Journal of Archaeological Science 45 (2014) 178–186

Contents lists available at ScienceDirect

Journal of Archaeological Science

journal homepage: http://www.elsevier.com/locate/jas

Proteomics evidence for kefir dairy in Early Bronze Age China

Yimin Yang [a,b,1], Anna Shevchenko [c,1], Andrea Knaust [c], Idelisi Abuduresule [d], Wenying Li [d], Xingjun Hu [d], Changsui Wang [a,*], Andrej Shevchenko [c,*]

[a] Department of Archaeometry, University of Chinese Academy of Sciences, Beijing 100049, PR China
[b] Key Laboratory of Vertebrate Evolution and Human Origins of Chinese Academy of Sciences, Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences, Beijing 100044, PR China
[c] MPI of Molecular Cell Biology and Genetics, 01307 Dresden, Germany
[d] Xinjiang Cultural Relics and Archaeology Institute, Ürümchi 830000, PR China

ARTICLE INFO

Article history:
Received 29 July 2013
Received in revised form
27 January 2014
Accepted 7 February 2014
Available online 18 February 2014

Keywords:
Archaeology
Ancient diet
Ancient dairy
Bronze Age
Eastern Eurasia
Proteomics

ABSTRACT

Cheese making has been inferred at several sites in northern Europe as early as the 6th millennium BC and was common in Egypt and Mesopotamia in 3rd millennium BC. However, the remains of ancient cheeses have never been found and recipes of ancient dairy, its production scale, social and economic impact remain poorly understood. Here we present direct proteomics evidence for the production of an earliest known cheese that was found as an organic mass associated with the mummies of Early Bronze Age cemetery of Xiaohe (1980–1450 BC) in Xinjiang, China. Kefir fermentation of ruminant milk by a symbiotic culture of *Lactobacillus kefiranofaciens* and other lactic acid bacteria and yeasts was the basis of robust, scalable, probiotic, lactose-free dairy and a key technological advance that introduced economic benefits of extensive herding into a semi-pastoral household of the Eastern Eurasia population already in the Early Bronze Age.

1. Introduction

Milk (Copley et al., 2003, Dunne et al., 2012, Evershed et al., 2008) and cheese (Salque et al., 2012) emerged in a human diet already in the 6th millennium BC. Despite being a milestone achievement in the nutrition history (Krausmann, 2004), social and economic impact of early cheese making is still poorly understood. Indeed, did it belong to a staple food already in the Bronze Age, or was it only made on specific occasions as a ritual or afterlife food? Sizable dairy implies extensive herding and collecting large season-dependent quantities of milk that should be rapidly processed under extreme unhygienic conditions. What production scale was achievable in antiquity and what labor efforts did it entail? Was the process efficient in utilizing raw milk and was it offering a palette of dairy products? What were the nutritional and economic (shelf life, transportability, or even the taste) properties of ancient cheeses? We argue that learning the technological aspects of ancient dairy holds a key for understanding its economic, social and cultural role by estimating the production scale, labor costs and nutritional properties of dairy products. Furthermore, commonalities between dairy recipes used in geographically distinct regions might be indicative of the cultural exchange.

Our understanding of ancient dairy remains poor (Salque et al., 2012) because no specimen of ancient cheeses suitable for the rigorous physicochemical characterization was available. Not surprisingly, the discovery of earliest known cheese making (Salque et al., 2012) relied upon the analyses of residual fats absorbed into pottery shards and was supported with circumstantial archeological and ethnographic evidence. However, the characterization of glycerolipids and free fatty acids by GC–MS as well as corresponding $\delta^{13}C$ and $\Delta^{13}C$ values could only establish their ruminant milk origin, yet their molecular compositions bore no hallmarks of milk processing activities. In contrast, the in-depth characterization of ancient cheese proteins could be revealing: different dairy recipes may specifically bias the curd composition as compared to raw milk or alter protein sequences in a recognizable process-dependent manner. Once the physicochemical analyses nail down the plausible dairy recipe, it could be reproduced and compositions of ancient and contemporary products compared.

Abbreviations: AMS, accelerator mass spectrometry; DTT, dithiothreitol; GC–MS, gas chromatography–mass spectrometry; FT IR, Fourier transform infrared spectroscopy; LAB, lactic acid bacteria; LC–MS/MS, liquid chromatography–tandem mass spectrometry; SDS, sodium dodecylsulfate.

* Corresponding authors.
E-mail addresses: cswang@ucas.ac.cn (C. Wang), shevchenko@mpi-cbg.de (A. Shevchenko).

[1] Equal contribution of Yimin Yang and Anna Shevchenko.

http://dx.doi.org/10.1016/j.jas.2014.02.005
0305-4403/

Proteomics was applied for the characterization of ancient samples in diverse analytical contexts (Buckley et al., 2013, Chambery et al., 2009, Hong et al., 2012, Leo et al., 2009, Solazzo et al., 2008). In general, proteomics technologies provide compositional information in several ways (reviewed in Aebersold and Mann, 2003): first (and the most commonly used) approach is identifying proteins by matching tandem mass spectra of individual peptides to cognate sequences in a protein database. Secondly, proteomics may provide a quantitative estimate of the protein composition: although it is impossible to quantify individual proteins without representative peptide standards, the intensities of peptides signals are reflective of their abundance and can be used for relative comparison between samples having similar protein compositions. Last but not least, proteomics could identify peptides specifically modified by phosphorylation, deamidation, oxidation *etc*. Accurate peptide mapping may also reveal how proteins were processed by limited proteolysis or non-enzymatic degradation. Modern analytical instrumentation enables the characterization of proteins at the low femtomole levels even in crude mixtures in which the abundance difference between major and minor components exceeds 1000-fold.

Here we report how a fortunate finding of extremely well preserved specimen of Early Bronze Age cheeses and their proteomics characterization shed light on a dairy technology that was conceived in antiquity and persisted almost unchanged till present times. We provide evidence that, despite being extraordinary simple, it possessed the necessary qualities for supporting the economic expansion of ruminant animal herding into Eastern Eurasia.

2. Materials and methods

2.1. Samples from Xiaohe tombs

Samples were collected during archaeological excavations in 2002–2004 (CRAIXAR, 2007). According to accelerator mass spectrometry (AMS) C14-dating, the calibrated date (68% confidence) of M12 is 1615–1530 BC, M29 is 1615–1515 BC, and M34 is 1610–1440 BC (Table 1) and relies upon the analyses of organic materials, *e.g.* plant seeds or animal tissues from the corresponding tombs. Details on samples characterization by FT IR, ion chromatography and elemental composition analyses are provided in Supplementary Materials.

