FEMS Yeast Research, 17, 2017, fox057

doi: 10.1093/femsyr/fox057
Advance Access Publication Date: 3 August 2017
Research Article

RESEARCH ARTICLE

Genomic and transcriptomic analysis of *Saccharomyces cerevisiae* isolates with focus in succinic acid production

Ricardo Franco-Duarte[1,*], Daniela Bessa[1], Filipa Gonçalves[1], Rosa Martins[2], António César Silva-Ferreira[2], Dorit Schuller[1], Paula Sampaio[1,†] and Célia Pais[1]

[1]CBMA (Centre of Molecular and Environmental Biology) / Department of Biology / University of Minho, 4710-057 Braga, Portugal and [2]Escola Superior de Biotecnologia, Universidade Católica Portuguesa, 4200-072 Porto, Portugal

*Corresponding author: CBMA (Centre of Molecular and Environmental Biology) / Department of Biology / University of Minho, Campus de Gualtar 4710-057 Braga, Portugal. Tel: (+351) 253 601,522; Fax: (+351) 253,678 980; E-mail: ricardofilipeduarte@bio.uminho.pt
One sentence summary: This is one of the first studies showing the capacity of natural yeast isolates to produce relevant amounts of succinic acid, giving clues for further strain improvement.
Editor: Cletus Kurtzman
†Paula Sampaio, http://orcid.org/0000-0002-1415-4428

ABSTRACT

Succinic acid is a platform chemical that plays an important role as precursor for the synthesis of many valuable bio-based chemicals. Its microbial production from renewable resources has seen great developments, specially exploring the use of yeasts to overcome the limitations of using bacteria. The objective of the present work was to screen for succinate-producing isolates, using a yeast collection with different origins and characteristics. Four strains were chosen, two as promising succinic acid producers, in comparison with two low producers. Genome of these isolates was analysed, and differences were found mainly in genes *SDH1*, *SDH3*, *MDH1* and the transcription factor *HAP4*, regarding the number of single nucleotide polymorphisms and the gene copy-number profile. Real-time PCR was used to study gene expression of 10 selected genes involved in the metabolic pathway of succinic acid production. Results show that for the non-producing strain, higher expression of genes *SDH1, SDH2, ADH1, ADH3, IDH1* and *HAP4* was detected, together with lower expression of *ADR1* transcription factor, in comparison with the best producer strain. This is the first study showing the capacity of natural yeast isolates to produce high amounts of succinic acid, together with the understanding of the key factors associated, giving clues for strain improvement.

Keywords: succinic acid; *Saccharomyces cerevisiae*; natural isolates; genomics; transcriptomics

INTRODUCTION

Several platform chemicals that can be produced from carbohydrates and converted to high-value bio-based chemicals or materials using biorefinery approaches have been previously identified. Succinic acid (SA) is one of these 12 building blocks that can be produced from sugars via biological of chemical conversions,

Received: 21 March 2017; **Accepted:** 28 July 2017

and plays an important role as a precursor for the synthesis of biodegradable polyesters, resins, dyestuff, pharmaceuticals and as food industry additive. Although bacterial synthesis of succinate from glucose can reach impressive final yields of 1 mol/mol, their application as production hosts is difficult because some of them are obligate anaerobes, potential pathogens or do not tolerate low pH values, osmotic stress and high glucose levels (Song and Lee 2006). However, the main bottleneck of SA production from renewable resources using bacterial species is the cost of purification since the final fermentation broth must be acidified to obtain the free acid (Beauprez, De Mey and Soetaert 2010). For this reason, some SA-producing companies started, already, using yeasts for SA production, which allows the possibility of production under lower pH values, and recurring mainly to metabolic engineering strategies to enhance its production.

The yeast *Saccharomyces cerevisiae* stands out as a platform microorganism for dicarboxylic acid production. Due to its role as a model organism in research, their physiology and genetics is extensively documented and a well-developed metabolic engineering toolbox is available. In addition, a huge number of genetic mutant strains are already available, such as the Euroscarf collection with around 5100 mutant strains, covering 82% of the ~6200 annotated yeast ORFs. Moreover, wild-type *S. cerevisiae* strains have GRAS (Generally Regarded As Safe) status, so that modified strains are more likely to be allowed in the production of these C4-dicarboxylic acids for food and pharmaceutical applications. It is known since long time ago the capacity of *S. cerevisiae* to achieve high concentrations of SA used in the manufacture of wine to enhance its quality (Wakai, Shimazaki and Hara 1980). This yeast species grows well under acidic conditions, and is thus an attractive alternative as a biocatalyst for SA production.

Although *S. cerevisiae* wild-type strains are able to produce SA, the optimisation of strains via metabolic engineering for a more efficient production is currently being gathering a lot of attention, as mentioned earlier. The production of organic acids begins in yeast via glycolysis, differing then downstream of pyruvate formation. From pyruvate, three pathways are possible for the succinate formation: via oxidative tricarboxylic acid (TCA) cycle, via reductive branch of TCA cycle or via oxidative pathway of the glyoxylate cycle. The reductive pathway has a theoretical yield of two molecules of succinate for every glucose molecule, due to the fixation of carbon dioxide, and is energetically unfavourable. The oxidative pathways have half the yield due to carbon dioxide production. One efficient alternative to produce succinate is the combined use of both the reductive TCA cycle and glyoxylate shunt by joining the oxidative and reductive route (Raab and Lang 2011), providing in this way an even redox balance, a higher maximum succinate yield and a fixation of CO_2 instead of its release. However, its implementation will be a challenge in the following years, since it requires that oxidative and fermentative metabolism run simultaneously in yeasts. This strategy was already successfully accomplished in *Escherichia coli* strains, with an increase of the theoretical yield (Vemuri, Eiteman and Altman 2002a,b; Rezaei *et al.* 2015).

Different strategies have been used to manipulate yeast strains to an enhanced production of SA, as reviewed in Cheng *et al.* (2013). The challenge in metabolic engineering *of S. cerevisiae* for the efficient production of SA involves at least three levels: (i) elimination of alcoholic fermentation, which occurs irrespective of the availability of oxygen; (ii) engineering fast and efficient metabolic pathways that link the high-capacity glycolytic pathway to the TCA cycle, taking into account NADH and ATP constraints, and (iii) engineering of product export. One of the first attempts to improve succinate production was reported by Arikawa *et al.* (1999) using sake yeast strains with TCA cycle gene deletions. The simultaneous deletion of *SDH1* and *FUM1* led to a 2.7-fold higher production of succinate in comparison with wild-type strains. However, these enhancements were not observed in the anaerobic conditions of sake fermentation. Another metabolic engineering strategy for the oxidative production of SA was the quadruple gene deletion (*SDH1*, *SDH2*, *IDH1*, *IDP1*) described by Raab *et al.* (2010). In glucose-grown shake-flask cultures, the mutant strain produced a titer of 3.62 g/L. Multigene deletion followed by directed evolution was used by Otero *et al.* (2013) to select a succinate producer mutant. The strategy included deletion of *SDH3* and interruption of glycolysis-derived serine by deletion of 3-phosphoglycerate dehydrogenase (*SER3/SER33*) and overexpression of native *ICL1*. The mutant strain presented a 43-fold increase in succinate yield on biomass compared to the reference strain. A modified pathway for succinate production was established by the deletion of genes *FUM1*, *PDC* and *GPD1* and reported by Yan *et al.* (2014). The authors obtained a succinate titer of 8.09 g/L, which was further improved to 9.98 g/L through regulation of biotin and urea levels, and to 12.97 g/L through optimal CO_2 conditions in a bioreactor. Another attempt to engineer *S. cerevisiae* strains with increased succinate production involved the disruption of *SDH1* and *SDH2* genes together with the expression of *Schizosaccharomyces pombe* malic acid transporter *MAE1*. The expression of the transporter to export intracellular acid outside of the cell successfully improved SA production (Ito, Hirasawa and Shimizu 2014).

The objective of the present work was to identify natural *S. cerevisiae* isolates with improved capacity to produce SA, and their characterisation, using genomic and transcriptomic approaches. The final goal was to understand the key factors that may contribute to the ability to produce SA in natural isolates, in order to design new strategies for genetic manipulations leading to an enhanced production.

