Title: Genotypic and phenotypic analysis of dairy *Lactococcus lactis* biodiversity in milk: Volatile Organic Compounds as discriminating markers.

**Running title:** Dairy *Lactococcus lactis* diversity in fermented milks

Amandine DHAISNE$^{1,2,3}$, Maeva GUELLERIN$^{4,5}$, Valérie LAROUTE$^{1,2,3}$, Sandrine LAGUERRE$^{1,2,3}$, Muriel COCAIGN-BOUSQUET$^{1,2,3, #}$, Pascal LE BOURGEOIS$^{4,5}$, Pascal LOUBIERE$^{1,2,3}$

$^1$ Université de Toulouse; INSA, UPS, INP; LISBP, 135, avenue de Rangueil, 31077 Toulouse, France, $^2$ INRA, UMR792 Ingénierie des systèmes biologiques et des procédés, 31400 Toulouse, France $^3$ CNRS, UMR5504, 31400 Toulouse, France

$^4$ Laboratoire de Microbiologie et Génétique Moléculaires, UMR5100, CNRS.

$^5$ Université de Toulouse ; Université Paul Sabatier, Toulouse, France.

# Corresponding author: Muriel Cocaign-Bousquet

Tel: +33 561 55 94 65

E-mail: cocaign@insa-toulouse.fr

Keywords: *lactococcus lactis*, dairy, biodiversity, phenotype, genotype, Volatile Organic components (VOCs), fermented milk.
ABSTRACT

The diversity of nine dairy strains of *Lactococcus lactis* subsp. *lactis* was investigated in fermented milks by both genotypic and phenotypic analyses. Pulsed field gel electrophoresis [PFGE] and multilocus sequence typing [MLST] were used to establish an integrated genotypic classification. This classification was coherent with discrimination of the biovar *diacetylactis* lineage, and reflected Clonal Complex phylogeny and the uniqueness of the genomes of these strains. To assess phenotypic diversity, 82 variables were selected as important dairy features; they included physiological descriptors, and the production of metabolites and volatile organic compounds (VOCs). Principal component analysis (PCA) demonstrated the phenotypic uniqueness of each of these genetically closely related strains, allowing strain discrimination. A method of variable selection was developed to reduce the time-consuming experimentation. We therefore identified 20 variables, all associated with VOCs, as phenotypic markers allowing discrimination between strain groups. These markers are representative of the three metabolic pathways involved in flavour: lipolysis, proteolysis and glycolysis. Despite large phenotypic diversity, the strains could be divided in four robust phenotypic clusters depending on their metabolic orientations. Inclusion of genotypic diversity in addition to phenotypic characters in the classification led to five clusters rather than four being defined. However, genotypic characters have a smaller contribution than phenotypic variables (no genetic distances selected among the most contributory variables). This work proposes an original method for phenotypic differentiation of closely related strains in milk and may be the first step towards a predictive classification for the manufacture of starters.
The mesophilic lactic acid bacterium *Lactococcus lactis* is one of the most extensively exploited micro-organisms: it is used in particular in the manufacture of dairy products. Because of its industrial importance, *L. lactis* has been used as a model bacterium for academic and application-oriented studies. Taxonomically, *L. lactis* has three subspecies, *L. lactis* subsp. *horniae*, *L. lactis* subsp. *lactis* (including the biovar *diacetylactis*) and *L. lactis* subsp. *cremoris* (1). Only *L. lactis* subsp. *lactis* and *L. lactis* subsp. *cremoris* are used in starter cultures for dairy production. *L. lactis* subsp. *lactis* colonises a wide ecological niche (dairy products, as well as animals and plant surfaces) (2), whereas *L. lactis* subsp. *cremoris* is mostly found in dairy environments (3, 4). Recent genotyping methods involving DNA fingerprinting analysis, including pulsed-field gel electrophoresis (PFGE) and comparative genomic hybridization (CGH), have been evaluated for their suitability for characterising genomic diversity and identifying reliable genetic markers of the phenotypic subspecies differentiation (5–9). Multilocus sequence typing (MLST), at the level of the gene, has proved to be a powerful method for describing *L. lactis* population structure and phylogeny with a limited number of genes sequenced (5, 6). Recent MLST analysis of the *lactis* subspecies (6) confirmed *L. lactis* population adaptation to the milk niche, and led to the proposal of a new classification into two ecotypes: one corresponding to “domesticated” strains essentially isolated from dairy starters or fermented products, and the other corresponding to “environmental” strains isolated from various sources such as plants, animals, and raw milk. The “domesticated” strains make only a small contribution to the genetic diversity of the *lactis* subspecies; indeed, phylogenetic analysis indicates that they essentially form two clonal complexes (CC), which probably emerged only recently from a single founder event (6).

The characteristics of the *L. lactis* subsp. *lactis* used in starter cultures for fermented milk dairy production determine their dairy phenotype which is related to an efficient growth...
in milk associated to fast coagulation. Coagulation is due to acidification and preserves milk from unwanted bacteria or mould growth. However *L. lactis* subsp. *lactis* also contributes to organoleptic quality (texture and sensory) of the fermented products. Rather than being based on phenotypic standards for subspecies classification, strain selection has focused on diverse but specific technical characteristics: acidification (10), redox potential (11), texture (12), and physiological features like aroma production (13, 14) and nisin synthesis (15), autolysis (16, 17) and enzymatic activities (18). In these previous studies, selection has been restricted to one or a limited number of criteria, and the behaviour of the strains as a whole cannot be predicted. Phenotypic starter screening requires exhaustive study in a controlled dairy environment to evaluate strain performances, but this is a very complex and time-consuming approach, incompatible with screening strain collections. Genotyping and genomic sequencing are efficient and have now been largely automated; there was an expectation that this approach could be used to classify and discriminate between strains, and to predict their phenotypes (19). Several studies combining genomic, transcriptomic and phenotypic data for *L. lactis* have been conducted in recent years (20, 7, 21, 22, 8, 23) but they have failed to establish strong exploitable links between genotype, its expression and dairy phenotypes. Rather, they have highlighted significant transcriptomic polymorphisms among *Lactococcus lactis* subsp. *lactis* strains with the same dairy origin (21) and the difficulties associated with explaining and predicting a phenotype other than the presence or absence of relations involving the gene (19, 20). Understanding the determinants of the phenotypic diversity of dairy strains is a prerequisite for predictive approaches and requires phenotypic screening in environmental conditions that are as close as possible to those encountered during the process of interest (18).

