Dynamics of microbial contaminants is driven by selection during ethanol production

Luciano Lopes Queiroz¹, Maria Silveira Costa², Alcilene de Abreu Pereira², Marcelo de Paula Avila³, Patrícia Silva Costa³, Andréa Maria Amaral Nascimento³, Gustavo Augusto Lacorte^{1,2}

¹Department of Food Sciences and Experimental Nutrition, School of Pharmaceutical Sciences, University of São Paulo, São Paulo, Brazil;

²Molecular Biology Lab, Federal Institute of Minas Gerais, Bambuí Campus, Bambuí, MG, Brazil;

³Laboratório de Genética de Microrganismos, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil;

Abstract

Brazil is the second largest ethanol producer in the World and largest using sugarcane feedstock. Bacteria contamination is one the most important issues faced by ethanol producers that seek to increase production efficiency. Each step of production is a selection event due to the environmental and biological changes that occur. Therefore, we evaluated the influence of the selection arising from the ethanol production process on diversity and composition of Bacteria. Our objectives were to test two hypothesis, (1) that species richness will decrease during the production process and (2) that Lactic Acid Bacteria will become dominant with the advance of ethanol production due to selection. Bacterial community assemblage was accessed using 16S rRNA gene sequencing from 19 sequential samples. Temperature is of great importance in shaping microbial communities. Species richness increased between the Decanter and Must steps of the process. Low Simpson index values were recorded at the fermentation step, indicating a high dominance of *Lactobacillus*. Interactions between *Lactobacillus* and yeast may be impairing the efficiency of industrial ethanol production.

Keywords: ethanolic fermentation, sugarcane, distillery, selection, microbial ecology

Corresponding author: Gustavo Augusto Lacorte; e-mail: gustavo.lacorte@ifmg.edu.br; ORCID iD: 0000-0002-6866-348X; Phone number: +55 34 3431-5417.

30 1. Introduction

31 Ethanol is produced in industries worldwide by the microbial fermentation of the
32 sugars. The yeast *Saccharomyces cerevisiae* is the main species used in ethanol
33 industries, due to their ability of converting carbohydrates into ethanol, energy, cellular
34 biomass, CO₂, and other products [1]. Brazil is the second largest ethanol producer in
35 the world, and has sugarcane as the main feedstock [2]. Since 1970, the use of ethanol
36 as fuel for automotive vehicles has increased in Brazil, and decreased in 1990's due the
37 reduction of subsidies aimed at ethanol industries. A new increase was recorded in the
38 years 2000's with the advance of flex-fuel cars. Today most cars circulating in Brazil
39 are flex-fuel [3].

40 There are two types of distilleries used in ethanol production plants in Brazil, fed-batch
41 and continuous fermentation processes. At fed-batch fermentations, tanks are filled one
42 at a time, and tank management and cleaning occur individually, as discrete events.
43 However, in continuous fermentations all tanks are connected in a row and the process
44 occurs simultaneously. The sugarcane juice, after undergoing consecutive treatments
45 steps, is used as substrate for yeast growth and ethanol production [4]. In general,
46 ethanol production consists of cane preparation, juice extraction, juice clarification,
47 juice evaporation, must production, fermentation using yeasts and wine production. The
48 wine follows to other distillation steps and yeast recycling [1].

49 Maintaining the equipment used in ethanol production within the industry clean is very
50 important to avoid contamination by other microorganisms. Despite the caution in
51 avoiding contamination, it happens and is tolerated [4]. Sugarcane juice has its own
52 associated microbiome and, these microorganisms will be present in some steps of
53 ethanol production. In addition, the soil from sugarcane roots also an important
54 contamination source [1]. Microbial contaminant species may interfere directly in
55 ethanol production competing with the yeast for nutrients during the fermentation
56 process, producing components that may be toxic to the yeast (e.g., acetic and lactic
57 acid)[5], and reducing the carbohydrate consumption and its conversion to ethanol by
58 the yeast. Consequently, high costs in clean up the equipment arise, reducing the
59 productivity in ethanol production [4–6]. The main contaminant taxon is a
60 *Lactobacillus*, specifically a Lactic Acid Bacteria (LAB) found in high proportion
61 during the fermentation procedure [7, 8].