2.2. Proteomics of ancient dairy foods

A piece of 5–15 mg was cut from a sample, transferred into 1.5 ml Eppendorf tube and disintegrated into fine powder by stirring with a pestle. Then 50–80 μl (depending on the sample size) of 65 mM Tris HCl buffer (pH 6.8) containing 10% sodium dodecylsulfate (SDS) and 10 mM dithiothreitol (DTT) were added and the tube was sonicated for 45 min. Then the slurry (note that insoluble debris was not removed) was loaded on a pre-cast 1 mm 12% polyacrylamide gel (BioRad Laboratories GmbH, Munich, Germany). To avoid carry-over of the protein material we loaded one sample per each gel and ran gels individually. Once the front migrated to *ca*. 4 cm, electrophoresis was terminated and the gel slab was stained with Coomassie, destained in 50% methanol in 5% acetic acid and cut into 4 or 5 slices each of which was independently digested with trypsin (Shevchenko et al., 2006) and recovered peptides analyzed by LC–MS/MS (see Supplementary Materials for details). Samples of contemporary milk, kefir starter grains and self-made kefir were dried in a vacuum centrifuge and processed in the same way.

Proteins were identified by Mascot v.2.2.04 software (Matrix Sciences Ltd, London, UK) by searching against a comprehensive (all species) NCBI protein sequences database (compiled in September 2012 from 20,308,369 entries) considering typical age-related protein modifications (Leo et al., 2011, Shevchenko et al., 2001). Identifications were accepted if proteins were matched with two or more peptides: each peptide comprised more than seven amino acid residues and its peptide ion score exceeded the MASCOT homology threshold and also was above the value of 30. Relative abundance of protein groups was determined by geLC–MS/MS label-free proteomics (Vasilj et al., 2012) using Progenesis software (NonLinear Dynamics, Newcastle). For each sample the abundances of peptides detected by LC–MS/MS of digests of several gel slices were summed up (Reidel et al., 2011). The abundance of peptides originating from proteins of each group (such as milk

Table 1
Composition of Xiaohe dairy foods.[a]

Tomb/gender	Salt content, wt%[e]	Protein composition[c]		Proteins from microorganisms			
		Protein content,[b] wt %	Identified proteins: from ruminant milk/in total	Lactic acid bacteria (LAB) Identified proteins: LAB species	Yeasts Identified proteins/ yeast species	Relative abundance LAB and yeasts, %	Relative abundance Mold, %
M11/f		74	16/39	4: LK, LB	8: KM, KL, SC, Y	0.5	0.1
M12/f	0.3	71	18/21	n.d.	1: KM	<0.1	<0.1
M13a/f		71	15/34	1: LK	1: KM	<0.1	0.2
M13b/f	0.6	70	38/40	n.d.	n.d.	n.d.	<0.1
M22a/m?		70	17/26	2: LK	3: KM, SC, KL, Y	<0.1	<0.1
M22b/m?		72	15/31	3: LK, LB	5: KM, KL, SC	0.5	0.4
M24/m	1.0	73	19/68	3: LK, LB	6: KM, Y	0.3	1.6
M25/m	0.5	74	31/52	7: LK, LB	4: KM	0.1	0.2
M28/f	0.6	65	14/34	1: LB	2: KM	<0.1	0.5
M29/m		63	20/59	3: LK	12: KM, SC, Y	0.2	1
M33/m		75	16/25	1: LK	2: KM	<0.1	0.1
M34/m	1.6	76	18/33	5: LK, LB	1: KM	0.2	0.2
Cattle milk[d]			79/79	n.d.	n.d.	n.d.	n.d.
Kefir curd[d]			80/153	15: LK, LB	56: KM, KL, SC, Y	1.0	n.d.

[a] Sample M22c was identified as a goat adipose fat and is not included in this table; in tombs names a and b indicate samples independently collected from different locations in the same tomb; m and f stand for male and female gender of the tomb owner, respectively.
[b] Assuming *ca* 16 wt% nitrogen content in proteins.
[c] Includes proteins from ruminant milk, LAB, yeasts and mold. M28 contained proteins from wheat grains. LAB species: LK – *L. kefiranofaciens*; LB – *Lactobacillus* sp.; Yeast species: KM – *K. marxianus*; KL – *Kluyveromyces lactis*; SC – *S. cerevisiae*; Y – yeast species undistinguished by proteomics; n.d. – not detected.
[d] Self-made dairy products are included as a reference.
[e] Only provided for samples having the stoichiometric content of sodium and chloride.

