Title: Microbial biogeography of the transnational fermented milk Matsoni

Nicholas A. Bokulich^{1,2,3,†}

Lia Amiranashvili⁴

Karine Chitchyan⁵

Narine Ghazanchyan⁵

Karen Darbinyan⁵

Nino Gagelidze⁴

Tinatin Sadunishvili⁴

Vigen Goginyan⁵

Giorgi Kvesitadze⁴

Tamas Torok⁶

David A. Mills^{1,2,3}

¹Department of Viticulture and Enology, University of California, Davis, CA 95616, USA

²Department of Food Science and Technology, University of California, Davis, CA 95616, USA

³Foods for Health Institute, University of California, Davis, CA 95616, USA

⁴S. Durmishidze Institute of Biochemistry and Biotechnology, Agricultural University of Georgia, 0159 Tbilisi, Georgia

⁵Scientific and Production Center “Armbiotechnology” NAS, Yerevan, Republic of Armenia

⁶Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

†Current address: Department of Medicine, NYU Langone Medical Center, New York, NY 10010, USA

#Corresponding author: David A. Mills, Department of Viticulture and Enology, University of California, Davis, One Shields Ave., Davis, CA 95616. Tel: 530-754-7821; Fax: 530-752-0382. damills@ucdavis.edu.

28 ABSTRACT

29 The fermented milk matsoni is a traditional, national food product of both Georgia and Armenia.
30 Little is known about the effects of biogeography and milk type on the microbial biodiversity of
31 matsoni or the fungal composition of matsoni fermentations. High-throughput marker-gene
32 sequencing was used to survey the bacterial and fungal communities of matsoni from different
33 milk types and regions throughout Armenia and Georgia. Results demonstrate that both
34 production region and milk type influence matsoni microbiota, suggesting that the traditional
35 production methods preserve the transfer of unique regional microbiota from batch to batch.
36 Bacterial profiles were dominated by *Lactobacillus* and *Streptococcus* species. Yeast profiles
37 varied dramatically, with *Kluyveromyces marxianus*, *Candida famata*, *Saccharomyces*
38 *cerevisiae*, *Lodderomyces elongisporus*, and *Kluyveromyces lactis* being the most important
39 species distinguishing production regions and milk types. This survey will enable more detailed
40 capture and characterization of specific microbiota detected within these fermentations.

41

42

INTRODUCTION

Matsoni (*syn. mazun, matsoon*) is a traditional Caucasian fermented milk product largely used from ancient times. It is considered an analog of yogurt and prepared from milk of cow, sheep, goat, buffalo, or a mixture thereof. Matsoni is made from heated or pasteurized milk, cooled to 35-42°C, and inoculated with a portion (approximately 1% wt/wt) of finished matsoni, thereby maintaining an ongoing culture consortium of lactic acid bacteria and yeasts. It is fermented at 35-42°C overnight and the finished product has a characteristic pleasant, cultured milk taste and aroma. It is a national food throughout the Caucasus, where it is widely considered to have beneficial health effects, particularly for intestinal disorders, and to increase longevity.

In contrast to yogurt, which is fermented by *Lactobacillus delbrueckii* subsp. *bulgaricus* and *Streptococcus thermophilus*, matsoni microbiota comprise more abundant and different microbial species and their symbiotic associations. Essentially important is that the microbial composition of matsoni apparently differs from various Caucasian areas but is very stable and characteristic for the region of origin (Afrikyan 2009, Afrikyan 2012). Several studies have described the microbiota of matsoni fermentations, variously reporting *Lactobacillus delbrueckii* subsp. *lactis*, *L. delbrueckii* subsp. *bulgaricus*, *L. acidophilus*, *Streptococcus thermophilus*, *Lactococcus lactis*, *L. lactis* subsp. *cremoris*, *Geotrichum candidum*, *Saccharomyces*, *Candida*, and other species of yeasts (Saroukhanian 1960, Yerzinkian 1965, Yerzinkian 1971, Ter-Ghazarian 1993, Afrikyan 2009, Afrikyan 2012, Quero et al. 2014, Uchida et al. 2007, Reddy et al. 1986, Merabishvili and Chanishvili 2001). Other bacteria, including *L. helveticus*, *L. paracasei*, and *Leuconostoc lactis* are also involved in some fermentations (Quero et al. 2014). This inconsistent bacterial composition has led some to conclude that the bacterial diversity of matsoni fermentations may be influenced by region of production (Uchida et al. 2007, Afrikyan 2009, Afrikyan 2012). Such

biogeographical patterns have been described previously in other food fermentations (Bokulich and Mills 2013a, Bokulich et al. 2014b, Yu et al. 2011) and may explain regional differences in matsoni product qualities, but a comprehensive survey of regional matsoni fermentations has yet to be described. The effect of milk type on microbial composition is similarly uncharacterized in matsoni or similar dairy products.

