

IN SILICO STUDY OF ARO GENES INVOLVED IN THE EHRlich PATHWAY: COMPARISON BETWEEN *SACCHAROMYCES CEREVISIAE* AND *KLUYVEROMYCES LACTIS*

Zlatyo Georgiev Uzunov, Ventsislava Yankova Petrova, Samuil Lyubomirov Ivanov and Anna Vangelova Kujumdzieva
Sofia University "St. Kliment Ohridski", Faculty of Biology, Sofia, Bulgaria
Correspondence to: Anna Vangelova Kujumdzieva
Email: kujumdzieva@biofac.uni-sofia.bg

ABSTRACT

At the present state of the art, genome sequencing techniques are advancing with a faster pace than the experimental proof of predicted genes and their functions. Many approaches for clustering annotated genes based on the degree of their sequence homology have emerged during the last decade. In the current comparative genomic study a parallel has been drawn between a set of ARO genes involved in the aromatic alcohol synthesis in the model yeast organism *Saccharomyces cerevisiae* and their orthologs in *Kluyveromyces lactis*. Through a combination of sequence comparison, online resources for homology and topology-based clustering, and a tool for prediction of the cellular localisation of proteins, it has been shown that the genes involved in the first two steps of the 2-phenylethanol synthesis in *K. lactis* are generally conserved when compared to their *S. cerevisiae* orthologs. They show 40-62% identity of the encoded protein as well as conserved synteny and cellular localisation. It is also likely that these species utilise the same type of regulation mechanisms for the process. The presence of a second ARO8 ortholog in *K. lactis* suggests a more environmentally flexible and also effective production of 2-phenylethanol. This comparison aims to facilitate the design of further experiments that can reveal the regulation of fusel alcohol formation in *K. lactis* and improve their production.

Biotechnol. & Biotechnol. Eq. 2011, 25(4), Suppl., 133-137

Keywords: Ehrlich pathway, 2-phenylethanol, ARO genes

Introduction

Saccharomyces cerevisiae and *Kluyveromyces lactis*, both producing fusel alcohols, are phylogenetically distant yeast species with a common ancestor in their evolutionary history (14). Their phylogenetic distance is comparable to that found between *Homo sapiens* and the fish *Takifugu rubripes* and indicates that the parallel between them is quite a difficult task (6). In addition, it is already proved that the entire genome of *S. cerevisiae* has undergone duplication throughout its evolutionary history and this event has generated a lot of paralogous gene copies and facilitated genome rearrangements (14).

At the present state of the art, genome sequencing techniques are advancing with a faster pace than the experimental proof of predicted genes and their functions (16). This is one of the reasons that make comparative genomics a powerful tool for the studying of phylogenetically-related species and the transfer of knowledge from well-known model organisms to less known newly sequenced genomes. In the last decade many methods for clustering of annotated genes in protein families based on automated or semi-automated approach emerged (19, 25). We used some of these methods and genetic database resources to identify and verify the orthologs of genes involved in the phenylethanol and tryptophol synthesis with the aim to

WHO WE ARE AND WHAT WE ACHIEVED

facilitate the design of further experiments that can reveal the regulation of higher alcohol formation in *K. lactis*.

Higher alcohols, sometimes referred to as fusel alcohols, are, from a biotechnological point of view, one of the most important groups of volatile compounds produced by yeast during alcoholic fermentation of sugars. They are characterised as containing more than two carbon atoms and include the branch chained alcohols (isobutanol, amyl alcohol, isoamyl alcohol), and the aromatic alcohols 2-phenylethanol, tryptophol and tyrosol. According to Rapp and Versini (20) concentrations of higher alcohols below 300 mg L⁻¹ add desirable complexity of wine, whereas higher concentrations can be detrimental to wine quality. Conversely, 2-phenylethanol is generally a positive contributor to wine aroma, being characterized by a pleasant rose-like aroma (22). Moreover, this compound is the most used fragrance chemical in perfume and cosmetics, as well as in food industry (4, 7, 8). Many factors can affect the quantity of the formed 2-phenylethanol: yeast species and strain, carbohydrates concentration, fermentation temperature, pH, assimilable nitrogen, aeration and etc. Furthermore, the production of 2-phenylethanol in the model microorganism *Saccharomyces cerevisiae* is cell density dependent (3).

