



RESEARCH ARTICLE

Open Access

Genetic basis of the highly efficient yeast *Kluyveromyces marxianus*: complete genome sequence and transcriptome analyses

Noppon Lertwattanasakul^{1,5†}, Tomoyuki Kosaka^{2†}, Akira Hosoyama^{3†}, Yutaka Suzuki^{4†}, Nadchanok Rodrussamee^{1,6}, Minenosuke Matsutani², Masayuki Murata¹, Naoko Fujimoto¹, Suprayogi¹, Keiko Tsuchikane³, Savitree Limtong⁵, Nobuyuki Fujita³ and Mamoru Yamada^{1,2*}

Abstract

Background: High-temperature fermentation technology with thermotolerant microbes has been expected to reduce the cost of bioconversion of cellulosic biomass to fuels or chemicals. Thermotolerant *Kluyveromyces marxianus* possesses intrinsic abilities to ferment and assimilate a wide variety of substrates including xylose and to efficiently produce proteins. These capabilities have been found to exceed those of the traditional ethanol producer *Saccharomyces cerevisiae* or lignocellulose-bioconvertible ethanologenic *Scheffersomyces stipitis*.

Results: The complete genome sequence of *K. marxianus* DMKU 3-1042 as one of the most thermotolerant strains in the same species has been determined. A comparison of its genomic information with those of other yeasts and transcriptome analysis revealed that the yeast bears beneficial properties of temperature resistance, wide-range bioconversion ability, and production of recombinant proteins. The transcriptome analysis clarified distinctive metabolic pathways under three different growth conditions, static culture, high temperature, and xylose medium, in comparison to the control condition of glucose medium under a shaking condition at 30°C. Interestingly, the yeast appears to overcome the issue of reactive oxygen species, which tend to accumulate under all three conditions.

Conclusions: This study reveals many gene resources for the ability to assimilate various sugars in addition to species-specific genes in *K. marxianus*, and the molecular basis of its attractive traits for industrial applications including high-temperature fermentation. Especially, the thermotolerance trait may be achieved by an integrated mechanism consisting of various strategies. Gene resources and transcriptome data of the yeast are particularly useful for fundamental and applied researches for innovative applications.

Keywords: *Kluyveromyces marxianus*, Thermotolerant yeast, Complete genome sequence, Transcriptome analysis, Xylose fermentation

Background

Along with rising concern about global warming and the rapid increase in fuel consumption, there is worldwide interest in the production of bioethanol from renewable resources [1]. For economically sustainable production of bioethanol, it is necessary to increase the types of biomass

such as lignocellulosic materials that can be used without competing with food supplies. Accordingly, microbes that can efficiently convert various sugars in these kinds of biomass to ethanol must be developed.

A high-temperature fermentation (HTF) technology is expected to help reduce cooling cost, efficiently achieve simultaneous saccharification and fermentation, reduce the risk of contamination, and offer stable fermentation even in tropical countries [2,3]. Thermotolerant yeast *Kluyveromyces marxianus*, which is able to ferment various sugars, may be a suitable microbe for HTF with lignocellulosic hydrolysates [4,5].

* Correspondence: m-yamada@yamaguchi-u.ac.jp

[†]Equal contributors

¹Applied Molecular Bioscience, Graduate School of Medicine, Yamaguchi University, Ube 755-8505, Japan

²Department of Biological Chemistry, Faculty of Agriculture, Yamaguchi University, Yamaguchi 753-8515, Japan

Full list of author information is available at the end of the article

K. marxianus is a haploid, homothallic, thermotolerant, hemiascomycetous yeast [6,7] and a close relative of *Kluyveromyces lactis*, a model Crabtree-negative yeast [8-11]. Both yeasts share the assimilating capability of lactose, which is absent from *Saccharomyces cerevisiae*. *K. marxianus* has a number of advantages over *K. lactis* or *S. cerevisiae*, including the intrinsic fermentation capability of various sugars at high temperatures [4,12,13], weak glucose repression that is preferable for mixed sugars such as hemicellulose hydrolysate, and fermentability of inulin [13,14]. However, its fermentation activity from xylose is extremely low compared to that of glucose. Recently, we developed a procedure that improves this disadvantageous trait (unpublished data) and increases the fermentation activity to slightly less than that of *Scheffersomyces stipitis* at around 30°C and a much higher activity at higher temperatures. Many biotechnological applications of *K. marxianus* have so far been achieved: production of various enzymes including heterologous proteins, aroma compounds or bioingredients, reduction of lactose content in food products, production of ethanol or single-cell protein, and bioremediation [13]. In addition, novel methods and genetic tools for genetic engineering have been developed on the basis of its high nonhomologous end-joining activity [15,16]. *K. marxianus* is thus a highly competent yeast for future developments. In order to facilitate such developments, its genomic information is essential. Draft genome sequences of three *K. marxianus* strains have been published [7,17,18], but no detailed analysis is available.

This study provides core information on *K. marxianus* DMKU 3-1042, which is one of the most thermotolerant strains in the same species isolated (3, unpublished data), including its ability to assimilate various sugars and

the molecular basis of its thermotolerance and efficient protein productivity, in addition to the complete genome sequence.

Results

Genomic information and comparative genomics

The genome sequence of *K. marxianus* DMKU 3-1042 was precisely determined (less than one estimated error per chromosome) by nucleotide sequencing with three different sizes of shotgun libraries. Telomeric regions were further analyzed by transposon-insertion sequencing of corresponding fosmid clones. This strategy allowed us to determine the complete genome sequence of 11.0 Mb including all centromeric regions and boundary regions containing up to one to several sequence repeats (GGTGTACGGATTGATTAGTTATGT) of telomeres. Optical mapping confirmed the genome organization except for three inverted regions, which were fixed in the final complete sequences (Additional file 1: Figure S1). There are eight chromosomes ranging in size from 0.9 to 1.7 Mb and a mitochondrial genome of 46 kb. The annotation process predicted 4,952 genes (Table 1), of which 98.0% were predicted to consist of a single exon (Additional file 2). The average gene density is 68.0% (Table 2). The average gene and protein lengths are 1.5 kb and 501 amino acids, respectively (Table 1).

Eukaryotic orthologous groups (KOG) database analysis led to the assignment of protein functions of about 72.4% of predicted genes (Additional file 1: Table S1) and protein domains were predicted in 3,584 gene models. UniProt and KAAS assignments led to the assignment of homologous genes of about 86.4% of predicted genes and KEGG Orthology number of 50.5%, respectively. The yeast shares 1,552 genes with *K. lactis*, *Ashbya gossypii*,

Table 1 General information of nuclear and mitochondrial genomes of *K. marxianus* DMKU 3-1042

	Length	CDS	Intron-containing CDS	tRNA	rDNA	na_length average	na_length maximum	aa_length average	aa_length maximum
Total	10,966,467	4,952	172	202	8 ^b				
Average						1,505		501	
Chromosome									
1	1,745,387	803	25	31	0	1,473	12,174	491	4,058
2	1,711,476	808	29	28	0	1,440	6,168	480	2,056
3	1,588,169	706	21	24	0	1,553	8,004	517	2,668
4	1,421,472	624	26	25	0	1,593	14,742	531	4,914
5	1,353,011 ^a	611	20	30	6 ^b	1,529	9,018	509	3,006
6	1,197,921	537	18	17	0	1,500	8,709	500	2,903
7	963,005	438	19	10	0	1,454	8,310	484	2,770
8	939,718	414	12	15	0	1,523	9,381	507	3,127
Mitochondrion	46,308	11	2	22	2	978	1,491	326	497

^aThe length does not include that of most of rDNA. ^bSix rDNA copies in the genome sequence in database, but 140 rDNA copies by optical mapping. CDS, coding DNA sequence.