Therefore, this work reports an integrated approach to assess the phenotypic biodiversity of nine *L. lactis* subsp. *lactis* strains exhibiting dairy phenotype and selected as
being representative of the “domesticated” ecotype in milk. Strain phenotypic signatures were compared with strain genotypic diversity, revealing discrepancies between the two classifications, highlighting the need for an integrated genotypic and phenotypic classification that takes into account both aspects.

**MATERIALS AND METHODS**

**Bacterial strains and growth conditions**

Nine strains of *Lactococcus lactis* subsp. *lactis* were selected from commercial starter cultures (Danisco-France), and from various laboratory collections (LMGM-Toulouse, France for Sx and EIPx strains; LMA-Caen, France for UCMAx strains). Strain origins and characteristics are listed in Table 1. Stock cultures were kept frozen (-20°C) in M17 broth (Oxoid Ltd., Basingstoke, Hampshire, England) containing 2% lactose (wt/vol) and 20% (vol/vol) glycerol. Individual strains were first grown at 37°C in sterilised (110°C, 15 min) reconstituted skimmed milk (10.9% wt/vol). These overnight cultures were used to inoculate (10%, vol/vol) 500 mL of pasteurised standardised cow’s milk (fat: 35 g/kg, protein: 40 g/kg) in 1 L flasks. The pH was unregulated (initial value of 6.60). Fermentations were continued at 28 °C until the pH reached a value of 4.60, and then transferred to 4°C. The fermented milks were stored in the dark at 4°C for 1 (sample A) or 14 (sample B) days before analysis. Six biological replicates (independent batches) were performed for each strain.

**Growth measurements and acidification activity**

Bacterial growth in milk (μ, h⁻¹), as well as cell cultivability after 1 and 14 days of storage, were determined in triplicate by plating appropriate dilutions onto M17 agar, incubating at 30°C for 48 h and counting colonies (cfu.mL⁻¹). Cell cultivability was calculated as the ratio between viable counts after 14 days to that after 1 day.

Acidification activity of lactococcal cells was monitored using the CINAC system (Alliance Instrument, France). For each culture, four descriptors were defined to characterise the
acidification activity during fermentation: the maximal acidification rate ($V_{\text{max}}$), expressed in $10^{-3}$ pH units per min [UpH.min$^{-1}$] and calculated from the slope of the pH curve as a function of time; the time necessary to reach $V_{\text{max}}$ ($T_{V_{\text{max}}}$, expressed in h); the pH at which $V_{\text{max}}$ is attained ($pH_{V_{\text{max}}}$, expressed in UpH); and the time necessary for the pH to decrease to 4.60 ($T_{pH4.60}$, expressed in h).

Fermentation analysis

Substrate (lactose and citrate) and fermentation end-product (lactate, formate, acetate, acetoin, diacetyl, and ethanol) concentrations in fermented milk after 14 days of storage were determined by high-pressure liquid chromatography (HPLC) as previously described (24). Briefly, a solution of 0.3 M barium hydroxide and zinc sulphate was used to precipitate proteins from the samples; then the samples were analysed in a Bio-Rad HPX87H column maintained at a temperature of 48 °C in a 1200 Series preparative HPLC (Agilent Technologies, Waldbronn, Germany), and 5 mM H$_2$SO$_4$ was used as the eluent, at a flow rate of 0.5 mL.min$^{-1}$. Free amino acid concentrations in milk culture supernatants at 14 days were measured by HPLC as follows: proteins in the samples were precipitated by adding 4 volumes of methanol to 1 volume of the sample and incubating the mixture overnight at 4°C. The mixture was then centrifuged, and the supernatant was kept for amino acid analysis as previously described (25). The amino acids were automatically derived with ortho-phthalic aldehyde (OPA) and 9-fluorenylmethyl chloroformate (FMOC) and the derivatives were separated on a Hypersil AA-ODS column (Agilent Technologies, Waldbronn, Germany) at 40°C by using a linear gradient of acetate buffer (pH 7.2) with triethylamine (0.018%) tetrahydrofuran (0.03%) and acetonitrile (60%). A diode array detector was used at 338 nm for OPA derivatives and at 262 nm for FMOC derivatives. These analyses were performed on the six biologically independent batches for each strain.
Identification and semi-quantification of volatile organic compounds (VOCs) in fermented milk

Volatile organic compounds in fermented milk were identified and semi-quantified by the Analytical Chemical Service of ISVV (Institut des Sciences du Vin et de la Vigne)-Bordeaux-France. The separation and semi-quantification were carried out by solid-phase-microextraction (SPME)/gas chromatography (GC)/mass spectrometry (MS) analysis.

Internal standards used for gas chromatography analysis were [2H5] ethyl acetate (ethyl acetate-d5) and [2H15] octanoic acid (octanoic acid-d15) from Sigma–Aldrich (Saint-Quentin-Fallavier, France), and [2,2,2-2H] 3-methylbutanal (3-methylbutanal-d2) from Euriso-top (Saint-Aubin, France). Solutions were prepared at 100 mg.L⁻¹ with milliQ water and kept at 4°C.

A 5 g sample of fermented milk was mixed with 250 μL of internal standard mix solution (at 10, 20 and 20 mg.L⁻¹ for ethyl acetate-d₅, octanoic acid-d₁₅ and 3-methylbutanal-d₂, respectively) in a 20 mL glass sample vial. Then, the vial was tightly sealed and the contents homogenised with a vortex shaker at 55°C for 30 min. The solid-phase-microextraction (SPME) fibre (Supelco, Bellefonte, PA, USA) used was coated with a 50/30 μm layer of divinylbenzene-carboxen-polydimethylsiloxane (57299-U, StableFlex) and maintained in the headspace at 55°C for 90 min. The volatile analytes absorbed to the SPME fibre were analysed by GC-MS using a HP 5890 GC system coupled to a HP 5972 quadrupole mass spectrometer (Agilent technologies, Wilmington, DE) equipped with a Gerstel MPS2 autosampler. The compounds were separated on a capillary column BP 21 (60 m, 0.32 mm, 0.25 μm film thickness, SGE, Courtaboeuf, France). The injection port was programmed to heat at 250°C. The oven temperature was programmed at 40°C for 5 min, then raised to 240°C at 3°C.min⁻¹, and held at that temperature for 10 min. Volatile compounds were identified in SCAN mode (mass range from 29 to 200 m/z using a scan rate of 1.49 scans.seconds⁻¹) and the ionization energy was set at 70eV. Only compounds for which area...
was at least twofold the background noise were selected. The eluted compounds were identified by their retention times and by comparison of their mass spectra with those in the NIST database (http://www.nist.gov/srd/ version of July 2002).

