Characterization of relative abundance of lactic acid bacteria species in French organic sourdough by cultural, qPCR and MiSeq high-throughput sequencing methods

Elisa Michel, Clarisse Monfort, Marion Deffrasnes, Stéphane Guezenec, Emilie Lhomme, Matthieu Barret, Delphine Sicard, Xavier Dousset, Bernard Onno

Characterization of relative abundance of lactic acid bacteria species in french organic sourdough by cultural, qPCR and MiSeq high-throughput sequencing methods

Elisa Michel, Clarisse Monfort, Marion Deffrasnes, Stéphane Guezenec, Emilie Lhomme, Matthieu Barret, Delphine Sicard, Xavier Dousset, Bernard Onno

PII: S0168-1605(16)30392-0
DOI: doi: 10.1016/j.ijfoodmicro.2016.07.034
Reference: FOOD 7324

To appear in: *International Journal of Food Microbiology*

Received date: 4 February 2016
Revised date: 25 July 2016
Accepted date: 27 July 2016

Please cite this article as: Michel, Elisa, Monfort, Clarisse, Deffrasnes, Marion, Guezenec, Stéphane, Lhomme, Emilie, Barret, Matthieu, Sicard, Delphine, Dousset, Xavier, Onno, Bernard, Characterization of relative abundance of lactic acid bacteria species in french organic sourdough by cultural, qPCR and MiSeq high-throughput sequencing methods, *International Journal of Food Microbiology* (2016), doi: 10.1016/j.ijfoodmicro.2016.07.034

Characterization of relative abundance of Lactic Acid Bacteria species in French organic sourdough by cultural, qPCR and MiSeq high-throughput sequencing methods

Authors

Elisa Michel [a,b], Clarisse Monfort [a], Marion Deffrasnes [a], Stéphane Guezenec [b], Emilie Lhomme [a], Matthieu Barret [c], Delphine Sicard [b], Xavier Dousset [d], Bernard Onno [a]

[a] Oniris, Laboratoire de Microbiologie Alimentaire et Industrielle, Rue de la Géraudière CS 82225, 44322 Nantes Cedex 3, France

[b] INRA, UMR Sciences pour l'œnologie, 2 place Pierre Viala, 34060 Montpellier Cedex 01, France

[c] INRA, UMR1345 Institut de Recherche en Horticulture et Semences, 42 rue Georges Morel, 49007 Beaucouzé Cedex, France

[d] SECALIMUnit UMR1014,Oniris, INRA, Université BretagneLoire, Nantes, France.UMR INRA 1014 Secalim, La Chantrerie CS, 40307 Nantes Cedex 3, France

Corresponding author: Bernard Onno, Xavier Dousset

Oniris, Laboratoire de Microbiologie Alimentaire et Industrielle, CS 82225, 44322 Nantes Cedex 3

E-mail address: bernard.onno@oniris-nantes.fr, xavier.dousset@oniris-nantes.fr

Abstract

In order to contribute to the description of sourdough LAB composition, MiSeq sequencing and qPCR methods were performed in association with cultural methods. A panel of 16 French organic bakers and farmer-bakers were selected for this work. The lactic acid bacteria (LAB) diversity of their organic sourdoughs was investigated quantitatively and qualitatively combining *(i) Lactobacillus. sanfranciscensis*-specific qPCR, *(ii)* global sequencing with MiSeq Illumina technology and *(iii)* molecular isolates identification. In addition, LAB and yeast enumeration, pH, Total Titratable Acidity, organic acids and bread specific volume were analyzed. Microbial and physico-chemical data were statistically treated by Principal Component Analysis (PCA) and Hierarchical Ascendant Classification (HAC). Total yeast counts were 6 $\log_{10}$ to 7.6 $\log_{10}$ CFU/g while LAB counts varied from 7.2 $\log_{10}$ to 9.6 $\log_{10}$ CFU/g. Values obtained by *L. sanfranciscensis*-specific qPCR were estimated between 7.2 and 10.3 $\log_{10}$ CFU/g, except for one sample at 4.4 $\log_{10}$ CFU/g. HAC and PCA clustered the sixteen sourdoughs into three classes described by their variables but without links to bakers' practices. *L. sanfranciscensis* was the dominant species in 13 of the 16 sourdoughs analyzed by Next Generation Sequencing (NGS), by the culture dependent method this species was dominant only in only 10 samples. Based on isolates identification, LAB diversity was higher for 7 sourdoughs with the recovery of *L. curvatus, L. brevis, L. heilongjiangensis, L.xiangfangensis, L. koreensis, L.*

pontis, Weissella sp. and Pediococcus pentosaceus, as the most representative species. *L. koreensis, L. heilongjiangensis* and *L. xiangfangensis* were identified in traditional Asian food and here for the first time as dominant in organic sourdough. This study highlighted that *L. sanfranciscensis* was not the major species in 6/16 sourdough samples and that a relatively high LAB diversity can be observed in French organic sourdough.

Keywords: organic sourdough, Lactic Acid Bacteria, molecular characterization, qPCR, MiSeq

1. Introduction

Bread is a staple in many European countries and the production of sourdough breads is part of a cultural and geographical identity (De Vuyst and Neysens, 2005). Despite a number of technical constraints, natural sourdough has advantages such as enhancing bread flavor, prolonging the shelf life, improving the dough structure as well as increasing the nutritional value (Minervini et al., 2014). The microbiota of a stable sourdough principally consists of lactic acid bacteria (LAB) and yeasts (Huys et al., 2013). New species of the genus *Lactobacillus* have been isolated from traditional sourdoughs and various studies on sourdough from different countries have been conducted to isolate a wide range of LAB (De Vuyst and Vancanneyt, 2007). Although they mostly belong to the genus *Lactobacillus*, other genera such as *Leuconostoc*, *Weissella*, *Pediococcus* and *Enterococcus* have also been identified (Gobbetti and Gänzle, 2012). Many species have been found, such as *L.sakei*, *L. crustorum, L. nantensis, L. mindensis, L. reuteri,, L. brevis, L. plantarum, L. buchneri, L. curvatus, L. panis, L. pontis, L. sanfranciscensis, L. spicheri, L. kimchi, L. amylovorus, L. casei, W. cibaria*, and *W. confusa* (De Vuyst et al., 2014; Gobbetti and Gänzle, 2012). For yeasts, the six most common yeast species found in stable sourdoughs are *S. cerevisiae, Kazachstania exigua, Candida humilis, Pichia kudriavzevii, Torulaspora delbrueckii*, and *Wickerhamomyces anomalus* (De Vuyst et al., 2014). Regarding French ones, *L. plantarum* and *Pediococcus pentosaceus* were found to be dominant (Robert et al., 2009). Other studies revealed a higher diversity with species such as *L.hammesii* and *L.nantensis* (Valcheva et al., 2006, 2005). For French organic and conventional sourdoughs, recent studies demonstrated that their microbiota contain mainly *L. sanfranciscensis* (Lhomme et al., 2015a, 2015b) and *Kazachstania bulderi* and *K. unispora* as dominant yeast species (Lhomme et al., 2016). Recently, the rise in culture-independent methods has provided additional data to culture-dependent techniques. Several culture-independent techniques such as pyrosequencing (Bessmeltseva et al., 2014; Ercolini et al., 2013; Lattanzi et al., 2013; Lhomme et al., 2015b), MiSeq ((Minervini et al., 2015), HRM-qPCR (quantitative High Resolution Melting PCR) (Lin and Gänzle, 2014), and quantitative PCR (Lee et al., 2015; Scheirlinck et al., 2009) have been used to study sourdough microbial diversity. The aim of this paper was to analyze quantitatively and qualitatively the LAB relative abundance of organic sourdoughs. A panel of 16 bakers using natural sourdough and organic

flours were selected for this study. They were characterized by their practices (farmer-bakers, artisan-bakers and industrial-bakers) and their geographical location. To describe LAB diversity, we combined (i) sourdough isolate identification, (ii) *L. sanfranciscensis* quantitative PCR and (iii) global sequencing with MiSeq Illumina technology.

2. Materials and Methods

2.1. Sourdough and bread sampling

Sixteen bakers located in different regions of France were selected (Table 1). They were also chosen because of their bakers' status and related practices: farmer-bakers who produce their own flour (B15, B20, B21, B22, B25 and B27), artisan bakers (B16, B17, B18, B19 and B24) or industrial bakers (B23, B26, B28, B29 and B30). Sourdoughs were collected at the end of the last backslopping as mature sourdoughs (final leavened dough) and final breads were also sampled. Both were stored in sterile vials and conserved at 4 °C until analyses.

