

RESEARCH ARTICLE

Open Access

Genome sequencing and genetic breeding of a bioethanol *Saccharomyces cerevisiae* strain YJS329

Dao-Qiong Zheng, Pin-Mei Wang, Jie Chen, Ke Zhang, Tian-Zhe Liu, Xue-Chang Wu*, Yu-Dong Li* and Yu-Hua Zhao

Abstract

Background: Environmental stresses and inhibitors encountered by *Saccharomyces cerevisiae* strains are the main limiting factors in bioethanol fermentation. Strains with different genetic backgrounds usually show diverse stress tolerance responses. An understanding of the mechanisms underlying these phenotypic diversities within *S. cerevisiae* populations could guide the construction of strains with desired traits.

Results: We explored the genetic characteristics of the bioethanol *S. cerevisiae* strain YJS329 and elucidated how genetic variations in its genome were correlated with specified traits compared to similar traits in the S288c-derived strain, BYZ1. Karyotypic electrophoresis combined with array-comparative genomic hybridization indicated that YJS329 was a diploid strain with a relatively constant genome as a result of the fewer Ty elements and lack of structural polymorphisms between homologous chromosomes that it contained. By comparing the sequence with the S288c genome, a total of 64,998 SNPs, 7,093 indels and 11 unique genes were identified in the genome of YJS329-derived haploid strain YJSH1 through whole-genome sequencing. Transcription comparison using RNA-Seq identified which of the differentially expressed genes were the main contributors to the phenotypic differences between YJS329 and BYZ1. By combining the results obtained from the genome sequences and the transcriptions, we predicted how the SNPs, indels and chromosomal copy number variations may affect the mRNA expression profiles and phenotypes of the yeast strains. Furthermore, some genetic breeding strategies to improve the adaptabilities of YJS329 were designed and experimentally verified.

Conclusions: Through comparative functional genomic analysis, we have provided some insights into the mechanisms underlying the specific traits of the bioethanol strain YJS329. The work reported here has not only enriched the available genetic resources of yeast but has also indicated how functional genomic studies can be used to improve genetic breeding in yeast.

Keywords: Bioethanol, *Saccharomyces cerevisiae*, Stress, Genome, RNA-Seq

Background

Bioethanol is an important adjunct to fossil fuel because it is renewable, relatively environmentally innocuous, and compatible with the current fuel transport facilities. To date, bioethanol is mainly produced through the yeast-based fermentation of carbohydrates at about 33°C to give a final product concentration of 8–15% (v/v) [1,2]. Some novel processes, including high-gravity fermentation, high-temperature fermentation, and production from cellulose, intended to increase the economic and social

benefits of ethanol, have been proposed and widely studied [2-6]. These processes, however, share the problem that they impose severe environmental stresses or inhibitors on yeast cells which greatly reduces their production efficiency. In addition, these stresses induce the formation of more by-products (mainly glycerol and acetic acid), consuming up to 5% of the carbon source [2-5].

The *Saccharomyces cerevisiae* strain S288c was, in 1996, the first eukaryotic genome to be sequenced [7]. In the 15 years that have passed since then, many functional genomic studies using the S288c genome as a reference sequence have greatly enriched our knowledge of how yeast cells respond to and resist various environmental stresses [8-16]. The information that has been

* Correspondence: mblab@163.com; youdng@gmail.com
Institute of Microbiology, College of Life Sciences, Zhejiang University,
Hangzhou 310058, Zhejiang Province, P R China

produced cannot always be extrapolated to other yeast strains because of their diverse genomes and phenotypes [8,17,18]. Compared with laboratory strains, industrial strains generally show higher adaptability to specific environments; however, the genetic basis for their improved characteristics is not well understood. Comparisons of the genomes of strains with different backgrounds should help identify the sequence changes that play important roles in the tolerance of particular stresses. Because of the progress in genome sequencing technology, some industrial yeast strains, including AWRI1631, EC1118, JAY270, Vin13 and FostersO, have now been sequenced [19,20]. Comparisons of the publicly available *S. cerevisiae* genome sequences have revealed the clear signatures (single nucleotide polymorphisms (SNPs), insertions and deletions (indels), and novel ORFs) of different strains [18,20,21]. However, further studies are needed to explore how the genetic variations confer the specific phenotype of each strain. Of these industrial strains, JAY270 (PE-2 derived) which uses sugar cane as feedstock, is the only bioethanol strain [1]. Little is known about the genome structure and characteristics of other bioethanol strains.

In this study, we investigated the genetic characteristics of a bioethanol strain, YJS329, and the molecular mechanisms that underlie its phenotypic differences from the laboratory strain, BYZ1 (S288c-derived). YJS329 exceeded BYZ1 in fermentation rate and ethanol yield under different stress conditions, consistent with its greater tolerance of multiple stresses. Comparative genomic hybridization array and whole genome sequencing revealed many differences in the genomes of these two strains, including SNPs, indels, novel ORFs and changes in chromosome structure. Finally, we used RNA-Seq to determine how the genetic differences might affect the transcriptional profile and physiological metabolism of the two strains. Our study enriches the genetic resources for *S. cerevisiae* and deepens our knowledge of the effects of genetic variation on phenotypic diversity.

Results

Phenotypic and physiological characteristics of YJS329

In comparisons of fermentation performance, YJS329 had a slightly higher fermentation rate than BYZ1 but they each produced similar amounts of ethanol in a 38-h period under standard conditions (Table 1 and Additional file 1). At higher temperatures and under higher gravity conditions, the ethanol yield of YJS329 was 16.6% and 12.1% (t test, $P < 0.001$) higher than that of BYZ1, respectively (Table 1 and Additional file 1). In addition, under the three fermentation conditions tested, YJS329 produced more glycerol, whereas BYZ1 produced more acetic acid (Table 1). Consistent with the fermentation tests, YJS329 grew faster than BYZ1 when exposed to stress factors

(ethanol, high temperature, osmotic stress, and oxidative stress; Figure 1A) and YJS329 also exceeded BYZ1 in tolerance to the furan derivative hydroxymethylfurfural (HMF), a major inhibitory compound in the fermentation of lignocellulosic hydrolysates.

We compared YJS329 and BYZ1 using some of the main anti-stress indicators, including trehalose accumulation, antioxidation factors, HMF reductase, and membrane compositions. YJS329 accumulated 1.29-fold (t test, $P < 0.05$) more intracellular trehalose, a nonspecific protectant that can maintain the function of macromolecules and membrane integrity under multiple stresses (Figure 1B) [22]. Consistent with its better menadione tolerance, YJS329 showed 1.32-fold (t test, $P < 0.05$) higher glutathione content and 5-fold (t test, $P < 0.001$) catalase (CAT) activity than BYZ1. In yeast cells, glutathione and CAT are important for the elimination of the reactive oxygen species that are caused by oxidizing agents or by other stresses [23]. HMF is formed as a result of hexose degradation during the process of lignocellulosic hydrolysis [24]. The chemical toxicity of HMF can be reduced by HMF reductase which converts the aldehyde functional group into an alcohol group in yeast cells [24,25]. Compared to BYZ1, the higher intracellular HMF reductase activity (t test, $P < 0.05$; Figure 1B) of YJS329 might partly contribute to its increased resistance to HMF. The results in Figure 1B show that, of the various membrane compounds, more ergosterol, palmitoleic acid ($C_{16:1}$), oleic acid ($C_{18:1}$), and linoleic acid ($C_{18:2}$) were detected in YJS329 (t test, $P < 0.05$). These findings indicated that there was significant variation in cellular components and physiological state between the YJS329 and BYZ1 strains.

Genome structure of YJS329

The DNA content of YJS329 was less than that of a triploid strain ZTW3 but close to that of BYZ1 (Figure 1C). After being grown in sporulation medium for 3–5 days, YJS329 showed an overall sporulation efficiency of 92%, producing mostly asci with two or three ascospores (Figure 1D). The pulse-field gel electrophoresis (PFGE) results revealed that YJS329 and BYZ1 differed distinctly in the length of their chromosomes; the exceptions were chromosomes 9, 10 and 14 (Figure 2A). The karyotype of YJS329 is more regular than the karyotypes of some other industrial strains [1,20], because the two homologs of each of the YJS329 chromosomes were the same length. The array-comparative genomic hybridization revealed that there were no big chromosomal aberrations in the genome of YJS329. The regions of the chromosomes that were underrepresented in the YJS329 genome (the green regions in Figure 2B) compared with in BYZ1 contain 267 ORFs (Figure 2C and Additional file 2). Most of these ORFs are located near the

Table 1 Performance of the yeast strains, BYZ1 and YJS329, under different fermentation conditions

