Genotyping of *Saccharomyces cerevisiae* strains by interdelta sequence typing using automated microfluidics

Ricardo Franco-Duarte[1], Inês Mendes[1] Ana Catarina Gomes[2], Manuel A.S. Santos[2,3], Bruno de Sousa[4], Dorit Schuller[1]

[1] CBMA (Centre of Molecular and Environmental Biology) / Department of Biology / University of Minho, Braga, Portugal

[2] BIOCANT - Biotechnology Innovation Center, Cantanhede, Portugal

[3] RNA Biology Laboratory, CESAM, Biology Department, Aveiro University, Campus Universitário de Santiago, Aveiro

[4] Centre for Malaria & Tropical Diseases Associated Laboratory, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa, Portugal.

*For correspondence:

Dorit Schuller

e-mail: dschuller@bio.uminho.pt

Tel.: (+351) 253 604 310/17

Fax: (+351) 253 678 980

Abbreviations: mtDNA - mitochondrial DNA; MLST - multi locus sequence typing

Keywords: capillary electrophoresis, interdelta sequences, non-parametric methods, *Saccharomyces cerevisiae*

Number of words: 5764

Abstract

Amplification of genomic sequences flanked by delta elements of retrotransposons TY1 and TY2 is a reliable method for characterization of *Saccharomyces cerevisiae* strains. The aim of this study is to evaluate the usefulness of microfluidic electrophoresis (Caliper LabChip®) to assess the factors that affect interlaboratory reproducibility of interdelta sequence typing for *S. cerevisiae* strain delimitation. We carried out experiments in two laboratories, using varying combinations of *Taq* DNA polymerases and thermal cyclers. The reproducibility of the technique is evaluated using non-parametric statistical tests and we show that the source of *Taq* DNA polymerase and technical differences between laboratories have the highest impact on reproducibility, whereas thermal cyclers have little impact. We also show that the comparative analysis of interdelta patterns is more reliable when fragment sizes are compared, than when absolute and relative DNA concentrations of each band are considered. Interdelta analysis based on a smaller fraction of bands with intermediate sizes between 100 and 1000 bp yield the highest reproducibility.

Introduction

Biotechnological processes conducted by *Saccharomyces cerevisiae* strains are gaining increasing importance. Tracking inoculated strains throughout productive processing is necessary for quality assurance in fermentative processes such as bioethanol production or wine fermentation. Besides, yeast has been identified as an emerging human pathogen capable of causing clinically relevant infections in immune compromised patients [1, 2]. Therefore, quick and accurate methods for yeast strains delimitation that rely on high-throughput genotyping methods based on microfluidics systems can be of interest in both industrial and clinical contexts.

Numerous molecular methods have been developed for yeast strain characterization, such as chromosome separation by pulsed field electrophoresis [3, 4], restriction fragment length polymorphism analysis of mitochondrial DNA (mtDNA RFLP) [5-8], random amplified polymorphic DNA (RAPD) [9], PCR fingerprinting followed by enzymatic restriction of amplified DNA [10], multi locus sequence typing (MLST) [11], microsatellite analysis [12-14], real-time PCR [15, 16] and PCR-amplification of inter-delta sequences [17, 18]. Delta sequences are flanking sequences (300 bp) of retrotransposons TY1 and TY2 that are dispersed throughout the genome (particularly in terminal chromosomal regions), but can also be found as single elements. About 300 delta elements were described in the genome of the laboratory strain S288c. Since the number and location of delta elements have a certain intraspecific variability they are appropriate genetic markers for the identification of polymorphisms. Amplification of interdelta regions between neighbouring delta sequences results in a mixture of differently sized strain-specific fragments. This PCR-based method is easy to perform, cheap and rapid, and therefore suitable for the characterization of high number of strains.

More recently, the interdelta method was improved by the use of alternative primers (δ12 and δ21) [17] that bind close to the initially described binding sites for primers δ1 and δ2 [18]. The combined use of these improved primer combinations (δ12 / δ21 or δ12 / δ2) revealed greater banding pattern polymorphism and improved discriminatory power [13]. The use of primer pairs δ12 / δ2 showed the same discriminatory power of other methods for strain delimitation, such as mtDNA RFLP, microsatellite analysis and karyotyping [19]. However, this method requires careful standardization of DNA concentration [20]. Occasional non-reproducible "ghost bands" are present due to the low annealing temperature (43 ºC), which is a disadvantage of the interdelta method. Increasing the annealing temperature to 55 ºC reduced ghost bands, but lead to poorer banding pattern and consequently reduced discriminatory power [21]. In summary, PCR profiling analysis of delta sequences is associated with good discriminatory power for the analysis of commercial strains [22], but the use of this typing method for routine analysis of yeast strains requires careful evaluation [21, 23-26]. It is therefore advisable to use additional methods such as mtDNA RFLP or microsatellite analysis to confirm ambiguous results.

Fluorescent primers and automated DNA sequencers improve significantly banding patterns containing weakly amplified fragments [27], decreasing experimental error and increasing data throughput, scoring and reliability [28]. When interdelta sequences are amplified with fluorescent primers, followed by capillary electrophoresis, the resolution of the obtained profiles is considerably increased in comparison with standard agarose gel electrophoresis [29].

The efficiency of PCR amplification is affected by numerous factors namely annealing temperature, the concentration of $MgCl_2$, primers and template DNA. Even slight variations in these parameters may affect results compromising data comparisons and

sharing between experiments and laboratories [30]. The optimal reaction conditions need to be optimized for each PCR application.

Microfluidics are gaining notoriety across broads research fields, e.g., forensics, clinical and genetic analysis [31-33]. Miniaturized reactions economize DNA samples, reagents and analytical time considerably, and increase sensitivity, throughput and automation possibilities [34, 35]. In the microfluidic chips for DNA analysis of the Caliper´s LabChip® system, DNA samples are electroosmotically transported and fragmented inside the chip, separated by capillary electrophoresis and finally analyzed using fluorescence detection [36].