Selected volatile compounds were detected and semi-quantified with the mass spectrometer operating in selected-ion-monitoring (SIM) mode with electron ionization at 70 eV. VOCs were semi-quantified by determining the ratio of the Total Ion Count (TIC) of the compounds to the TIC of their corresponding internal standard for one gram of fermented milk. The internal standard was chosen according to the chemical function and on the retention time.

For each strain, analytical triplicates were performed. Detection and quantification limits, and coefficients of variation for each compound were determined.

**DNA manipulation and phylogenetic analysis**

Bacteria were grown at 30 °C on M17 broth. Genomic DNA was extracted from overnight cultures using the “DNeasyTM tissue” kit according to the manufacturer’s instructions (Qiagen, Hilden, Germany). Genetic markers of important industrial traits (*lacE*, encoding the lactose-specific Enzyme II of the PTS system; *prtP*, encoding the cell envelope associated serine proteinase; and *citP*, encoding the membrane bound citrate permease involved in citrate uptake) were detected by PCR amplification and standard agarose gel electrophoresis. PCR primers for *lacE* (*lacE*-F2, 5’-AGCGTCTATGTTAGGGTCC-3’; *lacE*-R2, 5’-GATGGCACGGTTACGATCTG-3’; PCR product size of 606 bp), *prtP* (*prtP*-F2, 5’-GAGGCAGTGAAACTGTTAGTC-3’; *prtP*-R2, 5’-TCATTCGCAGCAGTACATC-3’; PCR product size of 713 bp), and *citP* (*CitP*1, 5’-ATGATGAATCACCCG-3’; *CitP*2, 5’-ACTTCATGAATATGAC-3’; PCR product size of 1327 bp) genes were designed by standard procedures using Clone Manager version 9.0 software (Sci-Ed Software). Cycling conditions were 95°C for 5 min followed by 30 cycles of 95°C for 30 s, 55°C for 1 min and 72°C for 1 min for *lacE* and *prtP* genes, and 98°C for 5 min followed by 30 cycles of 98°C
for 10 s, 42°C for 30 s and 72°C for 1 min for citP gene, using a MJ Mini thermocycler (Bio-Rad, Hercules, USA); each PCR involved a 25 µL-mixture containing 10 ng of genomic DNA, 0.5 mM of each primer, and 12.5 µL of iQ™ SYBR® Green Supermix (Bio-Rad, Hercules, USA).

Multilocus Sequence Typing (MLST) analysis was performed exactly as described in (6), using the 6-locus MLST scheme described. The concatenated sequences generated were used for phylogenetic analysis with MEGA5 software (26).

For pulsed field gel electrophoresis (PFGE), preparation of genomic DNA embedded in agarose matrix, digestion of DNA by restriction endonucleases, and electrophoresis were performed as previously described (27). The genomic relatedness of bacterial strains was estimated from pairwise comparisons of PFGE SmaI-macrorestriction patterns, and a matrix of binary data was constructed from the presence/absence of each band. Dice coefficients (SD) and corresponding genomic distances (1-SD) for each pairwise comparison were calculated from the matrix of binary data using the WINDIST program (28). UPGMA dendrograms were constructed with the NEIGHBOR program of the PHYLIP package v3.69 (29).

**Data Analysis**

All variables (physiological descriptors, metabolite concentrations and genetic distances) were first normalised (centring: subtracting the population mean and then scaling: dividing this difference by the standard deviation of the population) to allow an unbiased comparison of these heterogeneous data. Missing data in COVs semi-quantification for a third of biological replicates were filled in by applying the geometrical mean of the available replicates for that strain. To investigate the relationships between strains and the phenotypic variables of the fermented milks, an analysis of variance (ANOVA) was performed for the 82 phenotypic variables. The level of significance for all statistical analyses was set to a $p$-value cut-off of 0.05. Spearman’s rank-order correlations with $p$-value adjusted by Benjamini and
Hochberg to control the false discovery rate (FDR) (30) were performed to investigate pairwise associations between the variables involved in strain phenotype in milk. Multidimensional methods such as Principal Component Analysis (PCA), Hierarchical Ascendant Classification (HAC), Partial Least Squares Discriminant Analysis (PLS-DA) and Sparse PCA, were carried out using R free statistical software (2.15.0 version) and the mixOmics package (31). The Euclidian distance metric and the Ward’s criterion were used for HAC in phenotypic and genotypic classifications to constitute hierarchical groups of mutually exclusive subsets in which members are maximally similar with respect to their specified characteristics (32). Bootstrap analysis was applied with 1000 simulations with cluster analyses. Consensus tree were built by bootstrapping (bootstrap value > 0.50). To identify discriminatory variables, a variable selection method was developed based on the principal components, PCs. The variable selection combined a test of strain dependency (removing of one strain from the sample in order to differentiate markers highly dependent of a strain from common markers) and a selection of the most contributory variables for this set of individual. The contribution of a variable for the selected principal components is obtained by the ratio of the sum of the squared factor score divided by the sum of the eigenvalues. The contributory threshold was set to 75% for this procedure, therefore only variables with a contribution above the third quartile of the contributory distribution were selected as most contributory variables. This procedure was done iteratively for each strain and a strain clustering was then performed based on the selected variables. The cluster robustness and consequently, the reliability of the selected variable set were assessed by bootstrapping. In this work, the selected variables represented the 25% of total variables (i.e. twenty variables for phenotypic analyses, and twenty-two for phenotype-genotype investigation, respectively) with the highest headcount among each set of the most contributory variables. In order to provide a deeper
analysis of the robustness of the developed variable selection method, results were compared
to variable selection through sparse PCA (33).