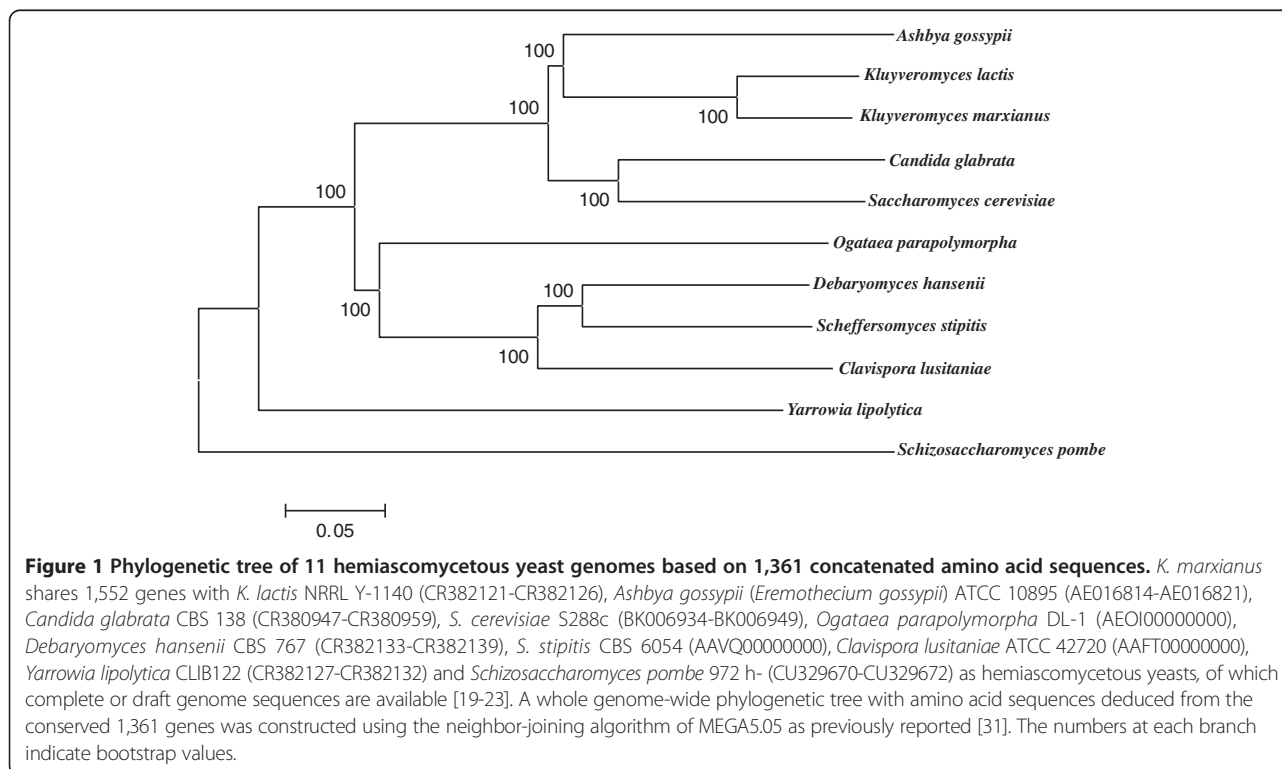
Table 2 General characteristics of 11 hemiascomycetous yeast genomes

Species	Genome size (Mb)	Average G + C content (%)	Total CDS	Total tRNA genes	Average gene density (%)	Average G + C in CDS (%)	Source
<i>Kluyveromyces marxianus</i>	10.97	40.12	4,952	202	68.00	41.69	This study
<i>Kluyveromyces lactis</i>	10.60	38.70	5,329	162	71.60	40.10	[20]
<i>Saccharomyces cerevisiae</i>	12.10	38.30	5,807	274	70.30	39.60	[20]
<i>Candida glabrata</i>	12.30	38.80	5,283	207	65.00	41.00	[20]
<i>Yarrowia lipolytica</i>	20.50	49.00	6,703	510	46.30	52.90	[20]
<i>Scheffersomyces stipitis</i>	15.40	41.10	5,841	-	55.90	42.70	[21]
<i>Ashbya gossypii</i>	9.12	51.70	4,776	220	77.10	52.80	[19]
<i>Ogataea parapolymorpha</i>	8.87	47.83	5,325	80	84.58	49.13	This study ^a
<i>Debaryomyces hansenii</i>	12.18	36.34	6,290	225	74.31	37.45	[20]
<i>Clavispora lusitaniae</i>	12.11	44.50	5,936	217	68.07	46.80	[22]
<i>Schizosaccharomyces pombe</i>	12.59	36.04	5,133	195	57.17	39.63	[23]

^aValues were summarized by us using data from Joint Genome Institute (JGI, <http://gold.jgi-psf.org>). G + C, guanine + cytosine; CDS, coding DNA sequence.

Candida glabrata, *S. cerevisiae*, *Ogataea parapolymorpha*, *Debaryomyces hansenii*, *S. stipitis*, *Clavispora lusitaniae*, *Yarrowia lipolytica*, and *Schizosaccharomyces pombe* as hemiascomycetous yeasts [19-23]. The phylogenetic tree exhibits the closest location of *K. marxianus* to *K. lactis* and closer to *A. gossypii*, *C. glabrata*, and *S. cerevisiae* in the 11 yeasts (Figure 1). Consistent with this, *K. marxianus* shares 4,676; 3,826; 3,672; and 3,853 genes with *K. lactis*, *A. gossypii*, *C. glabrata*, and *S. cerevisiae*, respectively. On

the other hand, there are 193 genes specific for *K. marxianus* (Additional file 1: Table S2), which may be responsible for its species-specific characteristics, of which two thirds of the genes could not be assigned by the KOG database (Additional file 1: Table S3). There are 422 genes shared only between *K. marxianus* and *K. lactis* (Additional file 1: Table S4), which may be related to their genus-specific characteristics, such as production of β -galactosidase [24], assimilation of a wide variety of inexpensive



substrates [25], efficient productivity of heterologous proteins [26–28], and synthesis of a killer toxin against certain ascomycetous yeasts [29,30].

The two most attractive traits of *K. marxianus* for fermentation applications are thermotolerance and pentose assimilation capability, which are also found in *O. parapolymorpha* and *S. stipitis*, respectively. The number of genes that are shared between the two thermotolerant yeasts but absent from *K. lactis* is 30, including genes for three siderophore-iron transporters and three vacuolar proteins (Additional file 1: Table S5). Notably, there are 27 putative sugar transporters in the *K. marxianus* genome (Additional file 1: Table S6). Like *S. stipitis*, the initial xylose catabolism after its uptake in *K. marxianus* is accomplished by three genes, *XYL1*, *XYL2*, and *XKS1*, which are involved in the conversion of xylose to xylulose-5-phosphate as an intermediate in the pentose phosphate pathway (PPP). Genes for utilization of various other sugars and alcohol dehydrogenases are listed (Table 3).

Chromosomal segments including 4,277 genes, which retain the ancestral gene groupings, were found between *K. marxianus* and *K. lactis* (Additional file 1: Figure S2). The average of mapped segments to the chromosome is 57.9%, and the *K. marxianus* chromosome best covered by *K. lactis* chromosomal segments is chromosome 6, with 60.8% coverage.

Ribosomal DNA (rDNA) copy number and thermotolerance

Optical mapping allowed us to estimate at least 140 copies of the rDNA gene as a cluster on chromosome 5 (Additional file 1: Figure S1), which occupies 67.5% (0.9 Mb) of the chromosome. To examine the relationship between the rDNA copy number and its thermotolerance among *K. marxianus* strains, strains exhibiting different growth at different temperatures (Additional file 1: Figures S3, S4) were subjected to a test to determine the copy number of rDNA (Additional file 1: Table S7). As a result, the rDNA copy number is not correlated to the thermotolerance of the yeast, and at least 31 copies of rDNA are sufficient to support its thermotolerance.

Genes regulated under a static condition

The alteration of genome-wide gene expression was analyzed by transcription start site sequencing (TSS Seq) under four different conditions (Additional file 3): shaking condition in yeast extract peptone dextrose (YPD) medium at 30°C (30D) or 45°C (45D), static condition in YPD medium at 30°C (30DS), and shaking condition in yeast extract peptone xylose (YPX) medium at 30°C (30X) and was expressed as the ratio of 30DS/30D, 45D/30D, and 30X/30D using 30D as a control condition, which was evaluated by a statistical test (FDR < 0.05) (Additional file 1: Table S8) and summarized (Additional file 1: Figure S5).

The growth of *K. marxianus* under the static condition is much slower than that under the shaking condition [4,12]. Under the 30DS condition, there were 159 significantly upregulated genes (Additional file 1: Table S8), and the top-five significantly enriched GO terms were ribosome biogenesis, ribonucleoprotein complex biogenesis, rRNA processing, rRNA metabolic process, and noncoding RNA processing (Additional file 1: Table S9); their individual gene details are shown in Additional file 1: Table S10. Interestingly, 55% of upregulated gene products are located in the nucleus (Additional file 1: Figure S6), some of which are factors for ribosome biogenesis, ATP-dependent RNA helicases, RNA polymerases subunits, nucleolar complex proteins, components of exosome complex, DNA polymerase subunits, and chromatin assembly factor 1 subunits. Conversely, there were 154 significantly downregulated genes (Additional file 1: Table S8), and their most significantly enriched GO terms were ascospore formation, sexual sporulation, sexual sporulation resulting in the formation of a cellular spore, cell development, and reproductive process in single-celled organisms (Additional file 1: Tables S11, S12). The largest population consists of 43 genes for membrane proteins including 15 transporters for amino acids and other metabolites. Taken together, under the static condition, *K. marxianus* may increase the turnover of RNAs and proteins in addition to suppression of transporters and spore formation that depends on mitochondrial respiration activity [32].

