

# Microbial interactions shape cheese flavour formation

Received: 24 February 2023

Accepted: 15 August 2023

Published online: 21 December 2023

 Check for updates

Chrats Melkonian <sup>1,2,3</sup> , Francisco Zorrilla<sup>4</sup>, Inge Kjærbølling<sup>1</sup>, Sonja Blasche<sup>4,5</sup>, Daniel Machado <sup>5,6</sup>, Mette Junge<sup>7</sup>, Kim Ib Sørensen<sup>7</sup>, Lene Tranberg Andersen<sup>8</sup>, Kiran R. Patil <sup>4,5</sup> & Ahmad A. Zeidan <sup>1</sup> 

Cheese fermentation and flavour formation are the result of complex biochemical reactions driven by the activity of multiple microorganisms. Here, we studied the roles of microbial interactions in flavour formation in a year-long Cheddar cheese making process, using a commercial starter culture containing *Streptococcus thermophilus* and *Lactococcus* strains. By using an experimental strategy whereby certain strains were left out from the starter culture, we show that *S. thermophilus* has a crucial role in boosting *Lactococcus* growth and shaping flavour compound profile. Controlled milk fermentations with systematic exclusion of single *Lactococcus* strains, combined with genomics, genome-scale metabolic modelling, and metatranscriptomics, indicated that *S. thermophilus* proteolytic activity relieves nitrogen limitation for *Lactococcus* and boosts de novo nucleotide biosynthesis. While *S. thermophilus* had large contribution to the flavour profile, *Lactococcus cremoris* also played a role by limiting diacetyl and acetoin formation, which otherwise results in an off-flavour when in excess. This off-flavour control could be attributed to the metabolic re-routing of citrate by *L. cremoris* from diacetyl and acetoin towards  $\alpha$ -ketoglutarate. Further, closely related *Lactococcus lactis* strains exhibited different interaction patterns with *S. thermophilus*, highlighting the significance of strain specificity in cheese making. Our results highlight the crucial roles of competitive and cooperative microbial interactions in shaping cheese flavour profile.

Fermented foods based on microbial consortia (e.g., cheese, kefir and kombucha) constitute a large part of the modern diet with many reported health benefits<sup>1–3</sup>. In cheese, well defined Starter Lactic Acid Bacteria (SLAB) cultures are used to ensure consistent and flavourful products. During cheese making, microbes encounter dynamic conditions characterized by different nutrient availability

and different stresses, starting with an initial feast during milk fermentation, followed by a long period of famine during cheese ripening. Cheese-making thus provides a controlled system with well characterized dynamics to understand the role of microbial interactions in shaping the characteristics and quality of fermented foods<sup>4</sup>.

<sup>1</sup>Bioinformatics & Modeling, R&D Digital Innovation, Chr. Hansen A/S, 2970 Hørsholm, Denmark. <sup>2</sup>Theoretical Biology and Bioinformatics, Science for Life, Utrecht University, Utrecht, the Netherlands. <sup>3</sup>Bioinformatics Group, Wageningen University and Research, Wageningen, the Netherlands. <sup>4</sup>Medical Research Council Toxicology Unit, University of Cambridge, Cambridge, UK. <sup>5</sup>European Molecular Biology Laboratory (EMBL), Meyerhofstrasse 1, 69117 Heidelberg, Germany. <sup>6</sup>Department of Biotechnology and Food Science, Norwegian University of Science and Technology, Trondheim 7491, Norway. <sup>7</sup>Strain Improvement, R&D Food Microbiology, Chr. Hansen A/S, 2970 Hørsholm, Denmark. <sup>8</sup>Global Application, Food Cultures & Enzymes, Chr. Hansen A/S, 2970 Hørsholm, Denmark.  e-mail: [chrats.melkonian@gmail.com](mailto:chrats.melkonian@gmail.com); [dkahze@chr-hansen.com](mailto:dkahze@chr-hansen.com)

In industrial cheese making, SLAB cultures are responsible for milk acidification via lactose fermentation. They are predominantly composed of mesophilic strains of *Lactococcus*, including the species *L. lactis* and *L. cremoris* as well as thermophilic strains of *Streptococcus thermophilus*. A rich literature exists on the physiology of the species in mono-cultures<sup>5–8</sup>, and genome-scale metabolic models have been used to model their metabolism<sup>9–11</sup>. In contrast, the interactions between the three taxa is still largely unknown from a mechanistic perspective despite indications of strong inter-dependencies<sup>12,13</sup>. Further, the current knowledge on cheese microbial interactions is limited to the scope of simplified artificial milk media, concerns pairwise associations, or focus only on the short time interval during the milk acidification. Less is known about the interactions involving more than two SLAB strains and how interaction networks evolve during the cheese ripening steps. Metabarcoding and metagenomics approaches have also been used to study the cheese microbiome<sup>14–16</sup>. While these studies provide a good overview on the community dynamics during cheese-making using higher-order taxonomy, they are limited in uncovering inter-species interactions and strain-level diversity.

In this study, we investigated the role of microbial interactions and interacting agents in cheese flavour formation. We employed a strain dropout strategy, which involves omitting either a single strain or a group of strains from the starter culture. Subsequently, we conducted a comprehensive characterization of the entire cheese-making process. We used an industrially relevant SLAB culture, containing: one *Streptococcus thermophilus* (ST), two major *L. lactis* (LLm1 & LLm2), one major *L. cremoris* (LC), and a mixture of 21 *L. cremoris* and *L. lactis* strains in smaller fractions (hereafter, *Lactococcus* blend (LB)). We investigated the effect of leaving out *S. thermophilus* and *Lactococcus* blend in a year-long Cheddar-making experiment (Fig. 1a). Subsequently, controlled milk experiments were conducted to delve into the potential role of interactions, aiming to elucidate the observed growth advantage of *Lactococcus* from *S. thermophilus*. For the latter, we used

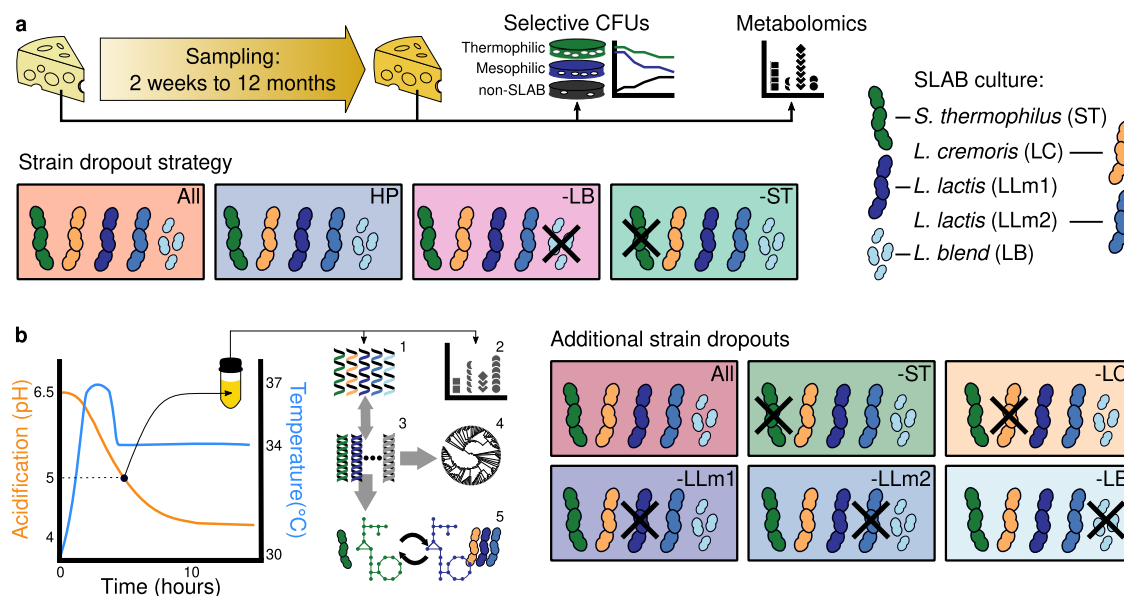
an integrative systems biology approach that combined multiple layers of biological information from genomics, genome-scale metabolic modeling, metatranscriptomics to metabolomics (Fig. 1b).

## Results and discussion

### *S. thermophilus* supports the growth of lactococci and shapes the metabolite profile during cheese ripening

To study the interactions between the members of the SLAB culture, we started by quantifying the population dynamics during a year-long ripening cheese making experiment (see Methods). We used four variations of starter cultures: (i) containing all member species of an industrial starter culture, *viz.*, *S. thermophilus* (ST), *L. cremoris* (LC), *L. lactis* (LLm1 and LLm2), and *Lactococcus* blend (LB); (ii) the same as (i) but prepared independently (HP); (iii) excluding LB; and (iv) excluding ST (Fig. 1a). In the three conditions that included *S. thermophilus*, we observed a trend of slow decline in the population of lactococci and *S. thermophilus*, starting with 9.25 and ending at 8 log<sub>10</sub> CFU/g (Fig. 2a–c). In the single condition that excluded *S. thermophilus*, we observed a much steeper decline in lactococci population, ending at 6.5 log<sub>10</sub> CFU/g (Fig. 2a, b). Also, the declining lactococci population exhibited different trajectories between the two batches, albeit both converging at 9 and 12 months (Fig. 2a, b). The non-SLAB count increased from 0 to 7.5 log<sub>10</sub> CFU/g, in the highest case (Fig. 2d). Based on 95% confidence interval estimates, we observed high variability among the non-SLAB populations, but with no appreciable overall difference between the conditions (Fig. 2d). Leaving out *Lactococcus* blend or changing the way of packing the different strains in the culture did not result in any visible response to the community dynamics. Thus, inter-species interactions, especially involving *S. thermophilus*, appear to be driving the overall population dynamics.

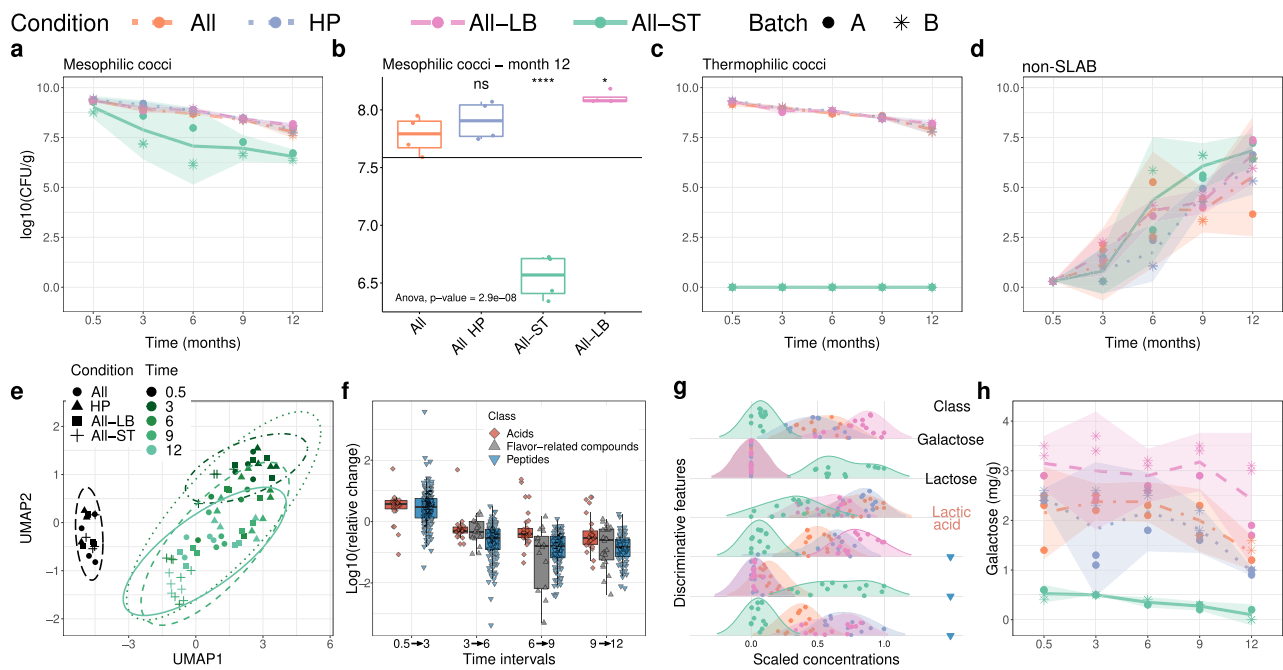
To measure the effect of the microbial interactions in the resulting phenotype of the cheese, we used targeted metabolomics analyses. Across conditions, the biggest change in the metabolic profiles of



**Fig. 1 | Schematic representation of the experimental design and methods.**

**a** The 1 year long cheddar-making experiment. The SLAB culture is composed of one *S. thermophilus* (ST), two major *L. lactis* strains (LLm1 & LLm2), one major *L. cremoris* (LC), and a blend of *Lactococcus* strains (LB). Microbial population dynamics were quantified during cheese ripening using a selective method of viable cell counting, which discriminates between thermophilic cocci, mesophilic cocci and non-Starter Lactic Acid Bacteria (non-SLAB). Additionally, metabolic changes in the cheeses were measured using several targeted analytical chemistry approaches measuring acids, sugars, flavour-related organic compounds and peptides. **b** Schematic representation of the controlled milk experiment in the

laboratory as well as of the methods (1–5) used in the controlled milk experiment. This second experiment involves the removal of additional strains, namely the individual exclusion of three major *Lactococcus* strains. The numbers indicate the different types of data and analysis as follows: 1. metatranscriptomics, 2. metabolomics (i.e., acids, sugars, flavour-related organic compounds), 3. genomics, 4. phylogenomics and 5. genomes-scale metabolic models (GEMs) and community simulations. Our integrative systems biology approach combined: (i) the analysis of the SLAB community's genomes, (ii) the generation and simulation of their respective GEMs, (iii) the analysis of the metatranscriptomes across the different strain removal conditions and (iv) the quantification of key metabolites.