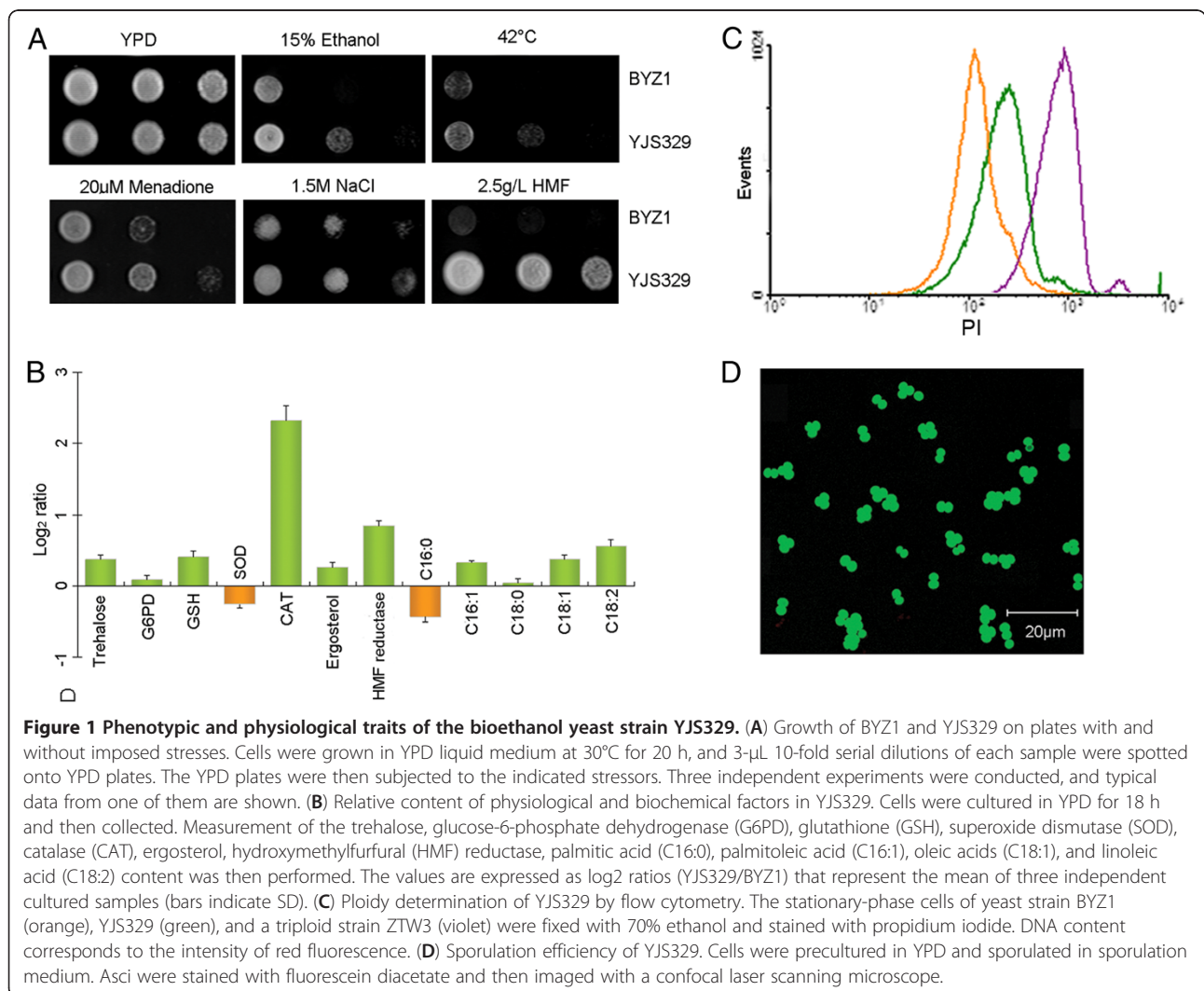
Fermentation conditions	Strain	Ethanol produced (g/L)	Glycerol produced (g/L)	Acetic acid produced (g/L)	Residual glucose (g/L)	Dry weight (g/L)
Normal	BYZ1	71.89 ± 0.67	3.95 ± 0.08	1.06 ± 0.02	0.13 ± 0.00	9.4 ± 0.14
	YJS329	71.15 ± 0.72	4.36 ± 0.07 ^a	0.44 ± 0.01 ^A	0.03 ± 0.00	9.1 ± 1.02
High temperature	BYZ1	60.18 ± 0.53	3.68 ± 0.09	1.21 ± 0.02	25.07 ± 0.34	3.9 ± 0.11
	YJS329	70.16 ± 0.55 ^B	4.90 ± 0.07 ^B	0.78 ± 0.01 ^B	1.11 ± 0.04 ^B	3.5 ± 0.09 ^b
High gravity	BYZ1	111.58 ± 1.19	5.93 ± 0.04	1.50 ± 0.04	21.45 ± 0.25	9.2 ± 0.13
	YJS329	125.10 ± 1.13 ^C	9.04 ± 0.14 ^C	0.61 ± 0.02 ^C	1.24 ± 0.01 ^C	8.8 ± 0.08 ^c

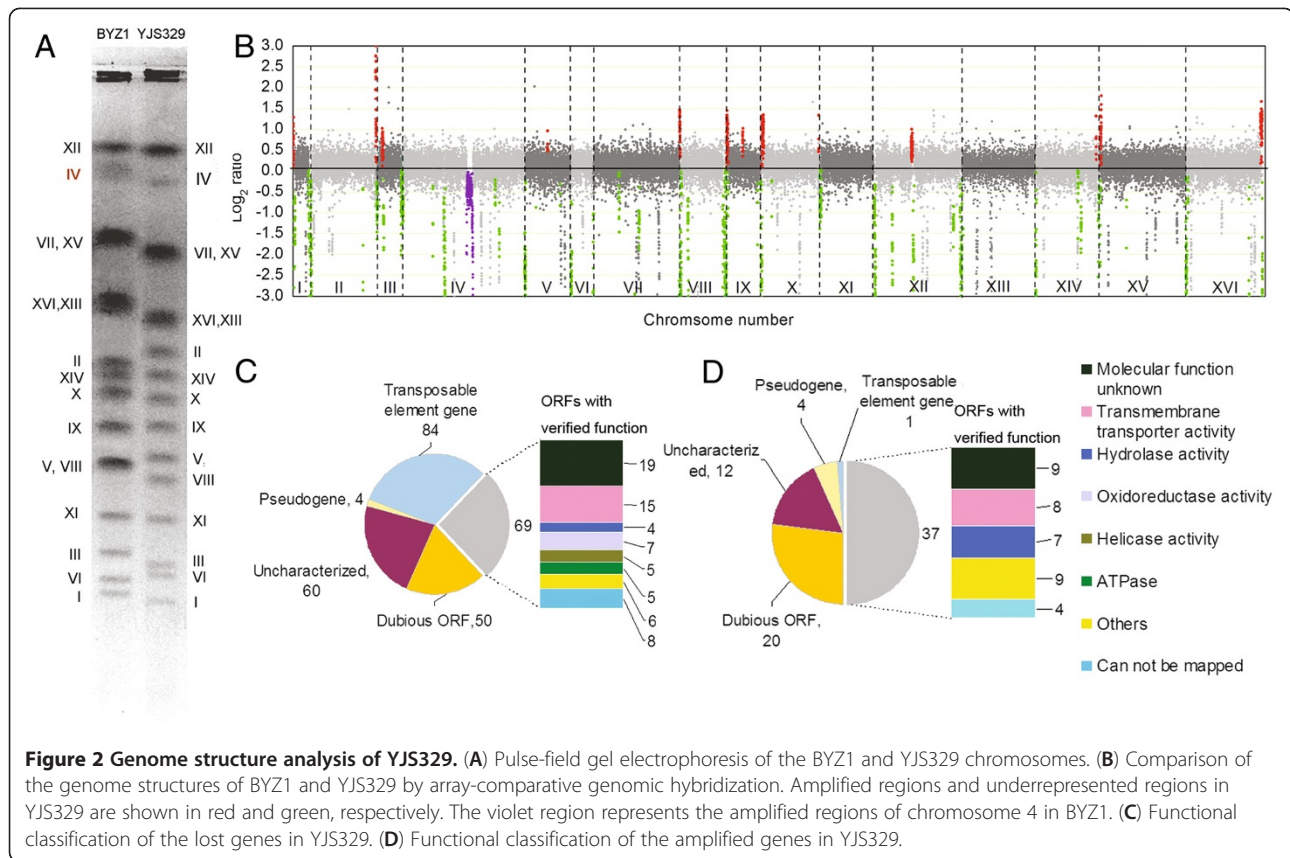
^{a-c}indicate a significant difference between YJS329 and BYZ1 under normal (letter a), high temperature (letter b), and high gravity (letter c) conditions at the 0.05 level using *t* test.

^{A-C}indicate a significant difference between YJS329 and BYZ1 under normal (letter A), high temperature (letter B), and high gravity (letter C) conditions at the 0.01 level using *t* test.

telomeres, long terminal repeat retrotransposons, or on tandemly repeated arrays. The regions of the chromosomes that were amplified in YJS329 relative to BYZ1 are shown in red in Figure 2B. Expressed products were identified for up to 50% of the ORFs in the amplified

segments (Figure 2D and Additional file 2). The expressed genes include three hexose transport genes (*HXT8*, *HXT9*, and *HXT11*), four genes involved in maltose metabolism (*MAL12*, *MAL31*, *MAL32*, and *MAL33*), and four alpha-glucosidase genes (*IMA2-5*).





The region of chromosome 4 that was amplified in BYZ1 (shown in purple in Figure 2B) led to the size differences between the homologs of chromosome 4 in this strain (Figure 2A). The RT-qPCR results confirmed that this amplification was present in the parent strain BY4742 before the generation of BYZ1 in the present work (See Additional file 3). This rearrangement was apparently Ty-derived as this region is flanked by the Ty elements YDR180W-A and YDRCTy1-3. Although the laboratory strains BY4741 and BY4742 have been used extensively in genetic research, this amplification has not been reported until now.