The goal of this study was to better characterize how production region and milk type shape the bacterial and fungal constitution of matsonis in Armenia and Georgia. We used high-throughput amplicon sequencing as a culture-independent surveillance tool to characterize regional and milk-derived patterns in the bacterial and yeast populations of matsoni. Results indicate that regional and substrate-driven conditions shape the bacterial and fungal consortia of matsoni fermentations, and indicate that matsonis are a diverse source of microbial cultures for dairy fermentations.

79 METHODS

80 Sample Collection

81 Samples were collected aseptically on-site at the farms or production facilities producing
82 matsoni, or at local markets, and transported on ice to the laboratory for immediate processing. A
83 total of 194 samples were collected from different sites within 18 distinct regions in Georgia,
84 Armenia, and Nagorno Karabakh (Figure 1, Table 1). Matsoni fermentations typically last 24 h
85 and all samples were collected from finished fermentations. DNA was extracted using the ZR-96
86 Fecal DNA MiniPrep Kit (Zymo Research, Irvine, CA), with bead beating in a FastPrep-24 bead
87 beater (MP Bio, Solon, OH), and stored at –20°C until further processing.

88 Marker-Gene Sequencing Library Construction

89 Amplification and sequencing was performed as described previously for bacterial (Bokulich et
90 al. 2012) and yeast communities (Bokulich and Mills 2013b). Briefly, the V4 domain of bacterial
91 16S rRNA genes was amplified using primers F515 (5'–
92 *NNNNNNNN***GT**GTGCCAGCMGCCGCGGTAA–3') and R806 (5'–
93 GGACTACHVGGGTWTCTAAT–3') (Caporaso et al. 2011), with the forward primer modified
94 to contain a unique 8 nt barcode (italicized poly-N section of primer above) and 2 nt linker
95 sequence (bold, underlined portion) at the 5' terminus. PCR reactions contained 5-100 ng DNA
96 template, 1X GoTaq Green Master Mix (Promega, Madison, WI), 1 mM MgCl₂, and 2 pmol of
97 each primer. Reaction conditions consisted of an initial 94°C for 3 min followed by 35 cycles of
98 94°C for 45 s, 50°C for 60 s, and 72°C for 90 s, and a final extension of 72°C for 10 min. Fungal
99 internal transcribed spacer (ITS) 1 loci were amplified with primers BITS (5'–
100 *NNNNNNNN***CT**ACCTGCGGARGGATCA–3') and B58S3 (5'–

GAGATCCRTTGYTRAAAGTT–3') (Bokulich and Mills 2013b), with a unique 8 nt barcode and linker sequence incorporated in each forward primer. PCR reactions contained 5-100 ng DNA template, 1X GoTaq Green Master Mix (Promega, Madison, WI), 1 mM MgCl₂, and 2 pmol of each primer. Reaction conditions consisted of an initial 95°C for 2 min followed by 40 cycles of 95°C for 30 s, 55°C for 30 s, and 72°C for 60 s, and a final extension of 72°C for 5 min. Amplicons were combined into two separate pooled samples (keeping bacterial and fungal amplicons separate) at roughly equal amplification intensity ratios, purified using the Qiaquick spin kit (Qiagen, Valencia, CA), and submitted to the UC Davis Genome Center DNA Technologies Core for Illumina paired-end library preparation, cluster generation, and 250-bp paired-end sequencing on an Illumina MiSeq sequencer (Illumina, San Diego, CA).

Data Analysis

Raw fastq files were demultiplexed, quality-filtered, and analyzed using QIIME v.1.8.0 (Caporaso et al. 2010b). The 250-bp reads were truncated at any site of more than three sequential bases receiving a quality score < Q10, and any read containing ambiguous base calls or barcode/primer errors were discarded, as were reads with < 75% (of total read length) consecutive high-quality base calls (Bokulich et al. 2013b). Reverse primer sequences were trimmed from the ends of ITS sequences following demultiplexing. Operational taxonomic units (OTUs) were clustered at 97% identity using the QIIME subsampled reference OTU-picking pipeline using UCLUST-ref (Edgar 2010) against either the Greengenes 16S rRNA gene database (May 2013 release) (McDonald et al. 2012) or the UNITE fungal ITS database (Abarenkov et al. 2010, Koljalg et al. 2005), modified as described previously (Bokulich and Mills 2013b). OTUs were classified taxonomically against these same databases using RDP classifier (Wang et al. 2007). Any OTU comprising less than 0.01% of total sequences for each