RESULTS

I. Genotypic characterisation of L. lactis subsp. lactis strains

To investigate the phenotypic diversity of L. lactis subsp. lactis strains in growth
conditions close to those of relevant industrial processes, nine strains isolated from raw milk
and starter cultures (Table 1) were selected according to their dairy phenotype as pure culture
ability to grow efficiently in milk inducing fast coagulation). All these strains were
considered to belong to the “domesticated” ecotype (6). The following genetic markers of
industrial traits were checked by PCR amplification: the lacE gene, encoding the lactose-
specific Enzyme II of the PTS system, and the prtP gene, encoding the cell envelope-
associated serine protease. Both were present in all nine strains. L. lactis subsp. lactis biovar
diacetylactis isolates are widely used in the dairy industry because of their ability to take up
and consume citrate to produce diacetyl and acetoin, two components essential for creamy
and buttery aroma (8, 18, 22). Four strains of our panel were assigned to the biovar
diacetylactis based on results of PCR amplification of citP, a plasmid-borne gene encoding
the permease involved in citrate uptake (34).

The “domesticated” L. lactis subsp. lactis currently contains 13 sequence types (ST)
organised into two main Clonal Complexes (CC) (https://www-
mlst.biotoul.fr/Lactococcuslactissubsplactis/). The main CC (CC1) contains 11 STs (ST1, ST6, ST9, ST10, ST15, ST16, ST18, ST22, ST23, ST34, and ST36), with ST15 predicted to
be the ancestor genotype, whereas CC2 comprises ST7 and ST4. MLST analysis revealed that
eight of the nine strains used in this study belonged to CC1 (4 strains from ST15, 2 strains
from ST34, 1 strain from ST10, and 1 strain from ST18) and that one was ST7 in CC2 (Table
We therefore considered our strain sample as genetically representative of the “domesticated” ecotype. As strain redundancy in bacterial collections cannot be excluded, especially for strains belonging to the same ST, Smal-macrorestriction analysis and PFGE were used to confirm that each of these strains was unique, and different from the other eight. The genomic relatedness of the selected strains was assessed by computing Dice coefficients (SD) from pairwise comparisons of the Smal-macrorestriction patterns (Fig. S1A, provided as supplementary data): 52 % of the calculated SD values were lower than 0.35, the value observed when comparing the two different subspecies cremoris and lactis (6). These findings confirmed previous descriptions of the substantial genomic variability within L. lactis subsp. lactis (6). The consensus tree built from UPGMA-based clustering of PFGE distances yielded to three major clusters (bootstrap value > 0.50) and clustered the four strains from the biovar diacetylactis together (bootstrap value = 0.61) (Fig. S1B, provided as supplementary data). However, strains from the same ST or the same CC were not necessarily grouped together.

We thus considered unsupervised hierarchical clustering integrating both genetic (MLST) and genomic (PFGE) datasets as an alternative approach to improve strain clustering by taking into account genotypic aspects. For this purpose, Hierarchical Ascendant Classification (HAC) with the Ward’s criterion was performed to constitute hierarchical groups of mutually exclusive subsets (32). The accuracy of the genotypic dendrogram was assessed by bootstrapping and the consensus tree was built (bootstrap value > 0.50) (Fig. 1). This integrated genotype classification is coherent with the mean features of gene phylogeny, biovar diacetylactis discrimination, and strain genomic uniqueness (Fig. 1). However, it failed to group strains from ST15 in one class. This last result can be explained from evolutionary concepts. Knowing that clonal diversification of lactococcal strains is mostly dependent on genome rearrangements that have large effects on PFGE fingerprints (35), the ST15 may be
more prone to macrorestriction polymorphism than other STs of its clonal complex (CC1)
because it is the ancestor genotype.

II. Phenotypic biodiversity and subpopulation structure

II.1. Phenotypic dataset

Eighty-two variables selected as being descriptive of dairy performance, including
physiological indicators (growth, acidification) and extra-cellular metabolic products (sugars,
free amino acids, organic acids and volatile organic compounds: VOCs) were assayed for the
nine L. lactis subsp. lactis strains grown in pasteurised standardised cow’s milk. Data for
some physiological indicators (acidification and growth kinetics) were collected at various
times during the milk fermentations, but most were collected after 14 days of storage
(corresponding to the estimated half-life of fermented milks). In all conditions, growth started
immediately after inoculation with maximal growth rates ranging from 0.60 to 0.80 h\(^{-1}\). No
significant difference was found for cell populations: there were 2.1 \(\times\) \(10^9\) to 2.6 \(\times\) \(10^9\)
CFU.mL\(^{-1}\) at the end of the culture (when the pH reached 4.60). The cell cultivability after 14
days was between 50 and 70 percent, confirming the growth similarities for all nine strains.

Acidification properties of the nine strains were compared. As expected, the mean
acidification rate for the four biovar diacetylactis strains was 40% lower than that for the
other strains; this was due to a fourfold higher acetoin concentration and a lower lactic acid
production by the biovar diacetylactis strains. The time taken for the fermentation to reach the
pH of 4.60 was two hours longer for the biovar diacetylactis strains than the other strains; a
delay considered to be significant for the dairy industry. Supernatants from 14-day fermented
milks were assayed for 18 free amino acids (FAAs): all were detected and their concentrations
could be quantified. However, the biogenic amine \(\gamma\)-amino butyric acid (GABA) was detected
only for three strains: EIP33A (0.1 mM), DIA-A and MC70 (1 mM). A total of 47 different
Volatile Organic Compounds (VOCs), belonging to the hydrocarbon, alcohol, aldehyde,
ketone, ester, sulphide compound and free fatty acid (FFA) families, were identified and quantified in the samples. To investigate the relationships between strains and the phenotypic variables of the fermented milks, an Analysis Of Variance (ANOVA) was performed for the 82 variables. This analysis revealed that only VOCs were strongly dependent on the strains ($p$-value < 0.05 for the ANOVA test).

