1   **Genotyping of *Saccharomyces cerevisiae* strains by interdelta sequence typing**

2   **using automated microfluidics**

3   Ricardo Franco-Duarte[1], Inês Mendes[1] Ana Catarina Gomes[2], Manuel A.S. Santos[2,3],

4   Bruno de Sousa[4], Dorit Schuller[1]

5

6   [1] CBMA (Centre of Molecular and Environmental Biology) / Department of Biology /

7   University of Minho, Braga, Portugal

8   [2] BIOCANT - Biotechnology Innovation Center, Cantanhede, Portugal

9   [3] RNA Biology Laboratory, CESAM, Biology Department, Aveiro University, Campus

10   Universitário de Santiago, Aveiro

11   [4] Centre for Malaria & Tropical Diseases Associated Laboratory, Instituto de Higiene e

12   Medicina Tropical, Universidade Nova de Lisboa, Portugal.

13

14   *For correspondence:

15   Dorit Schuller

16   e-mail: dschuller@bio.uminho.pt

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

18   Fax: (+351) 253 678 980

19

26 **Abstract**

27 Amplification of genomic sequences flanked by delta elements of retrotransposons TY1

28 and TY2 is a reliable method for characterization of *Saccharomyces cerevisiae* strains.

29 The aim of this study is to evaluate the usefulness of microfluidic electrophoresis

30 (Caliper LabChip[®]) to assess the factors that affect interlaboratory reproducibility of

31 interdelta sequence typing for *S. cerevisiae* strain delimitation. We carried out

32 experiments in two laboratories, using varying combinations of *Taq* DNA polymerases

33 and thermal cyclers. The reproducibility of the technique is evaluated using non-

34 parametric statistical tests and we show that the source of *Taq* DNA polymerase and

35 technical differences between laboratories have the highest impact on reproducibility,

36 whereas thermal cyclers have little impact. We also show that the comparative analysis

37 of interdelta patterns is more reliable when fragment sizes are compared, than when

38 absolute and relative DNA concentrations of each band are considered. Interdelta

39 analysis based on a smaller fraction of bands with intermediate sizes between 100 and

40 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.

66   More recently, the interdelta method was improved by the use of alternative primers

67   (δ12 and δ21) [17] that bind close to the initially described binding sites for primers δ1

68   and δ2 [18]. The combined use of these improved primer combinations (δ12 / δ21 or

69   δ12 / δ2) revealed greater banding pattern polymorphism and improved discriminatory

70   power [13]. The use of primer pairs δ12 / δ2 showed the same discriminatory power of

71   other methods for strain delimitation, such as mtDNA RFLP, microsatellite analysis and

72   karyotyping [19]. However, this method requires careful standardization of DNA

73   concentration [20]. Occasional non-reproducible "ghost bands" are present due to the

74   low annealing temperature (43 ºC), which is a disadvantage of the interdelta method.

75   Increasing the annealing temperature to 55 ºC reduced ghost bands, but lead to poorer

76   banding pattern and consequently reduced discriminatory power [21]. In summary, PCR

77   profiling analysis of delta sequences is associated with good discriminatory power for

78   the analysis of commercial strains [22], but the use of this typing method for routine

79   analysis of yeast strains requires careful evaluation [21, 23-26]. It is therefore advisable

80   to use additional methods such as mtDNA RFLP or microsatellite analysis to confirm

81   ambiguous results.

82   Fluorescent primers and automated DNA sequencers improve significantly banding

83   patterns containing weakly amplified fragments [27], decreasing experimental error and

84   increasing data throughput, scoring and reliability [28]. When interdelta sequences are

85   amplified with fluorescent primers, followed by capillary electrophoresis, the resolution

86   of the obtained profiles is considerably increased in comparison with standard agarose

87   gel electrophoresis [29].

88   The efficiency of PCR amplification is affected by numerous factors namely annealing

89   temperature, the concentration of $MgCl_2$, primers and template DNA. Even slight

90   variations in these parameters may affect results compromising data comparisons and

91   sharing between experiments and laboratories [30]. The optimal reaction conditions

92   need to be optimized for each PCR application.

93   Microfluidics are gaining notoriety across broads research fields, e.g., forensics, clinical

94   and genetic analysis [31-33]. Miniaturized reactions economize DNA samples, reagents

95   and analytical time considerably, and increase sensitivity, throughput and automation

96   possibilities [34, 35]. In the microfluidic chips for DNA analysis of the Caliper´s

97   LabChip® system, DNA samples are electroosmotically transported and fragmented

98   inside the chip, separated by capillary electrophoresis and finally analyzed using

99   fluorescence detection [36].

100 Genome-wide studies of yeast inter-strain variability require bio-databanks for

101 biodiversity conservation, sustainable development of genetic resources and equitable

102 sharing of genotypic data among laboratories. We consider interdelta sequences

103 amplification as a very useful method for high-throughput characterization of *S.*

104 *cerevisiae* strains, which is easy to perform, cheap and rapid in comparison to other

105 molecular methods. The aim of this study is to evaluate the impact of two different *Taq*

106 polymerases on the interlaboratory reproducibility of interdelta sequence typing for

107 yeast strain delimitation using microfluidics electrophoresis (Caliper´s LabChip®).