62 Microbial communities are shaped by four main process: selection, ecological drift,
63 dispersal and speciation [9] (or mutation [10]). Selection is the ability of species to
64 survive and reproduce in different habitats influenced by physical, chemical and biotic
65 factors. Ecological drift depicts the frequencies of species at a given location as a
66 consequence of demographic fluctuations changes. Dispersal is the ability of species to
67 reach and colonize a new location, and, speciation is the origin of new species by
68 speciation events [9, 10]. We believe that the microorganism community structure in the
69 ethanol production process will be determined by selection and ecological drift due the
70 spatial isolation among production steps, which hinder dispersal. In addition, the time

from one ethanol production stage to the next is too short for speciation to occur, even in organisms of high reproduction rates [11]. Therefore, each step of production may be considered a selection event due the environmental and biological changes inherent to each stage. In this study we investigated the influence of selection in diversity and composition of Bacteria during ethanol production process.

Our objectives in this study were test two hypothesis (1) species richness will decrease during the process and (2) Lactic Acid Bacteria will become dominant with advance of the ethanol production by selection. Bacterial community assemblage was accessed using 16S rRNA gene sequencing from 19 samples. Microbial contamination is of great importance in the ethanol plant factory. This study can shed light as to how these microbial communities interact and how species succession occurs with advancing of ethanol production.

2. Material and Methods

2.1. Sampling and ethanol production process

Samples were aseptically collected from each step of the ethanol production process in August 2016 at the Bambuí Bioenergia SA distillery localized in rural area of Bambuí, state of Minas Gerais. We collected one sample from each production step. All samples were stored at – 20 °C until further processing.

The consecutive steps of ethanol production (Figure 1) are sugarcane input, washing, and preparation, followed by sugarcane mill. A same sugarcane sample will pass through six sequential mills. The sugarcane juice is stored in different tanks in each mill, and is latter mixed in one tank. We collected seven samples in this stage: JuiceI, JuiceII, JuiceIII, JuiceIV, JuiceV, JuiceVI, and Mixed Juice. The juice was mixed from sample JuiceVI to Juice II, one by one, resulting JuiceI and a mixture of Juices (II – VI). JuiceI and the previous mixture of Juices (II – VI) were then mixed resulting in the Mixed Juice (Figure 1 – Mixed juice). After the milling step, mixed juice was sent to a heat exchanger (increasing temperature up to ≈50 °C), where liming was conducted. The liming process was carried out to regulate the pH of the solution to 6.2-6.6. The limed juice passed through two other heater exchange processes simultaneously, reaching the temperature of ≈ 105 °C. In this stage we collected the samples HeaterI and HeaterII.

The high-temperature juice was sent to the evaporation step, where the non-condensable gases were removed from the juice and acrylamide was added as it promotes the agglomeration of flakes, increases sedimentation speed, etc. These treatments improve the efficiency of the next step – decantation. We collected four decantation samples, two before the material entered the decanter, and two after (DecanterI input, DecanterII input, DecanterI output, and Decanter II output). After decantation, the product was sifted and filtered, and the must (juice treated and evaporated) was sent to the heat exchanger to reduce temperature from ≈ 85 °C to 33-35 °C, temperature needed for yeast growth in fermentation tanks. We collected samples before and after the must

111 passed through the heat exchanger (MustI input, MustI output and MustII output). We
112 were unable to collect must samples before it entered the heat exchanger II.

113 The low-temperature must was transferred to the first fermentation tank which was
114 filled with yeast solution to up to 25% of its volume capacity. The volume of the tanks
115 is then completed with must (3-4 hours), and the fermentation process takes place for
116 the next 4-5 hours. The must runs through three sequential fermentation tanks, each of
117 which were sampled: FermentationI, FermentationII and FermentationIII.

118 2.2. DNA extraction, library construction and sequencing

119 Total DNA was extracted from 250 µl of homogenized juice using an E.Z.N.A.® Soil
120 DNA Kit (Omega Bio-tek, Inc., Norcross, GA, USA) following the manufacturer's
121 instructions with adaptations. DNA was quantified by measuring the absorbance of the
122 sample at 260 nm using a NanoDrop ND-1000 spectrophotometer (Thermo Scientific,
123 Waltham, MA, EUA) and Qubit® dsDNA HS (High Sensitivity) Assay (Life
124 Technologies).

125 The primer set 341F (5'- CCTACGGGNGGCWGCAG-3') and 785R (5'-
126 GACTACHVGG-GTATCTAATCC-3') [12] was used to amplify the hypervariable
127 regions V3 and V4 of gene 16S. The Illumina adapter was used to build the 16S
128 sequence library following the protocol provided by Illumina (Illumina, 2013). PCR
129 mixtures contained 0.5 µM each primer, 0.7U of Taq DNA Polymerase (Life
130 Technologies, Carlsbad, CA., EUA), 1X Buffer, 4mM MgCl₂, 0.2 mM dNTP, 0.3 mg/ml
131 BSA (Bovine Serum Albumin) and 20 ng of the DNA template. 16S rDNA amplification
132 began with a denaturation time of 3 min at 95 °C, followed by 35 cycles of 45 s
133 denaturing at 95 °C, 1 min primer annealing at 57 °C, 45 s extension at 72 °C and final
134 extension of 10 min at 72 °C. Amplifications were conducted in a Thermal cycler
135 GeneAmp PCR System 9700 (Applied Biosystems, Foster City, CA). All 19 samples
136 were sequenced on the MiSeq platform (Illumina, Inc., USA), using 2 × 250 bp
137 MiSeq V2 reagent kit.