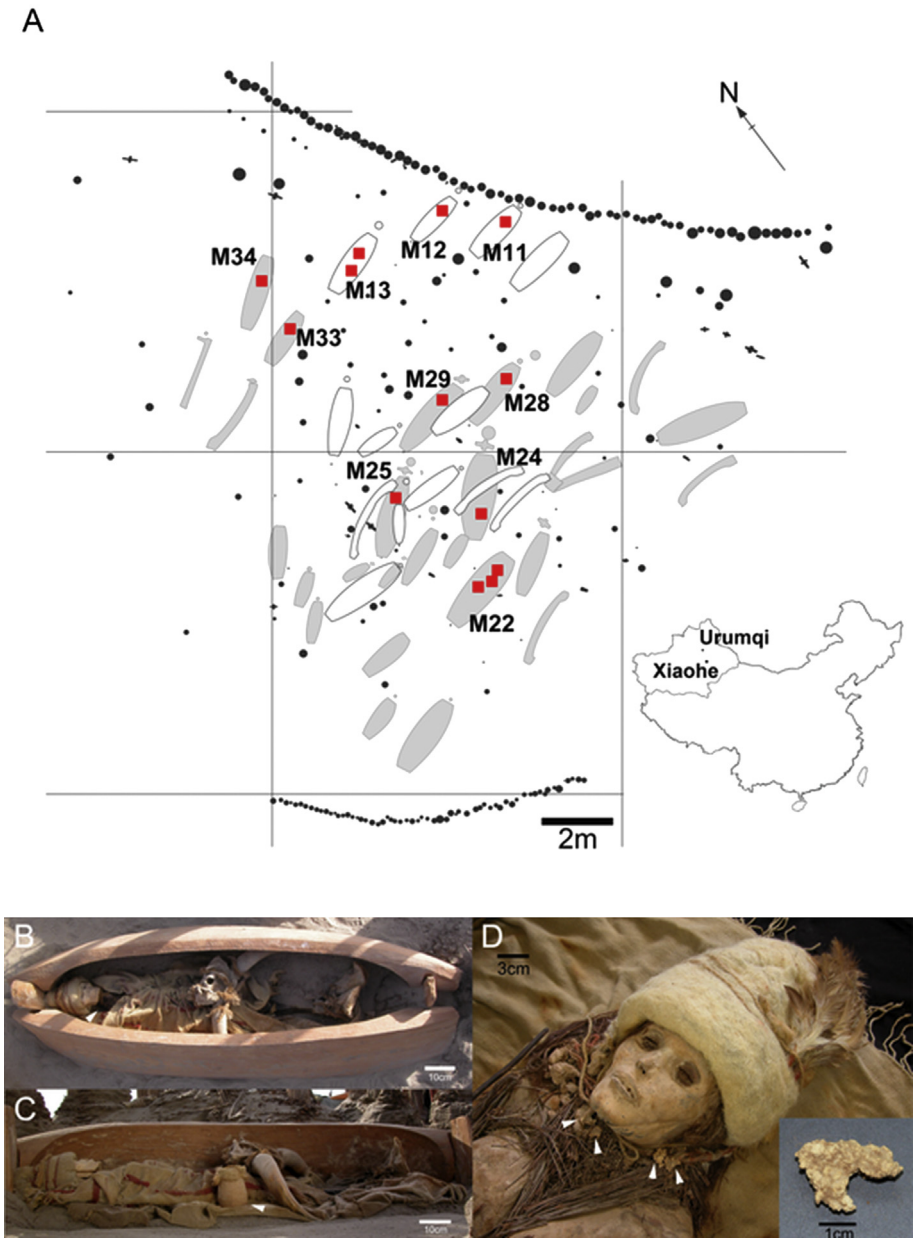

Fig. 1. Dairy foods from the tombs of Xiaohe. A: Scheme of the Xiaohe cemetery showing the location of tombs of the first (unfilled shapes) and second (shaded shapes) layers (1650–1450 BC), respectively. The tombs in which organic residues were collected are designated with red squares; black dots indicate other cultural artifacts (CRAIXAR, 2007). Inset at the bottom right corner indicates the location of the Xiaohe site at the territory of modern China. Panel B (top view) and C (side view) of a female mummy from the tomb M13. Panel D: detailed view of the head and chest of a female mummy from the tomb M11. Lumps of organic material are indicated with arrowheads, see inset for an enlarged view. Samples M13a and M11 were collected from the neck and chest of mummified bodies, while the sample M13b came from the woven grass basket found next to the mummy's waist (as indicated by the arrowhead in panel C).

proteins; yeast and LAB proteins, caseins, *etc*) was combined and the abundance of each group was normalized to the total abundance of all groups. Full details on the protein identification and quantification methods are provided in Supplementary Materials.

2.3. Reference dairy products made from contemporary raw milk

Milk and "feta" cheese were purchased from a local food store. Tibetan kefir starter grains were obtained from a private vendor in China; Caucasian starter from GUNKA (Hartenstein, Germany). Contemporary (self-made) kefir was prepared as follows: 4.5 g of kefir grains were added to 100 ml of pasteurized non-homogenized milk (Hof Mahlitzsch, Nossen, Germany) and incubated for 24 h at 20 °C. Then milk was replaced with a fresh 500 ml portion and further incubated for 72 h. Curdled proteins and whey were separated by filtering. Collected samples were dried in a vacuum centrifuge and analyzed as described above.

3. Results

3.1. Protein-rich foods found in Xiaohe tombs

Xiaohe cemetery (40°20′11″N; 88°40′20.3″E), also known as Ördek's necropolis or Small River Cemetery Number 5 was discovered by Sweden archaeologist Folke Bergman in 1934 (Bergman et al., 1939). It is located in the Taklamakan desert of the Tarim

Basin near a dried-up riverbed of the Kongque river in Xinjiang region, China. The cemetery was built on a large natural dune and, approximately at its center, divided by a wooden palisade into northern and southern areas. Tombs at the northern area sustained severe damage, while in the southern area well preserved tombs were arranged in five archaeologically distinct layers with the lowest layer of occupation dated to *ca.* 3980 ± 40 BP (Li et al., 2010). Dead bodies were placed into massive wooden coffins resembling overturned boats, which were covered by several layers of cowhide effectively sealing them from air, water and sand. The arid climate together with saline soil preserved wooden coffins with "caucasoid" mummies, woolen textiles, bundles of ephedra twigs, bits of bovine or ovicaprid ears, plant seeds and woven grass baskets (reviewed in Mair, 2006). The abundance of woolen cloths, cattle and goat horns indicated that Xiaohe population extensively practiced herding.

Altogether, 33 tombs were found in the layers 1 and 2 and dated by carbon-14 to 1650–1450 BC (CRAIXAR, 2007) (Fig. 1A). In several tombs 1–2 cm sized lumps of a yellowish organic material of irregular shape and uneven surface were found at the necks and chests of the mummified bodies (Fig. 1B and Supplementary Materials Fig. S1), while the site contained no pottery that might be associated with making or consuming the foods.

We collected 13 samples from 10 tombs in layers 1 and 2 at the southern part of Xiaohe (Fig. 1, Table 1, Supplementary Materials Fig. S1) and examined their composition by FT IR spectroscopy and elemental analyses, which both indicated the mostly proteinaceous origin of 12 samples, while lipids there were undetectable (Table 1and Supplementary Materials Table S1, Fig. S2). Although M22c sample (not included into Table 1) looked similar to others, FT IR suggested it was enriched in lipids and depleted in proteins (Supplementary Materials Fig. S2).

3.2. Protein-rich samples consist of strained cattle milk curd

From a few milligrams of each sample we extracted proteins by 10% sodium dodecylsulfate (SDS) and subjected the extracts to quantitative proteomic analyses by geLC–MS/MS (Vasilj et al., 2012). In each sample we identified 21–68 proteins of diverse organismal origin that fell into four major groups (Supplementary Dataset S1): the bulk of each sample consisted of 12–16 cattle milk proteins (Table 2) with alpha S1-, alpha S2-, beta- and kappa-caseins and beta-lactoglobulin being the most abundant components with the exception of M22a and M25, which also contained proteins from goat and/or sheep milk. The second group comprised lactic acid bacteria *Lactobacillus* proteins; the third contained yeast proteins from *Saccharomycetaceae* family (Table 3) and the fourth group contained mold proteins mainly of the *Aspergillus* genus. Wheat proteins were additionally found in M28, which contained plant seeds (Supplementary Dataset S1 and Fig. S1). Although some proteins were highly homologous between organisms, we were able to identify unique peptides unequivocally attributing their species of origin: MS BLAST search tool (Habermann et al., 2004; Junqueira et al., 2008) was applied to rule out occasional cross-species peptide identities (Fig. 2). Human background proteins

Table 2
Ruminant milk proteins along with the number of matched peptides identified in Xiaohe foods.

Protein name	Gene identifier	M11	M12	M13a	M13b	M22a	M22b	M24	M25	M29	M28	M33	M34
Caseins													
Casein alpha-S1	gi7c159793209	21	16	13	11	15 + 5[a]	12	13	14	14	13	12	11
Casein alpha-S2	gi7c27806963	17	13	14	7	18	8	16	24 + 13[a]	21	15	16	15
Casein beta	gi7c115660	14	13	10	11	13	9	9	12 + 9[a]	12	13	9	9
Casein kappa	gi7c162811	10	7	8	7	9	4	9	7	10	7	8	5
Whey proteins													
Lactadherin	gi7c296475529	15	8	12	15	13	8	15	14	13	12	9	12
Butyrophilin	gi7c2266432	12	7	10	7	6	6	11	13 + 10[a]	10	8	7	7
Xanthine dehydrogenase	gi7c1620375	11	11	13	6	9	12	15	16	11	9	9	12
Lactoglobulin beta	gi7c520	8	11	8	13	9	7	8	8	8	6	6	6
Complement component C3	gi7c83764016	9		11	18	7		13	15	11		9	
Lipoprotein lipase	gi7c163305	4	5			5		5	5	3		5	
Alpha-lactalbumin	gi7c68	2			2	2				2	2	2	2
Fatty acid-binding protein	gi7c27805809	4	2	4		3		5	5	4	2	4	
Adipose differentiation-related protein-like	gi7c18032251	2	5					4			2		
Component PP3	gi7c741536	2	4	6		5	3	3	6	6		2	6
Mucin	gi7c16755594		2				3		3				
ATP-binding cassette sub-family G	gi7c112817615		4			4	4	7	6	3	3	3	3
Serum albumin	gi7c229552		10	10	10	6		9	3	3		3	7
Perilipin-2	gi7c27806759		2	2		5			7	5			2
Secretoglobin family 1D	gi7c118150406	2	2					2					
Alpha-1B-glycoprotein	gi7c114053019				9								
Immunoglobulins *(total)*					81		3		3	4			
Factor XIIa inhibitor	gi7c27807349				6				2				
Serotransferrin	gi7c114326282				5								
Serpin A3-1	gi7c31340900				3								
Alpha-2-macroglobulin	gi7c157954061				4								
Kininogen I	gi7c488				4								
Antigen CD14	gi7c41386760				3								
Alpha-1-antiproteinase	gi7c27806941				2								
Gamma-glutamyltranspeptidase	gi7c329664306				2								
Polipoprotein E	gi7c312893				2								
Fatty acid synthase	gi7c60592790							5	3				
Lactoperoxidase	gi7c27806851							2	2				2
Cathelicidin	gi7c27807341								2	2			
Glycoprotein 2	gi7c113912055								2				
Glycosylation-dependent cell adhesion molecule 1	gi7c27807339										4		
Lactoferrin	gi7c408926				5								