Genome-wide studies of yeast inter-strain variability require bio-databanks for biodiversity conservation, sustainable development of genetic resources and equitable sharing of genotypic data among laboratories. We consider interdelta sequences amplification as a very useful method for high-throughput characterization of *S. cerevisiae* strains, which is easy to perform, cheap and rapid in comparison to other molecular methods. The aim of this study is to evaluate the impact of two different *Taq* polymerases on the interlaboratory reproducibility of interdelta sequence typing for yeast strain delimitation using microfluidics electrophoresis (Caliper´s LabChip®). Besides, we also evaluate the impact of different thermal cyclers on the patterns obtained. The study demonstrates that the reproducibility of the technique is most affected by the source of *Taq* DNA polymerase and technical differences between laboratories such as different operators. Interlaboratory reproducibility is highest when fragment sizes between 100 and 1000 bp are compared, rather than absolute and relative DNA concentrations of each band.

Materials and methods

Yeast strains and culture

Saccharomyces cerevisiae strains used in this work were collected in the Vinho Verde wine region (northwest Portugal) during three consecutive vintages (2001-2003). From a collection of 300 isolates, the 12 strains with highest genetic heterogeneity, according to their allelic microsatellite combinations for loci ScAAT1-ScAAT6 [37], were selected using neuronal networks [38]. Strains were named as follows: R8, R16, R20, R21, R30, R58, R60, R61, R62, R81, R88 and R101.

Interdelta sequences amplification and analysis

Yeast cells were cultivated (36 h, 28 ºC, 160 rpm) in 1 mL of YPD medium (yeast extract 1% w/v, peptone 1% w/v, glucose 2% w/v) and the DNA isolation was performed as previously described [6]. Briefly, cells were suspended in a sorbitol containing buffer in the presence of lyticase for cell wall degradation. Cells were then lysed by SDS addition, followed by DNA purification with sodium acetate and isopropanol to eliminate proteins, polysaccharides, RNA or other cell constituents. Subsequently, DNA was precipitated with ethanol, resuspended in TE and quantified (Nanodrop, Thermo Scientific). DNA amplification was performed recurring to primers δ12 (5'-TCAACAATGGAATCCCAAC - 3') and δ2 (5'- GTGGATTTTTATTCCAAC - 3') [17]. Thirty μL of reaction mixture was prepared with 120 ng of DNA, *Taq* buffer (10 mM Tris-HCl, 50 mM KCl, 0.08 % Nonidet P40), 50 pmoles of each primer, 0.4 mM of each dNTP, 3 mM $MgCl_2$ (MBI Fermentas) and 1.0 U of *Taq* DNA polymerase. After initial denaturation (95 ºC for 2 min), the reaction mixture was cycled 35 times using the following settings: 95 ºC for 30 s, 43.2 ºC for 1 min, 72 ºC for 1 min, followed by a final extension at 72 ºC during 10 min. Characteristic PCR profiles of the 12 strains are shown in Figure 1.

An experimental strategy was devised to study the reproducibility of the interdelta sequence amplification as a typing method for yeast strains using 96-well PCR plates and the following combinations of *Taq* DNA polymerase, thermal cyclers and laboratories: Plate 1 - commercial *Taq* (MBI Fermentas recombinant *Taq,* Ref. EP0402), BioRad MyCycler thermal cycler, laboratory 1 (8 replicates per strain); Plate 2 - in-house cloned and produced *Taq*, BioRad MyCycler thermal cycler, laboratory 1 (8 replicates per strain); Plate 3 - in-house cloned and produced *Taq*, Eppendorff Mastercycler thermal cycler, laboratory 1 (8 replicates per strain); Plate 4 - commercial *Taq* (MBI Fermentas recombinant *Taq* Ref. EP0402) or in-house cloned and produced *Taq* (4 replicates per strain), BioRad MyCycler thermal cycler, laboratory2. This approach resulted in 32 replicates for each strain and a total of 384 electrophoretic banding patterns. The four microplates thus included the following conditions to be compared: A - Commercial *Taq*, BioRad thermal cycler, laboratory 1; B - In-house *Taq*, BioRad thermal cycler, laboratory 1; C - In-house *Taq*, Eppendorff thermal cycler, laboratory 1; D - Commercial *Taq*, BioRad thermal cycler, laboratory 2; E - In-house *Taq*, BioRad thermal cycler, laboratory 2. Both laboratories used the same DNA samples and the same in-house cloned and commercial *Taq* enzymes. Amplifications were carried out with the same PCR buffer (MBI Fermentas, Ref. B33). PCR products were analyzed using a high-throughput automated microfluidic electrophoresis system (Caliper LabChip® 90 Electrophoresis System) and a 96-well plate format, according to the manufacturer's instructions. The tolerance of the sizing resolution for this system is ±15% (from 25 to 100 bp), ±10% (from 100 to 150 bp), ±5% (from 150 to 700 bp), ±10% (from 700 to 1000 bp).

Statistical analysis of electrophoretic data

The size (bp) and concentration (ng of DNA) of each band was determined using the LabChip® HT software (version 2.6) and exported to the software SPSS 18.0 package for the composition of a matrix containing data for each band of the 32 replicates banding patterns from each strain. Each band was analyzed and compared in terms of fragment sizes (bp), absolute DNA concentration (ng/μl) and relative DNA concentrations (%) (absolute concentration value was divided by the sum of all concentration values of all bands contained in a replicate banding pattern). An exploratory data analysis was performed, where normality distribution (Kolmogorov-Smirnov and Shapiro-Wilk tests) and variance homogeneity (Levene's test) were tested using SPSS 18.0. After several unsuccessful transformations of the data, non-parametric tests were performed, such as "Kruskall-Wallis one-way analysis of variance" test, to check for the equality of treatment medians among the different groups. More precisely, the null hypothesis (H_0) assuming equality of all medians was tested against the alternative hypothesis (H_1), which assumes that at least two of the strains show differences in their medians, as outlined below:

$$H_0: \theta_1=\theta_2=\cdots=\theta_{12} \quad \text{vs} \quad H_1: \exists_{(i,j)}:\theta_i\neq\theta_j \text{ for some } i\neq j, \qquad (1)$$

where θ_i represents the median concentration (or percentage of concentration) for the i^{th} strain, i=1,…,12.