Whole genome sequencing of YJS329

To investigate the genetic traits of YJS329, we isolated the haploid strain YJSH1 which, under certain conditions, is indistinguishable in ethanol yields from its parent strain YJS329 (See Additional file 4), for whole genome sequencing (See Additional file 5).

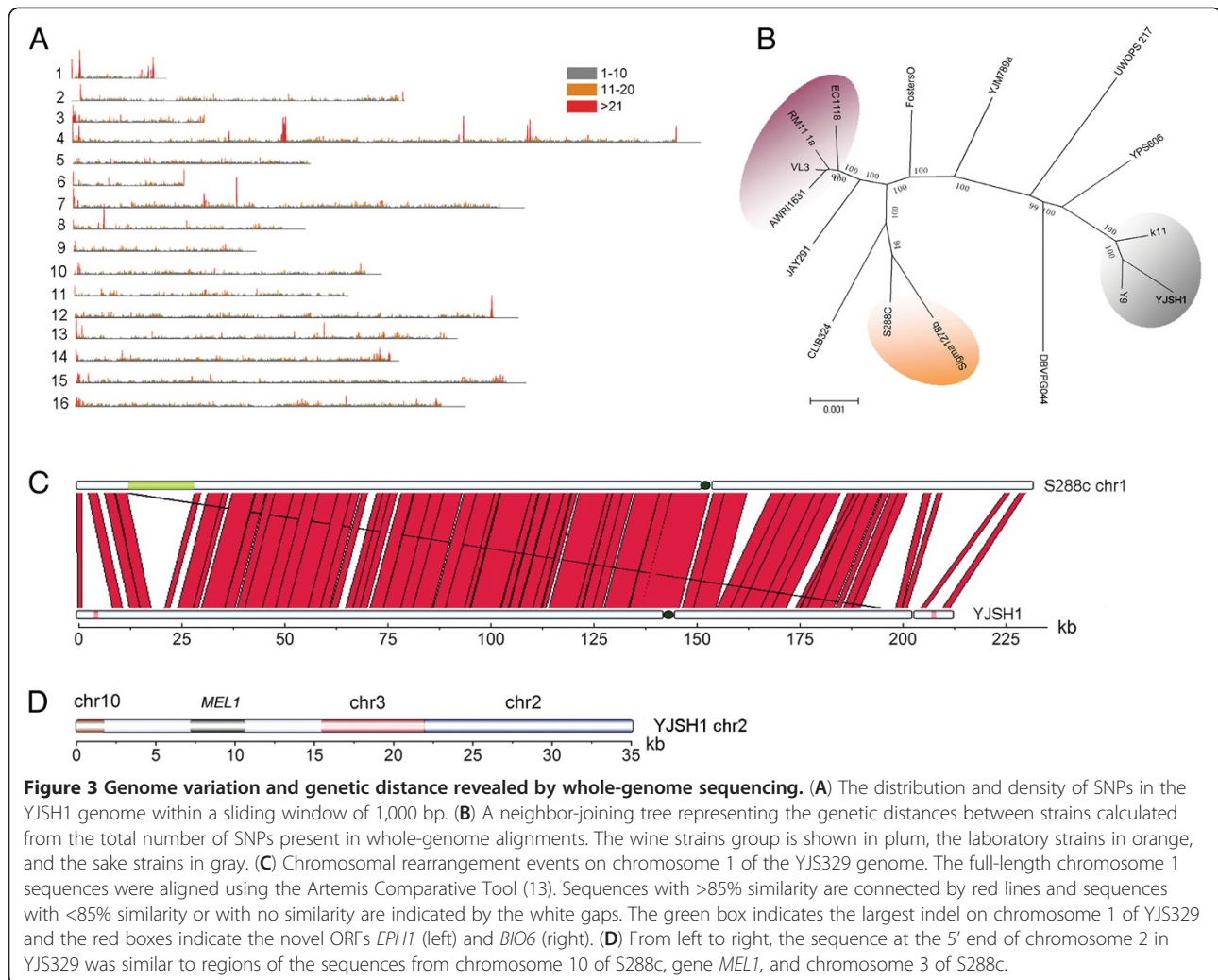
SNPs

We identified 64,998 SNPs within the aligned regions of the YJSH1 and S288c genomes (the location of the SNPs and their annotations are listed in Additional file 6). The average SNP density was 5.73 per kilobase throughout the genome but the density was not constant across

individual chromosomes (Additional file 5 and Figure 3A). A total of 39,098 SNPs were found in the ORFs and 38.7% of them resulted in non-synonymous mutations. We observed that genes (e.g. *HXT6*, *HXT7*, and *ARO3*) with redundant functions tended to accumulate more SNPs, which was consistent with their lower hybridization signals in the array-comparative genomic hybridization. Using the number of SNPs separating any two isolates as an estimation of their relatedness, we constructed a neighbor-joining tree that represented the genetic distances among 16 yeast strains. The tree shows that the bioethanol strains JAY291 and YJS329 displayed the closest evolutionary relatedness to the wine and sake strains, respectively (Figure 3B).

Indels

Based on the consensus YJSH1 genomic sequence, 412,794 bp that were absent in YJSH1 were identified in the S288c genome and 174,269 bp that were absent in S288c were identified in the YJSH1 genome (the location of the indels and their annotations were listed in Additional file 6). This analysis confirmed that some of the underrepresented regions in YJS329 genome (Figure 2B) were sequences that either were lost in this industrial strain or acquired in S288c. For example, the YJS329 genome had only one copy of *CUP1* and *ENA1*, and



none of the *ASP3* genes found in S288c. We also identified 21 Ty elements in the YJ5329 assembly (9 Ty1, 6 Ty2, 4 Ty3, 1 Ty4, and 1 Ty5), whereas 50 Ty elements have been identified in the S288c genome. The amplification of the Ty3 elements was consistent with the results of comparative genome hybridization for YJ5329 (See Additional file 2).

ORFs

A total of 5,602 ORFs (common to S288c and excluding dubious ORFs) were predicted for the nuclear genome of YJ5329 (the location of the ORFs and their annotations were listed in Additional file 7). Predictions indicated that 142 ORFs had in-frame stop codons, 129 ORF were affected by frame shifts, and 27 ORFs had lost start or stop codons because of the presence of SNPs or indels. For example, the *HO* gene of YJ5329 had both an in-frame termination (the C at 238 bp was changed to T) and frame shift (the C at 413 bp was missing) (verified by PCR using YJ5329 DNA as the template) that explained the

heterothallic life cycle of YJ5329. In addition, the YJ5329 genome has some ORF sequences that were not present in S288c (Additional file 7); however, nearly all of these ORFs could be found in the genomes of other *S. cerevisiae* strains. One such example is the ORF *EPH1* that encodes the epoxide hydrolase (E.C. 3.3.2.3) that catalyzes the hydration of chemically reactive epoxides to their corresponding dihydrodiol products. A recent study suggested that *EPH1* in the *S. cerevisiae* genome was the result of an introgression event from *S. paradoxus* and the *S. paradoxus* *EPH1* gene may itself be a result of horizontal transfer from bacteria [26].

Structural variations

Compared to the strictly diploid *S. cerevisiae* S288c, many industrial yeast strains display chromosomal copy number variations (CNVs). Whole-chromosome amplifications had been observed in the AWRI796, VL3, FostersO and FostersB strains [20]. Although no large chromosomal aneuploidy or length polymorphisms were

observed in the genome of YJSH1, some chromosomal rearrangement events in the YJSH1 genome were observed. The largest indel in the YJS329 genome was the 12.5-kb deletion in chromosome 1 region (11,872–24,331 bp; Figure 3C). The 5' end of chromosome 2 in YJS329 was apparently subjected to constant remodeling (verified by PCR using YJS329 DNA as the template). In this region two elements from the S288c genome, chromosome 10 (729,223–727,336 bp) and chromosome 3 (315506–307348 bp), and a region that is absent in S288c genome (a BLASTN search showed that this region contained a *MEL1* gene that has been found in *S. carlsbergensis* and in other *S. cerevisiae* strains), were found in YJSH1 (1–23,308 bp; Figure 3D).