MATERIALS AND METHODS

Saccharomyces cerevisiae strains

Seventeen *S. cerevisiae* isolates from different geographical origins and applications were selected from a previously established yeast collection (Mendes *et al.* 2013) and used in this study. All strains were stored at –80°C in cryotubes containing 1 mL glycerol (30% v/v).

Phenotypic screening was performed using two sets of tests: the first consisting of growth tests performed using liquid medium in 96-well microplates and the second one using agar plates. Detailed experimental conditions can be found in Mendes *et al.* (2013). Strains have also been genetically characterised regarding allelic combinations for described microsatellites ScAAT1, ScAAT2, ScAAT3, ScAAT4, ScAAT5, ScAAT6, YPL009, ScYOR267c, C4, C5 and C11 (Field and Wills 1998; Pérez *et al.* 2001; Legras *et al.* 2005; Franco-Duarte *et al.* 2014). DNA isolation was performed as previously described (Schuller *et al.* 2004) and used for microsatellite analysis. Multiplex PCR mixtures and cycling conditions were optimised and performed in 96-well PCR plates as described in Franco-Duarte *et al.* (2009).

The first set of individual fermentations of each strain, in anaerobic conditions, was carried out at 18°C using white grape must (variety Loureiro) in Erlenmeyer flasks (100 mL) with rubber stoppers perforated with a syringe needle to allow CO_2 release.

The used must had the following composition (w/v), determined by high-performance liquid chromatography (HPLC): glucose—84.05 g/L, fructose—54.36 g/L, tartaric acid—1.22 g/L, glycerol—0.19 g/L, acetic acid—0.05 g/L, ethanol—0.14 g/L, total acidity—6.20 g/L, pH—3.56. When glucose concentration was below 5 g/L and no weight variations were noted, samples were collected and frozen (–20°C) for metabolomic analysis. HPLC with refractive index was used to quantify ethanol, glycerol and organic acids (tartaric, malic, acetic and succinic) in an EX Chrome Elite HPLC using a Rezex® Ion Exclusion column. Column and refractive index detector temperatures were 60°C and 40°C, respectively, and the flow rate was 0.50 mL min^{-1} for 0–9 min, 0.25 mL min^{-1} for 10–14 min and 0.50 mL min^{-1} for 15–35 min. Relevant metabolites known to account for interstrain differences and that are related to volatile compounds (higher alcohols, esters, fatty acids) were determined by gas chromatography—mass spectrometry (GC-MS). Analyses were performed by solid-phase microextraction, using a divinylbenzene/carboxen/polydimethylsiloxane (DVB/CAR/PDMS) 50/30 μm (Supelco, Sigma) fibre for 15 min under continuous agitation and heating at 40°C. 3-Octanol (Sigma-Aldrich, 99% purity, St. Louis, Missouri, EUA) was used as internal standard. Compounds were then desorbed from the solid-phase microextraction fibre directly and analysed using a Varian CP-3800 gas chromatograph (Walnut Creek, CA, USA), equipped with a Varian Saturn 2000 mass selective detector, as previously described (Silva Ferreira *et al.* 2004).

Individual fermentations in aerobic conditions and HPLC analysis

New individual fermentations were accomplished with each of the 17 strains, at 30°C (200 rpm) using 60 mL of MS medium (Bely, Sablayrolles and Barre 1990), but using sucrose instead of glucose, as carbon source, in Erlenmeyer flasks (250 mL), with rubber stoppers that were perforated with a syringe needle for CO_2 release, during 96 h. The fermentative profile of each strain was monitored by absorbance measurement and weight loss determination of the flasks due to CO_2 liberation. HPLC analysis was performed using the same parameters referred in the previous subsection.

Genome sequencing and analysis

To investigate the extent of strain variation and the genomic differences underlying specific traits, whole genome of strains was sequenced by Illumina next-generation sequencing, according to the manufacture's protocols (Illumina 2009), in paired-end 104 bp mode, using an Illumina HiSeq2000 analyser. All demultiplexed reads were aligned to the sacCer3 assembly of the yeast reference genome (S288c) using BWA (bwa-aln and bwa-sampe; version 0.7.5a) with default parameters. Sequences were aligned using SAMtools (version 1.1) using the commands view, sort, index and mpileup (Li and Durbin 2009). All possible variants including frameshift insertions/deletions (Indels) and single nucleotide polymorphisms (SNPs) were then called from the aligned sequences, using Annovar (Wang, Li and Hakonarson 2010), with the following filters: QUAL $\geq$ 30 ('phred-scaled quality score for the assertion made in the alternate allele'), DP $\geq$ 15 ('raw read depth' or 'coverage'), MQ $\geq$ 40 ('root-mean-squared mapping quality of coverage reads') and GQ $\geq$ 50 ('genotype quality or phred-scaled confidence that the true genotype is the one provided').

RNA isolation

Saccharomyces cerevisiae yeast cells were inoculated in 10 mL of YPD and grown overnight at 30°C. Each overnight culture was used to inoculate 20 mL of YPD to an initial OD_{600} of 0.1, in triplicate and incubated at 30°C (200 rpm) during 96 h. Cells were harvested at two different time points (24 and 72 h) and immediately stored at –80°C. RNA extraction was performed using the RiboPure Yeast kit (Ambion) according to the manufacturer's instructions.

Real-time PCR

Transcript levels of some selected genes (*ADH1, ADH2, ADH3, SDH1, SDH2, IDH1, KGD1, ADR1, HAP4* and *GCR1*) were determined using a quantitative real-time PCR (rt-PCR) approach. Oligonucleotides used are listed in Supplementary data S1 (Supporting Information). Total RNA was incubated with RNAse-free DNAase I (Invitrogen) for 15 min at room temperature to eliminate genomic DNA contamination. DNase was inactivated according to the manufacturer's instructions. The Superscript III Platinium two-step rtPCR with SYBR green (Invitrogen) was used to generate first-strand cDNA from each DNase I-treated RNA sample as follows: 10 min at 25°C, 50 min at 42°C and 5 min at 85°C and then incubated 20 min at 37°C for RNase H treatment. Two microliters of each cDNA sample was added to a 20 mL PCR mixture containing 10 mL of Platinum SYBR green qPCR. SuperMix-UDG, 0.4 mL of 10 mM specific forward and reverse primers and 7.2 mL of RNase-free water were used. Each reaction was performed with an rtPCR detection system (BIO-RAD). Thermocycling conditions for rtPCR were 2 min at 50°C (UDG incubation) and 2 min at 95°C, followed by 45 cycles of 95°C for 15 s, 60°C for 30 s and 72°C for 30 s. The specificity of each primer pair was verified by the presence of a single melting temperature peak. The efficiency of the primers was assessed in titration experiments using cDNA in serial dilutions. Gene expression was normalised to the housekeeping gene *ACT1* and analysed by using the comparative threshold cycle (DDCT) method. Data were presented as the fold difference in gene expression in one strain relative to the expression in another.