In cases where the test produced statistical significant differences between strains, multiple pairwise comparisons were performed to trace the origin of such differences. The method proposed by Conover and Iman [39] searches for comparative magnitudes of the means based on the rank data, and assumes the t-student distribution. The test is based on the following expression:

$$\left|\frac{R_i}{n_i}-\frac{R_j}{n_j}\right| \geq t_{1-\frac{\alpha}{2}} \sqrt{\frac{S^2(N-1-H_c)}{N-k}.\left(\frac{1}{n_i}+\frac{1}{n_j}\right)} \quad (2)$$

with $t_{1-(\alpha/2)}$ the $(1-\alpha/2)$ quantile of a t-student distribution with (N-k) degrees of freedom, k the number of groups, H_c the value for the test statistic of the Kruskal-Wallis test corrected for ties and S^2 the corresponding variance.

Results

Electrophoretic profile of the Saccharomyces cerevisiae *strains*

Interdelta fragments of 12 genetically heterogeneous strains were amplified, using primers δ12 and δ2 and were analyzed using automated microfluidics electrophoresis (Caliper LabChip® 90 Electrophoresis System). In order to evaluate the inter-laboratorial reproducibility of the banding patterns and to determine which combination of *Taq* DNA polymerase and thermal cycler produced the most reproducible banding patterns between both laboratories, the experimental design included different combinations of the mentioned factors, as described in the Materials and Methods section. Unique banding patterns were obtained for each strain (Figure 1). The most common band was present in 9 out of the 12 strains and had a size of approximately 400 bp. Quantitative and qualitative analysis of each band was performed using the software package of the electrophoresis system, using the values of the co-injected internal markers (gel bands at 15 and 7000 bp) as a reference. The analysis presented herein is based on the length of the amplified fragments (bp), and the absolute and relative (%) values of DNA concentration (ng/μL) of each band, as outlined in the Material and Methods section.

Figure 2 shows an example of 32 replicate banding patterns of a representative strain tested under the conditions indicated in the first paragraph of Material and Methods section. Fragment sizes showed high reproducibility between replicates of the same condition and between conditions. Considerable differences were observed when, for each experimental condition, DNA concentrations were compared. The most intense banding patterns were obtained in laboratory 1, using in-house cloned and produced *Taq* and the Eppendorff thermal cycler (condition C), followed by condition B and A. The in-house produced *Taq* polymerase (C) amplified PCR products more efficiently than

commercial *Taq* (B). This agrees with the slightly stronger banding patterns of condition E compared to condition D in laboratory 2. These trends were similar for the other eleven strains (data not shown). One of eight replicates of condition A (corresponding to the 8^{th} lane of Figure 2) failed amplification for most strains due to lateral evaporation of the PCR reaction mixture during cycling in the 96-well plates. These replicates were excluded from further analysis.

Reproducibility of PCR-based interdelta typing

Our main goal in this study was to identify statistically significant differences between the banding patterns of yeast strains, generated under conditions A-E (see above), to enhance reproducibility of interdelta sequence analysis between laboratories. In the first step of the statistical analysis the data was verified for normality between the 12 strains and the corresponding homogeneity of variances. Kolmogorov-Smirnov and Shapiro-Wilk tests were used to investigate the normality assumption. The results (data not shown) revealed that our data did not follow a normal distribution since all *p*-values were approximately zero (<0.001) and, therefore, smaller than any of the usual levels of significance considered (1 %, 5 % and 10 %). Homogeneity of variances between strains was tested using Levene's test. This condition was also not satisfied by the data (data not shown), as *p*-values were approximately zero (<0.001) for both variables in the study. In an attempt to satisfy both normality and homogeneity of variances, data were transformed using logarithm of base 2 and inverse values of absolute or relative concentrations. New variables were created in SPSS, both for absolute and relative values. Once again, the normality and homogeneity of variance assumptions were rejected (data not shown), which lead us to use non-parametric tests.

The Kruskal-Wallis one-way analysis of variance was used to test equality of medians among the groups of strains corresponding to each of the previously mentioned

condition (A-E) , using the formula (1) shown in the Material and Methods section. The median was the measure of centrality for this test. It was expected that, in case of reproducibility, all strains should have similar results, meaning that the values of concentration (absolute or relative) and of fragment sizes (bp) should not differ in terms of the median values. However, the Kruskal-Wallis test rejected the equality of medians between groups, because once again the *p*-values were approximately 0 (<0.001). The following approach consisted in searching for differences in terms of the median values of fragment sizes (bp) and concentration values (absolute and relative) between strains. This approach was repeated for the distinct experimental conditions used (A-E) in order to search for the factors that most affect the reproducibility of the technique among the conditions A-E. Based on the results from the Kruskall-Wallis one-way analysis of variance, we assumed that at least two strains showed a difference in the medians. In order to identify the strains that lead to the rejection of the equality of the medians, Multiple Pairwise Comparisons, pooling the data for all 32 replicates per strain, were performed. All 3892 values (the total number of observations regarding all experiments, i.e. all bands of the 32 replicates of the 12 strains), were ordered by increasing numbers and a rank score was calculated for identical values of absolute and relative concentrations. Then, the formula (2) shown in the Material and Methods section was applied for pairwise strain comparisons, based on a *t*-student distribution to search for the origins of the differences between experimental conditions. The results of this test are summarized in Table 1, for each pair compared, for each strain and using the fragment size (bp), as well as absolute and relative DNA concentration values. Statistical significant differences were observed when comparing all 3892 records against each other, being the significant ones (based on a *t*-student significance test) represented with gray squares in Table 1. In the bottom part of this Table (last three

