JOURNAL OF PROTEOMICS 105 (2014) 363–371

Available online at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/jprot

Proteomics identifies the composition and manufacturing recipe of the 2500-year old sourdough bread from Subeixi cemetery in China☆

Anna Shevchenko[a,1], *Yimin Yang*[b,c,1], *Andrea Knaust*[a], *Henrik Thomas*[a], *Hongen Jiang*[b], *Enguo Lu*[d], *Changsui Wang*[b,*], *Andrej Shevchenko*[a,*]

[a]*MPI of Molecular Cell Biology and Genetics, 01307 Dresden, Germany*
[b]*Department of Archaeometry, University of Chinese Academy of Sciences, Beijing 100049, PR China*
[c]*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*
[d]*Xinjiang Cultural Relics and Archaeology Institute, Ürümchi 830000, PR China*

ARTICLE INFO

Available online 26 November 2013

Keywords:
geLC–MS/MS
Ancient organic residues
Ancient foods
Cereals
Fermentation
Archeometry

ABSTRACT

We report on the geLC–MS/MS proteomics analysis of cereals and cereal food excavated in Subeixi cemetery (500–300 BC) in Xinjiang, China. Proteomics provided direct evidence that at the Subexi sourdough bread was made from barley and broomcorn millet by leavening with a renewable starter comprising baker's yeast and lactic acid bacteria. The baking recipe and flour composition indicated that barley and millet bread belonged to the staple food already in the first millennium BC and suggested the role of Turpan basin as a major route for cultural communication between Western and Eastern Eurasia in antiquity.
This article is part of a Special Issue entitled: Proteomics of non-model organisms.

Biological significance
We demonstrate that organic residues of thousand year old foods unearthed by archeological excavations can be analyzed by geLC–MS/MS proteomics with good representation of protein source organisms and coverage of sequences of identified proteins. In-depth look into the foods proteome identifies the food type and its individual ingredients, reveals ancient food processing technologies, projects their social and economic impact and provides evidence of intercultural communication between ancient populations. Proteomics analysis of ancient organic residues is direct, quantitative and informative and therefore has the potential to develop into a valuable, generally applicable tool in archaeometry.
This article is part of a Special Issue entitled: Proteomics of non-model organisms.

1. Introduction

Cereals are a cornerstone of human diet since prehistory. Grain processing technologies improving nutritional properties, shelf life, texture, taste and aroma of cereal foods and beverages have been conceived already in antiquity (reviewed in [1]). The earliest known fermented beverage was made of rice and fruits at about 7000 BC in China [2] and the earliest known wine some 7000 years ago in Georgia [3]. In the second millennium BC in Egypt beer and wheat bread leavened with malt or yeast

☆ This article is part of a Special Issue entitled: Proteomics of non-model organisms.
* *Corresponding authors.*
E-mail addresses: cswang@ucas.ac.cn (C. Wang), shevchenko@mpi-cbg.de (A. Shevchenko).
[1] Equal contribution of these authors.

1874-3919/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jprot.2013.11.016

belonged to a staple food [4] and making sourdough bread was also practiced in Mesopotamia [5,6].

The exact determination of cereal species found in ancient food residues is an important step in the understanding of ancient diet, plant domestication, cooking recipes and tools, and even social organization [7]. Traditionally, microfossils of crops are identified by a microscope-assisted visual examination. The morphology of intact ancient seeds as well as their starch grains and phytolith is compared against reference specimen of modern cereals [8,9]. However, the outcome of morphological analysis is becoming less certain if grains were processed – ground, mixed with diverse components in complex foods, cooked, fermented, and also aged [7]. Thermal treatment (cooking or baking) of starch grains is recognizable by microscopy investigations. However, it does not reveal if and how grains were fermented giving little insight into the processing technology and nutritional qualities of ancient foods. Technological advances are important evidences of both the social development and cultural exchange between ancient populations. For example, the transition from raw milk consumption to cheese making was an important step towards developing and expanding the semi-pastoral household and economy [10]. Reducing lactose content in cheese compared to raw milk made herding of ruminant livestock attractive to lactose-intolerant populations of Eastern Eurasia [11].