Under the static condition, the oxygen level in cells may become low as cells proliferate so that the condition may affect oxygen-requiring biosynthetic pathways, such as those for heme, sterols, unsaturated fatty acids, pyrimidine, and deoxyribonucleotides [33]. As expected, almost all genes related to ergosterol biosynthesis, sterol biosynthesis, unsaturated fatty acids production, pyrimidine synthesis, and ribonucleotide reductase were largely upregulated under the 30DS condition (Additional file 1: Figure S7). However, unlike *S. cerevisiae*, the expression of *MDL1* for a putative mitochondrial heme carrier was not significantly altered [34]. In addition, *NPT1* for nicotinate phosphoribosyl transferase (Npt1) involved in the nicotinamide adenine dinucleotide phosphate (NAD) salvage pathway, *CYC7* for cytochrome *c* and *AAC* for ADP/ATP carrier in mitochondrial inner membrane were upregulated. The ADP/ATP carrier functions to exchange cytoplasmic adenosine diphosphate (ADP) for mitochondrial adenosine triphosphate (ATP) under aerobic conditions and vice versa under anaerobic conditions [33]. Taken together, these results suggest that the enhanced expression of most genes for several oxygen-dependent biosynthetic pathways or some genes related to the production and management of energy is crucial for the cellular metabolism of *K. marxianus* under the static condition. Notably, almost all genes described above

Table 3 Genes for utilization of sugars at their initial catabolism and genes for alcohol dehydrogenases in *K. marxianus*

	Product	UniProt gene	Sugar
KMLA_60412	Hexokinase	HXK2 (RAG5)	Glucose, fructose, mannose
KLMA_10763	Glucose-6-phosphate dehydrogenase	RAG2	
KLMA_50384	Mannose 6-phosphate isomerase	PMI40	Mannose
KLMA_20333	Galactokinase	GAL1	Galactose
KLMA_20331	Galactose-1-phosphate uridylyltransferase	GAL7	
KLMA_30099	Phosphoglucosmutase	GAL5 (PGM2)	
KLMA_10683	Xylose reductase	XYL1	Xylose
KLMA_70044	Xylitol dehydrogenase	XYL2	
KLMA_80066	Xylulokinase	XKS1	
KLMA_30577	Arabinose dehydrogenase [NADP ⁺ dependent]	ARA1	Arabinose
KLMA_40310	Arabinose dehydrogenase [NAD dependent]	ARA2	
KLMA_80176	Xylose/arabinose reductase	YJR096W	
KLMA_10558	D-Arabitol-2-dehydrogenase	ARD2	
KLMA_10157	Probable ribokinase	RBK1	D-Ribose
KLMA_10176	Ribose-phosphate pyrophosphokinase 5	PRS5	
KLMA_10783	Sorbose reductase	SOU1	Mannitol, glucitol, L-sorbose
KLMA_10649	D-Lactate dehydrogenase [cytochrome], mitochondrial	DLD1	Lactate
KLMA_40583	D-Lactate dehydrogenase [cytochrome] 1, mitochondrial	DLD1	
KLMA_50301	D-Lactate dehydrogenase [cytochrome], mitochondrial	DLD1	
KLMA_60482	D-Lactate dehydrogenase [cytochrome] 2, mitochondrial	DLD2	
KLMA_10179	Glycerol-3-phosphate dehydrogenase [NAD(+)] 1	GPD1	Glycerol
KLMA_30722	Glycerol-3-phosphate dehydrogenase, mitochondrial	GUT2	
KLMA_60361	Glycerol uptake protein 1	GUP1	
KLMA_80411	Glycerol uptake/efflux facilitator protein	FPS1	
KLMA_80412	Glycerol kinase	GUT1	
KLMA_10427	Galactose/lactose metabolism regulatory protein GAL80	GAL80	Lactose
KLMA_20830	Lactose permease	LAC12	
KLMA_30010	Lactose permease	LAC12	
KLMA_30728	Lactose permease	LAC12	
KLMA_30011	Beta-glucosidase	-	Cellobiose
KLMA_20184	Endo-1,3(4)-beta-glucanase 1	DSE4	1,3-β-D-Glucan
KLMA_50517	Endo-1,3(4)-beta-glucanase 2	ACF2	
KLMA_10518	Inulinase	INU1	Sucrose, raffinose, inulin
KLMA_40102	Alcohol dehydrogenase 1	ADH1	
KLMA_40220	Alcohol dehydrogenase 2	ADH2	
KLMA_80306	Alcohol dehydrogenase 3	ADH3	
KLMA_20005	Alcohol dehydrogenase 4a	ADH4a	
KLMA_20158	Alcohol dehydrogenase 4b	ADH4b	
KLMA_40624	Alcohol dehydrogenase	ADH	
KLMA_80339	Alcohol dehydrogenase 6	ADH6	

NAD, nicotinamide adenine dinucleotide; NADP, nicotinamide adenine dinucleotide phosphate.

were upregulated not only under the 30DS condition but also under the 45D condition, suggesting that cells suffer from oxygen deficiency under the two conditions.

Metabolic changes were further analyzed by KEGG assignment. A number of genes for glycolysis after 1,3-bisphosphoglycerate, PPP, and tricarboxylic acid (TCA)

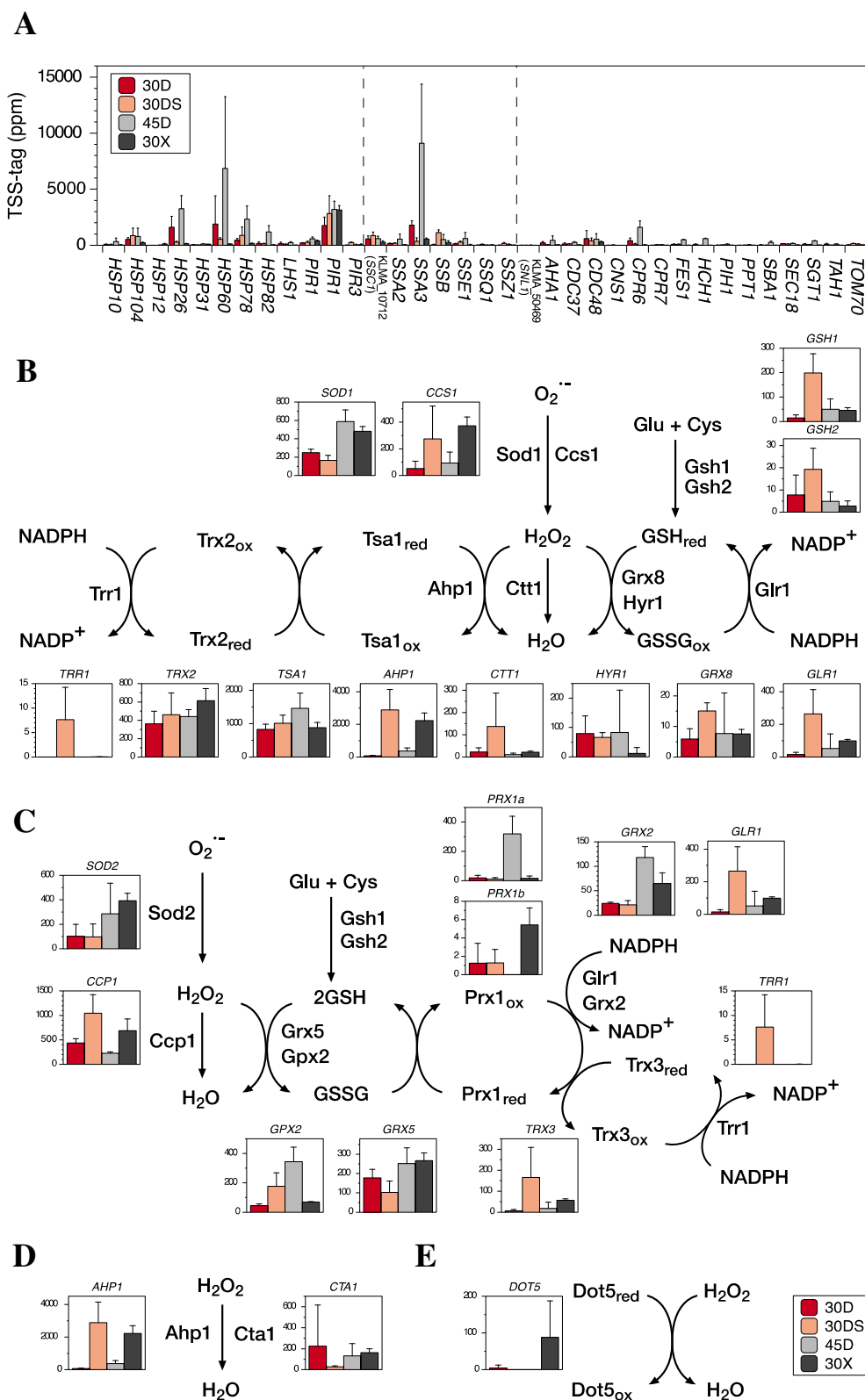


Figure 3 (See legend on next page.)

(See figure on previous page.)