lines), overall percentages are represented considering the differences between strains and between conditions, both for fragment size base pairs and absolute and relative DNA concentrations values. The inter-laboratory banding patterns reproducibility was rather low as observed by the distribution of gray squares in the corresponding main columns. Significant differences were found between strains analyzed in the two laboratories. The lack of reproducibility of these experiments between laboratories was not visible when analyzing the intervals of overall percentages. One could see that these intervals were very comprehensive (including 0 and 100%) and that this analysis was inconclusive for these comparisons. The reasons for this could be due to strain specific effects and also to the extreme values included in the statistical. For example, strain R101 was associated with 0% of statistically significant differences regarding absolute DNA concentration, while for strain R88, regarding fragment size 100% of significant differences were obtained. The cloned and in-house produced *Taq* increased reproducibility between laboratories relative to commercial *Taq*. The comparison between *Taq* polymerases produced data heterogeneity between laboratories. Low and high reproducibility was found between enzymes for laboratory 1 and 2, respectively (columns 3 and 4). This was shown by the higher number of gray squares in column 3 in comparison to column 4, and also by the intervals of overall percentages of significant differences (75-100% comparing to 8-50% regarding fragment length; 16-100% comparing to 0-42% regarding absolute concentration values; 83-100% in comparison to 0-58% regarding relative concentration values).

Regarding the different thermal cyclers used, experimental variation in laboratory 2 lead to more reproducible results, as shown by the comparison of fragment sizes. This reproducibility was not so evident when comparing absolute and relative concentration values.

When analyzing all conditions together, the comparison of absolute DNA concentration values produced the most reproducible results, followed by fragment size and relative DNA concentration values. Relative concentration values should not be used, however, because in replicate analysis of strains under different experimental conditions, distinct numbers of fragments were obtained, affecting the ratios of relative concentration.

Comparison of different experimental conditions for strains delimitation

To identify the experimental condition that best differentiate the 12 yeast strains, statistical analysis of the differences between group medians for each experimental condition was performed. For each experimental condition (from A to E), the percentage of significant differences between the strains was calculated (excluding the comparisons between the same strain for each experimental condition). Figure 3 shows that combination C (in-house cloned *Taq*, Eppendorff thermal cycler, laboratory 2) lead to the highest percentages regarding size, absolute and relative DNA concentration values. This suggests that this is the most suitable combination of experimental conditions for strain delimitation using interdelta banding patterns. Regarding fragment size and relative DNA concentration, these percentages were almost 100, meaning that the 12 electrophoretic patterns would correspond to 12 different strains. On the contrary, combinations A (Commercial *Taq*, BioRad thermal cycler, laboratory 2), D (Commercial *Taq*, BioRad thermal cycler, laboratory 1), and E (in-house *Taq*, BioRad thermal cycler, laboratory 1) were less capable of differentiating strains with only 28.79 %, 51.52 % and 40.91 % of correctly delimited strains regarding fragment sizes, respectively. Similar results were observed when comparisons were performed based on absolute and relative DNA concentrations. In general terms, the use of in-house cloned *Taq* polymerase led to better results than the use of commercial *Taq* polymerase, as can be observed when comparing combination A and D (commercial *Taq*) with

combinations B, C and E (in-house *Taq*). Regarding the laboratories where the PCR reactions were carried out, the strain patterns in laboratory 2 were better separated than those obtained in laboratory 1 (combinations A, B and C versus combinations D and E). The best results regarding strains differentiation were obtained when using relative DNA concentration values (100 % with combinations B and C), however the latter produced biased results. This is explained by the fact that, to calculate the relative DNA concentration values, the absolute values were divided by the sum of all concentration values of all bands contained in a banding pattern. In replicate analysis of different experimental conditions, distinct numbers of fragments were obtained affecting the ratios of relative concentration, leading to overestimated strain delimitation. Due to this we consider that the percentages obtained for the analysis of absolute DNA concentrations are more realistic to delimitate strains than relative DNA concentration value. Fragment length analysis is the preferable measure for typing of yeast strains using interdelta fragments amplification, even though the reproducibility associated was smaller compared to absolute values of concentration (Table 1), but producing more consistent results without introducing biases in the reproducibility of the technique.

Determination of identical banding patterns for each strain in all conditions

To gain further insight into the reproducibility of the interdelta sequence typing method, we tried to identify for each strain the bands that were amplified across the A-E experimental conditions. Strain R60, which showed a very different banding pattern was excluded from this analysis. As shown in Table 2, three to seven bands in the range of 100 – 900 bp were apparent in all 32 replicates of each strain. The respective standard deviations were rather low, ranging from 1.3 to 15.6 bp. Additional bands were mostly found for fragment sizes between 1000 and 1500 bp or below 100 bp, and were not represented because of lack of reproducibility. Some intermediate fragments were also

not included in Table 3 because they were represented only in some experimental conditions. Reproducibility would approximate to 100%, if only the bands included in Table 2 would be used for comparison of fragment sizes.

Discussion

The improved interdelta method [17] is suitable for the typing of yeast strains [19]. This method is simple, rapid and less expensive than others, such as sequencing and microsatellite amplification. Although less rigorous than other techniques as MLST or microsatellite amplification, the PCR-based interdelta method is suitable for high-throughput analysis of large strain collections using microfluidic electrophoresis. The amplification of interdelta regions results in a mixture of differently sized-specific fragments. As previously shown by BLAST analysis [17], the sequences of fragments obtained by amplification with primers δ12 and δ21 matched the predicted interdelta regions. We have designed an inter-laboratory approach to evaluate the performance and the reproducibility of this method as a high-throughput typing approach for the genetic characterization of yeast strains. The comparative approaches that we describe herein can contribute to the constitution of bio-databanks for equitable sharing of genotypic data among laboratories in the context of biodiversity conservation and sustainable development of genetic resources. However, it is crucial to find a set of parameters leading to most reproducible patterns between laboratories.

As outlined in the Materials and Methods section, interdelta sequences of 12 strains were amplified, under varying conditions (*Taq* DNA polymerase, thermal cycler and laboratory). Interdelta sequence typing showed the reproducibility necessary for implementation as a typing method for multiple (4 or 8) replicates of one strain, under identical experimental conditions. The use of the microfluidic LabChip® system greatly contributed to achieve very precise data with a high resolution, as reported in previous works [28, 29].