RESULTS

Saccharomyces cerevisiae strain characterisation

From a previously established strain collection (Mendes *et al.* 2013), a group of 17 most heterogeneous *S. cerevisiae* strains was selected. Figure 1 illustrates the established strain collection, for which a deep phenotypic, genetic and metabolic characterisation was performed. In Fig. 1A–D, results of the principal component analysis (PCA) loadings (distribution of variables) are presented, being the PCA scores omitted from the visualisation. In panel A, results obtained from HPLC analysis are shown, being accomplished with samples obtained at the end of fermentation, to evaluate the chemical compounds and conclude about the metabolic profiles of the 17 strains, especially regarding their capacity to produce SA. Strain-dependent differences could be observed concerning organic acids (malic, succinic and acetic), glycerol and ethanol, being the strain variability mainly influenced by succinic and acetic acid concentrations. Results allowed the separation of the strains as good and poor SA producers as influenced by the second principal component (PC-2). An interesting anticorrelation was observed between succinic and acetic acid production (arrow in panel A),

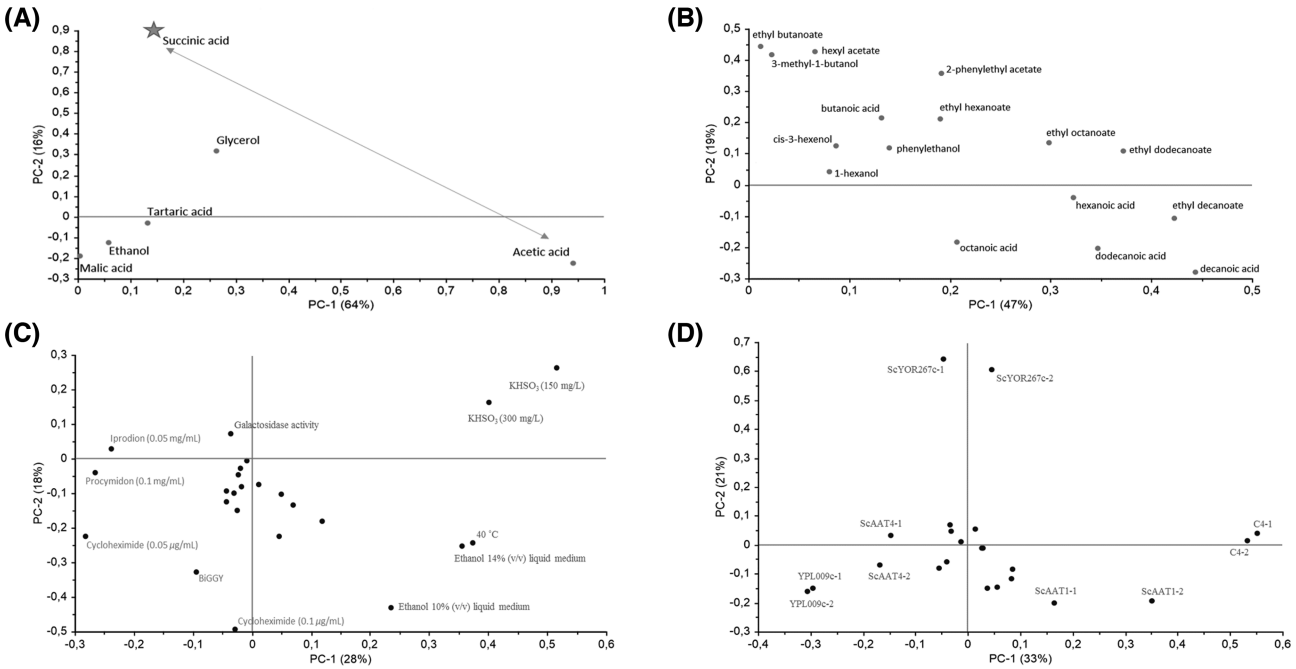

Figure 1. Principal component analysis results for 17 *S. cerevisiae* strains, regarding loadings distribution. Scores were omitted for each panel. (**A**) Concentration of six metabolites quantified by HPLC; (**B**) concentration of 16 volatile compounds determined by GC-MS; (**C**) 30 tests used in phenotypic characterisation; (**D**) 11 microsatellites alleles used in genetic characterisation (1 and 2 refers to the presence of the microsatellite in both copies of the chromosome).

once that for the best SA producer strains, the amounts of acetic acid produced were very low. This fact is very important since it suggests an antagonism in the production pathways of these two acids, as desirable, which is not always obtained when using bacteria for the purpose. Regarding GC-MS, phenotypic and genetic characterisation (panels B, C and D, respectively), a high diversity was obtained, showing the heterogeneity of the chosen collection, with some interesting features contributing to explain strain diversity, as for example: (i) separation of esters and acids positioning in the PCA of panel B; (ii) higher contribution for strain separation of ethanol resistance and capacity to grow in the presence of potassium bisulphite and at 40°C, as shown in panel C; (iii) finding of higher alleles in some strains for microsatellites ScYOR267c and C4.

Bioanalytical analysis

Individual fermentations were again performed with the 17 strains in order to identify promising *S. cerevisiae* isolates with capacity for an improved production of SA, using MS medium with sucrose as carbon source. Samples from the endpoint of fermentation were collected, and HPLC results were used to identify the best and worst SA producers, together with other compounds of interest (ethanol, glucose, fructose, glycerol, malic and acetic acids). SA production (Fig. 2A) varied between 0 g/L (strain Z86; reference S288c strain, a laboratorial non-fermentative strain) and 1.13 g/L (strain Z28). This heterogeneity was also observed regarding the other analysed metabolites, namely glycerol (panel B), ethanol (panel C), glucose and fructose (panel D), and malic and acetic acids (panel E).

From these results, four strains were chosen for further studies, being these four strains marked in Fig. 2A by coloured circles. We have chosen the two higher SA producers (strain Z28—1.13 g/L and strain Z56—0.98 g/L), the lowest producer (strain Z63—0.02 g/L) and a strain with a low production but showing some capacity to produce this acid (strain Z12—0.50 g/L). Other secondary criteria were also considered in the choice of these four strains, in particular, the fact that strain Z12 showed interesting amounts of the remaining tested compounds (panels B to D), strain Z63 was also the lowest producer of almost all other tested metabolites (panels B to E), and strains Z28 and Z56 showed an heterogeneous behaviour regarding the production of the remaining compounds, not always consistent with being the higher SA producers.

Genomic analysis with focus in SA metabolic pathways

To investigate the extent of genomic variation among the strains, we sequenced the whole genome from the 17 strains using next-generation sequencing techniques by Illumina sequencing. Quantification of SNPs and insertions/deletions (Indels) was performed by comparison of each strain with the reference (strain S288c). Exclusive SNPs were identified for each strain, having each one a unique genomic pattern. A set of 42 genes (39 genes plus three transcription factors) known as important and relevant in the SA production pathway was chosen to be further analysed (Supplementary data S2). These genes were studied in detail in the genome of the four chosen strains, in terms of SNPs, Indels and gene copy number. Figure 3 summarises the main findings obtained when comparing the genomes of the two best producers strains (Z28 and Z56) with the two strains with the lower capacity to produce SA (Z12 and Z63). No relevant changes were detected when looking to the total amount of SNPs and Indels between both groups (Fig. 3A). Also, when looking at panel B, in which the number of SNPs per 1 kb in the good SA-producing strains in comparison with the ones producing reduced amounts of this acid is presented, no particular differences attributed to the SA production are observed. However, the detailed distribution of the SNPs along the 16 chromosomes (panel C) allowed the understanding of some

