PHYLOGENETIC SELECTION GUIDED SACCHAROMYCES CEREVISIAE S288C GLUCOSE FERMENTATION MODELING

ASHISH RUNTHALA[1*], ANKITASH TULYANI[2], DHIRAJ SHARMA[3] AND MAHAVEER SINGH[4]

[1]Biological Sciences, Birla Institute of Technology & Science, Pilani, Rajasthan, India
[2]Department of Pharmacy, Birla Institute of Technology & Science, Pilani, Rajasthan, India
[3]Department of Bioinformatics, AdiBiosolutions, Chandigarh, Punjab, India
[4]Department of Pharmacy, Birla Institute of Technology & Science, Pilani, Rajasthan, India
*Corresponding Author: Email- ashish.runthala@gmail.com

Abstract- Fermentation products are indigenous to many civilizations, and they have been produced by industries since a long time. *Saccharomyces cerevisiae S288C* (commonly known as baker's yeast) is the strain mainly used in the Glucose based fermentation industries. We have seen the use of same yeast strain at different places with different Phenotypic Constraints. The way to improve the adaptability of considered strain for desired phenotypic conditions, using smart selection of genes through cybernetic modeling is illustrated. Phylogenetic homologues for all *S. cerevisiae S288c* Glucose Fermentation pathway genes were screened to search evolutionarily related functional domains in other yeast strains like *Saccharomyces cerevisiae YJM789, Candida glabrata CBS138, Kluyveromyces lactis NRRL Y-1140, Ashbya gossypii ATCC10895* etc., which are adapted naturally in different set of environment. We observed that *Saccharomyces cerevisiae YJM789, Candida glabrata CBS138, Ashbya gossypii ATCC10895, Kluyveromyces lactis NRRL Y-1140* possess highly conserved functional domains, which can be carefully selected based on usage. This study aims at designing an algorithm to select and incorporate evolutionary homologues for genes of a considered strain, which mostly show sub-optimal performance in the desired set of experimental constraints. Such a consideration of native microenvironment and evolutionary closeness in the selection of functional homologues of the entire genetic set can thus be significantly fruitful.

Key words - Cybernetic, Phylogeny, Domains, Homologues, Micro-environment

Background

One of most important yeast, *Saccharomyces cerevisiae* has been a very useful fungus for many millennia [1,25]. This single-celled model organism is used for studying cellular and molecular processes in eukaryotes and is used for making bread, beer, wine, enzymes, and pharmaceuticals [2]. The genome is composed of about 12 million base pairs (Mb) and contains 6,275 genes [3]. *Saccharomyces cerevisiae S288C* (Considered Strain) is commonly used model yeast in fungal molecular research including sequence analysis, genetic mechanism in the metabolic pathways, resistance to antifungal drugs, and the investigation of factors of pathogenicity, such as adhesion [4]. This simple eukaryote (also known as bakers' yeast) has many advantages as a research system: small size, rapid growth, complete sexual life cycle, safety, well-characterized genetics, a completely sequenced genome, and not least, the world-wide community of yeast genetics researchers as a resource base. Fermentation based thermodynamic parameters have been studied have also revealed different enthalpic compensations for the Yeast Hexokinase (HXK) Isozymes HXK1 & HXK2 [5]. Differential Requirement of Yeast Sugar Kinases is

already established for the Catabolite Repressed State [6]. All these studies pave a way to study the complete mechanism of the fermentation pathway of yeast *Saccharomyces cerevisiae S288C* to find its suitable and biologically significant homologues from known structures to consider them within other yeast species like *Saccharomyces cerevisiae YJM789, Candida glabrata CBS138, Kluyveromyces lactis NRRL Y-1140, Ashbya gossypii ATCC10895* based on the desired constraint set of requirements [7]. Recently a phylogenetic analysis was attempted on *Pistacia* L. for a completely different objective to extract evolutionarily conserved information, based on morphological data [8]. Ultimate objective of this multi-component, or multiple gene based optimization analysis is to scratch the homologous variants of all possible genes which might be functioning at sub-optimal expression levels under selected set of constraints. This study thus suggests the algorithm to select genetic variants of the considered glucose pathway genes to highlight the fact that micro-organisms always propagate best in the conditions which are closer to its actual native

growth conditions, and can thus be significantly fruitful.