run were removed prior to further analysis (Bokulich et al. 2013b). Bacterial 16S rRNA gene sequences were aligned using PyNAST (Caporaso et al. 2010a) against a reference alignment of the Greengenes core set (McDonald et al. 2012). From this alignment, chimeric sequences were identified and removed using ChimeraSlayer (Haas et al. 2011) and a phylogenetic tree was generated from the filtered alignment using FastTree (Price et al. 2010). Sequences failing alignment or identified as chimera were removed prior to downstream analysis.

Beta-diversity (between-sample community dissimilarity) estimates were calculated within Phyloseq (McMurdie and Holmes, 2013) using weighted UniFrac (Lozupone and Knight 2005) distance between samples for bacterial 16S rRNA reads (evenly sampled at 1500 sequences per sample) and Bray-Curtis dissimilarity for fungal ITS reads (evenly sampled at 100 reads per sample). Non-metric multidimensional scaling (NMDS) coordinates were computed from the resulting distance matrices to compress dimensionality into two-dimensional NMDS plots, enabling visualization of sample relationships. In order to determine whether regional factors (country, region, or district of production) or substrate factors (milk type or starter type) related to differences in phylogenetic or OTU diversity, permutational MANOVA (Anderson 2001) with 999 permutations was used to test the null hypothesis that sample groups were not statistically significant. For all classifications rejecting this null hypothesis, a Kruskal-Wallis test was used to determine which taxa differed significantly (with Bonferroni error correction) between sample groups.

Results

Finished matsoni fermentation samples were collected from 194 locations spanning 18 distinct regions in Georgia, Armenia, and Nagorno Karabakh (Figure 1, Table 1). Regional variation was assessed across three different geographic scales: between countries (including Nagorno Karabakh as a separate entity), between production regions, and between individual administrative districts. The effects of ingredients on microbial composition were also tested, comparing matsonis produced from different milks (cow, goat, buffalo, sheep) and different starter cultures (i.e., inoculated with matsoni made from one of these four milks). As we expected a high degree of interaction between regional factors and ingredients, the data were dissected to test individual factors in isolation within single geographic units and within individual milk types.

Marker-gene sequencing revealed a complex relationship between regional and substrate-driven (i.e., source milk and starter type) factors and the microbial consortia of matsoni fermentations. Country, region, and district of production all influenced microbial community composition across all samples, within individual countries and regions, and within individual milk types (Table 2). Source milk and starter type also impacted diversity across all countries, within individual regions, and within individual districts. However, these effects were not universally observed and some regions and milks displayed more homogeneous bacterial or fungal diversity than others. In particular, sheep milk fermentations and Georgian samples exhibited fewer patterns related to production region, milk type, and starter type.

The complex interactions between regional factors and ingredients are apparent using Non-metric Multidimensional Scaling (NMDS) to visualize how these factors affected microbial

community similarity at multiple geographical scales and within different types of matsoni (Figure 2). Comparing all samples, matsonis from Georgia and Nagorno-Karabakh are separated by NMDS, but Armenian samples bridge both clusters; individual regions within each country are more highly differentiated (Figure 2A). Regional variation is even clearer when comparing matsonis from individual milk types (Figure 2B-D), highlighting production region as a salient factor driving the microbial composition of matsoni fermentations and suggesting that milk type also contributes. Milk type did not significantly alter microbial diversity across all samples (Table 2) and does not explain sample ordination by NMDS (data not shown). However, matsonis segregate by milk type within smaller geographic scales, e.g., within Armenia, Nagorno Karabakh, and individual production regions (Figure 2E-G). In most of these cases, regional factors still explain sample ordination to a greater degree (e.g., Armenian samples cluster by production region), but sub-clusters based on milk type are apparent within many of these regional divisions (Figure 2E-G). In some cases, starter culture type (i.e., inoculated with finished cow matsoni, goat matsoni, or matsoni made from another milk type) exerted a greater effect on microbial composition than milk type, such as in Tavush (a region of Armenia) where samples clearly cluster by starter type rather than milk type (Figure 2H).