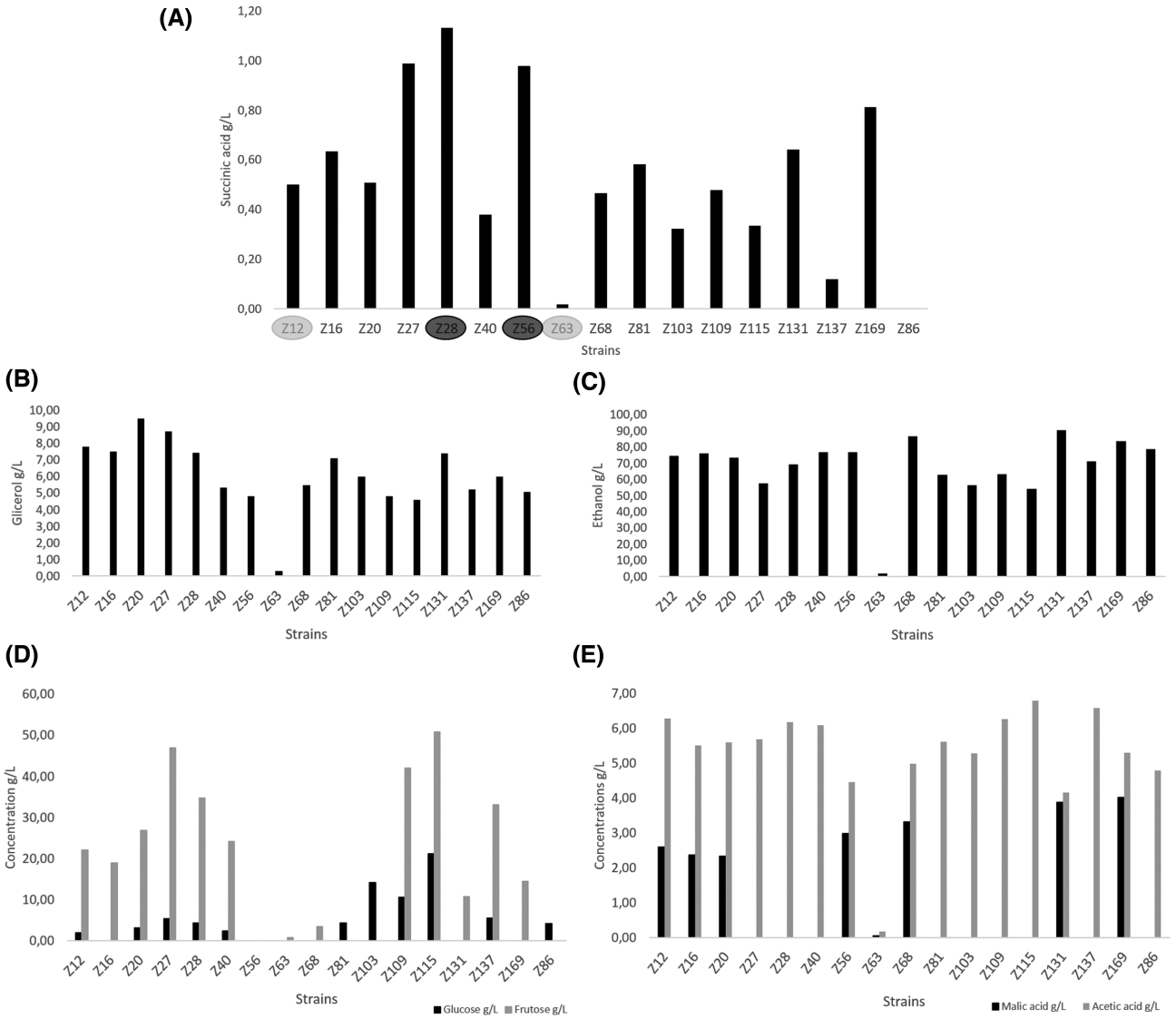

Figure 2. HPLC analysis results obtained with 17 *S. cerevisiae* strains, regarding succinic acid (**A**), glycerol (**B**), ethanol (**C**), fructose and glucose (**D**) and malic and acetic acids (**E**). Black and grey circles in panel A refer to good and poor succinic acid producer strains, respectively, chosen to be further characterised.

particular differences between the two groups of strains. As can be seen in this panel, strains with poor SA production capacity show, in general, lower number of SNPs (green), although in some chromosomes the good producers (marked in blue) have higher amounts of SNPs, mainly in the last part of chromosomes XIII, XIV and XV, and in the first third of chromosome III. In order to understand if these differences could be extensible to a higher or lower copy number of some genes, the full genome of these strains was analysed in terms of gene copy number in all chromosomes (Fig. 4). Results showed that strains with poor capacity to produce SA have higher number of copies in the entire chromosome XI, and also one of them (Z12) in almost the entire chromosome I. On the contrary, strains showing higher SA production have lower number of copies in these chromosomes, being this value equal to the reference strain in the case of strain Z28 or even lower than the reference strain in the case of strain Z56. The full genome of these isolates was scanned, in terms of SNPs and gene copy numbers, and the main differences found in key genes related with SA production pathway, previously chosen, are summarised in Table 1. Indels were not included in the table since no significant discriminatory results were found. From the 42 analysed genes, 33 showed at least one SNP difference in the four tested strains. Only SNPs leading to a translation different from the original were considered and included in this table, and the majority of the changes (underlined SNPs) also led to an amino acid from a different functional category.

Transcriptomic analysis

From the comparative genomic analysis, a set of 10 genes was selected to be further studied by transcriptomic analysis, along the fermentation, in the bad and good SA producer isolates, to help to further understand the mechanisms underlying the differences between these strains and give clues for their future improvement. To perform transcriptomic analysis, new fermentations were accomplished with these four strains, in triplicate, collecting samples from two different time points (Fig. 5) to be used for RNA extraction. Although the fermentative conditions were the same as previous, this time the fermentations were

(A)

	Total number of InDels	Total number of SNP	Total number of unique SNP
Higher succinic acid producers	527	29712	7699
Lower succinic acid producers	479	30274	8927

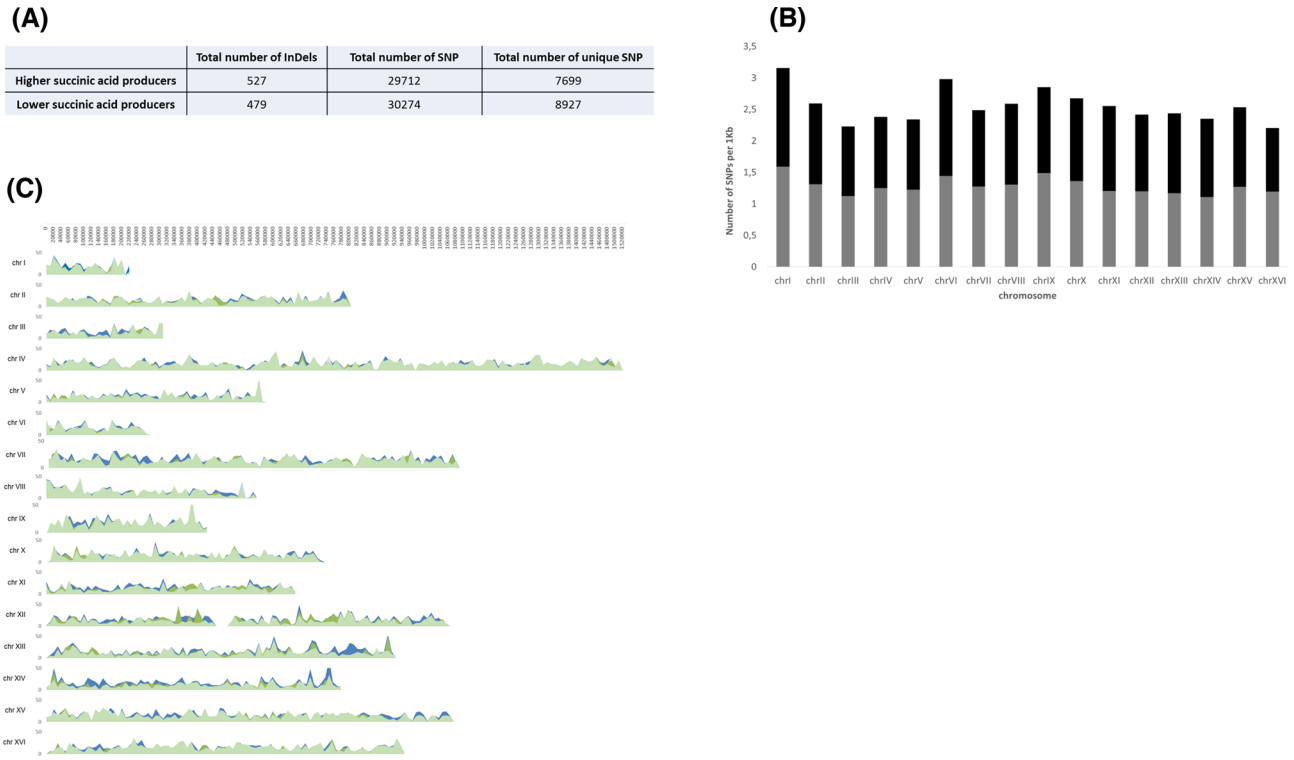

Figure 3. Comparative genomic results obtained after whole-genome sequencing of the four *S. cerevisiae* isolates. (**A**) Total number of SNPs, Indels and unique SNPs in both good and poor succinic acid producers; (**B**) number of SNPs per 1 kb in the good and poor succinic acid producers; (**C**) comparative analysis of SNPs distribution in the four strains. For each chromosome, the total amount of SNPs per 1 kb is shown. Strains chosen as poor succinic acid producers are marked in green and the ones suggested as good producers are marked in blue.

accomplished only until 72 h in order to preserve RNA integrity for transcriptomic study. This explains why the final amounts of SA obtained were lower than the ones obtained in the previous analysis (Fig. 2), which was calculated from samples obtained at 96 h of growth. No differences in the growth curves of the four strains (Fig. 5A) were visible, although the SA production was clearly different between them (Fig. 5B). Strains chosen as having a higher capacity to produce SA (Z28 and Z56) showed the highest accumulation at 72 h time point (T3), with a maximum of 0.4 g/L obtained. The remaining two strains with lower capacity to produce SA (Z12 and Z63) obtained a maximum of 0.2 g/L of SA. If the experiment was extended further, certainly higher amounts of SA accumulation would be obtained in the good producing strains, although it was not possible to obtain good quality RNA from these time points. However, the objective of this task was to compare the differences between the higher and lower producers of SA, being these differences clearly observed at the determined time points, especially at 72 h.

