Chapter 5
Comparative Genomics and Evolutionary Genetics of Yeast Carbon Metabolism

Zhenguo Lin and Wen-Hsiung Li

Abstract Yeasts display highly diversified physiological characteristics. The most distinct physiological character in many yeast species is their special sugar metabolism. In particular, the baker's yeast *Sacharomyces cerevisiae* and its relatives predominantly ferment sugars into ethanol even in the presence of oxygen, which is known as Crabtree effect or aerobic fermentation. It has been postulated that this unusual carbohydrate metabolism provides these yeasts selective advantages in sugar-rich environments. However, it has long been a mystery as to genetically how these yeasts evolved a predominantly fermentative lifestyle. The rapid accumulation of genomic, transcriptomic, and epigenetic data in many yeast species in recent years has greatly increased our understanding of the genetic basis and molecular mechanism for the diversified sugar metabolisms among yeasts. In this chapter, we provide a review of recent comparative genomics and evolutionary studies related to the metabolisms of glucose and galactose, whose metabolic pathways have been extensively studied in yeasts. A series of studies suggested that the evolution of aerobic fermentation involved many different factors, including increases in copy numbers of genes involved in glucose transport, glycolysis and ethanol production; sequence divergence; and transcriptional reprogramming of genes involved in mitochondrial functions through changes of *cis*-regulatory elements and promoter structures. It has also been found that the different abilities among yeasts to use galactose is strongly correlated with the presence of the galactose pathway genes in their genomes. These studies revealed that the adaptation of yeasts to specific niches has greatly shaped the genomic content and the regulatory program.

Z. Lin
Department of Ecology and Evolutionary Biology, Rice University, Houston, TX 77005, USA

W.-H. Li (✉)
Biodiversity Research Center, Academia Sinica, 128, 115 Taipei, Taiwan
e-mail: wli@uchicago.edu

W.-H. Li
Department of Ecology and Evolution, University of Chicago, Chicago, IL 60637, USA

J. Piškur and C. Compagno (eds.), *Molecular Mechanisms in Yeast Carbon Metabolism*,
DOI: 10.1007/978-3-642-55013-3_5, © Springer-Verlag Berlin Heidelberg 2014

5.1 Introduction

Although yeasts are single cell organisms, they have evolved highly diversified physiological characters, especially in carbohydrate metabolism (Barnett et al. 2000). The monosaccharide glucose is the basic carbohydrate unit of cellular metabolism and is the most important carbohydrate source of cellular energy. In the presence of oxygen, most eukaryotic species fully degrade glucose into CO_2 and H_2O through the respiration pathway for the maximum energy yield. However, many yeast species, including *Saccharomyces cerevisiae* and its close relatives, have evolved a remarkable ability to predominantly ferment sugars even under aerobic conditions to produce ethanol. This characteristic is called aerobic fermentation or Crabtree effect (De Deken 1966). These fermentative yeasts are able to tolerate a high concentration of extracellular ethanol and to utilize ethanol as the carbon source after depletion of sugars. This fermentative lifestyle was termed the "make–accumulate–consume" strategy, which in natural habitats enables *Saccharomyces* yeasts to outcompete other microorganisms (Piskur et al. 2006). How *S. cerevisiae* evolved to be a good fermenter has become a subject of intense research in the last decade.

Thanks to the availability of genomic sequences in many yeast species and the abundant high-throughput transcriptomic and epigenetic data, recent studies have greatly advanced our knowledge about the yeast carbon metabolism. As one of the most intensely studied eukaryotic model organisms, *S. cerevisiae* is the first eukaryotic species to have a completely sequenced genome (Goffeau et al. 1996). The second completely sequenced yeast genome came 6 years later from the fission yeast *Schizosaccharomyces pombe*, which diverged from *S. cerevisiae* probably more than 300 million years ago (Wood et al. 2002). With advances in sequencing technology, the number of completely sequenced genomes has been increasing rapidly. To date, the complete genomes of at least 50 different yeast species have been registered at Genbank. In addition, the genomes of 33 different wild and domestic strains of *S. cerevisiae* have also been sequenced (Engel and Cherry 2013). Compared to other eukaryotes, yeasts have streamlined genomes ranging from 9 to 20 megabases in haploid, containing 4,700–6,500 protein-coding genes (Dujon 2010). In addition to the genomic data, the first eukaryotic genome-wide gene expression data was completed in *S. cerevisiae* by microarray technology in 1997 (DeRisi et al. 1997). Since then, large amounts of transcriptomic data have been generated in many yeast species by micro-array, tiling-array, and next-generation sequencing technologies (Ferea et al. 1999; Ihmels et al. 2002; Yuan et al. 2005; Miura et al. 2006; Field et al. 2008; Tsankov et al. 2010). Furthermore, genome-wide epigenetic data and protein-protein interaction data have also been accumulated in several yeast species (Ferea et al. 1999; Ihmels et al. 2002; Yuan et al. 2005; Miura et al. 2006; Field et al. 2008, Tsankov et al. 2010). These various types of data and the availability of powerful bioinformatics tools for data analyses paved the way for comparative genomic and transcriptomic studies and for elucidating the genetic basis underlying the evolution of phenotypic

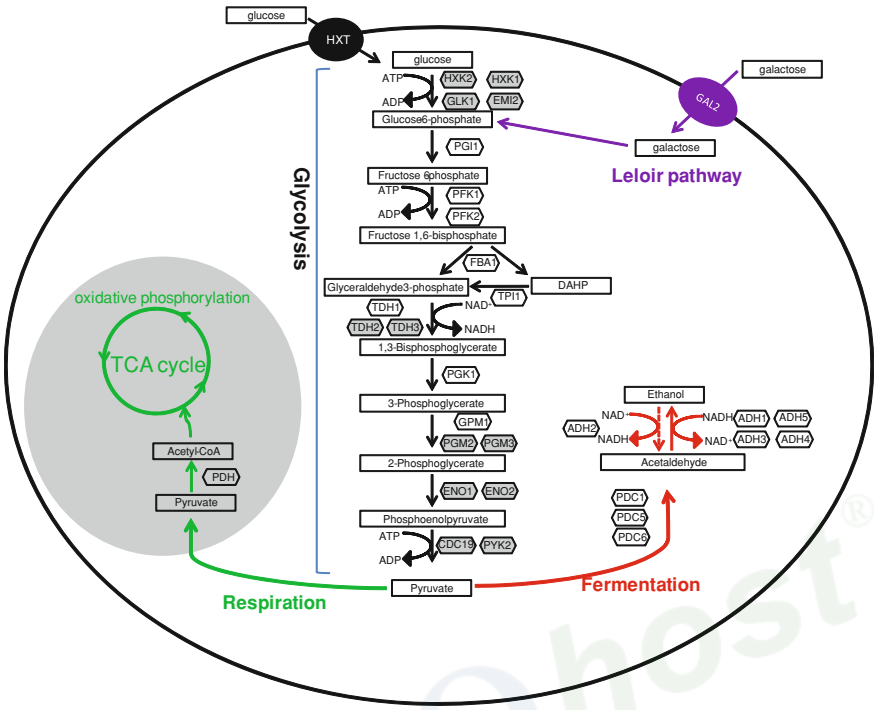

Fig. 5.1 Schematic illustration of major carbohydrate metabolic pathways in *S. cerevisiae*: glycolysis, alcohol fermentation, the TCA cycle, and the galactose Leloir pathway. Enzymes catalyzing each reaction are denoted by gene names in hexagons. Hexagons shaded in *gray* indicate paralogous enzymes derived from the whole genome duplication

traits. Many comparative genomics and evolutionary genetics studies relevant to yeast carbohydrate metabolism focused on the glucose and galactose pathways. This chapter will provide a brief overview of recent advances in our understanding of the genetic basis of yeast glucose and galactose metabolisms.

5.2 Copy Number Variation and Sequence Divergence of Genes Involved in Glucose Metabolism

5.2.1 Copy Number Variation of Hexose Transporter Genes

Glucose does not freely permeate cellular membranes, so the first step in glucose metabolism is to transport glucose across cellular membranes, which requires the aid of transporters (Fig. 5.1). In *S. cerevisiae*, glucose uptake is carried out by a large group of hexose transporters (Hxts) (Boles and Hollenberg 1997; Ozcan and

Johnston 1999). The hexose transporters belong to a superfamily of monosaccharide facilitators that are highly conserved in eukaryotes (Reifenberger et al. 1995). Twenty *HXT* genes have been identified in *S. cerevisiae*: *HXT1–HXT17* encode 17 glucose transporters, *GAL2* encodes a galactose transporter, and *SNF3* and *RGT2* encode two glucose sensors (Boles and Hollenberg 1997).