Proteomics could identify a few proteins in a variety of cultural relics from pottery shards to old paintings [12–17], however in-depth characterization of complex solid residues (like, foods) unearthed during archeological excavations, remains technically challenging. Conventional database searches rely on the identity of sequences of analyzed proteins and sequences deposited in a database, while any discrepancy between these sequences is penalized by the software. Therefore, the identification of proteins from organisms with unsequenced genomes, wild-bred species showing strong genome polymorphism or heavily modified and degraded proteins is technically challenging and result in promiscuous assignments (reviewed in [18,19]). While, to some extent, it is still possible to rely upon cross-species matching of identical peptide sequences between highly homologous proteins, such identifications might be misleading. When analyzing even simple protein mixtures this approach is heavily biased towards matching highly conserved proteins, which might only constitute a minor component, while the major component comprising less conserved proteins might remain unidentified [20].

Methods of homology-driven proteomics have been developed to overcome these technical hurdles [21,22]. They rely on automated de novo interpretation of tandem mass spectra and specialized search engines that can process queries comprising thousands of degenerate, redundant and partially accurate peptide sequence candidates [23,24]. Importantly, spectra acquired by LC–MS/MS experiment can be interpreted by both conventional and homology-driven data mining. These two approaches can also be employed in a layered manner by first identifying the known or highly homologous proteins via stringent searches, while unmatched MS/MS spectra could be further interpreted de novo and submitted to sequence – similarity searches by engines like MS BLAST [23,24]. Using label-free proteomics software, the abundances of peptide precursors identified by MS BLAST can also be used for quantitative inferences.

We hypothesized that a combination of conventional and homology driven proteomics might help to overcome the technical hurdles in the characterization of the ancient foods composition. Expectantly the scope of identified proteins will span through broader organism representation and the negative impact of protein degradation by fermentation, cooking or natural aging might be reduced. Beyond major components, it should be possible to uncover minor (yet important) proteins that might provide evidence on food fermentation, preservation or nutrition quality improvements.

To the best of our knowledge, this work is the first report on the proteomics analysis of ancient cereals and cereals foods that provides direct evidence that sourdough bread was made in Subeixi, China already at the first millennium BC. It belonged to the staple food of the local population and also had a ritual meaning in funeral ceremonies.

2. Materials and methods

2.1. Chemicals

Porcine trypsin sequencing grade was purchased from Promega (Mannheim, Germany); HPLC grade acetonitrile and water, from Merk (Darmstadt, Germany); gel electrophoresis buffers from Invitrogen (Carlsbad, CA); other chemicals (reagent grade) were from Sigma-Aldrich (Munich, Germany).

2.2. Archeological samples

Samples were collected during excavations of the cemetery no. 3 of the Subexi site, Turpan Basin, Shanshan County, Xinjiang Uighur Autonomous Region [25]. Both samples were recovered from the earthenware bowls IIIM27:8 and IIIM27:9 from the lower layer of the tomb IIIM27 (see [26] for the detailed sketch).

Modern broomcorn millet used as a reference was obtained from K.-O. Werz Naturmühle GmbH+Co.KG, Heidenheim, Germany; modern foxtail millet was from Xinjiang region.

2.3. Sample preparation for proteomics analyses

A clump of 8 mg of organic material from the IIIM27:9 bowl and six cereal grains from the M27:8 bowl were transferred into separate 1.5 mL Eppendorf tubes (Hamburg, Germany) and disintegrated into fine powder using disposable pestles (Argos Technologies, IL, USA). Then 50 μl of 65 mM Tris HCl buffer (pH 6.8) containing 2% sodium dodecylsulfate (SDS) were added and tubes sonicated for 45 min. The slurry together with insoluble debris was loaded on a pre-cast 1 mm 12% polyacrylamide gel (BioRad, Munich, Germany). To avoid carryover, each sample was loaded on a separate gel and electrophoresis run in separate chambers [27]. Once front migration distance reached ca. 4 cm, electrophoresis was stopped and the gel slab 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 [28].