2.2. Microbial analysis

For each sample, sourdough was ten-fold diluted in TS (0.1% tryptone, 0.85% NaCl) and mixed for 2 min with a Stomacher (AES Laboratoire, France). Cascade dilutions were performed from 10^{-2} to 10^{-6} and then plated with a spiral plater (Interscience, Saint-Nom-la-Bretèche, France) on MRS5 characterized by a vitamin mix addition (Meroth et al., 2003; Vera et al., 2009) for LAB enumeration and YPD (4 g/L Yeast Extract, 8 g/L glucose, 6.8 g/L agar) for yeast enumeration. Plates were incubated for 48 h at 30 °C under anaerobic conditions (Anaerocult A, Merck, Darmstadt, Germany) for LAB and at 26 °C under aerobic conditions for yeast. After LAB enumeration, approximately 15 bacterial colonies from each sample were selected and, after overnight culture on MRS, maltose (0.05%) and cysteine (1%), these were stored at -80 °C with 40% glycerol. LAB isolates were identified by partial or whole 16S rDNA sequencing. The 16S rDNA (about 1500 bp) of the pure LAB isolates was amplified by PCR as described previously (Jaffrès et al., 2009) from chromosomal DNA using primers fD1 and rD1 (Weisburg et al., 1991). The sequencing primers SP1, SP2, SP3, and SP5 targeting two conserved regions of the 16S rRNA gene were used (Lhomme et al., 2015a). The partial nucleotide sequence (about 700 bp) of the amplified 16S rDNA gene was determined using the sequencing primer SP1. To confirm species group identity, the whole 16S rDNA gene was sequenced (about 1500 bp) using sequencing primers SP. Forward and reverse sequences obtained with SP1 and SP2 or with SP3 and SP5 (Table 2) were concatenated with BioEdit. A similarity of at least 97.6% to 16S rDNA gene sequences of type strains was used as the criterion for identification (Stackebrandt and Ebers, 2006). As the 16S rDNA sequence did not discriminate *Lactobacillus plantarum*, *Lactobacillus paraplantarum* and *Lactobacillus pentosus*, a multiplex PCR targeting the *recA* gene (Torriani et al., 2001) was performed. Three forward primers (para-F, pent-F and F-Plan) specific for

each species and a common reverse primer (pREV) were used (Table 2), expecting an amplicon size depending on the species: 107, 218 and 318 bp for *L. paraplantarum*, *L. pentosus* and *L. plantarum*, respectively. In addition, a PCR targeting the *katA* gene (407 bp) was used (Ammor et al., 2005) to discriminate 16S rDNA identification of *Lactobacillus sakei* and *Lactobacillus curvatus* (Table 2). Similarly, as the 16S rDNA sequence did not discriminate all lactic acid bacteria, *rpoA* and *pheS* genes sequencing was performed for *L. koreensis* and *L.heilongjiangensis* identification, using the sequencing primers (Table 2) rpoA-21-F/rpoA-23-R and pheS-21-F and pheS-23-R (Naser et al., 2007). Identification queries were fulfilled by a BLAST search against the National Center for Biotechnology Information (NCBI, Bethesda, USA), the BIBI (Devulder et al., 2003) and the Ribosomal Data Project (Cole et al., 2011) databases.

Single colonies from pure bacterial cultures grown on MRS plates were transferred onto Biolog universal growth (BUG™) agar (Biolog Inc., Hayward,CA) and incubated at 28°C for 24 h. Colonies were picked using a sterile moistened Biolog cotton swab, suspended in sterile inoculating fluid, IF-A (Biolog Inc., Hayward, CA) and concentration adjusted to match Biolog GEN III turbidity standards. Aliquots of100µL of bacterial suspensions were loaded into each well of the Microplates. Each bacterial isolate was inoculated on a separate Biolog GEN III MicroPlates™. Readings of inoculated Biolog GEN III microplates was performed at 580nm (Tecan Infinite Pro 200, Salzburg, Austria) after 24–48 h of incubation at 30°C. All the wells start out colorless when inoculated. During incubation, there was increased respiration in the wells where cells could use a carbon source and/or grow. Increased respiration causes reduction of the tetrazolium redox dye, forming a purple color. Negative wells remain colorless as did the negative control well with no carbon source. A positive control well was also used as reference for the chemical sensitivity assays.

2.3. Determination of TTA (Total Titratable Acidity), pH, bread mass density and organic acids

Sourdough samples were ten-fold diluted with distilled water and mixed with a Stomacher for 2 min (AES Laboratoire, France). Each sample was analyzed in triplicate with an automatic titrator (pH-Matic 23, Grosseron, Saint-Herblain, France), using a 10-mL volume and N/10 NaOH solution concentration. The results of the TTA assay were expressed in volume of 0.1 M sodium hydroxide used to neutralize a 10 g sample (Romanian Standard Methods 90/2007). To measure the bread specific volume, rapeseed density was used. The bread was added in the same container and the excess seed mass was calculated, expressed in dm^3/kg. To measure organic acids, 6 ml of the homogenized mixture was centrifuged at 13,000 rpm for 5 min at room temperature. The samples were clarified by Carrez reagents I and II (250 µL) and centrifuged at 13,000 rpm for 5 min. The supernatant (water-soluble extract) was then analyzed for its organic acid concentration; it was diluted with 10 mM H_2SO_4 and analyzed by liquid chromatography using an HPLC system (Zeppa et al., 2001). Acid amounts were expressed as g/kg of sourdough, dough or bread.

2.4. Statistical analysis

LAB and yeast microbial counts, *L. sanfranciscensis* concentration estimation and biochemical characteristics were used as variables for statistical analyses. Principal Component Analysis (PCA) was carried out to describe the relations between sourdough, bread physico-chemical and microbiological variables. Hierarchical Ascendant Classification (HAC) was carried out to cluster sourdoughs and build a partition into homogeneous clusters. These tests were performed using XLstat (Addinsoft, 2007) and Rstudio Version 3.2.5 http://www.rstudio.com/ (Team, 2015).

2.5. Sourdough DNA extraction

Sourdough samples were five-fold diluted in TS (0.1% tryptone, 0.85% NaCl), then Tween80 was added and mixed for 2 min in a stomacher (AES Laboratoire, France). The matrix was transferred to a NucleoSpin® tube. After centrifugation at 8500 rpm for 10 min, the bacterial pellet was taken up in a lysis solution (1 M Tris-HCl, pH 8, 0.1 M Na-EDTA, Triton X100, 100 mg lysozyme/ml, mutanolysin 200 U/ml).The solution was then incubated at 37 °C and a vibratory lysis step using glass beads was performed (mini BeadBeater-8®, Biospec, California, USA). The samples were placed in an ice tray to burst the cells by thermal shock. After addition of 25 µL of proteinase K and 200 µL of AL buffer, the samples were incubated at 56 °C for 1 h then centrifuged for 3 min at 10000 rpm. Two hundred µL of ethanol was added and the mix was transferred to a Qiagen column. Two successive steps of washing with 500 µL of Buffer AW1 and Buffer AW2 were carried out. After each centrifugation (8000 rpm for 1 min), the filtrates were discarded. DNA was eluted in 100 µL of elution solution by centrifugation at 8000 rpm for 1 min. DNA was stored at -20 °C and its concentration and purity were checked using a spectrophotometer (Nanodrop ND-100; Nanodrop Technologies).

2.6. Quantitative PCR to detect *L. sanfranciscensis*

The *pheS* gene was amplified from sourdough DNA with primers LS-1F and LS-1R, as described by Scheirlinck et al. (2009). Previously, a standard curve was constructed on a target strain *L. sanfranciscensis* ATCC 43332, in the range 4–10 log CFU/g (Lhomme et al., 2015b). In addition, one non-inoculated dough was used as the negative control. The different bacterial cell densities were plotted against the corresponding CT values (Opticon Monitor Software 3®). Amplification efficiency (E) was calculated as $E = 10^{-1}/\text{slope}$ (Klein, 2002). *L. sanfranciscensis* in the sourdough samples was quantified by qPCR, using the mean values of CT obtained in three independent qPCR experiments.