In the experimental ground for this analysis are ARO genes involved in the Ehrlich pathway and in the biosynthesis of the mentioned aromas (Fig. 1). It is known that this pathway is also responsible for the degradation of amino acids and their use as a source of nitrogen (5). The chemical degradation of phenylalanine, tryptophan and tyrosine leads to the synthesis of phenylethanol and tryptophol through a three-step biochemical

reaction comprising transamination, decarboxylation and reduction to the corresponding aromatic alcohol (5, 9) (**Fig. 1**). It has been proved that this process is involved in the quorum sensing mechanism in *Saccharomyces cerevisiae* and its regulation is dependent on the concentration of ammonia in the culture medium and on a feedback signal from the end products phenylethanol and tryptophol (3, 26). Both control mechanisms are acting on the first two stages of the reactions – the transamination and the decarboxylation (26). The transamination is catalysed by aromatic aminotransferases I and II encoded by the genes ARO8 and ARO9 in *S. cerevisiae* (13). The decarboxylation step in *S. cerevisiae* is also assured by an even greater redundancy – five decarboxylases are capable of catalysing this step, encoded by the genes ARO10, PDC1, PDC5, PDC6, and TIH3 (24) (**Fig. 1**). The ARO80 gene encodes a transcription factor that is activated by the presence of tryptophol and, in turn, enters the nucleus and activates the expression of ARO8 and ARO9 genes (12, 26). It is also known that high population density expresses production of higher amount of 2-phenylethanol and at the same time its biosynthetic mechanism overlaps with the nitrogen sensing pathway. Thus, the expression of ARO8 and ARO9 strongly depends on the ammonium source. The latter is able to repress the filamentous growth of *Saccharomyces cerevisiae* (26). Nevertheless, this communication molecule does not seem to affect the elements of nitrogen activated protein kinase – protein kinase A (MAPK-PKA) pathway involved in information transition for nitrogen storage in the same microorganism.

Having in mind the importance of 2-phenylethanol for the aroma profile of different commercial products and the characteristic findings for ARO8 and ARO9 genes in the model organism *Saccharomyces cerevisiae*, a comparative investigation of genes involved in the first two stages of the 2-phenylethanol synthesis of *S. cerevisiae* and *Kluyveromyces lactis* is shown. Moreover, this investigation was also motivated by the phylogenetic distance between the two species, having a common ancestor, and the opportunity to look into the evolution events of the Ehrlich pathway and their impact on the 2-phenylethanol synthesis.

Materials and Methods

Selection of gene sets

Target genes related to the Ehrlich pathway were selected through the SGD Pathway tools (<http://pathway.yeastgenome.org>) and related publication data (5, 26).

Identification of orthologous genes

Protein sequences of selected genes have been downloaded from the site of the Génolevures project (<http://www.genolevures.org>) (21). The values of protein identity were downloaded from the BLASTp alignment data of the Génolevures database (21) and the Yeast Gene Order Browser (YGOB - <http://wolfe.gen.tcd.ie/ygob/>) (2). Candidates for orthologous genes of the *S. cerevisiae* set were selected as best identity matches and then verified with comparison of the annotation data of each gene.

Detection of synteny

The initial set of *K. lactis* orthologs was verified through version 5 of YGOB (<http://wolfe.gen.tcd.ie/ygob/>) (2). The YGOB search tool was launched for each of the *S. cerevisiae* genes with the following parameters: Post WGD species: “*S. cerevisiae*”, Pre WGD species: “*K. lactis*”, RNA features: “ON”, Window: “±4”; All searches were executed with the Saccharomyces Genome Database (SGD, <http://www.yeastgenome.org>) systematic name of the *S. cerevisiae* gene. Only syntenic genes that span through the same chromosomes as the genes in the gene pair were considered as valid.

Intracellular localisation

The intracellular localisation of the proteins encoded by the orthologous genes in *K. lactis* were analysed through PSORT II Prediction software (<http://psort.hgc.jp>) (18). The cellular targeting data for *S. cerevisiae* were downloaded from SGD (<http://www.yeastgenome.org>).