**Fig. 2 | *S. thermophilus* benefit the growth of lactococci community and influence the final flavour of cheese. a–d** Microbial population dynamics during cheese ripening. Relationship between CFUs and time (months) for (a) Mesophilic, (c) Thermophilic cocci, and (d) non-Starter Lactic Acid Bacteria (non-SLAB). The mean values are represented by lines, symbols (circle and asterisk) indicate batch experiments, and the shaded area reflect the 95% confidence interval estimates (similarly in h). b Presents the boxplot comparison of the different condition on Mesophilic cocci at 12 months. The box edges represent the interquartile range, the ends of the whiskers represent  $\pm 1.5 \times$  the interquartile range and central lines are the median values (similarly in f). The  $p$ -values of two-sided  $t$  test presented as ns:  $p > 0.05$ , \* $p < 0.05$ , \*\*\*\* $p < 0.0001$ . Note, significant reduction only on Mesophilic cocci is observed when *S. thermophilus* is absent. Condition All corresponds to whole culture which is composed of 25 number of strains. HP corresponds to an alternative method (hand packed) of the whole culture inoculation. -LB corresponds to the removal of a *L. lactis* blend population. -ST corresponds to the removal of the *S. thermophilus* strain. e–h Metabolome dynamics during cheese ripening. e 2-dimensional representation with the usage of UMAP of all samples based on the metabolomics measurements (acids, carbohydrates and peptides).

The colour indicates the time when the sample was taken and the shape indicates the four different conditions. Note, the stronger change is observed between 2 week and 3 months while the removal of ST (cross symbol) has a strong effect on the later time of cheese ripening. f Relative change of the metabolites on different time intervals including all the conditions, both colour and shape indicates the class of the metabolite. Note, the higher relative change of peptides followed by acids at the interval between 2 week and 3 months, later acids exhibit higher changes. The three measured sugars (glucose, lactose and galactose) are excluded from this panel. g Selection of the six most discriminative metabolites out of 50 for the four conditions. Colour indicate the condition while the right side shape the class of the metabolite. Note, the majority of peptide concentrations are significant different when *S. thermophilus* is not present, as well as with galactose, lactose and lactic acid. The amino acid sequences of the peptides presented (g) in top-down order are EEEKNRLNF, VNELSKD and ELSKDIGSESTE. h Galactose concentration over time highlights the absence of galactose when *S. thermophilus* is absent. The galactose concentration remains steady from the start till the 9 month of cheese ripening and then shows signs of decline. All panels are based on  $n = 4$  biologically independent samples.

cheese was observed between the two-week mark and three months (Fig. 2e). This can be attributed to changes in peptide composition followed by the general increase in amino acids concentrations (Fig. 2f). Furthermore, cheese metabolomes continue to change at a slower pace between the three- to twelve-month marks, resulting in different end profiles. A notable time-dependent accumulation of compounds after the third month was observed on acetic acid, tyramine,  $\gamma$ -amino butyric acid, putrescine and cadaverine (Fig. 2f and Supplementary Fig. 1). The cheeses produced without *S. thermophilus* formed a separate cluster relative to the rest (Fig. 2e). This pattern can be ascribed to compounds that, over time, result in significantly different concentrations across strain removal conditions, especially when *S. thermophilus* was absent. The majority of these compounds are peptides that were either not accumulated, accumulated in lower amounts or accumulated in higher amounts without the addition of *S. thermophilus* (Supplementary Fig. 2). Lactose, galactose and lactic acid concentrations were also found to be significantly different when *S. thermophilus* was absent (Fig. 2g, ANOVA  $F(3,36) = 208 \pm 334$ ,  $p < 0.001$ ,  $\eta^2 = 0.86 \pm 0.08$ ). Specifically, lactose was not fully consumed and galactose was not produced when *S. thermophilus* was absent (Supplementary Fig. 3c). This difference was already notable in the samples from the second week of cheese ripening and remained

relatively constant until the end of cheese-making (Fig. 2h and Supplementary Fig. 3a, c and e). These results point towards the significant influence of *S. thermophilus* growth during milk fermentation, while the over-time differentiation, for example on peptide composition, indicates the important effect of non-growing yet active cells during long-term cheese ripening.

The presence of *S. thermophilus* was found to benefit both the growth of the *Lactococcus* community and the final metabolic profile of the cheddar cheese. Notably, the growth benefit is not clearly visible before two weeks to three months, a time-frame that may be considered prohibitively long for experiments in the laboratory. Instead, these effects become more evident as the experimental time-frame expands to one year. The observed effect may be attributed to the early *S. thermophilus* overflow metabolites during milk fermentation or to alteration of its metabolism at near-zero growth rates during cheese ripening<sup>7,17,18</sup>. In both cases leading to potentially promoting cross-feeding interactions<sup>6</sup>. It cannot be ruled out that the release of distinct cytoplasmic enzymes through the lysis of *S. thermophilus* could enhance the growth of *Lactococcus*<sup>19,20</sup>. Consequently, these factors could result in the identified differences in peptide composition in the final cheese. In accordance with the literature, the presence of *S. thermophilus* reveals a few key metabolic changes from the milk

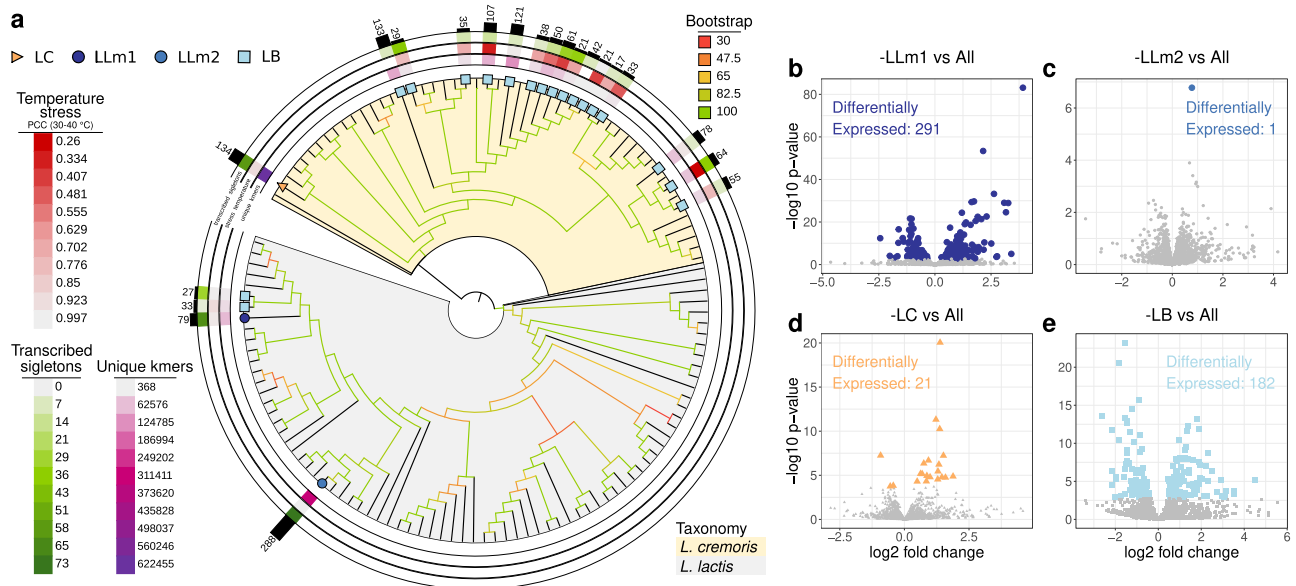
fermentation; the complete consumption of lactose and the production of galactose<sup>17,18</sup>. Together, the compositional and metabolic profiles led us to conclude that both early and late *S. thermophilus* activity is critical for the long term effect on *Lactococcus* population and cheese flavour development.

### *S. thermophilus* has strain-specific interactions with the members of *Lactococcus lactis* community

To assess the role of *S. thermophilus*, we performed controlled experiments wherein additional strains were removed (Fig. 1b). Samples were taken when milk fermentation reached the transition from exponential to stationary phase (at pH 5) and gene expression was monitored using metatranscriptomics. To further investigate the microbial interactions within the SLAB culture, we also analyzed the genomes of the individual strains in the culture and their phylogenetic relationship (Fig. 3a). The phylogenomic analysis separated the species *L. lactis* and *L. cremoris* into two distinct groups. Two of the main *Lactococcus* strains, LC and LLm2, have no close relatives within the culture community. For the third main *Lactococcus* strain, LLm1, we identified two close relatives (LL-LB01 and LL-LB02), belonging to the *Lactococcus* blend (Fig. 3 and Supplementary Table 1). The rest of the strains in the *Lactococcus* blend formed distinct sub-clusters within the *cremoris* clade (Fig. 3). By employing pan-genome analysis on *Lactococcus* strains, we identified singletons (i.e., a gene found only in one genome) for each member of the culture. The highest number of singletons was harbored by the main LLm2 followed by LC, and LLm1 where its number of singletons was in the same range as the *Lactococcus* blend strains, with numbers 288, 134, 79 and average of  $53.6 \pm 18.6$ , respectively (Fig. 3a). In parallel, we evaluated both the presence and activity level (using metatranscriptomics) of the

*Lactococcus* strains by calculating percentile of the transcribed singletons per genome. We found a high number of transcribed singletons on all the main *Lactococcus* strains, while this number varied considerably among the strains of the *Lactococcus* blend, raising questions regarding whether or not these strains were active (Fig. 3a and Supplementary Table 2). We hypothesized that the temperature dynamics during the experiment could influence the strain acidification potential. To test this, by carrying out milk fermentation using each of the individual *Lactococcus* strains at different temperatures, we calculated a proxy for their temperature stress tolerance (Pearson correlation between curves of 30 °C against 37 °C, 40 °C and 43 °C). The results showed that the *L. lactis* strains have higher temperature stress tolerance in the tested ranges, including *Lactococcus* blend strains LL-LB01 and LL-LB02 (PCC  $0.96 \pm 0.05$ ). The *L. cremoris* strains exhibited low temperature stress tolerance (PCC  $0.7 \pm 0.24$ ), with the main LC being one of the most stress tolerant strains within this group (PCC range 0.84–0.99) (Fig. 3a and Supplementary Fig. 4a, b).