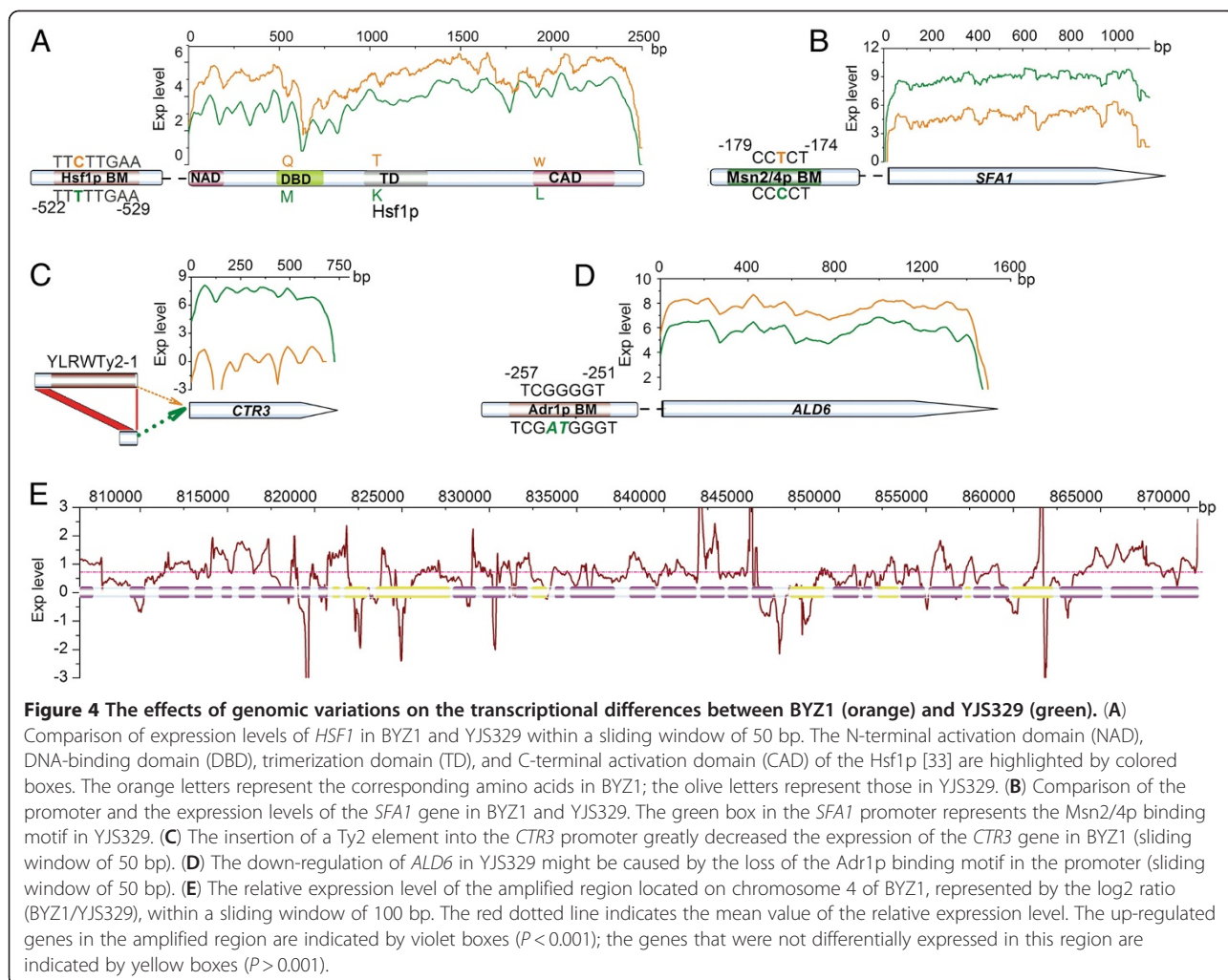
Comparison of BYZ1 and YJS329 transcription using RNA-Seq

To investigate transcription differences at single-nucleotide resolution between BYZ1 and YJS329, poly (A)-enriched mRNAs from BYZ1 and YJS329 were used for high-throughput Illumina sequencing. Overall, 90.9% of the reads mapped to unique genomic regions; 81% mapped to known reference genes when 2-bp mismatches were allowed (Additional file 8). Compared to BYZ1, 888 of the YJS329 genes were up-regulated and 1,433 were down-regulated ($P < 0.001$; Additional file 9). The functions of the up-regulated genes mainly fell within the oxidoreductase, peptidase activity and transporter-related processes categories (Additional file 10). For example, *SFA1* which is involved in the detoxification of formaldehyde and long-chain and complex alcohols formation [24,27] displayed more than a 15-fold increase in mRNA abundance in YJS329. The fair number of the up-expressed genes involved in transport processes in the YJS329 sample suggested that this strain might have higher adaptability to multiple nutrition shortages than BYZ1. The down-regulated genes were mainly involved in the functional categories of DNA/protein binding, ribosome biogenesis, and structural molecules (Additional file 10).

Among these differentially expressed genes, we focused specifically on the transcriptional activity of the genes that are closely related to the anti-stress factors. Consistent with the analyses at the physiological and biochemical levels, the genes in ergosterol and fatty acid biosynthesis, and the genes encoding catalases were highly expressed but to different degrees in YJS329 (Additional file 11). We also found five transcription factors (*HAPI*, *MSN2/4*, *ARR1*, and *HSF1*) which are known to be major regulators that control critical cellular processes and response to environmental conditions [12,13,16], that displayed significantly different expression patterns in the two strains (Additional file 10). Transcription regulation network analyses (Additional

file 10) revealed that a large proportion of the up-regulated genes (*TPS2* and *TSL1* in trehalose metabolism, *OLE1* and *ELO1* in oleic acid biosynthesis, and the catalase coding gene *CTT1*) was regulated by the Msn2/4p transcription factor, whose expression is itself dependent on or induced by other transcription factors (such as Hap1p) [12]. Although the zinc-finger transcription factor *HAPI* has a larger number of reads per kilobase of exon region per million mapped reads (RPKM) in BYZ1, the Hap1 protein is inactivated by a Ty1 insertion in the carboxy terminus [28]. The absence of this interrupting Ty1 element in the YJS329 protein may explain why the *HAPI*-regulated genes involved in the synthesis of fatty acids and ergosterol, such as *FAS1*, *FAS2*, *ERG2*, *ERG5*, *ERG11*, and *ERG25*, were expressed at a higher level in this strain (t test, $P < 0.001$). Except for *ZIMI7*, most of the genes that code for heat-shock proteins and the transcription factor Hsf1p showed less mRNA expression activity in YJS329 compared to BYZ1 (Additional file 10). An in-vitro experiment showed that the efficiency of the HSF1-promoter in YJS329 was 16% lower than in BYZ1 (t test, $P < 0.05$; Additional file 12). Compared to BYZ1, a SNP in the *HSF1* promoter in YJS329 resulted in the loss of the Hsf1p binding motif which may be important for the variations in *HSF1* expression (Figure 4A). Furthermore, three amino acid substitutions in the functional domains of Hsf1p may impede its interaction with the promoters of heat shock proteins; however, this supposition needs further experimental verification.

As well as the destruction of binding motifs in transcription factors, SNPs can also create new binding motifs. The Msn2/4p and Cat8p binding sites in the promoter of *SFA1* from YJS329 are examples of new motifs that may strengthen the expression of the *SFA1* gene (Figure 4B and Additional file 12) which plays a role in the detoxification of furan derivatives [24]. Indels were also important contributors to transcription differentiation among the two strains. An obvious example in BYZ1 is the interruption of *CTR3* (which encodes a high-affinity copper transporter responsible for copper uptake when environmental copper is low [29]) by the insertion of a Ty2 element [30]. This insertion might explain the much lower expression activity of *CTR3* in BYZ1 compared to YJS329 (Figure 4C). Further, small indels in the trans-elements can directly modify mRNA expression and phenotypic traits in different strains. The down-regulated expression of *ALD6* in YJS329 (whether grown in YPD medium or under fermentation conditions and verified by RT-qPCR; t test, $P < 0.001$), a major gene in acetic-acid formation, probably resulted partly from the insertion of two bases in the Adr1p binding motif in the *ALD6* promoter (Figure 4D and Additional file 12). In BYZ1, when the two copies of *ALD6* were deleted, the



strain produced 56% less acetic acid and 17% more glycerol under the normal fermentation conditions (Additional file 13). This result indicated that the lower expression of *ALD6* in YJS329 could be one of the causes of the different patterns of by-product (acetic acid and glycerol) production in YJS329 and BYZ1. Chromosomal aneuploidy accompanied by CNVs in large DNA regions is a ubiquitous phenomenon in yeast populations [20,31]. As indicated in Figure 4E, the expression levels of regions with CNVs apparently dependent on gene dosage. The average read depth of the amplified region on chromosome 4 of BYZ1 was 1.59 times that in YJS329, close to the increased DNA dosage.

Using RNA-Seq, we detected the expression of the unique ORFs at the whole-transcription-profile level. Among these ORFs (Annotation details are in Additional file 7), *MEL1* had the highest RPKM; others, such as *YJM-GNAT* [32], showed minimal expression. Additional file 13 shows the expression level and boundary of the predicted ORF *chr06.orf003*, which provides further evidence of the existence of this novel ORF which is absent

in other *S. cerevisiae* strains. RT-qPCR analyses revealed that the expression of some unique ORFs depended on the growth phase and other conditions (Additional file 14). When grown in YPD medium, all five of the selected genes (especially *BIO6*) showed the highest expression at the exponential phase. The ORFs *YJS-HE* and *MEL1* were significantly up-regulated under ethanol fermentation, whereas the others were down regulated, indicating the different psychological roles of these unique genes.