108 Besides, we also evaluate the impact of different thermal cyclers on the patterns

109 obtained. The study demonstrates that the reproducibility of the technique is most

110 affected by the source of *Taq* DNA polymerase and technical differences between

111 laboratories such as different operators. Interlaboratory reproducibility is highest when

112 fragment sizes between 100 and 1000 bp are compared, rather than absolute and relative

113 DNA concentrations of each band.

114

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.

141 An experimental strategy was devised to study the reproducibility of the interdelta

142 sequence amplification as a typing method for yeast strains using 96-well PCR plates

143 and the following combinations of *Taq* DNA polymerase, thermal cyclers and

144 laboratories: Plate 1 - commercial *Taq* (MBI Fermentas recombinant *Taq,* Ref.

145 EP0402), BioRad MyCycler thermal cycler, laboratory 1 (8 replicates per strain); Plate

146 2 - in-house cloned and produced *Taq*, BioRad MyCycler thermal cycler, laboratory 1

147 (8 replicates per strain); Plate 3 - in-house cloned and produced *Taq*, Eppendorff

148 Mastercycler thermal cycler, laboratory 1 (8 replicates per strain); Plate 4 - commercial

149 *Taq* (MBI Fermentas recombinant *Taq* Ref. EP0402) or in-house cloned and produced

150 *Taq* (4 replicates per strain), BioRad MyCycler thermal cycler, laboratory2. This

151 approach resulted in 32 replicates for each strain and a total of 384 electrophoretic

152 banding patterns. The four microplates thus included the following conditions to be

153 compared: A - Commercial *Taq*, BioRad thermal cycler, laboratory 1; B - In-house *Taq*,

154 BioRad thermal cycler, laboratory 1; C - In-house *Taq*, Eppendorff thermal cycler,

155 laboratory 1; D - Commercial *Taq*, BioRad thermal cycler, laboratory 2; E - In-house

156 *Taq*, BioRad thermal cycler, laboratory 2. Both laboratories used the same DNA

157 samples and the same in-house cloned and commercial *Taq* enzymes. Amplifications

158 were carried out with the same PCR buffer (MBI Fermentas, Ref. B33). PCR products

159 were analyzed using a high-throughput automated microfluidic electrophoresis system

160 (Caliper LabChip[®] 90 Electrophoresis System) and a 96-well plate format, according to

161 the manufacturer's instructions. The tolerance of the sizing resolution for this system is

162 ±15% (from 25 to 100 bp), ±10% (from 100 to 150 bp), ±5% (from 150 to 700 bp),

163 ±10% (from 700 to 1000 bp).

164

165 *Statistical analysis of electrophoretic data*

166   The size (bp) and concentration (ng of DNA) of each band was determined using the

167   LabChip® HT software (version 2.6) and exported to the software SPSS 18.0 package

168   for the composition of a matrix containing data for each band of the 32 replicates

169   banding patterns from each strain. Each band was analyzed and compared in terms of

170   fragment sizes (bp), absolute DNA concentration (ng/μl) and relative DNA

171   concentrations (%) (absolute concentration value was divided by the sum of all

172   concentration values of all bands contained in a replicate banding pattern). An

173   exploratory data analysis was performed, where normality distribution (Kolmogorov-

174   Smirnov and Shapiro-Wilk tests) and variance homogeneity (Levene's test) were tested

175   using SPSS 18.0. After several unsuccessful transformations of the data, non-parametric

176   tests were performed, such as "Kruskall-Wallis one-way analysis of variance" test, to

177   check for the equality of treatment medians among the different groups. More precisely,

178   the null hypothesis (H_0) assuming equality of all medians was tested against the

179   alternative hypothesis (H_1), which assumes that at least two of the strains show

180   differences in their medians, as outlined below:

181

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

183   where θ_i represents the median concentration (or percentage of concentration) for the ith

184   strain, $i=1,\ldots,12$.

185

186   In cases where the test produced statistical significant differences between strains,

187   multiple pairwise comparisons were performed to trace the origin of such differences.

188   The method proposed by Conover and Iman [39] searches for comparative magnitudes

189   of the means based on the rank data, and assumes the t-student distribution. The test is

190   based on the following expression:

191

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

192

193

194   with $t_{1-(\alpha/2)}$ the (1-α/2) quantile of a t-student distribution with (N-k) degrees of freedom,

195   k the number of groups, H_c the value for the test statistic of the Kruskal-Wallis test

196   corrected for ties and S^2 the corresponding variance.