Phylogenetic analyses of the *HXT* genes from 23 completely sequenced fungal genomes revealed that the 20 *S. cerevisiae HXT* genes were separated into two groups prior to the divergence of major fungal groups (Lin and Li 2011b). According to functional characterizations of these *HXT* genes in *S. cerevisiae*, the two groups play distinct roles in glucose metabolism. One group comprises *S. cerevisiae SNF3* and *RGT2*, which encode sensors that recognize the concentration of extracellular glucose for the induction of *HXT* expression (Boles and Hollenberg 1997; Ozcan and Johnston 1999); this group is called the *Sensor* subfamily (Lin and Li 2011b). All of the yeast species examined, except for the fission yeast *Sch. pombe*, have 1–4 *Sensor* genes and most species contain two copies (Table 5.1), suggesting that the number of *Sensor* genes remain largely constant during the evolution of yeast species. *S. cerevisiae SNF3* and *RGT2* appear to be derived from the whole genome duplication (WGD) event that occurred about 100 million years ago (mya) (Wolfe and Shields 1997; Kellis et al. 2004; Lin and Li 2011b). The *Sensor* gene is absent from the genome of *Sch. pombe*, probably because *Sch. pombe* primarily detects glucose via a cAMP-signaling pathway (Hoffman 2005).

The remaining 18 *S. cerevisiae HXT* genes (*HXT1–17* and *GAL2*) form the other *HXT* group. Because the products of these genes are directly involved in transporting glucose or galactose across cellular membrane, this group was named the *Transporter* subfamily (Lin and Li 2011b). In contrast to the *Sensor* group, the number of *Transporter* genes varies substantially among yeast species (Table 5.1). Evolutionary analyses showed that the *Transporter* genes originated from a single gene in the common ancestor of hemiascomycete (Saccharomycotina) yeasts. The copy number of *Transporter* genes has continually increased during the evolution of the *S. cerevisiae* lineage, starting from a single copy in *Yarrowia lipolytica*, which is most distantly related to *S. cerevisiae* in hemiascomycetes, to 2–8 copies in *Kluyveromyces lactis* and *Lachancea kluyveri* (*Saccharomyces kluyveri*), and to 18–19 copies in the *Saccharomyces sensu stricto* species (including *S. cerevisiae*, *S. paradoxus* and *S. bayanus*) (Table 5.1). All post-WGD species examined contained at least 10 *Transporter* genes except for *Vanderwaltozyma polyspora* (*Kluyveromyces polysporus*), a species most distantly related to *S. cerevisiae* among the post-WGD yeasts (Kurtzman and Robnett 2003), which contains only 5 *Transporter* genes.

Some studies proposed that the WGD event was a major source for the expansion of *Transporter* genes. As the first study that proposed WGD in the ancestor of *S. cerevisiae*, Wolfe and Shields noticed that sugar transporters are among the few gene families that are enriched with WGD gene pairs (Wolfe and Shields 1997). After examining the hexose transporter genes in 7 hemiascomycete yeasts, Conant and Wolfe (2007) found that all of the post-WGD species have at least twice as

Table 5.1 Copy number variations of the *Sensor* and *Transporter* subfamily genes in the HXT gene family

Phylum/Class	Species	Sensors	Transporters
Hemiascomycetes	*Saccharomyces cerevisiae*[*]	2	18
	Saccharomyces paradoxus[*]	2	19
	Saccharomyces mikatae[*]	2	19
	Saccharomyces bayanus[*]	2	18
	Naumovozyma castellii[*]	2	16
	Candida glabrata[*]	2	10
	Vanderwaltozyma polyspora[†]	1	5
	Zygosaccharomyces rouxii[†]	1	4
	Lachancea thermotolerans[†]	4	5
	Lachancea kluyveri[†]	2	7
	Lachancea waltii[†]	2	8
	Kluyveromyces lactis	1	2
	Eremothecium gossypii	1	4
	Candida albicans	2	4
	Debaryomyces hansenii	2	3
	Scheffersomyces stipitis	2	4
	Yarrowia lipolytica	2	1
Dothideomycetes	*Phaeosphaeria nodorum*	1	3
Sordariomycetes	*Gibberella zeae*	1	3
	Magnaporthe grisea	2	1
Schizosaccharomycetes	*Schizosaccharomyces pombe*[*]	0	8
Eurotiomycetes	*Aspergillus fumigatus*	3	3
Basidiomycota	*Ustilago maydis*	2	1

Species name underlined are post-WGD yeasts. The data of gene copy number were retrieved from (Lin and Li 2011b)

[*] Crabtree-positive species

[†] Medium Crabtree effect species (van der Sluis et al. 2000; Moller et al. 2002; Christen and Sauer 2011; Hagman et al. 2013)

many hexose transporter genes as the three pre-WGD species, indicating the impact of WGD. However, based on phylogenetic analysis and gene syntenic structures from 23 species, Lin and Li showed that only two pairs of *Transporters* genes were produced by WGD and most duplicate pairs had become lost immediately after the WGD (Lin and Li 2011b). Most of the *Transporters* genes present in the extant *S. cerevisiae* genome were generated by tandem duplication or can be attributed to the dynamic pattern of telomeric regions where 8 *Transporter* genes are located (Lin and Li 2011b). The *Transporter* genes were also expanded from a single copy to eight copies (*GTH1-GTH8*) in the *Sch. pombe* lineage. Similar to *S. cerevisiae*, *Sch. pombe* is capable of aerobic fermentation in the presence of excess sugars (Alexander and Jeffries 1990). However, *Sch. pombe* did not experience a WGD, suggesting that other mechanisms rather than WGD led to its increase in the number of *Transporter* genes. Four of the eight *Transporter* gene in *Sch. pombe* are tandemly arrayed on chromosome III, indicating that they were produced by a series of tandem duplication events (Lin and Li 2011b).

Past studies have suggested that glucose uptake is the major rate-limiting step in glycolysis and largely controls glucose metabolism activities (Gancedo and Serrano 1989; Diderich et al. 1999; Ye et al. 1999; Pritchard and Kell 2002; Elbing et al. 2004; Otterstedt et al. 2004; Conant and Wolfe 2007). Otterstedt et al. found that an *S. cerevisiae* strain with very limited capacity to transport hexoses switched to respiration in the presence of oxygen (Otterstedt et al. 2004). Replacing the *S. cerevisiae HXT1–17* genes by a chimera *HXT* gene decreased its ethanol production or even caused a switch to fully respiratory metabolism due to reduced glucose consumption rates (Elbing et al. 2004; Otterstedt et al. 2004). Moreover, when yeast cells were grown under glucose limitation, spontaneous duplication of hexose transporters was observed (Brown et al. 1998). The significant expansion of *Transporter* genes had independently occurred in both *Saccharomyces* and *Schizosaccharomyces* lineages, which in parallel evolved aerobic fermentation (Lin and Li 2011b). Furthermore, there is a significant positive correlation between the number of *Transporter* genes and efficiency to produce ethanol from glucose (Lin and Li 2011b). In cells that are operating near their maximal glucose uptake rates, an increase in *Transporter* genes confers a selective advantage in glucose-rich environments to support higher growth rates (Brown et al. 1998). These lines of evidence support the view that the expansion of *Transporter* genes had facilitated the evolution of aerobic fermentation in the two different lineages.

5.2.2 *Copy Number Variation of Genes Involved in Glycolysis*

A glucose molecule is converted into two molecules of pyruvate through a series of reactions in glycolysis, which is believed to be among the oldest biochemical pathways and is highly conserved in prokaryotes and eukaryotes. Several studies have shown that the occurrence of WGD in the hemiascomycete lineage has had profound impacts on the enzyme dosages involved in glycolysis. Soon after the WGD, there was a period of rapid losses of duplicate genes (Scannell et al. 2006). Most duplicate genes produced by the WGD have been lost in the post-WGD species, and less than 10 % of WGD gene pairs have remained in the genome of *S. cerevisiae* (Wolfe and Shields 1997; Kellis et al. 2004; Thomson et al. 2005; Conant and Wolfe 2007). Conant and Wolfe (2007) found that there are six WGD duplicate pairs of genes that have been maintained in the five out of the ten reactions of glycolysis in *S. cerevisiae* (Fig. 5.1). The retained WGD pairs are not the same among post-WGD species. In general, each post-WGD species has preserved 5–6 WGD pairs, but only one duplicate pair (*GLK1* and *EMI2*) is retained in all pot-WGD species examined (Table 5.2). As only 551 WGD duplicate pairs were preserved in *S. cerevisiae*, it is unlikely that the glycolysis genes were preserved in duplicates at the same frequency as the remainder of the genome (Conant and Wolfe 2007).