Peptides recovered from *in-gel* digests were dissolved in 15 μl of 5% aqueous formic acid spiked with 10 fmol of Glu-1

Fibrinopeptide B (reference standard for alignment and normalization of chromatograms). A 5 μl aliquot of the peptide digest was analyzed by LC–MS/MS on an Ultimate3000 nanoLC system (Dionex, Amsterdam, The Netherlands) equipped with a 75 μm i.d. × 20 mm trap column and 75 μm × 15 cm analytical column (Acclaim PepMap100 C18 3 μm/100A, both from Dionex) interfaced on-line to a LTQ Orbitrap Velos hybrid tandem mass spectrometer (Thermo Fisher Scientific, Bremen, Germany). A 180 min elution program was applied as the following: linear gradient of 0% to 30% B was delivered in 145 min and then B% was increased to 100% within 10 min and maintained for another 5 min, dropped to 0% in 10 min and maintained for another 10 min. Solvent A was 0.5% aqueous formic acid and B was neat acetonitrile. The data-dependent acquisition (DDA) cycle consisted of FT MS survey spectrum followed by 6 MS/MS spectra with a fragmentation threshold of 5000 ion counts and dynamic exclusion time of 25 s, singly charged and charge-unassigned precursor ions were excluded. FT survey scans were acquired within *m/z* range of 350 to 1600. Lock mass was set to the singly charged ion of dodecamethylcyclohexasiloxane ion ($(Si(CH_3)_2O))_6$; *m/z* =445.120025).

2.4. Protein identification by MASCOT and MS BLAST searches

Spectra were converted to .mgf format using extract_msn converter (ver.5, Thermo Fisher Scientific, Bremen, Germany) and mgf files obtained for individual gel slabs were merged. Proteins were identified by MASCOT v.2.2.04 software (Matrix Sciences) by searching against a comprehensive (all species) NCBI protein sequences database (compiled in August 2013 from 31,601,460 entries) considering typical age-related protein modifications [15,29] under the following settings: 5 ppm and 0.5 Da mass accuracy for precursor and fragment ions respectively; enzyme specificity: trypsin; number of allowed miscleavage sites: two; variable modifications: methionine oxidized, cysteine propionamide, asparagine and glutamine deamidated. Protein identifications were accepted with two or more peptides (each comprising more than seven amino acid residues) matching above homology threshold according to red bold rule under $p < 0.05$ and having the peptide ion scores above the value of 30. Single peptide identifications were accepted for proteins below 10 kDa with the minimum peptide ions score of 50. The protein hits were then used in MASCOT seconds pass searches, in which no trypsin specificity restrictions were applied to match semi- and non-tryptic peptides.

MS/MS spectra that were not matched by MASCOT were extracted from the .mgf query by an in-house developed script and interpreted de novo by PepNovo software as described [21,22]. Up to 7 candidate peptide sequences for each interpreted tandem spectra were considered and only candidates with the sequence quality score of 6 or above were used for subsequent MS BLAST searches. PepNovo outputs were pasted directly into MS BLAST query window and searched against the nr database using the LC–MS/MS Presets option of the MS BLAST web server (http://genetics.bwh.harvard.edu/msblast/); only high scoring segment pairs (HSPs) with the scores exceeding 55 were considered.

2.5. Protein composition of the ancient food specimen

Relative abundance of protein groups of plant proteins, yeast, lactic acid bacteria and caseins was determined by geLC–MS/MS label-free proteomics [30] using Progenesis v.2.6 software (NonLinear Dynamics, Newcastle). For each sample the abundances of peptides detected in LC–MS/MS of digests of several gel slices were summed up [31]. The abundance of peptides originating from proteins of each group was combined and the abundance of each group was normalized to the total abundance of all groups using in-house scripts. Abundances of peptide precursors matched by MASCOT and sequence similarity searched were included in calculation. The abundance of chromatographic peaks for precursors matched by MS BLAST were extracted from the list of Progenesis (NonLinear Dynamics, Newcastle) unmatched features using mass tolerance of 5 ppm and retention time tolerance of 0.5 min.

