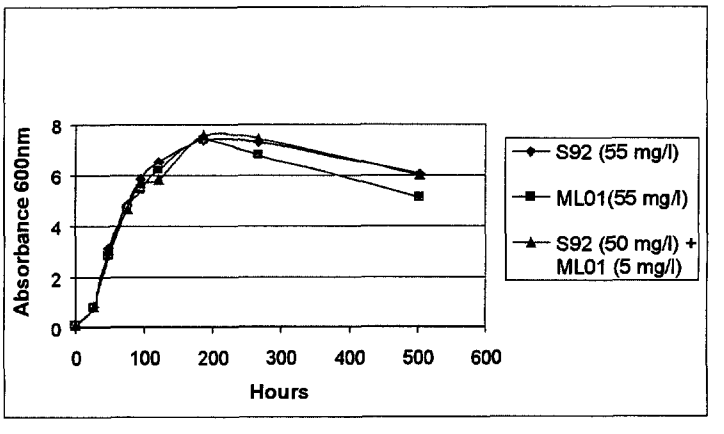
000123

<u>Second set of experiments</u> : III- Fermentations performed with 10 % of ML01 yeast cells per total inoculum

Results represent the mean of two experiments.





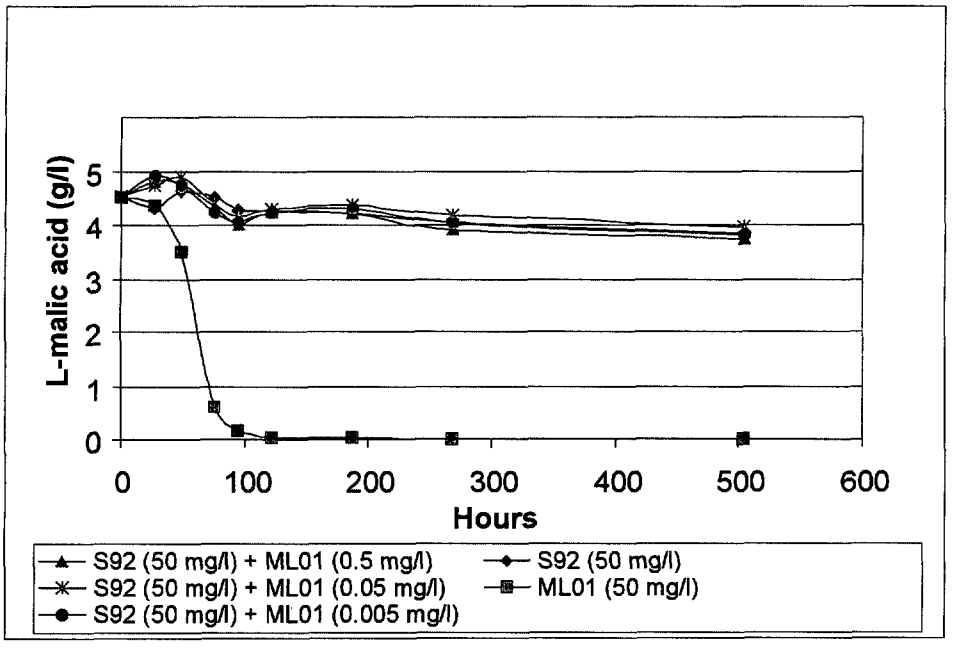


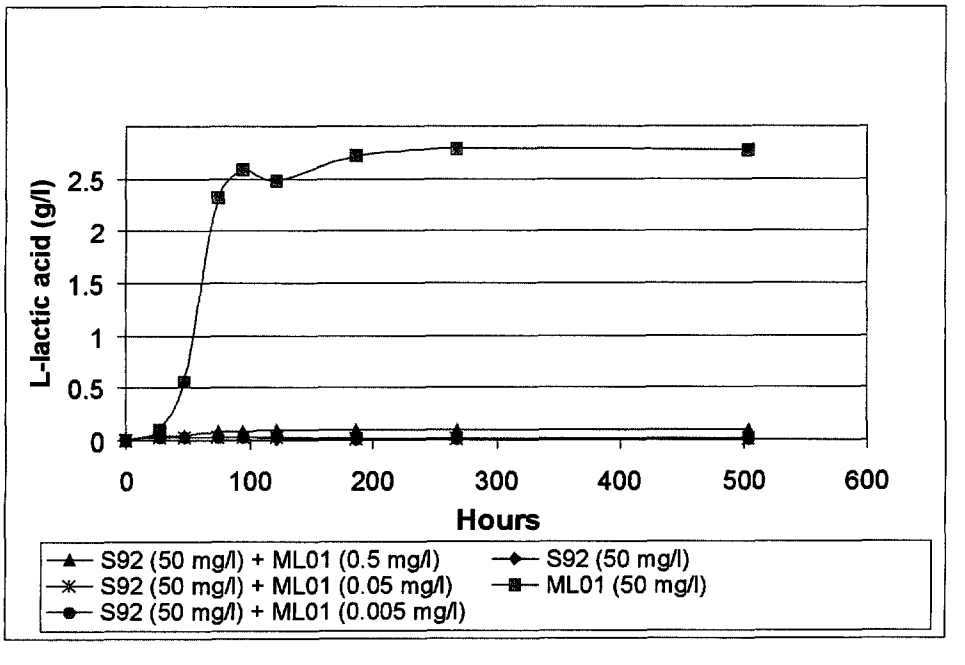
	% of ML01 colonies		