2.7. Sequencing on MiSeq Illumina

After sourdough DNA extraction, amplicon libraries were constructed following two rounds of PCR amplification. The first step was performed with the PCR primers 515f and 806r (Caporaso et al., 2011) which target the V4 region of the 16S rRNA gene. Forward and reverse primers carried the

following 5'-CTTTCCCTACACGACGCTCTTCCGATCT-3'and 5'-GGAGTTCAGACGTGTGCTCTTCCGATCT-3' tails, respectively. All PCR reactions were performed with a high-fidelity polymerase (AccuPrime *Taq* DNA Polymerase System; Invitrogen) using the manufacturer's protocol and x µl of environmental DNA (approximately x ng). Cycling conditions for 515f/806r were adapted from Caporaso et al. (2011). Briefly, reactions were held at 94 °C for 2 min, followed by 30 cycles of amplification at 94 °C (30 s), 50 °C (60 s) and 68 °C (90 s) with a final extension step of 10 min at 68 °C. All amplicons were purified with the Agencourt AMPure XP system and quantified with QuantIT PicoGreen. A second round of amplification was performed with 5 µl of purified amplicons and primers containing the Illumina adapters and indexes. PCR cycling conditions were: 94 °C (2 min), followed by 12 cycles of amplification (94 °C for 1 min, 55 °C for 1 min, 68 °C for 1 min) and a final extension step at 68 °C (10 min). All amplicons were purified and quantified as previously described. The purified amplicons were then pooled in equimolar concentrations and the final concentration of the library was determined using a Kapa qPCR Quantification Kit. Amplicon libraries were mixed with 5% PhiX control according to the Illumina protocols. The sequencing run was performed with MiSeq Reagent Kit v3 (600 cycles) (Barret et al., 2015).

2.8. Sequence analyses

Raw reads were analyzed using the steps described in the standard operating procedure of Mothur at http://www.mothur.org/wiki/MiSeq_SOP (Kozich et al., 2013). Briefly, 16S rRNA gene sequences were aligned against the 16S rRNA gene SILVA alignment using Mothur v1.33 (Schloss et al., 2009). Chimeric sequences were detected with UCHIME (Edgar et al., 2011) and subsequently removed from the dataset. Taxonomic affiliation of 16S rRNA genes was performed with a Bayesian classifier (Wang et al., 2007) (80% bootstrap confidence score) against the 16S rRNA gene training set (v9) of the Ribosomal Database Project (Cole et al., 2011). Unclassified sequences or sequences belonging to Eukaryota, Archaea, chloroplasts or mitochondria were discarded. Sequences were assigned to operational taxonomic units (OTUs) at 97% identity. Only abundant OTUs representing at least 0.1% of the library size were conserved for microbial community analyses (Barret et al., 2015). Taxonomic affiliation of these OTUs was done using BLAST+ and RDP against the different databases: NCBI, BIBI and RDP (Cole et al., 2011; Devulder et al., 2003; *The NCBI Handbook*, 2013).

3. Results and discussion

3.1. Enumeration of total LAB and yeast microflora

The number of total LAB enumerated on MRS5 medium ranged from 7.19 $\log_{10}$ to 9.62 $\log_{10}$ CFU with a mean value of 8.92 (Table 3). Sourdough B29 presented a low concentration, 7.19 $\log_{10}$ CFU/g of LAB, in comparison with other sourdoughs. The yeast population ranged from 5.92 $\log_{10}$ to 7.68

log_{10} CFU/g, with a mean value of 7.19. The yeast/LAB ratio ranged from 0.15% to 66.07%. The vast majority of sourdoughs present a low ratio < 10%, except for B22 and B29 (38.9% and 66.07% respectively). These results are consistent with those previously found for French and European sourdoughs (Gobbetti et al., 1994; Lhomme et al., 2015b).

3.2. Physical and chemical characteristics

TTA acidity, pH and bread specific volume are presented in Table 3. The pH of the sourdoughs ranged from 3.62 to 4.03, the breads from 3.98 to 4.96. Similarly, the TTA varied from 13.3 to 21.3 and 6.9 to 10.5 mL NaOH 0.1N for sourdoughs and bread, respectively. These are expected values. Dough acidity is mostly due to the metabolic activity of LAB (Gänzle and Gobbetti, 2015). This lower bread acidity is related to the dilution of the sourdough during its incorporation into the dough. The short fermentation time of the final dough (approximately 4 hours) prevents sufficient production of organic acids and a part is lost by volatilization during baking. TTA values vary depending on the nature of flours (Banu et al., 2011; Katina et al., 2005) and the fermentation temperature: at higher temperature, LAB produce organic acids more efficiently (Minervini et al., 2014). The bread specific volume varied between 1.81 and 2.80 dm^3/kg. These variations are due to the amount and activity of yeasts present in the sourdough and gas production. However, the dough retention capacity and the final volume are also related to flour properties, proteolytic activity (Thiele et al., 2002) and baking practices. Lactic acid values varied from 3.37 to 11.96 g/kg and acetic acid values from 0.91 to 4.44 g/kg. The fermentative quotient varied from 1.06 to 5.66; these are expected values consistent with previous observations (Lhomme et al., 2015a; Valmorri et al., 2010). The sourdough and bread samples studied here come from different bakers using different flours, with different baking practices in terms of temperature and the proportion of sourdough used in the final dough. These variable conditions explain the physico-chemical differences observed.

3.3. Estimated number of *L. sanfranciscensis* in sourdough by quantitative PCR

Because of the predominance of *L. sanfranciscensis* in French organic sourdough (Lhomme et al., 2015b), the qPCR technique was used to assess the *L. sanfranciscensis* population in sourdoughs (Gobbetti and Corsetti, 1997; Lee et al., 2015). Scheirlinck et al. (2009) showed the value of this method to estimate quickly the presence and amount of this species in a sample. Here, *L. sanfranciscensis* qPCR estimates varied from 4.41 to 10.29 log_{10} CFU/g (Table 4). For 13/16 sourdoughs (all but B23, B25, B28), the number of *L. sanfranciscensis* estimated by qPCR represented more than 90% (qPCR / total LAB ratio > 0.9) of the total number of LAB (Table 4), showing that *L. sanfranciscensis* was the main LAB species in these sourdoughs. These results are consistent with those found by Lhomme et al. (2015b), in which *L. sanfranciscensis* was predominant in 10/15 sourdoughs analyzed.

The observed overestimation (ratio > 1) of *L. sanfranciscensis* by qPCR compared to the total number of LAB for 7/16 samples (Tables 3 and 4), may be related to an underestimation of the LAB enumeration due to the selectivity of the medium MRS5. However, this medium is the most suitable to enable isolation of specific LAB species in bread sourdough (Vera et al., 2009). Alternatively, the qPCR method can enumerate viable and non-cultivable *L. sanfranciscensis* cells, i.e. living and dead bacteria. In order to differentiate living and dead cells specific PCR and sample treatment with PMA (propidium monoazide) may be combined to determine the viable population quantitatively (Macé et al.; 2013). For Hierro et al. (2006), RT-qPCR could be an alternative method to rule out the counting of dead cells.

3.4. Sourdough isolate identification

To complete and investigate LAB diversity, isolates were identified using 16S rDNA and *rpoA* sequence analysis. The results are shown in Table 4. For 7 sourdoughs (B17, B18, B19, B20, B22, B24 and B27) 100% of isolates belonged to *L. sanfranciscensis* species. For 3 sourdoughs (B26, B21 and B1), *L. sanfranciscensis* represented more than 50% of isolates. For 6 sourdoughs of the 16 studied, *L. sanfranciscensis* was not dominant or absent (B16, B23, B25, B28, B29 and B30).

Sourdough B16 isolates belonged to *L. curvatus* (90%) and *L. plantarum* (10%) species. For sourdough B25, isolates were identified as *L. brevis* (69%), *L. plantarum* (15%), *L. sanfranciscensis* (8%) and *L. paralimentarius* (8%). Sourdough B28 isolates were classified as *L. sanfranciscensis* (40%), *L. pontis* (33%), *L. diolivorans* (13%), *L. paracasei* (7%), and *L. parabuchneri* (7%). Greater genera diversity was observed in the sourdough B30 with *P. pentosaceus* and *Weissella confusa* as the dominant species. In sourdough B29, *L. xiangfangensis* was isolated for the first time in sourdough as the prevailing LAB species.