Across regions and milk types, marker-gene sequencing detected similar bacterial compositions. Most samples were dominated by *Lactobacillus* and *Streptococcus* species (> 80% combined relative abundance) with minor populations of other bacteria (Figure 3A). *Lactococcus* and *Enterococcus* species were prevalent secondary populations observed in many samples. *Pseudomonas*, *Rummeliibacillus*, *Enterobacteriaceae*, *Erwinia*, *Wautersiella*, and *Yersinia* were also frequently observed.

Yeast populations exhibited greater variation between samples (Figure 3B). *Kluyveromyces*

marxianus and *Saccharomyces cerevisiae* displayed the greatest average abundance across all samples. Some samples were clearly dominated by other yeasts, often distinguishing individual regions and milk types. For example, *Lodderomyces elongisporus* dominated a subset of Armenian sheep's milk matsonis; *Kluyveromyces lactis*, *Candida zeylanoides*, and *C. famata* were prominent in Georgian goat matsonis; *K. lactis* and *Naumovia castellii* in some Armenian goat matsonis; *Pichia kudriavzevii* in some Armenian sheep matsonis; and *C. parapsilopsis*, *C. albicans*, and *P. guilliermondii* in various goat matsonis from Nagorno Karabakh.

Many of these microbial populations were differentially abundant between milk types and especially between regional zones, becoming defining features of certain local matsonis (Figure 4). Most *Firmicutes* (including *Lactobacillales*), *Actinobacteria*, *Enterobacteriaceae*, and other bacterial clades were differentially abundant between countries, regions, and districts at multiple geographic scales and in different milk types. Buffalo and goat milks exhibited particularly large number of regionally defining taxa, whereas regional cow and sheep matsonis had fewer defining features.

Milk and starter types had fewer differential taxa compared to regions, but several trends were still observed. Species of *Lactococcus*, *Klebsiella*, *Salinicoccus*, *Planococcaceae*, *Acinetobacter*, *Enhydrobacter*, and *Micrococcaceae* were consistently differential between milk types at multiple geographic scales. Broad swathes of *Lactobacillales* and bacilli, *Gluconobacter*, *Ralstonia*, *Enterobacteriaceae*, *Enhydrobacter*, *Pseudomonas*, and *Bifidobacterium* species defined starter types at different geographic scales. Between starters in Tavush district, where starter cultures were the primary factor driving microbial composition, various *Clostridia*, *Turicibacter*, *Brevibacterium*, *Bifidobacterium*, *Enterobacteriaceae*, *Prevotella*, and *Wautersiella* were the most defining bacterial taxa. Many of these taxa were detected at < 1 %

212 relative abundance, and influence on flavor development during active fermentation is unlikely.

213 Even fewer yeast and other fungal species were differentially abundant between milk and starter

214 types. *K. marxianus*, *K. lactis*, *Issatchenkia* spp., *Alternaria* spp., *Davidiella tassiana*, and

215 *Cladosporium cladosporioides* differentiated milk types in Nagorno Karabakh and *C.*

216 *parapsilopsis* differentiated milk types in Kotayk, but no species were differentially abundant

217 between larger geographic areas. *C. cladosporioides*, *N. castellii*, *P. kudriavzevii*, *K. marxianus*,

218 *S. cerevisiae*, and *Hanseniaspora uvarum* differentiated starter types variously across all regions

219 as well as at smaller scales. In Tavush, *Botryotinia* spp., *L. elongisporus*, and *Zygoascus*

220 *hellenicus* defined starter types.

221

Discussion

Matsoni is an interesting case study for food biogeography, as fermentation is still performed by inoculating milk with portions of previous fermentations instead of using defined starter cultures. Hence, locally unique cultures likely establish and persist, potentially lending to the regional character of this transnational product. Matsoni fermentation microbiota exhibit a large degree of biodiversity, influenced by both geography and milk type. The basic techniques for producing matsoni are highly similar in all areas but the combination of local ingredients, local techniques, and microbial biogeography inherent across these regions distinguish individual source locations. These findings explain the noted inconsistencies between earlier surveys of matsoni bacteria (Uchida et al. 2007). Matsoni fermentations are not conducted by one standardized culture, but instead display a large degree of heterogeneity, linked to the source milk and production region. Several core taxa dominated most matsoni fermentations — *Lactobacillus* spp., *Streptococcus* spp., and *K. marxianus* — but other taxa are present and distinguish matsonis of unique provenance. The bacterial profiles, consisting predominantly of a mixture of *Lactobacillus* and *Streptococcus* species, are similar to other yogurt-type fermentations, though the minor bacterial taxa (e.g., *Enterococcus*, *Pseudomonas*, *Enterobacteriaceae*) are not typical. Another factor that distinguishes matsonis from other yogurts is that the fermentations contain large populations of yeasts. *K. marxianus* was the most common, dominant yeast observed, followed by *S. cerevisiae* and *C. famata*. *K. lactis* was dominant in some Georgian samples — particularly from Shida Kartli — while *L. elongisporus*, *C. albicans*, and *C. zeylanoides* were prevalent in matsonis from various regions of Armenia. Yeasts, including species of *Kluyveromyces*, *Saccharomyces*, and *Candida*, can significantly alter the sensory profiles of dairy products through the production of alcohols, organic acids, CO₂, and proteolytic and lipolytic activities, distinguishing yeast-