Figure 3 Transcript abundance of genes related to oxidative stress response in *K. marxianus*. (A) Transcript abundance of heat shock genes and related genes listed in Additional file 1: Table S18 is represented as TSS-tag ppm. Enzyme reactions to scavenge ROS in (B) cytoplasm, (C) mitochondria, (D) peroxisome, and (E) nucleus are shown. Each column in the graph shows the transcript abundance of each gene as described in Figure 2.

The oxaloacetate-malate shuttle may contribute to the oxidation of NADH in cytoplasm. Many genes for the respiratory chain were downregulated, though compensatorily most of the chaperone-coding genes for cytochrome *c* oxidase were upregulated (Additional file 1: Figure S8).

Genes regulated under a high-temperature condition

To clarify the thermotolerant mechanism of *K. marxianus*, it is necessary to consider the expressional alteration of whole genomic genes at high temperature. Under the 45D condition, 508 genes were significantly downregulated (Additional file 1: Table S8) and the top-five significant GO terms were carbohydrate metabolic process, isoleucine biosynthetic process, small molecule metabolic process, monosaccharide metabolic process, and branched chain family amino acid biosynthetic process (Additional file 1: Tables S13, S14). Conversely, in 199 upregulated genes (Additional file 1: Table S8), the most significant GO terms were noncoding RNA processing, ribosome biogenesis, rRNA processing, ribonucleoprotein complex biogenesis, and rRNA metabolic process (Additional file 1: Tables S15, S16), and 45 genes were related to translation, transcription, DNA replication and repair, and protein and RNA degradations. Interestingly, *LYS21* for homocitrate synthase, which is linked to the key process of DNA damage repair in a nucleus [35] in addition to its involvement in lysine biosynthesis in the cytoplasm, was upregulated. Several genes for homologous recombination and nonhomologous end joining, which function in the repair of DNA double-stranded breaks, were also upregulated (Additional file 1: Figure S9). Therefore, it is assumed that *K. marxianus* copes with high temperatures by reducing central metabolic activities and reinforcing the synthesis and degradation of proteins and DNA repair.

In further analysis of the subcellular localization of products (Additional file 1: Figure S6), 21% of all upregulated genes are located in the nucleolus or nucleus (Additional file 1: Table S17), and are related to 18S rRNA preprocessing, 60S ribosomal subunit biogenesis, transcription process, and pre-rRNA processing (Additional file 1: Tables S15, S16). Conversely, products of several downregulated genes exist in the nucleolus, which are involved in 40S ribosomal subunit biogenesis and pre-18S rRNA processing (Additional file 1: Tables S13, S14). Notably, seven genes for mitochondrial ribosome subunits were downregulated. In addition, genes for DNA repair in nuclei and mitochondria, including a DNA damage sensor or

chromosome transmission fidelity protein, were significantly upregulated (Additional file 1: Table S16).

Genes for glycolysis were remarkably downregulated, except for *GLK1* and *FBP1*, under the 45D condition (Figure 2A), which is consistent with relatively slow growth speed and low ethanol productivity at high temperatures [4]. Conversely, *ZWF* and *SOL1* in PPP in addition to *GLK1* were upregulated, indicating an increase in NADPH amount. In TCA cycle, downregulation of *CIT1*, *LSC2*, *FUM1*, and *MDH1* and upregulation of *IDH1*, *IDH2*, *KGD1*, and *KGD2* were found, which might lead to the accumulation of intermediates (Figure 2B and Additional file 1: Table S16). Most mitochondrial genomic genes were selectively expressed (Additional file 1: Figure S8). Additionally, some genes for succinate dehydrogenase (Sdh) and *bc1* complex were upregulated, whereas some genes for NADH dehydrogenase (Ndh), Coenzyme Q biosynthesis, and cytochrome *c* oxidase including heme *a* synthesis were downregulated (Additional file 1: Figure S8). These findings allow us to speculate that *K. marxianus* scavenges H₂O₂ by the pathway of Sdh-*bc1* complex-cytochrome *c* peroxidase and prevents the production of ROS by reduction of gene expression for Ndh and Coenzyme Q biosynthesis at high temperatures.

Heat shock proteins (Hsps) and chaperones are expected to be crucial for survival at high temperatures. The transcription of *HSP26*, *HSP60*, *HSP78*, *HSP82*, *SSA3*, and *CPR6* was enhanced under the 45D condition (Figure 3A and Additional file 1: Table S18), suggesting that both mitochondrial and cytoplasmic compartments need such Hsps at that temperature.

Genes regulated under a xylose-utilizing condition

Under the 30X condition, the top-five significant GO terms of significantly upregulated genes were fatty acid catabolic process, monocarboxylic acid catabolic process, cellular lipid catabolic process, fatty acid β -oxidation, and fatty acid oxidation (Additional file 1: Tables S19, S20). Conversely, the most significant GO terms of significantly downregulated genes were α -amino acid metabolic process, carboxylic acid metabolic process, lysine biosynthetic process, small molecule biosynthetic process, and oxoacid metabolic process (Additional file 1: Tables S21, S22). Therefore, the 30X condition may stimulate the degradation of lipid in peroxisome and keep a low level of amino acid synthesis, which is consistent with the slow growth of the yeast in xylose [4].

The phylogenetic tree of sugar transporters revealed that KLMA_50360, KLMA_50361, KLMA_50362, KLMA_50363, and KLMA_50364 share high similarity with xylose transporters predicted in *S. stipitis* (Figure 4) [21]. Unlike *S. stipitis*, however, these genes were not induced specifically under the 30X condition. Genes of several transporters exhibited specific induction under the 30X or 30D condition, suggesting that KLMA_60073, KLMA_80101, and KLMA_70145 are involved in xylose uptake.

Genes for the initial catabolism of xylose, PPP, the conversion of PEP to ethanol, the mitochondrial conversion of acetaldehyde to acetyl-CoA and TCA cycle were relatively upregulated (Figure 2A,B). A part of the ethanol produced in the cytoplasm may thus be consumed by conversion to acetyl-CoA in mitochondria through the ethanol-acetaldehyde shuttle followed by TCA cycle, which is consistent with low ethanol productivity in xylose medium [4]. In contrast, *ADH2* and *ADH4b* for alcohol dehydrogenases, *GUT1* for glycerol kinase and *GUT2* for the glycerol-3-phosphate shuttle were downregulated. TSS results also indicate the generation of acetyl-CoA from the fatty acids through the peroxisomal β -oxidation pathway (Figure 2C), indicating the possibility that fatty acids could be a subsidiary intracellular carbon source in xylose medium. Such supply of acetyl-CoA in xylose medium might result in more NADH production and generate more ATP, which is required for phosphorylation of xylulose and dihydroxyacetone. Additional ATP may also be supplied as a result of the *DAK1* upregulation in the cytoplasm (Figure 2A).

Discussion

As the first step to understanding the genetic basis of the highly efficient yeast *K. marxianus*, complete genome sequence and transcriptome analyses of its most thermotolerant strain were performed. The former analysis revealed many gene resources for the ability to assimilate various sugars in addition to species-specific genes. The latter clarified the molecular basis of attractive traits of the yeast. All information obtained here about the yeast will be useful for fundamental and applied research for innovative applications.

The thermotolerance as an attractive trait was investigated under the 45D condition, which revealed that *K. marxianus* seems to drastically change metabolic pathways from those under the 30D condition, that is, the enhancement of PPP and the attenuation of TCA cycle after the fumarate-producing step. The changes lead to the speculation that the former provides NADPH for scavenging ROS and that the latter deals with H_2O_2 via the electron transfer from Sdh to cytochrome *c* peroxidase (Figures 2 and 3). Consistent with these conjectures, a higher temperature generates more ROS, which causes

DNA damage [36]. Notably, the findings of the upregulation of genes for DNA double-stranded break repair and removal of uracil in DNA molecules suggest the occurrence of enhancement of double-stranded break or deamination of cytosines in DNA at high temperatures. In addition, ATP synthesis via oxidative phosphorylation may be greatly reduced due to the repression of *ATP3* for the gamma subunit of ATP synthase. The existence of additional strategies is guessed from the TSS analysis data for survival at high temperature: alteration of ribosome biogenesis including pre-rRNA processing presumably for stable and efficient protein synthesis, reduction of mitochondrial ribosome biogenesis probably for saving energy, reinforcement of checkpoints of DNA replication and spindle assembly, minimization of electron leakage in respiratory chain by reduction of NADH dehydrogenase, and Coenzyme Q or enhanced expression of Hsps and chaperones. Taken together, the thermotolerance of *K. marxianus* is likely achieved by systematic mechanisms consisting of various strategies. Especially, the yeast would mainly acquire ATP from glycolysis rather than TCA cycle at high temperatures, which could prevent the generation of ROS by minimization of mitochondrial activity.