197 **Results**

198 *Electrophoretic profile of the* **Saccharomyces cerevisiae** *strains*

199 Interdelta fragments of 12 genetically heterogeneous strains were amplified, using

200 primers δ12 and δ2 and were analyzed using automated microfluidics electrophoresis

201 (Caliper LabChip® 90 Electrophoresis System). In order to evaluate the inter-

202 laboratorial reproducibility of the banding patterns and to determine which combination

203 of *Taq* DNA polymerase and thermal cycler produced the most reproducible banding

204 patterns between both laboratories, the experimental design included different

205 combinations of the mentioned factors, as described in the Materials and Methods

206 section. Unique banding patterns were obtained for each strain (Figure 1). The most

207 common band was present in 9 out of the 12 strains and had a size of approximately 400

208 bp. Quantitative and qualitative analysis of each band was performed using the software

209 package of the electrophoresis system, using the values of the co-injected internal

210 markers (gel bands at 15 and 7000 bp) as a reference. The analysis presented herein is

211 based on the length of the amplified fragments (bp), and the absolute and relative (%)

212 values of DNA concentration (ng/μL) of each band, as outlined in the Material and

213 Methods section.

214 Figure 2 shows an example of 32 replicate banding patterns of a representative strain

215 tested under the conditions indicated in the first paragraph of Material and Methods

216 section. Fragment sizes showed high reproducibility between replicates of the same

217 condition and between conditions. Considerable differences were observed when, for

218 each experimental condition, DNA concentrations were compared. The most intense

219 banding patterns were obtained in laboratory 1, using in-house cloned and produced *Taq*

220 and the Eppendorff thermal cycler (condition C), followed by condition B and A. The

221 in-house produced *Taq* polymerase (C) amplified PCR products more efficiently than

222 commercial *Taq* (B). This agrees with the slightly stronger banding patterns of

223 condition E compared to condition D in laboratory 2. These trends were similar for the

224 other eleven strains (data not shown). One of eight replicates of condition A

225 (corresponding to the 8th lane of Figure 2) failed amplification for most strains due to

226 lateral evaporation of the PCR reaction mixture during cycling in the 96-well plates.

227 These replicates were excluded from further analysis.

228

229 ***Reproducibility of PCR-based interdelta typing***

230 Our main goal in this study was to identify statistically significant differences between

231 the banding patterns of yeast strains, generated under conditions A-E (see above), to

232 enhance reproducibility of interdelta sequence analysis between laboratories. In the first

233 step of the statistical analysis the data was verified for normality between the 12 strains

234 and the corresponding homogeneity of variances. Kolmogorov-Smirnov and Shapiro-

235 Wilk tests were used to investigate the normality assumption. The results (data not

236 shown) revealed that our data did not follow a normal distribution since all *p*-values

237 were approximately zero (<0.001) and, therefore, smaller than any of the usual levels of

238 significance considered (1 %, 5 % and 10 %). Homogeneity of variances between

239 strains was tested using Levene's test. This condition was also not satisfied by the data

240 (data not shown), as *p*-values were approximately zero (<0.001) for both variables in the

241 study. In an attempt to satisfy both normality and homogeneity of variances, data were

242 transformed using logarithm of base 2 and inverse values of absolute or relative

243 concentrations. New variables were created in SPSS, both for absolute and relative

244 values. Once again, the normality and homogeneity of variance assumptions were

245 rejected (data not shown), which lead us to use non-parametric tests.

246 The Kruskal-Wallis one-way analysis of variance was used to test equality of medians

247 among the groups of strains corresponding to each of the previously mentioned

11

248   condition (A-E) , using the formula (1) shown in the Material and Methods section. The

249   median was the measure of centrality for this test. It was expected that, in case of

250   reproducibility, all strains should have similar results, meaning that the values of

251   concentration (absolute or relative) and of fragment sizes (bp) should not differ in terms

252   of the median values. However, the Kruskal-Wallis test rejected the equality of medians

253   between groups, because once again the p-values were approximately 0 (<0.001). The

254   following approach consisted in searching for differences in terms of the median values

255   of fragment sizes (bp) and concentration values (absolute and relative) between strains.

256   This approach was repeated for the distinct experimental conditions used (A-E) in order

257   to search for the factors that most affect the reproducibility of the technique among the

258   conditions A-E. Based on the results from the Kruskall-Wallis one-way analysis of

259   variance, we assumed that at least two strains showed a difference in the medians. In

260   order to identify the strains that lead to the rejection of the equality of the medians,

261   Multiple Pairwise Comparisons, pooling the data for all 32 replicates per strain, were

262   performed. All 3892 values (the total number of observations regarding all experiments,

263   i.e. all bands of the 32 replicates of the 12 strains), were ordered by increasing numbers

264   and a rank score was calculated for identical values of absolute and relative

265   concentrations. Then, the formula (2) shown in the Material and Methods section was

266   applied for pairwise strain comparisons, based on a t-student distribution to search for

267   the origins of the differences between experimental conditions. The results of this test

268   are summarized in Table 1, for each pair compared, for each strain and using the

269   fragment size (bp), as well as absolute and relative DNA concentration values.