We next explored the extent to which we could discriminate between the different strains. The number of unique k-mers in each genome ranged from 368 to 622,455 and they were found to be dependent on the presence of close phylogenetic relative strains (Fig. 3a). The k-mer transcription abundance indicated the feasibility of discriminating between the three main *Lactococcus* strains (LC, LLm1 and LLm2) and the one *S. thermophilus* strain, but not between the strains in the *Lactococcus* blend (Supplementary Fig. 5). Using differential expression analysis, we investigated the interaction effect on *S. thermophilus* transcriptome by leaving out each *Lactococcus* strain/blend from the culture. We found the strongest response when LLm1 was left out followed by *Lactococcus* blend, LC and LLm2 with 291, 182, 21 and 1 gene(s) significantly differential expressed, respectively



**Fig. 3 | The removal of different *Lactococcus* strains has a distinct effect on gene expression in *S. thermophilus*.** **a–e** relationship of the *Lactococcus* phylogeny with the *S. thermophilus* transcription change on different dropout conditions. **a** *Lactococcus* phylogenetic tree separates the *L. cremoris* and *L. lactis* strains. The two clades are coloured with transparent yellow and gray, respectively. The colours and shape of the tree tips indicate the strains presents in the SLAB culture. Orange triangles represent *L. cremoris* (LC), dark and light blue circles represent *L. lactis* (LLm1 and LLm2), and a teal square represents the *Lactococcus* blend (LB). Tree empty tips correspond to the obtained complete NCBI genomes of *Lactococcus*. Additional information is presented in the outer layers of the tree. The first layer presents the unique k-mer content based on the SLAB culture's genomes (purple-scale colour) and indicates higher number for the strains who have no close relative within the culture. The second layer presents the proxy of

temperature stress tolerance based on Pearson Correlation Coefficient (PCC) between individual acidification curve of 30 °C and 40 °C (red-scale colour). High values correspond to high temperature tolerance. The clade *L. cremoris* shows lower temperature tolerance, with few exceptions including the main LC strain (more details presented in Supplementary Fig. 5). The third layer presents the percentile of transcribed singletons (green-scale colour) as well as the total number of the strains singletons (outer bars). **b–e** Volcano plots show the *S. thermophilus* transcription between the whole community (All) and the different *Lactococcus* dropout conditions, respectively. The colours correspond to the 4 *Lactococcus* components. The *S. thermophilus* transcription was observed to change more when LLm1 was left out followed by LB, with 291 and 182 number of total significant up/down regulated genes, respectively. A smaller effect of 21 genes was observed when LC was left out and only one gene when LLm2 was left out.



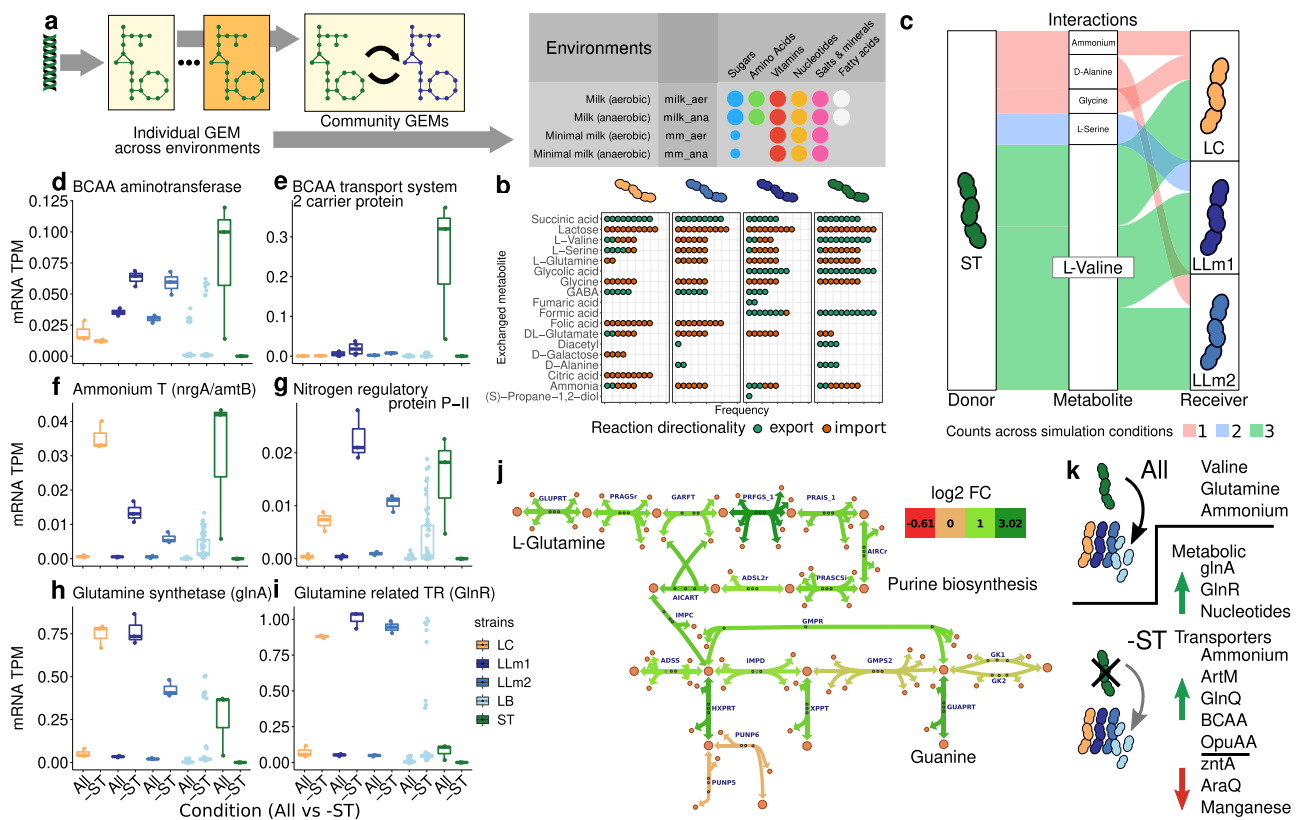
(Fig. 3b–e, Supplementary Table 3 and sections below). Although both LLm1 and LLm2 belong to *L. lactis* clade, the latter strain stands out by having the highest number of singletons as well as having the second highest number of unique k-mers (Fig. 3). By removing strains with a different degree of relatedness we identified that different *Lactococcus* strains have distinct effects on gene expression of *S. thermophilus*. This finding can be attributed to the different genetic background and the diverse phylogenetic relationships within the *Lactococcus* community<sup>21</sup>, highlighting the diversity of potential microbial interactions between *S. thermophilus* and the *Lactococcus* community.

### *S. thermophilus* proteolytic activity may benefit *Lactococcus* community by providing nitrogen sources

To investigate mechanisms of inter-species interactions, we generated genome-scale metabolic models for the individual strains in the culture and performed flux balance analysis-based simulations. To explore the effect of changes in growth medium composition, we generated models gap-filled on relevant variations of milk-mimicking

media. Supplementary Fig. 6a shows overall model statistics for the different species, while Supplementary Fig. 6b shows the specific reactions added through gap-filling on different media across the models (see Supplementary Note 1). Flux balance simulations with individual models were carried out using the different model sets in aerobic and anaerobic variants of fermented and non-fermented milk (Fig. 4a, b and Supplementary Fig. 6c, d). All species consistently consumed lactose in simulations where growth was feasible. Several differences in amino acid uptake/secretion were observed among the strains. The branched chain amino acid valine was exported by *S. thermophilus* in all but one simulation condition, whereas LC, LLm1, and LLm2 consumed this amino acid in four, six, and three simulations conditions, respectively. Additional amino acid uptake/secretion predictions are reported at Supplementary Note 1.

To assess potential cross-feeding, we performed community simulations using the SMETANA framework (see Supplementary Note 1). We used the calculated SMETANA scores, ranging from 0 to 1, as a measure of predicted interaction confidence (0 being lowest



**Fig. 4 | *S. thermophilus* provides nitrogen in form of amino acids to *L. lactis* community, which is necessary for de novo nucleotide biosynthesis.**

**a** Summary of metabolic modeling analysis including individual model and community-based simulations. **b** Selected exchange fluxes predicted across individual flux balance analysis simulations carried in milk media variations for each set of metabolic models. In particular fermentation products, amino acids, and carbon sources highlighting differences in metabolic strategies across species. **c** Alluvial diagram showing predicted metabolic exchanges from community simulations in milk media variants, highlighting the fact the amino acid valine is strongly predicted across all simulation conditions where metabolic cross-talk is expected. The four metabolic models used correspond to *L. cremoris* (LC), coloured orange; *L. lactis* (LLm1 and LLm2), coloured dark and light blue; and *S. thermophilus* (ST), coloured green. Refer to Supplementary Fig. 6 for more details on the metabolic models and simulation. **d–i** Transcriptomic profiles between All and -ST (*S. thermophilus* is left out) conditions of key functions for all the strains in SLAB culture. The box edges represent the interquartile range, the ends of the whiskers represent  $\pm 1.5 \times$  the interquartile range and central lines are the median values ( $n = 3$

biologically independent samples). **d–f** Transcriptomics boxplots of branched chain amino acids (BCAA) aminotransferase, BCAA transport system 2 carrier protein and ammonium transporter (T) supports the predictions of metabolic exchanges. Note, **(e)** the higher transcription profile of *S. thermophilus* on BCAA transport system. **g–j** Up-regulation of *Lactococcus* glutamine and nucleotide metabolism when *S. thermophilus* is left out. **g–i** Boxplots present the activities of nitrogen regulatory protein P-II, glutamine synthetase (glnA) and related transcriptional regulator (TR) (GlnR) for all community members. Note, the up-regulation of all the enzymes derived from *Lactococcus* strains when *S. thermophilus* is left out. **j** Metatranscriptomics represents the increased activity of *Lactococcus* reactions, which are incorporated into Escher maps. The pattern shows a coordinate up-regulation (with green) of the enzymes towards guanine biosynthesis (and uridine). **k** Schematic compilation of the compounds *S. thermophilus* may provide to *Lactococcus* community as well as selected transcriptional changes of the *Lactococcus* community when *S. thermophilus* was left out. The colour and direction of the arrows represent the up- and down- regulation with green/up and red/down, respectively.

confidence and 1 being the highest)<sup>22</sup>. These simulation indicate that *S. thermophilus* provides valine to *Lactococcus* strains in all three community simulation conditions (gap-filled on rich aerobic milk & simulated on minimal aerobic milk, gap-filled on rich anaerobic milk & simulated on minimal aerobic milk, and gap-filled on rich anaerobic milk & simulated on minimal anaerobic milk) where exchanges were predicted between community members (Fig. 4c). Not only was this interaction consistent, it also occurred with a SMETANA score of 1 in every case, suggesting that this is a key ecological interaction for maintaining the composition and function of the community. Serine exchanges from *S. thermophilus* to LLm1 were predicted in 2 out of 3 community simulation conditions, with an average SMETANA score of  $0.38 \pm 0.01$ . Exchanges from *S. thermophilus* to LC involving glycine and ammonium were predicted in 1 out of the 3 simulation conditions, with SMETANA scores of 0.3 and 0.29, respectively (Fig. 4c). Finally, an exchange from *S. thermophilus* to LLm2 involving alanine was predicted under only 1 simulation condition with a SMETANA score of 0.42. Overall, the metabolic simulations strongly indicate cross-feeding between the community members, with *S. thermophilus* standing out as a key donor (Fig. 4c).

To complement the genomic and metabolic modeling analysis, we investigated the changes in *S. thermophilus* and *Lactococcus* community transcriptome. The principal component analysis (Supplementary Fig. 7) showed pattern consistency with the differential transcriptome analysis (Fig. 3b–e, Supplementary Note 2). Regarding changes in *S. thermophilus*'s transcriptome, we found up-regulated genes annotated as oligopeptide-binding protein (AmiA) and oligopeptide transport system permease protein (OppB, OppC, OppD & OppF) (see Supplementary Note 2). Seven down-regulated genes are annotated as transporters, such as glutamine ABC transporter permease protein (GlnP), cadmium, cobalt and zinc/H(+)-K(+) antiporter (Supplementary Note 2). Further, transcriptional repressors such as the catabolite control protein A (CcpA) and nitrogen-metabolism-regulating-proteins such as GTP-sensing transcriptional pleiotropic repressor (CodY) were found to be transcriptionally active. In line with the putative interactions predicted using metabolic modeling, we found a gene annotated as branched-chain amino acids (BCAA) transaminase expressed in *S. thermophilus*, suggesting that valine biosynthesis is active. Moreover, we found a gene annotated as BCAA transport system 2 carrier protein (BrnQ) as well expressed, suggesting that *S. thermophilus* is secreting valine. Complementary differential pathway enrichment analysis supported further that valine biosynthesis was enriched in *S. thermophilus* metatranscriptome (Fig. 4d, e). Overall, the metabolic modeling and metatranscriptome analysis together suggests that *S. thermophilus* can act as a branched chain amino acid donor to *Lactococcus*. It is well known that *Lactococcus* harbor a number of amino acid auxotrophies including valine<sup>1,20,23,24</sup>. In addition, the SMETANA simulations provide insights into the metabolic strategies that may be employed by community members under challenging media conditions.