Table 5.2 Number of glycolytic gene paralogs in hemiascomycete yeasts

SCE Gene 1	SCE Gene 2	SMI	SBA	CGL	KAF	KNA	NCA	TBL	VPO	ZRO	KLA	EGO	LKL	LTH	LWA
HXK1	HXK2	2	2	2	1	2	2	2	1	1	1	1	1	1	1
GLK1	EMI2	2	2	2	2	2	2	2	2	1	1	1	1	1	1
PGI1		1	1	1	1	1	1	1	1	1	1	1	1	1	1
PFK1		1	1	1	1	1	1	1	1	1	1	1	1	1	1
PFK2		1	1	2	2	2	1	1	2	1	1	1	1	1	1
FBA1		1	1	1	1	1	1	2	2	1	1	1	1	1	1
TPI1		1	1	1	1	1	1	1	1	1	1	1	1	1	1
TDH1		1	1	0	0	0	0	0	0	0	0	0	0	0	0
TDH2	TDH3	2	2	1	2	1	2	2	1	1	1	0	1	1	1
PGK1		1	1	1	2	2	2	1	1	1	1	1	1	1	1
GPM1		1	1	1	1	1	1	1	1	1	1	1	1	1	1
GPM2	GPM3	2	2	1	1	1	1	1	1	1	1	1	1	1	1
ENO1	ENO2	2	2	2	2	1	2	2	2	1	1	1	1	1	1
CDC19	PYK2	2	2	2	1	2	2	1	2	1	1	1	1	1	1

Species with the name underlined are post-WGD yeasts. The data were retrieved from yeast Gene Order Browser (Byrne and Wolfe 2005). SCE *Saccharomyces cerevisiae*; SMI *S. mikatae*; SBA *S. bayanus*; CGL *Candida glabrata*; KAF *Kazachstania africana*; KNA *Kazachstania naganishii*; NCA *Naumovozyma castellii*; TBL *Tetrapisispora blattae*; VPO *Vanderwaltozyma polyspora*; ZRO *Zygosaccharomyces rouxii*; KLA *Kluyveromyces lactis*; EGO *Eremothecium gossypii*; LKL *Lachancea kluyveri*; LTH *L. thermotolerans*; LWA *L. waltii*

The significant higher survival rate of WGD pairs in glycolysis than other pathways might have increased the relative levels of glycolytic enzymes. Papp et al. proposed that the retention of gene duplicates is better explained by selection for high enzymatic flux (Papp et al. 2004). A simulation of the effect of increased concentration of enzymes on glycolytic flux revealed that, when the concentration of glycolytic enzyme increases from 65 to 100 %, the end product of glycolysis pyruvate is increased by 17 % (Conant and Wolfe 2007). This observation is consistent with several studies proposing that the WGD event enhanced *S. cerevisiae*'s ability to metabolize glucose (Wolfe and Shields 1997; Wolfe 2004; Liti and Louis 2005). Pyruvate is the branching point between respiration and fermentation (Fig. 5.1). Pyruvate decarboxylase (Pdc) and pyruvate dehydrogenase (Pdh) compete for pyruvate, and the destiny of pyruvate depends on the consequence of substrate competition. Due to the different inherent kinetics between the two enzyme complexes, increasing the pyruvate concentration increases relative flux through Pdc and thus the fermentation pathway (Conant and Wolfe 2007). Therefore, more pyruvate molecules are directed to the fermentation pathway in post-WGD species as a result of increased glycolytic flux. Another reason for increased pyruvate rerouting to the fermentation pathway is probably the limited availability of mitochondria. Unlike fermentation which occurs in the cytosol, respiration takes place in mitochondria. The increased glycolytic flux might exceed the capacity of mitochondria due to the limitation in the number and size of mitochondria (Pronk et al. 1996). Indeed, it appears that during aerobic respiration, yeast mitochondria are larger and closer to the cell membrane than during anaerobic growth, possibly because this location is more efficient for oxygen uptake (Hoffmann and Avers 1973; Jensen et al. 2000). The WGD did not increase the number of mitochondria or make mitochondria larger, so the increased glycolytic flux might have exceeded the respiratory capacity of mitochondria. Furthermore, unlike the high retention rate of WGD pairs in glycolysis pathway genes, the WGD pairs were preserved in only two out of nine reactions in the TCA cycle in *S. cerevisiae* (Conant and Wolfe 2007).

A survey of over 40 yeast species both with and without the WGD indicates that the presence of the Crabtree effect is strongly associated with yeasts with the WGD (Merico et al. 2007). Another study also found a general, though weak, trend for higher rates of ethanol production in post-WGD yeasts than in pre-WGD yeasts (Blank et al. 2005). These studies suggested that the WGD event played a significant role in the adaptation of *S. cerevisiae* toward aerobic fermentation (Wolfe and Shields 1997; Kellis et al. 2004; Thomson et al. 2005; Conant and Wolfe 2007). However, not all aerobic fermentative species have experienced WGD. For example, *Dekkera bruxellensis*, which is a pre-WGD species separated from the *Saccharomyces* lineage more than 200 mya, also efficiently makes, accumulates and consumes ethanol (Rozpedowska et al. 2011). In addition, the fission yeast *Sch. pombe*, which also predominantly assimilates glucose through the fermentation pathway, diverged from the hemiascomycete lineage about 300 mya and has not experienced WGD (Wood et al. 2002). Based on the analysis of the Crabtree

effect for over 40 species in 12 genera of hemiascomycete yeasts, Hagman et al. found that many pre-WGD species demonstrate an intermediate level of Crabtree effect (Hagman et al. 2013). They argued that the evolution of Crabtree effect is gradual process or at least a two-step "invention". The progressive evolution of aerobic fermentation coincides with gradual duplication of hexose transporter genes (Lin and Li 2011b). The WGD event and regulatory rewiring of respiration-related genes, which occurred at different time points, have further strengthened the Crabtree effect in the lineages of *S. cerevisiae* (Hagman et al. 2013). Therefore, WGD might have facilitated the evolution of aerobic fermentation, but apparently it is not a prerequisite factor.

5.2.3 Copy Number Variation of Genes Involved in Fermentation

In aerobic fermentative species, most pyruvate molecules remain in cytosol and are converted into acetaldehyde by pyruvate decarboxylase (Pdc). The *S. cerevisiae* genome contains three copies of the *Pdc* encoding gene (*PDC1*, *PDC5* and *PDC6*), though *PDC1* encodes the major enzyme in this reaction and is highly expressed in rich medium. The evolutionary history of *PDC* genes in hemiascomycetes suggests that the three *PDC* genes were generated by two consecutive duplication events in the common ancestor of the *sensu stricto* species (Fig. 5.2a). After duplications, loss of *PDC* genes have been detected in some *sensu stricto* species. For example, only *PDC1* is present in the genome of *S. mikatae* (Fig. 5.2a), and loss of *PDC6* genes was observed in *S. kudriavzevii* (Scannell et al. 2011). However, comparing with other *sensu stricto* species, no significant difference in ability of ethanol production was detected in *S. mikatae* (Hagman et al. 2013). In general, 2–3 copies of *PDC* genes are present in most aerobic fermentative species. The number of *PDC* genes in the respiratory species varies from 1 to 4 copies. For example, only a single *PDC* gene is found in *K. lactis* (Bianchi et al. 1996), while three copies are found in *Debaryomyces hansenii* (Fig. 5.2a). Therefore, there is no significant increase in the number of *PDC* genes in fermentative species, in agreement with a previous finding that the number of *PDC* genes is not correlated with the intensity of alcoholic fermentation (Moller et al. 2004).

The second and the last step of reaction in the fermentation pathway is converting acetaldehyde into ethanol and recycling the NADH generated during glycolysis (Fig. 5.1). In most species, this reaction is catalyzed by alcohol dehydrogenase (*Adh*). In *S. cerevisiae*, there are five genes that encode *Adh*, *ADH1–ADH5*. *Adh1p* is the major enzyme responsible for converting acetaldehyde into ethanol, while Adh2p catalyzes the reverse reaction to consume ethanol (Leskovac et al. 2002). *ADH1* and *ADH2* were derived from a duplication event prior to the divergence of *sensu stricto* yeast species (Thomson et al. 2005). Thomson et al. (2005) have reconstructed the last common ancestor of *Adh1p* and *Adh2p*, which is

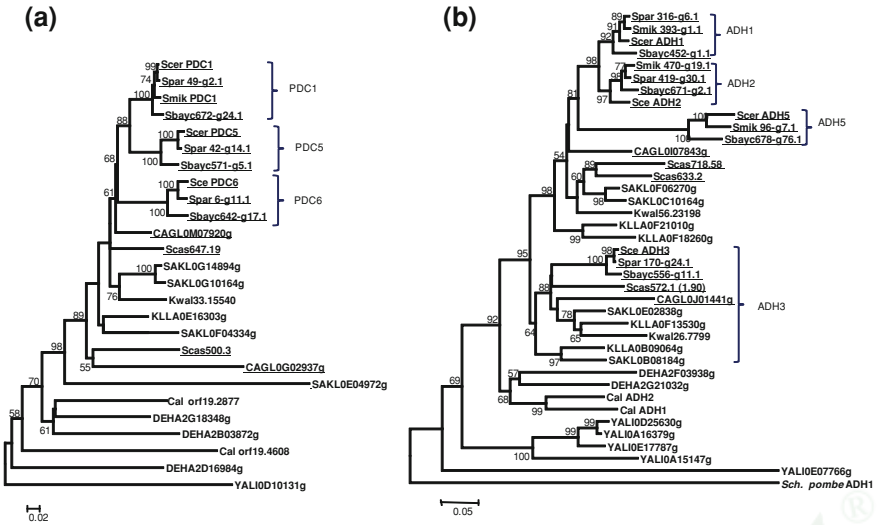

Fig. 5.2 The evolutionary history and copy number variations of the PDC and ADH gene families in hemiascomycete yeasts. **a** Phylogenetic tree of the PDC gene family. **b** Phylogenetic tree of the ADH gene family. The three PDC genes (*PDC1*, *PDC5* and *PDC6*) and the three ADH genes (*ADH1*, *ADH2* and *ADH5*) in *S. cerevisiae* were produced by two consecutive duplications prior to divergence of the *Senso stricto* group. The *S. cerevisiae ADH4* is a distant relative to the other four ADH genes and was not included in the tree. Both phylogenetic trees were constructed using the Neighbor-Joining (NJ) method with 1,000 bootstrap replicates. The evolutionary distances were computed using the JTT matrix-based method. Species names are abbreviated as follows: Scer: *Saccharomyces cerevisiae*; Spar: *S. paradoxous*; Smik: *S. mikatae*; Sbay: *S. bayanus*; CAGL: *Candida glabrata*; Scas: *S. castellii* (*Naumovozyma castellii*); KLLA: *Kluyveromyces lactis*; SAKL: *S. kluyveri* (*Lachancea kluyveri*); Kwal: *K. waltii* (*Lachancea waltii*); Cal: *C. albicans*; DEHA: *Debaryomyces hansenii*; YALI: *Yarrowia lipolytica*