Yeast strain	**Day 0**	**Day 8**	**Day 21**
S92	0	-	0
ML01	100	-	100
90 % S92 10 % ML01	6.52	3.26	2.17

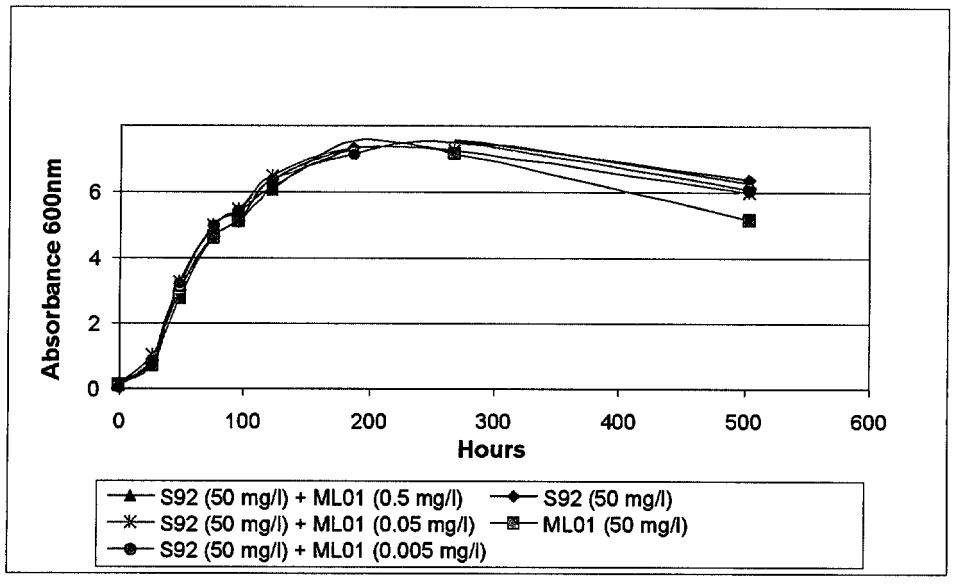
000124

<u>Second set of experiments</u> : IV- Fermentations performed with 1 % , 0.1 % and 0.01 % of ML01 yeast cells per total inoculum

Results represent the mean of two experiments.