[a] Peptides from bovine, goat and sheep proteins.

typically associated with sample handling (Supplementary Dataset S2) were disregarded.

We further employed label-free proteomics to determine the relative abundance of each protein group (Table 1, Fig. 3). Since milk proteins might have been extensively degraded, we accounted for the abundances of both tryptic and non-tryptic peptides and the abundances of peptides detected in the digests of several gel slices were combined. Then the abundances of peptides originating from proteins of each group were summed and the abundance of each group was normalized to the total abundance of all groups.

Proteomic analyses indicated that Xiaohe samples (except M13b) consisted of strained and delipidated curd: compared to raw milk, they were highly enriched in caseins and *ca.* 100-fold depleted in major whey proteins: serum albumin, beta-lactoglobulin and alpha-lactalbumin (Fig. 3). We next asked how was this curd made? Apparently, curdling was not performed by boiling the acidified milk because rapid precipitation of denatured proteins usually yields the curd with no significant compositional bias as compared to raw milk. Contrary, fermentation coagulates and enriches caseins in the curd, while globulins and albumins largely remain in a soluble (whey) fraction. Rennet – an enzyme complex from calf intestine (Kumar et al., 2010) has been used in cheese-making since ancient times (Geetha et al., 1996) and it is conceivable it might have been known also to the herdsmen of Xiaohe. Chymosin, a major proteolytic enzyme of the ruminant rennet, cleaves kappa-casein protein (gi|36988716) between ^{126}Phe –^{127}Met amino acid residues triggering its aggregation, disintegration of micelles and milk curdling (Jolles et al., 1968). We determined the ratio of abundances of chymosin-cleaved peptides to all peptides covering ^{126}Phe–^{127}Met in the kappa-casein sequence (Supplementary Dataset S3 and Fig. 4). In a contemporary "Feta" cheese that was made by rennet curdling the corresponding peptide was cleaved almost completely (>97%), compared to less than 7% in all Xiaohe samples (Fig. 4) except the sample M28 that also contained wheat proteins (Supplementary Dataset S1).

Apparently, at Xiaohe milk was not curdled by a ruminant rennet, while some chymosin-like cleavage of kappa-casein might occur during food preparation and storage (Hinz et al., 2012). We therefore hypothesized that instead, in Xiaohe dairy foods milk could have been microbiologically fermented.

3.3. Xiaohe dairy food is a kefir cheese made from skimmed milk

We further noticed that ten Xiaohe samples contained proteins from LAB including *Lactobacillus kefiranofaciens*, and various *Saccharomycetaceae* yeasts, a compositional signature of the symbiotic culture of microorganisms traditionally used for making kefir – light alcoholic carbonated dairy beverage (Garrote et al., 2001; Hui, 2004; Miguel et al., 2010). Usually, kefir starter grains – a lumpy mass of kefiran polysaccharide and curdled milk, in which more than 80% of total proteins originate from *Lactobacillus* sp. and *Saccharomycetaceae* yeasts (Supplementary Dataset S1) are incubated with raw milk for several days and then recovered by decantation and re-used (Motaghi et al., 1997), while the fermented milk (a kefir) is either consumed as a beverage or further processed into cheese-like products. We purchased starter grains of Tibetan and Caucasian origin from local providers, made kefir according to a conventional recipe, strained the curdled protein mass and

Table 3
The number of peptides matched to proteins from LAB and yeasts by the proteomics analysis of Xiaohe foods.

Protein name	Gene identifier[a]	Organism[b]	M11	M13a	M22a	M22b	M24	M25	M28	M29	M33	M34
LAB proteins												
Glyceraldehyde-3-phosphate dehydrogenase	gi7c336054463	*L. kefiranofaciens*	7	4	6	5	8	7	2	7	3	9
Surface layer protein	gi7c336055311	*L. kefiranofaciens*	3					5				
Surface layer protein	gi7c1054802	*L. helveticus*						2				
Phosphopyruvate hydratase	gi7c227893416	*Lactobacillus ultunensis*	2				2	3				
Pyruvate kinase	gi7c336054221	*Lactobacillus sp.*	2		3	2	2	4		3		3
D-lactate dehydrogenase	gi7c104773324	*Lactobacillus sp.*						2				
Fructose-bisphosphate aldolase	gi7c104774377	*Lactobacillus sp.*						2				
ATP-dependent protease	gi7c161508175	*Lactobacillus sp.*										3
Conserved hypothetical protein	gi7c260101505	*L. helveticus*										3
Elongation factor Tu	gi7c124377108	*L. kefiranofaciens*								2		
Glutamine synthetase	gi7c336053793	*L. kefiranofaciens*										3
Yeasts proteins												
Glyceraldehyde-3-phosphate dehydrogenase 2	gi7c1245703	*K. marxianus*	10			6	6	5	2	8	2	2
Glyceraldehyde-3-phosphate dehydrogenase 1	gi7c116247787	*K. marxianus*	7	2	5	7	5	5	2	5	2	
Pyruvate decarboxylase	gi7c416888	*K. marxianus*	7							2		
Enolase 2	gi7c6321968	*S. cerevisiae*	7		2	3		2		4		
Alcohol dehydrogenase 2	gi7c12229579	*K. marxianus*	5			3				3		
70 kDal heat shock protein	gi7c172713	*Yeast sp.*	4									
Hypothetical protein	gi7c50312181	*K. lactis*	3		5	3	3	2		3		
Unnamed protein	gi7c2867	*K. lactis*	3									
Hypothetical protein	gi7c365984825	*Naumovozyma sp.*			2							
Enolase 1	gi7c260944716	*Candida lusitaniae*					3					
Hypothetical protein	gi7c156836648	Yeast sp.					2					
40S ribosomal protein S9	gi7c213402101	Yeast sp.								2		
tsa1p	gi7c401624401	Yeast sp.								2		
Histone H2b	gi7c6320430	Yeast sp.					2			4		

[a] If homologous proteins from multiple species could not be distinguished only one GI is shown.
[b] *Lactobacillus sp.* and yeast sp. indicate that homologous proteins from various *Lactobacillus* or yeast species were matched.