In general, DNA amplification depends on numerous factors such as the method of DNA isolation, the concentrations of DNA, primers, $MgCl_2$, dNTPs, the *Taq*

polymerase and the annealing temperature. In the present work only one DNA extraction was performed for each strain, and the same DNA was used by both laboratories, being therefore no variable in our experiments. Our (unpublished) results showed that the DNA extraction protocol used is the most appropriate and leads to much better results than an extraction method using phenol. DNA quantification was performed in the Nanodrop^{TM} system, which allowed unambiguous evaluation of the DNA quality. In recent publications [17, 19, 23, 26, 29, 40], DNA concentration values were in the range of 0.1 - 2.5 ng/µl (final concentration). Fernandez-Espinar (2001) showed that the optimal DNA quantities ranged from 0.6 to 2.5 ng/µl (final concentration). The highest number of bands was amplified using the concentration of 2.5 ng/µl, which is similar to the concentration used throughout this work (4 ng/µl). In the publications mentioned above, optimal $MgCl_2$ concentrations ranged from 1.5 to 3.0 mM, whereas the primer and dNTP concentrations were in the range of 1 to 1.67 μM and 200 to 400 μM, respectively. In our (unpublished) optimization approaches, we found that more fragments were amplified when using 3.0 mM $MgCl_2$, 400 μM dNTPs and 1.67 μM of each primer. We suppose that these higher concentrations of primers and dNTPs are necessary to amplify a group of fragments, contrarily to a PCR reaction where just one single band is amplified.

The main objective of the present work was to show the extent of variation due to factors such as the DNA polymerase or the thermal cycler. A commercial *Taq* DNA polymerase and an in-house cloned and produced *Taq* were used, and different amplification patterns were found. In our (unpublished) optimization approaches several commercial *Taq* enzymes were tested, whereas the *Taq* polymerase used in this study revealed to be most suitable for interdelta amplification. The choice of the polymerase is therefore important before setting up PCR reactions. Several factors can contribute to

403 the differences found between the commercial and the in-house cloned *Taq*, such as the
404 preparation method (residual salt content), and/or an inaccurately measured enzymatic
405 activity of the in-house *Taq*. Besides, this *Taq* might be less purified and contain
406 residual cellular compounds that could contribute to better performance. All references
407 regarding interdelta amplification report a quite low annealing temperature
408 (predominantly 43ºC to 46ºC) [17-20, 22, 26, 29, 41]. Higher temperatures (55ºC) lead
409 to a more stable fragment profile, but reduce significantly the number of bands that are
410 amplified [21]. Our previous (unpublished) data revealed that 43.2ºC was the best
411 temperature to achieve both a high number of amplified bands and increased
412 reproducibility of the electrophoretic profiles.
413 Although the DNA samples used for interdelta fragments amplification were the same
414 for both laboratories, the accomplishment of experiments in different laboratories, the
415 use of different *Taq* DNA polymerases and thermal cyclers reduced reproducibility. In
416 fact, the same isolate could be considered as a different strain if typed in different
417 laboratories, due to the experimental variation associated with the conditions A-E. The
418 highest variability was associated with the source of *Taq* DNA polymerase and to
419 laboratory specific technical details, whereas the effect of the thermal cycler was low.
420 Both laboratories used the same aliquot of *Taq* polymerase. If different batches from the
421 same supplier were used in both laboratories, it is possible that the reproducibility
422 would be even more affected. Despite the mentioned limitations, PCR amplification of
423 interdelta sequences is most indicated for the typing of large strain collections, and a
424 high reproducibility is achieved for replicates within the same experimental conditions.
425 When considering interlaboratory experiments, a careful standardization of all the
426 factors that can interfere with the PCR reaction is mandatory to eliminate variability
427 caused by the source of *Taq* DNA polymerase and minor experimental differences

between laboratories. This study also demonstrates that, for reliable data sharing between laboratories, comparative interdelta sequence analysis should be based on a reduced number of bands that lead to reproducible banding pattern profiles.

ACKNOWLEDGEMENTS

This work was funded by the fellowship SFRH/BD/48591/2008 and by the projects POCI/AGR/56102/2004, PTDC/BIA-BCM/64745/2006 and PTDC/AGR-ALI/103392/2008 from the Portuguese Research Agency (Fundação para a Ciência e Tecnologia). The research leading to these results has also received funding from the European Community´s Seventh Framework Programme (FP7/2007-2013) under grant agreement nº 232454, and MCI grant MTM2008-01603.

References

[1] Aucott, J. N., Fayen, J., Grossnicklas, H., Morrissey, A., *et al.*, *Rev Infect Dis* 1990, *12*, 406-411.

[2] Hazen, K. C., *Clin Microbiol Rev* 1995, *8*, 462-478.

[3] Blondin, B., Vezinhet, F., *Revue Française d' Oenologie* 1988, *28*, 7-11.

[4] Carle, G. F., Olson, M. V., *PNAS (USA)* 1985, *82*, 3756-3760.

[5] Dubordieu, D., Sokol, A., Zucca, J., Thalouarn, P., *et al.*, *Connais Vigne Vin* 1984, *21*, 267-278.

[6] Lopez, V., Querol, A., Ramon, D., Fernandez-Espinar, M. T., *Int J Food Microbiol* 2001, *68*, 75-81.

[7] Querol, A., Barrio, E., Huerta, T., Ramon, D., *Appl Environ Microbiol* 1992, *58*, 2948-2953.

[8] Vezinhet, F., Blondin, B., Hallet, J. N., *Appl. Microbiol. Biotechnol.* 1990, *32*, 658–671.

[9] Corte, L., Lattanzi, M., Buzzini, P., Bolano, A., *et al.*, *J Appl Microbiol* 2005, *99*, 609-617.

[10] Baleiras Couto, M. M., Eijsma, B., Hofstra, H., Huis in't Veld, J. H., van der Vossen, J. M., *Appl Environ Microbiol* 1996, *62*, 41-46.