Under a static condition, the growth and ethanol production of *K. marxianus* were low compared to those under a shaking condition [4], probably due to low ATP yield in mitochondria, which may be related to the enhanced expression of *AAC*. *K. lactis* bearing a null mutation of *AAC2* exhibits growth defect on glycerol, galactose, maltose, and raffinose [37]. In addition, the expression of *RAG5* for hexokinase was relatively low. In *K. lactis*, *RAG5* mutations [38] abolish the expression of *RAG1* for a low-affinity glucose transporter [39] and decrease the level of the 2.0-kb mRNA species of *HGT1* for a high-affinity glucose transporter [40]. Furthermore, NADH would be accumulated in the cytoplasm because of the downregulation of *GAP1* for glyceraldehyde-3-phosphate dehydrogenase. Contrarily, respiratory genes kept their transcriptional levels similar to those under the 30D condition (Additional file 1: Figure S8). Such situations may raise reactive oxygen species from the respiratory chain to cause oxidative stress. Consistently, cytoplasmic oxidative stress response genes were relatively strongly expressed (Figure 3B), especially glutathione-related genes depicted at the cytoplasm side were highly induced. Almost *HSPs*, however, were not upregulated under the 30DS condition as under the 30D condition (Figure 3A). These metabolic activities may lead the low level of cell proliferation under a static condition.

Regulation of gene expression in response to the level of oxygen is achieved via several transcription factors. Under a static condition, *K. marxianus* seems to increase glucose metabolism and shift to fermentation, implying a

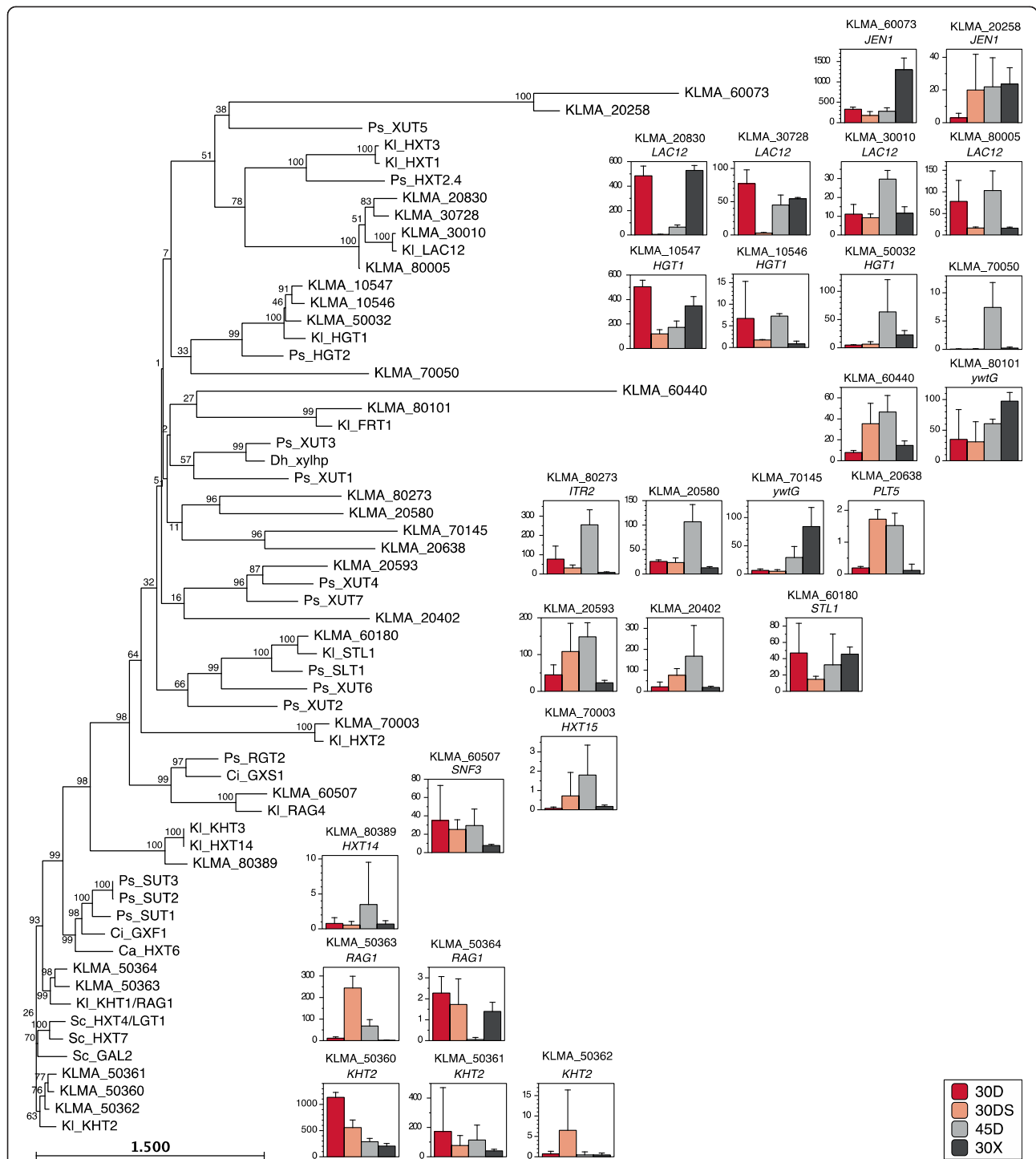


Figure 4 Phylogenetic tree of predicted sugar transporters with transcript abundance in *K. marxianus*. The phylogenetic tree was constructed with amino acid sequences of predicted sugar transporters by the neighbor-joining algorithm using CLC Sequence Viewer. Each column in the graph shows the transcript abundance of each gene as described in Figure 2. The accession numbers of amino acid sequences of sugar transporters are as follows: Ps_HGT2, XP_001382755; Sc_HXT7, NP_010629; Sc_GAL2, NP_013182; Dh_xylhp, AY347871; Ci_GXS1, GN107181; Ci_GXF1, GN107179; Ps_SUT1, XP_001387898; Ps_SUT2, XP_001384295; Ps_SUT3, XP_001386019; Ps_HXT2.4, XP_001387757; Ps_XUT1, XP_001385583; Ps_XUT2, XP_001387242; Ps_XUT3, XP_001387138; Ps_XUT4, XP_001386715; Ps_XUT5, XP_001385962; Ps_XUT6, XP_001386589; Ps_XUT7, XP_001387067; Ps_RGT2, XP_001386588; Ps_SLT1, XP_001383774; Ki_KHT1/RAG1, XP_453656; Ki_KHT2, GN107317; Ki_KHT3, XP_454897; Ki_FRT1, XP_454356; Ki_HGT1, XP_451484; Ki_HXT1, XP_455078; Ki_HXT14, XP_454897; Ki_HXT2, XP_453960; Ki_HXT3, XP_453088; Sc_HXT4/LGT1, NP_011960; Ki_STL1, XP_456249; Ki_RAG4, XP_455315; Ki_LAC12, XP_452193; Ca_HXT6, XP_719472.

connection between the oxygen- and glucose-sensing pathways. In *S. cerevisiae* and *K. lactis*, the transcription factor Hap1 mediates the induction of genes involved in their respiration, lipid metabolism, and oxidative stress response [11]. *K. lactis* Hap1 negatively regulates fermentation [41]. In contrast, the *HAP1* expression in *K. marxianus* was upregulated under the 30DS condition, and consistently, several genes related to ATP synthase and chaperones for respiratory chain components were upregulated (Additional file 1: Figure S8). These lines of evidence and the enhanced expression of genes for glycolytic pathway suggest differences in regulation of oxygen-responsive genes from those in *S. cerevisiae* and *K. lactis*.

The oxidative stress-response genes were found to be highly induced under the three conditions tested (Figures 2C and 4B,C,D,E), indicating that ROS is accumulated in the cytoplasm, mitochondria, and peroxisome under the 30DS and 30X conditions and in the cytoplasm and mitochondria under the 45D condition. Notably, Ahp1 in addition to Cta1 and Dot5 may be responsible for H₂O₂ detoxification in the peroxisome and nucleus, respectively.

Xylose assimilation capability as the second trait was examined under the 30X condition. It is known that *K. marxianus* tends to suffer from cofactor imbalance in xylose medium [42,43], and thus, its growth strongly depends on mitochondrial respiratory activity [44]. Interconvertibility of NAD⁺ species for maintaining the redox balance is essential for growth efficiency and metabolite excretion [45], but the yeast is incapable of directly converting NAD⁺ and NADPH into NADP⁺ and NADH owing to lack of transhydrogenases [46]. Instead, redox-balancing mechanisms between the cytoplasm and mitochondria are probably used to resolve the NADH/NADPH imbalance. Reoxidation of cytosolic NADPH and its strong connection to oxidative stress in *K. lactis* have been reported [47,48]. In *S. cerevisiae*, five cytosolic-mitochondrial redox shuttles have been proposed [49]. Of these, genes for enzymes related to ethanol-acetaldehyde, citrate-oxoglutarate, and oxaloacetate-malate shuttles were relatively upregulated under the 30X condition (Figures 2A, B). The GABA shunt from 2-oxoglutarate to succinate that has been proposed in *S. cerevisiae* and *S. stipitis* [21,50] may not be so important due to the low expression of related genes (Additional file 1: Figure S10), suggesting that *K. marxianus* uses different shuttles for resolving the cofactor imbalance from those of the two yeasts. In addition to TCA cycle intermediates by these shuttles, acetyl-CoA might be transferred from peroxisome on the basis of TSS data. Eventually, mitochondrial activity may be enhanced, which tends to increase the leakage of electrons to generate ROS, which is consistent with elevation of expression of oxidative stress-response genes.