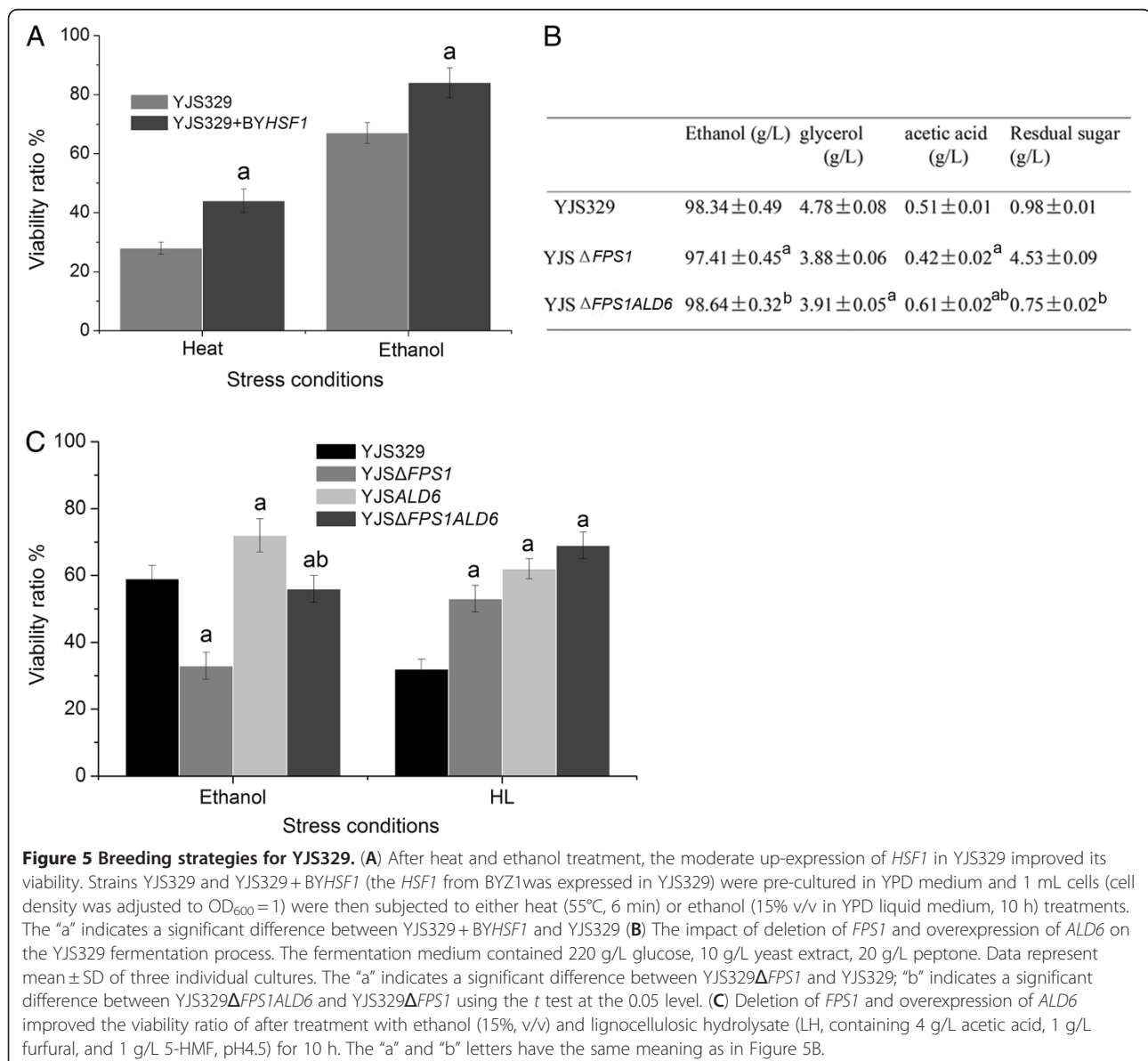
Genetic breeding strategies for YJS329

Hsf1p is a conserved transcription factor that regulates hundreds of targets in response to multiple stresses [33]. Optimized expression of Hsf1p is important for yeast cells because either the deletion or overexpression of this gene leads to growth arrest [15]. To evaluate whether the lower expression activity of Hsf1p and related heat shock proteins was beneficial or detrimental to YJS329 under stress conditions, we expressed the *Hsf1p* gene from BYZ1 in YJS329 using a low-copy

plasmid. This genetic manipulation enhanced the cell viability of YJS329 by 57% and 25% after heat or ethanol treatment (*t* test, $P < 0.05$; Figure 5A), respectively, indicating that the appropriate readjustment of the expression of important transcription factors can contribute to the adaptability of yeast strains.

More glycerol might improve the taste of alcoholic beverages but is undesirable for bioethanol production. When *FPS1* (involved in efflux of glycerol; this gene showed lower expression in YJS329 compared with BYZ1) was deleted in YJS329 to produce the YJS Δ *FPS1* strain, the production of glycerol and acetic acid decreased and the conversion rate of glucose to ethanol improved by 1% compared with YJS329; however, the final concentration of ethanol was slightly less than in

YJS329 because of the higher residual sugar in YJS Δ *FPS1* (*t* test, $P < 0.05$; Figure 5B). Inspired by the different regulatory roles of *ALD6* in YJS329 and BYZ1, we explored the possibility to further reduce the production of glycerol in YJS Δ *FPS1* by overexpression of *ALD6*. Beyond our expectation, strain YJS Δ *FPS1ALD6* produced similar amounts of glycerol but 1.3% more ethanol (*t* test, $P < 0.05$) than YJS Δ *FPS1* as a result of consuming more sugar than YJS Δ *FPS1*. We found that the over-expression of *ALD6* could enhance the tolerance of ethanol in both YJS329 and YJS Δ *FPS1* (*t* test, $P < 0.05$; Figure 5C), which may explain the higher fermentation ability of strain YJS Δ *FPS1ALD6*. In addition, the over-expression of *ALD6* and deletion of *FPS1* significantly improved the tolerance of lignocellulosic hydrolysate (LH,



contains inhibitors acetic acid, furan, and 5-HMF) in YJS329 (*t* test, $P < 0.05$; Figure 5C), suggesting that this strategy may be useful for breeding industrial yeast strains with the ability to increase ethanol production from lignocellulosic biomass.

Discussion

The genomic structural analysis (DNA content, PFGE, and aCGH analysis) indicated that YJS329 retained a diploid karyotype and had much lower structural polymorphisms than the bioethanol strain JAY270 and some other industrial strains [1,20]. We also sequenced the genome of YJSH2 (a haploid spore derived from the same tetrad as YJSH1) using the Illumina paired-ends method. After mapping the reads of YJSH2 to the YJSH1 genome, we estimated that the YJS329 genome had about 0.6 SNP/kb between allelic regions in homologous chromosomes (unpublished data). These results indicated that the YJS329 strain was genetically very stable, a desirable phenotype for industry practice. Although S288c has been widely used in scientific research, because of the high number of Ty elements, its genome seems to be more plastic [31,34]. High expression activity of Ty elements in genes was confirmed in the S288c-derived strain BYZ1 as a result of a dose effect (Additional file 4). The duplicated region on chromosome 4 in BYZ1 is probably the result of chromosomal translocations by ectopic recombination mediated by the flanking Ty elements. Strikingly, no dosage-compensation mechanisms acted to normalize the expression from each gene because the higher expression (1.59-fold) of this duplicated region almost matched the higher gene dose (1.5-fold). These results indicated that spontaneous Ty-driven rearrangements could be quite common and, if ignored, could easily lead to incorrect experimental results in genetic studies, especially for the S288c-derived strains.

Second-generation sequencing technology has proven to be an effective tool for the investigation of the genome sequences and structures of yeast strains and has provided many new insights into genome evolution and phenotypic effects [1,17,20,21,35,36]. The level of nucleotide polymorphisms between YJSH1 and S288c (0.57%) is very similar to the level separating S288c and AWRI1631 (wine strain), YJM789 (pathogenic strain), M22 (vineyard strain) or YPS163 (oak tree strain) [21,36], but, interestingly, YJSH1 was grouped closely with sake strains, consistent with their geographical distributions. To the best of our knowledge, YJS329 is the first bioethanol strain for which a high-quality assembled genome has been completed. The SNPs and indels that we have identified in the aligned regions of YJSH1 and S288c constitute the main genome mutations in these two strains. Mutation frequencies were found to be

higher in the intergenic regions than in the coding regions, we found that up to 40% of the SNPs and 88% of the indels were located in intergenic sequences (accounting for about 27% of the genome). This pattern could arise from the sequence characteristics of intergenic regions (for example: the abundance of repeated sequences). However, we also observed a considerable number of mutations in the ORFs that play important roles in specified physiological activities. Remedying some of these mutations may improve the capabilities or change the specified phenotype of YJS329. A total of 11 ORFs were predicted in the YJS329 genome that are absent from the S288c genome. Remarkably, some of these ORFs may be very similar to those in other *Saccharomyces* species, including *S. paradoxus*, *S. carlsbergensis*, and *S. mikatae*. Therefore, during the evolution of the YJS329 genome, repeated yeast hybridization events that were followed by the gradual loss of one of the contributing genomes might have occurred. Undoubtedly, the genotypic characteristics of YJS329 that have been revealed in the present study will enrich the genetic resources of this species, which will be valuable for breeding strains with the desired phenotypes.