Results and Discussion

Aromatic aminotransferases

The ARO8 gene codes for aromatic aminotransferase I (**Fig. 1**). This enzyme catalyses the first step of tryptophan, phenylalanine, and tyrosine catabolism (5). ARO8 is situated on chromosome VII in *S. cerevisiae* and its expression is regulated by the general control of amino acid biosynthesis (13). ARO8 has two orthologous genes in the genome of *K. lactis* – KLLA0A04906g situated on chromosome I and KLLA0F10021g on chromosome VI. KLLA0A04906g has 40% identity with the protein sequence of ARO8 and KLLA0F10021g, 62% (**Fig. 2**). Both genes are acceptable candidates for orthologs but the higher identity as well as the conserved synteny (there is a similar position of other neighbouring orthologs in both species) make KLLA0F10021g a better candidate. As for KLLA0A04906g, it could be a duplicated paralogous copy of KLLA0F10021g that has emerged in the evolutionary path of *K. lactis* after the separation from the common ancestor and the *S. cerevisiae* lineage. The hypothesis is supported by the fact that these genes had syntenic orthologs in other species that have not undergone whole genome duplication – *Kluyveromyces thermotolerans*, *Eremothecium (Ashbya) gossypii*, *Saccharomyces kluyveri*, and *Kluyveromyces waltii* (10). The product of ARO8 in *S. cerevisiae* is localised in the cytosol (11) and from the two candidate orthologs PSORT II predicted unambiguous cytosol localisation only for the product of KLLA0A04906g (**Table 1**). The better ortholog candidate KLLA0F10021g is the only one in the set that showed an ambiguous result in the localisation prediction different from the annotation in *S. cerevisiae* – extracellular and cell wall localisation, cytoplasmic, and nucleic (**Table 1**).

The ARO9 gene codes for another aminotransferase – aromatic aminotransferase II (**Fig. 1**), which also catalyses the first step of tryptophan, phenylalanine, and tyrosine catabolism. In *S. cerevisiae* it is localised on chromosome

TABLE 1

Comparison of the cellular localisation of *S. cerevisiae* genes and their *K. lactis* best orthologs. The localisation of *K. lactis* genes is predicted by the in silico analysis. The percentage of probability of the prediction is indicated when more than one possible localisations with difference less than 20% was returned. The values of localisation for *S. cerevisiae* genes are displayed according to the annotation on the SDG website

Systematic Names in the genome of <i>K. lactis</i>	Systematic Names in the genome of <i>S. cerevisiae</i>	Orthologue Standard Names	Predicted localisation in <i>K. lactis</i>	Localisation in <i>S. cerevisiae</i>
KLLA0A04906g	YGL202W	ARO8	cytoplasm	cytoplasm
KLLA0F10021g		ARO8	33.3%: extracellular, including cell wall 22.2%: cytoplasm 22.2%: nucleus	
KLLA0D11110g	YHR137W	ARO9	cytoplasm	cytoplasm nucleus
KLLA0E02707g	YDR380W	ARO10	cytoplasm	cytoplasm
KLLA0A01804g	YDR421W	ARO80	nucleus	nucleus