2.6. Synchrotron X-ray microtomography (SR-μCT)

A piece of sample was cut and scanned by SR-μCT at Shanghai Synchrotron Radiation Facility, Shanghai city, China. The scanned object was put on an open sample platform. Parallel SR X-rays with height of 4 mm and width of 2 cm was directed at the object with source energy settings of 14 keV. The CCD detector has space resolution of 13 μm. In each scan, 339 slices were imaged. Scan time was about 10 min.

2.7. Starch grains microscopic analysis

A few mg of the IIIM27:9 organic material were disintegrated, overlaid with 4 ml deionized water and dispersed. A drop of the slurry was then fixed with glycerin at a microscope slide and the starch grains were imaged on Olympus BX51 microscope (Olympus, Japan) in both transmitted and polarized light at a magnification of 100×.

3. Results and discussion

3.1. Ancient foods from Turpan basin: discovery and site description

Ancient foods were discovered at the Subeixi site located in the Turpan Basin, Xinjiang Uighur Autonomous Region of China [25]. The site comprises the remains of three houses and also three cemeteries which have been dated to around 500–300 BC and attributed to the Subeixi culture [32]. The arid climate of Turpan basin contributed to good preservation of mummified bodies: out of 19 excavated mummies 16 belong to Caucasians. According to archeological findings (houses, pottery, grinding stones, crops, noodles and cakes, bows, bridles, leather, wool clothing, sheep remains, etc.), the inhabitants of Subeixi had adopted a semi-agricultural and semi-pastoral (stock raising) household [26].

The tomb IIIM27 – one of altogether 29 tombs of the 3rd cemetery, which is located at the Western part of the site is an earth-pit burial consisting of three layers. The skeletons of an adult male, a female, and an infant were found in the upper layer; the skeleton of another adult female was buried in the

middle layer; an elderly man was interred on the lowest layer together with two pottery bowls IIIM27:8 and IIIM27:9 at the left side of his waist [33]. In IIIM27:8 broomcorn millet (*Panicum miliaceum*) grains were identified by paleobotanical analysis [26] (Fig. 1). IIIM27:9 contained dark clumpy mass that was tentatively attributed to foods (Fig. 1; Fig. 2). Both earthenware bowls showed no signs of fire exposure and were attributed to serving pottery. Among other artifacts wooden utensils, ironware, bow and arrows, as well as roasted sheep's head were found in the same tomb [26,33].

We further collected 6 grains from the bowl IIIM27:8 and 8 mg of the presumed food material from the bowl IIIM27:9 and subjected them to proteomics analyses.

3.2. *Protein composition of the food from IIIM27:9 earthenware bowl*

Previous investigation of the food residues from the earthenware bowl IIIM27:9 (Figs. 1, 2) revealed trace amounts of milk [33], whereas the identity of the main compounds remained unknown. The organic material was collected, disintegrated in a 1.5 Eppendorf tube, processed and analyzed by LC–MS/MS as described in Section 2.

MASCOT search against a comprehensive protein database (NCBI) identified 106 proteins that, according to their organismal origin, were divided into four groups: i) plant proteins; ii) yeasts of Saccharomycetaceae family; iii) lactic acid bacteria (LAB) from Leuconostocaceae and Lactobacilli families and iv) human background proteins (keratins and other proteins typical to human skin and sweat). In concordance with the previous analyses, a few peptides from bovine S1 alpha-casein were also detected. The full list of identified proteins is provided in Supplementary Table 1S.

We then used label-free proteomics software to determine the abundances of peptide precursors matched to proteins of the groups i) to iii). Since proteins smeared along the gel lane, the same proteins (and corresponding peptides) were found in different gel slices. In these instances, the abundances of the same peptides were summed. We then calculated the relative abundance of each protein group by normalizing it to the total abundance of all matched peptides (Fig. 3). Note that here we did not determine the abundance of individual proteins, but rather estimated the averaged abundances of large protein groups to find out which of them dominated in sample.

In this way, we established that the group of 32 plant proteins was most abundant. There were no RuBisCo proteins, which are

Fig. 1 – Images of IIIM27:8 and IIIM27:9 samples excavated in the III27 tomb of the 3rd Subeixi cemetery. Upper panel: earthenware bowls with organic materials. Lower panel: detailed view of the bowls content.