fermented products from other thermophilic dairy fermentations (Roostita and Fleet 1996, Fleet 1990, Wouters et al. 2002). These behaviors may explain many of the unique characteristics of regional matsonis, including effervescence and fruity and yeasty flavors. However, more work must be done to examine the role of these cultures in flavor production during matsoni fermentation and storage. The biodiversity of matsoni contributes to the rich variety of flavors and textures typical of different matsonis (Quero et al. 2014), and may be a rich source to mine for dairy starter cultures (Wouters et al. 2002).

Milk variety also shapes matsoni microbiota, though the effect is most apparent within defined geographical scales (e.g., within a single production region). It is unclear whether these microbiota are influenced by milk chemistry, animal-associated microbiota, or other factors, and to what degree these effects reflect site-specific differences. Although some villages purvey matsonis from different milk types, these different animals may be raised on different farms, leading to differences in both local microbial inputs, as well as differences in milk handling. Milks from different animal species or breeds can contain widely different contents of fat, protein, lactose and other sugars, oligosaccharides, and other macronutrients, as well as lactoferrin, immunoglobulin, and other antimicrobial compounds (Jensen 1995, Peterson 2013). Milk composition could thus influence fermentation microbiota in matsoni or other dairy fermentations. Further studies are needed to assess what role milk chemistry may play in driving fermentation microbiota.

More work is needed to understand several aspects of microbial activity in matsoni production. Seasonal variation in ingredients can influence the microbiota of dairy fermentations (Wullschleger et al. 2013), potentially explaining variation between batches of matsoni (Quero et al. 2014). Repeated sampling of matsoni fermentations at multiple individual sites will clarify the

effect of seasonal variation on matsoni microbiota. The equipment and production environment in which matsonis are made could be important vectors for microbial transmission between batches, similar to other foods (Bokulich and Mills 2013a, Bokulich et al. 2014b, Bokulich et al. 2013a, De Filippis et al. 2013), Studying the production environment in matsoni production sites will be important to understand how processing areas contribute to the biogeographical diversity of matsoni fermentations. Finally, the influence of the microbiota on matsoni sensory properties is poorly described, particularly for low-abundance microbiota and yeasts. Thus, it is currently unknown whether regional biogeography of matsoni microbiota is actually related to the regional flavor characteristics of different matsoni fermentations. Even the dominant yeasts in these matsoni fermentations have undetermined roles that deserve better characterization for their utility in these thermophilic dairy fermentations.

This study describes the first complete culture-independent survey of matsoni bacteria, yeasts, and other fungi in a range of milk types throughout Armenia and Georgia. Results demonstrate that matsoni microbiota are shaped by region of production and milk type, indicating that the traditional method of serial inoculation preserves the transfer of unique regional microbiota from batch to batch. This survey will enable more detailed capture and characterization of specific microbiota detected within these fermentations.

Acknowledgments

The authors thank Chad Masarweh, Morgan Lee, Khatuna Varsimashvili, Lana Tolordava, Lela Tinikashvili, and Marika Gamkrelidze for technical support, and Astghik Harutyunyan and Evrik Afrikyan for guidance and advice on this work. This work was supported, in part, by the U.S.

290 Department of Energy Global Initiatives for Proliferation Prevention (GIPP) program (LBNL-
291 0225-GE and LBNL-0231-AM). The project in Georgia (P509) was funded through the Science
292 and Technology Center in Ukraine (STCU). The International Science and Technology Center
293 (ISTC) in Moscow, Russia provided the financial support to Armenia (A-1957). This work was
294 also supported in part by funding from the Peter J. Shields Endowed Chair in Dairy Food
295 Science (DAM). NAB was supported by the 2012-2013 Dannon Probiotics Fellow Program (The
296 Dannon Company, Inc.), an American Wine Society Educational Foundation Endowment Fund
297 scholarship, the Brian Williams and Samuel Adams Scholarships (American Society of Brewing
298 Chemists Foundation), a Wine Spectator scholarship, and Grant T32-GM008799 from NIH-
299 NIGMS during the completion of this work.