called *AdhA*. The *AdhA* protein has a high Michaelis constant K_M for ethanol, which is similar to *Adh1p*, suggesting that *AdhA* was optimized to make ethanol (Thomson et al. 2005). Therefore, *Adh1p* maintains the ancestral function for making ethanol, while *Adh2p* has evolved a much lower K_M for ethanol, which is optimized for consuming ethanol. Because many microorganisms cannot grow in a high concentration of ethanol, accumulating ethanol may help yeasts to outcompete their competitors for fruit resources (Piskur et al. 2006). Thus, the birth of *ADH2* might have enabled yeasts to tolerate a higher concentration of ethanol produced from fermentation by efficiently consuming ethanol after depletion of sugars (Thomson et al. 2005). By including more hemiascomycete species, we show here (Fig. 5.2b) that the gene duplication event that produced *ADH1* and *ADH2* occurred after the split between the common ancestor of *sensu stricto* yeast species and other post-WGD yeasts including *Candida glabrata* and *N. castelli*. Because *C. glabrata* and *N. castelli* are both Crabtree-positive species (Merico

et al. 2007), it means that *ADH2* is not necessary for the Crabtree effect, suggesting that the birth of *ADH2* might have increased ethanol tolerance, but it was probably not essential for aerobic fermentation.

5.2.4 Elevated Evolutionary Rates and Biased Codon Usage of Respiration-Related Genes

In view of the fact that post-WGD yeasts predominantly assimilate glucose through the fermentation pathway, the role of mitochondria in generating energy for cellular growth appears to be weakened (Merico et al. 2007). Kellis et al. (2004) noticed that some WGD gene pairs have accelerated evolutionary rate at nucleotide level, but not at amino acid level. For example, the pyruvate kinase genes *CDC19* and *PYK2* were produced by WGD, and *PYK2* shows a three-fold acceleration in substitution rate at degenerate third-codon positions (Kellis et al. 2004). Jiang et al. (2008) calculated the rate of non-synonymous substitution (d_N) for 2,603 one-to-one orthologous genes, including 296 nuclear genes for mitochondrial proteins, among six post-WGD yeast species and three closely related pre-WGD yeast species. They found that the rates of sequence divergence of mitochondrial genes are very similar within the post-WGD species and within the pre-WGD species. However, the average evolutionary distance for mitochondrial genes for the post-WGD species pairs is about 13 % higher than that for the pre-WGD species pairs (Jiang et al. 2008), supporting the view that genes involved in mitochondrial functions have experienced relaxation of functional constraints in post-WGD yeasts.

It was postulated that to ensure efficient and accurate translation, highly expressed genes tend to have strong codon usage bias (Ikemura 1981, 1982). The codon usage bias can be reduced if the gene product undergoes a reduction in functional constraint (Akashi 1997; Gu et al. 2005). Therefore, if mitochondrial energy production became less important to the post-WGD species, a decreased codon usage bias is expected for the mitochondrial genes of these species. The codon usage bias for six genes encoding the electron transport chain cytochrome-c (CYC) from five yeasts was found to be stronger in aerobic respiration species than in fermentative species (Freire-Picos et al. 1994). The difference in codon usage of *CYC* genes was correlated with their difference in mRNA level between the two types of yeasts (Freire-Picos et al. 1994). Jiang et al. also found that mitochondrial genes displayed significantly stronger codon usage bias than non-mitochondrial genes in all three studied pre-WGD species. In contrast, there was no significant difference in codon usage bias between mitochondrial and non-mitochondrial genes for all the studied post-WGD species (except for *V. polyspora*). Therefore, the accelerated evolution of mitochondrial function genes in post-WGD yeasts at the nucleotide level appears to reflect a relaxation in selection on the codon usage.

5.3 Comparative Studies of Gene Regulation in Carbohydrate Metabolism

5.3.1 Differential Expression of Genes Involved in Carbohydrate Metabolism Among Yeasts

The major difference between the two glucose metabolism styles depends on how pyruvate is degraded. In respiratory yeasts, most pyruvate enters mitochondria and is completely degraded to CO_2 and H_2O under aerobic conditions, while in fermentative species, most pyruvate remains in the cytosol and is converted into ethanol and CO_2. Because the respiration-related genes have been retained in both respiratory and fermentative species, it means that the evolution of aerobic fermentation required modifications of the regulation of respiration-related genes. Differential expression of genes involved in glucose metabolism between fermentative and respiratory species has been observed in several studies of individual genes. The respiration-related genes, such as *CYC1*, *QCR7,* and *QCR8*, are highly expressed in the presence of oxygen in the Crabtree-negative species *K. lactis* (Freire-Picos et al. 1995; Mulder et al. 1995). In contrast, although the expression of glycolysis and fermentation-related genes is induced in *S. cerevisiae* under growth on glucose, the expression of respiration-related genes is repressed (Holland and Holland 1978; Schmitt et al. 1983; Forsburg and Guarente 1989; DeRisi et al. 1997; Carlson 1999).

The global modification of regulatory control of respiration-related genes has been confirmed by recent studies based on large sets of genome-wide gene expression data from yeasts. Significant expression differences in genes related to carbohydrate metabolism and respiratory functions have been detected by heterologous DNA arrays between *S. cerevisiae* and *K. lactis* growing in a sugar-rich medium (Becerra et al. 2004). Ihmels et al. (2005) compared datasets of 1,000 and 198 published genome-wide expression profiles between *S. cerevisiae* and the human pathogen *Candida albicans* (Ihmels et al. 2002). *C. albicans*, which diverged from the *S. cerevisiae* lineage approximately 100–300 million years ago, is predominantly Crabtree negative. Because the large number of cytosolic ribosomal proteins (CRP) genes are coherently expressed under different conditions and show a strong correlation with cell growth (Mager and Planta 1991; Gasch et al. 2000), they can be used as a good proxy to evaluate the expression profiles for different sets of genes. Ihmels et al. found that genes coding for mitochondrial ribosomal proteins (MRP) and CRP display a strongly correlated expression pattern in *C. albicans*, but this correlation is lost in the fermentative yeast *S. cerevisiae* (Ihmels et al. 2005). Instead, the expression of the 72 MRP genes in *S. cerevisiae* exhibits a distinct correlation with that of genes induced in response to environmental stress conditions. Because the Crabtree positive yeasts evolved from respiratory yeasts, the authors concluded that the regulation of MRP genes in

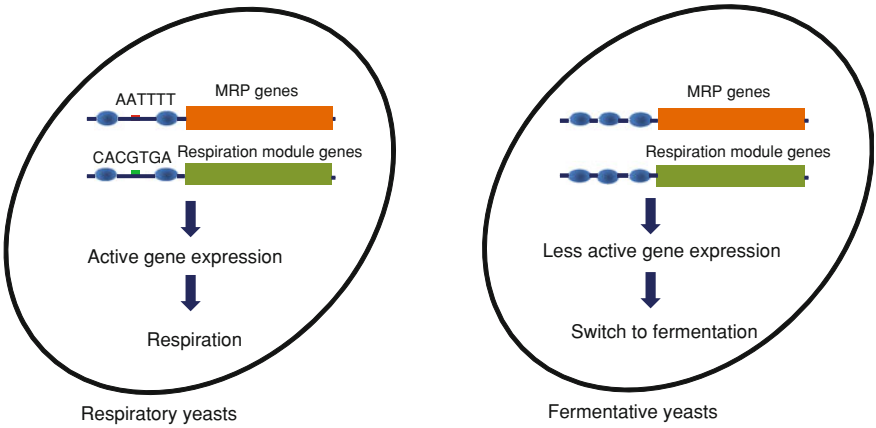

Fig. 5.3 A schematic illustration of the genetic basis underlying the regulatory rewiring of respiration-related genes in fermentative yeasts. In respiratory yeasts, the expression of MRP genes is activated by a transcription factor through binding to the AATTTT motif. The Respiration module genes, such as CYC1, ATP4, QCR7, and QCR and QCR8, are activated by Cbf1p through its binding to the CACGTGA motif. Both groups of genes appear to have nucleosome-depleted type of promoters. The AATTTT and Cbf1 motifs have been lost and the promoter became nucleosome-occupied in the fermentative species. As a consequence of the changes in promoter chromatin structure and *cis*-regulatory elements, the respiration-related genes are not actively expressed, so that most pyruvate is directed to the fermentation pathway in the Crabtree-positive species

S. cerevisiae has been rewired during the evolution of aerobic fermentation (Ihmels et al. 2005).

Field et al. (2009) reanalyzed the same sets of gene expression data using the Gene Ontology (GO) groups as units and calculated their expression correlation with CRP genes in both *S. cerevisiae* and *C. albicans* (Field et al. 2009). Based on the gene expression correlation with CRP genes, the authors identified 13 GO groups of genes (called "category III" genes) that are differentially expressed between the two species. The "category III" set includes 157 and 146 genes in *S. cerevisiae* and *C. albicans*, respectively. The *S. cerevisiae* "category III" genes include 34 MRP genes as well as genes related to cellular respiration and mitochondrial functions, such as the TCA cycle and oxidative phosphorylation. Similar to the MRPgenes, a high expression correlation between "category III" genes and CRP genes was observed in *C. albicans* but not in *S. cerevisiae* (Field et al. 2009). Therefore, in addition to the MRP genes, the transcriptional regulation of other respiration-related genes has also been reprogrammed during the evolution of *S. cerevisiae* (Fig. 5.3).