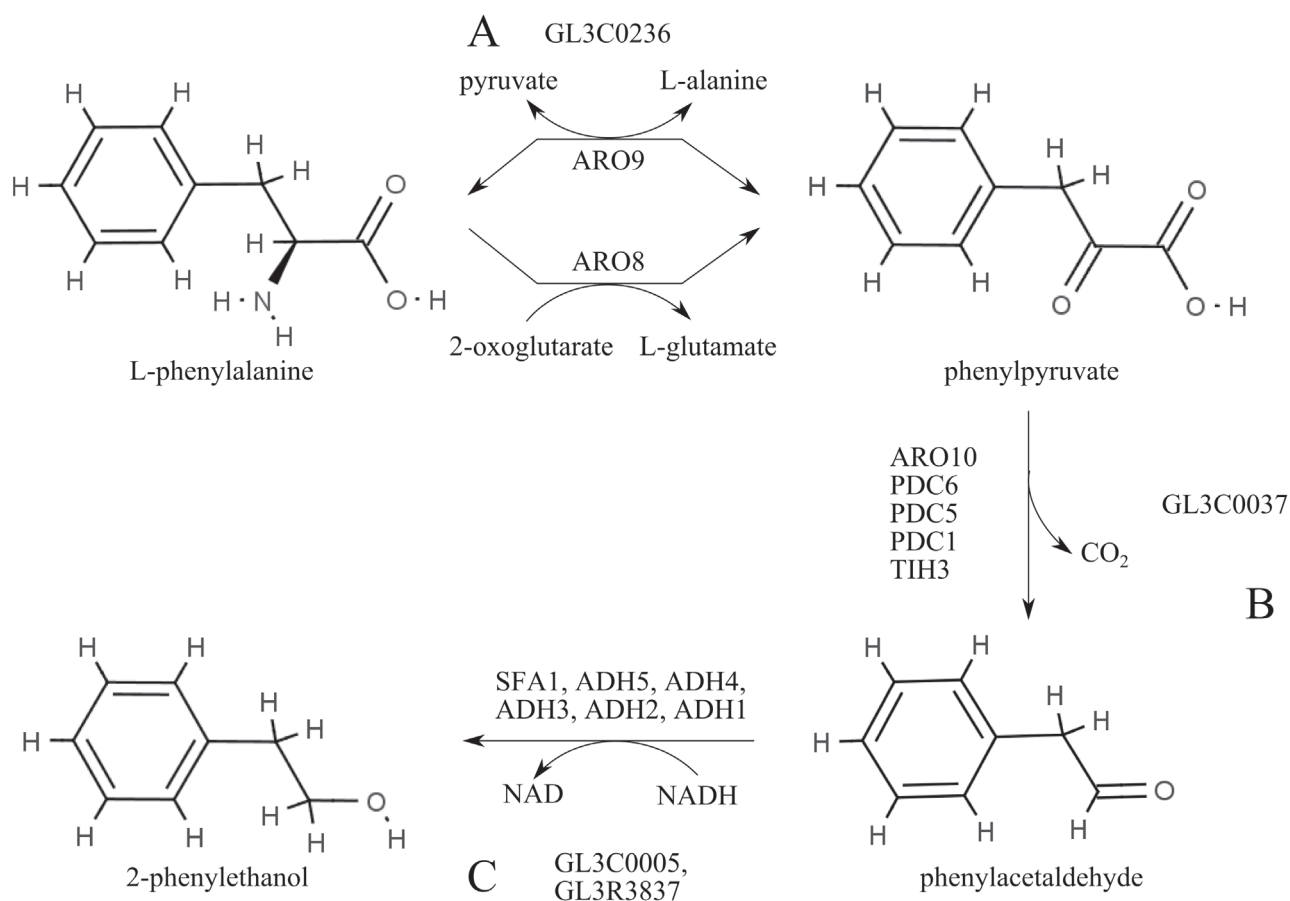


Fig. 1. Schematic representation of the three stages of 2-phenylethanol synthesis and the genes involved in them: (A) Aminotransferase reaction, (B) Decarboxylation, (C) Reduction to aromatic alcohol. Standard names of genes coding for the enzymes are shown on the inner side next to each reaction. The Génolevures families of the genes involved in each reaction are shown on the outer side (family names start with “GL3”).

VIII. ARO9 expression is induced when aromatic amino acids are present in the growth medium and also in *aro8* mutants grown on minimal ammonia medium (13). In the genome of *K. lactis* KLLA0D11110g is situated on chromosome IV and was identified as an ARO9 ortholog with 44% identity (Fig. 2). These genes also had conserved synteny. PSORTII predicted the same localisation for the product of KLLA0D11110g as for ARO9 in *S. cerevisiae* (Table 1).

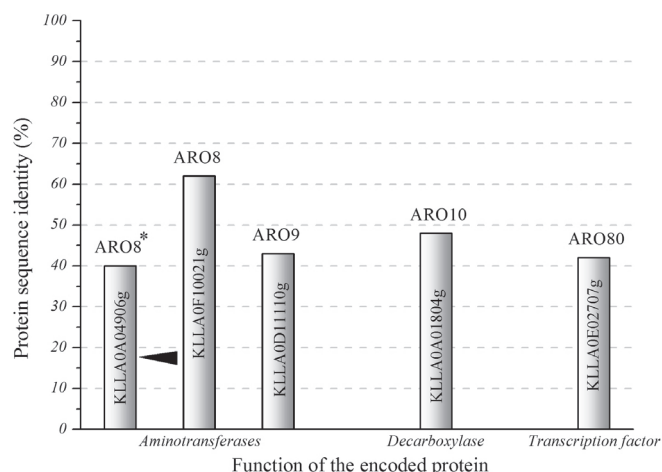


Fig. 2. Protein sequence identity of orthologous ARO genes between *S. cerevisiae* and *K. lactis*. The histograms are grouped by the function of the protein encoded by the gene. Columns connected with “◄” denote genes in *K. lactis* that have the same ortholog in *S. cerevisiae*; “*” denotes genes that do not have a conserved synteny in *K. lactis* according to the YGOB web application. Systematic names of *K. lactis* orthologs are displayed in the columns.