[11] Ayoub, M. J., Legras, J. L., Saliba, R., Gaillardin, C., *J Appl Microbiol* 2006, *100*, 699-711.

[12] Hennequin, C., Thierry, A., Richard, G. F., Lecointre, G., *et al.*, *J Clin Microbiol* 2001, *39*, 551-559.

[13] Legras, J. L., Ruh, O., Merdinoglu, D., Karst, F., *Int J Food Microbiol* 2005, *102*, 73-83.

[14] Perez, M. A., Gallego, F. J., Martinez, I., Hidalgo, P., *Lett Appl Microbiol* 2001, *33*, 461-466.

[15] Martorell, P., Querol, A., Fernandez-Espinar, M. T., *Appl Environ Microbiol* 2005, *71*, 6823-6830.

[16] Hierro, N., Esteve-Zarzoso, B., Gonzalez, A., Mas, A., Guillamon, J. M., *Appl Environ Microbiol* 2006, *72*, 7148-7155.

[17] Legras, J. L., Karst, F., *FEMS Microbiol Lett* 2003, *221*, 249-255.

[18] Ness, F., Lavalee, F., Dubordieu, D., Aigle, M., Dulau, L., *J. Sci. Food Agric* 1993, *62*, 89-94.

[19] Schuller, D., Valero, E., Dequin, S., Casal, M., *FEMS Microbiol Lett* 2004, *231*, 19-26.

[20] Fernandez-Espinar, M. T., Lopez, V., Ramon, D., Bartra, E., Querol, A., *Int J Food Microbiol* 2001, *70*, 1-10.

[21] Ciani, M., Mannazzu, I., Marinangeli, P., Clementi, F., Martini, A., *Ant Leeuwenhoek* 2004, *85*, 159-164.

[22] Lavallée, F., Salvas, Y., Lamy, S., Thomas, D. Y., *et al.*, *Am J Enol Viticult* 1994, *45*, 86-91.

[23] Pramateftaki, P. V., Lanaridis, P., Typas, M. A., *J Appl Microbiol* 2000, *89*, 236-248.

[24] Lopes, C. A., van Broock, M., Querol, A., Caballero, A. C., *J Appl Microbiol* 2002, *93*, 608-615.

[25] Cappello, M. S., Bleve, G., Grieco, F., Dellaglio, F., Zacheo, G., *Journal of Applied Microbiology* 2004, *97*, 1274–1280.

[26] Demuyter, C., Lollier, M., Legras, J. L., Le Jeune, C., *J Appl Microbiol* 2004, *97*, 1140-1148.

[27] Terefework, Z., Kaijalainen, S., Lindstrom, K., *J Biotechnol* 2001, *91*, 169-180.
[28] Papa, R., Troggio, M., Ajmone-Marsan, P., Nonnis Marzano, F., *J Anim Breed Genet* 2005, *122*, 62-68.
[29] Tristezza, M., Gerardi, C., Logrieco, A., Grieco, F., *J Microbiol Methods* 2009, *78*, 286-291.
[30] Vilioen, G. J., Nel, L. H., Crowther, J. R., *Molecular Diagnostic PCR Handbook*, Springer, P.O. Box 17, 3300 AA Dordrecht, The Netherlands. 2005.
[31] Tudos, A. J., Besselink, G. A. J., Schasfoort, R. B. M., *Lab on a Chip* 2001, *1*, 83-95.
[32] Verpoorte, E., *Electrophoresis* 2002, *23*, 677-712.
[33] Ryley, J., Pereira-Smith, O. M., *Yeast* 2006, *23*, 1065-1073.
[34] Whitesides, G. M., *Nature* 2006, *442*, 368-373.
[35] Lion, N., Reymond, F., Girault, H. H., Rossier, J. S., *Current Opinion in Biotechnology* 2004, *15*, 31-37.
[36] Mark, D., Haeberle, S., Roth, G., von Stetten, F., Zengerle, R., *Chem Soc Rev* 2010, *39*, 1153-1182.
[37] Schuller, D., Casal, M., *Ant Leeuwenhoek* 2007.
[38] Aires-de-Sousa, J., Aires-de-Sousa, L., *Bioinformatics* 2003, *19*, 30-36.
[39] Conover, W. J., Iman, R. L., *Technical Report, LA-7677-MS. Los Alamos Scientific Laboratory* 1979.
[40] Fernandez-Gonzalez, M., Espinosa, J. C., Ubeda, J. F., Briones, A. I., *Syst Appl Microbiol* 2001, *24*, 634-638.
[41] Masneuf, I., Dubourdieu, D., *Journal International Des Sciences De La Vigne Et Du Vin* 1994, *28*, 153-160.

Figure 1

Electrophoretic profile of the PCR-amplified interdelta regions of 12 *Saccharomyces cerevisiae* strains. Amplification was performed using primers δ12 and δ2, and PCR products were analyzed in the Caliper LabChip® 90 Electrophoresis System. The darker bands at 15 and 7000 bp represent co-injected internal markers.