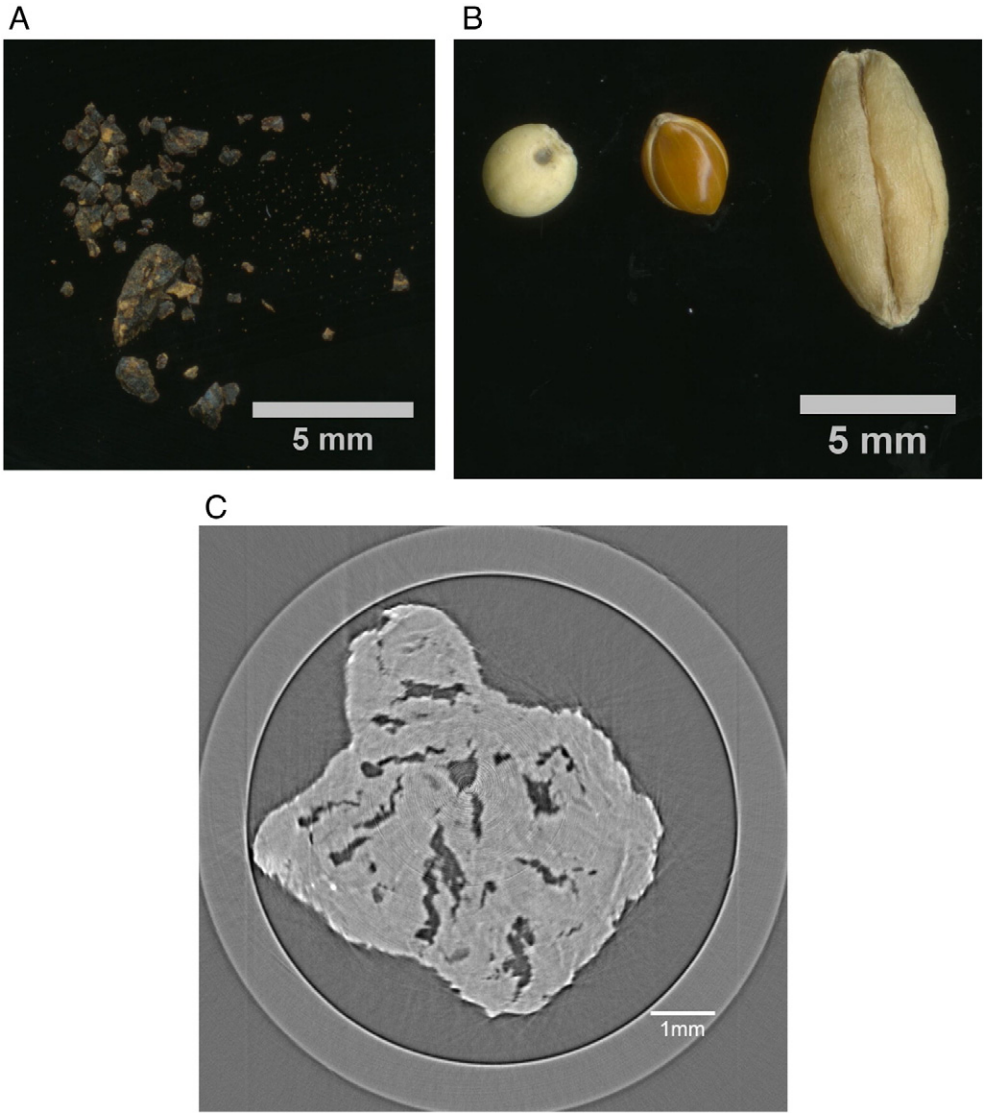

Fig. 2 – Close-up view of the food sample III M27:9. Panel A: ca 20 mg sample taken from III M27:9 bowl. Panel B: contemporary broomcorn millet naked and with husk and barley grains, respectively. Panel C: micro CT-image showing internal porosity of the III M27:9 material.

common in leaves or stems of green plants, however there were many seed enzymes and storage proteins suggesting the crop grains origin of the analyzed food. Considering homogeneous surface texture of the IIIM27:9 residue the seeds were ground. Identified plant proteins belong to both groups of true grasses Poaceae family: 23 proteins – to the BEP-clade (the clade includes crops like barley, wheat, oat and rye), five proteins – to the Panicoideae subfamily of the PACC-clade (to which millet crops belong); the remaining were matching plants from both groups (Fig. 3).