The recently developed RNA-Seq approach was used to explore the transcription profiles of the YJS329 and BYZ1 *S. cerevisiae* strains. Among the 2,611 differently expressed genes in these two strains, many were involved in the trehalose metabolism pathways, antioxidative factors, and membrane composition biosynthesis that are closely related to multiple stress-tolerance and fermentation characteristics. For example, consistent with the higher oleic acid content of membranes, the genes encoding the subunits of fatty acid synthetase (*FAS1* and *FAS2*), the acetyl-CoA carboxylase gene (*ACCI*), and the genes that function in fatty-acid desaturation and elongation (*ELO1* and *OLE1*) were considerably up-regulated in YJS329. Our results indicated that most of the differences in the physiological factors were consistent with the mRNA transcription differences between these two strains. Transcription –regulatory network analyses revealed that the transcription factors Msn2/4p, Hap1p, Hsf1p, and Arr1p might give prominence to the differently expressed genes and phenotypic differences between the two strains. This result was consistent with the observation that the *trans* variation is more common in expression polymorphism in yeast [37–39]. In spite of this, the contributions of *cis* variations on the divergence of mRNA expression and physiological metabolism should not be neglected because our results confirmed that mutations in the promoters of some important transcription factors and genes could directly affect the efficiency of their promoter efficiency. Overall, the molecular mechanisms underlying the mRNA expression differences between

YJS329 and BYZ1 might involve: (i) SNPs and indels in the cis-acting elements that affect the expression efficiency of the genes; (ii) the inactivation of transcription factors by SNPs or indels; and (iii) changes in gene copy number. Remarkably, the discrepancies between the transcriptional profile (for example, of Hap1p) and the phenotype in the two strains might reflect variations in the activities of homologous proteins or posttranscriptional regulation, which deserve further assessment. In addition, here, for the first time, the expression activities of some novel ORFs under different conditions have been determined. Our study shows that whole-genome sequencing combined with RNA-Seq is a powerful tool for linking genotypes and phenotypes in functional genomic studies.

Conclusions

A thorough understanding of the genetic variations and how these variations contribute to phenotypic diversities is vital for the development of excellent yeasts for industrial applications. In this study, functional genomics has revealed the genetic characteristics of a bioethanol strain YJS329 and compared it to the laboratory strain BYZ1. From the results of this study, targeted genetic strategies for YJS329 could be constructed. These strategies might include the introduction of wild type genes to remedy deleterious mutations in some of the strains, a heightening of the effects of beneficial mutations by gene deletion or overexpression, and the expression of novel genes to obtain specified functions. We expect that functional genomics studies of industrial microorganisms, such as those reported here, will, in the future, provide more effective means of improving breeding strategies to obtain the desired production traits.

Methods

Yeast strains and culture conditions

The S288c-isogenic strain BYZ1 (*MATa/MAT α his3 Δ 1/his3 Δ 1 leu2 Δ 0/leu2 Δ 0 lys2 Δ 0/+ met15 Δ 0/+ ura3 Δ 0/ura3 Δ 0*) was generated from a cross between BY4741 and BY4742 (gift from Oliver Valerius, University of Göttingen, Germany). The yeast strain YJS329 (CCTCC 2011275) was isolated from a soil sample and was used for bioethanol production in Henan Tianguan Group Co., Ltd., China. Strain ZTW3 is a triploid strain that is stored in our laboratory. The growth medium (YPD) contained 10 g/L yeast extract, 20 g/L peptone, and 20 g/L glucose and had a pH of 5.5.

Fermentation test

The fermentation medium contained 10/L yeast extract, 20 g/L peptone, and 160 or 280 g/L glucose. Yeast cells were precultured in YPD for 20 h at 30°C and transferred to the fermentation medium with an initial OD₆₀₀ of 1. Three fermentation conditions were used: (i)

160 g/L glucose at 30°C; (ii) 160 g/L glucose at 40°C; and (iii) 280 g/L glucose at 30°C. Glucose and ethanol were measured as previously described [3].

Analyses of physiological and biochemical factors

Yeast cells were cultured in 25 mL YPD with an initial OD₆₀₀ of 0.05 and then collected at the early stationary phase (18 h, most genes involved in the stress response are induced at this phase). Trehalose, catalase, superoxide dismutase, and ergosterol were measured as previously described [3]. Glutathione was measured using a Glutathione Assay Kit according to the manufacturer's instructions (Nanjing Jiancheng Bioengineering Institute, China). Fatty acid was extracted by the method of Hama et al. [40] and then analyzed with a FOCUS GC Gas Chromatograph [41].

PFGE and Array-comparative genomic hybridization

Yeast chromosomes were prepared as described by Argueso et al. [42] and separated by PFGE as described previously [41].

Total genomic DNA from BYZ1 and YJS329 was isolated with the yeast DNA kit (OMEGA, GA, USA) and then sonicated. The shearing DNA (200–1000 bp) was labeled with Cy5/Cy3 and hybridized to *S. cerevisiae* CGH 385 K Whole-Genome Tiling Arrays (NimbleGen). Scanning was performed with the Axon GenePix 4000B Microarray Scanner (Axon, USA). Raw data were extracted as pair files using NimbleScan software. Log₂-ratio data were calculated and normalized by spatial correction and qspline fit normalization. DNA segments that contained three or more continuous probes with CNVs ($|\text{Log}_2\text{-ratio}| \geq 0.35$) were considered over- or under-represented regions. The microarray data have been deposited in the NCBI Gene Expression Omnibus [GEO:GSE31872].

Whole genome sequencing and data analysis

Strain YJS329 was previously cultured in sporulation medium for 5 days, and an ascus with four ascospores was dissected to obtain four haploid strains (named YJSH1-4). YJSH1 was chosen for genome sequencing. Whole genome sequencing was performed on the 454 Life Sciences Genome Sequencer FLX (Roche) platform according to the manufacturer's standard recommended sample preparation procedures. A shotgun sequencing library was constructed and a total of 718,904 reads were generated. 98.01% of the reads were assembled into 314 contigs using the Newbler software with the default parameters (minimum overlap length 40, minimum overlap identity 90%). The assembled sequences were manually checked, and some of the gaps were closed by Sanger sequencing reactions (contigs were first mapped to the corresponding chromosome and the sequences in

gaps were amplified by PCR) to build the scaffolds. The 16 nuclear YJSH1 chromosomes were covered by 16 scaffolds including 30 contigs (Additional file 5). The sequences of the final contigs and scaffolds have been deposited with DDBJ/EMBL/GenBank under the Whole Genome Shotgun project [GenBank:AGAW00000000]. The version of the sequences described here is the first version of the sequences [GenBank:AGAW01000000].

SNPs were detected using the public BLASTN software [43] after the YJSH1 contig sequences were aligned to the individual S288c chromosome sequences [32]. The BLASTN parameters were adjusted as match=4, mismatch=-5, gapopen=3, gapextend=5. Indels between the YJSH1 scaffolds and S288c chromosomes were detected using BLAT [44] (with default parameter) to reveal the physical gaps. The sizes and types (deletion or insertion) of indels were identified using the block sizes, qstarts, and tstarts information in the BLAT results file. Potential ORFs were predicted in two steps: (i) direct mapping of S288c ORFs from the *Saccharomyces* genome database by BLAT with the match length >95%, and (ii) using the Glimmer software (with the default parameters) to predict the ORFs located in unaligned regions of the YJSH1 contigs and S288c chromosomes [45]. The predicted ORFs were annotated by searching for their homologs in the NCBI non-redundant protein database. To predict structural variations, the YJSH1 scaffolds were aligned to the S288c chromosomes using the Artemis Comparative Tool [46]. The YJSH1 sequences that could not be aligned to the S288c genome were then compared against the contigs in the Whole Genome Shotgun database using BLASTN. Finally, PCRs were used to verify the predicted structural variations.

RNA-Seq

The total RNA of each sample (three individual cultures of yeast cells) was extracted by the hot phenol method (growth conditions and time of extraction were identical to those used in the physiological factor analysis). cDNA libraries were prepared using the methods described by Pan and co-workers [47]. The cDNA library products were sequenced on the Illumina HiSeq™ 2000. The raw Illumina sequencing data have been deposited in NCBI's GEO database [GEO:GSE31601]. After removing reads containing sequencing adapters and reads of low quality (reads in which the percentage of low quality bases (quality value ≤5) was more than 50%), the remaining clear reads were aligned to the *S. cerevisiae* S288c or YJSH1 genes with SOAPAligner [48]. The expression level was normalized by reads per kilobase of exon region per million mapped reads (RPKM) [49]. Screening of differentially expressed genes and *P*-value calculations were performed using the method proposed by Audic

and Claverie [50]. The accuracy of the RNA-Seq experiment was verified by RT-qPCR.

RT-qPCR

RNA extraction and quantitative PCR were performed as described by Tao et al. [41]. The primers that were used for quantitative PCR are listed in Additional file 15.

Promoter efficiency evaluation

The promoters of *HSF1* (826 bp), *SFA1* (1250 bp), and *ALD6* (1199 bp) from BYZ1 were cloned into Sac I and Xho I sites before the Cre gene of plasmid pSH47 [GenBank:AF298782.1]. Inverse PCR was used to introduce the sequence mutations of YJS329 shown in Figure 4. The efficiency of the promoters was evaluated by the expression activity (RT-qPCR) of the report gene Cre. The values were represented by the log₂ ratio of YJS329/BYZ1. The primers that were used for promoter cloning and RT-qPCR are listed in Additional file 15.

Genetic manipulation

The full-length *HSF1* ORF along with 807 bp of the sequence upstream of the ORF was cloned into the CEN6 plasmid, pGFP-ble (derived from pGFP-N-FUS; the *URA3* marker was replaced by *ble*^r). Deletion of the two copies of *FPS1* in YJS329 was performed as previously described [51]. In all cases, homozygous gene deletions were confirmed by diagnostic PCR. Overexpression of *ALD6* was carried out by cloning the *ALD6* ORF plus 1,005 bp of upstream sequence and 407 bp of downstream sequence into plasmid pYZ, which is derived from pYES2 (Invitrogen) but with *ble*^r replacing the *URA3* marker.