As a whole it seems that the first step in the phenylalanine and tryptophan degradation in *K. lactis* is assured at least by one aromatic aminotransferase I encoded by KLLA0F10021g because its percentage of identity was close to the mean evolutionary distance between the orthologs of these two species (6). In addition, it was the most conserved gene in the aromatic aminotransferase group and showed a partially conserved synteny. The hypothetical paralogous copy of ARO8 – KLLA0A04906g, was also found and suggested the possibility for production of couple isozymes, functioning under different environmental and physiological conditions, thus ensuring increased cellular robustness of this yeast species. However, further experimental proof of its function and regulation mechanisms is recommended as a second paralogous copy of ARO8 which lacks in the model strain. This finding could predict robustness against deletion (1, 15), higher expression, and increased production of 2-phenylethanol by analogy with the result from the overexpression of BAT1 and BAT2 genes leading to increased synthesis of odorous compounds (17).

Phenylpyruvate decarboxylases

Another ARO gene, ARO10, is also involved in the process of regulation of phenylethanol biosynthesis. It represents a phenylpyruvate decarboxylase catalysing the decarboxylation of phenylpyruvate to phenylacetaldehyde, which is the first specific step in the Ehrlich pathway (Fig. 1). Its function is

coupled with that of the ARO9 gene product, but its product catalyses a decarboxylation reaction. The decarboxylation step in *S. cerevisiae* is assured by a more redundant set of genes coding for isoenzymes with decarboxylase function – ARO10, PDC1, PDC5, PDC6, and TIH3 that is active only in the presence of the product of one of the PDC genes (23). This specific “burst” of gene paralogs is most probably also a consequence of the whole genome duplication event (14) and is not observed in *K. lactis*. Transcript level analysis revealed that among the five thiamine-pyrophosphate-dependent decarboxylases (ARO10, PDC1, PDC5, PDC6, and TIH3), only ARO10 was transcriptionally up-regulated when phenylalanine, leucine, or methionine was used as a nitrogen source compared to growth on ammonia, proline, and asparagines (23). In the genome of *S. cerevisiae* ARO10 is situated on chromosome X. Its ortholog in *K. lactis*, KLLA0E02707g, is situated on chromosome V and shows conserved synteny with the neighbouring region of ARO10 in *S. cerevisiae*. The protein sequences of the orthologs have 42% identity (Fig. 2) and the cellular localisation for the product of KLLA0E02707g predicted by PSORT II is the same as that of the ARO10 product (Table 1).

Transcription factors

The expression of ARO9 and ARO10 in *S. cerevisiae* is regulated by the transcription factor Aro80 (12). Thus the expression of ARO80, in turn, is positively regulated by tryptophol, which is the end product of tryptophan degradation in the Ehrlich pathway. ARO10 is located on chromosome IV in the genome of *S. cerevisiae*. Its ortholog in *K. lactis*, KLLA0A01804g, is on chromosome I and has relatively conserved synteny. The identity of their protein sequences is 42% (Fig. 2). The prediction of PSORT II for the cellular localisation of the protein encoded by KLLA0A01804g is coherent with that of the *S. cerevisiae* product (Table 1).

Regarding the lower identity percentage (Fig. 2) and the more complicated regulatory relation between ARO9, ARO10 and ARO80 orthologs, an experimental proof of their function in *K. lactis* is required. The conserved synteny for the orthologs of these genes in *K. lactis* strongly suggests common expression patterns and regulation mechanisms.

Conclusions

Prediction, control and profiling of yeast metabolic routes responsible for flavour formation is of particular interest. Moreover, the availability of the sequence of the *Saccharomyces* genome as well as those of several other ascomycetous yeasts in combination with the development of chemical analytical technologies with dynamic ranges sensitive enough to detect volatile aromatic compounds has generated renewed interest in defining the role of yeast in the biosynthesis of different aroma and flavour. Genetic differences among aroma-producing strains are well documented and aroma profiles also appear to vary, implying that specific allelic alterations may exist and impact the production of compounds associated with flavour. In this respect the performed *in silico* analysis of

the transaminases' gene family represented by ARO8, ARO9 genes as well as of the decarboxylase coding ARO10 gene in fermentative *Saccharomyces cerevisiae* and respiratory *Kluyveromyces lactis* strains suggested the same initial steps for the biosynthesis of biotechnologically important aroma compound 2-phenylethanol. These genes are present in both yeasts and are generally conserved. However, some differences have appeared. In *K. lactis* yeasts two putative paralogs of the ARO8 gene were detected with different cellular localisation (**Table 1**). Thus, on the basis of the already discussed data about the branched-chain amino acid transaminase genes, the role of gene duplication, as well as the *in silico* results for the second ARO8 gene copy in the genome of *K. lactis*, it could be suggested that there is enhanced expression of aromatic aminotransferase I and hence higher production of 2-phenylethanol. The established differences between the ARO genes of *S. cerevisiae* and *K. lactis* yeasts constitute new promising research fields and applied biotechnological implications could be also envisaged.