Materials & Methods

Methodology includes getting information from different sources to analyze certain key parameters which can be used for improving the fermentation process efficiency. This methodology was developed in the 7 months tenure from August 6th 2009 to March 03rd 2010, in Birla Institute of Technology & Science College, INDIA on my system. One such algorithm was recently proposed to identify *Streptomyces noboritoensis* TBG-V20 variants with cellulase production based on the Neighbor-Joining phylogenetic analysis of 16S-rRNA and morphological features [9]. This algorithm considers fermentation pathway genes as different variables and the optimization problem can thus be broken down to predict a set of genes which can proliferate and respond well under the desired set of constraints, as bulleted below:

i.   **Sequence retrieval:** Saccharomyces Genome Database (SGD) was used for the study, which provides scientific database to molecular biology and genetics of the yeast. SGD provides many resources to compare and integrate information on genomic sequences and associated information.

ii. **Sequence comparison:** Sequence Similarity Query Tool employing PSI-BLAST Sequence Alignment Algorithm was used for Sequence comparisons to predict the evolutionary relationship of specific protein sequences. Genome-wide Protein Similarity View program and Fugal Alignment Viewers were then used for analyzing the alignment between the considered Sequences.

iii. **Cluster Probable Homologues:** Sacch3D was then employed to organize and present structural information about the considered set of fermentation pathway proteins and their putative homologues.

iv. **Estimating degree of closeness:** From the clusters, Phylogeny trees were computed to find the biologically close sequences for the considered genes. These tools helped in estimating the functional relationships between different considered genes and thus the possible Evolutionary conserved nature of their protein folds was analyzed.

v.   **CYGD verification:** Comprehensive Yeast Genome Database (CYGD) was also used finally to compile genetic data to verify the functional relationships of the *Saccharomyces cerevisiae*'s considered genes and the other related species. Thus

we confirmed genes with scores showing the degree of closeness for the considered gene and thus 25 trees were drawn for all genes of the Glucose Fermentation Pathway. The details for considering these genes individually as each of them is there for a specific purpose is elaborated in Table I and this Methodology is briefly given as Flowchart in Fig.(I).

Results

Cybernetic modeling approach used in the article streamlined the optimal requirements and resultant productive end-products of the individual reaction steps involved in the glucose fermentation pathway. Sequential utilization of substrates with the preferential utilization supporting the higher growth rate is well characterized and can be optimally employed to stabilize the dynamics of diauxic growth. Objective for this study solved the purpose and highlighted that we can use the enzymes which we want in for a specific reaction, instead of extra cell survival pressure burden caused by the transcriptions and translations of additional genes not required for the economic scaled fermenter runs. Thus instead of studying and optimizing u_i and v_i, variables respectively for enzyme synthesis and activity for each reaction step involved in the pathway, we can make a combined set constant of 2 variables for all the reactions considered in the fermentation pathway, as indicated in example below.

$$0.21A_1 + 0.16A_2 + 0.76A_3 + 0.48A_4 + 0.27A_5 + 0.11A_6 \longrightarrow 0.76P_1 + 0.23P_2$$
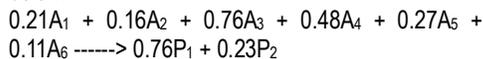
In this equation for example, 6 reactants form two products. If we just require Product P1, we can essentially substitute constant 0.23 with 0. This will essentially redirect cell chemical load for just the production of Product P1. Similarly, we can also remove unwanted genes or think of substituting the genes with more efficient genes with better productive rate constants from related genera.

Biological process optimization control can thus be decomposed into a sequence of elementary components. Each elementary component actually steers the reaction toward its physiological objectives in an optimal manner. The cells normally utilize limited pool of resources in an optimal manner. So, if genetic alterations were made feasible, i.e. if certain genes for existing pathway enzymes, not required for the considered fermentation pathway were deleted, then we could actually think of further enhancing overall productivity of the fermentation pathway, by setting the corresponding cybernetic variables to zero in the linear optimization equation for the deleted genes. Otherwise also, if there would have been genetic alteration of the gene(s) encoding key

179

enzyme(s), we could have used fractional cybernetic variable for activity.