138 2.4. Bioinformatics

139 We performed a bioinformatics analysis using the Brazilian Microbiome Project (BMP)
140 pipeline [13]. BMP pipeline is a combination of the VSEARCH [14] and QIIME [15]
141 software. The software VSEARCH was used to remove barcodes and primer sequences
142 from a *fastq* file, filter sequences by length (fastq_truncLen 400) and quality
143 (fastq_maxEE 0.5), sort by abundance and remove singletons. After that, OTU was
144 clustered and chimeras were removed. Two files were generated through a *fastq* file
145 with filtered sequences and an OTU table in .txt. We assigned the taxonomy to OTU
146 using the *uclust* method in QIIME version 1.9.0 and the SILVA 16S Database (version
147 n128) as reference sequences [16]. The OTU table file was converted to BIOM and
148 taxonomy metadata was added. Sequences were aligned, filtered and the phylogenetic
149 tree constructed. Then, alpha and beta diversity were calculated using QIIME

150 commands. The nucleotide sequence data reported are available in the NCBI under
151 BioProject PRJNA508648.

152 2.5. Statistical analyses

153 The mean Observed richness, Shannon, Simpson and Equitability values were
154 compared among sampling groups using the Wilcoxon-Mann Whitney test [17]. A
155 heatmap was built with the 20 most abundant OTUs using Ward's hierarchical
156 clustering method (ward.d2) [18]. Moreover, we used a Principal Coordinates Analysis
157 (PCoA) to compare similarities among samples and tested differences using a
158 Permutational analysis of variance (PERMANOVA) [19]. All analyses were carried out
159 using the statistical software R [20] (*qiimer*, *ggplot2* [21], *phyloseq* [22], and *vegan* [23]
160 packages).

161 3. Results

162 3.1. Bacterial community diversity

163 A total of 51.047 sequences and 222 OTU_{0.03} were obtained after bioinformatics
164 analysis. The number of sequences in each sample had a high variability, ranging from
165 951 sequences in SugarV to 11.159 in MustII-output (mean=2.687±2.338). The last step
166 of the fermentation process had the highest richness (100 species), followed by MustI-
167 input (82), and FermentationI (67). SugarIV was the sample with lowest richness (37).
168 Shannon diversity index (H') showed values between 0.865 and 2.990. Simpson and
169 Equitability values were similar among samples, considering the similar assumptions of
170 the analysis. Simpson values ranged from 0.288 to 0.905 and Equitability from 0.238 to
171 0.679 (Table 1).

172 Species richness was consistent in the first three steps (Sugar, Heater, and Evaporation),
173 but we observed that richness increased in the Must step, followed by a low richness
174 increase during Fermentation. Richness of the Evaporation and Must groups differed
175 (Wilcoxon-Mann Whitney test, $p = 0.044$) (Figure 2). The diversity indices measured
176 (Shannon, Simpson and Equitability) exhibited similar results as regards abundance
177 distribution among samples and groups of ethanol production steps. Diversity index
178 values were high in the Sugar, Heater, Evaporation, and Must steps with mean values of
179 1.977±0.546 (Shannon), 0.711±0.182 (Simpson), and 0.506±0.123 (Equitability). When
180 must was transferred to the fermentation tanks, these values decreased (mean values
181 1.208±0.102, 0.423±0.033, and 0.287±0.411, respectively; Suppl. figures 1-3 for
182 diversity indices). These differences were tested by Wilcoxon-Mann Whitney test and
183 no significant difference was found.

184 3.2. Bacterial community composition

185 The community composition of all samples was comprised mostly Firmicutes (76.43%),
186 Proteobacteria (19.46%), and Cyanobacteria (3.71%) (Suppl. Figure 4). The Firmicutes
187 phylum was the most abundant in several samples and the sequences from this phylum
188 were in the Bacilli class (75.31%; Suppl. Figure 5). Over 90% of the sequences in all

189 Juice samples were Bacilli, except in JuiceI, where 53.58% of the sequences were
190 Bacilli, 21.83% were Alphaproteobacteria (Proteobacteria) and 13.56% of Chloroplast
191 (Cyanobacteria). Bacilli were highly dominant in Heater samples (on average 94.62% of
192 the sequences), and the mean was 81.45% in Fermentation samples. The relative
193 abundance for the Evaporation and Must samples was a similar to that recorded for
194 JuiceI.