	% of ML01 colonies		
Yeast strain	**Day 0**	**Day 8**	**Day 21**
S92	0	-	0
ML01	100	-	100
99 % S92 1 % ML01	1.09	1.09	1.09
99.9 % S92 0.1 % ML01	0*	0*	0*
99.99% S92 0.01 % ML01	0*	0*	0*

* the number of colonies analysed does not permit the estimation the percentage of ML01 colonies.

000126

Second set of experiments : V- Conclusions

As shown previously, when the ML01 strain represents a small proportion of the total yeast inoculum, only a reduced amount of malate is consumed. Malate degradation becomes negligible when ML01 represents 1 % or less of the total yeast inoculum.

Colony counts show that in each case, the ML01 yeast strain never takes over the fermentation. The presence of malate in the media and its degradation do not represent a selective advantage for the ML01 strain. This confirms as it was predictable, that malate and its subsequent degradation into lactate through malolactic fermentation do not serve as carbon or energy sources for the recombinant strain.

Colony counts show that the proportion of ML01 cells at day 21 (504 hours) is below that at day 0. This indicates that the ML01 strain by no means takes over the fermentation when in co-culture with the S92 host strain, and even has a tendency to die faster than the host strain. These results are reinforced by the Absorbance readings at 600 nm which show an accentuated decrease of yeast cell concentration in ML01 fermentations compared to S92 fermentations.

Malolactic fermentation only takes place during the yeast exponential growth phase. Cells need to be actively dividing to be able to degrade malate.
At ML01 cells concentration below 1 % of the total inoculum, no malate degradation will occur. Hence, there will be no unwanted yeast malolactic fermentation when the ML01 strain represents less than 1 % of the total yeast inoculum, which is a high level of contamination.
When the yeast population is a mix of the ML01 strain and the S92 strain, the ML01 yeast cells in no case take over the fermentation.

End Submission

AM

Ricker, Karin

From: Bob.Biwersi@lesaffreyeastcorp.com
Sent: Monday, June 16, 2003 8:33 AM
To: Ricker, Karin
Subject: Active Dry Yeast Process Description

pic19169.pcx (12 KB)

Yeast Cultures (Laboratory)

Yeast cultures are maintained frozen in glycerol for many years. When needed for production, the yeast is transferred to agar slants that contain sugar and nutrients. The yeast growth on the slants in anaerobic

Pasteur Flasks (Laboratory)

A media of molasses and nutrients is prepared and autoclaved. Once cooled, the media is inoculated with the yeast from the culture slant. The flask is incubated at optimal yeast growth temperature for a specific period time. This is an anaerobic fermentation designed to produce approx 100 grams of yeast solids.

Pure Culture Fermentation (Production)

The pure culture fermentation is a set batch fermentation. This means that all growth media and nutrients are added to the fermenter, sterilized and cooled prior to being inoculated with the yeast from the Pasteur flask(s). This fermentation is an aerobic air is bubbled through the fermenter. The aerobic process is used to force the yeast to propagate instead of going into an alcohol production mode. The set batch process is used to minimize any possible bacterial contamination at this stage of the yeast production process.

Stock Fermentation

The stock fermentation is the first of a series of fermentations that are considered "fed-batch" fermentations. In a fed batch fermentation, water is added to the fermentation vessel along with required vitamins and minerals. The yeast from the pure culture fermentation is then added to the fermenter as "seed". The stock fermentation is an aerobic fermentation, where large volumes of air are used to force yeast replication and minimize alcohol production. The molasses, used as the carbohydrate source, is added to the fermenter over a period of 12-15 hrs. The amount of molasses added is controlled by measuring the alcohol level in the exhaust gas. If the alcohol in the exhaust gas goes above a predetermined setpoint, the molasses feed will be reduced. If it drops too low, the molasses feed will be increased.

Separation

At the end of a set amount of time, the molasses feed and air will be discontinued and the contents of the fermenter pumped to centrifugal separators. The separators are used to separate the yeast solids form the fermenter liquid (Unfermentable molasses residual and water). The process includes a direct pass of the fermenter contents through the separators, diluting the yeast solids with water (washing), and running the diluted yeast through the separators a second time. The washing process provides a light tank cream yeast product (Approx 18% solids).