Finally, sourdough B23 was unique here because of its low value of *L. sanfranciscensis* (4.41 $\log_{10}$ CFU/g) estimated by qPCR compared to a total of 9.02 $\log_{10}$ LAB. Isolates were sorted into 4 distinct groups according to their rDNA 16S sequence homology. The first group was identified as *L. curvatus* by rDNA 16S and *katA* gene absence by PCR amplification (Ammor et al., 2005). For the second group, three species were identified on the basis of 16S sequencing (*L. plantarum*, *L. pentosus*, *L. paraplantarum*). Using *recA* amplicon size variation (Torriani et al., 2001), the isolates were identified as being *L. plantarum* (amplicon size of 318 bp). Isolates from group 3 were identified by rDNA 16S sequencing as belonging to *L. heilongjiangensis* (100%). Using *rpoA and pheS* genes (Gu et al., 2013; Naser et al., 2007), *L. heilongjiangensis* identity was confirmed with 99% and 96% homology respectively and distinguished from closely-related LAB species as *L. nantensis*, *L. mindensis*, *L. crustorum* or *L. farciminis*. Finally, for group 4, rDNA 16S sequences showed 97% of homology with *L. koreensis* and 95% of homology with *L. parabrevis* using RDP database. With BIBI database, the closest sequence based on patristic distances was *Lactobacillus koreensis* T FJ904277

with 100% of identity. Adding sequencing of *rpoA* gene and *pheS* gene, the presence of *L. koreensis* in sourdough B23 was confirmed.

For 7 sourdough samples (B16, B21, B23, B25, B28, B29 and B30), LAB diversity was higher than previously found in French organic sourdoughs (Lhomme et al., 2015a).

The sourdough LAB diversity observed here was characterized by new species isolated in sourdough: *L. xiangfangensis*, *L. heilongjiangensis* and *L. koreensis*. Regarding *L. xiangfangensis*, it has been isolated in Chinese pickles (Gu et al., 2012). Already identified in French sourdough (Lhomme et al., 2015a), its occurrence was confirmed here in two sourdoughs (B21 and B29). It belongs to the *L.plantarum* species group and produces acid from diverse carbohydrate substrates such as ribose, D xylose, glucose, sucrose, fructose, and maltose. Its genome has been sequenced (Sun et al., 2015). *L. heilongjiangensis* had previously been found in traditional Chinese pickles (Gu et al., 2013) and was isolated here for the first time in sourdough (B23). It is a Gram-positive, catalase-negative, non-spore-forming rod and produces acid from many carbon substrates. The metabolic profile of *L. heilongjiangensis* strain revealed the use of mainly glucose, fructose, mannose and sucrose (Table S1). The complete genome is 2.79 Mbp, devoid of plasmids, and with a GC content of 37.5%. Pathways involved in the biosynthesis of riboflavin and folate were identified in the genome of *L. heilongjiangensis* DSM 28069T, which are present in some other probiotic LAB strains (Zheng et al., 2015). Considering *L. koreensis*, previously isolated from Kimchi (Bui et al., 2011), it was also isolated in sourdough (B23) for the first time. Closely related to *L. brevis*, *L. koreensis* is an heterofermentative species and metabolizes mainly fructose, glucose, and maltose (Table S2). Its genome has been sequenced recently (Sun et al., 2015). The genomic diversity of *L. xianfangensis*, *L. heilongjiangensis* and *L. koreensis* could be described by analyzing and comparing genome sequences. Their genomes were sequenced with 210 other *Lactobacillus* strains by Sun et al., (2015): they presented a robust phylogenomic framework of existing species and for classifying new species.

3.5. Relative abundance of LAB species estimated by Illumina MiSeq Sequencing

To complement the cultural and molecular approaches, the relative abundance of species in the 16 samples was determined by NGS with Illumina MiSeq technology. A total of 777,750 quality-trimmed sequences of 16S RNA gene amplicons were obtained. The average number of sequences per sample was 51,850. After analysis, a total of 20 OTUs were identified, with a range of 1 to 8 OTUs per sourdough. Table 4 shows the OTUs found with a relative abundance of at least 0.1% for each sourdough. For 12 sourdoughs, B15, B17, B18, B19, B20, B21, B22, B24, B26, B27, B29, and B30, *L. sanfranciscensis* was present at more than 98%. This OTU was also detected at much lower abundance in B25 and B28 sourdough samples (36.1% and 88.8%, respectively) and at very low abundance in B23 and B16 samples (0.2% and 0.1%, respectively). With a relative abundance of 75.2%, *L. curvatus*

was the predominant species in B16 sourdough. For sourdough B25, the predominant species (61.9%) corresponded to an OTU that was identified only at the taxonomic genus level, *Lactobacillus* sp. Similarly, other OTUs in sourdough B15 (0.1%), B16 (12.0%), B21 (0.1%), B26 (0.6%) and B30 (0.1%) were affiliated only to the genus *Lactobacillus*. Finally, the OTU corresponding to 80.07% in B23 sourdough could not be formally related to a species, but was either *L. koreensis* or *L. brevis*. In sourdoughs B21, B29 and B30, the same OTU was detected with a relative abundance of 0.1%, 0.2% and 0.1%, respectively. In B23 sourdough, the OTU corresponding to 16.7% could not be related to a species, but was either *L. heilongjiangensis*, *L. farciminis* or *L. crustorum*. Other sub-dominant species detected in the sourdoughs (relative abundance of 0.1%-4.3%) were *Weissella* found in B16 (0.2%) and B30 (0.4%), *Pediococcus pentosaceus* in B30 (0.2%), *L. sakei* in B16 (0.7%), *L. brevis* in B25 (0.9%), *L. nantensis* in B23 (0.8%), *Leuconostoc* in B23 (0.4%), and *L. parabrevis* in B25 (0.6%) and B28 (2.4%). *L. pontis* was detected at a relative abundance of 4.3%, *L. buchneri* at 0.8% and *L. diolivorans* at 2.3%, in B28 sourdough. These Miseq analysis observations could reveal a relatively higher biodiversity considering subdominant LAB species.

3.6. Comparison of LAB diversity analyzed by the different approaches

Quantitative PCR showed that *L. sanfranciscensis* was the dominant species in 13 of the 16 sourdoughs studied. For 12 of the 13 samples, *L. sanfranciscensis* was also present at a high relative abundance estimated by MiSeq analysis (98.6%), consistent with the qPCR quantification (Table 4). Moreover, a significant correlation was observed between these two methods (R^2 =0.68, data not shown) in PCA analysis (Figure 1C). For 10 sourdoughs, culture-dependent method confirmed *L.sanfranciscensis* dominance (>50%) and were in agreement with culture-independent methods. Likewise, in sourdough B16, *L. curvatus* was predominant by the molecular method and MiSeq analysis confirming the possible occurrence of this species in sourdough as already described (Lhomme et al., 2015b; Robert et al., 2009). For sample B28, MiSeq results were partly consistent with the molecular characterization of isolates including *L. sanfranciscensis* dominance. Thus *L. pontis*, *L. diolivorans* and *L. buchneri* related sp. were also identified by these two methods.

Regarding sourdough B25, the results were not totally consistent between MiSeq and the cultural method. While *L. sanfranciscensis* and *L. brevis* were found by both methods, this was not the case for *L. plantarum* and *L.paralimentarius*, only identified as isolates. However, these species might have been classified as *Lactobacillus* species and could not be characterized more precisely by the MiSeq method. The same analysis and comments can be made for B26 and B30. Surprisingly, sourdough B29 analysis revealed a discrepancy regarding MiSeq results and isolate identification. Bias in the DNA extraction, sequence analyses or in sequencing might have occurred. Furthermore, the databases used could be incomplete or may contain poor quality sequences (Ercolini, 2013). Medium composition and strain- specific requirements could also contribute to this divergence.

Finally, the most original sourdough of our study was sample B23. *L. sanfranciscensis* was found with a relative abundance of 0.17% in MiSeq analyses, corresponding to the low population (log 4.41 CFU/g) estimated by qPCR. By molecular analysis of isolates, 37.5% of *L. heilongjiangensis* was found, 25% of *L. plantarum*, 25% of *L. koreensis* and 12.5% of *L. curvatus*. This species might also have been detected by MiSeq analysis. However, the MiSeq method could not discriminate between three closely-related species, due to a too short sequence used. MiSeq analysis could be extended to obtain more detailed and specific results by searching for primers amplifying a longer and more discriminating 16S rDNA sequence, or using housekeepings genes as *gyrB* or *rpoA*.