From our analysis, we studied the reasonably conserved nature of PGI1 signature 1 and 2 domains across the species. Similar was the case for FBA1 signature 1 & 2 domains. We also investigated that almost all genes have their more productive homologue copy across the species, possibly evolved because of differential phenotypic & chemical constraints on them. More productive copy of TPII was also found as homologous in *Candida, Valterwaltozyma, Ashbya* strains. TDH1, TDH2 and TDH3 genes were also found to have variants in different strains based on the available survival pressures on them. An almost exact copy of PGK1 was also visualized. ENO1 and ENO2 domains were found 100% conserved in *Candida, Vanderwaltozyma, Ashbya* strains. PYK2 and CDC19 were also shown to catalyze production of phosphoenolpyruvate to pyruvate, and this property was found to be evolutionary closest in *Candida, Vanderwaltozyma, Ashbya* strains. PDC6, PDC5 and PDC1 were also found structurally closest in *Candida, Ashbya and Kluveromyces* strains. Such strains were also found closest for ADH5, ADH4, ADH3, and ADH2 genes of *Saccharomyces Cerevisiae* 288C genes.

Discussion

In the detail analysis of the genes involved in glucose fermentation in *Saccharomyces cerevisiae S288C* we found that each gene contains a specific function, based on the domain region. It was revealed that other yeast species contain the homologous genes, being almost similar with the genes involved in glucose fermentation in *Saccharomyces cerevisiae S288C* and also share complete homology at the conserved domain regions. This indicates the conserved nature of the functional domains across the yeast species and hence can be used for careful selection of genes. In the Entire Glucose Fermentation Pathway, Hexokinase-1(HXK1), Hexokinase-2(HXK2) and Glucokinase-1(GLK1) were found to contain conserved hexokinase domain to catalyze the phosphorylation of keto- and aldohexoses (e.g. Glucose, mannose and fructose) using Mg-ATP as phosphoryl group donor [6]. Similarly, homologous copies of the following genes were found in different strains like PGI1 (Phosphoglucose isomerase) signature 1 and 2 domains, which catalyze the reversible isomerization of glucose-6-phosphate and fructose-6-phosphates. PFK1 and PFK2 (Phosphofructokinase domain) which catalyzes phosphorylation of fructose-6-phosphate to fructose-1,6-bisphosphate by ATP was also reasonably conserved across the species. An almost similar case was observed by [10, 11] for FBA1(Fructose-bisphosphate aldolase) class-II Signature 1 and 2 domains, which was found to

catalyze the reversible aldol cleavage or condensation of Fructose-1, 6-bisphosphate into DHAP and glyceraldehyde 3-phosphate, and investigated TPII (Triosephosphate isomerase) domain [12], which catalyzes the reversible interconversion of glyceraldehyde 3-phosphate and DHAP was found as homologous in *Candida, Valterwaltozyma, Ashbya* strains Triose Phosphate Dehydrogenase -1,2,3 genes. Triosephosphate Dehydrogenase-1,2,3 genes [13] (TDH1, TDH2 and TDH3) containing Glyceraldehyde 3-phosphate dehydrogenase domain, which is involved in forming a covalent phosphoglycerol thioester intermediate, were also found to have variants in different strains based on the available phenotypic conditions to them for survival [14]. An almost exact copy of PGK1 (Phosphoglycerate kinase), catalyzing second step in the second phase of glycolysis, i.e. reversible conversion of 1, 3-diphospho-glycerate to 3-phosphoglycerate and an ATP molecule was also visualized [15]. Phosphoglycerate mutase, which catalyze transfer of phosphate groups between the three carbon atoms of phosphoglycerate was also investigated by [16]. ENO1 and ENO2 (Enolase) domain which catalyzes dehydration of 2-phospho-D-glycerate to phosphoenolpyruvate were found 100% conserved in *Candida, Vanderwaltozyma, Ashbya* strains [17] PYK2 and CDC19 (Pyruvate kinase), were also shown to catalyze conversion of phosphoenolpyruvate to pyruvate with the concomitant phosphorylation of ADP to ATP, and this property was found to be evolutionary closest in *Candida, Vanderwaltozyma, Ashbya* strains. PDC6, PDC5 and PDC1, which contain Thiamine pyrophosphate; and which requires thiamine pyrophosphate (TPP) (vitamin B1) as a cofactor [18]. *Candida, Ashbya and Kluveromyces* strains were found to be closest for ADH5, ADH4, ADH3, and ADH2 genes of Saccharomyces Cerevisiae 288C genes which contain Alcohol dehydrogenase to catalyze reversible oxidation of ethanol to acetaldehyde with the concomitant reduction of NAD [19].