Gene expression analysis was performed in order to compare the expression of genes when growing amounts of SA are considered as being produced by the strains (raw data—Supplementary data S3 and S4; schematic summary of results—Fig. 6). Schematic representation in Fig. 6 indicates (indicative of over or underexpression, respectively) comparative expression of the mentioned genes by using red and green colours, at 72 h. Due to being similar, results of the 24 h were omitted from the scheme. In particular, expression of 10 genes (including three transcription factors) was considered (*HAP4*, *ADR1*, *GCR1*, *ADH1*, *ADH2*, *ADH3*, *IDH1*, *KGD1*, *SDH1* and *SDH2*), comparing the non-producer strain (Z63) with the intermediate producer (Z12) in order to understand which one triggers the beginning of SA production (Fig. 6A), and then comparing the best SA producer strain with the intermediate one, to evaluate what enhances its production (Fig. 6B). Results showed that when SA production is higher (panel B versus panel A), a lower expression of *SDH* genes is detected, as expected, together with a lower expression, in general, of *ADH* genes. No relevant changes were detected regarding the expression of *KGD1* and *GCR1*. When considering the expression of transcription factors *HAP4* and *ADR1*, interesting results were observed. *HAP4* was higher expressed in panel A, being highly expressed in the non-producer strain, which is in accordance with the previous shown increase in number of copies of this gene. On the contrary, *ADR1* showed higher expression in the best producer strain, when comparing with the intermediate one, suggesting its involvement in the enhancement of SA production. This gene is involved in the aerobic oxidation of carbon sources after the diauxic transition, being in accordance with the higher SA concentrations produced by strain Z28.

DISCUSSION

Along the last few years, several publications discussed the advantages of microbial production of SA over the chemical method (Zeikus, Jain and Elankovan 1999; Carole, Pellegrino and Paster 2004; McKinlay, Vieille and Zeikus 2007; Bechthold *et al.* 2008; Beauprez, De Mey and Soetaert 2010; Bozell and Petersen 2010). Although the majority of processes were developed using bacteria, yeasts, in particular *Saccharomyces cerevisiae*, have drawn the attention due to the possibility of production under lower pH values, with less by-product formation, using simple media and with an easier downstream processing needed.

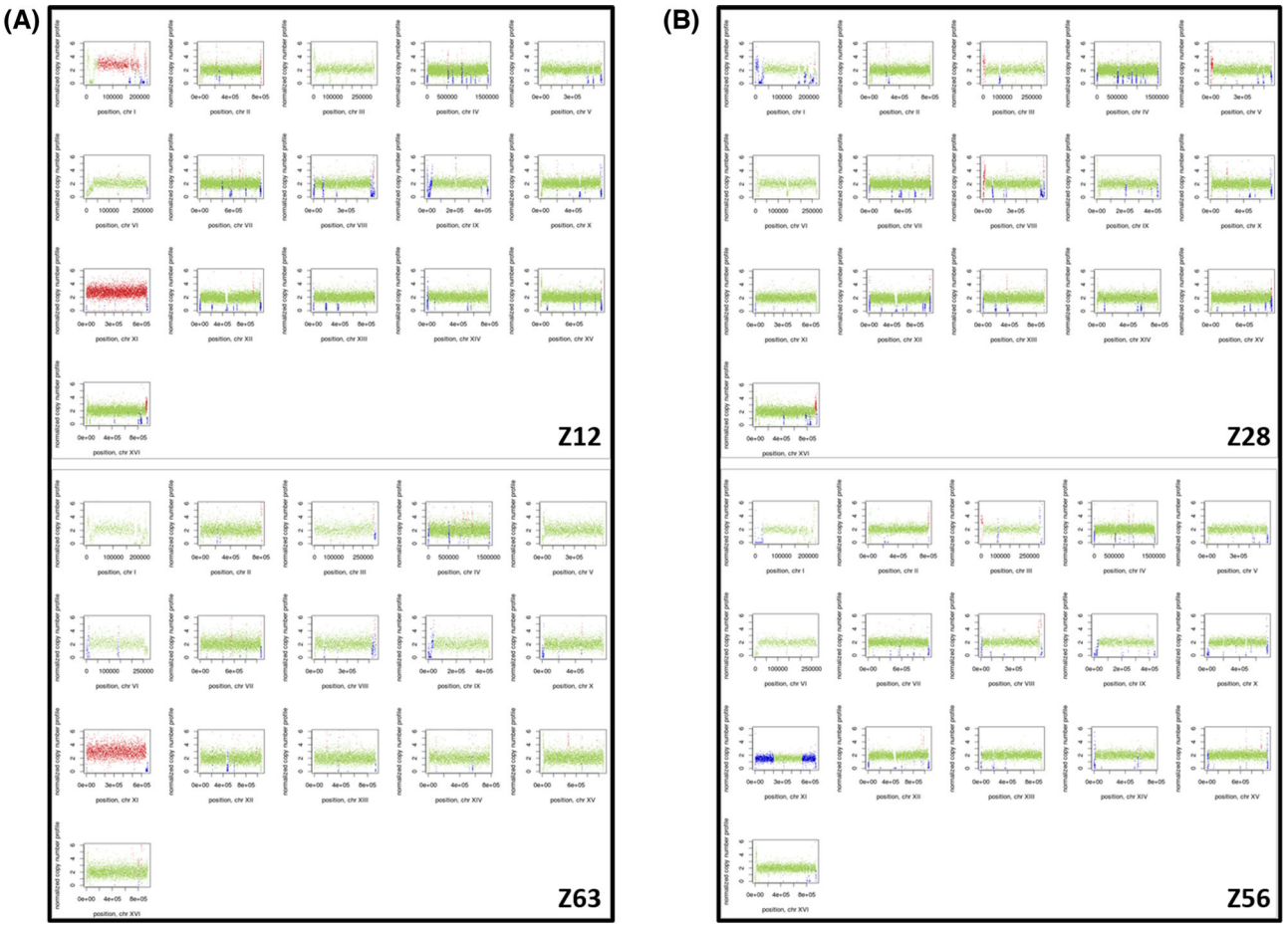

Figure 4. Copy number profiles of each chromosome of the two strains chosen as poor producers of succinic acid (**A**: Z12, Z63), and the two strains chosen as the best producers (**B**: Z28, Z56) in comparison with the reference genome. Green—no copy number changes; blue—decrease in the copy number profile; red—increase in the copy number profile.

Although there is no native predisposition for production and accumulation of SA in *S. cerevisiae*, in the present work natural yeast isolates were tested and showed a good capacity to produce SA. By exploring natural diversity, without recurring to genetic manipulation for an enhanced production, food safety concerns and population acceptance can be overcome.