The results obtained here by the three methods partially confirm previous studies for traditional Italian sweet leavened baked goods (Lattanzi et al., 2013) and French organic sourdoughs (Lhomme et al., 2015b). One major point of this work is that LAB species diversity was higher than previously described (Lhomme et al. 2015a) for French organic sourdough. Our results are in agreement with the observations of Rizzello et al. (2015), showing that *Firmicutes* diversity was the highest for sourdough made with organic durum flours.

3.7. Relationships between bread and sourdough physico-chemical and microbiological characteristics by Hierarchical Ascendant Classification

Hierarchical Ascendant Classification (HAC) was carried out to cluster sourdoughs according to microbial counts, *L. sanfranciscensis* amount estimation (qPCR and MiSeq) and physico-chemical characteristics (Table 3 and 4). Three clusters were differentiated (Figure 1B). Cluster 1 grouped 6 sourdoughs, B15, B22, B24, B26, B27 and B28. Compared to the other clusters, this cluster was characterized by the lowest N LAB values counts and surprisingly the highest sourdough/bread acidity characteristics TTA, pH, acetic acid (Table 3). Four sourdoughs were associated in cluster 2, B16, B19, B23 and B25. They were characterized by the highest N LAB concentration counts, associated with low values for *L. sanfranciscensis* estimation (qPCR, Miseq Relative Abundance). For this cluster low acidity values were observed (sourdough /bread pH and TTA, lactic acid concentration). Enumeration media selectivity and intraspecific LAB diversity could explain these results (Foschino et al., 2001; Venturi et al., 2012). Finally cluster 3 grouped the remaining sourdoughs B17, B18, B20, B21, B29 and B30. They were characterized by yeast counts, high *L. sanfranciscensis* estimation (qPCR, Miseq Relative Abundance), low acetic acid concentration and high Fermentative Quotient. Furthermore, in this cluster the low number of yeasts could be related to the low bread volume. Regarding acetic acid concentration, the low number of yeasts for cluster 3 could have an impact on fructose production through sucrose metabolism (Gobetti et al., 1994; Meignen et al., 2001; Stolz et al., 1995). Overall, it is quite difficult to evaluate the relative contribution of LAB and yeasts for a given sample in terms of lactic and acetic acids balance.

HAC clustering (Table 5, Figure 1B) was completed by PCA analysis (Figure 1C). Miseq and qPCR *L. sanfranciscensis* estimation are statistically linked and also correlated with sourdough TTA. These results might be explained by the acidification capacity of *L. sanfranciscensis* in sourdough (Gobbetti and Corsetti, 1997). Surprisingly, acetic acid concentration was not related to microbiological counts and *L. sanfranciscensis* estimation. Yeast LAB interactions might explain these unexpected results. Furthermore, bakers' practices or status as defined in this study - farmer-bakers, industrial-bakers and artisan-bakers – were not significantly different on the physicochemical or microbiological responses considered here (Figure 1D). LAB diversity observed here, could partially explain these results. However, backslopping frequency might influence microbiota and sourdough characteristics (Huys et al., 2013). Moreover, Minervini et al. (2014) had observed relationships between bread physico-chemical characteristics and sourdough microbiological data and baking practices during the traditional sourdough process.

4. Conclusion

The aim of this study was to investigate the bacterial biodiversity of 16 French organic sourdoughs. The lactic acid bacteria diversity of these sourdough was investigated quantitatively and qualitatively combining *(i) L. sanfranciscensis*-specific qPCR, *(ii)* global sequencing with MiSeq Illumina technology and *(iii)* molecular isolate identification. Although *L. sanfranciscensis* was still the dominant species in 10 of the 16 sourdoughs studied, LAB species diversity was higher than previously described for French organic sourdough, with 16 different species identified considering all the methods used. Regarding both cultural and non-cultural methods, 6 samples showed a non *L. sanfranciscensis* dominance, whatever the characteristics of baker' practices. This study confirmed that *L. curvatus* could be a predominant species in sourdough. Furthermore, *L. xianfangensis*, *L. heilongjiangensis* and *L. koreensis* were highlighted for the first time in sourdough, having previously only been found in traditional Asian food. Genomics experiments focusing on the genomic diversity of these species will be performed. Finally, bakers' practices and status, as considered here, did not appeared to be significantly related to LAB diversity and sourdough features.

Acknowledgements

The authors acknowledge financial support from ANR Bakery–13-ALID-0005 (French National Research Agency), INRA and Oniris. We wish to thank the sixteen bakers as well as Alain and Françoise Bourgeois (Apiliger Earl) for their collaboration on the project, sharing with us their bread-making knowledge and pleasure in good bread. Thanks are also addressed to Marie-Noëlle Madec and Florence Valence-Bertel from INRA Rennes UMR 1253 Sciences et Technologie du Lait et de l'Œuf, Isabelle Hue and Héloïse Debroise from Oniris for their assistance in LAB microbiology, Marc Perez and Thibault Nidelet from INRA Montpellier Science pour l'œnologie for HPLC settings and statistical analyses respectively.

References

Addinsoft, T.M., 2007. XlSTAT Your data analysis solution. Versión.

Ammor, S., Dufour, E., Zagorec, M., Chaillou, S., Chevallier, I., 2005. Characterization and selection of *Lactobacillus sakei* strains isolated from traditional dry sausage for their potential use as starter cultures. Food Microbiol. 22, 529–538.

Banu, I., Vasilean, I., Aprodu, I., 2011. Effect of select parameters of the sourdough rye fermentation on the activity of some mixed starter cultures. Food Biotechnol. 25, 275–291.

Barret, M., Briand, M., Bonneau, S., Préveaux, A., Valière, S., Bouchez, O., Hunault, G., Simoneau, P., Jacques, M.-A., 2015. Emergence Shapes the Structure of the Seed Microbiota. Appl. Environ. Microbiol. 81, 1257–1266.

Bessmeltseva, M., Viiard, E., Simm, J., Paalme, T., Sarand, I., 2014. Evolution of bacterial consortia in spontaneously started rye sourdoughs during two months of daily propagation. PloS One 9, e95449.

Bui, T.P.N., Kim, Y.-J., In, J.-G., Yang, D.-C., 2011. *Lactobacillus koreensis* sp. nov., isolated from the traditional Korean food kimchi. Int. J. Syst. Evol. Microbiol. 61, 772–776.

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Lozupone, C.A., Turnbaugh, P.J., Fierer, N., Knight, R., 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc. Natl. Acad. Sci. 108, 4516–4522.

Cole, J.R., Wang, Q., Chai, B., Tiedje, J.M., 2011. The Ribosomal Database Project: Sequences and Software for High-Throughput rRNA Analysis. Handb. Mol. Microb. Ecol. Metagenomics Complement. Approaches 313–324.

De Vuyst, L., Neysens, P., 2005. The sourdough microflora: biodiversity and metabolic interactions. Trends Food Sci. Technol. 16, 43–56.

De Vuyst, L., Van Kerrebroeck, S., Harth, H., Huys, G., Daniel, H.-M., Weckx, S., 2014. Microbial ecology of sourdough fermentations: diverse or uniform? Food Microbiol. 37, 11–29.

De Vuyst, L., Vancanneyt, M., 2007. Biodiversity and identification of sourdough lactic acid bacteria. Food Microbiol. 24, 120–127.

Devulder, G., Perrière, G., Baty, F., Flandrois, J.P., 2003. BIBI, a bioinformatics bacterial identification tool. J. Clin. Microbiol. 41, 1785–1787.

Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C., Knight, R., 2011. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 27, 2194–2200.

Ercolini, D., 2013. High-Throughput Sequencing and Metagenomics: Moving Forward in the Culture-Independent Analysis of Food Microbial Ecology. Appl. Environ. Microbiol. 79, 3148–3155. doi:10.1128/AEM.00256-13

Ercolini, D., Pontonio, E., De Filippis, F., Minervini, F., La Storia, A., Gobbetti, M., Di Cagno, R., 2013. Microbial ecology dynamics during rye and wheat sourdough preparation. Appl. Environ. Microbiol. 79, 7827–7836.

Foschino, R., Arrigoni, C., Mora, D., Galli, A., others, 2001. Phenotypic and genotypic aspects of *Lactobacillus sanfranciscensis* strains isolated from sourdoughs in Italy. Food Microbiol. 18, 277–285.