A recent study measured the genome-wide gene expression levels growing under the same rich medium in 12 completely sequenced yeasts using tiling arrays (Tsankov et al. 2010). The 12 yeasts include six aerobic fermentative species and six respiratory species, offering an ideal opportunity to identify the genes that have highest significant expression differences between the two types of yeasts. Using

the 82 transcriptional modules (Ihmels et al. 2002) as units, Lin et al. (2013) compared the difference in gene expression level for each module between the two types of yeasts. They found that the modules that include the MRP genes showed only the sixth largest expression difference among the 82 modules based on the two-sample Kolmogorov-Smirnov test. In contrast, a module (Module 5) that includes genes involved in mitochondrial energy generation and phosphorylation oxidation has the largest expression difference between the two types of species. For convenience, Module 5 is thereafter called the Respiration module. *S. cerevisiae* respiration module includes 49 genes, 23 of which overlap with the "category III" genes identified by Field et al. (2009). Thus, the genes associated with mitochondrial energy production, instead of MRP genes, have experienced the most significant changes in gene expression levels during the evolution of aerobic fermentation (Lin et al. 2013).

The fission yeast *Sch. pombe* independently evolved the Crabtree effect. So, it was interesting to see if gene expression reprogramming had also occurred during the evolution of *Sch. pombe*. Lin and Li (2011a) conducted pairwise comparisons for the ~1,000 sets of genome-wide gene expression profiles in *Sch. pombe*, and gene expression data in *S. cerevisiae* and *C. albicans*. They found that the two fermentative species *S. cerevisiae* and *Sch. pombe* are more similar to each other on the genome-wide gene expression patterns than to the respiratory yeast *C. albicans*, although *S. cerevisiae* is evolutionarily closer to *C. albicans*. Lin and Li identified a group of genes that are differentially expressed between *Sch. pombe* and *C. albicans* and most of them are involved in mitochondrial respiration process. In summary, similar to what happened in the *S. cerevisiae* lineage, the evolution of aerobic fermentation in the *Sch. pombe* lineage was also associated with regulatory rewiring of genes involved in the mitochondrial respiration process (Lin and Li 2011a).

5.3.2 Genetic Basis for Gene Expression Reprogramming

Although there are some discrepancies about what genes have experienced regulation reprogramming among studies (Ihmels et al. 2005; Field et al. 2009; Lin et al. 2013), there is a general agreement that all these genes are involved in respiration-related processes. However, with respect to the genetic basis underlying these gene expression divergences, these studies have reached different conclusions. Ihmels et al. found a sequence motif "AATTTT" significantly overrepresented in the promoters of the MRP genes in *C. albicans*, but not in their orthologous genes in *S. cerevisiae* (Ihmels et al. 2005). The AATTTT motif was proposed to be involved in the regulation of rRNA processing genes in *S. cerevisiae*, but the protein binding to this motif remains to be identified (Tavazoie and Church 1998). Therefore, Ihmels et al. concluded that the loss of the AATTTT sequence was associated with global regulatory reprogramming of MRP genes in the *S. cerevisiae* lineage, and contributed to its evolution of aerobic fermentation (Ihmels et al. 2005). The loss of AATTTT

motif was also observed in other species that independently evolved aerobic fermentative ability. As mentioned above, aerobic fermentation evolved independently in the pre-WGD hemiascomycete yeast *D. bruxellensis* (Rozpedowska et al. 2011). Similar to what was observed in *S. cerevisiae*, the AATTTT motif did not exhibit any positional conservation in the promoters of the MRP genes in *D. bruxellensis* (Rozpedowska et al. 2011). The authors suggested that the AATTTT element underwent independent massive losses in the promoter of MRP genes in both *S. cerevisiae* and *D. bruxellensis* (Rozpedowska et al. 2011).

It is worth noting that only intermediate numbers of the AATTTT motif in the promoters of MRP genes were observed in the post-WGD yeast *V. polyspora* (Fekete et al. 2007; Chen et al. 2008; Jiang et al. 2008). Moreover, there are several other genetic and physiological characteristics in *V. polyspora* that are more similar to aerobic respiratory yeasts than to the rest of post-WGD yeasts. For example, only five hexose transporter genes are present in *V. polyspora*, compared to 10–19 copies in other post-WGD species (Lin and Li 2011b). *V. polyspora* also shows a different pattern of mitochondrial gene codon usage bias from the other post-WGD species (Fekete et al. 2007; Chen et al. 2008; Jiang et al. 2008). In addition, other post-WGD species are petite positive, the ability to tolerate the loss of mtDNA, whereas *V. polyspora* is petite negative (Fekete et al. 2007; Chen et al. 2008; Jiang et al. 2008). *V. polyspora* was thus speculated as an intermediate fermentative species that cannot carry out efficient aerobic fermentation (Fekete et al. 2007; Chen et al. 2008; Jiang et al. 2008). Surprisingly, a recent study revealed that *V. polyspora* actually has high fermentation efficiency (Hagman et al. 2013). Hagman et al. suggested that the evolution of aerobic fermentation is a progressive process which involves multiple genetic modifications that gradually remodel the yeast carbon metabolism. It is likely that *V. polyspora* represents an intermediate lineage where some traits are still in transition.

On the other hand, Lin et al. (2013) found that in respiratory yeasts, a motif with the core consensus sequence CACGTGA is prevalent in the promoters of Respiration module genes, but present at a much lower frequency in the promoters of their orthologous genes in the aerobic fermentation species (Fig. 5.3). This motif matches that of Cbf1p binding sites in *S. cerevisiae*; Cbf1p is a transcription factor that regulates chromatin modification (Cai and Davis 1990). The homolog of *S. cerevisiae CBF1* has been characterized in *K. lactis* and the Cbf1p proteins from the two species are functionally interchangeable (Mulder et al. 1994). However, despite the functional conservation of *CBF1*, unlike in *S. cerevisiae*, inactivation of the *CBF1* gene in *K. lactis* is not viable, indicating that the normal function of *CBF1* is essential for *K. lactis* (Mulder et al. 1994). Lin et al. (2013) suggested that Cbf1p is a general activator for the respiration-related genes in respiratory yeasts and the massive losses of the Cbf1 motif in fermentative species have led to the reduced expression of respiration-related genes.

In addition to changes in *cis*-regulatory elements, it was suggested that the distinct nucleosome organizations in the promoters of respiration-related genes between the aerobic respiration and the fermentation yeasts are partly responsible for their expression divergences (Field et al. 2009; Tsankov et al. 2010). In

eukaryotes, DNA is repetitively wrapped around nucleosomes. The presence of nucleosome may hinder the direct interaction between a transcription factor and its binding sites and may, therefore, obstruct the transcriptional initiation of a gene. In *S. cerevisiae*, the binding sites of transcription factors are highly enriched in the nucleosome depleted region of promoters (Lin et al. 2010). Several studies have found that genes with different expression profiles are associated with distinct nucleosome occupancy patterns in the promoter regions (Tirosh and Barkai 2008; Jiang and Pugh 2009). The promoters of constantly expressed genes usually contain a nucleosome-depleted region where most transcription factor-binding sites are located (Yuan et al. 2005; Lee et al. 2007). In contrast, conditionally expressed genes, such as stress-response genes, are associated with nucleosome-occupied promoters (Tirosh and Barkai 2008). Field et al. (2009) compared the promoter nucleosome occupancy patterns among 12 hemiascomycete yeasts and found that the promoters of respiration-related genes tend to be more depleted of nucleosomes in respiratory yeasts than that in aerobic fermentative species. They concluded that in aerobic fermentation yeasts, respiration-related gene promoters have evolved from the nucleosome-depleted type to the nucleosome-occupied type and that this change has contributed to regulatory reprogramming of respiration-related genes and the evolution of aerobic fermentation in the hemiascomycete lineage (Field et al. 2009).

However, it is not clear whether the nucleosome reorganization was the leading or a minor cause for the evolution of aerobic fermentation (Tirosh et al. 2010). Comparative studies of nucleosome occupancy between *S. cerevisiae* and its close relative *S. paradoxus* showed that genes that are associated with diverged nucleosome positions are not more likely to diverge in expression and genes that are differentially expressed are not more likely to diverge in nucleosome positioning (Tirosh et al. 2010). Consistent with Tiroshi et al.'s observation, Lin and Li (2011a) found that changes in nucleosome organization were not coupled with the expression reprogramming of respiration-related genes in *Sch. pombe*. Specifically, although the expression regulation of the respiration-related genes in *Sch. pombe* has been reprogrammed during the evolution of aerobic fermentation, their promoter nucleosome organization remains depleted as in aerobic respiration species (Lin and Li 2011a).