Figure 2

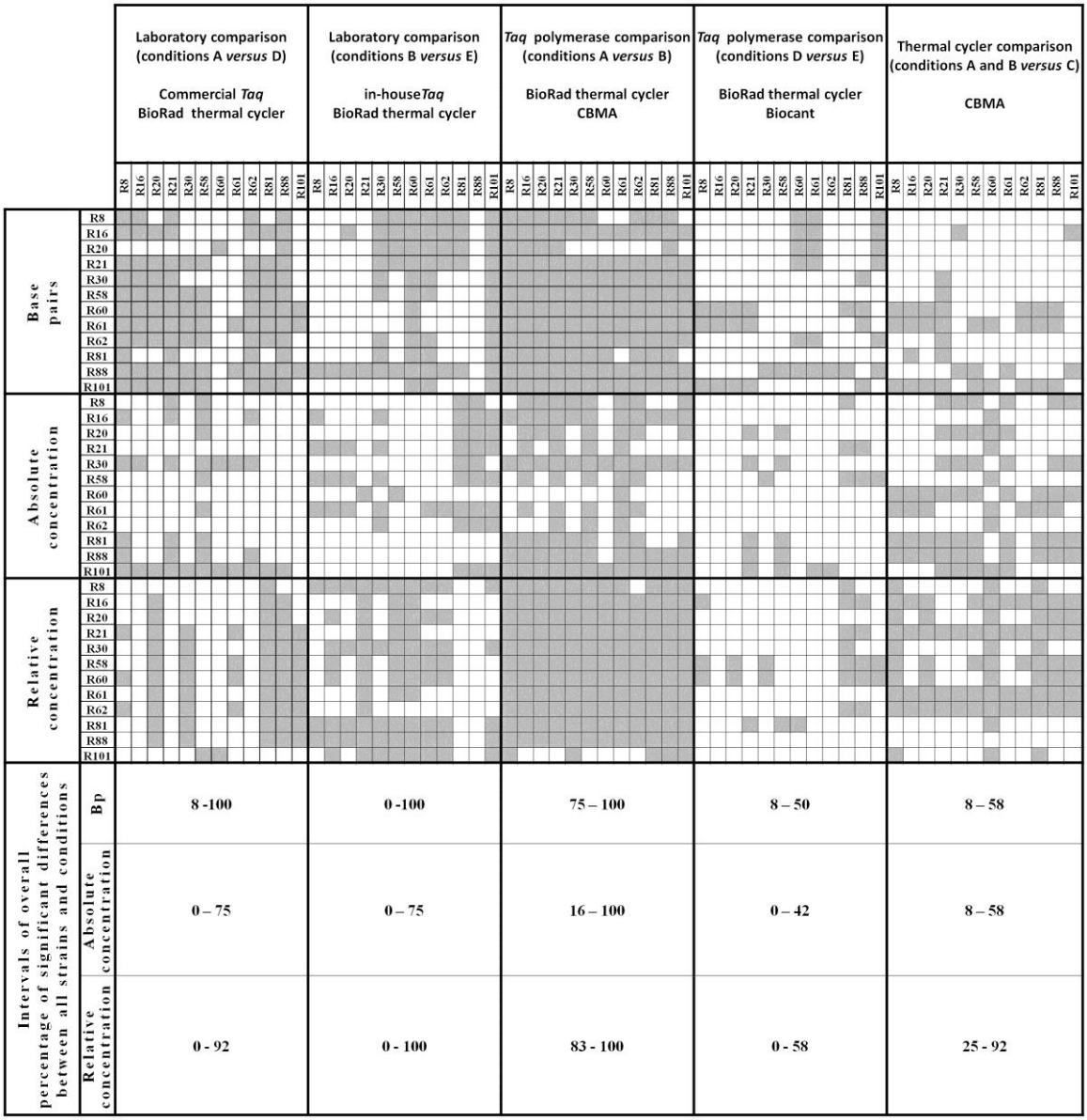
Replicates of the interdelta banding patterns of *S. cerevisiae* strain R81, obtained under different amplification conditions. **A** - Commercial *Taq*, BioRad thermal cycler, laboratory A; **B** - In-house *Taq*, BioRad thermal cycler, laboratory A; **C** - In-house *Taq*, Eppendorff thermal cycler, laboratory A; **D** - Commercial *Taq*, BioRad thermal cycler, laboratory B; **E** - In-house *Taq*, BioRad thermal cycler, laboratory B.

Figure 3

Comparison between the tested conditions for the delimitation of 12 yeast strains, regarding fragment sizes (in bp), absolute and relative DNA concentration values.

Percentages indicate the differences found between strains when performing statistical analysis of the differences between group medians considering each experimental condition: **A** - Commercial *Taq*, BioRad thermal cycler, laboratory A; **B** - In-house *Taq*, BioRad thermal cycler, laboratory A; **C** - In-house *Taq*, Eppendorff thermal cycler, laboratory A; **D** - Commercial *Taq*, BioRad thermal cycler, laboratory B; **E** - In-house *Taq*, BioRad thermal cycler, laboratory B.

Table 1: Comparison between experimental conditions (enzymes, thermal cyclers and laboratories) for each strain, based on the fragment sizes (bp), absolute and relative DNA concentration of each band of each strain, using Multiple Pairwise Testing based on a *t*-student distribution. Colored squares represent statistical significant differences. Each square is associated with one pair of comparisons.



562 **Table 2:** Fragment sizes (bp, average value and standard deviation) that were present in all 32 replicates of each strain

563

Average size (bp) of reproducible fragments	Strains										
	R8	**R16**	**R20**	**R21**	**R30**	**R58**	**R61**	**R62**	**R81**	**R88**	**R101**
97	97 ± 2,1	96 ± 2,4		96 ± 2,1	96 ± 2,1	96 ± 2,2	96 ± 1,9	96 ± 2	96 ± 2,1	96 ± 1,9	107 ± 1,8
134		134 ± 2								134 ± 1,9	
161			156 ± 1,7				167 ± 2	157 ± 1,3			162 ± 3
188							189 ± 2,1				186 ± 1,3
205						205 ± 1,7					
231			232 ± 2						231 ± 1,5	231 ± 4,4	
262	262 ± 2,1										
285					285 ± 2						
320									326 ± 3,5		314 ± 4
348	348 ± 8,7				349 ± 4,5			347 ± 4,4			346 ± 4,4
371									371 ± 3,7		
425	425 ± 4	425 ± 7	427 ± 5,7	427 ± 3,5	424 ± 3,7	427 ± 3,9		423 ± 3,4	426 ± 3,2		421 ± 4,8
458			453 ± 6,2	462 ± 3,5							
486					482 ± 5,8				489 ± 5,3		
531						531 ± 13,2					
680									680 ± 8,7		
721			721 ± 18,5								
899	899 ± 15,6										