A **gi|336055311 surface layer protein [Lactobacillus kefiranofaciens]**

```
  1 MKKNLRIVSV AAAALLAVAP IAATAMPVNA ATTATTATTA TNKPTVDLSG AGSVSESKDT VNVTPSFTLT SAIATNSGVV TTPATLQGSI EASLNGTSVT
101 ADVADVAKDV TLKNGKQVVY SYDNESKKLT NNLGAFDSQA KKAYVEAGTS YTMTLSGVGF SFGKANANKT LTFKLPKGVT VEGANYKDGK VTLDQYGNVT
201 GLKFTLDVKA YNSENTSAVS FYDAKSGLVA PQGSYMTLAQ NGDLNVNTLL TELKGKYEAM QFQSGKFETV NVNTTADDVK AELEKAGIKV DAAGNFEAPD
301 TFTVTLNAKS DDNGKTASLP VVVTVPNGKS TVAPSVSKTI MHNAYYYNKE GKRVGTEKAT RYDSVTVAPK TTTINGKTYY QVVENGKAVD KYINAGNIDG
401 TKRTLKHNAY VYKTSKKRAN KVVLKKGDTV TTYGASYKFK NGKRYYKIGN NTDKTYVKVA NFQ
```

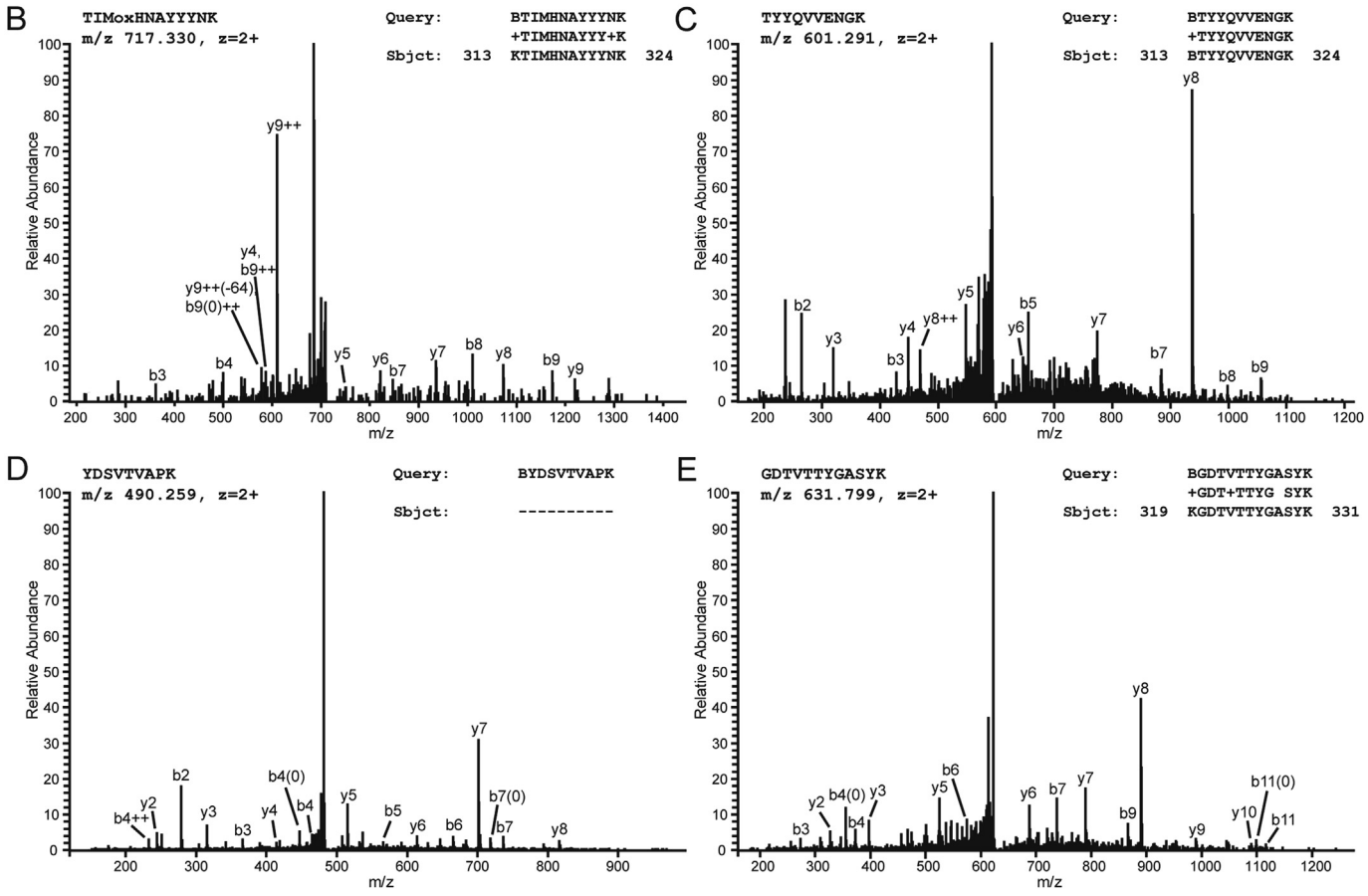

Fig. 2. Xiaohe dairy foods contain proteins from *L. kefiranofaciens*, a hallmark of the kefir fermentation. Proteomics evidence acquired for M25 sample are presented here as an example. Four tryptic peptides matched Surface Layer Protein gi| 336055311 from *L. kefiranofaciens* (highlighted in its sequence in panel A) and their MS/MS spectra are presented in panels B–E. Peptide hits were further checked for closely-related sequences by MS BLAST and the produced sequence alignments are shown next to the spectra. Query stands for the *L. kefiranofaciens* sequence and Subject for the aligned sequence from a homologous protein. In panel B the protein homologues are from *Lactobacillus helveticus*, *Lactobacillus crispatus*, *Lactobacillus amylovorus*; in panel C from 7 species of *Lactobacillus sp*; in panel E from *L. helveticus*; the peptide in panel D produced no significant alignments. In the Query symbol B stands for a trypsin cleavage site preceding the peptide sequence. Altogether, 3 out of 4 peptides were unique for *L. kefiranofaciens* and may only be similar to peptides from other LAB.