A group of 17 *S. cerevisiae* isolates was screened regarding phenotypic, genetic and metabolic data (Fig. 1). Results showed a high strain diversity, pointing to the importance of the initial constitution of an heterogeneous collection. Regarding SA production, one of the advantages of using yeasts for SA production, that is, the lower amounts of acetic acid obtained as a by-product, was demonstrated. As can be seen in Fig. 1, a very interesting anti-correlation was observed between succinic and acetic acids production, once that for the best SA producer strains, the amounts of acetic acid produced were very low. This fact is very important since it suggests an antagonism in the production pathways of these two acids, as desirable, which is not always obtained when using bacteria for the purpose. Using these 17 isolates as the core collection, new individual fermentations were performed. HPLC results were used to identify the best and worst producers of SA, together with other compounds of interest (ethanol, glucose, fructose, glycerol, malic and acetic acids). Under the tested conditions, the maximum concentration of SA obtained was 1.13 g/L. Although not so high as in some of the studies using engineered yeasts, it is, until the extent of our knowledge, the first study showing the capacity of natural isolates to produce this amount of SA. Strain Z56 revealed itself as a good fermentative isolate, completely consuming all carbon present in the medium, and leading to a high production of succinic, acetic and malic acids, suggesting a metabolism directed for the production of acids via TCA cycle. When comparing with results obtained by Otero *et al.* (2013), obtaining 0.9 g/L of SA using gene deletions (*SDH*, *SER3/SER33*, overexpression of native *ICL1*), or with the ones obtained by Agren, Otero and Nielsen (2013), obtaining 0.23 g/L when deleting gene *DIC1*, our results show a great potential. However, it is still far from the amounts obtained in some engineering strategies, such as the triple gene deletion (Δ*FUM1*, Δ*PDC*, Δ*GPD1*) tested by Yan *et al.* (2014), which led to an accumulation of 8.09 g/L of succinate, the highest amount obtained so far, to our knowledge. One should take into attention that these concentration values also depend on the growth conditions, media, pH, nitrogen concentrations, time, etc., although their analysis were not included in the objective of this work.

Comparative genomics of the two best SA producer strains and the two poor producers was performed in order to understand differences in the genome that could be related with the different concentrations obtained by them. Number and distribution of SNPs together with analysis of gene copy number were studied. A higher number of SNPs were detected in good producer isolates, in the last part of chromosomes XIII,

XIV and XV, and in the first third of chromosome III (Fig. 3C). On the contrary, poor-producing strains showed higher number of copies in the genes of chromosome XI, and also one of them (Z12) in almost the entire chromosome I. A set of 42 genes (Supplementary data S2) was chosen as relevant in the pathway of SA production, and studied in detail regarding SNPs' presence and gene copy number differences (Table 1). When considering the total number of SNPs in the analysed genes for the two groups, a similar number was obtained: 93 SNPs in the two higher producing strains versus 95 SNPs in the lower producers. Although if one compares only the higher producer (Z28) with the non-producer (Z63, producing almost 0 g/L of SA), a larger difference is observed: 64 versus 44 SNPs, respectively. This fact proves the importance of the SNP presence, which could indicate that some genes could be mutated in the higher SA-producing strains leading to an accumulation of SA, instead of its transformation in other subproducts of the same pathway. Another interesting and related fact is that the two strains indicated as best producers of SA showed a slightly higher number of SNPs in the succinate dehydrogenase genes (with the exception of *SDH1*, *SDH3* and *SDH8*). This fact again could be related with mutated SDH genes in these strains, which allows the accumulation of succinate instead of its oxidation to fumarate in the TCA cycle. Regarding gene copy number, four genes were identified as having different copy number profile between both groups: *SDH1, SDH3, MDH1* and the transcription factor *HAP4*. Strains chosen as good SA producers showed a decreased number of copies of these four genes, in comparison with the poorly producing strains that have an increase in the number of copies when compared with the reference strain. These genes are important in the degradation of SA in the TCA cycle. In particular, genes *SDH1* and *SDH3* are involved in the oxidation of succinate to fumarate, not allowing their accumulation. Gene *MDH1* codifies the enzyme responsible for the interconversion of malate to oxaloacetate, which occurs in the TCA cycle after the degradation of succinate. In this way, the different number of copies of these genes confirms their importance to explain the higher or lower production of SA in natural strains. Transcription factor *HAP4*, also having its copy number increased in the less-producing strains, constitutes one of the most relevant results obtained. This gene is a transcriptional activator and a global regulator of respiratory gene expression, which is involved in the diauxic shift from fermentative (ethanol production) to oxidative status, and its higher activity will redirect flux over oxidative part of TCA cycle. Although the fact that the poor-producing strains also have *SHD* genes with higher number of copies, it will prevent the accumulation of succinate, oxidizing it to fumarate, and force the TCA cycle to continue. These results are in accordance with previously shown role of *HAP4* in the higher accumulation of succinate in strains with deletions in *SDH* genes (Cimini *et al.* 2009; Raab *et al.* 2011).

In order to confirm results from genomic analysis, a set of 10 genes was selected from the previous group, in particular (*HAP4, ADR1, GCR1, ADH1, ADH2, ADH3, IDH1, KGD1, SDH1 and SDH2*), and comparative transcriptomic analysis of these genes was done comparing the non-producer strain (Z63) with the intermediate producer (Z12) in order to understand which one triggers the beginning of SA production (Fig. 6A), and then comparing the best SA producer strain with the intermediate one, to evaluate which enhances its production (Fig. 6B). Results showed for non-producing strain (panel A) a higher expression of *SDH1*, *SDH2*, *ADH1, ADH3, IDH1* and *HAP4* genes, and a lower expression of *ADR1* transcription factor, in comparison with the higher producer (panel B). Results obtained for *SDH* and *ADH* genes are in accordance with the expected, showing first a shift

Table 1. Gene copy number (GCN) and SNP position (SNP-Pos) of each strain in comparison to the reference strain (S288c), for the 39 genes analysed and three transcription factors.

		SDH1 chr XI: 356024 to 357928	SDH2 chr XII: 53131 to 53931	SDH3 chr XI: 179667 to 180263	SDH4 chr IV: 817950 to 818495	SDH5 chr XV: 196507 to 196995	SDH6 chr IV: 1233278 to 1233517	SDH7 chr IV: 1470017 to 1470418	SDH8 chr II: 742160 to 742576	FUM1 chr XVI: 47336 to 48802	IDH1 chr XIV: 557920 to 559002	IDP1 chr IV: 334835 to 336121	LSC1 chr XV: 593057 to 594046	KGD1 chr IX: 122689 to 125733	MDH1 chr XI: 279123 to 280127	MDH2 chr XV: 81787 to 82920
Z28	GCN	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
	SNP-Pos	0	53316	179949	815632, 815797, 815851, 815880	196663	1233357	1470381	0	0	0	0	593104	125455	0	0
Z56	GCN	1	2	1	2	2	2	2	2	2	2	2	2	2	2	2
	SNP-Pos	0	53316	179949	816125	196663	0	1470381	0	0	0	0	593982	125455	0	82904
Z12	GCN	3	2	3	2	2	2	2	2	2	2	2	2	2	3	2
	SNP-Pos	0	0	179887, 179949	0	196663	0	1470381	0	48340	0	0	593104	125455	0	0
Z63	GCN	3	2	3	2	2	2	2	2	2	2	2	2	2	3	2
	SNP-Pos	0	53316	179703, 179887, 179949	814779, 815632, 815851	0	0	0	0	47946	0	0	0	125455	0	0

Table 1. (*Continued*).

		MDH3 chr IV: 315357 to 316388	CIT1 chr XIV: 629622 To 631061	CIT2 chr III: 120946 to 122328	CIT3 chr XVI: 556377 to 557837	ACO1 chr XII: 735212 to 737548	ADH1 chr XV: 159548 to 160594	ADH2 chr XIII: 873291 to 874337	ADH3 chr XIII: 434788 to 435915	ADH4 chr VII: 15159 to 16307	ADH5 chr II: 533762 to 534817	PDC1 chr XII: 232390 to 234081	PDC5 chr XII: 410723 to 412414	PDC6 chr VII: 651290 to 652981	TPI1 chr IV: 555726 to 556472	GPD1 chr IV: 411825 to 413000
Z28	GCN	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
	SNP-Pos	0	630632	121590	556411, 556756, 556778, 557406	0	159655, 160153, 160213	873817, 873884, 873930	435664	15244, 15311, 15337, 15420, 16087, 16175, 16202	0	232935	411161, 411841	0	0	0
Z56	GCN	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
	SNP-Pos		0	0	0	0	160153, 160213, 160419, 160420, 160549	873817, 873884, 874247	0	15337, 15420, 16087, 16175, 16202	0	0	0	0	0	0
Z12	GCN	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
	SNP-Pos	315716, 315761	630961	121026	556778, 557035	0	159583, 159655, 160213	873817, 873884, 874247	0	15337, 15420, 16087, 16175, 16202	534201	0	411841	0	0	412079
Z63	GCN	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
	SNP-Pos	0	0	0	0	736619	159583, 159655, 160213, 160419, 160420	873817, 873896, 873930	0	15337, 15420	0	232935	411818, 411841, 412172, 412236	0	0	0

Table 1. (*Continued*).