564

Figure1

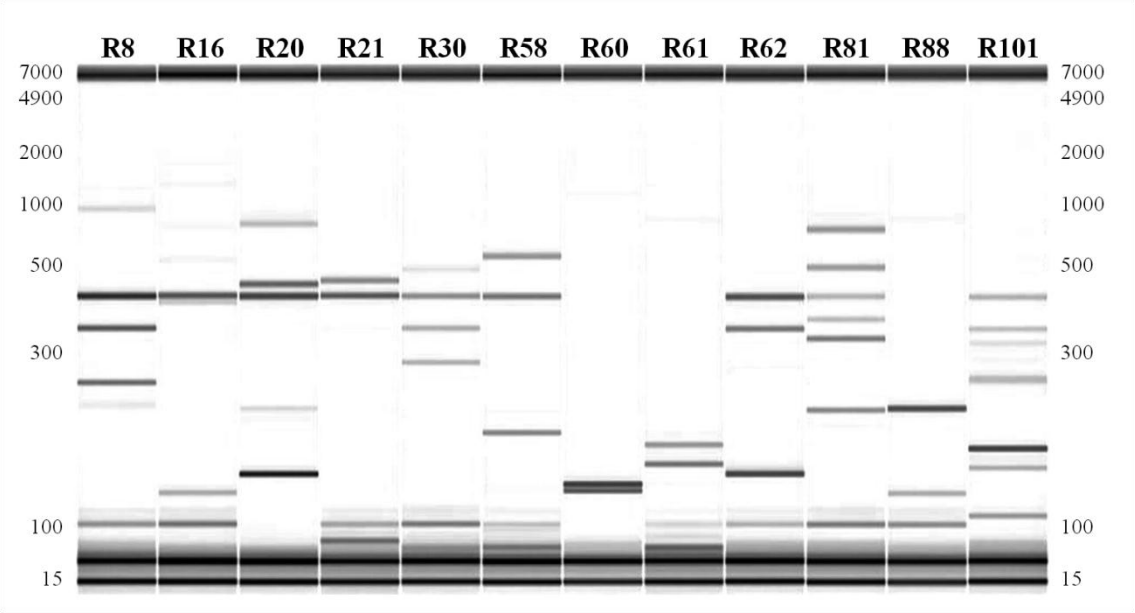
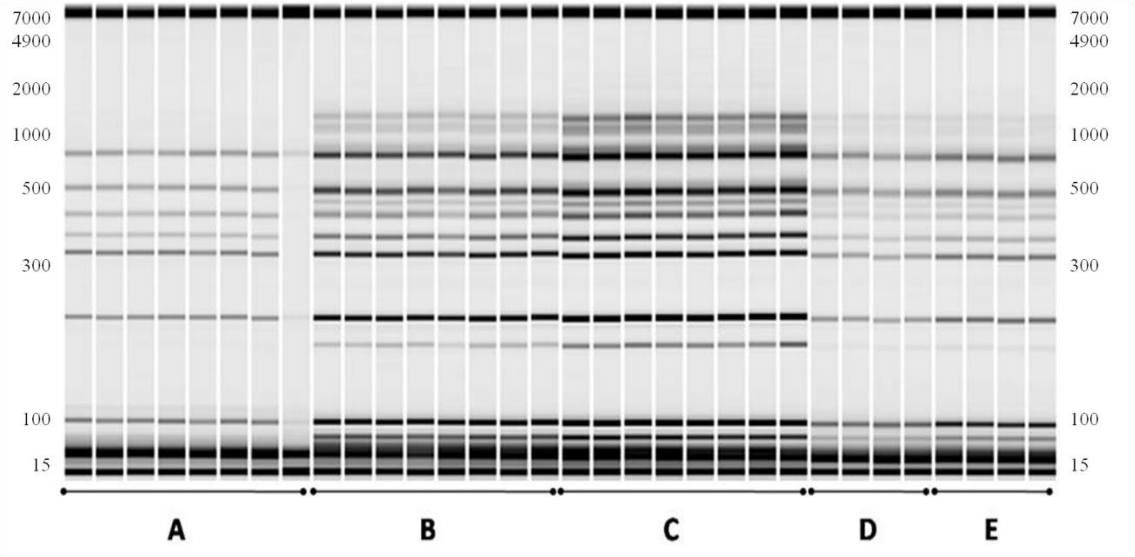
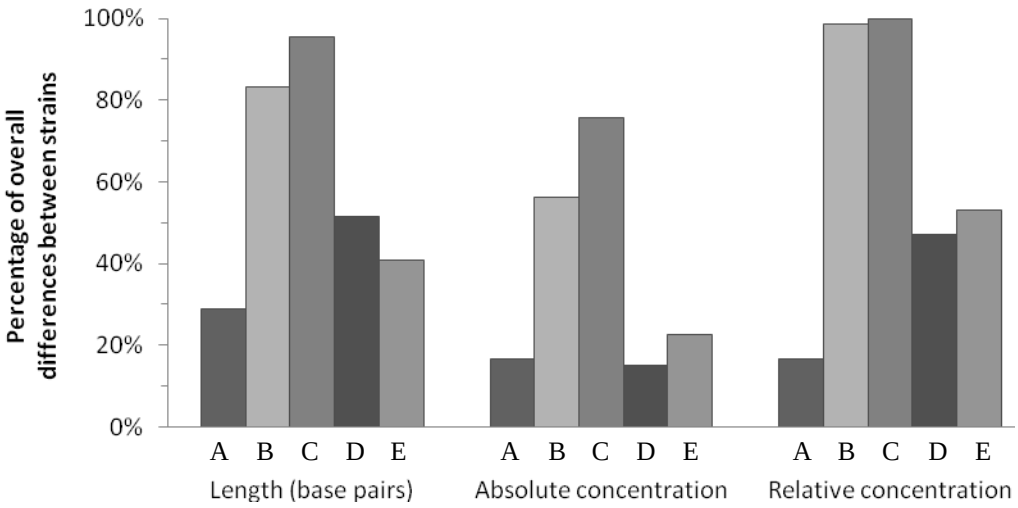

R8
R16
R20
R21
R30
R58
R60
R61
R62
R81
R88
R101
7000
4900
2000
1000
500
300
100
15

Figure 2



588 **Figure 3**



589

590

591

592

593

594

595

596