Further 19 proteins were unique for species of the *Hordeum* genus. Although it was not possible to distinguish if grains originated from wild or domesticated forms, *Hordeum vulgare subsp.vulgare* and *Hordeum vulgare*, the presence of barley proteins in the food was certain. Proteins from *Panicoideae* genus might be attributed to two millet species: *Setaria italica* (*foxtail millet*) and *Panicum miliaceum* (broomcorn millet). Since the genome of broomcorn millet is not yet available, the exact identification of millet species remained unclear.

Contrary to typical *Aspergillus* mold, it is unlikely that yeast and lactic acid bacteria – another major group of proteins identified in III27:9 (Supplementary Table 1S) were associated with food rotting. None of them was detected in the III27:8 that was located next to the bowl III27:9 in the same tomb. Therefore we reasoned that both yeast and LAB together with ground cereals and traces of milk were bona fide components of the ancient food specimen.

3.3. *Food from the Subeixi tomb is compositionally similar to sourdough bread*

We next asked if the food type and food processing recipe could be inferred from its protein composition. Both yeast and lactic acid bacteria belong to a traditionally used symbiotic starter

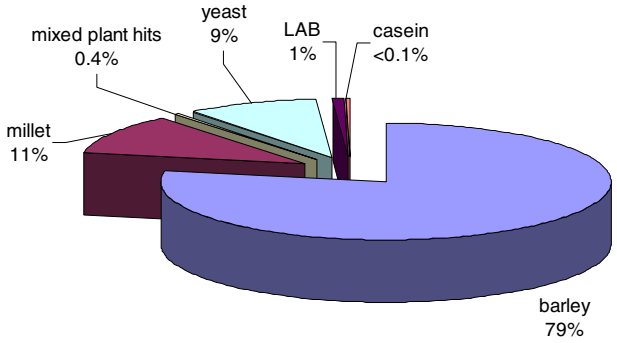

Fig. 3 – Overview of the protein composition of the food from III M27:9. The relative abundances of protein groups were calculated by summing up the abundances of all peptides matched to protein within each groups to the total abundance of all matched peptides. Background proteins of human origin and mold proteins were excluded; the abundances of peptides matched by MS BLAST were summed up.

for fermenting milk and cereals. The evidence of fermentation in the material from IIIM27:9 was corroborated by Micro-CT imaging: it revealed its pronounced porous inner structure that is typical for the leavened cereal foods made using yeasts, suggesting that the ground cereals were further fermented (Fig. 2). Dough from wheat flour, which is rich in starch and gluten can be leavened with yeast alone in a relatively short time. In comparison, barley or rye flours need longer time for fermentation and require yeast and LAB starter [34]. These symbiotic bacteria introduced in cereal flour and water matrix is called sourdough [35] and is compositionally similar to the major components of the Subeixi food: milled cereals, yeast and LAB.

Relatively low content of ground millet (Fig. 3) indicates that the ancient inhabitants of Subexi possessed good knowledge of culinary properties of both cereals. Millet belongs to gluten-free grains and therefore its flour alone can hardly be fermented for leavened bread, while mixing with barley flour accelerates its fermentation significantly. Barley can be fermented alone [36] and we can only speculate why it was mixed with millet in the Subeixi food. Perhaps, at the Subeixi barley was a more valuable crop that was used together with more abundant millet to improve the taste and nutritional value of the baked food.

Bacterial component in sourdoughs vary to great extent depending on flour, fermentation conditions or additives (salt, sugar or milk) and there is no strict correlation of its composition with the geographical location [35]. Microbiota identified in the Subeixi food – baker's yeast *S. cerevisiae* and LAB from *Lactobacilli* genus and *Leuconostocacea* genus (*Weissella cibaria*, *Leuconostoc pseudomesenteroides*) – are also common to contemporary sourdoughs.