195 The relative genera abundance showed a great variability among the ethanol production
196 steps (Figure 3). The production steps are linear starting with sugarcane mill and juice
197 production. The first juice obtained was JuiceI and comprised the genera *Weissella*
198 (27.3%), *Leuconostoc* (22.01%), and *Oryza mayeriana* (rice; 19.67%). The abundance
199 of *Lactobacillus* increased slightly from JuiceII to JuiceIV (11.65%), *Weissella* showed
200 a high increase to 79.27%, and *Leuconostoc* a high decrease to 0.55%. JuiceV had a
201 high abundance of *Lactobacillus* (55.31%), followed by *Weissella* (28.18%) and
202 *Streptococcus* (9.57%). JuiceVI had different proportions of these genera, specifically a
203 *Lactobacillus* reduction to 29.61%, *Weissella* reduction to 8.67% and *Streptococcus*
204 increase to 32.7%. The last step of juice production is mixing all juices in a same tank.
205 The Mixed Juice sample had a high abundance of *Streptococcus* (43.54%), followed by
206 *Lactobacillus* (37.34%) and a high decrease of *Weissella* abundance (0.37%). These
207 patterns may also be observed in the heatmap figure (Figure 4).

208 The next step of ethanol production is the Heater step, in which the Mixed Juice passes
209 through two heaters simultaneously (HeaterI and II). The *Lactobacillus* abundance in
210 the juice after it passed through HeaterI was 49.55%, followed by *Anoxybacillus* with
211 17.03% and *Streptococcus* with 12.63%. After HeaterII, *Lactobacillus* abundance
212 increased to 61.91%, *Streptococcus* to 24.95% and *Anoxybacillus* decreased to 6.41%.
213 The Juice passed through the heaters was sent to the decantation tank for the
214 evaporation step (two tanks). We collected samples from the evaporation step before
215 juice was directed to the tanks (Evaporation I input and Evaporation II input) and after
216 the end of the evaporation process (Evaporation I output and Evaporation II output).
217 Evaporation I input and Evaporation II input had a similar genera pattern, specifically,
218 *Anoxybacillus* was the most abundant (34.38% and 42.19%, respectively), followed by
219 *Oryza mayeriana* (19.64% and 28.86%, respectively). We observed a dominance of the
220 genus *Bacillus* after evaporation processes, comprising nearly 80% of the sequenced
221 genes Evaporation I output and 50% in Evaporation II output. *Acinetobacter* and
222 *Geobacillus* were also abundant in Evaporation II output sample (21.48 and 14.32%,
223 respectively).

224 The must produced after evaporation was sent to the heat exchanger to reduced
225 temperature to the reach optimum temperature for the yeast. The most abundant genus
226 in MustI input was *Acinetobacter* (31.44%), followed by *Lactobacillus* (17.43%) and
227 *Streptococcus* (4.06%). The pattern changed in MustI output, with an increased
228 *Acinetobacter* (46.43%) and *Streptococcus* abundance (24.09%) and decreased
229 *Lactobacillus* abundance (2.01%). MustII output is a mixture of MostI and II output,

230 and showed a high abundance of *Zymomonas* (37.34%), followed by *Lactobacillus*
231 (10.45%), *Streptococcus* (9.71%), and prominent reduction of *Acinetobacter* (0.95%).

232 The fermentation process occurs in three different and sequential tanks, FermentationI,
233 II and III. *Lactobacillus* was the most abundant genus in all fermentation steps,
234 specifically accounting for 80.39% of the sequences in FermentationI, 77.95% in
235 FermentationII, and 79.83% in FermentationIII, followed by *Acinetobacter* and
236 *Zymomonas*.

237 A Principal Coordinates Analysis (PCoA) was conducted with four distance matrices:
238 Bray-Cutis (Figure 5), Jaccard, and Weighted and Unweighted UniFrac (Suppl. Figure
239 6). The samples were sorted into five groups according to the step of production.
240 Microbial communities present in samples from the same group, when using Bray-
241 Curtis similarity matrix, were highly similar (PERMANOVA, $F = 4.667$, $R^2 = 0.55$, $p =$
242 0.001). We observed a high proximity between samples from Fermentation step (blue
243 dots, figure 5). Juice samples were grouped in left side of the ordination plot (near the -
244 0.2 position of Axis 1), and Must and Decanter samples were grouped in the center.