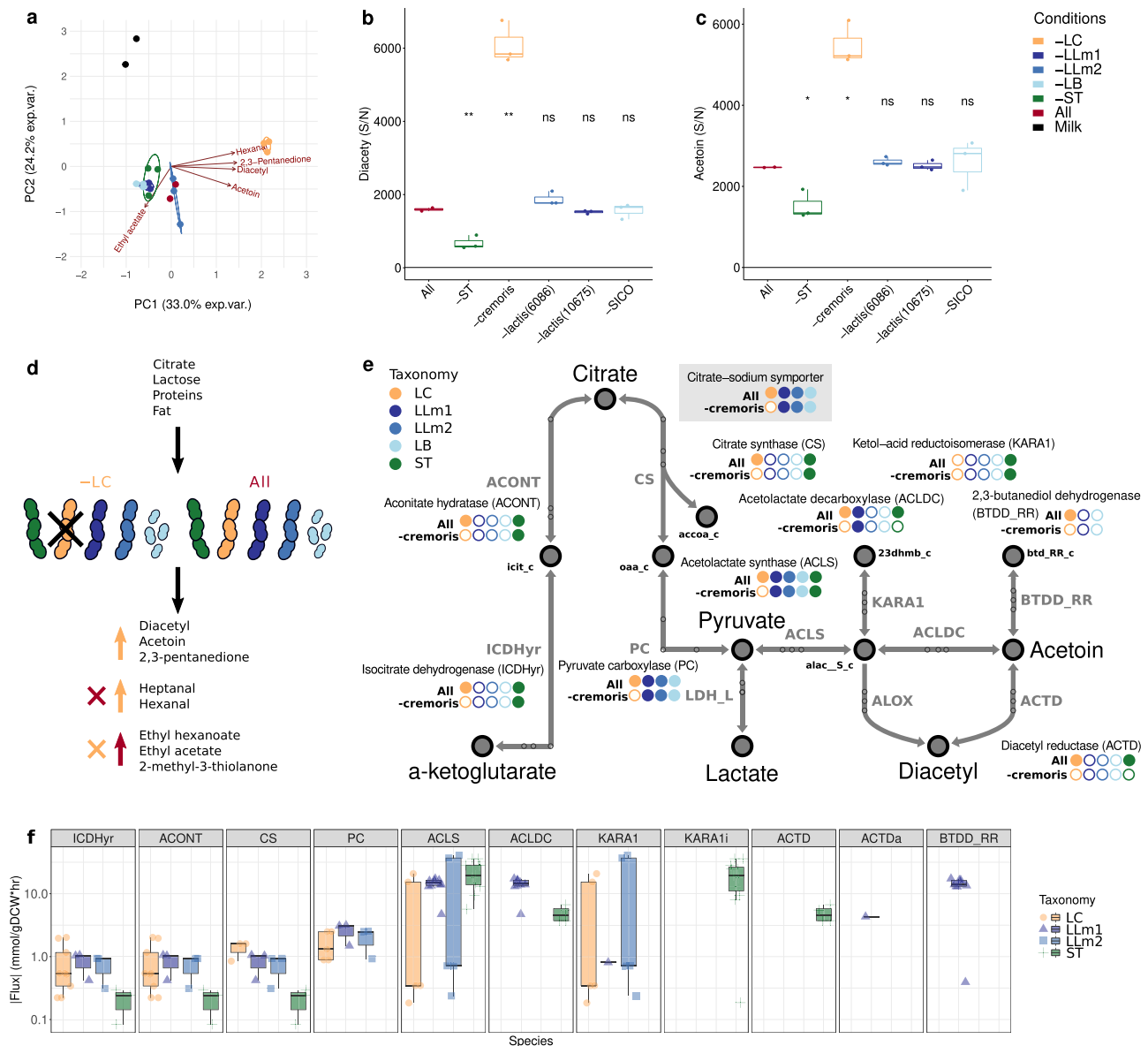
We next investigated transcriptional change in the *Lactococcus* community in the presence and absence of *S. thermophilus*. As it is not possible to distinguish between the transcriptome profiles of all the different *Lactococcus* strains, we merged these into one group for pan-genome analysis and pan-metabolic modeling. Strain-to-strain transcriptional changes were investigated when possible (see Supplementary Note 2). The most up-regulated OGs in *Lactococcus* when *S. thermophilus* is left out are involved in nitrogen assimilation, namely nitrogen regulatory protein P-II, related transcriptional regulator (GlnR), ammonium transporter (amtB), glutamine synthetase (glnA) and glutamine transport ATP-binding protein (glnQ). Additionally, we confirmed that GlnR in *Lactococcus* regulates the three operons, namely amtB, glnA, and glnQ (Fig. 4f–i, Supplementary Table 4 and Supplementary Note 2). The purine and pyrimidine biosynthesis pathways are differentially expressed in a coordinated manner. Several

genes in the pathways were up-regulated, in particular the genes associated with the reactions leading to uridine and guanine synthesis (Fig. 4j and Supplementary Note 2). Both pathways are connected via L-glutamine synthetase. Expression patterns in amino acid metabolism show a mixed pattern of both up- and down- regulation (see Supplementary Note 2). Overall, the metatranscriptomics analysis further supports that *S. thermophilus* cross-feed *Lactococcus* community. Co-culture studies traditionally emphasized competitive interactions<sup>25</sup>, but recent evidence highlights the importance of cross-feeding cooperative interactions in community assembly<sup>26</sup>. Here, one hypothesis that could explain the identified cross-feeding interaction is that *S. thermophilus* has a higher proteolytic activity, thereby providing more peptides and amino acids (available nitrogen source) in the culture which *Lactococcus* takes advantage of<sup>27,28</sup>. When *S. thermophilus* is absent, there is a lower degree of nitrogen availability causing up-regulation of the nitrogen assimilation and nucleotide synthesis pathway in *Lactococcus*. A previous study conducted in a chemically-defined medium has also identified proteolysis and amino acid exchange as significant interactions between *S. thermophilus* and *L. lactis*<sup>13</sup>.

### Cheese flavour compounds are strongly influenced by the interactions within *Lactococcus* community

To elucidate the role of interactions between the SLAB strains in the development of cheese flavour, we performed a targeted metabolomics analysis. Leaving out *L. cremoris* (LC) led to the strongest metabolic response followed by a weaker and different response when leaving out LLm1, *S. thermophilus* and the *Lactococcus* blend (see Supplementary Note 3). Surprisingly, leaving out LLm2 did not lead to a marked change in the metabolic profile of the acidified milk (Fig. 5a). Six flavour compounds, namely heptanal, hexanal, 2-ethyl-furan, 2,3-pentanedione, diacetyl and acetoin, were either detected at significantly higher concentrations or only produced when this main *L. cremoris* strain was removed (Fig. 5b, c and Supplementary Fig. 8). A graphical summary of the metabolome differences between the whole culture and the culture lacking LC is shown in Fig. 5d. Many additional flavour compounds found were significantly altered by the strain removal strategy (Supplementary Fig. 9–10 and Supplementary Note 3), with the majority corresponding to byproducts of secondary metabolism with unknown microbial biosynthesis pathways or gene clusters. Yet, investigation of their patterns highlights the numerous ways these compounds may cross-feed within the SLAB culture. For example, 2,3-pentanedione could be cross-fed by *S. thermophilus* to LC, as it was found in milk only when *S. thermophilus* was present and in higher amounts when LC was absent (Supplementary Fig. 8d). Alternatively, the observed pattern could be a result of substrate competition between the two strains. On the other hand, ethyl hexanoate, ethyl acetate and 2-methyl-3-thiolanone are likely produced by LC (Supplementary Fig. 10a, b and d).

C4 aroma compounds such as diacetyl and acetoin are known derivatives of citrate metabolism with characteristic contribution to buttery-like aroma. Although desirable in small quantities, e.g less than 0.05 mg. per 100 g of diacetyl content, higher amounts could lead to off-flavours<sup>29,30</sup>. Therefore, we investigated whether the increase in diacetyl and acetoin concentrations triggered by the removal of LC could be explained by the metatranscriptome profiles. Indeed, we found differential expression patterns in genes related to diacetyl and acetoin metabolism. We examined the gene expression levels in the metabolic route from citrate towards diacetyl and acetoin through pyruvate as well as an alternative route towards  $\alpha$ -ketoglutarate. The citrate-sodium symporter that is present only on *Lactococcus* strains was highly expressed implying active uptake of the citrate available in milk by *L. lactis*. From citrate, *Lactococcus* strains showed expression of the genes associated with the pyruvate carboxylase and acetolactate synthase reactions, whereas only LC and LLm1 had the genes related to



**Fig. 5 | Cheese flavour compounds are strongly influenced by the interactions within *Lactococcus* community.** The measured metabolome from the controlled milk experiments in response to strain removal conditions (a–d). **a** The PCA of key carbohydrates, acids and cheese flavour compounds. The strongest response was observed from the removal of *L. cremoris*, followed by a coordinated response of LLm1, LB and *S. thermophilus*. Removal of LLm2 does not affect the measured metabolome. The arrows represent the top 5 compounds responsible for the observed changes (2,3-pentanedione, hexanal, diacetyl, acetoin and ethyl acetate). **b, c** Represents the signal-to-noise ratio (S/N) of diacetyl and acetoin, respectively. The absence of *S. thermophilus* in the culture resulted in a lower accumulation of those compounds, whereas the absence of *L. cremoris* led to their higher accumulation. The black horizontal line indicates the average value of the two compounds at milk prior to acidification. The box edges represent the interquartile range, the ends of the whiskers represent  $\pm 1.5 \times$  the interquartile range and central lines are the median values (as in f). Panels based on  $n = 3$  biologically independent samples. Two-sided  $t$  test were performed for each condition against the whole community (All). The symbols indicate the statistical significance results with ns:  $p > 0.05$ , \* $p < 0.05$  and \*\* $p < 0.01$ . **d** Schematic compilation of the compounds that changed between the whole SLAB culture and when *L. cremoris* was left out

(for detailed boxplots see Supplementary Fig. 8 & 10). The colour represents the removal conditions; orange indicates the removal of *L. cremoris* while red represents the whole SLAB culture. The upwards arrow indicates the increase of the compounds relative to their compounds concentration in milk. The X symbol indicates the lack of production of the compounds in the respective condition. **e** Simplified metabolic Escher maps from citrate to  $\alpha$ -ketoglutarate as well as to diacetyl through pyruvate and (S)-2-acetolactate. Metabolites and reactions are indicated by black and gray colour in the graph, respectively. Full coloured circles and open coloured circles indicate the high and low transcription per strain, respectively. The absence of circle indicates the lack of the corresponding orthologous group genes. Note that the transcriptional fluxes from the citrate-sodium symporter towards diacetyl are present in all members of the *Lactococcus* community. In addition, LC and LLm1 manifest higher transcription among the *Lactococcus* community on the genes related to acetolactate decarboxylase. Only LC strain within the *Lactococcus* community manifests transcriptional flux through diacetyl reductase and butanediol dehydrogenase as well as in reactions towards  $\alpha$ -ketoglutarate. **f** Predicted fluxes based on metabolic modeling across the different simulated media.

acetolactate decarboxylase expressed at high levels. On the other route towards  $\alpha$ -ketoglutarate and further, genes associated with aconitate hydratase and isocitrate dehydrogenase are only expressed in LC and *S. thermophilus* (Fig. 5e and Supplementary Figs. 11, 12.