Was sourdough fermentation spontaneous or the starter culture was maintained and re-used? Spontaneous fermentation of non-roasted flour can be initiated just by pouring in some water because of endogenous microbiota resting on flour granules. However, it is much slower and the product quality is poorly controlled. Using a stable starter culture alleviates these problems. Fermented flour/water or milk mix can be used as a living bacterial starter, which should be regularly renewed by adding water or milk and another portion of flour. We detected three peptides from the most abundant milk protein alpha casein S1 (Supplementary Table 1S), indicating that milk was present in the food, albeit in a relatively low amount (<0.01% of the abundance of all other proteins, Fig. 3), suggesting that a bacterial starter prepared on milk was used for initiating fermentation. Furthermore the artifact IIIM27:9 was a bowl with a ritual afterlife food, which also speaks against spontaneous fermentation because of its unfavorable timing.

3.4. *Subeixi cereal foods contain broomcorn millet*

In the sample IIIM27:9 we detected proteins from broomcorn *P. miliaceum* and foxtail *S. italica* millets. Both crops belong to Panicoideae family and are among most economically important cereals in Eastern Asia. *P. miliaceum* was a staple food in Nord China already in 8000 BC and was later substituted by *S. italica*. However, due to stronger drought resistance, broomcorn millet remained prominent in inland arid areas [37,38]. Distinguishing these two taxons in ancient foods could help to determine the borders of plant cultivation and trading.

Altogether, the current release of NCBI database comprises more than 292,000 protein sequences from plants of Panicoideae family including ca 36,000 sequences of from *S. italica*. The genome of broomcorn millet *P. miliaceum* has not yet been sequenced and currently only *ca* 100 of redundant protein sequences are available. To identify the exact millet specie in IIIM27:9 food we extracted proteins from contemporary broomcorn and foxtail millets and from six ancient millet grains from the bowl IIIM27:8 and compared their protein profiles. Grains in the bowl IIIM27:8 were previously attributed to broomcorn millet (*P. miliaceum*) by the morphology of their starch grains and by phytolith analysis (Fig. 1) [26].

From 600 proteins identified in the foxtail millet none could be cross-matched to the broomcorn millet. Contrary, modern broomcorn millet proteins were matched to multiple *Panicoideae* species including 200 hits from *S. italica* and only three proteins originated from *P. miliaceum* itself: granule-bound starch synthase 1 (gi 281333915), aspartate aminotransferase (gi 2059) and alanine aminotransferase (gi 461498).

Ancient broomcorn millet grains (sample IIIM27:8) contained, among other cereal proteins, 39 proteins common to Panicoideae plants (Supplementary Materials Fig. 1S, Table 1S), including 11 proteins unique for *S. italica* and, similar to modern grains, same three proteins originated from *P. miliaceum*. The granule-bound starch synthase 1 with 22 matching peptides was the most abundant among the three and has a very close homologue (gi 526117365) in *S. italica* sharing 98% sequence identity which, however, was not matched by the database search. Close inspection revealed that there are three unique peptides matching the *Panicum* sequence (Fig. 4) while other 19 are identical for both *P. miliaceum* and *S. italica* homologues. We reasoned that these unique peptides solely matching the *P. miliaceum* granule-bound starch synthase 1 protein can unequivocally distinguish the millet type also in complex foods.