The last trait is efficient protein productivity, which meets the demand for fast growth and high yield biomass [51]. The yeast has been exploited as a cell factory to obtain valuable enzymes, showing retention of activity over a large temperature range [52]. TSS results under the 30D and 30X conditions reveal high expression of *INUI* for inulinase, which is useful for the production of recombinant proteins in culture medium, as described in previous studies [26,27,53]. These useful characteristics may allow simultaneous production of ethanol and valuable proteins, thus, reducing the cost of ethanol production.

Conclusions

The complete sequences of *K. marxianus* DMKU 3-1042 nuclear and mitochondrial genomes have been determined, which reveal many genes for the cells to cope with a high temperature and to assimilate a wide variety of sugars including xylose and arabinose in addition to species-specific genes. The present study thus provides the molecular basis of attractive traits of the yeast for industrial applications including high-temperature fermentation, and information of its gene resources and transcriptome data, which are particularly useful for fundamental and applied researches for innovative applications.

Methods

Strains, media, and culture conditions

The yeast strain used in this work was *K. marxianus* DMKU 3-1042 strain, which has been deposited in the NITE Biological Resource Center (NBRC) under the deposit numbers NITE BP-283 and NBRC 104275. Media used were YP (1% w/v yeast extract and 2% w/v peptone) supplemented with one of two different carbon sources: YPD, with 2% w/v glucose, or YPX, with 2% w/v xylose.

Genome sequencing, assembly, and annotation

Two plasmid libraries with average insert sizes of 3 and 5 kb were generated in pTS1 (Nippon Gene, Tokyo, Japan) and pUC118 (Takara Bio. Inc., Otsu, Shiga, Japan) plasmid vectors, respectively, while a fosmid library with an average insert size of 40 kb was constructed in pCC1FOS (EPI-CENTRE, Illumina Inc., San Diego, CA, USA). Shotgun sequencing was performed on an ABI 3730xl DNA Analyzer (Applied Biosystems Co., Thermo Fisher Scientific, Inc., Foster City, CA, USA). Gaps were closed by the sequencing of gap-spanning PCR products. Telomeric regions were further analyzed by transposon-insertion sequencing of corresponding fosmid clones with Template Generation System II (Finnzymes, Thermo Fisher Scientific, Inc., Foster City, CA, USA). Genome assemblies were validated by optical mapping (OpGen, Gaithersburg, MD, USA). The genome was finally assembled into nine ungapped

contigs corresponding to eight chromosomes and a mitochondrion with an average coverage of 11.1x. The mean error rate was estimated to be less than 6×10^{-8} .

The rRNA-encoding regions were identified by a BLASTN program using the ribosomal sequences of *K. marxianus*, while the tRNA-coding regions were predicted by the ARAGORN program [54]. Proteins-coding gene prediction was performed by combining the Glimmer 3.02 program with a self-training dataset and six-frame prediction by using *in silico* Molecular Cloning (in silico biology, Inc., Yokohama, Kanagawa, Japan) and manual identification [55,56]. Intron prediction was performed by the AUGUSTUS program using the coding DNA sequences (CDSs) of *Ashbya gossypii* as the reference for pattern learning and manual correction of each CDS [57,58]. Functional annotation of the predicted CDSs was performed by BLASTP searching against the nonredundant (nr) database [59] with an E-value threshold of 10^{-10} . Protein domains were predicted using the InterProScan program against various domain libraries (Prints, Prosite, PFAM, ProDom, SMART). Protein functions were assigned by KOG database [60]. Assignment of UniProt number was performed by a BLASTP program against the UniProt database [61] with an E-value threshold of 10^{-10} . UniProt gene name, cellular localization, and Gene Ontology were assigned based on UniProt database information. The assignment of KO number of KEGG Orthology was performed by the KAAS program [62,63]. Individual annotations were then summarized according to KEGG Orthology and KEGG metabolic pathways.

Transcriptome analysis in *K. marxianus*

Cells grown in 100-ml Erlenmeyer flasks in 30 ml YPD medium at 30°C, 160 rpm for 18 h were inoculated at the initial OD₆₆₀ of about one into sequential batch culture, which was conducted in 300-ml Erlenmeyer flasks with 100 ml YPD at 30°C or 45°C under the shaking condition (30D and 45D) or the static condition at 30°C (30DS) or with 100 ml YPX at 30°C under the shaking condition (30X). The cells were further cultivated under each condition for 6 h and immediately subjected to RNA isolation. Each culture condition for TSS Seq experiments is under the same condition as performed previously [4]. Total RNA from cells was isolated by the hot phenol method [64] and was purified using an RNeasy Midi Kit (QIAGEN, Hilden, Germany) with RNase free DNase I (QIAGEN) according to the manufacturer's instructions. Transcriptional starting site (TSS) Seq analysis was performed using the extracted total RNA, providing precise information on TSSs and their expression levels in a high-throughput manner [65]. TSS-tag counts were divided by the total number of uniquely

and perfectly (with no mismatch) mapped TSS-tags to calculate TSS-tag ppm (parts per million). Experiments on each culture condition and TSS Seq analysis were performed in triplicate. Statistical testing was performed by edgeR of R package using the TSS-tag counts data of TSS analysis with a cutoff value as a false discovery rate smaller than 0.05 indicating significantly changed genes [66]. The GO enrichment test was performed by topGO of R package [67]. The TSS data was submitted to the Gene Expression Omnibus database (GEO, <http://www.ncbi.nlm.nih.gov/geo/>) under the accession number GSE66600.

Nucleotide sequence accession numbers

The complete genome sequence of *K. marxianus* DMKU 3-1042 has been deposited in DDBJ/EMBL/GenBank under accession no. AP012213-AP012221.

Quantitative real-time PCR (qPCR) analysis

Primers for qPCR were designed by using Primer Express software version 3.0 (Applied Biosystems Co.). A pair of primers, Km-rDNA-F1: 5'-GATCGGGTGGTGT TTTTCTTATG-3' and Km-rDNA-R1: 5'-TCCCCCA-GAACCCAAAG-3', was designed to amplify the 18S rDNA gene. The reaction produced a 71-bp PCR product. Probe for 18S rDNA, Km-rDNA-probe1: 5'-CCC ACTCGGCACCTTACGAGAAATCA-3' was labeled at the 5'-end with 6-carboxyfluorescein (FAM) as a reporter and at the 3'-end with dihydrocyclopyrroloindole tripeptide minor groove binder (MGB) as a quencher. Real-time PCR was performed using a TaqMan® Universal Master Mix II (Applied Biosystems Co.) and Applied Biosystems 7300 Real-Time PCR system (Applied Biosystems Co.). The 18S rDNA was amplified using genomic DNA isolated from 10 strains of *K. marxianus*. The C_T value was determined by the instrument's software and adjusted manually as necessary. Concentration and DNA quality were measured by using Qubit™ dsDNA HS Assay Kits (Invitrogen Ltd., Paisley, UK) with Qubit® Fluorometer and by gel electrophoresis and converted to the number of copies by using the molecular weight of the DNA. The equation $C_T = m (\log \text{quantity}) + b$ from the equation for a line ($y = mx + b$) was constructed by plotting the standard curve of log quantity versus its corresponding C_T value. The 18S rDNA copy numbers were determined by the absolute quantitation method, by which total copies were first calculated using the following equation: total 18S rDNA copies = $10^{((C_T - b)/m)}$. The number of 18S rDNA copies per genome was then determined by the following equation: 18S rDNA copies per genome = (Total copies of 18S rDNA)/(Total copy of genomic DNA). The genome size of 11.0 Mb of *K. marxianus* DMKU 3-1042 was used for all calculations.

Additional files

Additional file 1: Supplementary information. A file containing 10 supplementary figures and 22 supplementary tables.

Additional file 2: Gene annotation. A file that contains locus tags and their corresponding products.

Additional file 3: Transcriptome analysis results. A file that contains locus tags, their corresponding products and TSS-tag count in part per million (ppm) as used under the different conditions described in the study.