270   Statistical significant differences were observed when comparing all 3892 records

271   against each other, being the significant ones (based on a t-student significance test)

272   represented with gray squares in Table 1. In the bottom part of this Table (last three

12

273   lines), overall percentages are represented considering the differences between strains

274   and between conditions, both for fragment size base pairs and absolute and relative

275   DNA concentrations values. The inter-laboratory banding patterns reproducibility was

276   rather low as observed by the distribution of gray squares in the corresponding main

277   columns. Significant differences were found between strains analyzed in the two

278   laboratories. The lack of reproducibility of these experiments between laboratories was

279   not visible when analyzing the intervals of overall percentages. One could see that these

280   intervals were very comprehensive (including 0 and 100%) and that this analysis was

281   inconclusive for these comparisons. The reasons for this could be due to strain specific

282   effects and also to the extreme values included in the statistical. For example, strain

283   R101 was associated with 0% of statistically significant differences regarding absolute

284   DNA concentration, while for strain R88, regarding fragment size 100% of significant

285   differences were obtained. The cloned and in-house produced *Taq* increased

286   reproducibility between laboratories relative to commercial *Taq*. The comparison

287   between *Taq* polymerases produced data heterogeneity between laboratories. Low and

288   high reproducibility was found between enzymes for laboratory 1 and 2, respectively

289   (columns 3 and 4). This was shown by the higher number of gray squares in column 3

290   in comparison to column 4, and also by the intervals of overall percentages of

291   significant differences (75-100% comparing to 8-50% regarding fragment length; 16-

292   100% comparing to 0-42% regarding absolute concentration values; 83-100% in

293   comparison to 0-58% regarding relative concentration values).

294   Regarding the different thermal cyclers used, experimental variation in laboratory 2 lead

295   to more reproducible results, as shown by the comparison of fragment sizes. This

296   reproducibility was not so evident when comparing absolute and relative concentration

297   values.

13

298   When analyzing all conditions together, the comparison of absolute DNA concentration

299   values produced the most reproducible results, followed by fragment size and relative

300   DNA concentration values. Relative concentration values should not be used, however,

301   because in replicate analysis of strains under different experimental conditions, distinct

302   numbers of fragments were obtained, affecting the ratios of relative concentration.

303

304   *Comparison of different experimental conditions for strains delimitation*

305   To identify the experimental condition that best differentiate the 12 yeast strains,

306   statistical analysis of the differences between group medians for each experimental

307   condition was performed. For each experimental condition (from A to E), the

308   percentage of significant differences between the strains was calculated (excluding the

309   comparisons between the same strain for each experimental condition). Figure 3 shows

310   that combination C (in-house cloned *Taq*, Eppendorff thermal cycler, laboratory 2) lead

311   to the highest percentages regarding size, absolute and relative DNA concentration

312   values. This suggests that this is the most suitable combination of experimental

313   conditions for strain delimitation using interdelta banding patterns. Regarding fragment

314   size and relative DNA concentration, these percentages were almost 100, meaning that

315   the 12 electrophoretic patterns would correspond to 12 different strains. On the contrary,

316   combinations A (Commercial *Taq*, BioRad thermal cycler, laboratory 2), D

317   (Commercial *Taq*, BioRad thermal cycler, laboratory 1), and E (in-house *Taq*, BioRad

318   thermal cycler, laboratory 1) were less capable of differentiating strains with only 28.79

319   %, 51.52 % and 40.91 % of correctly delimited strains regarding fragment sizes,

320   respectively. Similar results were observed when comparisons were performed based on

321   absolute and relative DNA concentrations. In general terms, the use of in-house cloned

322   *Taq* polymerase led to better results than the use of commercial *Taq* polymerase, as can

323   be observed when comparing combination A and D (commercial *Taq*) with

14

324 combinations B, C and E (in-house *Taq*). Regarding the laboratories where the PCR

325 reactions were carried out, the strain patterns in laboratory 2 were better separated than

326 those obtained in laboratory 1 (combinations A, B and C versus combinations D and E).

327 The best results regarding strains differentiation were obtained when using relative

328 DNA concentration values (100 % with combinations B and C), however the latter

329 produced biased results. This is explained by the fact that, to calculate the relative DNA

330 concentration values, the absolute values were divided by the sum of all concentration

331 values of all bands contained in a banding pattern. In replicate analysis of different

332 experimental conditions, distinct numbers of fragments were obtained affecting the

333 ratios of relative concentration, leading to overestimated strain delimitation. Due to this

334 we consider that the percentages obtained for the analysis of absolute DNA

335 concentrations are more realistic to delimitate strains than relative DNA concentration

336 value. Fragment length analysis is the preferable measure for typing of yeast strains

337 using interdelta fragments amplification, even though the reproducibility associated was

338 smaller compared to absolute values of concentration (Table 1), but producing more

339 consistent results without introducing biases in the reproducibility of the technique.