**II.2. Metabolic exploration of fermented milk by correlation investigations**

The analysis of these 82 scaled variables as a single integrated phenotypic dataset allowed exploration of the metabolic network in fermented milks. Spearman’s rank-order correlations after $p$-value multiple testing corrections by Benjamini and Hochberg ($p$-value cut-off of 0.05) were then performed to investigate pairwise associations between the variables involved in strain signatures in milk (Fig. S2 provided as supplementary data). The principal finding was the strong correlation ($R \geq 0.70$) between VOCs of the free fatty acids (C$_4$ to C$_9$:2), methyl-ketone (2-undecanone, 2-tridecanone, 2-pentadecanone), aldehyde (hexadecanal), alcohol (heptanol, octanol, nonanol) and lactone ($\delta$-octalactone, $\delta$-decalactone, $\delta$-dodecalactone) families (Fig. S2A). All these metabolites are involved in the four main steps of lipolysis: release of free fatty acids (FFAs) from milk triglycerides, followed by $\beta$-oxidation of released FFAs to $\alpha$-keto acids, then decarboxylation to alkan-2-ones and finally reversible reduction to the corresponding alkan-2-ols (13, 36–38). The strong relationships between these metabolites confirmed that they are involved in the same pathway.

The metabolites derived from glycolysis and more particularly those involved in pyruvate metabolism (acetalddehyde, lactic acid, acetoin, diacetyl, and butanediol) displayed strong positive correlations with each other ($R \geq 0.67$). The contribution of citrate consumption to the production of diacetyl, acetoin, acetic and lactic acids, and butanediol was confirmed by strong negative correlations ($R \leq -0.75$) (Fig. S2B).
Several correlations related to proteolysis were also identified. Firstly, the catabolic link between benzaldehyde and its amino acids precursor, phenylalanine, was confirmed by strong negative correlation (R < -0.70). Metabolites of methionine catabolism (sulphide and thiol VOCs) were positively correlated (Fig. S2B). Lastly, 3-methylbutanal and 2-methylpropanol, the main products of branched chain amino acid catabolism, were positively correlated (R = 0.61).

II.3. High strain phenotypic diversity

A Principal Component Analysis (PCA) of the fifty-four fermented milks (6 cultures for each of the 9 strains) was used to study phenotypic strain diversity. This analysis was limited to the three first Principal Components (PCs) of the PCAs, because the cumulated variance analysis of PCs 1 to 3 accounted for 67.8 %, percentage sufficiently high to ensure that the PCA plots were representative of the main features in the dataset. Based on visual explorations of the score plots (Fig. 2A and 2B), nine independent signatures, i.e. one per strain, were observed. This result illustrates the high phenotypic diversity of genetically closely related strains, all exhibiting a dairy phenotype (efficient growth in milk associated to fast coagulation). This result also confirms the suitability of the 82 monitored variables for establishing a discriminatory phenotypic dataset for investigating the diversity of “domesticated” L. lactis subpopulation.

Thus, each of these nine strains has a unique phenotypic signature. We classified them according to their metabolic proximity. Although the PCA score plots allowed intuitive strain grouping, a clustering analysis was preferred to allow the robustness of the strain classification to be assessed. The 82 variables were used to build a dendrogram by hierarchical clustering (Fig. 3). The accuracy of the classification was checked by bootstrap analysis, which revealed that all biological replicates of a strain (six independent experiments) were repeatedly grouped together (bootstrap value > 0.94) confirming the quality of the data.
Thus, our integrated phenotypic approach allows robust and accurate strain identification and discrimination by analysis of fermented milk. It would be interesting to analyse how these nine strains could be grouped together based on the proximity of their phenotypes. To choose the relevant number of strain clusters in this classification, two criteria were examined: the decrease of inter-class variance (data not shown) and bootstrap analysis (Fig. 3). However, with these criteria, four, five and nine clusters were possible options, without it being possible to conclude as to which was the most appropriate. Consequently, further investigations were required to determine the number of clusters and by extension the organisation of “domesticated” L. lactis subsp. lactis subpopulation.

II.4. Subpopulation organisation and variable selection

To identify the relevant number of strain clusters for the “domesticated” L. lactis subsp. lactis subpopulation, a robustness analysis was carried out. This strain cluster analysis also aimed to reduce the number of variables used from 82, since strain discrimination is no longer required. This approach allowed the identification of the most significant phenotypic markers. We focused our study on the variables making the largest contribution: those most responsible for strain positioning. From the variable selection (details in materials and methods), twenty variables, corresponding to 25% of the 82 initial variables was found to be the minimal number of most contributory variables allowing robust strains clustering in each case (data not shown). Four strains clusters were thereby defined (bootstrap value > 0.82), and named I to IV. Their compositions are shown in a dendrogram (Fig. 3). A supplementary Partial Least Squares Discriminant Analysis (PLS-DA) was then applied to estimate the error rate of strain affiliation to the four clusters identified. Using only the 25% of variables that contributed the most to this clustering, a perfect accuracy (100%) was predicted in strain cluster affiliation.
Thus, 20 of the 82 variables studied were identified as the most contributory. Variable selection through sparse PCA confirmed the suitability of the set of discriminatory variables. All these selective variables were related to VOCs and were classified into three main groups (A-B-C) (Table 2). These groups are representative of the three major catabolic pathways involved in flavour formation, consistent with the findings of the correlation analysis. Group A is related to glycolysis (and citrate products except for nonan-2-one), group B is involved in proteolysis, and group C in lipolysis. Correlation analysis between these three groups of selected variables and the three first PCs of the PCA (Table 2), demonstrated the relevance of the PCs for the analysis of metabolism.

To interpret the metabolic significance of each PC, the major correlation coefficients between the three variables groups and each PC were investigated (Table 2). The first PC, with which all the group C variables (plus nonan-2-one) were positively correlated, could be interpreted as the lipolytic activity of the strains in milk. Group A variables were positively correlated with the second PC, suggesting that PC2 is related to glycolysis. The third PC, correlated only to group B variables, may be interpreted as an indicator of proteolytic activity.

From these results, the metabolic orientations of the signatures of single strains could easily be identified. Thus, UCMA5713 and S86 strains display the highest lipolytic activity; S87 produces the largest amounts of the end-products of the citrate pathway; and strain DIA-A displays the highest concentrations of VOCs from amino acid catabolism (Fig. 2A and Table 2).