Metabolic modeling simulations across the different medium conditions are largely corroborated by the metatranscriptomics results with the exception of three reactions involving acetoin. These are acetolactate decarboxylase (ACLDC), which does not carry a flux in the LC

model, and 2,3-butanediol dehydrogenase (BTDD\_RR) and diacetyl reductase (ACTD), which carry fluxes in LLM1 model but not in LC model (Fig. 5e, f). Taken together, the presence of the main *L. cremoris* strain in the cultures results in the prevention of an undesirable increase in the levels of some key flavour compounds, including diacetyl and acetoin. This can be attributed to the degradation of diacetyl, e.g., through the action of some alcohol dehydrogenases from LC. Alternatively, LC might be competing with the other strains in the culture on citrate and subsequently converting it to other products than diacetyl and acetoin, such as ethyl-hexanoate, ethyl-acetate and 2-methyl-3-thiolanone, which normally happens in mixed culture fermentation following the rapid drop of redox potential in the beginning of fermentation<sup>4,31,32</sup>. This would result in the reduction of the overall amount of citrate available for diacetyl and acetoin formation.

### Microbial interactions as a mechanism in Cheddar cheese flavour formation

Interactions in microbial communities are reported in a wide range of ecosystems<sup>33–37</sup>. While such interactions are widely reported also in food microbial communities<sup>4,38–42</sup>, very few studies have provided insights into the molecular agents that mediate the interactions. Uncovering interactions in situ is particularly challenging due to complexity of the communities involved as well as that of the medium such as milk. Here, we combined genomics, metatranscriptomics, metabolomics, and metabolic modeling to uncover key microbial interactions in Cheddar cheese-making. Notably, we used industrial strains and a full cycle of one-year long cheese ripening.

In presence of *S. thermophilus*, we report a significant one-year long growth benefit to *Lactococcus* strain population, leading to a distinct final cheese metabolome profile. Based on these findings, we conclude that cross-feeding interactions within the SLAB culture substantially contribute to the formation of flavour in cheese. Our analysis indicated that *S. thermophilus* plays a crucial role by providing the necessary nitrogen source to the *Lactococcus* community. This provides an explanation about the observed growth benefit of the *Lactococcus* population and establishes a connection between the cross-feeding interactions and the development of cheese flavour. Additionally, we identified competitive metabolic interactions within the *Lactococcus* community. *L. cremoris* strain competes with *L. lactis* for the available citrate, resulting in accumulation of key metabolites such as diacetyl and acetoin in the final product. Such interactions take place within the first 5 hours of milk fermentation and strongly influences the final cheese flavour. Lastly, we have identified that the strains of *Lactococcus* affect the activity of *S. thermophilus* differently. While the presence of one *L. lactis* strain hardly influences the activity of *S. thermophilus* and the development of the final cheese flavour profile, the presence of a different *L. lactis* strain influences both significantly. Often in literature the attempts to study microbial interactions disregard the strain diversity. Yet, recent studies highlight the importance of strain interactions, claiming the major role of those in predicting eco-evolutionary dynamics<sup>43</sup>. Our results show how strain-specific metabolic interactions between microbes shape the biochemical profile of cheese, and provide targets towards the rational design and assembly of microbial communities with the aim of fine-tuning cheese flavour. More broadly, the study provides a blue-print to uncovering in situ interactions in complex food microbial ecosystems.

## Methods

### Starter lactic acid bacteria culture

A cheese making experiment was designed to investigate the effect of removing selected strains from a defined-strain SLAB culture. Wild-type strains were obtained from the Chr. Hansen Culture Collection and were originally isolated from dairy products or cultures. By applying this strain removal strategy, it becomes possible to gain insights into the individual role of the strain but also its interactions

within the microbial community. The original defined-strain SLAB culture consisted of five different components: one component with single strain *S. thermophilus*, one component with a multi strain mixture of 21 *Lactococcus* strains belonging to both species of *L. lactis* and *L. cremoris*<sup>44</sup>, which were inoculated as a single component (hereafter, *Lactococcus* blend), two components with single strain *L. lactis*, and one component with single strain *L. cremoris*. The original defined-strain SLAB culture was composed of 25 strains and all bacterial and all strains applied are listed in Supplementary Table 1.

To apply the strain removal strategy four defined-strain SLAB cultures were designed. Condition ALL, corresponded to the original defined-strain SLAB composed of 25 strains packed under industrial conditions. Condition HP, corresponded to the original defined-strain SLAB composed of 25 strains packed under non-industrial condition by hand. Condition -LB, corresponded to the removal of the *Lactococcus* blend. Condition -ST, corresponded to the removal of the *S. thermophilus* strain. The amount added varied between the conditions to target the same acidification profile in the cheese vat while the cheeses making process was kept constant. The target was to obtain a pH of 5.35 at milling and a final moisture in non-fat solids of the cheeses at 54% ± 0.5% measured after 2 weeks of ripening. The designed defined-strain SLAB for conditions ALL, HP, and -LB were all dosed using 8.1 g of culture/100 L, whereas the designed defined-strain SLAB for condition -ST was dosed at a concentration of 17.3 g/100 L to compensate for missing contribution to acidification from *S. thermophilus*. The proportion of each component was adjusted based on the dosage recommended by the producer for the intended use expressed in terms of units of activity, as indicated on the component's information sheet. The final compositions of each component within the SLAB cultures are provided in Supplementary Table 14.

### Cheese making

Cheeses were produced at the Application and Technology Center of Chr. Hansen A/S (Hørsholm, Denmark). The composition of the pasteurized cheese milk (organic milk, Naturmælk, Denmark) was measured using a Milkoscan™ (FOSS, Hillerød, Denmark), and the fat level was adjusted with pasteurized cream (organic cream 38% fat, Naturmælk, Denmark) to obtain a protein to fat ratio of 0.90. The cheese milk was heated to 32 °C and each cheese vat was filled with 150 kg. The milk was ripened for 40 min with the starter culture before the coagulant (CHY-MAX® Plus, Chr. Hansen A/S) was added. After 30 min of coagulation, the gel was cut into cubes 10 × 10 mm. The whey and the cheese grains were stirred for 10 min before heating to 38 °C over 40 min and the final scalding and stirring period was 45 min before whey drainage at pH 6.4–6.5. After 20 min, the cheese curd was cut into 12 blocks and rearranged in blocks 3 times during the next 75 min until pH 5.35 was obtained. The blocks were then milled, and the chips were dry salted (1.7–2.0% salt in dry matter) by manually adding and mixing the salt 2 times during 15 min. The salted chips were mixed every 5 min over a period of 30 min. Following molding, the chips were pressed at 2 bars for 15 min followed by 5 bars for 17 h. The cheeses were vacuum-packed and stored at 9 °C until sampling after 0.5, 3, 6, 9 and 12 months. The cheese gross composition (moisture, fat, protein and total solids) was estimated after 2 weeks of ripening by FoodScan™ (FOSS, Hillerød, Denmark). The NaCl content was estimated by analyzing the chloride concentration by automated potentiometric endpoint titration (DL50, Mettler-Toledo A/S, Glostrup, Denmark). pH was measured potentiometrically (PHC2002-8, Radiometer Analytical SAS, Lyon, France) in a paste prepared by mixing 10 g of grated cheese with 10 ml of deionized water with a wooden spatula. All analyses were carried out in duplicate.

In summary, for the one year-long ripening cheese making experiment we inoculated three variations of the SLAB cultures into four milk tanks. Two milk tanks were inoculated with the complete set of strains, one used as the control and the other to test a different



approach of culture inoculations (All and All HP, respectively). The two other variations of the culture were prepared using the 'strain removal strategy' with the aim of testing the effect of *L. lactis* blend (All-LB) and of the single *S. thermophilus* (All-ST) during cheese-making. Following the acidification phase, we sampled the resulting cheddar cheeses in 5 time points over the one year period. We followed two batch experiments started on two consecutive days.

### Microbial community dynamics

Grated cheese (5.0 g) was mixed with 45.0 g of autoclaved 2% sodium citrate buffer, pH 7.5, 46 °C. The mixture was homogenized in a Stomacher® blender for 4 min at medium setting to dissolve the cheese and suspend the bacteria present. Sequential 10-fold dilutions were prepared as required in 0.1% peptone/0.15M NaCl in water, pH 7.0. The total population of Mesophilic cocci (*Lactococcus*) was determined on pour plated M17 Agar (Difco, USA) after aerobic incubation for 5 days at 30 °C. The total population of Thermophilic cocci (ST) was determined on pour plated M17 Agar after aerobic incubation for 3 days at 37 °C. The total population of non-SLAB was determined on overlaid Rogosa Agar (Sigma Aldrich, USA) after incubation for 7 days at 37 °C. The microbial analysis was carried out in duplicate at each sampling point.

### Carbohydrates and organic acids in cheese

The content of lactose, glucose, galactose, lactate, citrate and acetate were quantified using a Dionex ICS-3000 RFIC-EG™ dual system equipped with an amperometric detector (Dionex, Sunnyvale, CA, USA). The separation was performed using an anion-exchange column (CarboPac® PA20, 3 × 150 mm, 6.5 μm) and an ion-exclusion column (IonPac® ICE-AS6, 9 × 250 mm, 8 μm). Grated cheese (3.0 g) was mixed with 15 ml 83 mM PCA containing mM Na-EDTA, 30 mM arabinose and 48 mM 2-hydroxyisobutyric acid (internal standards) and rotated for 30 min at room temperature. This extract was subsequent centrifuged (5000 × g, 30 min, 4 °C) and the supernatant was filtered (0.45 μm). The supernatant was diluted (600-fold) before analysis. For analysis of carbohydrates 25 μL was injected and the separation was performed at 26 °C with a flow rate of 1.3 mL/min increasing the KOH concentration as follow: 1 mM KOH for 5 min, from 1 mM to 20 mM KOH over 0.5 min, then kept at 20 mM KOH for 6.6 min before re-equilibration to 1 mM KOH over 5 min. For analysis of organic acids 50 μL was injected and the separation was performed at 30 °C with a flow rate of 1 mL/min using 0.4 mM heptfluorobutyric acid in 5% acetonitrile as eluent. The analysis was performed in duplicate at each sampling point.

### Proteolysis in cheese

Grated cheese (2.0 g) was mixed with 18 ml distilled water and blended for 2 min at 25,000 rpm (Ultra-Turrax model T25). This suspension was used for determination of total nitrogen (TN), Non-Casein nitrogen (NCN), and non-protein nitrogen (NPN) using the fractionation procedure described in standard NF ISO 27871 (ISO, 2011). The nitrogen content for each fraction was determined using the Kjeldahl method as described in standard NF EN ISO 8968-1 (ISO, 2014). The level of primary proteolysis (NCN) and secondary proteolysis (NPN) were expressed as the percentage of TN found in the cheese. The analysis was performed in duplicate at each sampling point.

### Peptides in cheese

The peptide profile was analyzed by liquid chromatography-mass spectrometry (LC-MS/MS). Grated cheese (3.0 g) was mixed with 15 ml urea solution (8 M urea, 0.01 M HCL, 0.1% DTT) and rotated for 30 min at room temperature. This extract was subsequent centrifuged (5000 × g, 20 min, 4 °C). 1 ml of the supernatant was transferred to an Eppendorf tube and centrifuged (12,000 × g, 5 min, 4 °C). The supernatant was transferred to vial ready for analysis. Separation was performed using an Agilent 1290 Infinity UHPLC (Agilent, Santa Clara,

USA). Injection volume was 3 μL and the peptides were separated using a reversed-phase column (Waters Acquity CSH C18 peptide column (1.7 μm, 2.1 × 150 mm). Binary linear gradient elution at 50 °C was applied by mixing MilliQ water (0.1% formic acid) (A) with acetonitrile (0.1% formic acid)(B). The gradient used was as follows: 0–0.36 min 96% A, 70% A at 8 min, 54% A at 14 min, 20% A at 16 min, holding 20% A until 16.5 min, increasing to 96% A at 16.6 min and holding until 21 min. The flow rate was 0.3 mL/min. Detection was performed using an Agilent 6540A quadrupole time-of-flight (QTOF) mass spectrometer (Agilent, Santa Clara, USA). Mass spectra were acquired using electrospray ionization in positive mode in the mass range m/z 100–3000. MS/MS analysis was performed using collision induced fragmentation using data-dependent acquisition. For each MS full scan, the two highest intensity ions were selected for fragmentation. The mass of the peptides was annotated using MassHunter Bioconfirm (ver. B.10.0, Agilent Technologies) and quantification by MassHunter Qualitative analysis (ver. B.09.00, Agilent Technologies). The peptide masses were assigned to particular sequences of dairy proteins using Peak Studios (Bioinformatics Solutions Inc.). The data base search was performed using a 50 ppm threshold on mass accuracy and a 0.3 Da threshold on the fragment ions. The digestion enzyme was set to none leading to an unspecified digestion pattern. The analysis was performed in duplicate at each sampling point.