In addition, a recent study suggested that changes in the length of the 5′ untranslated region (5′UTR) were linked to the gene expression divergence of respiration-related genes in *S. cerevisiae* (Lin and Li 2012). The length of 5′UTR varies considerably among the genes in a genome, ranging from a few base pairs to several thousand base pairs (Pesole et al. 2001; Mignone et al. 2002; Nagalakshmi et al. 2008). It has been noticed that genes with different functions show distinct 5′UTR lengths in vertebrates and yeasts (Kozak 1987; Hurowitz and Brown 2003; David et al. 2006; Nagalakshmi et al. 2008; Bruno et al. 2010). In general, genes with a long 5′UTR, such as those involved in development or meiosis, are generally highly and finely regulated, whereas genes with a reduced need for regulation, such as housekeeping genes and the ribosomal subunit genes, usually have shorter 5′UTRs. Lin and Li (2012) examined the association between UTR lengths

and the patterns of gene expression across various conditions in *S. cerevisiae* and *C. albicans* (Berman and Sudbery 2002; Sudbery et al. 2004). They calculated the expression correlation of all eligible GO group genes with CRP genes and found that such correlations are negatively associated with their 5′UTR lengths in both species (Lin and Li 2012). It suggests that genes with a longer 5′UTR tend to have higher expression plasticity under different conditions. Among a few GO groups with the largest increase in 5′UTR length in *S. cerevisiae* and the largest decrease in this gene expression correlation, most of these genes are involved in mitochondrial respiration process. It is well known that the expression of mitochondrial respiration genes has been reprogrammed in *S. cerevisiae* during the evolution of aerobic fermentation. These results suggest that the elongation of 5′UTR might have contributed to this global gene regulation modification process. As it is not clear how changes in 5′UTR length affect gene expression plasticity, it was speculated that an increase in 5′ UTR length may affect the nucleosome occupancy patterns in promoters (Lin and Li 2012).

5.4 Comparative Genomics of Galactose Metabolism

Although glucose is the preferred carbohydrate in yeasts, most yeasts are able to use galactose well. Galactose itself cannot be directly used for glycolysis. As shown in Fig. 5.1, after galactose is transported into cells by galactose permease Gal2p, it needs to be converted into glucose-6-phosphate to enter the glycolysis pathway. The pathway that catalyzes this conversion is also called the Leloir pathway (Johnston 1987; Bhat and Murthy 2001). In *S. cerevisiae*, the Leloir pathway is composed of five enzymes: galactose mutarotase, UDP-galactose-4-epimerase Gal10p, galactokinase Gal1p, galactose-1-phosphate uridyl transferase Gal7p, phosphoglucomutase Pgm1 and Pgm2. The genes encoding the Leloir pathway enzymes (often called the *GAL* genes) are tightly controlled at the transcriptional level in yeasts (Sellick et al. 2008). In the presence of glucose, the *GAL* genes are repressed. The repression is released when glucose is absent, and rapid and high-level activation of the *GAL* genes is triggered by three other proteins, Gal4p, Gal80p, and Gal3p (Sellick et al. 2008). The *GAL* genes are broadly distributed in all eukaryotes, bacteria, and archaea.

A number of yeast species lack the ability to use galactose (Naumov et al. 2000). Hittinger et al. examined the genomic data for seven species that can use galactose and four that cannot. The phylogeny of the 11 yeast species suggests that galactose utilization was present in their common ancestor and the loss of galactose utilization in the four species were due to at least three parallel losses (Hittinger et al. 2004). The seven *GAL* genes (*GAL1*, *GAL2*, *GAL3*, *GAL4*, *GAL7*, *GAL10* and *GAL80*) are present in all species that can use galactose. In contrast, these *GAL* genes are absent in the four species lacking the galactose utilization ability, suggesting the degeneration of the entire pathway (Hittinger et al. 2004).

Remarkably, different galactose utilization abilities due to presence/absence of *GAL* genes are also observed between different strains in a species. The Japanese strains of *Saccharomyces kudriavzevii*, a close relative of *S. cerevisiae*, lack the galactose utilization ability. All the seven *GAL* genes in the Japanese strains of *S. kudriavzevii* were heavily degenerated and became pseudogenes soon after the split between the *S. kudriavzevii* and *S. cerevisiae lineages* (Hittinger et al. 2004). However, the *S. kudriavzevii* strains isolated from Portugal were found to be capable of utilizing galactose because of the presence of six functional *GAL* genes in their genomes (the *GAL3* gene is absent) (Hittinger et al. 2010). Because none of the functional *GAL* genes in the Portuguese strains appeared to be acquired from other species, the polymorphisms in these *GAL* genes may have co-existed throughout the evolutionary history of *S. kudriavzevii*, providing a classic example of balancing selection on the multi-loci gene network.

In summary, these lines of evidence revealed a tight correlation between the ability to use galactose and the presence of the *GAL* genes in the genome. This strong correlation of genomic content and galactose utilization probably reflects the adaptation of yeasts to their own niches in which the galactose content varies substantially (Gross and Acosta 1991). The independent losses of all *GAL* pathway genes in multiple lineages were probably facilitated by the clustered organization of these genes. In most species, the *GAL* genes are not clustered. Slot and Rokas found that the genes encoding three major enzymes of the Leloir pathway (*GAL1*, *GAL7,* and *GAL10*) have independently become clustered in four different fungal lineages by different mechanisms (Slot and Rokas 2010). Moreover, a significant higher rate of *GAL* pathway gene loss in the species with clustered *GALs* than in those without clustered *GAL* genes was observed, suggesting that the adaptation of fungal species to different environments by gain or loss of galactose utilization ability could be facilitated by the clustering of *GAL* genes (Slot and Rokas 2010). However, even though the structural members and arrangement of *GAL* gene cluster are conserved among different yeast lineages, their regulatory circuits have been rewired in the Hemiascomycete lineage (Martchenko et al. 2007). In *S. cerevisiae*, the *GAL* genes are activated by Gal4p and repressed by Mig1p through binding to the Gal4 or Mig1 binding sites in their promoters (Lohr et al. 1995). Interestingly, the Gal4 and Mig1 binding sites are absent from the *C. albicans GAL* genes clusters (Martchenko et al. 2007). In contrast, it is found that the expression of *GAL* genes in *C. albicans* is activated by Cph1p, a homolog of *S. cerevisiae* transcription factor Ste12p (Martchenko et al. 2007). By comparing the promoter sequences of *GAL* genes in 11 yeast species, the author proposed that Gal4p and Mig1p were recruited to co-regulate *GAL* genes with Cph1p prior to WGD. The Cph1 binding sites were eventually lost during evolution of *S. cerevisiae* lineage, resulting in the switch of the control *GAL* gene from Chp1p to Gal4p and Mig1p (Martchenko et al. 2007).

As discussed above, the WGD event was considered to have a strong impact on the glucose metabolism. Similarly, it may also have enhanced the ability of galactose utilization. In *S. cerevisiae*, the galactokinase gene *GAL1* and the co-inducer gene *GAL3* are paralogous genes that arose from a single bifunctional

ancestral gene by the WGD (Wolfe and Shields 1997; Hittinger et al. 2004; Kellis et al. 2004). The bifuntional gene is still present in some pre-WGD yeasts, such as *K. lactis*. In *S. cerevisiae*, the expression of *GAL1* is highly induced up to 1,000 fold, in the presence of galactose, while *GAL3* is only induced three–five-fold. The sharply different regulations on the two paralogous genes are believed to be advantageous for galactose utilization. The single bifunctional gene in *K. lactis* may be subject to adaptive conflict at the level of transcriptional regulation. The adaptive conflict appeared to be resolved by the WGD event that produced an extra copy of the *GAL* gene. Specific modifications on the promoters of the *GAL1* and *GAL3* genes allowed them to have totally different ranges of transcription in post-WGD species (Hittinger and Carroll 2007).

5.5 Conclusions and Prospects

Different yeasts show highly distinct preferences in carbohydrate metabolism. Recent comparative genomics and bioinformatics studies revealed that different glucose and galactose metabolisms were associated with changes in genomic content and regulatory landscape. While the evolution of galactose utilization was mainly due to gain or loss of *GAL* genes and changes in *cis*-regulatory elements, the evolution of different glucose metabolisms in yeasts was influenced by many different factors including changes in gene copy number, *cis*-regulatory elements, promoter chromatin structure, and 5′UTR length. The expansion of glucose transporter genes and the high retention rate of WGD gene pairs in glycolytic enzymes have increased the glycolytic flux in the aerobic fermentative species. In addition, the switch from the respiration pathway to the fermentation pathway also required regulatory rewiring of genes involved in mitochondrial functions, so that most pyruvate is directed to the fermentation pathway. However, it would only make sense if the ancestral yeasts had already evolved the active and highly efficient fermentation pathway under aerobic conditions prior to global repression of mitochondrial function, or the yeast cells would suffer selective disadvantages due to shortage of energy and reduced growth rate. Most studies so far focused on the glycolysis and mitochondrial respiration pathways. It is not known if the genes involved in fermentation reactions have been activated in the presence of oxygen or the enzyme activities have been greatly improved at an early evolutionary stage of aerobic fermentation. Further work will be needed to study what evolutionary changes in the fermentation pathway have triggered the switch from the respiratory to the fermentative style.