The biological significance of these four clusters was investigated according to the metabolic interpretation of each of the PCs (as described with the 20 most contributory variables: groups A-B-C) and by positioning the four strain clusters in PCA score plots (Fig. 2A and 2B). In cluster I, strains UCMA5713 and S86 predominantly produced VOCs by intense lypolysis (group C) whereas the levels of these compounds were relatively low in...
fermentations with cluster III strains (DIA-A). In cluster II, strains EIP37F and S87 was mainly characterised by the production of acetoin, diacetyl and butanediol (group A). Noted that the DIA-A strain was a singleton (cluster III) emphasising its extreme phenotypic specificity. DIA-A was also associated with strong amino acid catabolism (group B). Finally, the four remaining strains (EIP41A, MC70, EIP33A and EIP33F) belonged to cluster IV; their phenotypic signatures are not characteristic of a particular metabolic pathway, with near mean concentrations of the most contributory VOCs. Therefore, we considered cluster IV to be characterised by no dedicated metabolic pathway.

These findings proved that within a single cluster, other phenotypic variables were responsible for strain unique signatures. To identify the variables that are potentially specific markers of each strain, a PCA was performed on the complete phenotypic dataset for each strain cluster. For the metabolically undefined cluster IV, this analysis revealed that EIP41A was further specifically identified by high sulphide VOC concentrations (responsible for onion and sulphur aroma) whereas EIP33F was distinguished by high residual amino acid concentrations (His-Ala-Pro-Gln-Val-Leu-Ile-Phe) in milk. Both high growth rate and high ethanol concentration were specificities of EIP33A whereas MC70 appeared to be characterised by high furfural and 2-furanmethanol concentrations and the presence of GABA. Furfural and 2-furanmethanol metabolites can be produced in the presence of L-alanine and glucose in acidic conditions (especially at pH 5 or lower), and are characterised by a “bread and burnt” aroma (39).

### III. Genotype-phenotype integration

Phenotype-based strain clustering (Fig. 2) appeared to correspond neither to MLST phylogeny (strains from the same ST did not cluster together), nor to genomic PFGE profiles (four clusters formed rather than the three groups of related strains suggested by PFGE) (Fig. 1). Moreover, the four strains belonging to the biovar diacetylactis were distributed between
different phenotypic clusters (clusters II, III, and IV, Fig. 3). Comparisons of the integrated
phenotypic classification (Fig. 3) with genotypic strain diversity (Fig. 1) confirmed the
discrepancies between the classifications. In particular, the specificity of strain MC70 is
underlined in genotypic classification due to its specific CC2 affiliation of the “domesticated”
strains, but its phenotypic signature was identified as not dedicated to a particular metabolic
pathway. These results highlight the need for a system of classification integrating both
genotype and phenotype to describe L. lactis subsp. lactis biodiversity.

Using the integrated genotype-phenotype dataset, strains were subjected to both PCA and
clustering analysis (Fig. 4): this led to a five-cluster organisation rather than the four strain
clusters generated with discriminatory phenotypic markers (I to IV). Two groups were similar
between these two classifications (clusters II and clusters III). The previous cluster IV was the
most affected by the new classification. In particular, MC70 previously classified in cluster IV
was now singled out. The variable analysis underlined the contribution of MLST phylogenic
distances to PC2 (data not shown) to this new classification and thus the position of MC70.

Using independent genetic (MLST) and genomic (PFGE) datasets in addition to phenotypic
variables revealed the relevance of MLST for the classification. Indeed, unlike MLST data,
the PFGE genomic diversity dataset did not change phenotypic PCA or clustering results
(data not shown). The other strains previously classified in cluster IV (EIP41A and EIP33F)
were classified in existing clusters (I and IV) by the integrated genotype-phenotype
classification.

Variable selection (as previously described) demonstrated that no genotypic variable (either
MLST or PFGE distances) was selected as being one of the most contributory variables.
Contributory variables were, again, mostly VOCs (21/23), and acetic acid and fermentation
time.

DISCUSSION
We report an analysis of the phenotypic diversity of nine strains of *L. lactis* subsp. *lactis* representative of the “domesticated” ecotype and exhibiting a dairy phenotype (efficient growth in milk and fast coagulation). Eighty-two variables related to important dairy features, including physiological descriptors (growth, acidification) (10), carbon and nitrogen metabolites and flavour determinants (14, 40) were monitored in fifty-four fermentations (6 per strain). Our integrated phenotypic approach confirms the suitability of these 82 variables for establishing well-characterised dairy strain signatures in fermented milks. Furthermore, this study reveals large phenotypic strain-to-strain divergences, being sufficient for accurate identification of the nine strains from their milk fermentation signatures. Consequently, these 82 variables provide a useful dataset for classifying dairy strains, and provide new possibilities for applications, and particularly for strain identification and selection.

Unexpectedly, these results illustrate the substantial phenotypic diversity in a limited genetic subpopulation of strains. Indeed, phylogenetic studies suggest that dairy strains are poor contributors to the overall genetic diversity of *L. lactis* subsp. *lactis* (8, 21, 41). “Domesticated” strains arose from a genetic bottleneck or were the result of the successful adaptation to milk of a common “environmental” ancestor, due to genomic flexibility (42, 6, 43). Consequently, it has been suggested that the phenotypic diversity of dairy strains is expected to be low, due to the small pool of strains (44, 45). However, despite the extensive genotypic similarities, we report substantial phenotypic divergence between “domesticated” strains: consequently, these strains have different potentials for different industrial applications.

The phenotypic biodiversity of these dairy strains could be subdivided into four robust clusters depending on metabolic orientations (lypolysis, proteolysis and glycolysis). These clusters may be helpful for assessing strain relatedness. The analysis of more *L. lactis* subsp. *lactis* collections, including strains representative of the “domesticated” ecotype from several
dairy environments, would probably lead to the compositions of these clusters being modified, and to the identification of additional clusters.

However, these findings imply that it would be useful to have appropriate markers to define the specific abilities of a strain. Indeed, a clear description of the differences between strain groups is necessary for a better control of starter choice. Therefore, we used the integrated phenotypic dataset to identify the 20 variables that contribute the most to discrimination between the strains. All these variables are related to VOCs. Our study confirms that selected VOCs can be used as discriminatory phenotypic markers to predict strain signatures in milk processing (46). Combined with the recent development of ultrafast GC/TOF-MS technology for fermented milks (46, 47), monitoring these markers should help high-throughput screening of strain collections. Further analyses of these 20 VOCs revealed that the three major pathways involved in flavour, i.e. lipolysis, proteolysis and glycolysis, were represented. Proteolysis and glycolysis in milk have been studied and described (13, 48, 49), whereas lipolysis in fermented milks has been regularly underestimated; this is largely because many experiments are carried out in skimmed milk media. This work underlines the importance of lipolysis, because more than half of the most discriminatory VOC markers belonged to this class.