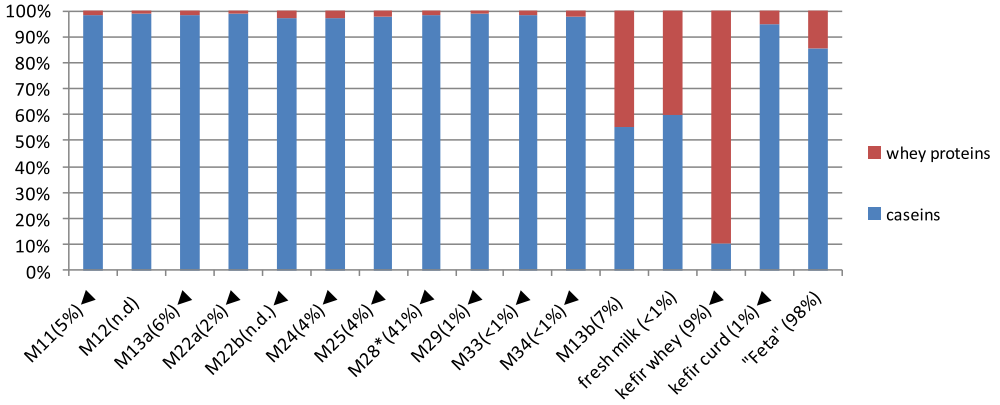

Fig. 3. Xiaohe dairy foods consist of strained ruminant curd. As compared to raw milk and milk whey, Xiaohe dairy foods were highly enriched in caseins, depleted of whey proteins and bore no hallmarks of the rennet fermentation (% of the chymosine cleavage of kappa casein is indicated at each sample). Samples designated with the filled arrowheads contained up to 1% of yeast and LAB proteins (see Table 1 for their species of origin and relative abundances). "Feta" is a contemporary soft cheese manufactured using rennet; kefir curd – the curd strained off a self-made kefir. *M28 also contained plant seeds.

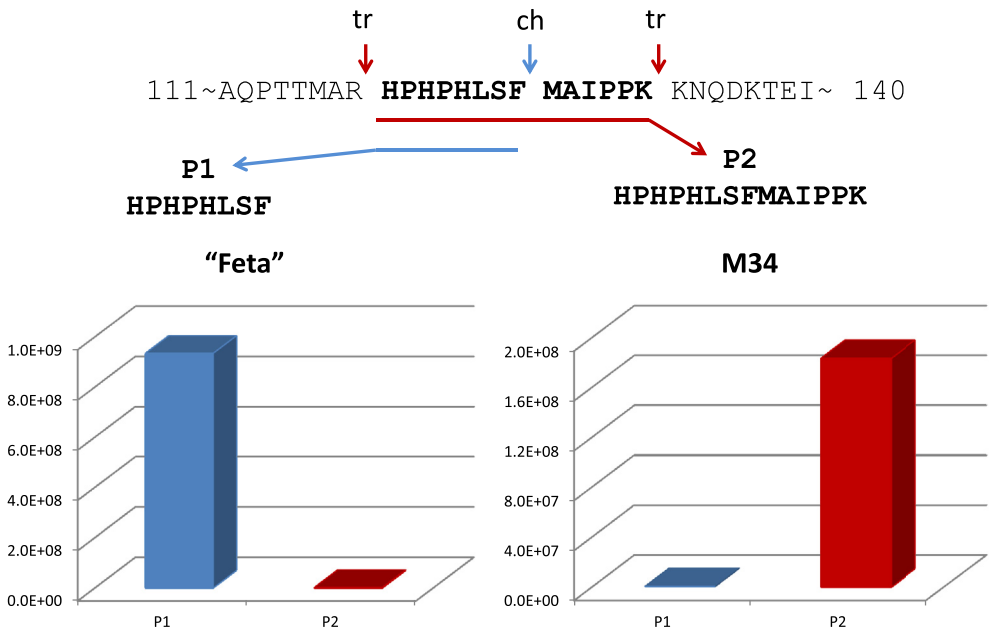

Fig. 4. Kappa-casein does not bear hallmarks of the rennet fermentation in Xiaohe dairy foods. Equal amounts of the sample M34 and, for comparison, "Feta" cheese were analyzed by geLC–MS/MS. Above is the stretch of kappa-casein protein sequences (gi|36988716); "tr" and "ch" indicate the expected cleavage sites of trypsin and chymosin, respectively. P2 represents a full tryptic peptide of intact casein; P1 is a semi-tryptic peptide produced by trypsin digestion of casein that had been cleaved by chymosin prior to trypsin digestion. The ratio of abundances of P1 and P2 in the digests of caseins from the "Feta" cheese and from M34 sample suggests that the latter was not treated by ruminant rennet.

compared its composition to the composition of ancient samples. In Xiaohe dairy foods as well as in the self-made strained kefir curd *L. kefiranofaciens* was the major source of LAB proteins. Yeast proteins were from *Kluyveromyces marxianus* (also called *Candida kefir*) and *Saccharomyces cerevisiae*, which are common to homemade kefirs in modern Tibet (Bai et al., 2010). The self-made kefir cheese contained *ca.* 80% (and up to 95%, if the curd was additionally pressed) of caseins and up to 1.5% of LAB and yeast proteins, while chymosin cleavage of kappa-casein remained at negligibly low (<1%) level (Table 1, Fig. 3).

Therefore, we recognized ten out of the total of twelve Xiaohe samples as kefir cheeses because they shared common compositional hallmarks with a self-made kefir curd (Figs. 3 and 4): they contained proteins from specific yeasts, LAB and milk in similar relative abundances and there was no evidence of the rennet fermentation. Caseins were highly enriched over whey proteins suggesting that the latter were extensively removed while straining the curdled casein mass into a solid cheese.

We further compared the protein compositions of two samples from the M13 tomb: like others, lumps of M13a were collected at the mummy's neck, while M13b was taken from a woven grass basket (or a vessel) found next to the mummy's waist (Fig. 1). Although both samples were dairy products (Table 1, Fig. 3), in M13b both the relative abundance of caseins and the chymosin cleavage rate were similar to the values observed in raw milk, but not in the kefir cheeses. In M13b we detected no yeast or LAB proteins, although mold proteins were found in both samples. At the same time, the composition of M22a and M22b, separately collected from the M22 tomb, was similar and consistent with the composition of kefir cheeses. We therefore concluded that the protein composition that we attributed to Xiaohe kefir cheeses is a *bona fide* dairy practice hallmark and that it did not arise spontaneously from unprocessed milk over thousands of years.

Was kefir cheese consumed shortly after it had been made or was it a product with extended shelf-life? Salted (brined) soft cheeses may contain up to 7% salt and are having superior storage properties compared to conventional cheeses having *ca.* 1% salt. Xiaohe dairy foods were found in the sealed wooden coffins and were well protected from the environment; however they might have been contaminated by saline soils during excavation and handling. Therefore, we extracted salts by dispersing Xiaohe foods in ultrapure water, determined the content of both sodium and chloride by ion chromatography and then only considered six out of 12 samples in which sodium and chloride were present in approximately stoichiometric ratios (±25 mol%). In five samples the estimated (w/w) salt content was below 1% and only in one sample it was 1.6% (Table 1), which is close to the values expected for raw milk. Contemporary brined cheeses, especially brands intended for extended shelf-life, are usually having 3–5-folds higher salt content. We therefore speculated that Xiaohe cheeses were not brined and likely poised for on-site consumption, rather than serving as a commodity of long-distance trade. However, this notion remains tentative and should be confirmed by broader analysis of the microelements spectrum and by comparison with representative control samples of earth particles collected from other artifacts.