340

341 ***Determination of identical banding patterns for each strain in all conditions***

342 To gain further insight into the reproducibility of the interdelta sequence typing method,

343 we tried to identify for each strain the bands that were amplified across the A-E

344 experimental conditions. Strain R60, which showed a very different banding pattern was

345 excluded from this analysis. As shown in Table 2, three to seven bands in the range of

346 100 – 900 bp were apparent in all 32 replicates of each strain. The respective standard

347 deviations were rather low, ranging from 1.3 to 15.6 bp. Additional bands were mostly

348 found for fragment sizes between 1000 and 1500 bp or below 100 bp, and were not

349 represented because of lack of reproducibility. Some intermediate fragments were also

350   not included in Table 3 because they were represented only in some experimental

351   conditions. Reproducibility would approximate to 100%, if only the bands included in

352   Table 2 would be used for comparison of fragment sizes.

Discussion

354 The improved interdelta method [17] is suitable for the typing of yeast strains [19]. This

355 method is simple, rapid and less expensive than others, such as sequencing and

356 microsatellite amplification. Although less rigorous than other techniques as MLST or

357 microsatellite amplification, the PCR-based interdelta method is suitable for high-

358 throughput analysis of large strain collections using microfluidic electrophoresis. The

359 amplification of interdelta regions results in a mixture of differently sized-specific

360 fragments. As previously shown by BLAST analysis [17], the sequences of fragments

361 obtained by amplification with primers δ12 and δ21 matched the predicted interdelta

362 regions. We have designed an inter-laboratory approach to evaluate the performance

363 and the reproducibility of this method as a high-throughput typing approach for the

364 genetic characterization of yeast strains. The comparative approaches that we describe

365 herein can contribute to the constitution of bio-databanks for equitable sharing of

366 genotypic data among laboratories in the context of biodiversity conservation and

367 sustainable development of genetic resources. However, it is crucial to find a set of

368 parameters leading to most reproducible patterns between laboratories.

369 As outlined in the Materials and Methods section, interdelta sequences of 12 strains

370 were amplified, under varying conditions (*Taq* DNA polymerase, thermal cycler and

371 laboratory). Interdelta sequence typing showed the reproducibility necessary for

372 implementation as a typing method for multiple (4 or 8) replicates of one strain, under

373 identical experimental conditions. The use of the microfluidic LabChip® system greatly

374 contributed to achieve very precise data with a high resolution, as reported in previous

375 works [28, 29].

376 In general, DNA amplification depends on numerous factors such as the method of

377 DNA isolation, the concentrations of DNA, primers, $MgCl_2$, dNTPs, the *Taq*

378 polymerase and the annealing temperature. In the present work only one DNA

379 extraction was performed for each strain, and the same DNA was used by both

380 laboratories, being therefore no variable in our experiments. Our (unpublished) results

381 showed that the DNA extraction protocol used is the most appropriate and leads to

382 much better results than an extraction method using phenol. DNA quantification was

383 performed in the Nanodrop^{TM} system, which allowed unambiguous evaluation of the

384 DNA quality. In recent publications [17, 19, 23, 26, 29, 40], DNA concentration values

385 were in the range of 0.1 - 2.5 ng/µl (final concentration). Fernandez-Espinar (2001)

386 showed that the optimal DNA quantities ranged from 0.6 to 2.5 ng/µl (final

387 concentration). The highest number of bands was amplified using the concentration of

388 2.5 ng/µl, which is similar to the concentration used throughout this work (4 ng/µl). In

389 the publications mentioned above, optimal $MgCl_2$ concentrations ranged from 1.5 to 3.0

390 mM, whereas the primer and dNTP concentrations were in the range of 1 to 1.67 μM

391 and 200 to 400 μM, respectively. In our (unpublished) optimization approaches, we

392 found that more fragments were amplified when using 3.0 mM $MgCl_2$, 400 μM dNTPs

393 and 1.67 μM of each primer. We suppose that these higher concentrations of primers

394 and dNTPs are necessary to amplify a group of fragments, contrarily to a PCR reaction

395 where just one single band is amplified.

396 The main objective of the present work was to show the extent of variation due to

397 factors such as the DNA polymerase or the thermal cycler. A commercial *Taq* DNA

398 polymerase and an in-house cloned and produced *Taq* were used, and different

399 amplification patterns were found. In our (unpublished) optimization approaches several

400 commercial *Taq* enzymes were tested, whereas the *Taq* polymerase used in this study

401 revealed to be most suitable for interdelta amplification. The choice of the polymerase is

402 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

428   between laboratories. This study also demonstrates that, for reliable data sharing

429   between laboratories, comparative interdelta sequence analysis should be based on a

430   reduced number of bands that lead to reproducible banding pattern profiles.