		GPD2 chr XV: 217126 to 218448	GUT2 chr IX: 51759 to 53708	SFA1 chr IV: 159604 to 160764	SER3 chr V: 322686 to 324095	SER33 chr IX: 221081 to 222490	FRD1 chr V: 65385 to 66797	ICL1 chr V: 285241 to 286914	AGX1 chr VI: 76831 to 77988	SHM1 chr II: 736264 to 737736	HAP4 chrXI: 232227 to 233891	ADR1 chr IV: 895035 to 899006	GCR1 chr XVI: 412254 to 415362
Z28	GCN	2	2	2	2	2	2	2	2	2	2	2	2
	SNP-Pos	218101, 218167	0	160513, 160520, 160523, 160615, 160642	0	222311	65944	285738, 286441	0	736876, 736946	232363 233373 233436	895677, 895969, 895993, 896774, 898087, 898191, 898456	413248, 413483, 413829, 413894, 414655, 414884, 415343
Z56	GCN	2	2	2	2	2	2	2	2	2	2	2	2
	SNP-Pos	0	0	160520, 160615	0	222311	0	0	0	0	232567 233436	898087	413894, 414785
Z12	GCN	2	2	2	2	2	2	2	2	2	3	2	2
	SNP-Pos	217612	0	160520, 160615	323117	222261, 222311	65944	285738	0	736861, 736946	233525	895993, 896664, 896761, 896774, 896785, 896814, 898087, 898114, 898399, 898456	413027, 413894, 415343
Z63	GCN	2	2	2	2	2	2	2	2	2	3	2	2
	SNP-Pos	218167	0	160520	0	222311	0	285738, 286686	0	0	232567	895093, 896181, 896352, 896774, 898087, 898456, 898533	413050, 413894, 414010, 414833, 414884, 415343

Each SNP is indicated in terms of its position in the respective chromosome (position in the genome). Underlined SNP positions refer to frameshift changes leading to an aminoacid from a different class (see the text for details).

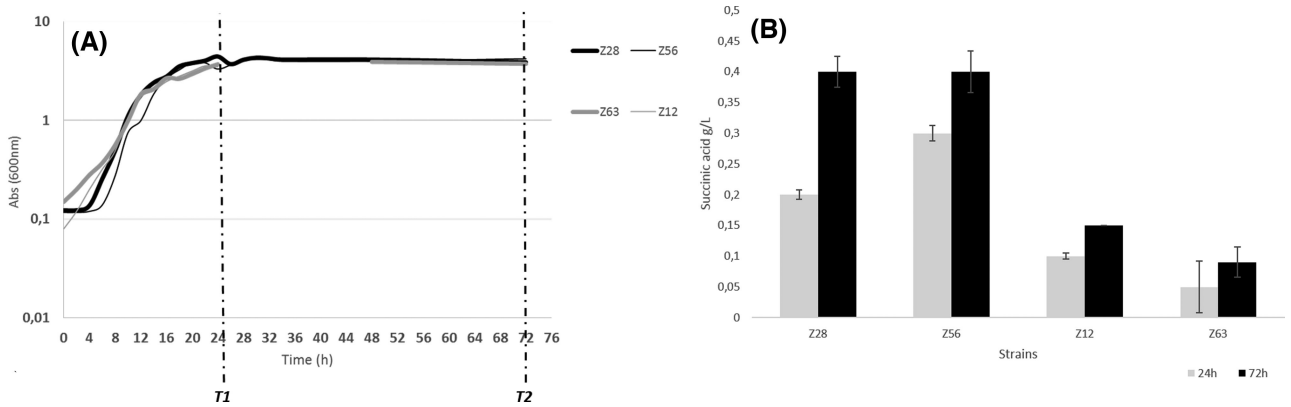

Figure 5. Succinic acid production in the four chosen *S. cerevisiae* strains. (**A**) Growth curve of the four strains. Time points of cell harvest for RNA extraction are marked. (**B**) Succinic acid production (g/L), quantified by HPLC in the four strains, in the two considered time points.

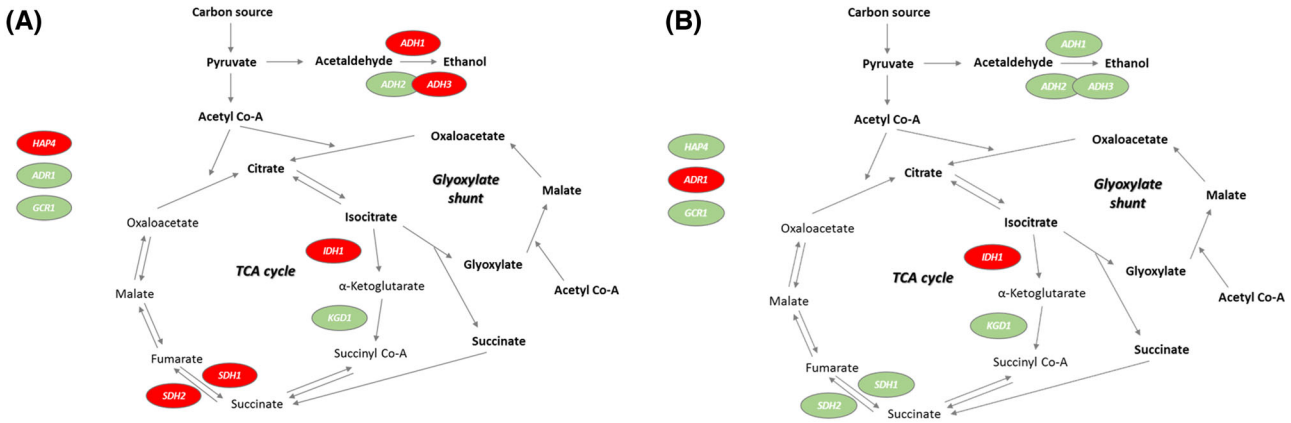

Figure 6. Schematic representation of succinate production metabolic pathways in *S. cerevisiae*. Not all enzymatic steps are shown. Genes studied by transcriptomic analysis are shown inside circles, coloured according to their expression values when comparing between (**A**) non-producer strain (Z63) and intermediate producer strain (Z12) of SA, and (**B**) best producer strain (Z28) and intermediate producer strain (Z12) of SA. Red and green colours correspond to overexpression or underexpression, respectively, of the mentioned gene in the non-producer or the best producer strains when comparing with the intermediate producer strain.

towards the production of ethanol in this strain (showed by the higher expression of *ADH1* and *ADH3* genes), and also a higher oxidation of succinate towards fumarate (showed by the higher expression of *SDH1* and *SDH2*), explaining in this way why this strain does not accumulate succinate. Regarding *ADR1* transcription factor, its involvement in the aerobic oxidation of carbon sources after the diauxic transition is known (Denis and Young 1983). After the diauxic shift, when glucose is residual, yeast growth will rely in the use of metabolic products derived from fermentative reactions, such as ethanol. As already shown, strain Z28 has its metabolism directed for the production of SA via TCA cycle, although also showing capacity to produce some ethanol (as shown in Fig. 1). When glucose is depleted, this strain will continue to produce SA via TCA cycle using this time ethanol as the major source of metabolites and energy, having transcription factor *ADR1* a major role in this part. In the case of *HAP4*, this transcription factor showed a higher expression in the non-producer strain. *HAP4* is a transcriptional activator and global regulator of respiratory gene expression, and is involved in the diauxic shift (Forsburg and Guarente 1989). The involvement and the mechanism of action of this gene in the succinate formation pathway are yet not completely understood, although in the last years some studies have suggested the need to study this gene deeply (Cimini *et al.* 2009; Raab *et al.* 2011). In this study, the higher expression of this gene in the non-producer strain suggests its involvement in the shift from by-product formation to biomass formation and growth, which is confirmed by the previous results obtained. We would expect a lower expression of transcription factor *GCR1* in the non-producer strain in comparison with the higher producer strain, based on studies showing that *gcr1* mutants showed an increase in the expression of TCA and respiratory genes.