The low content of triacylglycerols (TAG), including TAGs with saturated fatty acid moieties that are usually resistant towards environmental degradation (Evershed et al., 2002) (Supplementary Materials Fig. S2) may indicate that the kefir cheese was possibly made from skimmed, rather than raw milk, although no butter-like products were identified at Xiaohe. FT IR analysis of M22c sample indicated its low protein and high lipid content (Supplementary Materials Fig. S4). Its characterization by electrospray tandem mass spectrometry revealed that the internal environment of sealed coffins of Xiaohe tombs apparently preserved TAG from degradation (Supplementary Materials Fig. S4). Since TAG were also readily detectable in contemporary strained curd made from raw milk (data not shown), we speculated that in Xiaohe the bulk of milk fat might have been physically removed prior the fermentation, as now is commonly practiced in rural areas across the Eurasia steppe and also in Tibet. This notion, however, should be considered with caution since TAG might be depleted by growing mold.

Interestingly, mass spectrometric analysis of TAG extracted from this sample revealed TAG 42:0 and TAG 44:0 comprising molecular

species with short chain (C6–C12) fatty acid moieties and their relative abundance (Supplementary Materials Fig. S4) commonly regarded as a compositional hallmark of ancient milk fats (Dudd and Evershed, 1998; Mirabaud et al., 2007). However, subsequent proteomics analysis of its SDS extract identified a few typical adipose fat proteins, but no proteins common to milk fat or butter (Supplementary Materials Dataset S1). Therefore product origin identifications solely relying on glycerolipids profiles are less reliable compared to proteomics analyses.

4. Discussion

A kefir cheese found in the Xiaohe tombs is the earliest indication of extensive cattle and goat herding and milking practices in the Eastern Eurasia and the first direct material evidence of the ancient fermentation dairy. Despite being herdsmen, Xiaohe people curdled milk by a symbiotic culture of microorganisms, rather than by a ruminant rennet or by acidification/boiling, which argues against a popular notion on the origin of cheese-making tradition (Tamime, 2006). The same recipe persisted over the entire 200 years of Xiaohe and continues through present times: strained kefir cheeses (*labneh*) popular in the Middle East are still made in a similar way. While it is commonly believed that kefir was first mentioned by Marco Polo in 13th century, our findings indicated that probiotic dairy was already in use in 2nd millennium BC.

Why kefir dairy prevailed at Xiaohe? It was rapidly scalable, did not require livestock slaughtering to obtain the curdling enzyme (rennet) and was remarkably resistant towards spoilage bacteria and milk rancidity. On demand, the same recipe might either produce a probiotic beverage (kefir) poised for immediate consumption, or a kefir cheese with extended shelf life and high nutrition value. Fermentation reduced lactose content in kefir and also in the whey collected upon straining the curdled casein mass such that both could serve as protein-rich nutrients and sources of vitamins and minerals for a lactose-intolerant population (Wang et al., 1984) of ethnically admixed Xiaohe (Li et al., 2010). Last but not least, the probiotic beverage helped to maintain and restore human gut micorbiota (Beermann and Hartung, 2013) by adapting it to mixed and season-dependent diets common to a semi-pastoral household in the extreme environment of Taklamakan.

In a broader perspective, kefir cheese found at Xiaohe is among the earliest material evidences of using microorganisms for transforming food properties that, together with fermented alcoholic beverages (McGovern et al., 2009, McGovern et al., 2004), hallmarked the onset of modern biotechnology. This major technological advance supported the Eastward expansion of domestication and herding of ruminant animals into neighboring regions, such as Tibet, and therefore could have made major impact on the household and economy of Bronze Age East Eurasia.

5. Conclusions and perspectives

Proteomics characterization of well-preserved residues of ca. 3800-years old cheese from the tombs of Xiaohe identified them as a product of the kefir fermentation and provided direct material evidence on, to the best of our knowledge, the earliest known dairy practice that persists till present times in an almost unchanged way. The evidence of kefir dairy that occurred already at the Early Bronze Age helps to understand why milking was spreading over Eastern Eurasia despite lactose intolerance of the local population, its primitive nutrition habits and food processing skills; why cheese-making was the most economical way of processing milk and how kefir became an inherent element of the Tibetan nutritional culture and much later became known to Europeans as “Tibetan mushroom”.

Also we underscore an interesting methodical aspect of our work. Currently, proteomics is rarely used in archaeometry, likely because of a common notion that proteins are chemically unstable, very difficult to recover, handle and analyze in micro-quantities. While this generally holds true, modern analytical instrumentation and software may help to overcome these hurdles. Here we demonstrated that proteomics analysis of organic residues is direct, quantitative, and informative and therefore has the potential to develop into a valuable, generally applicable tool in archaeometry.

Acknowledgment

We are grateful for Dr. Henrik Thomas for expert technical support and Drs. James Saenz and Shirin Pocha for critical reading of the manuscript. This study was supported by the grants from the National Science Foundation in China (41172164) and the CAS Strategic Priority Research Program (XDA05130303). Andrej S and Anna S are in debt to Dr. Vitaly A. Kolotuhin† for introducing them to archaeology.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.jas.2014.02.005.

References

Aebersold, R., Mann, M., 2003. Mass spectrometry-based proteomics. Nature 422, 198–207.
Bai, M., Qing, M., Guo, Z., Zhang, Y., Chen, X., Bao, Q., Zhang, H., Sun, T.S., 2010. Occurrence and dominance of yeast species in naturally fermented milk from the Tibetan Plateau of China. Can. J. Microbiol. 56, 707–714.
Beermann, C., Hartung, J., 2013. Physiological properties of milk ingredients released by fermentation. Food Funct. 4, 185–199.
Bergman, F., Sylwan, V., Konow, S., Ljungh, H., 1939. Archaeological Research in Sinkiang, Bokfoerlags aktiebolaget Thule.
Buckley, M., Melton, N.D., Montgomery, J., 2013. Proteomics analysis of ancient food vessel stitching reveals >4000-year-old milk protein. Rapid Commun. Mass Spectrom. 27, 531–538.
Chambery, A., Di Maro, A., Sanges, C., Severino, V., Tarantino, M., Lamberti, A., Parente, A., Arcari, P., 2009. Improved procedure for protein binder analysis in mural painting by LC-ESI/Q-q-TOF mass spectrometry: detection of different milk species by casein proteotypic peptides. Anal. Bioanal. Chem. 395, 2281–2291.
Copley, M.S., Berstan, R., Dudd, S.N., Docherty, G., Mukherjee, A.J., Straker, V., Payne, S., Evershed, R.P., 2003. Direct chemical evidence for widespread dairying in prehistoric Britain. Proc. Natl. Acad. Sci. U. S. A. 100, 1524–1529.
Cultural Relics and Archaeology Institute of Xinjiang Autonomous Region (CRAIXAR), 2007. A brief excavation report on Xiaohe graveyard located in Luobupo, Xinjiang Autonomous Region. Cult. Relics 10, 4–42.
Dudd, S.N., Evershed, R.P., 1998. Direct demonstration of milk as an element of archaeological economies. Science 282, 1478–1481.
Dunne, J., Evershed, R.P., Salque, M., Cramp, L., Bruni, S., Ryan, K., Biagetti, S., di Lernia, S., 2012. First dairying in green Saharan Africa in the fifth millennium BC. Nature 486, 390–394.
Evershed, R.P., Dudd, S.N., Copley, M.S., Berstan, R., Stott, A.W., Mottram, H., Buckley, S.A., Crossman, Z., 2002. Chemistry of archaeological animal fats. Acc. Chem. Res. 35, 660–668.
Evershed, R.P., Payne, S., Sherratt, A.G., Copley, M.S., Coolidge, J., Urem-Kotsu, D., Kotsakis, K., Ozdogan, M., Ozdogan, A.E., Nieuwenhuyse, O., Akkermans, P.M., Bailey, D., Andeescu, R.R., Campbell, S., Farid, S., Hodder, I., Yalman, N., Ozbasaran, M., Bicakci, E., Garfinkel, Y., Levy, T., Burton, M.M., 2008. Earliest date for milk use in the Near East and southeastern Europe linked to cattle herding. Nature 455, 528–531.
Garrote, G.L., Abraham, A.G., De Antoni, G.L., 2001. Chemical and microbiological characterisation of kefir grains. J. Dairy Res. 68, 639–652.
Geetha, S., Lakshmi, G., Ranjithakani, P., 1996. An ethnic method of milk curdling using plants. Anc. Sci. Life 16, 60–61.
Habermann, B., Oegema, J., Sunyaev, S., Shevchenko, A., 2004. The power and the limitations of cross-species protein identification by mass spectrometry-driven sequence similarity searches. Mol. Cell. Proteomics 3, 238–249.
Hinz, K., O'Connor, P.M., Huppertz, T., Ross, R.P., Kelly, A.L., 2012. Comparison of the principal proteins in bovine, caprine, buffalo, equine and camel milk. J. Dairy Res. 79, 185–191.