431

439 **References**

440 [1] Aucott, J. N., Fayen, J., Grossnicklas, H., Morrissey, A*., et al.*, *Rev Infect Dis* 1990,
441 *12*, 406-411.
442 [2] Hazen, K. C., *Clin Microbiol Rev* 1995, *8*, 462-478.
443 [3] Blondin, B., Vezinhet, F., *Revue Française d' Oenologie* 1988, *28*, 7-11.
444 [4] Carle, G. F., Olson, M. V., *PNAS (USA)* 1985, *82*, 3756-3760.
445 [5] Dubordieu, D., Sokol, A., Zucca, J., Thalouarn, P*., et al.*, *Connais Vigne Vin* 1984,
446 *21*, 267-278.
447 [6] Lopez, V., Querol, A., Ramon, D., Fernandez-Espinar, M. T., *Int J Food Microbiol*
448 2001, *68*, 75-81.
449 [7] Querol, A., Barrio, E., Huerta, T., Ramon, D., *Appl Environ Microbiol* 1992, *58*,
450 2948-2953.
451 [8] Vezinhet, F., Blondin, B., Hallet, J. N., *Appl. Microbiol. Biotechnol.* 1990, *32*, 658–
452 671.
453 [9] Corte, L., Lattanzi, M., Buzzini, P., Bolano, A*., et al.*, *J Appl Microbiol* 2005, *99*,
454 609-617.
455 [10] Baleiras Couto, M. M., Eijsma, B., Hofstra, H., Huis in't Veld, J. H., van der
456 Vossen, J. M., *Appl Environ Microbiol* 1996, *62*, 41-46.
457 [11] Ayoub, M. J., Legras, J. L., Saliba, R., Gaillardin, C., *J Appl Microbiol* 2006, *100*,
458 699-711.
459 [12] Hennequin, C., Thierry, A., Richard, G. F., Lecointre, G*., et al.*, *J Clin Microbiol*
460 2001, *39*, 551-559.
461 [13] Legras, J. L., Ruh, O., Merdinoglu, D., Karst, F., *Int J Food Microbiol* 2005, *102*,
462 73-83.
463 [14] Perez, M. A., Gallego, F. J., Martinez, I., Hidalgo, P., *Lett Appl Microbiol* 2001,
464 *33*, 461-466.
465 [15] Martorell, P., Querol, A., Fernandez-Espinar, M. T., *Appl Environ Microbiol* 2005,
466 *71*, 6823-6830.
467 [16] Hierro, N., Esteve-Zarzoso, B., Gonzalez, A., Mas, A., Guillamon, J. M., *Appl*
468 *Environ Microbiol* 2006, *72*, 7148-7155.
469 [17] Legras, J. L., Karst, F., *FEMS Microbiol Lett* 2003, *221*, 249-255.
470 [18] Ness, F., Lavalee, F., Dubordieu, D., Aigle, M., Dulau, L., *J. Sci. Food Agric* 1993,
471 *62*, 89-94.
472 [19] Schuller, D., Valero, E., Dequin, S., Casal, M., *FEMS Microbiol Lett* 2004, *231*,
473 19-26.
474 [20] Fernandez-Espinar, M. T., Lopez, V., Ramon, D., Bartra, E., Querol, A., *Int J Food*
475 *Microbiol* 2001, *70*, 1-10.
476 [21] Ciani, M., Mannazzu, I., Marinangeli, P., Clementi, F., Martini, A., *Ant*
477 *Leeuwenhoek* 2004, *85*, 159-164.
478 [22] Lavallée, F., Salvas, Y., Lamy, S., Thomas, D. Y*., et al.*, *Am J Enol Viticult* 1994,
479 *45*, 86-91.
480 [23] Pramateftaki, P. V., Lanaridis, P., Typas, M. A., *J Appl Microbiol* 2000, *89*, 236-
481 248.
482 [24] Lopes, C. A., van Broock, M., Querol, A., Caballero, A. C., *J Appl Microbiol* 2002,
483 *93*, 608-615.
484 [25] Cappello, M. S., Bleve, G., Grieco, F., Dellaglio, F., Zacheo, G., *Journal of*
485 *Applied Microbiology* 2004, *97*, 1274–1280.
486 [26] Demuyter, C., Lollier, M., Legras, J. L., Le Jeune, C., *J Appl Microbiol* 2004, *97*,
487 1140-1148.

488 [27] Terefework, Z., Kaijalainen, S., Lindstrom, K., *J Biotechnol* 2001, *91*, 169-180.
489 [28] Papa, R., Troggio, M., Ajmone-Marsan, P., Nonnis Marzano, F., *J Anim Breed*
490 *Genet* 2005, *122*, 62-68.
491 [29] Tristezza, M., Gerardi, C., Logrieco, A., Grieco, F., *J Microbiol Methods* 2009, *78*,
492 286-291.
493 [30] Vilioen, G. J., Nel, L. H., Crowther, J. R., *Molecular Diagnostic PCR Handbook*,
494 Springer, P.O. Box 17, 3300 AA Dordrecht, The Netherlands. 2005.
495 [31] Tudos, A. J., Besselink, G. A. J., Schasfoort, R. B. M., *Lab on a Chip* 2001, *1*, 83-
496 95.
497 [32] Verpoorte, E., *Electrophoresis* 2002, *23*, 677-712.
498 [33] Ryley, J., Pereira-Smith, O. M., *Yeast* 2006, *23*, 1065-1073.
499 [34] Whitesides, G. M., *Nature* 2006, *442*, 368-373.
500 [35] Lion, N., Reymond, F., Girault, H. H., Rossier, J. S., *Current Opinion in*
501 *Biotechnology* 2004, *15*, 31-37.
502 [36] Mark, D., Haeberle, S., Roth, G., von Stetten, F., Zengerle, R., *Chem Soc Rev*
503 2010, *39*, 1153-1182.
504 [37] Schuller, D., Casal, M., *Ant Leeuwenhoek* 2007.
505 [38] Aires-de-Sousa, J., Aires-de-Sousa, L., *Bioinformatics* 2003, *19*, 30-36.
506 [39] Conover, W. J., Iman, R. L., *Technical Report, LA-7677-MS. Los Alamos Scientific*
507 *Laboratory* 1979.
508 [40] Fernandez-Gonzalez, M., Espinosa, J. C., Ubeda, J. F., Briones, A. I., *Syst Appl*
509 *Microbiol* 2001, *24*, 634-638.
510 [41] Masneuf, I., Dubourdieu, D., *Journal International Des Sciences De La Vigne Et*
511 *Du Vin* 1994, *28*, 153-160.
512
513