245 4. Discussion

246 4.1. Bacterial richness increase even after selection events

247 We observed that bacterial richness increased, instead decreasing as we expected in our
248 first hypothesis. In this study, each step in the ethanol production process was
249 considered a selection event for microbial interactions and/or environmental changes.
250 Bacterial species richness was constant during first three steps (Juice, Heater, and
251 Decanter). However, richness increased from the Decanter to the Must steps (Figure 2).
252 This result may be explained by a contamination with microorganisms after the
253 decantation process.

254 Industrial ethanol production is not a sterile process, microbial contamination is
255 common and identified as an important problem regarding ethanol production efficiency
256 [4, 24]. The increase of microbial richness after the decantation process, as observed in
257 this study, showed that selection events that occur before decantation may be efficient in
258 controlling microbial communities. The samples may become contaminated during the
259 transfer of decanted juice to must tanks, when the juice passes the rotary sieves, or
260 during the evaporation system.

261 The species richness in Must samples was constant during the Fermentation step.
262 However, evenness decreased in Fermentation samples, as we observed in the diversity
263 indices values. Relative abundance is altered by selection, such that the environmental
264 and biological condition may be ideal for the survival and reproduction of a species at a
265 given time, and not at another [10]. Diversity indices based on observed richness and
266 relative abundance, such as the Shannon, Simpson, and Equitability indexes, may be
267 used to understand certain species and their contribution in abundance [25, 26].

268 However, the differences in diversity index values between the Must and Fermentation
269 stages were not significant (Suppl. Figure 1-3).

270 4.2. Selection process shaping microbial community

271 The high variance of the relative abundance of OTUs among the ethanol production
272 steps may be influenced by temperature [4] and negative interactions with other taxon.
273 The importance of these two variables for selection may be noted when observing the
274 relative abundance of *Weissella*. *Weissella* exhibited a high abundance in Juice samples.
275 However, its abundance was lower in the juice mixture, possibly due to a negative
276 interaction with *Lactobacillus* and *Streptococcus*. This negative interaction may have
277 reduced *Weissella* while increasing *Lactobacillus* and *Streptococcus* abundance values.
278 The sequences rated as *Weissella* almost disappeared after the Heater steps (Figure 3). A
279 recent study found lower abundances of *Weissella* in fermentation tanks,[8] as recorded
280 in this study.

281 All of our sequences of *Weissella* were rated at species level as *Weissella confuse*.
282 *Weissella* species are non-spore and gram-positive coccobacilli [27]. *W. confuse* is
283 found in fermentation foods, but have also been associated to human diseases (sepsis
284 and bacteremia) as an opportunistic pathogen [28]. These bacteria are
285 heterofermentative, belong to the Lactic Acid Bacteria (LAB) group, some strain are
286 able to growth at 45 °C and consume several types of carbohydrates (Fructose,
287 Galactose, Maltose) [29]. *W. confuse* has been found in sugarcane samples [24, 30]. The
288 reduced abundance of *Weissella* may be explained by a low ability to uptake
289 carbohydrates from the environment when compared with other lactic acid bacteria,
290 such as *Lactobacillus* and *Streptococcus*.

291 On the next steps, temperature increased twice to approximately 50 °C and to 105 °C.
292 We observed that *Lactobacillus* and *Streptococcus* were the most abundant species in
293 Heater samples, even under high temperatures. The relative abundance of *Lactobacillus*
294 and *Streptococcus* decreased abruptly, while the abundance of *Anoxybacillus* and
295 *Geobacillus* increased, occupying the habitat when juice samples are stored in decanter
296 tanks, after undergoing the acrylamide treatment and being kept under constant
297 temperature at 105 °C (Decanter input samples).

298 *Anoxybacillus* and *Geobacillus* are gram-positive Bacillaceae family, and have the
299 morphology of rod or coccus [31, 32]. Both are moderately thermophilic, with optimal
300 growth temperature of 50-62 °C and 55-65 °C for *Anoxybacillus* and *Geobacillus*
301 respectively. These genera are found in several types of habitats, for example
302 geothermal springs, manure, and milk-processing plants. An interesting observation
303 about these two genera is their ability to form endospores. *Geobacillus* have a
304 widespread distribution in world, even in cold habitats [33]. *Anoxybacillus spp.* are
305 alkaliphilic or alkalitolerant, but are able to grow under neutral pH [34]. The high
306 abundance of *Anoxybacillus* and *Geobacillus* in the decantation step may be explained

307 by their capacity of tolerating high temperatures, unlike others species. Thus, they
308 dominated habitat due to the high temperature and absence of competitors.