*L. lactis* subsp. *lactis* biovar *diacetylactis* strains are used in the dairy industry to produce acetoin and notably diacetyl which imparts a high level of buttery flavour notes. Surprisingly, the four strains in our sample belonging to biovar *diacetylactis* (*DIA-A, EIP33F*, S87 and *EIP33F*) were not grouped in a single cluster at a phenotypic level but were classified together by integrated genotypic classification. Interestingly, this feature demonstrates that acetoin and diacetyl production is not relevant for discriminating biovar *diacetylactis* strain signatures in milk at a phenotypic level. A recent study on *L. lactis* diacetyl- and acetoin-producing strains isolated from diverse origins, indicates that this aroma
production is not restricted to strains able to grow in milk (50). These observations illustrate the intra-group metabolic diversity of these strains, and show the need for further metabolic investigations to characterise these strains thoroughly (amino acid and lipids catabolism notably).

The phenotypic classification was not entirely consistent with the genetic (MLST) and genomic (PFGE) dendrograms or with the new integrated genotypic classification. The proposed integrated genotypic classification for these dairy strains was coherent with the clonal complex phylogenic organisation of “domesticated” strains (MLST) (6), discrimination of the biovar *diacetylactis* lineage (based on results of PCR amplification of *citP*), and strain genomic uniqueness (non redundancy of *SmaI*-macrorestriction by PFGE).

Genotype and phenotype matching is not straightforward, as previously demonstrated, due to regulation of genome and proteome expression (18, 43, 21). However, we identified contributory phenotypic determinants that are promising for the discovery of reliable genetic markers. This approach has been used previously but only at the *lactis* and *cremoris* subspecies level (9, 22). Based on the three groups of VOCs identified as the most contributory variables, the corresponding genes, (*i.e.* involved in lipolysis, proteolysis and glycolysis and their regulation), could be sequenced and their variability analysed.

We also considered a genotype-phenotype classification to integrate all aspects of dairy *L. lactis* subsp. *lactis* diversity. We found that the genotype made only a small contribution to diversity which depends on the large phenotypic differences. Nevertheless, integrating genotypic diversity modified the phenotypic classification and led to a five-cluster organisation rather than the four strain clusters generated with only discriminatory phenotypic markers. In particular, this increased the diversity of the strains “not dedicated to a metabolic pathway” of cluster IV. Moreover, “domesticated” strains of *L. lactis* subsp. *lactis* are characterised by noticeable mobility and substantial genomic variability of their genomes (6).
This genotypic variability may contribute significantly to variations in properties of the starter culture. Adding genotypic characteristics would ensure that all strain potentialities encoded in the genotype would be covered, including those not expressed in our model dairy process but which may be of relevance to other technological applications (such as osmotic resistance in cheese ripening or phage resistance). The genotypic-phenotypic classification tends toward a predictive classification for starter manufacturing.

Finally, this study illustrates the feasibility of establishing well-characterised strain signatures, and this approach could be extended to mixed culture profiling. Indeed, it may be possible to assess the effect of a strain adjunction in co-cultures of lactococci by comparing single-strain signatures to the resulting mix signatures. This method may complement traditional trial-and-error methods used to design targeted starters for particular applications and facilitate the rational development of defined mixed cultures.

**ACKNOWLEDGMENTS**

We acknowledge Helene Tormo from INP-EI Purpan (Toulouse, France) for providing EIP bacterial strains. We thank Gilles De Revel and Laurent Riquier from the Analytical Chemical Service of ISVV (Institut des Sciences du Vin et de la Vigne-Bordeaux, France) for SPME-GC/MS analyses. This research was supported by Yoplait S.A.
REFERENCES


Figure 1: Integrative genotypic classification of nine domesticated *L. lactis* subsp. *lactis* strains from genetic (MLST) and genomic (PFGE) datasets. The Euclidian distance metric and Ward’s criterion for algorithms were used for Hierarchical Ascendant Clustering on scaled MLST and PFGE distances. From left-to-right: strain dendrogram, strain names, presence of the *citP* gene (black diamond), sequence type (ST), Clonal Complexes (CC), the three main clusters reported from PFGE classification (shading). Bootstrap values above 0.50 are indicated on each corresponding node (n=1000).

Figure 2: Principal Component Analysis (PCA) of domesticated *L. lactis* subsp. *lactis* strains from an integrative phenotypic dataset in milk. A-B: Score plots of PC1-PC2-PC3 scores. Diamonds of the same colour correspond to biological replicates of the same strain. Ellipses show the strain category models at a 95 % confidence level.

Figure 3: Integrative phenotypic classification of nine domesticated *L. lactis* subsp. *lactis* strains grown in milk. The Euclidian distance metric and Ward’s criterion for algorithms were used for Hierarchical Ascendant Clustering with the 82 variables. From left-to-right: strain dendrogram, names of the 54 experiments (9 strains with their 6 replicates, from A to F), sequence type from MLST and presence of the *citP* gene (black diamond). Bootstrap values above 0.50 are indicated on each corresponding node (n=500). Dotted squares with roman numeral in the left panel (I to IV) correspond to the four main conserved clusters identified by robustness analysis.

Figure 4: Integrative phenotypic and genotypic Principal Component Analysis (PCA) of nine domesticated *L. lactis* subsp. *lactis* strains grown in milk. Score plots of PC1-PC2 (A) and of PC1-PC3 (B). Diamonds with the same colour are biological replicates of the same strain and coloured ellipses show the strain category models (at a 95 % confidence level). Dotted squares with roman numerals (I to IV) correspond to the four phenotypic clusters (see Fig. 3) and black ellipses represent the five integrative genotype-phenotype clusters.
SUPPLEMENTARY FIGURE LEGENDS

Figure S1: Pulsed field gel electrophoresis (PFGE) of nine strains of *Lactococcus lactis* subsp. *lactis* exhibiting a dairy phenotype. A) Matrix of SD values for all pairwise comparisons of PFGE fingerprints. B) Consensus tree built from UPGMA classification of SD values. Bootstrap values above 0.5 (n=1000) are indicated on each corresponding node. First column corresponds to ST number, second column to strain name and the third to the three main clusters (shading). Biovar *diacetylactis* (assigned from PCR targeting the plasmid-borne *citP* gene) is indicated with asterisk.