Conclusion

This shows that glucose fermentation pathways also take place in these yeast species and they can be used to carryout the fermentation process in the industry at an enhanced production rate using cybernetic modeling of the variables. Some species like *Saccharomyces cerevisiae YJM789, Candida glabrata CBS138, Ashbya gossypii ATCC 10895, Kluyveromyces lactis NRRL Y-1140* were found to contain highly conserved domains of almost all genes with average scores closer to 0.0582, 0.09340, 0.11499, and 0.13436 respectively. But, *Aspergillus and Yarrowia* strains were found to be evolutionarily farthest, with no such obvious similar functionally conserved domains with their average

scores being almost 0.19 and 0.27 respectively. So, there is a need to check the evolutionary conserved nature and scores to select genes for desired strain development for considered environmental constraints for the fermentation process as they can be effectively used instead of *Saccharomyces cerevisiae S288* with additional evolved features, depending on our usage.

Acknowledgement

References

[1]   Maitra P.K. and Lobo Z. (1977) *Molecular Genetics and Genomics*, 156(1), 55-60.

[2]   Sprague G.F. Jr. (1977) *The Journal of Bacteriology*, 130(1), 232-241.

[3]   Brazma A., Vilo J., Ukkonen E. and Valtonen K. (1997) *'Proceedings of the Fifth International Conference Intelligent Systems for Molecular Biology' ISMB-97, Published by AAAI Press*, 65-74.

[4]   Zhenglong G., David L., Petrov D., Jones T., Davis R.W. and Steinmetz L.M. (2005) *The Proceedings of the National Academy of Science USA,* 102(4), 1092-1097.

[5]   Bianconi M.L. (2003) *The Journal of Biological Chemistry*, 278(21), 18709-18713.

[6]   Bisson L.F., and Fraenkel D.G. (1983) *The Proceedings of the National Academy of Science USA*, 80(6), 1730-1734.

[7]   Lowe S.L. and Reithel F.J (1975) *The Journal of Biological Chemistry*, 250(1), 94-99.

[8]   Al-Saghir M.G. (2010) *Journal of Plant Sciences*, 9(1), 28-35.

[9]   Arunachalam R., Wesely E.G., George J., and Annadural G. (2010) *Current Research in Bacteriology*, 3(1), 15-26.

[10]   Clifton D. and Fraenkel D.G. (1982) *Biochemistry*, 21(8), 1935-1942.

[11]   Compagno C., Ranzi B.M. and Martegani E. (1991) *FEBS Letters*, 293(1-2), 97-100.

[12]   Scott E.W. and Baker H.V. (1993) *Molecular and Cell Biology*, 13(1), 543-550.

[13]   Kumar A., Agarwal S., Heyman J.A., Matson S. and Heidtman M. (2002) *Genes and Development*, 16(6), 707-719.

[14]   McAlister L. and Holland M.J. (1985) *The Journal of Biological Chemistry*, 260(28), 15019-15027.

[15]   Klein C.J., Olsson L. and Nielsen J. (1998) *Microbiology*, 144(1), 13-24.

[16]   McAlister L. and Holland M.J. (1982) *The Journal of Biological Chemistry,* 257(12), 7181-7188.

[17]   Brewer J.M., Glover C.V., Holland M.J. and Lebioda L. (1997) *Biochimica et Biophysica Acta*, 1340(1), 88-96.

[18]   Boles E., Gubbles P.J. and Pronk J.T. (1998) *The Journal of Bacteriology*, 180(11), 2875-2882.