309 The juice passed through rotary sieves and another evaporation system during the
310 transportation of the juice from Decanter to Must tanks. The temperature of the
311 decanted juice before entering Must tanks ranged from 85-90 °C, and its microbial
312 communities were dominated by *Bacillus*. This temperature reduction was enough to
313 create conditions for *Acinetobacter* growth, as confirmed on the next step with a high
314 increase of its relative abundance. *Acinetobacter* is gram-negative Protobacteria and has
315 been found in soil environments [35]. These bacteria may have come with sugarcane
316 roots, despite the washing steps.

317 When the temperature of the MustI output sample was reduced to 33-35 °C, the
318 abundance of *Acinetobacter* remained high, but abundance of *Streptococcus* began to
319 increase. This sample was mixed with MustII output, and presented a high increase of
320 *Zymomonas* abundance and decrease of *Acinetobacter* under mesophilic conditions. The
321 genus *Zymomonas* has only one species, *Zymomonas mobilis*. *Z. mobilis* is a gram-
322 negative Alphaproteobacteria, facultative anaerobic and rod-shaped, it can produce
323 bioethanol by the Entner–Doudoroff pathway [36]. This species has been studied for
324 industrial ethanol production, and it has advantages over *Saccharomyces cerevisiae*,
325 such as the high ethanol tolerance [37]. The high abundance at this moment may be
326 explained by optimal temperature conditions and the absence of other fermentative
327 microorganisms in high abundance. *Lactobacillus* became dominant and *Zymomonas*
328 abundance reduced to a stable proportion when Must was sent to fermentation tanks and
329 yeast was added.

330 We also observed a high similarity in the composition of the microbial communities
331 within samples from the same group (i.e., fermentation step; Figure 5). Samples from
332 Juice group had low within-group similarity values (yellow-mustard dots). However,
333 when Mixed Juice was submitted to the first step of juice preparation (HeaterI and II),
334 the communities present became highly similar (red dots). Therefore, communities were
335 selected by the increased temperature and pH regulation with liming. These results may
336 be explained by the different environmental variables and inter-specific interactions
337 present at each community, in each production step.

338 4.3. High abundance of *Lactobacillus* during fermentation

339 Our second hypothesis was that Lactic Acid Bacteria (LAB) would become with the
340 advance of ethanol production, due to selection. *Lactobacillus* was the most abundant
341 bacteria in the Fermentation stage. Other LAB, such as *Weissella*, showed a lower
342 abundance and *Leuconostoc* was not present at this ethanol production step. Several
343 studies have recorded *Lactobacillus* as dominant species during the fermentation of
344 sugarcane juice. Costa et al. (2015) reported that almost 100% of the sequences were of
345 *Lactobacillus* species, while Bonatelli et al. (2017) reported a prevalence of 92 to 99%.

The main selection factors present during the fermentation step were temperature (30-32 °C), high abundance of yeast competing for carbohydrates, and ethanol produced during carbohydrate fermentation [4, 7]. Ethanol is a toxic compound for most of the microorganisms present in sugarcane juice during fermentation [24, 38]. *Lactobacillus*, *Zymomonas*, *Acinetobacter*, *Enterococcus* and *Weissella* were resistant, but *Lactobacillus* was a better competitor in carbohydrate uptake than the others and became dominant because of its fast growth [39].

4.4. Conclusions and remarks

Microbial communities varied greatly among the ethanol production process, with main changes occurring due to the influence of temperature. Thus, temperature was identified as an important selection factor. Our first hypothesis was rejected, in contrast with our expectations, given that species richness increased between Decanter and Must samples. Our second hypothesis was corroborated by the compositional data, which show a high abundance of *Lactobacillus* sequences in the Fermentation step. However, no significant differences were recorded among the diversity indices of the production steps. The effort made to avoid contamination in ethanol industries and the identification of temperature as important selection factor, is not enough to avoid the prevalence of *Lactobacillus*, a major competitor with yeast for resources in this system.

5. Acknowledges

We would like to thank Bruno Spacek for critical reviews of the manuscript.

Funding information

This work was granted by Instituto Federal de Minas Gerais - Edital de Pesquisa Aplicada no. 156/2013.