000129

First Generation and Trade Fermentations

These fermentations are similar to the stock fermentation. The yeast from the

stock fermentation can be used to seed 3-5 "1st generation" fermentations. Each of the 1st generation fermentation can then be used to seed 7-9 "Trade" fermentation. The series of fermentations is produced to provide a large quantity of yeast from the original culture slants.

NOTE: Wine yeast products are usually produced on a much smaller scale than bakers dry yeast products. Yeast from the stock fermentation is often used in the drying process.

Dewatering

Prior to drying, the yeast cream (18%) solids is stored in refrigerated tanks at < 40 F. This cream yeast is processed using a Rotary Vacuum Filter, where vacuum draws water from the yeast as it is sprayed on a rotating drum of starch. The dewatered cream yeast is then scraped off the drum, and conveyed to an extruder (26-27% solids, "cake yeast".

Extruding

The cake yeast is forced through a noodle plate producing fine strands of yeast (resembles angel hair pasta). This process is performed to provide more surface area to the high moisture yeast cake. As the strands of yeast exit the extruder, they are cut in to short sections. The yeast is extruded directly into the airlift dryer.

Drying

The dryer uses filtered and heated air to gradually dry the yeast to a moisture content of approximately 5%. Drying yeast and maintaining activity requires gentle drying conditions that do not damage the yeast. Once dried, the yeast is conveyed to packaging hoppers.