### Free amino acids in cheese

The content of free amino acids (not arginine, but including the breakdown product ornithine), γ-amino butyric acid (GABA), and biogenic amines (phenylethylamine, putrescine, putrescine, cadaverine, histamine, tyramine, spermidine) were quantified using gas chromatography-mass spectrometry (GC-MS) (7890A & 5975C, Agilent Technologies). Grated cheese (3.0 g) was mixed with 15.0 ml milli-Q-water and rotated for 30 min at room temperature and subsequent centrifuged (5000 × g, 30 min, 4 °C) and the supernatant was filtered (0.45 μm). For each sample, 25 μL were mixed with 225 μL milli-Q-water, 50 μL 1.72% norvaline as internal standard, 200 μL of methanol/pyridine 32/8% (v/v). Subsequently, 25 μL methyl chloroformate was added and mixed using a whirlmixer at 3000 rpm for 5 s. Then, 500 μL 1% mff/chloroformate was added and the sample mixed using a whirlmixer at 3000 rpm for 5 s. For the analysis 2 μL was injected onto a DB-XLB 15 m × 0.25 mm × 0.25 μm column (Agilent Technologies) using helium as carrier gas, flow of 1.26 mL/min and an own program of 110 °C (0 min) with a gradient of 20 °C/min until 320 °C (0 min). The analysis was performed in duplicate at each sampling point.

### Volatile compounds in cheese

Volatile compounds were analyzed using Solid Phase Micro Extraction (SPME) and GC-MS (7890B & 5977A, Agilent Technologies) equipped with a SPME autosampler (Gerstel). A plug of cheese (3.0 g) was transferred using a syringe to a 20 ml vial and 100 μL internal standard was added before closing the vial with a cap. The internal standard consisted of a 10 ppm solution of ethanol-d6, dimethyl-d6-disulfide, 2-methyl-3-heptanone and pyrazine-d4 solubilized in 1-methyl-2-pyrrolidone. The vial was heated at 60 °C for 10 min and volatiles were adsorbed to the SPME fiber (DVB/Car/PDMS Gray, 2 cm, Supelco) using an extraction time of 20 min. The fiber was transferred to the GS-inlet and kept at –50 °C for 2 min and then the temperature was increased at 16 °C/s to 150 °C and then 12 °C/s to 300 °C, leading to desorption of the volatiles. Volatiles were then separated on a DB-5, 30 m × 0.25 mm × 1 μm column (Agilent Technologies) using helium as a carrier gas 0.085 ml/min at 325 °C. The mass spectra of the volatile components were collected by running MS scans within the mass range of m/z 29–209. The volatiles were identified by comparison of mass spectra using a compound specific quantifier ion and 2 qualifier ions within a given retention index window. This was done using MassHunter (Version B.07.02.1938, Agilent Technologies). Results

were given as signal-to-noise (S/N) which was calculated as peak height divided by baseline noise. Cheeses from the sampling points 3, 6, 9 and 12 months were analyzed in duplicate. The measurements for the second week are missing for volatile compounds, as they are expected to be very limited. Finally, the metabolic profile of the cheeses was quantified by measuring 30, 3, 28 and 248 features belonging to the classes of acids, sugars, flavour-related organic compounds and peptides, respectively.

### Genome sequencing and assembly

Genomic DNA for de novo short read whole-genome shotgun sequencing was extracted from 1 mL of overnight culture (M17 Broth (Difco) supplemented with glucose) of each of the strains (at OD<sub>600</sub>=1) with DNeasy Blood and Tissue kit on QiaCube system (Qiagen, Germany) following the manufacturer's protocol. Prior to extraction, cell pellets were washed twice in TES buffer (50 mM TRIS pH 8.0, 1 mM EDTA pH 8.5, 20% sucrose) and then resuspended in 180 µL of pre-lysis TET buffer (20 mM TRIS-Cl pH 8.0, 2 mM EDTA pH 8.5, 1,2% Triton X-100, 20 mg/mL lysozyme, 2 µL of 25 U/µL mutanolysin, 4 µL of 100 mg/mL RNase A). Genomic libraries were prepared using KAPA LTP Library Preparation Kit KK8230 (Roche, Switzerland) on a Biomek 4000 Liquid Handler (Beckman Colter, USA). A portion of 500 ng of genomic DNA diluted in EB buffer (Tris-Cl, pH 8.0) was mechanically fragmented on Bioruptor® Standard (Diagenode, USA) with 12 sonication cycles (30 s ON/OFF) to obtain an average fragment size of 300 bp (CV%: 1.5). Fragmented DNA was processed following the KK8230 kit manufacturer's protocol. Following the adapter ligation step, adapter-modified DNA fragments were enriched by 8-cycle PCR. AMPure XP (Beckman Colter) paramagnetic beads were used for clean-ups to purify fragments at average size between 450 and 550 bp. Concentration of gDNA and double stranded DNA libraries were measured by Qubit® Fluorimeter using Qubit dsDNA Broad range and Qubit 1 × dsDNA HS assays (Thermo Fisher Scientific, USA), respectively. Average dsDNA library size distribution was determined using an Agilent HS NGS Fragment (1–6000 bp) kit on an Agilent Fragment Analyzer (Agilent Technologies, USA). Libraries were normalized and pooled in NPB solution (10 mM Tris-Cl, pH 8.0, 0.05% Tween 20) to the final concentration of 10 nM. Following denaturation in 0.2 N NaOH, 10 pM of pooled libraries in 600 µL ice-cold HT1 buffer were loaded onto the flow cell provided in the MiSeq Reagent kit v3 (600 cycles) and sequenced on a MiSeq platform (Illumina Inc., San Diego, USA) with a paired-end protocol and read lengths of 301 nucleotides.

All processing of the short reads was done in either CLC Genomics Workbench versions 9.5.3, 9.5.4 or 10.1.1. The short reads were mapped with default parameters to the reference sequence of the phage Phi × 174 using the tool "Map reads to reference". Unmapped reads from the mapping were trimmed for quality using the PHRED score 23 as the threshold and with the non-default parameter of discarding reads that were less than 50 base pairs long using the tool "Trim Sequences". The trimmed reads were de novo assembled with default parameters except for the minimum contig length was set to 600 base pairs using the tool "De Novo Assembly". Afterwards, a decontamination step was performed where contigs with low depth of coverage were removed using a custom plugin written by Qiagen. The decontamination step first removes all contigs where the depth of coverage is below 15X and afterwards removes all contigs where the depth of coverage is below 25% of the median depth of coverage for the entire genome assembly. Gene calling of the filtered contigs was done with Prodigal version 2.6.2 using the default parameters. Finally, the genome assemblies with annotated genes were functionally annotated with BLAST against a local annotation database using a custom plugin written by Qiagen.

### Comparative genomics and phylogenomics

A *L. lactis* and *L. cremoris* pan-genome was created without strain LC-LB16 which was found to have 6 times more singletons than all the

other strains. The number of core, accessory, and unique (singletons) genes in the pan-genome is 1247, 3308, and 2323, respectively. The OGs from this pan-genome were used in the comparative genomics and differential metatranscriptome analyses as well as in the creation of a pan-metabolic network of *L. lactis* and *L. cremoris* for microbial community modeling. We identified all k-mers of length up to 31 bps ( $k=31$ ) in the genomes of all the strains and unique k-mers in each strain were filtered. For the set of unique k-mers we mapped the metatranscriptomic reads to get their abundance and kept only those above seven counts. Normalization of the number of counts per strain performed based on the number of unique k-mers identified in each of the strains. From the total pool of 54667 genes 15.8% was removed as the cluster above 96 cd-hit identity was mixed among the three taxa. still 84.2% consider a significant number to look at *L. cremoris* and *L. lactis* differences. For the phylogenetic analysis 22 genomes from the SLAB culture were used alongside 107 complete genomes of *L. lactis* and *L. cremoris*, which were retrieved from NCBI results in total to 129 genomes. CD-HIT (v4.8) on 80% amino acid identity thresholds along with a custom-made R script were used to identify a set of 464 monocore marker gene sets<sup>45,46</sup>. The trees were constructed using PhyloPhlAn v3.0<sup>47</sup> using the PhyloPhlAn parameters "--accurate" and "--diversity low", which translate to the usage of a pfsam60 substitution matrix. In addition, multiple sequencing alignments (msa) were performed with muscle (v3.8.1551) and trimming was performed by removing columns with at least one nucleotide appearing above a threshold of 0.99. The final maximum likelihood tree was constructed on the concatenated DNA msa using raxmlHPC (v8.2.12), 100 bootstrap and GTRGAMMAI model<sup>48,49</sup>.

### Generation and simulation of genome-scale metabolic models

The metaGEM workflow's prokaryotic GEM reconstruction and simulation module was used for metabolic modeling<sup>50</sup>. Genome-scale metabolic models were generated for *S. thermophilus*, *Lactococcus* LLM1, LLM2, and LC using CarveMe v1.4.1<sup>51</sup> based on the protein sequences of assembled genomes. First, Prodigal v2.6.3<sup>52</sup> was used to generate open-reading-frame annotated protein sequence files from the corresponding DNA fasta files. Additionally, milk media from a recent kefir publication was used for gap-filling during model generation<sup>42</sup>. More specifically, we formulated four biologically relevant variations of the milk composition, including aerobic rich milk, anaerobic rich milk, aerobic depleted milk, and anaerobic depleted milk. The former two media represent the initial composition of milk, while the latter two media represent milk after it has been depleted by fermentation. A version of each of the four species was generated by gap-filling on each of the four milk media variations, as well as without gap-filling, resulting in a total of 20 models. All models were generated using the default CarveMe universal bacterial model template. Note that although gram positive bacteria are generally not modeled with a periplasmic compartment due to its smaller size relative to gram positive bacteria, all models generated with the automated CarveMe tool contain such a compartment. Individual model simulations were carried out for all models on each of the four media variations using the reframed v1.2.1 and cobrapy v.0.20.0 metabolic modeling packages<sup>53</sup>. Community simulation of genome-scale metabolic models was also carried out in the different media variations, using SMETANA v1.2.0<sup>22</sup>. All associated data, including code used for generating and plotting results, is available on GitHub.

### Inoculations in milk

All milk fermentations were carried out in organic low fat (1.5%) milk from the organic dairy NATURM/ELK, Tinglev, Denmark, which consists of skim milk powder at a level of dry matter of 9.5% (w/v) reconstituted in distilled water and pasteurized at 99 °C for 30 min, followed by cooling to 30 °C<sup>54</sup>. Starting material for all inoculations were concentrated F-DVS® (Direct Vat Set) strains and cultures for

direct inoculation into. All cultures were inoculated 0,02% with F-DVS material by firstly transferring 2.0 g F-DVS material to 200 ml cold milk. After mixing, 4 ml were transferred to 200 ml pre-warmed milk (32 °C) for the final fermentation. The percentages of the different strains and cultures for the 6 different culture blends is given in Supplementary Table 15.

### Milk acidifications and sampling

For the SLAB cultures, after inoculation into bottles with milk placed at 32 °C in a waterbath, pH measurement was started and acidification was followed using a CINAC system<sup>55</sup>. After 59 min, the temperature was raised to 38 °C and at 183 min, the temperature was reduced to the final temperature, 35 °C, for 20 h. The six different culture blends were each run in triplicate. After around 5 h of acidification during the exponential growth phase when reaching pH 5, 1 g fermented milk from each fermentation and 1 g unfermented milk were sampled and added 200 µl 4 N H<sub>2</sub>SO<sub>4</sub> to prepare for analysis for organic acids, Carbohydrates and Volatiles. In addition 1 g was sampled for mRNA extraction. Temperature stress of individual *Lactococcus* strains was assessed by following the acidification of the same milk on four temperatures (30 °C, 37 °C, 40 °C and 43 °C) by continuous monitoring the milk's pH levels.