6. References

1. Lopes ML, Paulillo SC de L, Godoy A, et al (2016) Ethanol production in Brazil: a bridge between science and industry. *Brazilian J Microbiol* 47:64–76. <https://doi.org/10.1016/j.bjm.2016.10.003>
2. Amorim H V., Lopes ML, De Castro Oliveira JV, et al (2011) Scientific challenges of bioethanol production in Brazil. *Appl Microbiol Biotechnol* 91:1267–1275. <https://doi.org/10.1007/s00253-011-3437-6>
3. Basso L, Basso T, Rocha S (2011) Ethanol production in Brazil: the industrial process and its impact on yeast fermentation. *Biofuel Prod - Recent Dev Prospect* 1530:85–100. <https://doi.org/10.5772/959>
4. Costa OYA, Souto BM, Tupinambá DD, et al (2015) Microbial diversity in sugarcane ethanol production in a Brazilian distillery using a culture-independent method. *J Ind Microbiol Biotechnol* 42:73–84. <https://doi.org/10.1007/s10295-014-1533-1>
5. Bischoff KM, Liu S, Leathers TD, et al (2009) Modeling bacterial contamination

- 384 of fuel ethanol fermentation. *Biotechnol Bioeng* 103:117–122.
385 <https://doi.org/10.1002/bit.22244>
- 386 6. Bayrock DP, Ingledew WM (2004) Inhibition of yeast by lactic acid bacteria in
387 continuous culture: Nutrient depletion and/or acid toxicity? *J Ind Microbiol*
388 *Biotechnol* 31:362–368. <https://doi.org/10.1007/s10295-004-0156-3>
- 389 7. Lucena BTL, Dos Santos BM, Moreira JLS, et al (2010) Diversity of lactic acid
390 bacteria of the bioethanol process. *BMC Microbiol* 10:.
391 <https://doi.org/10.1186/1471-2180-10-298>
- 392 8. Bonatelli ML, Quecine MC, Silva MS, Labate CA (2017) Characterization of the
393 contaminant bacterial communities in sugarcane first-generation industrial
394 ethanol production. *FEMS Microbiol Lett* 364:.
395 <https://doi.org/10.1093/femsle/fnx159>
- 396 9. Vellend M (2010) Conceptual Synthesis in Community Ecology. *Q Rev Biol*
397 85:183–206. <https://doi.org/10.1086/652373>
- 398 10. Hanson C a, Fuhrman J a, Horner-Devine MC, Martiny JBH (2012) Beyond
399 biogeographic patterns: processes shaping the microbial landscape. *Nat Rev*
400 *Microbiol* 10:497–506. <https://doi.org/10.1038/nrmicro2795>
- 401 11. Lenski RE (2017) Experimental evolution and the dynamics of adaptation and
402 genome evolution in microbial populations. *ISME J* 1–14.
403 <https://doi.org/10.1038/ismej.2017.69>
- 404 12. Klindworth A, Pruesse E, Schweer T, et al (2013) Evaluation of general 16S
405 ribosomal RNA gene PCR primers for classical and next-generation sequencing-
406 based diversity studies. *Nucleic Acids Res* 41:e1.
407 <https://doi.org/10.1093/nar/gks808>
- 408 13. Pylro VS, Roesch LFW, Morais DK, et al (2014) Data Analysis for 16S
409 Microbial Profiling from Different Benchtop Sequencing Platforms. *J Microbiol*
410 *Methods* 107:30–37. <https://doi.org/10.1016/j.mimet.2014.08.018>
- 411 14. Rognes T, Flouri T, Nichols B, et al (2016) VSEARCH: a versatile open source
412 tool for metagenomics. *PeerJ* 4:e2584. <https://doi.org/10.7717/peerj.2584>
- 413 15. Caporaso JG, Kuczynski J, Stombaugh J, et al (2010) QIIME allows analysis of
414 high-throughput community sequencing data. *Nat Methods* 7:335–6.
415 <https://doi.org/10.1038/nmeth.f.303>
- 416 16. Quast C, Pruesse E, Yilmaz P, et al (2013) The SILVA ribosomal RNA gene
417 database project: Improved data processing and web-based tools. *Nucleic Acids*
418 *Res* 41:590–596. <https://doi.org/10.1093/nar/gks1219>
- 419 17. Fay MP, Proschan M a (2010) Wilcoxon-Mann-Whitney or t-test? On
420 assumptions for hypothesis tests and multiple interpretations of decision rules.
421 *Stat Surv* 4:1–39. <https://doi.org/10.1214/09-SS051>
- 422 18. Murtagh F, Legendre P (2014) Ward ’ s Hierarchical Agglomerative Clustering
423 Method□: Which Algorithms Implement Ward ’ s Criterion□? *J Classif* 31:274–
424 295. <https://doi.org/10.1007/s00357->