300

301 **References**

- 302 Abarenkov, K., et al. 2010. The UNITE database for molecular identification of fungi - recent
303 updates and future perspectives. *New Phytol.* 186:1447-1452.
- 304 Afrikian, E.G. 2009. Biological and practical aspects of lactobacteria from national food
305 Matsoon. *Electronic Journal of Natural Sciences of the NAS of Armenia* 1(12):16-23.
- 306 Afrikian, E.G. 2012. Studies of lactic-acid bacteria in Armenia with emphasis on
307 radioprotective properties. *Environmentalist* 32(2):256-268. DOI 10.1007/s10669-
308 012-9387-4
- 309 Anderson, M.J. 2001. A new method for non-parametric multivariate analysis of variance.
310 *Austral. Ecol.* 26:32-46.
- 311 Bokulich, N.A., C.M.L. Joseph, G.R. Allen, A. Benson, and D.A. Mills. 2012. Next-generation
312 sequencing reveals significant bacterial diversity of botrytized wine. *PLoS ONE*
313 7:e36357. doi:10.1371/journal.pone.0036357.
- 314 Bokulich, N.A., and D.A. Mills. 2013a. Facility-specific "house" microbiome drives microbial
315 landscapes of artisan cheesemaking plants. *Appl. Environ. Microbiol.* 79:5214-5123.
- 316 Bokulich, N.A., and D.A. Mills. 2013b. Improved Selection of internal transcribed spacer-
317 specific primers enables quantitative, ultra-high-throughput profiling of fungal
318 communities. *Appl. Environ. Microbiol.* 79:2519-2526.
- 319 Bokulich, N.A., M. Ohta, M. Lee, and D.A. Mills. 2014a. Indigenous bacteria and fungi drive
320 traditional kimoto sake fermentations. *Appl. Environ. Microbiol.* 80.
- 321 Bokulich, N.A., M. Ohta, P. Richardson, and D.A. Mills. 2013a. Monitoring seasonal changes
322 in winery-resident microbiota. *PLoS ONE* 8:e66437.
- 323 Bokulich, N.A., S. Subramanian, J.J. Faith, D. Gevers, J.I. Gordon, R. Knight, D.A. Mills, and J.G.
324 Caporaso. 2013b. Quality-filtering vastly improves diversity estimates from Illumina
325 amplicon sequencing. *Nat. Methods* 10:57-59.
- 326 Bokulich, N.A., J.H. Thorngate, P.M. Richardson, and D.A. Mills. 2014b. Microbial
327 biogeography of wine grapes is conditioned by cultivar, vintage, and climate. *PNAS*
328 111:e139-e148.
- 329 Caporaso, J.G., K. Bittinger, F.D. Bushman, T. DeSantis, G.L. Andersen, and R. Knight. 2010a.
330 PyNAST: a flexible tool for aligning sequences to a template alignment.
331 *Bioinformatics* 26:266-267.
- 332 Caporaso, J.G., et al. 2010b. Qiime allows analysis of high-throughput community sequence
333 data. *Nat. Methods* 7:335-336.
- 334 Caporaso, J.G., C.L. Lauber, W.A. Walters, D. Berg-Lyons, C.A. Lozupone, P.J. Turnbaugh, N.
335 Fierer, and R. Knight. 2011. Global patterns of 16S rRNA diversity at a depth of
336 millions of sequences per sample. *PNAS* 108:4516-4522.
- 337 De Filippis, F., A. La Stora, F. Villani, and D. Ercolini. 2013. Exploring the sources of
338 bacterial spoilers in beefsteaks by culture-independent high-throughput
339 sequencing. *PLoS One* 8:e70222.
- 340 Edgar, R.C. 2010. Search and clustering orders of magnitude faster than BLAST.
341 *Bioinformatics* 26:2460-2461.
- 342 Fleet, G.H. 1990. Yeasts in dairy products. *J. Appl. Bacteriol.* 68:199-211.