[19]   Watson H.C., Walker N.P., Shaw P.J., Bryant T.N., and Wendell P.L., Fothergill L.A., Perkins L.A., Conroy S.C., Dobson M.J., Tuite M.F., Kingsman A.J. and Kingsman S.M. (1982) *The EMBO Journal*, 1(12), 1635-1640.

[20]   Heinisch J. (1986) *Molecular and General Genetics,* 202(1), 75-82.

[21]   Compagno C., Brambilla L., Capitanio D., Boschi F., Ranzi B.M. and Porro D. (2001) *Yeast*, 18(7), 663-670.

[22]   Rodicio R. and Heinisch J. (1987) *Molecular and General Genetics*, 206(1), 133-140.

[23]   Lam K.B. and Marmur J. (1977) *The Journal of Bacteriology*, 130(2), 746-749.

[24]   Dickinson J.R. and Williams A.S. (1986) *Journal of General Microbiology*, 132(9), 2605-2610.

[25]   Maitra P.K. and Lobo Z. (1977) *Molecular and General Genetics*, 152(2), 193-200.

[26]   Hohmann S. (1991) *The Journal of Bacteriology*, 173(24), 7963-7969.

[27]   Young E.T. and Pilgrim D. (1985) *Molecular and Cellular Biology*, 5(11), 3024-3034.

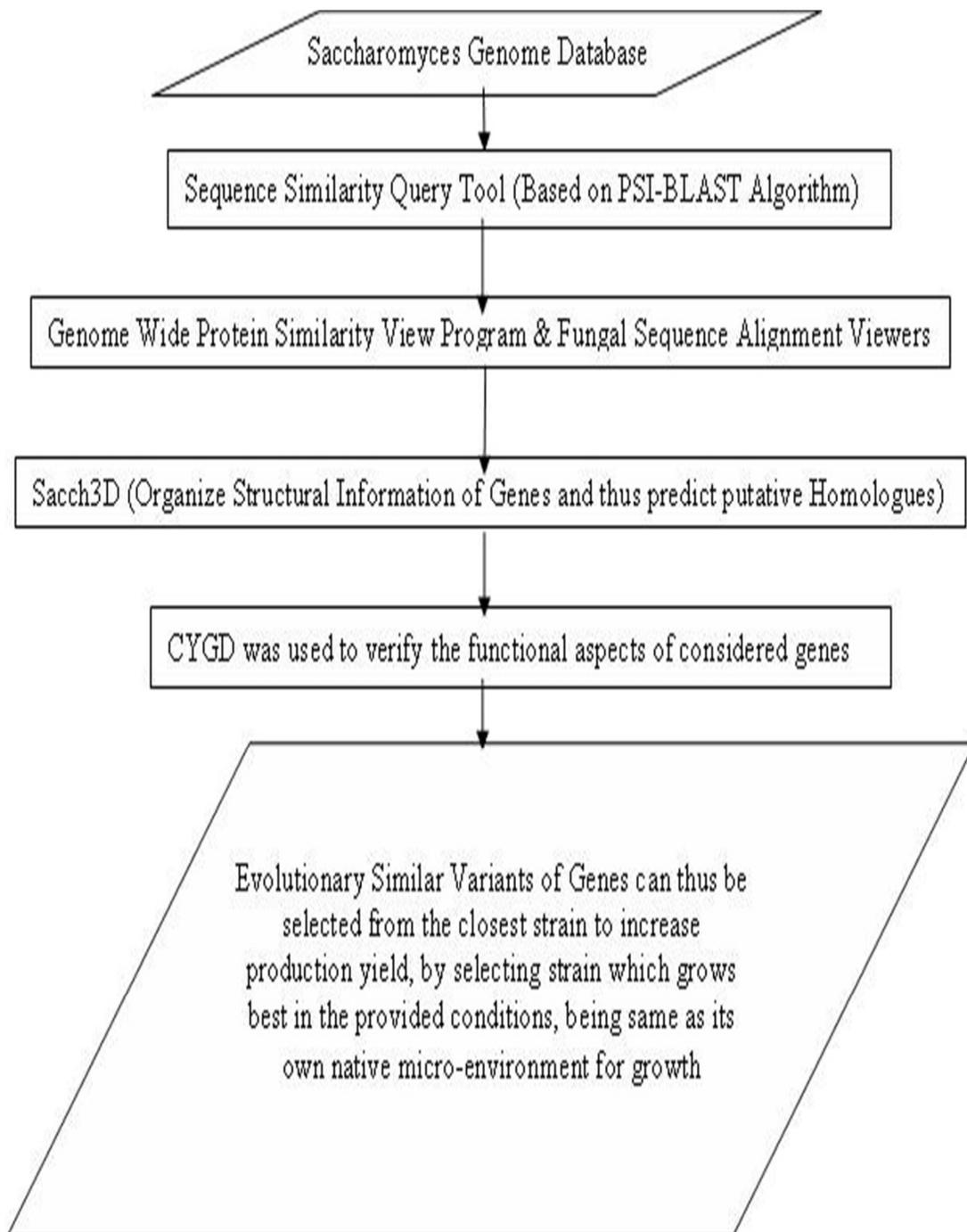[28]   Drewke C. and Ciriacy M. (1988) *Biochimica et Biophysica Acta*, 950(1), 54-60.

```
                    ┌─────────────────────────────────────────┐
                   /   Saccharomyces Genome Database           /
                  └─────────────────────────────────────────┘
                                    │
                                    ▼
        ┌───────────────────────────────────────────────────────┐
        │ Sequence Similarity Query Tool (Based on PSI-BLAST Algorithm) │
        └───────────────────────────────────────────────────────┘
                                    │
                                    ▼
        ┌───────────────────────────────────────────────────────────┐
        │ Genome Wide Protein Similarity View Program & Fungal Sequence Alignment Viewers │
        └───────────────────────────────────────────────────────────┘
                                    │
                                    ▼
        ┌─────────────────────────────────────────────────────────────┐
        │ Sacch3D (Organize Structural Information of Genes and thus predict putative Homologues) │
        └─────────────────────────────────────────────────────────────┘
                                    │
                                    ▼
        ┌───────────────────────────────────────────────────────┐
        │ CYGD was used to verify the functional aspects of considered genes │
        └───────────────────────────────────────────────────────┘
                                    │
                                    ▼
```