- 425 19. Anderson M (2001) A new method for non-parametric multivariate analysis of
426 variance. *Austral Ecol* 32–46
- 427 20. R Development Core Team (2014) R: A language and environment for statistical
428 computing.
- 429 21. Wickham H (2016) *ggplot2*. Springer International Publishing, Cham
- 430 22. McMurdie PJ, Holmes S (2013) phyloseq: an R package for reproducible
431 interactive analysis and graphics of microbiome census data. *PLoS One*
432 8:e61217. <https://doi.org/10.1371/journal.pone.0061217>
- 433 23. Oksanen J, Blanchet FG, Kindt R, et al (2013) *vegan: Community Ecology*
434 *Package*
- 435 24. Skinner KA, Leathers TD (2004) Bacterial contaminants of fuel ethanol
436 production. *J Ind Microbiol Biotechnol* 31:401–408.
437 <https://doi.org/10.1007/s10295-004-0159-0>
- 438 25. Hill M (1973) Diversity and evenness: a unifying notation and its consequences.
439 *Ecology* 54:427–432. <https://doi.org/10.2307/1934352>
- 440 26. Haegeman B, Hamelin J, Moriarty J, et al (2013) Robust estimation of microbial
441 diversity in theory and in practice. *ISME J* 7:1092–1101.
442 <https://doi.org/10.1038/ismej.2013.10>
- 443 27. Kamboj K, Vasquez A, Balada-Llasat JM (2015) Identification and significance
444 of *Weissella* species infections. *Front Microbiol* 6:1–7.
445 <https://doi.org/10.3389/fmicb.2015.01204>
- 446 28. Fairfax MR, Lephart PR, Salimnia H (2014) *Weissella confusa*: Problems with
447 identification of an opportunistic pathogen that has been found in fermented
448 foods and proposed as a probiotic. *Front Microbiol* 5:1–5.
449 <https://doi.org/10.3389/fmicb.2014.00254>
- 450 29. Fusco V, Quero GM, Cho GS, et al (2015) The genus *Weissella*: Taxonomy,
451 ecology and biotechnological potential. *Front Microbiol* 6:
452 <https://doi.org/10.3389/fmicb.2015.00155>
- 453 30. Hammes WP, Vogel RF (1995) The genus *Lactobacillus*. In: Wood BJB,
454 Holzapfel WH (eds) *The Genera of Lactic Acid Bacteria*. Springer US, Boston,
455 MA, pp 19–54
- 456 31. Goh KM, Gan HM, Chan KG, et al (2014) Analysis of *Anoxybacillus* genomes
457 from the aspects of lifestyle adaptations, prophage diversity, and carbohydrate
458 metabolism. *PLoS One* 9:. <https://doi.org/10.1371/journal.pone.0090549>
- 459 32. Studholme DJ (2015) Some (bacilli) like it hot: Genomics of *Geobacillus* species.
460 *Microb Biotechnol* 8:40–48. <https://doi.org/10.1111/1751-7915.12161>
- 461 33. Zeigler DR (2014) The *Geobacillus* paradox: Why is a thermophilic bacterial
462 genus so prevalent on a mesophilic planet? *Microbiol (United Kingdom)* 160:1–
463 11. <https://doi.org/10.1099/mic.0.071696-0>

- 464 34. Goh KM, Kahar UM, Chai YY, et al (2013) Recent discoveries and applications
465 of *Anoxybacillus*. *Appl Microbiol Biotechnol* 97:1475–1488.
466 <https://doi.org/10.1007/s00253-012-4663-2>
- 467 35. Doughari HJ, Ndakidemi PA, Human IS, Benade S (2011) The Ecology, Biology
468 and Pathogenesis of *Acinetobacter* spp.: An Overview. *Microbes Environ*
469 26:101–112. <https://doi.org/10.1264/jsme2.ME10179>
- 470 36. Rogers PL, Lee KJ, Skotnicki ML, Tribe DE (1982) Ethanol production by
471 *Zymomonas mobilis*. pp 37–84
- 472 37. Swings J, De Ley J (1977) The biology of *Zymomonas*. *Bacteriol Rev* 41:1–46
- 473 38. Narendranath N V., Power R (2005) Relationship between pH and medium
474 dissolved solids in terms of growth and metabolism of lactobacilli and
475 *Saccharomyces cerevisiae* during ethanol production. *Appl Environ Microbiol*
476 71:2239–2243. <https://doi.org/10.1128/AEM.71.5.2239-2243.2005>
- 477 39. Narendranath N V., Hynes SH, Thomas KC, Ingledew WM (1997) Effects of
478 lactobacilli on yeast-catalyzed ethanol fermentations. *Appl Environ Microbiol*
479 63:4158–4163. <https://doi.org/9361399>

480

481

489

490

Figure 2. Species observed in each step of ethanol production.

491

492

493

494

Figure 3. Relative abundance of the main genera found in each step of ethanol production.

Figure 4. Heatmap of the 20 most abundant OTU_{0.03} found in each step of ethanol production.

498
499 Figure 5. Principal Coordinates Analysis (PCoA) using Bray-Curtis dissimilarity matrix
500 for the samples of ethanol production process.