### Carbohydrates, acids and volatiles in milk

For the preparation of fermented milk samples used for carbohydrate and small organic acid analysis, 1 g or 1 mL of sample was quenched with 200 µL 4N sulfuric acid (H<sub>2</sub>SO<sub>4</sub>) in a 7 mL glass tube, mixed and stored at -20 °C until analysis. The analytes for carbohydrate analysis were extracted from the sample and proteins get deproteinated and precipitated by treatment with an aqueous perchloric acid (PCA) solution. The samples get further diluted to fit into the dynamic range of the quantification. Arabinose is added as an internal standard. The diluted samples are analyzed on a Dionex ICS-3000 system (Thermo Fischer Scientific, Waltham (MA), USA) using an analytical anion-exchange column and a pulsed amperometric detector. For quantification a one-point calibration curve is used. Concentrations are calculated based on the chromatographic peak heights after normalizing to the internal standard (arabinose). The analytes for analysis of small organic acids were extracted from the sample and proteins get deproteinated and precipitated adding an aqueous PCA solution. The samples get further diluted to fit into the dynamic range of the quantification. Adipic acid is added as an internal standard. The diluted samples are analyzed on a Dionex ICS-3000 or ICS-5000 system (Thermo Fischer Scientific, Waltham, MA, USA) using an analytical ion exclusion column and a suppressed conductivity detector. For quantification, an 8-point calibration curve is used. Concentrations are calculated based on the chromatographic peak heights after normalizing to the internal standard (adipic acid).

For the preparation of the volatile organic compounds (VOC's) samples, 1 g or 1 mL of each sample was transferred to a head space vial (20 mL) with 200 µL of 4N H<sub>2</sub>SO<sub>4</sub> and sealed with teflon-lined aluminum caps. The samples were then identified using a static head space sampler connected to a Gas Chromatograph with Flame Ionization Detector (GC-FID) (Perkin Elmer, MA, USA) and equipped with a HP-FFAP column. The identification of VOC's was based on retention time in comparison with that of standards. The injector and detector were maintained at 180 °C and 220 °C, respectively. The oven was initially heated to 60 °C and held for 2 min, then increased to 230 °C and held for 0.5 min. The calculation of the concentration of each compound in each sample was based on the peak height divided by the response factor. The response factor is established using standard solutions by the quotient of the peak height divided by the known sample concentration. All of the chemicals and analytes used for the standard solutions were provided by Sigma-Aldrich, Munich, Germany. Overall with these approaches, we measured a total of 43 metabolites

comprising five carbohydrates, eight organic acids and 30 important flavour compounds.

### RNA extraction, sequencing and analysis

Total RNA was extracted using the RNeasy Mini kit (Qiagen 74104) and rRNA was depleted with the NEBNext rRNA Depletion Kit (Bacteria). The sequencing library was then prepared using the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina and the libraries were sequenced by EMBL Genomic Core Facility using the Illumina NextSeq 500 system, read length 150 bp. The RNA-reads are preprocessed and filtered based on minimum quality and length using the NGLess pipeline and subtrim. Given a read it finds the longest substring, such that all bases are of at least the given quality. The parameters minimum quality and minimum length were set to 25 and 100, respectively. The filtered reads are mapped to the genomes of all strains present in the Culture using bbmap (v38.75)—if one read has multiple hits, a random distribution is used (selecting one top-scoring site randomly). The total number of reads ranges from 4,855,868 to 18,812,096 in samples 4B and 1B, respectively, and the percentage of mapped reads is in the range of 93–95% (Supplementary Table 16)<sup>56</sup>. htseq-count (v0.11.2) was used to count the number of reads per gene<sup>57</sup>. The resulting counts were grouped per orthology group (custom python script) based on a previously created pan-genome created for the pan-metabolic network. Normalization was performed and differential gene expression with DESeq2 (v1.26.0) and SARTools (v1.7.2) in R<sup>58,59</sup>. For the differential gene expression analysis, Wald significance tests were performed with BH *p*-value adjustments<sup>60</sup> to take into account multiple testing, and the level of controlled false positive rate was set to 0.01. Further differentially expressed genes, *Lactococcus* pan-metabolic network information and KEGG pathways were based on custom analysis<sup>35,61</sup>. The data was normalized using DESeq2 to correct systematic technical biases and make it possible to compare read counts across samples. The median scaling factor for each sample was used. This is done for the data containing all genes. RegPrecise database was used for validation of key functional regulation<sup>62</sup>.

### General statistical and computational analysis

We used R-packages Boruta (v8.0.0), uwot (v0.1.14), apcluster (v1.4.10), ggplot2 (v3.4.2), ggrepel (v0.9.3), ggfortify (v0.4.16), ggridges (v0.5.4), ggbiplot (v0.55), ggpubr (v0.6.0), ggalluvial (v0.12.5), scales (v1.2.1), tidyverse (v2.0.0), dplyr (v1.1.2), plyr (v1.8.8), reshape2 (v1.4.4), cowplot (v1.1.1), patchwork (v1.1.2), rstatix (v0.7.2), viridis (v0.6.3), grid (v4.2.2), Biostrings (v2.64.1) and readxl (1.4.2) for downstream analysis including, data parsing and visualization, clustering, dimensionality reduction and importance feature ranking analysis<sup>46,63–83</sup>. Specifically, Boruta algorithm is a wrapper around Random Forest<sup>84</sup> that performs randomization tests. It is used to identify discriminative metabolites in cheese ripening experiment. Features with confidence of importance above 0.99 (the default value in Boruta) were treated as informative. Also the maximal number of importance source runs was increased to 10,000. In this study, statistical analysis of variance (ANOVA) and pairwise comparisons between two groups using either the Wilcoxon rank-sum test (wilcox.test) or the Student's *t* test (*t* test) were conducted. The resulting *p*-values were reported in the figures. Temperature stress of *Lactococcus* strains was represented as a proxy based on the Pearson correlation coefficient (PCC) between the acidification curves of mono-cultures at 30 °C compared to mono-cultures at 37 °C, 40 °C and 43 °C. Metabolic networks were visualized using Escher web-tool (v1.5.0)<sup>85</sup>. Schematic representations and figures were created and polished, respectively, in Inkscape.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.



## Data availability

Source Data, including metabolomics, metatranscriptomics count tables are available in [https://github.com/Chrats-Melkonian/mi\\_cheese](https://github.com/Chrats-Melkonian/mi_cheese)<sup>86</sup>. Genomic and metatranscriptomics data have been deposited under the BioProject ID [PRJNA950467](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA950467). The genomes of SLAB culture has been deposited under BioSample accession SAMN34041181-SAMN34041202. The metatranscriptomes has been deposited under SRA accession SRR24029527-SRR24029544.

## Code availability

Code to reproduce the results is available in [https://github.com/Chrats-Melkonian/mi\\_cheese](https://github.com/Chrats-Melkonian/mi_cheese)<sup>86</sup>.

## References

- Teusink, B. & Molenaar, D. Systems biology of lactic acid bacteria: for food and thought. *Curr. Opin. Syst. Biol.* **6**, 7–13 (2017).
- Smith, J., Yeluripati, J., Smith, P. & Nayak, D. R. Potential yield challenges to scale-up of zero budget natural farming. *Nat. Sustain.* **3**, 247–252 (2020).
- Wastyk, H. C. et al. Gut-microbiota-targeted diets modulate human immune status. *Cell* **184**, 4137–4153 (2021). 4137.
- Blaya, J., Barzideh, Z. & LaPointe, G. Symposium review: interaction of starter cultures and nonstarter lactic acid bacteria in the cheese environment. *J. Dairy Sci.* **101**, 3611–3629 (2018).
- Song, A. A., In, L. L. A., Lim, S. H. E. & Rahim, R. A. Erratum to: a review on lactococcus lactis: from food to factory. *Microb. Cell Factor.* **16**, 139 (2017).
- Markakiou, S., Gaspar, P., Johansen, E., Zeidan, A. A. & Neves, A. R. Harnessing the metabolic potential of streptococcus thermophilus for new biotechnological applications. *Curr. Opin. Biotechnol.* **61**, 142–152 (2020).
- Rau, M. H. et al. Genome-scale metabolic modeling combined with transcriptome profiling provides mechanistic understanding of streptococcus thermophilus CH8 metabolism. *Appl. Environ. Microbiol.* **88**, <https://doi.org/10.1128/aem.00780-22> (2022).
- Douwenga, S., van Tatenhove-Pel, R. J., Zwering, E. & Bachmann, H. Stationary lactococcus cremoris: energetic state, protein synthesis without nitrogen and their effect on survival. *Front. Microbiol.* **12**, 794316 (2021).
- Oliveira, A. P., Nielsen, J. & Förster, J. Modeling lactococcus lactis using a genome-scale flux model. *BMC Microbiol.* **5**, 39 (2005).
- Pastink, M. I. et al. Genome-scale model of streptococcus thermophilus LMG18311 for metabolic comparison of lactic acid bacteria. *Appl. Environ. Microbiol.* **75**, 3627–3633 (2009).
- Flahaut, N. A. L. et al. Genome-scale metabolic model for lactococcus lactis MG1363 and its application to the analysis of flavor formation. *Appl. Microbiol. Biotechnol.* **97**, 8729–8739 (2013).
- Champagne, C., Gagnon, D., St-Gelais, D. & Vuilleumard, J. Interactions between lactococcus lactis and streptococcus thermophilus strains in cheddar cheese processing conditions. *Int. Dairy J.* **19**, 669–674 (2009).
- Özcan, E. et al. Dynamic co-culture metabolic models reveal the fermentation dynamics, metabolic capacities and interplays of cheese starter cultures. *Biotechnol. Bioeng.* **118**, 223–237 (2020).
- Duru, I. C. et al. Metagenomic and metatranscriptomic analysis of the microbial community in swiss-type maasdam cheese during ripening. *Int. J. Food Microbiol.* **281**, 10–22 (2018).
- Salazar, J. K. et al. Metagenomics of pasteurized and unpasteurized gouda cheese using targeted 16s rDNA sequencing. *BMC Microbiol.* **18**, 189 (2018).
- Walsh, A. M., Macori, G., Kilcawley, K. N. & Cotter, P. D. Meta-analysis of cheese microbiomes highlights contributions to multiple aspects of quality. *Nat. Food* **1**, 500–510 (2020).
- Hutkins, R., Halambeck, S. & Morris, H. Use of galactose-fermenting streptococcus thermophilus in the manufacture of swiss, mozzarella, and short-method cheddar cheese. *J. Dairy Sci.* **69**, 1–8 (1986).
- Proust, L. et al. Multi-omics approach reveals how yeast extract peptides shape streptococcus thermophilus metabolism. *Appl. Environ. Microbiol.* **86**, e01446–20 (2020).
- Crow, V. et al. The role of autolysis of lactic acid bacteria in the ripening of cheese. *Int. Dairy J.* **5**, 855–875 (1995).
- Nugroho, A. D. W., Kleerebezem, M. & Bachmann, H. Growth, dormancy and lysis: the complex relation of starter culture physiology and cheese flavour formation. *Curr. Opin. Food Sci.* **39**, 22–30 (2021).
- Zhai, Y. & Wei, C. Open pangenome of lactococcus lactis generated by a combination of metagenome-assembled genomes and isolate genomes. *Front. Microbiol.* **13**, 948138 (2022).
- Zelezniak, A. et al. Metabolic dependencies drive species co-occurrence in diverse microbial communities. *Proc. Natl. Acad. Sci.* **112**, 6449–6454 (2015).
- Godon, J. J. et al. Gene inactivation in lactococcus lactis: branched-chain amino acid biosynthesis. *J. Bacteriol.* **175**, 4383–4390 (1993).
- Delorme, C., Godon, J. J., Ehrlich, S. D. & Renault, P. Gene inactivation in lactococcus lactis: histidine biosynthesis. *J. Bacteriol.* **175**, 4391–4399 (1993).
- Ghoul, M. & Mitri, S. The ecology and evolution of microbial competition. *Trends Microbiol.* **24**, 833–845 (2016).
- Kehe, J. et al. Positive interactions are common among culturable bacteria. *Sci. Adv.* **7**, eabi7159 (2021).
- Kunji, E. R. S., Mierau, I., Hagting, A., Poolman, B. & Konings, W. N. The proteolytic systems of lactic acid bacteria. *Antonie van Leeuwenhoek* **70**, 187–221 (1996).
- Rodríguez-Serrano, G. M. et al. Proteolytic system of streptococcus thermophilus. *J. Microbiol. Biotechnol.* **28**, 1581–1588 (2018).
- Smid, E. & Kleerebezem, M. Production of aroma compounds in lactic fermentations. *Annu. Rev. Food Sci. Technol.* **5**, 313–326 (2014).
- Calbert, H. E. & Price, W. V. A study of the diacetyl in cheese. i. diacetyl content and flavor of cheddar cheese. *J. Dairy Sci.* **32**, 515–520 (1949).
- Hugenholtz, J. Citrate metabolism in lactic acid bacteria. *FEMS Microbiol. Rev.* **12**, 165–178 (1993).
- Wang, Y. et al. Metabolism characteristics of lactic acid bacteria and the expanding applications in food industry. *Front. Bioeng. Biotechnol.* **9**, 612285 (2021).
- Bäumler, A. J. & Sperandio, V. Interactions between the microbiota and pathogenic bacteria in the gut. *Nature* **535**, 85–93 (2016).
- Lara, E. G., van der Windt, I., Molenaar, D., de Vos, M. G. J. & Melkonian, C. Using functional annotations to study pairwise interactions in urinary tract infection communities. *Genes* **12**, 1221 (2021).
- Melkonian, C. et al. High biodiversity in a benzene-degrading nitrate-reducing culture is sustained by a few primary consumers. *Commun. Biol.* **4**, 530 (2021).
- Melkonian, C., Seidl, M. F., van der Hooft, J. J. & de Vos, M. G. Metabolic interactions shape a community's phenotype. *Trends Microbiol.* **30**, 609–611 (2022).
- Machado, D. et al. Polarization of microbial communities between competitive and cooperative metabolism. *Nat. Ecol. Evol.* **5**, 195–203 (2021).
- Irlinger, F. & Mounier, J. Microbial interactions in cheese: implications for cheese quality and safety. *Curr. Opin. Biotechnol.* **20**, 142–148 (2009).
- Jiang, N. et al. Multi-omics approaches to elucidate the role of interactions between microbial communities in cheese flavor and quality. *Food Rev. Int.* 1–13 <https://doi.org/10.1080/87559129.2022.2070199> (2022).