Additional files

Additional file 1: Comparison of fermentation rates (CO₂ production) of YJS329 (cycle) and BYZ1 (triangle). Fermentations were performed under (A) regular, (B) heat, and (C) high-gravity conditions motioned in the section of Material and Methods.

Additional file 2: Comparison of the regions with copy number variations (CNVs) between YJS329 and BYZ1. Format: XLS.

Additional file 3: Verification of the amplification of the DNA region of chromosome 4 in BY4742 genome. Two pairs of primers (sequences were showed in Additional file 15) specified to genes *HMO1* and *UME6* were designed to verify the copy number variations of the ~60 kb region of chromosome 4 in BY4741, YJS329, and BY4742 genomes by RT-qPCR.

Additional file 4: Comparison of ethanol yield of YJS329 and YJSH1. Fermentations were performed under regular, high gravity, and heat conditions described in the Methods.

Additional file 5: Status and distribution of polymorphisms in each of the YJS329 chromosomes.

Additional file 6: Details of the sequence variations detected in the YJSH1 genome.

Additional file 7: Details of the YJSH1 gene annotations.

Additional file 8: RNA-seq reads mapping to S288c genome and genes.

Additional file 9: Differentially expressed genes of YJS329 and BYZ1 revealed by RNA-Seq.

Additional file 10: Functional classification and transcriptional-regulation analysis of genes expressed differently in BYZ1 and YJS329. (A) GO functional enrichment analysis of genes expressed differently in BYZ1 and YJS329 (FDR < 0.05). Orange pillars represent the classification of up-regulated genes in YJS329, and olive pillars represent the down-regulated genes. (B) Regulation network analysis of some key trans-transcriptional factors and their target genes. These genes were grouped into five terms marked with different color borders, including trehalose metabolism (black), antioxidative factors (green), heat-shock proteins (red), and fatty-acid and ergosterol metabolism (blue). Regulation relationships are presented by the arrows linking the nodes. The genes up-regulated or down-regulated genes with respect to BYZ1 are shown in red and green, respectively, and the color gradient represents the extent of regulation.

Additional file 11: Comparison of the expression levels of stress-related genes between BYZ1 and YJS329.

Additional file 12: The different efficiencies of the promoters of HSF1, SFA1, and ALD6 between BYZ1 and YJS329. The efficiency of the promoters was evaluated by the expression activity of report gene Cre. The values were represented by log₂ ratio of YJS329/BYZ1. Error bars represent SD of three independent samples. Format: DOC.

Additional file 13: The effects of ALD6 deletion on metabolites yield of ethanol fermentation. Yeast cells were precultured in YPD overnight, and were then transferred to the fermentation medium (10/L yeast extract, 20 g/L peptone, and 160 g/L glucose) with the initial OD₆₀₀ of 1. Fermentations were performed at 30°C for 55 h.

Additional file 14: Verification of the transcription of some novel genes. (A) The expression level and boundary of the novel ORF chr06.003. (B) Relative expression of five novel ORFs under different conditions. YJS329 was grown in YPD medium with initial OD₆₀₀ of 0.05, and total RNA were then extracted at 7 h (exponential phase), 15 h (diauxic growth), and 25 h (stationary) for determination of the expression of these novel genes. "Fermentation" indicates the total RNA extracted at 20 h under ethanol-fermentation conditions (33°C) with corn mash as the feedstock (containing 270 g/L glucose).

Additional file 15: Primers used in this study.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

ZDQ and WXC designed the study and drafted the manuscript. ZDQ, WPM and LYD carried out the genome sequencing and molecular genetic studies. LTZ, LP, CJ and ZYH participated in the design of the study and performed the physiological and chemical analysis. All authors read and approved the final manuscript.

Acknowledgements

We thank Ming-Guan Feng, Hai-Chun Gao, Xiao-Hang Ma, Gen-Fu Wu and Zhen-Mei Lv (Institute of Microbiology, Zhejiang University, China); and Mu-Yuan Zhu (Institute of Genetics, Zhejiang University, China) for excellent technical assistance. This work was supported in part by a grant from the National Natural Science Foundation of China (No.31101339).

Received: 3 May 2012 Accepted: 14 August 2012

Published: 15 September 2012

References

- Argueso JL, Carazzolle MF, Mieczkowski PA, Duarte FM, Netto OVC, Missawa SK, Galzerani F, Costa GGL, Vidal RO, Noronha MF, Dominska M, Andrietta MGS, Andrietta SR, Cunha AF, Gomes LH, Tavares FCA, Alcarde AR, Dietrich FS, McCusker JH, Petes TD, Pereira GAG: **Genome structure of a *Saccharomyces cerevisiae* strain widely used in bioethanol production.** *Genome Res* 2009, **19**(12):2258–2270.
- Kollaras A, Kavanagh JM, Bell GL, Purkovic D, Mandarakas S, Arcenal P, Ng WS, Routledge KS, Selwood DH, Koutouridis P, Paras FE, Milic P, Tirado-Escobar ES, Moore MJ, Bell PJ, Atfield PV: **Techno-economic implications of improved high gravity corn mash fermentation.** *Bioresour Technol* 2011, **102**(16):7521–7525.
- Zheng DQ, Wu XC, Tao XL, Wang PM, Li P, Chi XQ, Li YD, Yan QF, Zhao YH: **Screening and construction of *Saccharomyces cerevisiae* strains with improved multi-tolerance and bioethanol fermentation performance.** *Bioresour Technol* 2010, **102**(3):3020–3027.
- Abdel-Banat BMA, Hoshida H, Ano A, Nonklang S, Akada R: **High-temperature fermentation: how can processes for ethanol production at high temperatures become superior to the traditional process using mesophilic yeast?** *Appl Microbiol Biotechnol* 2010, **85**(4):861–867.
- Almeida JRM, Runquist D, Nogue VSI, Liden G, Gorwa-Grauslund MF: **Stress-related challenges in pentose fermentation to ethanol by the yeast *Saccharomyces cerevisiae*.** *Biotechnol J* 2011, **6**(3):286–299.
- Nakamura T, Watanabe T, Srichuwong S, Arakane M, Tamiya S, Yoshinaga M, Watanabe I, Yamamoto M, Ando A, Tokuyasu K: **Selection of stress-tolerant yeasts for simultaneous saccharification and fermentation (SSF) of very high gravity (VHG) potato mash to ethanol.** *Bioresour Technol* 2010, **101**(24):9710–9714.
- Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, Louis EJ, Mewes HW, Murakami Y, Philippsen P, Tettelin H, Oliver SG: **Life with 6000 genes.** *Science* 1996, **274**(5287):546–567.
- Kvitek DJ, Will JL, Gasch AP: **Variations in stress sensitivity and genomic expression in diverse *S. cerevisiae* isolates.** *PLoS Genet* 2008, **4**(10):e1000223. doi:10.1371/journal.pgen.1000223.
- Ma MG, Liu ZL: **Comparative transcriptome profiling analyses during the lag phase uncover *YAP1*, *PDR1*, *PDR3*, *RPN4*, and *HSF1* as key regulatory genes in genomic adaptation to the lignocellulose derived inhibitor HMF for *Saccharomyces cerevisiae*.** *BMC Genomics* 2010, **11**:660.
- Causton HC, Ren B, Koh SS, Harbison CT, Kanin E, Jennings EG, Lee TI, True HL, Lander ES, Young RA: **Remodeling of yeast genome expression in response to environmental changes.** *Mol Biol Cell* 2001, **12**(2):323–337.
- Capaldi AP, Kaplan T, Liu Y, Habib N, Regev A, Friedman N, O'Shea EK: **Structure and function of a transcriptional network activated by the MAPK Hog1.** *Nat Genet* 2008, **40**(11):1300–1306.
- Gasch AP, Spellman PT, Kao CM, Carmel-Harel O, Eisen MB, Storz G, Botstein D, Brown PO: **Genomic expression programs in the response of yeast cells to environmental changes.** *Mol Biol Cell* 2000, **11**(12):4241–4257.
- Hahn JS, Hu ZZ, Thiele DJ, Iyer VR: **Genome-wide analysis of the biology of stress responses through heat shock transcription factor.** *Mol Cell Biol* 2004, **24**(12):5249–5256.
- Auesukaree C, Damnernsawad A, Kruatrachue M, Pokethitiyook P, Boonchird C, Kaneko Y, Harashima S: **Genome-wide identification of genes involved in tolerance to various environmental stresses in *Saccharomyces cerevisiae*.** *J Appl Genet* 2009, **50**(3):301–310.
- Giaever G, Chu AM, Ni L, Connelly C, Riles L, Veronneau S, Dow S, Lucau-Danila A, Anderson K, Andre B, Arkin AP, Astromoff A, El Bakkoury M, Bangham R, Benito R, Brachat S, Campanaro S, Curtiss M, Davis K, Deutschbauer A, Entian KD, Flaherty P, Foury F, Garfinkel DJ, Gerstein M, Gotte D, Guldener U, Hegemann JH, Hempel S, Herman Z, et al: **Functional profiling of the *Saccharomyces cerevisiae* genome.** *Nature* 2002, **415**(6896):387–391.
- Harbison CT, Gordon DB, Lee TI, Rinaldi NJ, Macisac AD, Danford TW, Hannett NM, Tagne JB, Reynolds DB, Yoo J, Jennings EG, Zeitlinger J, Pokholok DK, Kellis M, Rolfe PA, Takusagawa KT, Lander ES, Gifford DK, Fraenkel E, Young RA: **Transcriptional regulatory code of a eukaryotic genome.** *Nature* 2004, **431**(7004):99–104.
- Liti G, Carter DM, Moses AM, Warringer J, Parts L, James SA, Davey RP, Roberts IN, Burt A, Koufopanou V, Tsai IJ, Bergman CM, Bensasson D, O'Kelly MJT, van Oudenaarden A, Barton DBH, Bailes E, Ba ANN, Jones M, Quail MA, Goodhead I, Sims S, Smith F, Blomberg A, Durbin R, Louis EJ: **Population genomics of domestic and wild yeasts.** *Nature* 2009, **458**(7236):337–341.
- Dowell RD, Ryan O, Jansen A, Cheung D, Agarwala S, Danford T, Bernstein DA, Rolfe PA, Heisler LE, Chin B, Nislow C, Giaever G, Phillips PC, Fink GR, Gifford DK, Boone C: **Genotype to phenotype: a complex problem.** *Science* 2010, **328**(5977):469–469.
- Akao T, Yashiro I, Hosoyama A, Kitagaki H, Horikawa H, Watanabe D, Akada R, Ando Y, Harashima S, Inoue T, Inoue Y, Kajiwara S, Kitamoto K, Kitamoto