Abbreviations

1,3-DPG: 1,3-bisphosphoglycerate; 2PG: 2-phosphoglycerate; 3PG: 3-phosphoglycerate; 6P1,5R: 6-phospho-D-glucono-1,5-lactone; 6PG: 6-phosphogluconate; ADP: Adenosine diphosphate; ATP: Adenosine triphosphate; DHA: Dihydroxyacetone; DHAP: Dihydroxyacetone phosphate; E4P: Erythrose-4-phosphate; F6P: Fructose-6-phosphate; FDP: Fructose 1,6 bisphosphate; G6P: Glucose-6-phosphate; GAP: Glyceraldehyde-3-phosphate; Hsp: Heat shock protein; HTF: High-temperature fermentation; KOG: Eukaryotic orthologous groups; NAD: Nicotinamide adenine dinucleotide; NADP: Nicotinamide adenine dinucleotide phosphate; NBRC: NITE Biological Resource Center; Ndh: NADH dehydrogenase; NR: Nonredundant; PEP: Phosphoenolpyruvate; PPP: Pentose phosphate pathway; PYR: Pyruvate; qPCR: quantitative real-time PCR; R5P: Ribose-5-phosphate; rDNA: ribosomal DNA; ROS: Reactive oxygen species; Ru5P: Ribulose-5-phosphate; S7P: Sedoheptulose-7-phosphate; Sdh: Succinate dehydrogenase; TSS Seq: Transcription start site sequencing; Xul5P: Xylulose-5-phosphate.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

AH, N Fujit, and KT performed a whole-genome sequencing and optical mapping. NL, TK, NR, M Mu, N Fujim, and S contributed to the gene annotation. YS performed the transcription start site (TSS) analysis. M Ma and TK contributed to the analysis of TSS data and most of other bioinformatics analysis. NL performed most of other wet lab experiments. SL provided the *K. marxianus* DMKU 3-1042 strain and its general information. NL, TK, and MY prepared the manuscript. MY organized this research project. All authors have read and approved the final version of the manuscript.

Acknowledgements

We thank Drs. Shimoi, Matsushita, and Akada for their helpful discussions. This work was supported by the Program for Promotion of Basic Research Activities for Innovative Biosciences, New Energy and Industrial Technology Development Organization (NEDO), the Special Coordination Funds for Promoting Science & Technology, Ministry of Education, Culture, Sports, Science & Technology (MEXT), and the Advanced Low Carbon Technology Research and Development Program, Japan Science and Technology Agency (JST). This work was partially performed as a collaborative research in the Asian Core Program and in the Core to Core Program, which was supported by the Scientific Cooperation Program agreed by the Japan Society for the Promotion of Science (JSPS), the National Research Council of Thailand (NRCT), and universities involved in the program.

Author details

¹Applied Molecular Bioscience, Graduate School of Medicine, Yamaguchi University, Ube 755-8505, Japan. ²Department of Biological Chemistry, Faculty of Agriculture, Yamaguchi University, Yamaguchi 753-8515, Japan. ³National Institute of Technology and Evaluation, Shibuya-ku, Tokyo 151-0066, Japan. ⁴Department of Medical Genome Sciences, The University of Tokyo, Chiba 277-8562, Japan. ⁵Department of Microbiology, Faculty of Science, Kasetsart University, Bangkok 10900, Thailand. ⁶Department of Biology, Faculty of Science, Chiang Mai University, Chiang Mai 50200, Thailand.

Received: 16 September 2014 Accepted: 19 February 2015

Published online: 18 March 2015

References

- Hahn-Hägerdal B, Galbe M, Gorwa-Grauslund MF, Lidén G, Zacchi G. Bio-ethanol – the fuel of tomorrow from the residues of today. *Trends Biotechnol.* 2006;24:549–56.
- Banat IM, Nigam P, Singh D, Marchant R, McHale AP. Review: ethanol production at elevated temperatures and alcohol concentrations: Part I – Yeasts in general. *World J Microbiol Biotechnol.* 1998;14:809–21.
- Limtong S, Sringiew C, Yongmanitchai W. Production of fuel ethanol at high temperature from sugar cane juice by a newly isolated *Kluyveromyces marxianus*. *Biores Technol.* 2007;98:3367–74.
- Rodrussamee N, Lertwattanasakul N, Hirata K, Suprayogi, Limtong S, Kosaka T, et al. Growth and ethanol fermentation ability on hexose and pentose sugars and glucose effect under various conditions in thermotolerant yeast *Kluyveromyces marxianus*. *Appl Microbiol Biotechnol.* 2011;90:1573–86.
- Nonklang S, Abdel-Banat BM, Cha-aim K, Moonjai N, Hoshida H, Limtong S, et al. High-temperature ethanol fermentation and transformation with linear DNA in the thermotolerant yeast *Kluyveromyces marxianus* DMKU 3–1042. *Appl Environ Microbiol.* 2008;74:7514–21.
- Lachance MA. *Kluyveromyces van der Walt*. In: Kurtzman CP, Fell JW, Boekhout T, editors. *The Yeasts. A Taxonomic Study*. 5th ed. Amsterdam: Elsevier; 2011. p. 471–81.
- Llorente B, Malpertuy A, Blandin G, Artiguenave F, Wincker P, Dujon B. Genomic exploration of the hemiascomycetous yeasts: 12. *Kluyveromyces marxianus* var. *marxianus*. *FEBS Lett.* 2000;487:71–5.
- Fukuhara H. *Kluyveromyces lactis* - a retrospective. *FEMS Yeast Res.* 2006;6:323–4.
- Schaffrath R, Breunig KD. Genetics and molecular physiology of the yeast *Kluyveromyces lactis*. *Fungal Genet Biol.* 2000;30:173–90.
- González-Siso MI, Freire-Picos MA, Ramil E, González-Domínguez M, Rodríguez Torres A, Cerdán ME. Respirofermentative metabolism in *Kluyveromyces lactis*: insights and perspectives. *Enzyme Microb Technol.* 2000;26:699–705.
- Rodicio R, Heinisch JJ. Yeast on the milky way: genetics, physiology and biotechnology of *Kluyveromyces lactis*. *Yeast.* 2013;30:165–77.
- Lertwattanasakul N, Rodrussamee N, Suprayogi, Limtong S, Thanonkeo P, Kosaka T, et al. Utilization capability of sucrose, raffinose and inulin and its less-sensitiveness to glucose repression in thermotolerant yeast *Kluyveromyces marxianus* DMKU 3–1042. *AMB Express.* 2011;1:20.
- Fonseca GG, Heinzle E, Wittmann C, Gombert AK. The yeast *Kluyveromyces marxianus* and its biotechnological potential. *Appl Microbiol Biotechnol.* 2008;79:339–54.
- dos Santos VC, Bragança CRS, Passos FJV, Passos FML. Kinetics of growth and ethanol formation from a mix of glucose/xylose substrate by *Kluyveromyces marxianus* UFV-3. *Antonie Van Leeuwenhoek.* 2013;103:153–61.
- Abdel-Banat BM, Nonklang S, Hoshida H, Akada R. Random and targeted gene integrations through the control of non-homologous end joining in the yeast *Kluyveromyces marxianus*. *Yeast.* 2010;27:29–39.
- Hoshida H, Murakami N, Suzuki A, Tamura R, Asakawa J, Abdel-Banat BM, et al. Non-homologous end joining-mediated functional marker selection for DNA cloning in the yeast *Kluyveromyces marxianus*. *Yeast.* 2014;31:29–46.
- Jeong H, Lee DH, Kim SH, Kim HJ, Lee K, Song JY, et al. Genome sequence of the thermotolerant yeast *Kluyveromyces marxianus* var. *marxianus* KCTC 17555. *Eukaryot Cell.* 2012;11:1584–5.
- Suzuki T, Hoshino T, Matsushika A. Draft genome sequence of *Kluyveromyces marxianus* strain DMB1, isolated from sugarcane bagasse hydrolysate. *Genome Announcements.* 2014;2:e00733–14.
- Dietrich FS, Voegeli S, Brachat S, Lerch A, Gates K, Steiner S, et al. The *Ashbya gossypii* genome as a tool for mapping the ancient *Saccharomyces cerevisiae* genome. *Science.* 2004;304:304–7.
- Dujon B, Sherman D, Fischer G, Durrens P, Casaregola S, Lafontaine I, et al. Genome evolution in yeasts. *Nature.* 2004;430:35–44.
- Jeffries TW, Grigoriev IV, Grimwood J, Laplaza JM, Aerts A, Salamov A, et al. Genome sequence of the lignocellulose-bioconverting and xylose-fermenting yeast *Pichia stipitis*. *Nat Biotechnol.* 2007;25:319–26.
- Butler G, Rasmussen MD, Lin MF, Santos MA, Sakthikumar S, Munro CA, et al. Evolution of pathogenicity and sexual reproduction in eight *Candida* genomes. *Nature.* 2009;459:657–62.
- Wood V, Gwilliam R, Rajandream MA, Lyne M, Lyne R, Stewart A, et al. The genome sequence of *Schizosaccharomyces pombe*. *Nature.* 2002;415:871–80.
- Rubio-Teixeira M. Endless versatility in the biotechnological applications of *Kluyveromyces LAC* genes. *Biotechnol Adv.* 2006;24:212–25.