40. Parente, E., Zotta, T. & Ricciardi, A. A review of methods for the inference and experimental confirmation of microbial association networks in cheese. *Int. J. Food Microbiol.* **368**, 109618 (2022).
41. Ponomarova, O. et al. Yeast creates a niche for symbiotic lactic acid bacteria through nitrogen overflow. *Cell Syst.* **5**, 345–357.e6 (2017).
42. Blasche, S. et al. Metabolic cooperation and spatiotemporal niche partitioning in a kefir microbial community. *Nat. Microbiol.* **6**, 196–208 (2021).
43. Goyal, A., Bittleston, L. S., Leventhal, G. E., Lu, L. & Cordero, O. X. Interactions between strains govern the eco-evolutionary dynamics of microbial communities. *eLife* **11**, e74987 (2022).
44. Oren, A. & Garrity, G. M. Notification that new names of prokaryotes, new combinations and new taxonomic opinions have appeared in volume 71, part 3 of the IJSEM. *Int. J. Syst. Evol. Microbiol.* **71**, <https://doi.org/10.1099/ijsem.0.004812> (2021).
45. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).
46. R Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria <https://www.R-project.org/> (2020).
47. Asnicar, F. et al. Precise phylogenetic analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0. *Nat. Commun.* **11**, 2500 (2020).
48. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
49. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
50. Zorrilla, F., Buric, F., Patil, K. R. & Zelezniak, A. metaGEM: reconstruction of genome scale metabolic models directly from metagenomes. *Nucleic Acids Res.* **49**, e126–e126 (2021).
51. Machado, D., Andrejev, S., Tramontano, M. & Patil, K. R. Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. *Nucleic Acids Res.* **46**, 7542–7553 (2018).
52. Hyatt, D. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform.* **11**, 1–11 (2010).
53. Ebrahim, A., Lerman, J. A., Pálsson, B. O. & Hyduke, D. R. Cobrapy: constraints-based reconstruction and analysis for python. *BMC Syst. Biol.* **7**, 1–6 (2013).
54. Sørensen, K. I., Curic-Bawden, M., Junge, M. P., Janzen, T. & Johansen, E. Enhancing the sweetness of yoghurt through metabolic remodeling of carbohydrate metabolism in streptococcus thermophilus and lactobacillus delbrueckii subsp. bulgaricus. *Appl. Environ. Microbiol.* **82**, 3683–3692 (2016).
55. Corrieu, G., Spinnler, H., Jomier, Y. & Picque, D. Automated system to follow up and control the acidification activity of lactic acid starters. *French patent FR 2*, 612 (1988).
56. Bushnell, B. Bbmap: A fast, accurate, splice-aware aligner <https://www.osti.gov/biblio/1241166> (2014).
57. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2014).
58. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
59. Varet, H., Brillet-Guéguen, L., Coppée, J.-Y. & Dillies, M.-A. SAR-Tools: a DESeq2- and EdgeR-based r pipeline for comprehensive differential analysis of RNA-seq data. *PLOS One* **11**, e0157022 (2016).
60. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc.: Ser. B (Methodological)* **57**, 289–300 (1995).
61. Melkonian, C. et al. Finding functional differences between species in a microbial community: case studies in wine fermentation and kefir culture. *Front. Microbiol.* **10**, 1347 (2019).
62. Novichkov, P. S. et al. RegPrecise 3.0 – a resource for genome-scale exploration of transcriptional regulation in bacteria. *BMC Genom.* **14**, 745 (2013).
63. Melville, J. *uwot: The Uniform Manifold Approximation and Projection (UMAP) Method for Dimensionality Reduction* <https://CRAN.R-project.org/package=uwot> (2022). R package version 0.1.14.
64. Wickham, H. & Seidel, D. *scales: Scale Functions for Visualization* <https://CRAN.R-project.org/package=scales> (2022). R package version 1.2.1.
65. Bodenhofer, U., Kothmeier, A. & Hochreiter, S. APCluster: an r package for affinity propagation clustering. *Bioinformatics* **27**, 2463–2464 (2011).
66. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag New York, 2016). <https://ggplot2.tidyverse.org>.
67. Slowikowski, K. *ggrepel: Automatically Position Non-Overlapping Text Labels with 'ggplot2'* <https://CRAN.R-project.org/package=ggrepel> (2023). R package version 0.9.3.
68. Tang, Y., Horikoshi, M. & Li, W. ggfortify: unified interface to visualize statistical results of popular r packages. *R J.* **8**, 474 (2016).
69. Wilke, C. O. *ggridges: Ridgeline Plots in 'ggplot2'* <https://CRAN.R-project.org/package=ggridges> (2022). R package version 0.5.4.
70. Kassambara, A. *ggpubr: 'ggplot2' Based Publication Ready Plots* <https://CRAN.R-project.org/package=ggpubr> (2023). R package version 0.6.0.
71. Wickham, H. et al. Welcome to the tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).
72. Wickham, H., François, R., Henry, L., Müller, K. & Vaughan, D. *dplyr: A Grammar of Data Manipulation* <https://CRAN.R-project.org/package=dplyr> (2023). R package version 1.1.2.
73. Wickham, H. The split-apply-combine strategy for data analysis. *J. Stat. Softw.* **40**, 1–29 (2011).
74. Wickham, H. Reshaping data with the reshape package. *J. Stat. Softw.* **21**, 1–20 (2007).
75. Wilke, C. O. *cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'* <https://CRAN.R-project.org/package=cowplot> (2020). R package version 1.1.1.
76. Pedersen, T. L. *patchwork: The Composer of Plots* <https://CRAN.R-project.org/package=patchwork> (2022). R package version 1.1.2.
77. Kurs, M. B. & Rudnicki, W. R. Feature selection with the Boruta package. *J. Stat. Softw.* **36**, 1–13 (2010).
78. Kassambara, A. *rstatix: Pipe-Friendly Framework for Basic Statistical Tests* <https://CRAN.R-project.org/package=rstatix> (2023). R package version 0.7.2.
79. Garnier et al. *viridis(Lite) - Colorblind-Friendly Color Maps for R* <https://sjmgarnier.github.io/viridis/> (2023). Viridis package version 0.6.3.
80. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria <https://www.R-project.org/> (2022).
81. Brunson, J. *ggalluvial: layered grammar for alluvial plots*. *J. Open Source Softw.* **5**, 2017 (2020).
82. Pagès, H., Aboyoun, P., Gentleman, R. & DebRoy, S. *Biostrings: Efficient manipulation of biological strings* <https://bioconductor.org/packages/Biostrings> (2022). R package version 2.64.1.
83. Wickham, H. & Bryan, J. *readxl: Read Excel Files* <https://CRAN.R-project.org/package=readxl> (2023). R package version 1.4.2.
84. Breiman, L. *Machine Learning* **45**, 5–32 <https://doi.org/10.1023/a:1010933404324> (2001).
85. King, Z. A. et al. Escher: a web application for building, sharing, and embedding data-rich visualizations of biological pathways. *PLOS Comput. Biol.* **11**, e1004321 (2015).

86. Melkonian, C. & Zorrilla, F. Chrats-melkonian/mi\_cheese: code and data accompanying the manuscript ‘microbial interactions shape cheese flavor formation’ <https://zenodo.org/record/8177856> (2023).

## Acknowledgements

We would like to thank Mads Bennedsen for his efforts in funding acquisition and also in project management, together with Karsten Hellmuth and Eric Johansen. We thank Kosai Al-Nakeeb, Anna Koza, Martin Abel-Kistrup and Jacob Bælum for their significant contributions to the sequencing and assembly of the genomes of individual strains. We thank Lisandra Zepeda and Ida Bærholm Schnell for their help with cheese sampling. Finally, we thank Gunnar Øregaard for providing acidification data for the individual strains and Jannik Vindeloev and Ana Rute Neves for helpful discussions. This project has received funding from Innovation Fund Denmark (Grand Solution grant no. 6150-00033B; FoodTranscriptomics) and Chr. Hansen A/S. C.M acknowledges support from the Dutch Research Council, as part of the MiCROP Consortium (NWO/OCW grant no. 024.004.014). K.R.P. acknowledges support from the European Research Council (ERC) under the European Union’s Horizon 2020 Research and Innovation Program (Grant agreement no. 866028).

## Author contributions

C.M. organizing ideas, writing, comparative genomics, phylogenomics, transcriptomics and bioinformatics analyses. F.Z. genome-scale metabolic modeling, community simulations and writing. I.K. initial transcriptomics, bioinformatics analyses and genome-scale metabolic modeling. M.J. and K.L.Sr. carried out the experiments and drafted the manuscript part relating to sections “Inoculations in milk” and “Milk acidifications and sampling”. S.B. mRNA extraction of the milk fermentation samples for mRNA sequencing. D.M. analysed initial metatranscriptomics data, performed model reconstruction and community simulations. L.T.A. participated in conceptualization of the study, design and supervise the year-long cheddar-making experiment. K.R.P. participated in conceptualization of the study, funding acquisition, supervision of modeling and critically revising the manuscript. A.A.Z. participated in conceptualization of the study, funding acquisition, supervision of experimental design, data analysis and bioinformatics, and critically revising the manuscript.

## Competing interests

C.M., I.K., M.J., K.L.Sr., L.T.A. and A.A.Z. are present or previous employees at Chr. Hansen A/S, a global supplier microbial cultures for food fermentation. The authors’ views presented in this study, however, are solely based on scientific grounds and do not reflect the commercial interest of their employer. F.Z., D.M., S.B. and K.R.P. declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-023-41059-2>.

**Correspondence** and requests for materials should be addressed to Chrats Melkonian or Ahmad A. Zeidan.

**Peer review information** *Nature Communications* thanks Maria Kazou and the other, anonymous, reviewers for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

© Chr. Hansen 2023