- Haas, B.J., et al. 2011. Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Res.* 21:494-504.
- Jensen, R.G. 1995. *Handbook of milk composition*. San Diego Academic Press.
- Koljal, U., et al. 2005. UNITE: a database providing web-based methods for the molecular identification of ectomycorrhizal fungi. *New Phytol.* 166:1063-1068.
- Lozupone, C.A., and R. Knight. 2005. UniFrac: A new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* 71:8228-8235.
- McDonald, D., M.N. Price, J. Goodrich, E.P. Nawrocki, T.Z. DeSantis, A. Probst, G.L. Andersen, R. Knight, and P. Hugenholtz. 2012. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME* 6:610-618.
- McMurdie, P.J., and S. Holmes. 2013. Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* 8:e61217.
- Merabishvili, M., and N. Chanishvili. 2001. Identification of the lactic acid bacterial cultures isolated from the samples of the Caucasian Matsoni. *Proc. Georgian Acad. Sci.* 27:91-95.
- Peterson, R., Cheah, W.Y., Grinyer, J., and Packer, N. 2013. Glycoconjugates in human milk: protecting infants from disease. *Glycobiology* 23(12):1425-1438.
- Price, M.N., P.S. Dehal, and A.P. Arkin. 2010. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS ONE* 5:e9490.
- Quero, G.M., V. Fusco, P.S. Cocconcelli, L. Owczarek, M. Borcakli, C. Fontana, S. Skapska, U.T. Jasinska, T. Ozturk, and M. Morea. 2014. Microbiological, physico-chemical, nutritional and sensory characterization of traditional Matsoni: Selection and use of autochthonous multiple strain cultures to extend its shelf-life. *Food Microbiol.* 38:179-191.
- Reddy, N.S., T. Miyamoto, K. Kataoka, and T. Nakae. 1986. Isolation and identification of lactic acid bacteria from milk products in Eastern Europe. *Jap. J. Dairy Food Sci.* 35:5-12.
- Roostita, R., and G.H. Fleet. 1996. Growth of yeasts in milk and associated changes to milk composition. *Int. J. Food Microbiol.* 31:205-219.
- Saroukhanian, F. 1960. Microflora of main fermentative technologies of the Armenian SSR. Publishing house of the Academy of Science of the Armenian SSR, Yerevan, p 306 (in Russ.).
- Ter-Ghazarian, S. 1993. Ecological potential of the lactic acid bacteria species and principles of floristic bacteriology. Doctor Thesis, National Academy of Science, Yerevan, Armenia, p 365 (in Russ.).
- Uchida, K., T. Urashima, N. Chanishvili, I. Arai, and H. Motoshima. 2007. Major microbiota of lactic acid bacteria from Matsoni, a traditional Georgian fermented milk. *Animal Sci. J.* 78:85-91.
- Wang, Q., G.M. Garrity, J.M. Tiedje, and J.R. Cole. 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* 73.
- Wouters, J.T.M., E.H.E. Ayad, J. Hugenholtz, and G. Smit. 2002. Microbes from raw milk for fermented dairy products. *Int. Dairy J.* 12:91-109.
- Wullschleger, S. et al. 2013. Analysis of lactic acid bacteria communities and their seasonal variations in a spontaneously fermented dairy product (Malian fènè) by applying a

389 cultivation/genotype-based binary model. *Int. Dairy J.* 29:28-35.
390 Yerzinkian, L.A. 1965. Preparation and use of therapeutic acidophilic milk and “Narine”
391 milk. Publishing house of the Academy of Science of the Armenian SSR, Yerevan, p.
392 13 (in Russ.)
393 Yerzinkian, L.A. 1971. Biological peculiarities of some races of lactic acid bacteria.
394 Publishing house of the Academy of Science of the Armenian SSR, Yerevan, p 233 (in
395 Russ.).
396 Yu, J. et al. 2011. Diversity of lactic acid bacteria associated with traditional fermented
397 dairy products in Mongolia. *J. Dairy Sci.* 94:3229-3241.
398

399

400

401 **Table 1. Description of Matsoni Samples Collected**

	Cow	Buffalo	Goat	Sheep	Total
Armenia	21	22	46	32	121
Georgia	10	10	16	8	44
Nagorno Karabakh	4	11	14	0	29
Total	35	43	76	40	194