Gobbetti, M., Corsetti, A., 1997. Lactobacillus sanfranciscoa key sourdough lactic acid bacterium: a review. Food Microbiol. 14, 175–188.

Gobbetti, M., Gänzle, M., 2012. Handbook on sourdough biotechnology. Springer Science & Business Media.

Gobetti, M., Corsetti, A., Rossi, J., 1994. The sourdough microflora. Interactions between lactic acid bacteria and yeasts: metabolism of amino acids. World J Microbiol Biotechnol 10, 275–279.

Gu, C.T., Li, C.Y., Yang, L.J., Huo, G.C., 2013. *Lactobacillus heilongjiangensis* sp. nov., isolated from Chinese pickle. Int. J. Syst. Evol. Microbiol. 63, 4094–4099.

Gu, C.T., Wang, F., Li, C.Y., Liu, F., Huo, G.C., 2012. *Lactobacillus xiangfangensis* sp. nov., isolated from Chinese pickle. Int. J. Syst. Evol. Microbiol. 62, 860–863.

Hierro, N., Esteve-Zarzoso, B., González, Á., Mas, A., Guillamón, J.M., 2006. Real-time quantitative PCR (QPCR) and reverse transcription-QPCR for detection and enumeration of total yeasts in wine. Appl. Environ. Microbiol. 72, 7148–7155.

Huys, G., Daniel, H.-M., Vuyst, L.D., 2013. Taxonomy and Biodiversity of Sourdough Yeasts and Lactic Acid Bacteria, in: Gobbetti, M., Gänzle, M. (Eds.), Handbook on Sourdough Biotechnology. Springer US, pp. 105–154.

Jaffrès, E., Sohier, D., Leroi, F., Pilet, M.F., Prévost, H., Joffraud, J.J., Dousset, X., 2009. Study of the bacterial ecosystem in tropical cooked and peeled shrimps using a polyphasic approach. Int. J. Food Microbiol. 131, 20–29.

Katina, K., Arendt, E., Liukkonen, K.-H., Autio, K., Flander, L., Poutanen, K., 2005. Potential of sourdough for healthier cereal products. Trends Food Sci. Technol. 16, 104–112.

Klein, D., 2002. Quantification using real-time PCR technology: applications and limitations. Trends Mol. Med. 8, 257–260.

Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K., Schloss, P.D., 2013. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. Appl. Environ. Microbiol. 79, 5112–5120.

Lattanzi, A., Minervini, F., Di Cagno, R., Diviccaro, A., Antonielli, L., Cardinali, G., Cappelle, S., De Angelis, M., Gobbetti, M., 2013. The lactic acid bacteria and yeast microbiota of eighteen sourdoughs used for the manufacture of traditional Italian sweet leavened baked goods. Int. J. Food Microbiol. 163, 71–79.

Lee, H., Baek, H., Lim, S.B., Hur, J.S., Shim, S., Shin, S.-Y., Han, N.S., Seo, J.-H., 2015. Development of species-specific PCR primers and polyphasic characterization of *Lactobacillus sanfranciscensis* isolated from Korean sourdough. Int. J. Food Microbiol. 200, 80–86.

Lhomme, E., Lattanzi, A., Dousset, X., Minervini, F., De Angelis, M., Lacaze, G., Onno, B., Gobbetti, M., 2015a. Lactic acid bacterium and yeast microbiotas of sixteen French traditional sourdoughs. Int. J. Food Microbiol. 215, 161–170.

Lhomme, E., Orain, S., Courcoux, P., Onno, B., Dousset, X., 2015b. The predominance of *Lactobacillus sanfranciscensis* in French organic sourdoughs and its impact on related bread characteristics. Int. J. Food Microbiol. 213, 40–48.

Lhomme, E., Urien, C., Legrand, J., Dousset, X., Onno, B., Sicard, D., 2016. Sourdough microbial community dynamics: An analysis during French organic bread-making processes. Food Microbiol., 19th Meeting of the Club des Bactéries Lactiques 53, Part A, 41–50. doi:10.1016/j.fm.2014.11.014

Lin, X.B., Gänzle, M.G., 2014. Quantitative high-resolution melting PCR analysis for monitoring of fermentation microbiota in sourdough. Int. J. Food Microbiol. 186, 42–48.

Macé, S., Mamlouk, K., Chipchakova, S., Prévost, H., Joffraud, J.-J., Dalgaard, P., Pilet, M.-F., Dousset, X., 2013. Development of a rapid real-time PCR method as a tool to quantify viable *Photobacterium phosphoreum* bacteria in salmon (Salmo salar) steaks. Appl. Environ. Microbiol. 79, 2612–2619.

Meignen, B., Onno, B., Gélinas, P., Infantes, M., Guilois, S., Cahagnier, B., 2001. Optimization of sourdough fermentation with *Lactobacillus brevis* and baker's yeast. Food Microbiol. 18, 239–245.

Meroth, C.B., Walter, J., Hertel, C., Brandt, M.J., Hammes, W.P., 2003. Monitoring the bacterial population dynamics in sourdough fermentation processes by using PCR-denaturing gradient gel electrophoresis. Appl. Environ. Microbiol. 69, 475–482.

Minervini, F., De Angelis, M., Di Cagno, R., Gobbetti, M., 2014. Ecological parameters influencing microbial diversity and stability of traditional sourdough. Int. J. Food Microbiol. 171, 136–146.

Minervini, F., Lattanzi, A., De Angelis, M., Celano, G., Gobbetti, M., 2015. House microbiotas as sources of lactic acid bacteria and yeasts in traditional Italian sourdoughs. Food Microbiol. 52, 66–76. doi:10.1016/j.fm.2015.06.009

Naser, S.M., Dawyndt, P., Hoste, B., Gevers, D., Vandemeulebroecke, K., Cleenwerck, I., Vancanneyt, M., Swings, J., 2007. Identification of lactobacilli by *pheS* and *rpoA* gene sequence analyses. Int. J. Syst. Evol. Microbiol. 57, 2777–2789.

Rizzello, C.G., Cavoski, I., Turk, J., Ercolini, D., Nionelli, L., Pontonio, E., De Angelis, M., De Filippis, F., Gobbetti, M., Di Cagno, R., 2015. Organic cultivation of Tri*ticum turgidum*

subsp. durum is reflected in the flour-sourdough fermentation-bread axis. Appl. Environ. Microbiol. 81, 3192–3204.

Robert, H., Gabriel, V., Fontagné-Faucher, C., 2009. Biodiversity of lactic acid bacteria in French wheat sourdough as determined by molecular characterization using species-specific PCR. Int. J. Food Microbiol. 135, 53–59.

Scheirlinck, I., Van der Meulen, R., De Vuyst, L., Vandamme, P., Huys, G., 2009. Molecular source tracking of predominant lactic acid bacteria in traditional Belgian sourdoughs and their production environments. J. Appl. Microbiol. 106, 1081–1092.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., Robinson, C.J., others, 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl. Environ. Microbiol. 75, 7537–7541.

Stackebrandt, E., Ebers, J., 2006. Taxonomic parameters revisited: tarnished gold standards. Microbiol. Today 33, 152.

Stolz, P., Vogel, R.F., Hammes, W.P., 1995. Utilization of electron acceptors by lactobacilli isolated from sourdough. Z. Für Lebensm.-Unters. Forsch. 201, 402–410.

Sun, Z., Harris, H.M., McCann, A., Guo, C., Argimón, S., Zhang, W., Yang, X., Jeffery, I.B., Cooney, J.C., Kagawa, T.F., others, 2015. Expanding the biotechnology potential of lactobacilli through comparative genomics of 213 strains and associated genera. Nat. Commun. 6.

Team, Rs., 2015. RStudio: integrated development for R. RStudio, Inc., Boston, MA.

The NCBI Handbook, 2nd ed, 2013. . National Center for Biotechnology Information (US).

Thiele, C., Gänzle, M.G., Vogel, R.F., 2002. Contribution of sourdough lactobacilli, yeast, and cereal enzymes to the generation of amino acids in dough relevant for bread flavor. Cereal Chem. 79, 45–51.

Torriani, S., Felis, G.E., Dellaglio, F., 2001. Differentiation of *Lactobacillus plantarum*, *L. pentosus*, and *L. paraplantarum* by recA gene sequence analysis and multiplex PCR assay with recA gene-derived primers. Appl. Environ. Microbiol. 67, 3450–3454.