Figure S2: Correlation matrix for the integrative phenotypic dataset (Spearman coefficients with Benjamini & Hochberg multiple testing corrections, p-value cut-off of 5.10^-2). A) Integral view of correlation matrix. B) Selected variable clusters for the three main pathways: lipolysis, glycolysis and proteolysis.
Table 1: Dairy *L. lactis* subsp. *lactis* strains used in this study

<table>
<thead>
<tr>
<th>Name</th>
<th>Origin</th>
<th>Clonal Complex (CC)</th>
<th>Sequence Type (ST)</th>
<th>Gene of interest</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>S86</td>
<td>Starter</td>
<td>CC1</td>
<td>10</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>UCMA5713</td>
<td>Grassland (France)</td>
<td>CC1</td>
<td>18</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>S87</td>
<td>Starter</td>
<td>CC1</td>
<td>15</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>EIP41A</td>
<td>Raw milk (France)</td>
<td>CC1</td>
<td>15</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>EIP33A</td>
<td>Raw milk (France)</td>
<td>CC1</td>
<td>15</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>DIA-A</td>
<td>Starter Choozit™ DIA A FRO (France-Danisco)</td>
<td>CC1</td>
<td>15</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>EIP33F</td>
<td>Raw milk (France)</td>
<td>CC1</td>
<td>34</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>EIP37F</td>
<td>Raw milk (France)</td>
<td>CC1</td>
<td>34</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>MC70</td>
<td>Starter Choozit™ MC-70 FRO (France-Danisco)</td>
<td>CC2</td>
<td>7</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>
Table 2: Correlation coefficients for the twenty most contributory variables of the PCA to the three Principal Components (PC) (Significant level at p-value below 0.05). R: correlation coefficient of the variables to the PC. Hierarchical clustering of ascendant classification using the Euclidian distance metric and Ward’s criterion is shown for groups A to C (letters on the left).

<table>
<thead>
<tr>
<th>Variables</th>
<th>PC1</th>
<th></th>
<th>PC2</th>
<th></th>
<th>PC3</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R</td>
<td>p-value</td>
<td>R</td>
<td>p-value</td>
<td>R</td>
<td>p-value</td>
</tr>
<tr>
<td><strong>GROUP A</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acetoin</td>
<td>-0.55</td>
<td>1.45 x 10^{-5}</td>
<td>0.77</td>
<td>1.54 x 10^{-15}</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Acetaldehyde</td>
<td>-</td>
<td>-</td>
<td>0.86</td>
<td>5.51 x 10^{-16}</td>
<td>-0.37</td>
<td>5.37 x 10^{-3}</td>
</tr>
<tr>
<td>2-Nonanone</td>
<td>0.31</td>
<td>2.19 x 10^{-2}</td>
<td>0.89</td>
<td>9.87 x 10^{-20}</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Butan-2,3-diol D7</td>
<td>-0.46</td>
<td>5.24 x 10^{-4}</td>
<td>0.81</td>
<td>8.77 x 10^{-14}</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>GROUP B</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-Methylbutanal</td>
<td>-0.57</td>
<td>5.50 x 10^{-6}</td>
<td>0.29</td>
<td>3.05 x 10^{-7}</td>
<td>0.74</td>
<td>1.95 x 10^{-10}</td>
</tr>
<tr>
<td>Butyl acetate</td>
<td>-0.54</td>
<td>2.24 x 10^{-5}</td>
<td>-0.29</td>
<td>3.05 x 10^{-7}</td>
<td>0.76</td>
<td>3.83 x 10^{-11}</td>
</tr>
<tr>
<td>2-Methyl propan-1-ol</td>
<td>-0.33</td>
<td>1.54 x 10^{-2}</td>
<td>0.51</td>
<td>7.56 x 10^{-5}</td>
<td>0.73</td>
<td>2.64 x 10^{-10}</td>
</tr>
<tr>
<td>2-Hydroxypentan-3-one</td>
<td>-0.61</td>
<td>8.44 x 10^{-7}</td>
<td>0.43</td>
<td>1.10 x 10^{-5}</td>
<td>0.60</td>
<td>1.39 x 10^{-6}</td>
</tr>
<tr>
<td>Ethanol</td>
<td>-0.28</td>
<td>4.22 x 10^{-2}</td>
<td>-0.43</td>
<td>1.19 x 10^{-4}</td>
<td>0.90</td>
<td>8.02 x 10^{-3}</td>
</tr>
<tr>
<td><strong>GROUP C</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Octanol</td>
<td>0.79</td>
<td>1.28 x 10^{-12}</td>
<td>0.53</td>
<td>3.21 x 10^{-5}</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Undecan-2-one</td>
<td>0.90</td>
<td>1.94 x 10^{-20}</td>
<td>0.36</td>
<td>7.82 x 10^{-20}</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Butanoic acid</td>
<td>0.96</td>
<td>8.93 x 10^{-30}</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2-Tridecanone</td>
<td>0.88</td>
<td>8.68 x 10^{-19}</td>
<td>0.44</td>
<td>7.48 x 10^{-4}</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Hexanoic acid</td>
<td>0.93</td>
<td>5.21 x 10^{-24}</td>
<td>0.28</td>
<td>3.73 x 10^{-2}</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Heptanoic acid</td>
<td>0.92</td>
<td>3.38 x 10^{-23}</td>
<td>0.35</td>
<td>8.78 x 10^{-2}</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2-Pentadecanone</td>
<td>0.89</td>
<td>8.43 x 10^{-20}</td>
<td>0.42</td>
<td>1.43 x 10^{-4}</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Octanoic acid</td>
<td>0.92</td>
<td>3.71 x 10^{-22}</td>
<td>0.34</td>
<td>1.23 x 10^{-2}</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Nonanoic acid</td>
<td>0.89</td>
<td>8.73 x 10^{-20}</td>
<td>0.37</td>
<td>5.51 x 10^{-3}</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Decanoic acid</td>
<td>0.93</td>
<td>1.00 x 10^{-23}</td>
<td>0.28</td>
<td>3.94 x 10^{-2}</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>δ-Dodecalactone</td>
<td>0.90</td>
<td>2.95 x 10^{-29}</td>
<td>0.32</td>
<td>1.73 x 10^{-2}</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>