514 **Figure 1**

515 Electrophoretic profile of the PCR-amplified interdelta regions of 12 *Saccharomyces*

516 *cerevisiae* strains. Amplification was performed using primers δ12 and δ2, and PCR

517 products were analyzed in the Caliper LabChip® 90 Electrophoresis System. The darker

518 bands at 15 and 7000 bp represent co-injected internal markers.

519

520 **Figure 2**

521 Replicates of the interdelta banding patterns of *S. cerevisiae* strain R81, obtained under

522 different amplification conditions. **A** - Commercial *Taq*, BioRad thermal cycler,

523 laboratory A; **B** - In-house *Taq*, BioRad thermal cycler, laboratory A; **C** - In-house *Taq*,

524 Eppendorff thermal cycler, laboratory A; **D** - Commercial *Taq*, BioRad thermal cycler,

525 laboratory B; **E** - In-house *Taq*, BioRad thermal cycler, laboratory B.

526

527 **Figure 3**

528 Comparison between the tested conditions for the delimitation of 12 yeast strains,

529 regarding fragment sizes (in bp), absolute and relative DNA concentration values.

530 Percentages indicate the differences found between strains when performing statistical

531 analysis of the differences between group medians considering each experimental

532 condition: **A** - Commercial *Taq*, BioRad thermal cycler, laboratory A; **B** - In-house *Taq*,

533 BioRad thermal cycler, laboratory A; **C** - In-house *Taq*, Eppendorff thermal cycler,

534 laboratory A; **D** - Commercial *Taq*, BioRad thermal cycler, laboratory B; **E** - In-house

535 *Taq*, BioRad thermal cycler, laboratory B.

536

537 **Table 1**: Comparison between experimental conditions (enzymes, thermal cyclers and

538   laboratories) for each strain, based on the fragment sizes (bp), absolute and relative

539   DNA concentration of each band of each strain, using Multiple Pairwise Testing based

540   on a *t*-student distribution. Colored squares represent statistical significant differences.

541   Each square is associated with one pair of comparisons.

		Laboratory comparison (conditions A *versus* D) Commercial *Taq* BioRad thermal cycler	Laboratory comparison (conditions B *versus* E) in-house *Taq* BioRad thermal cycler	*Taq* polymerase comparison (conditions A *versus* B) BioRad thermal cycler CBMA	*Taq* polymerase comparison (conditions D *versus* E) BioRad thermal cycler Biocant	Thermal cycler comparison (conditions A and B *versus* C) CBMA
Intervals of overall percentage of significant differences between all strains and conditions	Bp	8 - 100	0 - 100	75 – 100	8 – 50	8 – 58
	Absolute concentration	0 – 75	0 – 75	16 – 100	0 – 42	8 – 58
	Relative concentration	0 - 92	0 - 100	83 - 100	0 - 58	25 - 92

The upper part of the table is a grid of pairwise comparisons (strains R8, R16, R20, R21, R30, R58, R60, R61, R62, R81, R88, R101) divided into three row groups — Base pairs, Absolute concentration, and Relative concentration — with colored (gray) squares representing statistically significant differences.