25. Lane MM, Morrissey JP. *Kluyveromyces marxianus*: a yeast emerging from its sister's shadow. *Fungal Biol Rev.* 2010;24:17–26.
26. Rocha SN, Abrahão-Neto J, Cerdán ME, Gombert AK, González-Siso MI. Heterologous expression of a thermophilic esterase in *Kluyveromyces* yeasts. *Appl Microbiol Biotechnol.* 2011;89:375–85.
27. Rocha SN, Abrahão-Neto J, Cerdán ME, González-Siso MI, Gombert AK. Heterologous expression of glucose oxidase in the yeast *Kluyveromyces marxianus*. *Microb Cell Fact.* 2010;9:4.
28. van Ooyen AJ, Dekker P, Huang M, Olsthoorn MM, Jacobs DI, Colussi PA, et al. Heterologous protein production in the yeast *Kluyveromyces lactis*. *FEMS Yeast Res.* 2006;6:381–92.
29. Abranches J, Mendonça-Hagler LC, Hagler AN, Morais PB, Rosa CA. The incidence of killer activity and extracellular proteases in tropical yeast communities. *Can J Microbiol.* 1997;43:328–36.
30. Jablonowski D, Schaffrath R. Zymocin, a composite chitinase and tRNase killer toxin from yeast. *Biochem Soc Trans.* 2007;35:1533–7.
31. Matsutani M, Hirakawa H, Yakushi T, Matsushita K. Genome-wide phylogenetic analysis of *Gluconobacter*, *Acetobacter*, and *Gluconacetobacter*. *FEMS Microbiol Lett.* 2010;315:122–8.
32. Codón AC, Gasent-Ramírez JM, Benítez T. Factors which affect the frequency of sporulation and tetrad formation in *Saccharomyces cerevisiae* baker's yeasts. *Appl Environ Microbiol.* 1995;61:630–8.
33. Snoek ISI, Steensma HY. Why does *Kluyveromyces lactis* not grow under anaerobic conditions? Comparison of essential anaerobic genes of *Saccharomyces cerevisiae* with the *Kluyveromyces lactis* genome. *FEMS Yeast Res.* 2006;6:393–403.
34. Kwast KE, Lai LC, Menda N, James III DT, Aref S, Burke PV. Genomic analyses of anaerobically induced genes in *Saccharomyces cerevisiae*: functional roles of Rox1 and other factors in mediating the anoxic response. *J Bacteriol.* 2002;184:250–65.
35. Scott EM, Pillus L. Homocitrate synthase connects amino acid metabolism to chromatin functions through Esa1 and DNA damage. *Genes Dev.* 2010;24:1903–13.
36. Hori A, Yoshida M, Shibata T, Ling F. Reactive oxygen species regulate DNA copy number in isolated yeast mitochondria by triggering recombination-mediated replication. *Nucleic Acids Res.* 2009;37:749–61.
37. Flores CL, Rodríguez C, Petit T, Gancedo C. Carbohydrate and energy-yielding metabolism in non-conventional yeasts. *FEMS Microbiol Rev.* 2000;24:507–29.
38. Prior C, Mamessier P, Fukuhara H, Chen XJ, Wésolowski-Louvel M. The hexokinase gene is required for transcriptional regulation of the glucose transporter gene *RAG1* in *Kluyveromyces lactis*. *Mol Cell Biol.* 1993;13:3882–9.
39. Chen XJ, Wésolowski-Louvel M, Fukuhara H. Glucose transport in the yeast *Kluyveromyces lactis*. II. Transcriptional regulation of the glucose transporter gene *RAG1*. *Mol Gen Genet.* 1992;233:97–105.
40. Billard P, Menart S, Blaissonneau J, Bolotin-Fukuhara M, Fukuhara H, Wésolowski-Louvel M. Glucose uptake in *Kluyveromyces lactis*: role of the *HGT1* gene in glucose transport. *J Bacteriol.* 1996;178:5860–6.
41. Bao WG, Guiard B, Fang ZA, Donnini C, Gervais M, Passos FML, et al. Oxygen-dependent transcriptional regulator Hap1p limits glucose uptake by repressing the expression of the major glucose transporter gene *RAG1* in *Kluyveromyces lactis*. *Eukaryot Cell.* 2008;7:1895–905.
42. Zhang B, Zhang L, Wang D, Gao X, Hong J. Identification of a xylose reductase gene in the xylose metabolic pathway of *Kluyveromyces marxianus* NBRC1777. *J Ind Microbiol Biotechnol.* 2011;38:2001–10.
43. Lulu L, Ling Z, Dongmei W, Xiaolian G, Hisanori T, Hidehiko K, et al. Identification of a xylitol dehydrogenase gene from *Kluyveromyces marxianus* NBRC1777. *Mol Biotechnol.* 2003;53:159–69.
44. Lertwattanasakul N, Suprayogi, Murata M, Rodrussamee N, Limtong S, Kosaka T, et al. Essentiality of respiratory activity for pentose utilization in thermotolerant yeast *Kluyveromyces marxianus* DMKU 3–1042. *Antonie Van Leeuwenhoek.* 2013;103:933–45.
45. Marres CA, de Vries S, Grivell LA. Isolation and inactivation of the nuclear gene encoding the rotenone-insensitive internal NADH: ubiquinone oxidoreductase of mitochondria from *Saccharomyces cerevisiae*. *Eur J Biochem.* 1991;195:857–62.
46. van Dijken JP, Scheffers WA. Redox balances in the metabolism of sugars by yeasts. *FEMS Microbiol Rev.* 1986;32:199–224.
47. Tarrío N, Becerra M, Cerdan ME, Gonzalez-Siso MI. Reoxidation of cytosolic NADPH in *Kluyveromyces lactis*. *FEMS Yeast Res.* 2006;6:371–80.
48. Gonzalez-Siso MI, Garcia-Leiro A, Tarrío N, Cerdan ME. Sugar metabolism, redox balance and oxidative stress response in the respiratory yeast *Kluyveromyces lactis*. *Microb Cell Fact.* 2009;8:46.
49. Bakker BM, Overkamp KM, van Maris AJ, Kötter P, Luttk MA, van Dijken JP, et al. Stoichiometry and compartmentation of NADH metabolism in *Saccharomyces cerevisiae*. *FEMS Microbiol Rev.* 2001;25:15–37.
50. Cao J, Barbosa JM, Singh NK, Locy RD. GABA shunt mediates thermotolerance in *Saccharomyces cerevisiae* by reducing reactive oxygen production. *Yeast.* 2013;30:129–44.
51. Groeneveld P, Stouthamer AH, Westerhoff HV. Super life – how and why 'cell selection' leads to the fastest-growing eukaryote. *FEBS J.* 2009;276:254–70.
52. Foukis A, Stergiou P, Theodorou LG, Papagianni M, Papamichael EM. Purification, kinetic characterization and properties of a novel thermotolerant extracellular protease from *Kluyveromyces marxianus* IFO 0288 with potential biotechnological interest. *Biores Technol.* 2012;123:214–20.
53. Raimondi S, Uccelletti D, Amaretti A, Leonardi A, Palleschi C, Rossi M. Secretion of *Kluyveromyces lactis* Cu/Zn SOD: strategies for enhanced production. *Appl Microbiol Biotechnol.* 2010;86:871–8.
54. Laslett D, Canback B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.* 2004;32:11–6.
55. Salzberg SL, Delcher AL, Kasif S, White O. Microbial gene identification using interpolated Markov models. *Nucleic Acids Res.* 1998;26:544–8.
56. Delcher AL, Bratke KA, Powers EC, Salzberg SL. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics.* 2007;23:673–9.
57. Stanke M, Waack S. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics.* 2003;19:i215–25.
58. Stanke M, Schöffmann O, Morgenstern B, Waack S. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics.* 2006;7:62.
59. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997;25:3389–402.
60. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, et al. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics.* 2003;4:41.
61. Apweiler R, O'nonov C, Magrane M, Alam-Faruque Y, Antunes R, Bely B, et al. Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Res.* 2012;40:D71–5.
62. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* 2007;35 Suppl 2:W182–5.
63. Aoki-Kinoshita KF, Kanehisa M. Gene annotation and pathway mapping in KEGG. *Methods Mol Biol.* 2007;396:71–91.
64. Aiba H, Adhya S, de Crombrughe B. Evidence for two functional gal promoters in intact *Escherichia coli* cells. *J Biol Chem.* 1981;256:11905–10.
65. Tsuchihara K, Suzuki Y, Wakaguri H, Irie T, Tanimoto K, Hashimoto S, et al. Massive transcriptional start site analysis of human genes in hypoxia cells. *Nucleic Acids Res.* 2009;37:2249–63.
66. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2010;26:139–40.
67. Alexa A, Rahnenfuhrer J. topGO: Enrichment Analysis for Gene Ontology. In: R Package Version 2. 2000.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