Acknowledgments We thank Ken Wolfe and Chris Hittinger for their valuable comments.

References

Akashi H (1997) Codon bias evolution in *Drosophila*. Population genetics of mutation-selection drift. Gene 205:269–278

Alexander MA, Jeffries TW (1990) Respiratory efficiency and metabolite partitioning as regulatory phenomena in yeasts. Enzyme and Microb Technol 12:2–19

Barnett JA, Payne RW, Yarrow D (2000) In yeasts: characteristics and identification, 3rd (edn). Cambridge University Press, Cambridge

Becerra M, Tarrio N, Gonzalez-Siso MI, Cerdan ME (2004) Genome-wide analysis of *Kluyveromyces lactis* in wild-type and rag2 mutant strains. Genome 47:970–978

Bhat PJ, Murthy TV (2001) Transcriptional control of the GAL/MEL regulon of yeast *Saccharomyces cerevisiae*: mechanism of galactose-mediated signal transduction. Mol Microbiol 40:1059–1066

Bianchi MM, Tizzani L, Destruelle M, Frontali L, Wesolowski-Louvel M (1996) The 'petite-negative' yeast *Kluyveromyces lactis* has a single gene expressing pyruvate decarboxylase activity. Mol Microbiol 19:27–36

Blank LM, Lehmbeck F, Sauer U (2005) Metabolic-flux and network analysis in fourteen hemiascomycetous yeasts. FEMS Yeast Res 5:545–558

Boles E, Hollenberg CP (1997) The molecular genetics of hexose transport in yeasts. FEMS Microbiol Rev 21:85–111

Brown CJ, Todd KM, Rosenzweig RF (1998) Multiple duplications of yeast hexose transport genes in response to selection in a glucose-limited environment. Mol Biol Evol 15:931–942

Bruno VM, Wang Z, Marjani SL, Euskirchen GM, Martin J, Sherlock G et al (2010) Comprehensive annotation of the transcriptome of the human fungal pathogen *Candida albicans* using RNA-seq. Genome Res 20:1451–1458

Byrne KP, Wolfe KH (2005) The yeast gene order browser: combining curated homology and syntenic context reveals gene fate in polyploid species. Genome Res 15:1456–1461

Cai M, Davis RW (1990) Yeast centromere binding protein CBF1, of the helix-loop-helix protein family, is required for chromosome stability and methionine prototrophy. Cell 61:437–446

Carlson M (1999) Glucose repression in yeast. Curr Opin Microbiol 2:202–207

Chen H, Xu L, Gu Z (2008) Regulation dynamics of WGD genes during yeast metabolic oscillation. Mol Biol Evol 25:2513–2516

Christen S, Sauer U (2011) Intracellular characterization of aerobic glucose metabolism in seven yeast species by 13C flux analysis and metabolomics. FEMS Yeast Res 11:263–272

Conant GC, Wolfe KH (2007) Increased glycolytic flux as an outcome of whole-genome duplication in yeast. Mol Syst Biol 3:129

David L, Huber W, Granovskaia M, Toedling J, Palm CJ, Bofkin L et al (2006) A high-resolution map of transcription in the yeast genome. Proc Natl Acad Sci U S A 103:5320–5325

De Deken RH (1966) The Crabtree effect: a regulatory system in yeast. J Gen Microbiol 44:149–156

DeRisi JL, Iyer VR, Brown PO (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. Science 278:680–686

Diderich JA, Teusink B, Valkier J, Anjos J, Spencer-Martins I, van Dam K et al (1999) Strategies to determine the extent of control exerted by glucose transport on glycolytic flux in the yeast *Saccharomyces bayanus*. Microbiology 145(Pt 12):3447–3454

Dujon B (2010) Yeast evolutionary genomics. Nat Rev Genet 11:512–524

Elbing K, Larsson C, Bill RM, Albers E, Snoep JL, Boles E et al (2004) Role of hexose transport in control of glycolytic flux in *Saccharomyces cerevisiae*. Appl Environ Microbiol 70:5323–5330

Engel SR, Cherry JM (2013) The new modern era of yeast genomics: community sequencing and the resulting annotation of multiple *Saccharomyces cerevisiae* strains at the Saccharomyces genome database. Database (Oxford) 2013, bat012

Fekete V, Cierna M, Polakova S, Piskur J, Sulo P (2007) Transition of the ability to generate petites in the *Saccharomyces*/*Kluyveromyces* complex. FEMS Yeast Res 7:1237–1247

Ferea TL, Botstein D, Brown PO, Rosenzweig RF (1999) Systematic changes in gene expression patterns following adaptive evolution in yeast. Proc Natl Acad Sci U S A 96:9721–9726

Field Y, Fondufe-Mittendorf Y, Moore IK, Mieczkowski P, Kaplan N, Lubling Y et al (2009) Gene expression divergence in yeast is coupled to evolution of DNA-encoded nucleosome organization. Nat Genet 41:438–445

Field Y, Kaplan N, Fondufe-Mittendorf Y, Moore IK, Sharon E, Lubling Y et al (2008) Distinct modes of regulation by chromatin encoded through nucleosome positioning signals. PLoS Comput Biol 4:e1000216

Forsburg SL, Guarente L (1989) Communication between mitochondria and the nucleus in regulation of cytochrome genes in the yeast *Saccharomyces cerevisiae*. Annu Rev Cell Biol 5:153–180

Freire-Picos MA, Gonzalez-Siso MI, Rodriguez-Belmonte E, Rodriguez-Torres AM, Ramil E, Cerdan ME (1994) Codon usage in *Kluyveromyces lactis* and in yeast cytochrome c-encoding genes. Gene 139:43–49

Freire-Picos MA, Hollenberg CP, Breunig KD, Cerdan ME (1995) Regulation of cytochrome c expression in the aerobic respiratory yeast *Kluyveromyces lactis*. FEBS Lett 360:39–42

Gancedo C, Serrano R (1989) In energy-yielding metabolism, 2nd (edn). Academic Press, London, England, vol. 3

Gasch AP, Spellman PT, Kao CM, Carmel-Harel O, Eisen MB, Storz G et al (2000) Genomic expression programs in the response of yeast cells to environmental changes. Mol Biol Cell 11:4241–4257

Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H et al (1996) Life with 6000 genes. Science 274(546):63–67

Gross KC, Acosta PB (1991) Fruits and vegetables are a source of galactose: implications in planning the diets of patients with galactosaemia. J Inherit Metab Dis 14:253–258

Gu Z, David L, Petrov D, Jones T, Davis RW, Steinmetz LM (2005) Elevated evolutionary rates in the laboratory strain of *Saccharomyces cerevisiae*. Proc Natl Acad Sci U S A 102:1092–1097

Hagman A, Sall T, Compagno C, Piskur J (2013) Yeast "make-accumulate-consume" life strategy evolved as a multi-step process that predates the whole genome duplication. PLoS ONE 8:e68734

Hittinger CT, Carroll SB (2007) Gene duplication and the adaptive evolution of a classic genetic switch. Nature 449:677–681

Hittinger CT, Goncalves P, Sampaio JP, Dover J, Johnston M, Rokas A (2010) Remarkably ancient balanced polymorphisms in a multi-locus gene network. Nature 464:54–58

Hittinger CT, Rokas A, Carroll SB (2004) Parallel inactivation of multiple GAL pathway genes and ecological diversification in yeasts. Proc Natl Acad Sci U S A 101:14144–14149

Hoffman CS (2005) Glucose sensing via the protein kinase a pathway in *Schizosaccharomyces pombe*. Biochem Soc Trans 33:257–260

Hoffmann HP, Avers CJ (1973) Mitochondrion of yeast: ultrastructural evidence for one giant, branched organelle per cell. Sci 181(4101):749–751

Holland MJ, Holland JP (1978) Isolation and identification of yeast messenger ribonucleic acids coding for enolase, glyceraldehyde-3-phosphate dehydrogenase, and phosphoglycerate kinase. Biochemistry (Mosc) 17:4900–4907

Hurowitz EH, Brown PO (2003) Genome-wide analysis of mRNA lengths in *Saccharomyces cerevisiae*. Genome Biol 5:R2

Ihmels J, Bergmann S, Gerami-Nejad M, Yanai I, McClellan M, Berman J et al (2005) Rewiring of the yeast transcriptional network through the evolution of motif usage. Science 309:938–940

Ihmels J, Friedlander G, Bergmann S, Sarig O, Ziv Y, Barkai N (2002) Revealing modular organization in the yeast transcriptional network. Nat Genet 31:370–377

Ikemura T (1981) Correlation between the abundance of Escherichia coli transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the E. coli translational system. J Mol Biol 151:389–409

Ikemura T (1982) Correlation between the abundance of yeast transfer RNAs and the occurrence of the respective codons in protein genes. Differences in synonymous codon choice patterns of yeast and Escherichia coli with reference to the abundance of isoaccepting transfer RNAs. J Mol Biol 158:573–597

Jensen RE et al (2000) Yeast mitochondrial dynamics: fusion, division, segregation, and shape. Microsc Res Tech 51(6):573–583

Jiang C, Pugh BF (2009) Nucleosome positioning and gene regulation: advances through genomics. Nat Rev Genet 10:161–172