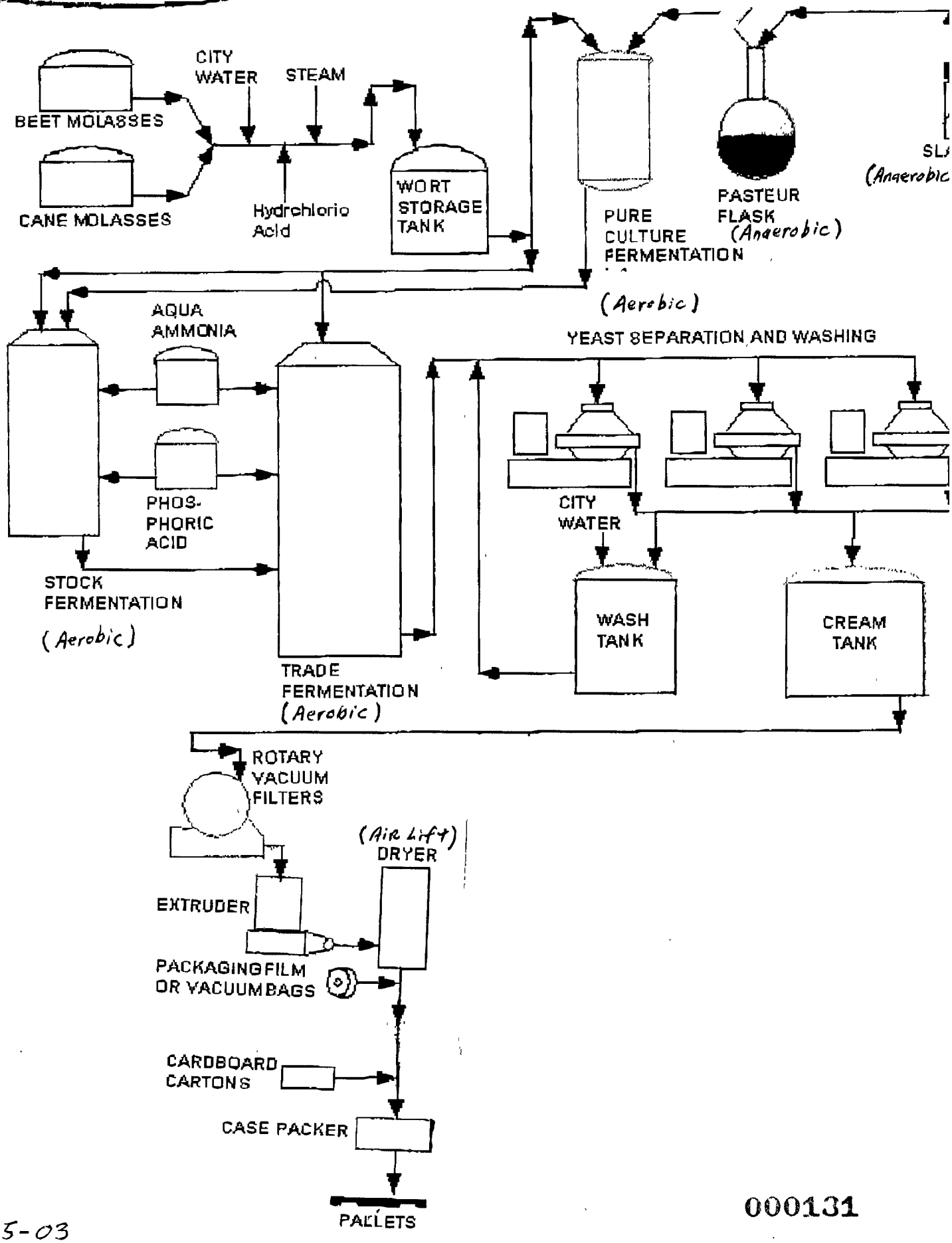
Packaging

Active dry wine yeast is packed in vacuum foil pouches, which are then packed in corrugated boxes for shipment. The vacuum and foil are required to eliminate exposure to oxygen that can shorten the shelf-life of the yeast. Active dry wine yeast have a shelf-life of 24 months @ 70 F.

[IMAGE](Embedded image moved to file: pic19169.pcx)

000130

ACTIVE DRY YEAST PROCESS



000131

6-5-03

AM

Ricker, Karin

From: Bob.Biwersi@lesaffreyeastcorp.com
Sent: Monday, June 16, 2003 9:47 AM
To: Ricker, Karin
Subject: MLO1 References - Wine Making

pic26589.pcx (12 KB)

I have found the following references that should be useful in understanding the wine making process. The "About Wine" website is especially useful in understanding the Malo-lactic fermentation that would be eliminated by using the ML01 wine yeast strain.

Book Reference:
Yeast Technology - second edition, 1991: Van Nostrand Reinhold Publisher, Authors Reed and Nagodawithana Chapter 4

Web Site References:
www.wine.about.com/library/bl_making.htm

www.geocities.com/napavalley/4908/winemaking.htm

[IMAGE](Embedded image moved to file: pic26589.pcx)

000132

LESAFFRE YEAST CORPORATION

FAX FORM

DATE: Monday, June 16, 2003

TO: Karin Ricker

C:

COMPANY: U.S. Food and Drug Administration

PAGES: (Including Coversheet)

MESSAGE:
Bob Biwersi, Director QA Lesaffre Yeast Corporation has requested me to fax this information to you. If you have any additional questions, please direct them to Bob at 1-877- 677-7000 extension 4085.

(b)(6)

Carole Paasch-Homan
QA Coordinator
414-615-4125

Fax No. 1-202-418-3131

Lesaffre Yeast Corporation
433 E. Michigan Street
Milwaukee, WI 53202
Phone: 414-221-6333

Pages 000134-000172 have been removed in accordance with copyright laws. Please see appended bibliography list of the references that have been removed from this request.

Blank Page Inserted for
Pagination

Computer Technology Services, Inc.

Reference List for Industry Submission, GRN 000120

Pages	*Author*	*Title*	*Publish Date*	*Source*	*BIB_Info*
000134-000172	Reed, Gerald; Nagodawithana, Tilak W.	Wine Yeast	1991	Yeast Technology	NA

NA- Not applicable