Hong, C., Jiang, H., Lu, E., Wu, Y., Guo, L., Xie, Y., Wang, C., Yang, Y., 2012. Identification of milk component in ancient food residue by proteomics. PloS One 7, e37053.
Hui, Y.H., 2004. Handbook of Food and Beverage Fermentation Technology. In: Food Science and Technology. Marcel Dekker, New York pp. xii, 919.
Jolles, J., Alais, C., Jolles, P., 1968. The tryptic peptide with the renin-sensitive linkage of cow's kappa casein. Biochim. Biophys. Acta 168, 591–593.
Junqueira, M., Spirin, V., Balbuena, T.S., Thomas, H., Adzhubei, I., Sunyaev, S., Shevchenko, A., 2008. Protein identification pipeline for the homology-driven proteomics. J. Proteomics 71, 346–356.
Krausmann, F., 2004. Milk, manure, and muscle power. Livestock and the transformation of preindustrial agriculture in Central Europe. Hum. Ecol. 32, 735–772.
Kumar, A., Grover, S., Sharma, J., Batish, V.K., 2010. Chymosin and other milk coagulants: sources and biotechnological interventions. Crit. Rev. Biotechnol. 30, 243–258.
Leo, G., Bonaduce, I., Andreotti, A., Marino, G., Pucci, P., Colombini, M.P., Birolo, L., 2011. Deamidation at asparagine and glutamine as a major modification upon deterioration/aging of proteinaceous binders in mural paintings. Anal. Chem. 83, 2056–2064.
Leo, G., Cartechini, L., Pucci, P., Sgamellotti, A., Marino, G., Birolo, L., 2009. Proteomic strategies for the identification of proteinaceous binders in paintings. Anal. Bioanal. Chem. 395, 2269–2280.
Li, C., Li, H., Cui, Y., Xie, C., Cai, D., Li, W., Mair, V.H., Xu, Z., Zhang, Q., Abuduresule, I., Jin, L., Zhu, H., Zhou, H., 2010. Evidence that a West-East admixed population lived in the Tarim Basin as early as the early Bronze Age. BMC Biol. 8, 15.
Mair, V.H., 2006. The rediscovery and complete excavation of Ordek's Necropolis. J. Indo-Eur. Stud. 34, 273–318.
McGovern, P.E., Mirzoian, A., Hall, G.R., 2009. Ancient Egyptian herbal wines. Proc. Natl. Acad. Sci. U. S. A. 106, 7361–7366.
McGovern, P.E., Zhang, J., Tang, J., Zhang, Z., Hall, G.R., Moreau, R.A., Nunez, A., Butrym, E.D., Richards, M.P., Wang, C.S., Cheng, G., Zhao, Z., Wang, C., 2004. Fermented beverages of pre- and proto-historic China. Proc. Natl. Acad. Sci. U. S. A. 101, 17593–17598.
Miguel, M., Cardoso, P.G., Lago, L.D., Schwan, R.F., 2010. Diversity of bacteria present in milk kefir grains using culture-dependent and culture-independent methods. Food Res. Int. 43, 1523–1528.
Mirabaud, S., Rolando, C., Regert, M., 2007. Molecular criteria for discriminating adipose fat and milk from different species by NanoESI MS and MS/MS of their triacylglycerols: application to archaeological remains. Anal. Chem. 79, 6182–6192.
Motaghi, M., Mazaheri, M., Moazami, N., Farkhondeh, A., Fooladi, M.H., Goltapeh, E.M., 1997. Kefir production in Iran. World J. Micorbiol. Biotechnol. 13, 579–581.
Reidel, B., Thompson, J.W., Farsiu, S., Moseley, M.A., Skiba, N.P., Arshavsky, V.Y., 2011. Proteomic profiling of a layered tissue reveals unique glycolytic specializations of photoreceptor cells. Mol. Cell. Proteomics 10, 1–14.
Salque, M., Bogucki, P.I., Pyzel, J., Sobkowiak-Tabaka, I., Grygiel, R., Szmyt, M., Evershed, R.P., 2012. Earliest evidence for cheese making in the sixth millennium bc in northern Europe. Nature 493, 522–525.
Shevchenko, A., Loboda, A., Ens, W., Schraven, B., Standing, K.G., Shevchenko, A., 2001. Archived polyarylamide gels as a resource for proteome characterization by mass spectrometry. Electrophoresis 22, 1194–1203.
Shevchenko, A., Tomas, H., Havlis, J., Olsen, J.V., Mann, M., 2006. In-gel digestion for mass spectrometric characterization of proteins and proteomes. Nat. Protoc. 1, 2856–2860.
Solazzo, C., Fitzhugh, W.W., Rolando, C., Tokarski, C., 2008. Identification of protein remains in archaeological potsherds by proteomics. Anal. Chem. 80, 4590–4597.
Tamime, A.Y., 2006. Brined Cheeses. Blackwell Publishing Ltd.
Vasilj, A., Gentzel, M., Ueberham, E., Gebhardt, R., Shevchenko, A., 2012. Tissue proteomics by one-dimensional gel electrophoresis combined with label-free protein quantification. J. Proteome Res. 11, 3680–3689.
Wang, Y.G., Yan, Y.S., Xu, J.J., Du, R.F., Flatz, S.D., Kuhnau, W., Flatz, G., 1984. Prevalence of primary adult lactose malabsorption in three populations of northern China. Hum. Genet. 67, 103–106.