In conclusion, this study reports, for the laboratorial conditions tested, the capacity of natural yeast isolates to produce high amounts of SA (1.13 g/L). Our approach attempted to obtain a full yeast characterisation using phenotypic, genetic and metabolic portrayal, combined with whole-genome sequencing and comparative transcriptomics. We could in this way obtain a holistic view of the key factors contributing to reveal the potential ability of natural isolates to be used as successful producers of SA and giving clues for a further strain improvement in order to be applied in industrial applications.

SUPPLEMENTARY DATA

Supplementary data are available at *FEMSYR* online.

ACKNOWLEDGEMENTS

We would like to acknowledge Vladimir Benes and Tobias Rausch from EMBL for all the help regarding whole-genome sequencing.

FUNDING

This work was supported by the TRANSBIO project from the European Community's Seventh Framework Programme (FP7/2007-2013, grant agreement No. 289603), by the EcoAgri-Food project (NORTE-01-0145-FEDER-000009) via the North Portugal Regional Operational Programme (Norte 2020) under the Portugal 2020 Partnership Agreement, and by FCT I.P. through the strategic funding UID/BIA/04050/2013.

Conflict of interest. None declared.

REFERENCES

Agren R, Otero JM, Nielsen J. Genome-scale modelling enables metabolic engineering of *Saccharomyces cerevisiae* for succinic acid production. *J Ind Microbiol Biot* 2013;**40**:735–47.

Arikawa Y, Kuroyanagi T, Shimosaka M *et al*. Effect of gene disruptions of the TCA cycle on production of succinic acid in *Saccharomyces cerevisiae*. *J Biosci Bioeng* 1999;**87**:28–36.

Beauprez JJ, De Mey M, Soetaert WK. Microbial succinic acid production: Natural versus metabolic engineered producers. *Process Biochem* 2010;**45**:1103–14.

Bechthold I, Bretz K, Kabasci S *et al*. Succinic Acid: A new platform chemical for biobased polymers from renewable resources. *Chem Eng Technol* 2008;**31**:647–54.

Bely M, Sablayrolles JM, Barre P. Description of alcoholic fermentation kinetics - its variability and significance. *Am J Enol Vitic* 1990;**41**:319–24.

Bozell JJ, Petersen GR. Technology development for the production of biobased products from biorefinery carbohydrates—the US Department of Energy's "Top 10" revisited. *Green Chem* 2010;**12**:539.

Carole T, Pellegrino J, Paster M. Opportunities in the industrial biobased products industry. *Appl Biochem Biotechnol* 2004;**113**:871–85.

Cheng KK, Wang GY, Zeng J *et al*. Improved succinate production by metabolic engineering. *Biomed Res Int* 2013;**2013**:538790, DOI: 10.1155/2013/538790.

Cimini D, Patil KR, Schiraldi C *et al*. Global transcriptional response of *Saccharomyces cerevisiae* to the deletion of *SDH3*. *BMC Syst Biol* 2009;**3**:17.

Denis CL, Young ET. Isolation and characterization of the positive regulatory gene *ADR1* from *Saccharomyces cerevisiae*. *Mol Cell Biol* 1983;**3**:360–70.

Field D, Wills C. Abundant microsatellite polymorphism in *Saccharomyces cerevisiae*, and the different distributions of microsatellites in eight prokaryotes and *S. cerevisiae*, result from strong mutation pressures and a variety of selective forces. *P Natl Acad Sci USA* 1998;**95**:1647–52.

Forsburg SL, Guarente L. Identification and characterization of *HAP4*: a third component of the CCAAT-bound *HAP2/HAP3* heteromer. *Gene Dev* 1989;**3**:1166–78.

Franco-Duarte R, Mendes I, Umek L *et al*. Computational models reveal genotype – phenotype associations in *Saccharomyces cerevisiae*. *Yeast* 2014;**31**:265–77.

Franco-Duarte R, Umek L, Zupan B *et al*. Computational approaches for the genetic and phenotypic characterization of a *Saccharomyces cerevisiae* wine yeast collection. *Yeast* 2009;**26**:675–92.

Illumina. Mate Pair Library v2 Sample Preparation Guide For 2–5 Kb Libraries. Illumina inc., 2009.

Ito Y, Hirasawa T, Shimizu H. Metabolic engineering of *Saccharomyces cerevisiae* to improve succinic acid production based on metabolic profiling. *Biosci Biotechnol Biochem* 2014;**78**:151–9.

Legras J-L, Ruh O, Merdinoglu D *et al*. Selection of hypervariable microsatellite loci for the characterization of *Saccharomyces cerevisiae* strains. *Int J Food Microbiol* 2005;**102**:73–83.

Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;**25**:1754–60.

McKinlay JB, Vieille C, Zeikus JG. Prospects for a bio-based succinate industry. *Appl Microbiol Biot* 2007;**76**:727–40.

Mendes I, Franco-Duarte R, Umek L *et al*. Computational models for prediction of yeast strain potential for winemaking from phenotypic profiles. *PLoS One* 2013;**8**:e66523.

Otero JM, Cimini D, Patil KR *et al*. Industrial systems biology of *Saccharomyces cerevisiae* enables novel succinic acid cell factory. *PLoS One* 2013;**8**:e54144.

Pérez M, Gallego F, Martínez I *et al*. Detection, distribution and selection of microsatellites (SSRs) in the genome of the yeast *Saccharomyces cerevisiae* as molecular markers. *Lett Appl Microbiol* 2001;**33**:461–6.

Raab AM, Gebhardt G, Bolotina N *et al*. Metabolic engineering of *Saccharomyces cerevisiae* for the biotechnological production of succinic acid. *Metab Eng* 2010;**12**:518–25.

Raab AM, Hlavacek V, Bolotina N *et al*. Shifting the fermentative/oxidative balance in *Saccharomyces cerevisiae* by transcriptional deregulation of Snf1 via overexpression of the upstream activating kinase Sak1p. *Appl Environ Microb* 2011;**77**:1981–9.

Raab AM, Lang C. Oxidative versus reductive succinic acid production in the yeast *Saccharomyces cerevisiae*. *Bioeng Bugs* 2011;**2**:120–3.

Rezaei MN, Aslankoohi E, Verstrepen KJ *et al*. Contribution of the tricarboxylic acid (TCA) cycle and the glyoxylate shunt in *Saccharomyces cerevisiae* to succinic acid production during dough fermentation. *Int J Food Microbiol* 2015;**204**:24–32.

Schuller D, Valero E, Dequin S *et al*. Survey of molecular methods for the typing of wine yeast strains. *FEMS Microbiol Lett* 2004;**231**:19–26.

Silva Ferreira AC, Guedes de Pinho P, César A *et al*. Norisoprenoids profile during port wine ageing—influence of some technological parameters. *Anal Chim Acta* 2004;**513**:169–76.

Song H, Lee SY. Production of succinic acid by bacterial fermentation. *Enzyme Microb Technol* 2006;**39**:352–61.

Vemuri G, Eiteman M, Altman E. Effects of growth mode and pyruvate carboxylase on succinic acid production by metabolically engineered strains of *Escherichia coli*. *Appl Environ Microb* 2002a;**68**:1715–27.

Vemuri GN, Eiteman MA, Altman E. Succinate production in dual-phase *Escherichia coli* fermentations depends on the time of transition from aerobic to anaerobic conditions. *J Ind Microbiol Biot* 2002b;**28**:325–32.

Wakai Y, Shimazaki T, Hara S. Formation of succinate during fermentation of sake mash and grape must. *Hakkokogaku Kaishi* 1980;**58**:363–8.

Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010;**38**:e164.

Yan D, Wang C, Zhou J *et al*. Construction of reductive pathway in *Saccharomyces cerevisiae* for effective succinic acid fermentation at low pH value. *Bioresource Technol* 2014;**156**:232–9.

Zeikus JG, Jain MK, Elankovan P. Biotechnology of succinic acid production and markets for derived industrial products. *Appl Microbiol Biot* 1999;**51**:545–52.