Valcheva, R., Ferchichi, M.F., Korakli, M., Ivanova, I., Gänzle, M.G., Vogel, R.F., Prévost, H., Onno, B., Dousset, X., 2006. *Lactobacillus nantensis* sp. nov., isolated from French wheat sourdough. Int. J. Syst. Evol. Microbiol. 56, 587–591.

Valcheva, R., Korakli, M., Onno, B., Prévost, H., Ivanova, I., Ehrmann, M.A., Dousset, X., Gänzle, M.G., Vogel, R.F., 2005. *Lactobacillus hammesii* sp. nov., isolated from French sourdough. Int. J. Syst. Evol. Microbiol. 55, 763–767.

Valmorri, S., Tofalo, R., Settanni, L., Corsetti, A., Suzzi, G., 2010. Yeast microbiota associated with spontaneous sourdough fermentations in the production of traditional wheat sourdough breads of the Abruzzo region (Italy). Antonie Van Leeuwenhoek 97, 119–129.

Venturi, M., Guerrini, S., Granchi, L., Vincenzini, M., 2012. Typing of *Lactobacillus sanfranciscensis* isolates from traditional sourdoughs by combining conventional and multiplex RAPD–PCR profiles. Int. J. Food Microbiol. 156, 122–126.

Vera, A., Rigobello, V., Demarigny, Y., 2009. Comparative study of culture media used for sourdough lactobacilli. Food Microbiol. 26, 728–733.

Wang, Q., Garrity, G.M., Tiedje, J.M., Cole, J.R., 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl. Environ. Microbiol. 73, 5261–5267.

Weisburg, W.G., Barns, S.M., Pelletier, D.A., Lane, D.J., 1991. 16S ribosomal DNA amplification for phylogenetic study. J. Bacteriol. 173, 697–703.

Zeppa, G., Conterno, L., Gerbi, V., 2001. Determination of organic acids, sugars, diacetyl, and acetoin in cheese by high-performance liquid chromatography. J. Agric. Food Chem. 49, 2722–2726.

Zheng, B., Jiang, X., Cheng, H., Xu, Z., Li, A., Hu, X., Xiao, Y., 2015. Complete genome sequence of *Lactobacillus heilongjiangensis* DSM 28069T: Insight into its probiotic potential. J. Biotechnol. 216, 65–66. doi:10.1016/j.jbiotec.2015.09.023

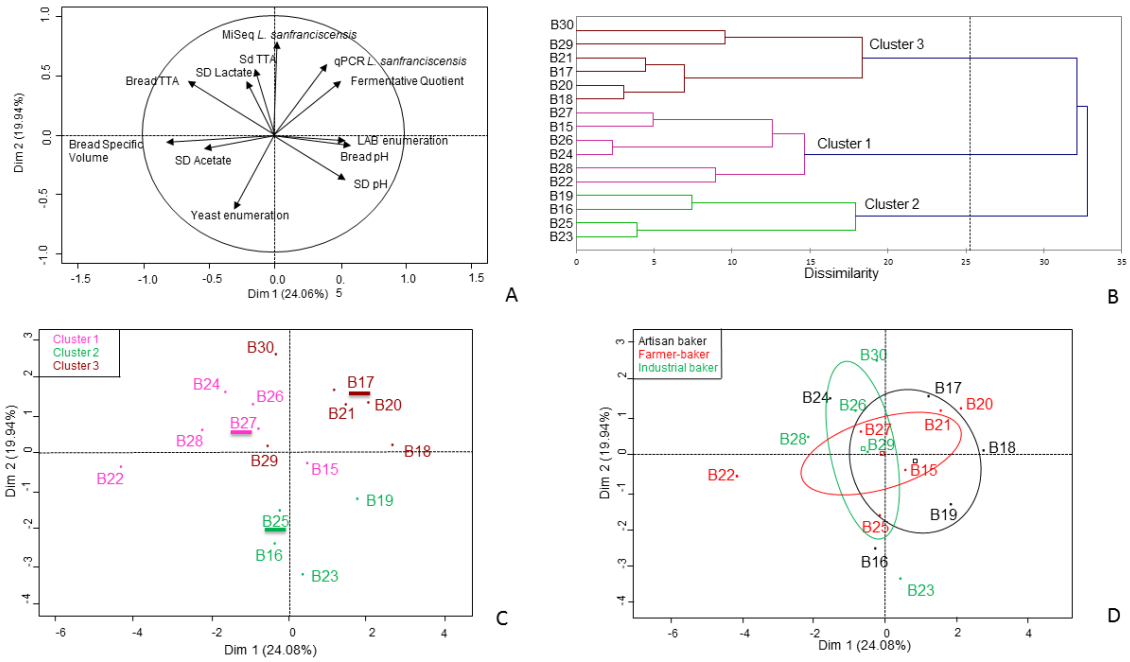

Figure 1

Figure captions

Figure 1: Principal Component Analysis and Hierarchical Ascendant Classification applied to sourdoughs (SD) B15 to B30 and breads characteristics (physicochemical results, yeast and LAB counts and *Lactobacillus sanfranciscensis* qPCR estimation and MiSeq relative abundance)

1A - Variables factor map of first and second principal components after Principal Component Analysis (PCA)

1B - Hierarchical Ascendant Classification (HAC) of sixteen French organic sourdoughs based on physico-chemical and microbiological criteria

1C - Score and loading plots of first and second principal components after Principal Component Analysis (PCA).Underlined data represented centroids samples of HAC clusters: Cluster 1O, Cluster 2O and Cluster 3 O

1D - Confidence ellipses based on bakers' groups

Sourdough	Location of the bakery in France	Bakers' class
B15	Hérault (South)	Farmer baker
B16	Yvelines (Paris area)	Artisan baker
B17	Somme (North)	Artisan baker
B18	Paris	Artisan baker
B19	Aveyron (South)	Artisan baker
B20	Alpes-Maritimes	Farmer baker
B21	Morbihan (West)	Farmer baker
B22	Lot-et-Garonne (South)	Farmer baker
B23	Bas-Rhin (East)	Industrial baker
B24	Vaucluse (South)	Artisan baker
B25	Haut-Rhin (East)	Farmer baker
B26	Côtes-d'Armor (West)	Industrial baker
B27	Puy-de-Dôme (Center)	Farmer baker
B28	Paris	Industrial baker
B29	Aude (South)	Industrial baker
B30	Savoie (East)	Industrial baker

Table 1: Location and classification of the 16 French bakers selected

Primer set	Oligonucléotide sequence	Target gene	Reference
fD1	5'-AGAGTTTGATCCTGGCTCAG-3'	16S rRNA gene	Weisburg *et al*., 1991
rD1	5'-TAAGGAGGTGATCCAGCC-3'		
Sp1	5'-ACTCCTACGGGAGGCAGCA-3'	rDNA 16S position 338 forward	Lhomme *et al*., 2015a
Sp2	5'-ACCGCGGCTGCTGGCACG-3'	rDNA 16S position 514 reverse	
Sp3	5'-GATACCCTGGTAGTCCACG	rDNA 16S position 810 forward	
Sp4	5'-CTCGTTGCGGGACTTAAC-3'	rDNA 16S position 1100	
SP5	5'-GGTACCTTGTTACGACTT-3'	rDNA 16S position 1089 reverse	
702-F	5'-AATTGCCTTCTTCCGTGTA-3'	*katA*	Ammor *et al*., 2005
310-R	5'-AGTTGCGCACAATTATTTTC-3'		
paraF	5'-GTCACAGGCATTACGAAA AC-3'	*recA*	Torriani *et al*., 2001
planF	5'-CCGTTTATGCGGAACACC TA-3'		
pentF	5'-CAGTGGCGCGGTTGATAT C-3'		
pREV	5'-TCGGGATTACCAAACATCAC-3'		
21-F	5'-ATGATYGARTTTGAAAAACC-3'	*rpoA*	Naser *et al*., 2007
23-R	5'-ACYTTVATCATNTCWGVYTC-3'		
LS-1F	5'-TGCAAGCACGTTCTCTAGAAA-3'	*pheS*	Scheirlinck *et al.*, 2009

LS-1R	5’-GAGTTGGATCGTCGGTATCA-3’		
515f	5’-CTTTCCCTACACGACGCTCTTCCGATCT-3’	MiSeq PCR 16S (V4)	Caporaso *et al*., 2011
806r	5’-GGAGTTCAGACGTGTGCTCTTCCGATCT-3’		