Jiang H, Guan W, Pinney D, Wang W, Gu Z (2008) Relaxation of yeast mitochondrial functions after whole-genome duplication. Genome Res 18:1466–1471

Johnston M (1987) A model fungal gene regulatory mechanism: the GAL genes of *Saccharomyces cerevisiae*. Microbiol Rev 51:458–476

Kellis M, Birren BW, Lander ES (2004) Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. Nature 428:617–624

Kozak M (1987) An analysis of 5′-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Res 15:8125–8148

Kurtzman CP, Robnett CJ (2003) Phylogenetic relationships among yeasts of the 'Saccharomyces complex' determined from multigene sequence analyses. FEMS Yeast Res 3:417–432

Lee W, Tillo D, Bray N, Morse RH, Davis RW, Hughes TR et al (2007) A high-resolution atlas of nucleosome occupancy in yeast. Nat Genet 39:1235–1244

Leskovac V, Trivic S, Pericin D (2002) The three zinc-containing alcohol dehydrogenases from baker's yeast, *Saccharomyces cerevisiae*. FEMS Yeast Res 2:481–494

Lin Z, Li WH (2012) Evolution of 5′ untranslated region length and gene expression reprogramming in yeasts. Mol Biol Evol 29:81–89

Lin Z, Li WH (2011a) The evolution of aerobic fermentation in *Schizosaccharomyces pombe* was associated with regulatory reprogramming but not nucleosome reorganization. Mol Biol Evol 28:1407–1413

Lin Z, Li WH (2011b) Expansion of hexose transporter genes was associated with the evolution of aerobic fermentation in yeasts. Mol Biol Evol 28:131–142

Lin Z, Wang TY, Tsai BS, Wu FT, Yu FJ, Tseng YJ et al (2013) Identifying cis-regulatory changes involved in the evolution of aerobic fermentation in yeasts. Genome Biol Evol 5(6):1065–1078

Lin Z, Wu WS, Liang H, Woo Y, Li WH (2010) The spatial distribution of cis regulatory elements in yeast promoters and its implications for transcriptional regulation. BMC Genom 11:581

Liti G, Louis EJ (2005) Yeast evolution and comparative genomics. Annu Rev Microbiol 59:135–153

Lohr D, Venkov P, Zlatanova J (1995) Transcriptional regulation in the yeast GAL gene family: a complex genetic network. FASEB J 9:777–787

Mager WH, Planta RJ (1991) Coordinate expression of ribosomal protein genes in yeast as a function of cellular growth rate. Mol Cell Biochem 104:181–187

Martchenko M, Levitin A, Hogues H, Nantel A, Whiteway M (2007) Transcriptional rewiring of fungal galactose-metabolism circuitry. Curr Biol 17:1007–1013

Merico A, Sulo P, Piskur J, Compagno C (2007) Fermentative lifestyle in yeasts belonging to the Saccharomyces complex. FEBS J 274:976–989

Mignone F, Gissi C, Liuni S, Pesole G (2002) Untranslated regions of mRNAs. Genome Biol 3, REVIEWS0004

Miura F, Kawaguchi N, Sese J, Toyoda A, Hattori M, Morishita S et al (2006) A large-scale full-length cDNA analysis to explore the budding yeast transcriptome. Proc Natl Acad Sci U S A 103:17846–17851

Moller K, Christensen B, Forster J, Piskur J, Nielsen J, Olsson L (2002) Aerobic glucose metabolism of *Saccharomyces kluyveri*: growth, metabolite production, and quantification of metabolic fluxes. Biotechnol Bioeng 77:186–193

Moller K, Langkjaer RB, Nielsen J, Piskur J, Olsson L (2004) Pyruvate decarboxylases from the petite-negative yeast *Saccharomyces kluyveri*. Mol Genet Genomics 270:558–568

Mulder W, Scholten IH, Grivell LA (1995) Distinct transcriptional regulation of a gene coding for a mitochondrial protein in the yeasts *Saccharomyces cerevisiae* and *Kluyveromyces lactis* despite similar promoter structures. Mol Microbiol 17:813–824

Mulder W, Winkler AA, Scholten IH, Zonneveld BJ, de Winde JH, Yde Steensma H et al (1994) Centromere promoter factors (CPF1) of the yeasts *Saccharomyces cerevisiae* and *Kluyveromyces lactis* are functionally exchangeable, despite low overall homology. Curr Genet 26(3):198–207

Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M et al (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing. Science 320:1344–1349

Naumov GI, James SA, Naumova ES, Louis EJ, Roberts IN (2000) Three new species in the *Saccharomyces sensu stricto* complex: *Saccharomyces cariocanus*, *Saccharomyces kudriavzevii* and *Saccharomyces mikatae*. Int J Syst Evol Microbiol 50(Pt 5):1931–1942

Otterstedt K, Larsson C, Bill RM, Stahlberg A, Boles E, Hohmann S et al (2004) Switching the mode of metabolism in the yeast *Saccharomyces cerevisiae*. EMBO Rep 5:532–537

Ozcan S, Johnston M (1999) Function and regulation of yeast hexose transporters. Microbiol Mol Biol Rev 63:554–569

Papp B, Pal C, Hurst LD (2004) Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast. Nature 429:661–664

Pesole G, Mignone F, Gissi C, Grillo G, Licciulli F, Liuni S (2001) Structural and functional features of eukaryotic mRNA untranslated regions. Gene 276:73–81

Piskur J, Rozpedowska E, Polakova S, Merico A, Compagno C (2006) How did Saccharomyces evolve to become a good brewer? Trends Genet 22:183–186

Pritchard L, Kell DB (2002) Schemes of flux control in a model of *Saccharomyces cerevisiae* glycolysis. Eur J Biochem 269:3894–3904

Pronk JT, Yde Steensma H, Van Dijken JP (1996) Pyruvate metabolism in *Saccharomyces cerevisiae*. Yeast 12:1607–1633

Reifenberger E, Freidel K, Ciriacy M (1995) Identification of novel HXT genes in *Saccharomyces cerevisiae* reveals the impact of individual hexose transporters on glycolytic flux. Mol Microbiol 16:157–167

Rozpedowska E, Hellborg L, Ishchuk OP, Orhan F, Galafassi S, Merico A et al (2011) Parallel evolution of the make-accumulate-consume strategy in Saccharomyces and Dekkera yeasts. Nat Commun 2:302

Scannell DR, Byrne KP, Gordon JL, Wong S, Wolfe KH (2006) Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. Nature 440:341–345

Scannell DR, Zill OA, Rokas A, Payen C, Dunham MJ, Eisen MB et al. (2011) The awesome power of yeast evolutionary genetics: new genome sequences and strain resources for the *Saccharomyces sensu stricto* genus. G3 (Bethesda) 1, 11–25

Schmitt HD, Ciriacy M, Zimmermann FK (1983) The synthesis of yeast pyruvate decarboxylase is regulated by large variations in the messenger RNA level. Mol Gen Genet 192:247–252

Sellick CA, Campbell RN, Reece RJ (2008) Galactose metabolism in yeast-structure and regulation of the leloir pathway enzymes and the genes encoding them. Int Rev Cell Mol Biol 269:111–150

Slot JC, Rokas A (2010) Multiple GAL pathway gene clusters evolved independently and by different mechanisms in fungi. Proc Natl Acad Sci U S A 107:10136–10141

Tavazoie S, Church GM (1998) Quantitative whole-genome analysis of DNA-protein interactions by in vivo methylase protection in E. coli. Nat Biotechnol 16:566–571

Thomson JM, Gaucher EA, Burgan MF, De Kee DW, Li T, Aris JP et al (2005) Resurrecting ancestral alcohol dehydrogenases from yeast. Nat Genet 37:630–635

Tirosh I, Barkai N (2008) Two strategies for gene regulation by promoter nucleosomes. Genome Res 18:1084–1091

Tirosh I, Sigal N, Barkai N (2010) Divergence of nucleosome positioning between two closely related yeast species: genetic basis and functional consequences. Mol Syst Biol 6:365

Tsankov AM, Thompson DA, Socha A, Regev A, Rando OJ (2010) The role of nucleosome positioning in the evolution of gene regulation. PLoS Biol 8:e1000414

van der Sluis C, Wolken WA, Giuseppin ML, Tramper J, Wijffels RH (2000) Effect of threonine, cystathionine, and the branched-chain amino acids on the metabolism of *Zygosaccharomyces rouxii**. Enzyme Microb Technol 26:292–300

Wolfe K (2004) Evolutionary genomics: yeasts accelerate beyond BLAST. Curr Biol 14: R392–R394

Wolfe KH, Shields DC (1997) Molecular evidence for an ancient duplication of the entire yeast genome. Nature 387:708–713

Wood V, Gwilliam R, Rajandream MA, Lyne M, Lyne R, Stewart A et al (2002) The genome sequence of *Schizosaccharomyces pombe*. Nature 415:871–880

Ye L, Kruckeberg AL, Berden JA, van Dam K (1999) Growth and glucose repression are controlled by glucose transport in *Saccharomyces cerevisiae* cells containing only one glucose transporter. J Bacteriol 181:4673–4675

Yuan GC, Liu YJ, Dion MF, Slack MD, Wu LF, Altschuler SJ et al (2005) Genome-scale identification of nucleosome positions in S. cerevisiae. Science 309:626–630