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

563

						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										

Average size (bp) of reproducible fragments

564

565 **Figure1**
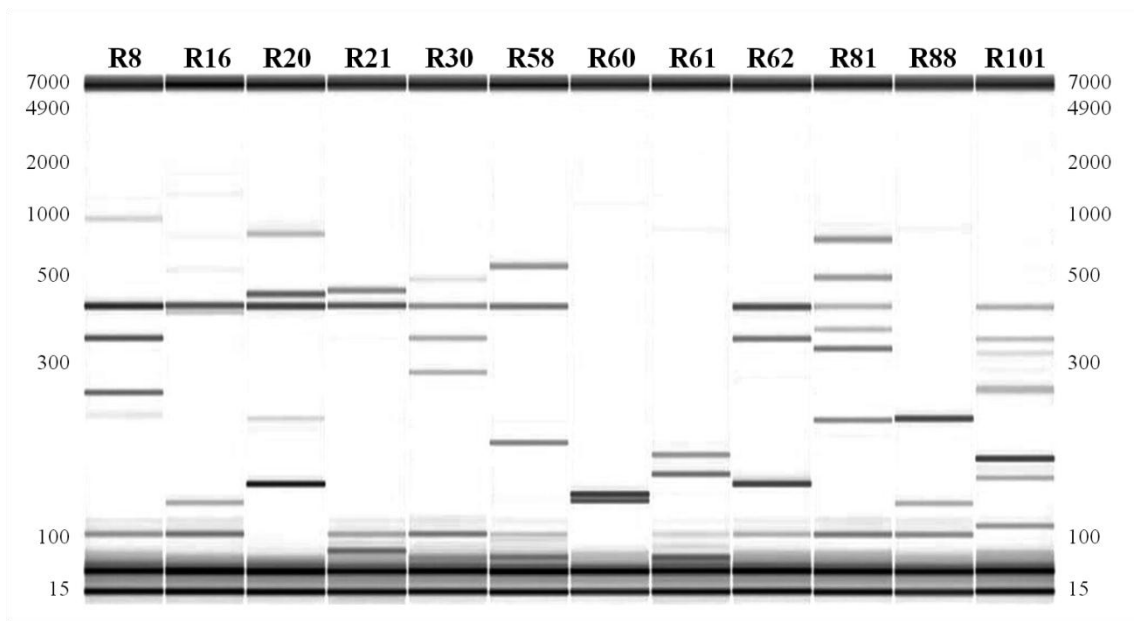
566

567

568

569

570

571

572

573
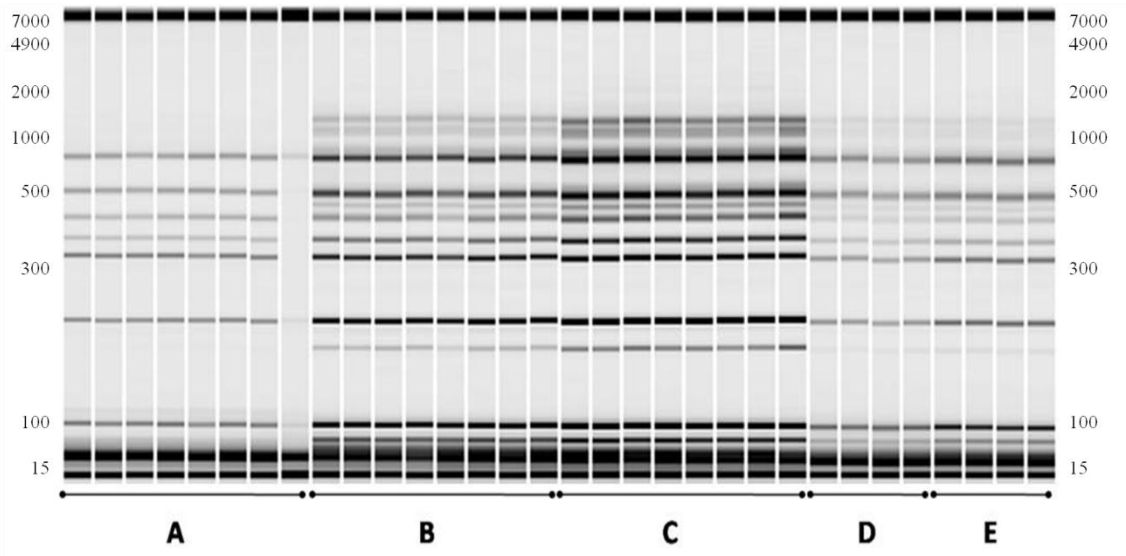
574

575

576

577

578
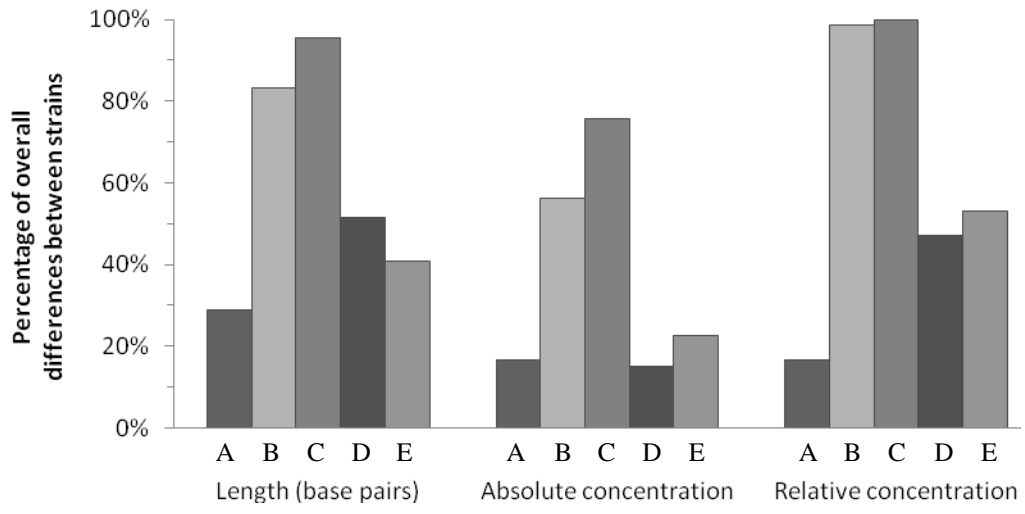
579   **Figure 2**

580

581

582

583

584

585

586

587

588   **Figure 3**

589

590

591

592

593

594

595

596