Table 2: Oligonucleotides primers used for PCR amplification and sequencing

Sourdough samples	LAB enumeration (log CFU/g)	Yeast enumeration (log CFU/g)	Ratio Yeast/ LAB (%)	Sourdough pH	Sourdough TTA	Sourdough Acetate (g/kg)	Sourdough Lactate (g/kg)	Sourdough Fermentative Quotient
B15	9.42 ± 0.03	7.64 ± 0.08	1.66	3.82 ± 0.01	18.3 ± 6.12	1.57 ± 0.08	3.37 ± 0.02	1.28
B16	9.11 ± 0.04	7.67 ± 0.12	3.63	4.02 ± 0.01	13.4 ± 2.45	ND	ND	ND
B17	9.16 ± 0.10	6.70 ± 0.08	0.35	3.84 ± 0.01	15.0 ± 0.07	0.91 ± 0.02	8.71 ± 0.38	5.66
B18	9.44 ± 0.11	6.99 ± 0.08	0.36	3.98 ± 0.01	14.4 ± 0.32	1.41 ± 0.09	8.74 ± 0.16	3.70
B19	9.62 ± 0.19	7.01 ± 0.04	0.25	4.03 ± 0.01	13.3 ± 0.31	1.53 ± 0.07	5.29 ± 0.07	2.06
B20	9.60 ± 0.10	6.78 ± 0.01	0.15	3.87 ± 0.01	19.8 ± 0.32	ND	ND	ND
B21	8.97 ± 0.14	7.20 ± 0.07	1.69	3.70 ± 0.01	18.1 ± 0.71	1.22 ± 0.01	9.17 ± 0.12	4.47
B22	8.09 ± 0.03	7.68 ± 0.10	38.9	3.70 ± 0.01	14.9 ± 0.3	4.44 ± 0.05	7.62 ± 0.04	1.06
B23	9.02 ± 0.03	7.69 ± 0.12	4.68	3.86 ± 0.00	14.5 ± 0.16	2.62 ± 0.04	7.82 ± 0.07	1.99
B24	8.81 ± 0.04	7.58 ± 0.14	5.89	3.85 ± 0.01	21.3 ± 1.26	2.48 ± 0.01	11.96 ± 0.07	2.87
B25	8.62 ± 0.05	7.43 ± 0.16	6.46	3.80 ± 0.01	14.1 ± 0.16	1.63 ± 0.03	7.10 ± 0.11	2.59
B26	9.29 ± 0.04	7.45 ± 0.13	1.45	3.77 ± 0.01	18.0 ± 0.00	2.19 ± 0.09	10.89 ± 0.13	2.97
B27	9.46 ± 0.01	7.52 ± 0.07	1.15	3.62 ± 0.00	16.5 ± 0.15	1.75 ± 0.03	7.86 ± 0.28	2.66
B28	8.54 ± 0.09	6.74 ± 0.02	1.58	3.64 ± 0.00	19.2 ± 0.16	3.34 ± 0.33	8.31 ± 0.79	1.48
B29	7.19 ± 0.09	7.01 ± 0.02	66.07	3.83 ± 0.00	14.8 ± 0.15	1.62 ± 0.07	8.09 ± 0.01	2.97
B30	8.38 ± 0.35	5.92 ± 0.06	0.35	3.81 ± 0.06	14.8 ± 1.18	1.09 ± 0.02	8.50 ± 0.16	4.65

Table 3: Total LAB and yeast enumeration (log CFU/g); physicochemical characteristics of sourdoughs and related breads (mean values ± standard deviation). ND : Not determined

Sourdough samples	Quantification of *L. sanfranciscensis* (log CFU/g) by qPCR	Ratio qPCR results/ LAB count	Species relative abundance (%) by NGS Miseq	Species identified in sourdough by biomolecular method (%)
B15	10.29	1.09	*L. sanfranciscensis* (99.8 %), *Lactobacillus sp.* (0.1%)	*L. sanfranciscensis* (86%), *L. plantarum* (14%)
B16	8.41	0.92	*L. sanfranciscensis* (0.1%), *Lactobacillus sp.* (12.0%), *L .curvatus* (75.2%), *L. sakei* (0.7%), *Weissella* (0.2%)	*L. curvatus* (90%), *L. plantarum* (10%)
B17	9.27	1.01	*L. sanfranciscensis* (99.8%)	*L.sanfranciscensis* (100%)
B18	9.47	1.01	*L. sanfranciscensis* (99.9%)	*L.sanfranciscensis* (100%)
B19	9.58	0.99	*L. sanfranciscensis* (99.9%)	*L.sanfranciscensis* (100%)
B20	10.06	1.05	*L. sanfranciscensis* (99.7%)	*L.sanfranciscensis* (100%)
B21	9.48	1.05	*L. sanfranciscensis* (99.8%), *Lactobacillus sp.* (0.1%), *L. koreensis* or *L. brevis* (0.1%)	*L.sanfranciscensis* (54%), *L. plantarum* (8%), *L. xiangfangensis* (15%), *L. brevis* (23%)
B22	7.85	0.97	*L. sanfranciscensis* (99.9%)	*L.sanfranciscensis* (100%)
B23	4.41	0.43	*L. koreensis* or *L. brevis* (80.0%), *L. heilongjiangensis* or *L. farminis* or *L. crustorum* (16.7%), *L. nantensis* (0.8%), *Leuconostoc (order)* (0.4%), *L. sanfranciscensis* (0.2%),	*L. heilongjiangensis* (37.5%), *L. koreensis* (25%), *L. plantarum* (25%), *L. curvatus* (12.5%),
B24	8.03	0.91	*L. sanfranciscensis* (99.9%)	*L sanfranciscensis* (100%)
B25	7.40	0.85	*Lactobacillus sp.* (61.9%), *L. sanfranciscensis* (36.1%), *L. brevis* (0.9%)	*L. brevis* (69%), *L. plantarum* (15%), *L. sanfranciscensis* (8%), *L. paralimentarius* (8%)
B26	8.99	0.97	*L. sanfranciscensis* (98.6%), *Lactobacillus sp.* (0.6%), *L. parabrevis* (0.6%)	*L.sanfranciscensis* (75%), *L. plantarum* (25%)
B27	9.09	0.96	*L. sanfranciscensis* (99.8%)	*L.sanfranciscensis* (100%)
B28	7.19	0.84	*L. sanfranciscensis* (88.8%), *L. pontis* (4.3%), *L. parabrevis* (2.4%), *L. buchneri* (0.8%), *L. diolivorans* (2.3%)	*L. sanfranciscensis* (40%), *L. pontis* (33%), *L. diolivorans* (13%), *L. paracasei* (7%), *L. parabuchneri* (7%)
B29	8.23	1.15	*L. sanfranciscensis* (99.6%), *L. koreensis or L. brevis* (0.2%)	*L. plantarum* (9%) , *L. xiangfangensis* (55%), *Leuconostoc citreum* (36%)
B30	9.46	1.13	*L. sanfranciscensis* (99.2%), *Weissella* (0.4%), *Pediococcus pentosaceus* (0.2%), *Lactobacillus sp.* (0.1%), *L. koreensis or L. brevis* (0.1%)	*P. pentosaceus* (55%), *Weisella confusa* (27%), *L. sanfranciscensis* (18%)

Table 4: Microbial diversity of 16 French organic sourdoughs. Estimation of *L. sanfranciscensis* by qPCR (log CFU/g), relative abundance (expressed as %) of bacterial taxonomic units by the MiSeq method and isolate identification by 16S rDNA sequencing (number of isolates per species / total number of isolates per sourdough expressed as %)

Class	N LAB (log CFU/g)	N Yeast (log CFU/g)	qPCR *L. sanfranciscensis* (log CFU/g)	Relative Abundance *L. sanfranciscensis* (%)	Sourdough pH	Sourdough TTA	Sourdough Acetic acid (g/kg)	Sourdo Lactic a (g/kg
1	8.34	7.44	8.57	97.8	3.73	18.03	2.63	8.34
2	9.09	7.45	7.45	34.08	3.93	13.83	1.83	6.38
3	8.79	6.77	9.33	99.67	3.84	16.15	1.28	8.66

Table 5: Cluster class centroids obtained by Hierarchical Ascendant Classification based on bread and sourdough physico-chemical and microbiological criteria