Acknowledgements

This work was supported by grant BG051PO001-3.3.04/32 financed by Operational Program Human Resources Development (2007 - 2013) and co-financed by the European Social Fund of the European Union. It complies with the topic of COST action FA0907 – BIOFLAVOUR.

REFERENCES

1. Bisson L.F., Karpel J.E. (2010) *Annu. Rev. Food Sci. Technol.*, **1**, 139-162.
2. Byrne K.P., Wolfe K.H. (2005) *GenomeRes.*, **15**, 1456-1461.
3. Chen H., Fink G. R., (2006) *Genes&Development*, **20**, 1150 -1161.
4. Clark G. S., (1990) *Perfumer&Flavorist*, **15**, 37-44.
5. Dickinson J. R., Salgado L. E. J., Hewlins M. J. E., (2003) *Journal of BiologicalChemistry*, **278**, 8028 -8034.
6. Dujon B., (2006) *TRENDS inGenetics*, **22**, 375-387.
7. Etschmann M., Bluemke W., Sell D., Schrader J., *AppliedMicrobiologyandBiotechnology*, **59**, 1-8.
8. Fabre C. E., Blanc P. J., Goma G., (1998) *Perfumer&Flavorist*, **23**, 43-45.
9. Hazelwood L. A., Daran J.-M., vanMaris A. J. A., Pronk J. T., Dickinson J. R., (2008) *Appl. Environ. Microbiol.*, **74**, 2259-2266.
10. Hébert A., Casaregola S., Beckerich J., (2011) *FEMS YeastResearch*, **11**, 366-378.
11. Huh W.-K., Falvo J. V., Gerke L. C., Carroll A. S., Howson R. W., Weissman J. S., O'Shea E. K., (2003) *Nature*, **425**, 686-691.
12. Iraqui I., Vissers S., Andre B., Urrestarazu A., (1999) *Mol. Cell. Biol.*, **19**, 3360-3371.
13. Iraqui I., Vissers S., Cartiaux M., Urrestarazu A., (1998) *Molecularand General Genetics MGG*, **257**, 238-248.
14. Kellis M., Birren B. W., Lander E. S., (2004) *Nature*, **428**, 617-624.
15. Maltsev N., Glass E. M., Ovchinnikova G., Gu Z., (2005) *Journal of Biochemistry*, **137**, 177 -187.
16. Mardis E. R., (2008) *TrendsInGenetics*, **24**, 133-141.
17. Mariska L., Bauer F. F., Styger G., Lambrechts M. G., Pretorius I. S., (2006) *FEMS YeastResearch*, **6**, 726-743.
18. Nakai, K., Kanehisa, M., (1992) *Genomics*, **14**, 897-911.
19. Nikolski M., Sherman D. J., (2007) *Bioinformatics*, **23**, e71 -e76.
20. Rapp A., Versini G., (1996) *Wein-Wissenschaft*, **51**, 193-203.
21. Sherman D. J., Martin T., Nikolski M., Cayla C., Souciet J.-L., Durrens P., for the Genolevures Consortium, (2009) *Nucl. AcidsRes.*, **37**, D550-554.
22. Swiegers J. H., Bartowsky E. J., Henschke P. A., Pretorius I. S., (2005) *Australian Journal of Grapeand Wine Research*, **11**, 139-173.
23. Vuralhan Z., Luttik M. A. H., Tai S. L., Boer V. M., Morais M. A., Schipper D., Almering M. J. H., Kotter P., Dickinson J. R., Daran J.-M., Pronk J. T., (2005) *Appl. Environ. Microbiol.*, **71**, 3276-3284.
24. Vuralhan Z., Morais M. A., Tai S.-L., Piper M. D. W., Pronk J. T., (2003) *Appl. Environ. Microbiol.*, **69**, 4534-4541.
25. Waterhouse R. M., Zdobnov E. M., Tegenfeldt F., Li J., Kriventseva E. V., (2010) *Nucleic Acids Research*, **39**, D283-D288.
26. Wuster A., Babu M. M., (2010) *Mol. BioSyst.*, **6**, 134.