- N, Kobayashi O, Kuhara S, Masubuchi T, Mizoguchi H, Nakao Y, Nakazato A, Namise M, Oba T, Ogata T, Ohta A, Sato M, Shibasaki S, Takatsume Y, Tanimoto S, Tsuboi H, Nishimura A, *et al.*: Whole-genome sequencing of sake yeast *Saccharomyces cerevisiae* Kyokai no. 7. *DNA Res* 2011, **18**(6):423–434.
20. Borneman AR, Desany BA, Riches D, Affourtit JP, Forgan AH, Pretorius IS, Egholm M, Chambers PJ: Whole-genome comparison reveals novel genetic elements that characterize the genome of industrial strains of *Saccharomyces cerevisiae*. *PLoS Genet* 2011, **7**(2):e1001287.
21. Borneman AR, Forgan AH, Pretorius IS, Chambers PJ: Comparative genome analysis of a *Saccharomyces cerevisiae* wine strain. *FEMS Yeast Res* 2008, **8**(7):1185–1195.
22. Gancedo C, Flores CL: The importance of a functional trehalose biosynthetic pathway for the life of yeasts and fungi. *FEMS Yeast Res* 2004, **4**(4–5):351–359.
23. Ikner A, Shiozaki K: Yeast signaling pathways in the oxidative stress response. *Mutat Res* 2005, **569**(1–2):13–27.
24. Petersson A, Almeida JRM, Modig T, Karhumaa K, Hahn-Hägerdal B, Gorwa-Grauslund MF, Liden G: A 5-hydroxymethyl furfural reducing enzyme encoded by the *Saccharomyces cerevisiae* *ADH6* gene conveys HMF tolerance. *Yeast* 2006, **23**(6):455–464.
25. Liu ZL: Molecular mechanisms of yeast tolerance and in situ detoxification of lignocellulose hydrolysates. *Appl Microbiol Biotechnol* 2011, **90**(3):809–825.
26. Dunn B, Richter C, Kvittek DJ, Pugh T, Sherlock G: Analysis of the *Saccharomyces cerevisiae* pan-genome reveals a pool of copy number variants distributed in diverse yeast strains from differing industrial environments. *Genome Res* 2012, **22**(5):908–924.
27. Wehner EP, Rao E, Brendel M: Molecular-structure and genetic-regulation of *Sfa*, a gene responsible for resistance to formaldehyde in *Saccharomyces cerevisiae*, and characterization of its protein product. *Mol Gen Genet* 1993, **237**(3):351–358.
28. Gaisne M, Bécam AM, Verdier J, Herbert CJ: A 'natural' mutation in *Saccharomyces cerevisiae* strains derived from S288c affects the complex regulatory gene *HAP1* (*CYP1*). *Curr Genet* 1999, **36**(4):195–200.
29. Peña MMO, Puig S, Thiele DJ: Characterization of the *Saccharomyces cerevisiae* high affinity copper transporter *Ctr3*. *J Biol Chem* 2000, **275**(43):33244–33251.
30. Knight SA, Labbe S, Kwon LF, Kosman DJ, Thiele DJ: A widespread transposable element masks expression of a yeast copper transport gene. *Genes Dev* 1996, **10**(15):1917–1929.
31. Chan JE, Kolodner RD: A genetic and structural study of genome rearrangements mediated by high copy repeat *Ty1* elements. *PLoS Genet* 2011, **7**(5):e1002089.
32. Wei W, McCusker JH, Hyman RW, Jones T, Ning Y, Cao Z, Gu Z, Bruno D, Miranda M, Nguyen M, Wilhelmy J, Komp C, Tamse R, Wang X, Jia P, Luedi P, Oefner PJ, David L, Dietrich FS, Li Y, Davis RW, Steinmetz LM: Genome sequencing and comparative analysis of *Saccharomyces cerevisiae* strain YJM789. *Proc Natl Acad Sci USA* 2007, **104**(31):12825–12830.
33. Eastmond DL, Nelson HCM: Genome-wide analysis reveals new roles for the activation domains of the *Saccharomyces cerevisiae* heat shock transcription factor (*Hsf1*) during the transient heat shock response. *J Biol Chem* 2006, **281**(43):32909–32921.
34. Mieczkowski PA, Lemoine FJ, Petes TD: Recombination between retrotransposons as a source of chromosome rearrangements in the yeast *Saccharomyces cerevisiae*. *DNA Repair* 2006, **5**(9–10):1010–1020.
35. Novo M, Bigey F, Beyne E, Galeote V, Gavory F, Mallet S, Cambon B, Legras JL, Wincker P, Casaregola S, Dequin S: Eukaryote-to-eukaryote gene transfer events revealed by the genome sequence of the wine yeast *Saccharomyces cerevisiae* EC1118. *Proc Natl Acad Sci USA* 2009, **106**(38):16333–16338.
36. Doniger SW, Kim HS, Swain D, Corcuera D, Williams M, Yang SP, Fay JC: A catalog of neutral and deleterious polymorphism in yeast. *PLoS Genet* 2008, **4**(8):e1000183.
37. Emerson JJ, Hsieh LC, Sung HM, Wang TY, Huang CJ, Lu HH, Lu MY, Wu SH, Li WH: Natural selection on cis and trans regulation in yeasts. *Genome Res* 2010, **20**(6):826–836.
38. Yvert G, Brem RB, Whittle J, Akey JM, Foss E, Smith EN, Mackelprang R, Kruglyak L: Trans-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors. *Nat Genet* 2003, **35**(1):57–64.
39. Brem RB, Yvert G, Clinton R, Kruglyak L: Genetic dissection of transcriptional regulation in budding yeast. *Science* 2002, **296**(5568):752–755.
40. Hama S, Yamaji H, Kaieda M, Oda M, Kondo A, Fukuda H: Effect of fatty acid membrane composition on whole-cell biocatalysts for biodiesel-fuel production. *Biochem Eng J* 2004, **21**(2):155–160.
41. Tao XL, Zheng DQ, Liu TZWPM, Zhao WP, Zhu MY, Jiang XH, Zhao YH, C WX: A Novel Strategy to Construct Yeast *Saccharomyces cerevisiae* Strains for Very High Gravity Fermentation. *PLoS One* 2012, **7**(2):e31235. doi:10.1371/journal.pone.0031235.
42. Argueso JL, Westmoreland J, Mieczkowski PA, Gawel M, Petes TD, Resnick MA: Double-strand breaks associated with repetitive DNA can reshape the genome. *Proc Natl Acad Sci USA* 2008, **105**(33):11845–11850.
43. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997, **25**(17):3389–3402.
44. Kent WJ: BLAT - The BLAST-like alignment tool. *Genome Res* 2002, **12**(4):656–664.
45. Delcher AL, Bratke KA, Powers EC, Salzberg SL: Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* 2007, **23**(6):673–679.
46. Carver TJ, Rutherford KM, Berriman M, Rajandream MA, Barrell BG, Parkhill J: ACT: the Artemis comparison tool. *Bioinformatics* 2005, **21**(16):3422–3423.
47. Wang B, Guo GW, Wang C, Lin Y, Wang XN, Zhao MM, Guo Y, He MH, Zhang Y, Pan L: Survey of the transcriptome of *Aspergillus oryzae* via massively parallel mRNA sequencing. *Nucleic Acids Res* 2010, **38**(15):5075–5087.
48. Li RQ, Yu C, Li YR, Lam TW, Yiu SM, Kristiansen K, Wang J: SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* 2009, **25**(15):1966–1967.
49. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B: Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 2008, **5**(7):621–628.
50. Audic S, Claverie JM: The significance of digital gene expression profiles. *Genome Res* 1997, **7**(10):986–995.
51. Zheng DQ, Wu XC, Wang PM, Chi XQ, Tao XL, Li P, Jiang XH, Zhao YH: Drug resistance marker-aided genome shuffling to improve acetic acid tolerance in *Saccharomyces cerevisiae*. *J Ind Microbiol Biot* 2011, **38**(3):415–422.

doi:10.1186/1471-2164-13-479

Cite this article as: Zheng *et al.*: Genome sequencing and genetic breeding of a bioethanol *Saccharomyces cerevisiae* strain YJS329. *BMC Genomics* 2012 **13**:479.

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