402

403

Table 2. One-Way Permutational MANOVA Comparisons Between Regions and Substrates

Comparison Type		Subset	Comparison	Bacterial Weighted UniFrac			Fungal Bray-Curtis		
				<i>R</i>	<i>P</i>		<i>R</i>	<i>P</i>	
Regional	All Samples	All Samples	Country	0.057	0.001	***	0.084	0.001	***
			Region	0.296	0.001	***	0.278	0.001	***
			District	0.690	0.001	***	0.521	0.001	***
	Within Country	Armenia	Region	0.279	0.001	***	0.223	0.001	***
			District	0.768	0.001	***	0.505	0.001	***
		Georgia	Region	0.186	0.275		0.338	0.006	**
			District	0.397	0.444		0.645	0.006	**
	Within Region	Nagorno Karabakh	District	0.246	0.003	*	0.125	0.11	
			Kotayk	0.808	0.001	***	0.350	0.001	***
			Tavush	0.382	0.017	*	0.623	0.001	***
	Within Milk Type	Buffalo	Country	0.147	0.001	***	0.271	0.001	***
			Region	0.706	0.001	***	0.696	0.001	***
			District	0.895	0.001	***	0.847	0.001	***
		Cow	Country	0.172	0.003	*	0.139	0.007	**
			Region	0.722	0.001	***	0.387	0.007	**
			District	0.849	0.001	***	0.636	0.009	**
		Goat	Country	0.098	0.002	**	0.167	0.001	***
			Region	0.483	0.002	**	0.392	0.001	***
			District	0.596	0.01	*	0.680	0.001	***
		Sheep	Country	0.062	0.043	*	0.043	0.118	
			Region	0.605	0.001	***	0.378	0.205	
			District	0.796	0.001	***	0.795	0.001	***
Substrate	All Samples	All Samples	Milk	0.227	0.001	***	0.050	0.001	***
			Starter	0.360	0.001	***	0.094	0.001	***
	Within Country	Armenia	Milk	0.328	0.001	***	0.064	0.001	***
			Starter	0.482	0.001	***	0.145	0.001	***
		Georgia	Milk	0.044	0.781		0.116	0.185	
			Starter	0.044	0.782		0.116	0.213	
	Within Region	Nagorno Karabakh	Milk	0.193	0.022	*	0.404	0.001	***
			Starter	0.241	0.001	***	0.106	0.021	*
		Kotayk	Milk	0.589	0.001	***	0.143	0.002	**
			Starter	0.531	0.002	**	0.171	0.001	***
		Tavush	Milk	0.298	0.023	*	0.175	0.152	
			Starter	0.371	0.001	***	0.553	0.001	***

FIGURE LEGENDS

Figure 1. Matsoni sample collection map. Each point represents the coordinates of a single sampling site within Armenia (red), Georgia (blue), or the Nagorno Karabakh region (green, bordered by dashed line).

Figure 2. Matsoni microbial communities vary by region, starter, and milk type. Non-metric multidimensional scaling plots of bacterial and fungal communities across different sample subsets. Each pair of panels depicts non-metric multidimensional scaling plots of bacterial (left panels) and fungal communities (right panels) within sample subsets. The title above each twin panel indicates the sample subset, followed by the metadata categories used to determine the color and shape of each point, respectively, in the following format: (subset): (color category) × (shape category). Individual samples are represented by a single point, and the proximity between points indicates the degree of similarity between those samples. In the plots comparing production regions (left-hand side of figure), regional variation is observed as clusters of samples from different regions (colors); the separation of these clusters indicates that the microbial communities within the samples identify this region (or regions) from other, disparate clusters. In the plots comparing substrates (milk and starter types) within sub-regions (right-hand side of the figure), samples cluster by milk type, starter type, or sub-clusters explained by milk type are apparent within the regional clusters (e.g., the Armenia and Kotayk plots).

Figure 3. Microbial composition of matsoni fermentations. Bar plots indicate the average relative abundances of bacteria (left) and fungi (right) in matsoni fermentations from different production regions, milk types, and starter types within the Armenian region of Tavush. Only taxa detected at 1% or more average abundance in at least one sample subset are shown. The

length of each bar represents the average relative abundance of that taxon in the designated sample subset; empty space represents the sum of other species detected at less than 1% average abundance in all sample subsets. NK = Nagorno Karabakh.

Figure 4. Microbial populations exhibit regional and substrate-driven patterns of abundance. Heat map indicates which bacteria and fungi are differentially abundant between production regions and substrate types across various sample subsets, according to Kruskal-Wallis tests. Microbes are ordered along rows according to taxonomic affiliation (as indicated by the taxonomic dendrogram). Category comparisons between regions and substrates are ordered by column, as are the sample subsets in which these comparisons were made, as indicated below. Bonferroni-corrected Kruskal-Wallis *P* values for each organism x comparison are indicated at the intersections between each row and column. Only organisms with at least one significant interaction are shown.

Figure 1

Figure 2

Figure 3

Figure 4