Evolutionary Similar Variants of Genes can thus be selected from the closest strain to increase production yield, by selecting strain which grows best in the provided conditions, being same as its own native micro-environment for growth

Fig. I: Methodology Flowchart applied in the Analysis
Flowchart indicating the steps to select genes for the alignment analysis to be finally used for Structural Similarity of Homologous and structurally conserved folds, ensuring the algorithm that genes with almost similar structural folds can be considered as Genetic Variables thus allowing the selection of best set of genes for the best possible Fermentation Yield based solely on the Micro-organism Growth Kinetics, mimicked to its natural native condition.

Table I: Inter-related Genetic Network of 25 Genes
It should be well simulated, so that all the genes express to the best possible optimized level, to give desired production rates in the altered conditions, feasible for us to implement. All these genes and their functional inter-relationships can be clearly understood from phylograms as shown in Figure 2 for GLK1 gene.

S. No.	Genes (NCBI Gene ID)	Encodes	Function	Inference [Reference]
1.	GLK1 (850317)	Glucokinase	Catalyzes glucose phosphorylation at C6 in the first irreversible step of glucose metabolism	HXK1 and HXK2p are predominant isozymes during growth on glucose. Figure 2 shows the phylogeny of GLK1 gene to cluster the genes similar to GLK1 for tracing the closest genetic copy [6]
2.	HXK1 (852639)	Hexokinase		
3.	HXK2 (850614)			
4.	PGII (852495)	Phospho-Glucose Isomerase	Catalyzes interconversion of glucose-6-phosphate and fructose-6-phosphate	This gene is required for cell cycle progression and completion of gluconeogenic events of sporulation [20].
5.	PFK1 (3636448)	Alpha subunit of phosphor-fructokinase	Encodes a glycolysis enzyme that is strongly induced by glucose and catalyzes the formation of fructose 1, 6-bisphosphate from fructose 6-phosphate and ATP.	Due to Glucose co-relation, their role in allosteric regulation in changing environments is indicated. Phosphofructokinase and pyruvate kinase (Cdc19p, Pyk2p) only function in the forward direction. So they are specific to only to glycolysis [10]
6.	PFK2 (3637497)	Beta subunit of phosphor-fructokinase		
7.	FBA1 (853805)	Fructose-1,6-Bisphosphate aldolase	Catalyzes the conversion of fructose 1, 6 bisphosphate into two 3-carbon products: lyceraldehyde-3-phosphate and dihydroxyacetone phosphate	FBA1 transcription is not regulated by glucose but its expression increases on a non-fermentable carbon source. Fba1p also catalyzes reverse reaction, being important for growth on non-sugar carbon source [11, 21] .
8.	TPII (851620)	Triose phosphate isomerase	Abundant glycolytic enzyme whose transcription is controlled by Reb1p, Gcr1p, and Rap1p activators.	Tpi1p catalyzes reversible interconversion of glyceraldehyde 3-phosphate and dihydroxy-acetone phosphate, required for growth on glucose as sole carbon source [15]
9.	TDH1 (853395)	Threonine dehydrogenase	These unlinked genes encode related polypeptides, forming active homotetramers. Tdh2p and Tdh3p are detected in log phase whereas Tdh1p is detected in stationary phase.	Pfk1p, Pfk2p and Cdc19p are functional forward only. So gluconeogenesis pathway replaces these steps with pyruvate carboxylase (Pyc1p, Pyc2p), phosphoenolpyruvate carboxykinase (Pck1p), generating oxaloacetate as an intermediate [17].
10.	TDH2 (853465)			
11.	TDH3 853106)			

183

12.	PGK1 (850370)	3-Phospho-glycerate Kinase	Catalyzes transfer of phosphoryl groups from 1, 3-bisphosphoglycerate to ADP to produce ATP	It was found that PGK1 transcription is increased by heat shock [19]. Transcription is activated by transcription factors Rap1p, Abf1p, and Reb1p, by binding to their respective sequences in PGK1 promoter [22].
13.	GPM1 (853705)	Tetrameric phosphor-glycerate mutase	Converts 3-phosphoglycerate to 2-phosphoglycerate during glycolysis and reverse reaction during gluconeogenesis	Homozygous diploid GPM1 mutant studied [21] failed to sporulate [24], which proved that gluconeogenesis was required for sporulation [16].
14.	ENO1 (853169)	Enolase I	Catalyzes the conversion of 2-phosphoglycerate to phosphoenolpyruvate during glycolysis and the reverse reaction during gluconeogenesis	These enzymes function as dimeric phosphopyruvate hydratase complexes. Log phase cells, grown on glucose contain 20-fold more Eno2p than Eno1p, whereas cells grown on ethanol or glycerol with lactate contain the similar amounts [17].
15.	ENO2 (856579)	Enolase II		
16.	PYK2 (854529)	Pyruvate kinase	Encodes pyruvate kinase which catalyzes the conversion of phosphoenolpyruvate to pyruvate	PYK2 is subjected to glucose repression and seems insensitive to FBP level, proving that it is active when FBP level is too low to activate CDC19 [1, 25].
17.	CDC19 (851193)			CDC19 deletion mutants cannot grow using glucose or other fermentable sugars, but grow well on ethanol or lactate indicating an alternate route for pyruvate synthesis [1,25].
18.	PDC1 (850733)	Isozymes of Pyruvate Decarboxylase	Catalyze pyruvate degradation into acetaldehyde and carbon dioxide. They can also decarboxylate 2-oxo acids such as indole-pyruvate, 2-keto-3-methyl-valerate.	Characterized proteins contributed in different catabolic activities [26] because of amino acids like isoleucine, phenylalanine, tryptophan, and valine. Involvement of specific amino acids in primary structure of *S. cerevisiae* alcohol dehydrogenase gene was also shown the same time.
19.	PDC5 (850825)			
20.	PDC6 (852978)			
21.	ADH1 (854068)	Alcohol Dehydrogenases	Function in the ethanol metabolism pathway.	ADH3 reduce acetaldehyde to ethanol [27] during glucose fermentation. ADH4 was purified later to study its activity [28]. It was revealed that except ADH2p all of these dehydrogenases reduce acetaldehyde to ethanol during glucose fermentation, whereas Adh2p catalyzes the reverse reaction of oxidizing ethanol to acetaldehyde.