Next we compared these results with findings in the IIIM27:9 cereals food. 35 peptides were matched to the homologues of granule-bound starch synthase 1 in barley, foxtail and broomcorn millet. Out of these 35, three were unique for the barley *H. vulgare* sequence and other two for *P. miliaceum*. Similar to the IIIM27:8

```
PM      2 AGMNVVFVGAEMAPWSKTGGLGDVLGGLPPAMAANGHRVMVVSPRYDQYKDAWDTSVVSE  61
SI     77 AGMNVVFVGAEMAPWSKTGGLGDVLGGLPPAMAANGHRVMVISPRYDQYKDAWDTSVVSE 136
          **************************************** ******************

PM     62 IKMGDRYETVRFFHCYKRGVDRVFIDHPSFLERVWGKTGEKIYGPDAGVDYKDNQLRFSL 121
SI    137 IKVGDRYERVRFFHCYKRGVDRVFIDHPSFLERVWGKTGEKIYGPDAGVDYKDNQLRFSL 196
          ** ***** ************************************************** 

PM    122 LCQAALEAPRILSLNNNPYFSGPYGEDVVFVCNDWHTGPLSSYLKSNYQSNGIYKNAKTA 181
SI    167 LCQAALEAPRILSLNNNPYFSGPYGEDVVFVCNDWHTGPLSSYLKSNYQSNGIYRNAKTA 256
          ********************************************************* *****

PM    182 FCIHNISYQGRFAFSDYPELNLPERFRSSFDFIDGYEKPVEGRKINWMKGGILEADKVLT 241
SI    257 FCIHNISYQGRFAFSDYPELNLPERFRSSFDFIDGYEKPVEGRKINWMKAGIIEADRVLT 316
          ********************************************** ** *** *** 

PM    242 VSPYYAEELISGIARGCELDNIMRLTGITGIVNGMDVSEWDPSKDKYIATKYDVSTAIAA 301
SI    317 VSPYYAEELISGIARGCELDNIMRLTGITGIVNGMDVSEWDPSKDKYIATKYDVSTAIAA 376
          ************************************************************

PM    302 KALNKEALQAAAGLPVDRKIPLVAFVGRLEEQKGPDVMAAAIPQLMEEDVQIVLLGTGKK 361
SI    377 KALNKEALQAAAGLPVDRKIPLVAFVGRLEEQKGPDVMAAAIPQLMEEDVQIVLLGTGKK 436
          ************************************************************

PM    362 KFERMLMSAEEKYPDKVRAVVKFNAALAHHIMAGADLLAVTSRFEPCGLIQLQGMRYGTP 421
SI    437 KFERMLMSAEEKYPDKVRAVVKFNAAVAHHIMAGADLLAVTSRFEPCGLIQLQGMRYGTP 496
          ************************* ********************************

PM    422 CVCASTGGLVDTVIEGKTGFHMGRLSVDCKVVEPADVQKVATTLKRAIKVVGTPAYEEMV 481
SI    497 CVCASTGGLVDTVIEGKTGFHMGRLSVDCKVVEPADVQKVASTLKRAIKVVGTPAYEEMV 556
          **************************************** ******************

PM    482 RNCMIQDLSWKGPAKNWENVLLSLGVAGSQPGIEGEEIAPLA 523
SI    557 RNCMIQDLSWKGPAKNWENVLLSLGVAGSQPGIEGEEIAPLA 598
          ******************************************
```

Fig. 4 – The flour in IIIM27:8 food was made from broomcorn (*Panicum miliaceum*), but not foxtail (*Setaria italic*) millet grains. The figure presents the sequence alignment of granule-bound starch synthases from *P. miliaceum* (gi 281333915) and *S. italica* (gi 526117365). Three sequenced peptides (in red) were unique for the *P. miliaceum* sequences, while no peptides unique for *S. italica* were identified. Amino acid residues different in the sequenced peptides are highlighted.

and modern broomcorn millet, no peptides unique for foxtail millet were found suggesting that broomcorn millet was another cereal in IIIM27:9 (Supplementary Table 1S; Supplementary Figure 1S).

Surprisingly, no alpha-prolamins were found by MASCOT searches in broomcorn millet samples, although they are major storage proteins in seeds accounting for over 50% of the total protein content [39]. Their sequences, however, contain very few tryptic cleavage sites. To identify prolamis and estimate their content, all MS/MS spectra confidently matched by MASCOT were removed from .mgf, the remaining (unmatched) spectra were interpreted de novo by PepNovo software and obtained candidate sequence proposals submitted to MS BLAST search [22]. Expectantly, many predicted peptides were matched to the sequences of storage proteins from Panicoideae plants such as zeins (*S. italica*, *Z. mays*, *P. sumatrense*); kafirin (*S. bicolor*); caneins (*S. officinarum*) or coixin (*C. lacryma-jobi*), although none of them was found by MASCOT (Fig. 5). The total abundance of these non-trypitc peptides exceeded the abundance of peptides matched by MASCOT to all plant proteins by twofold.

Was the mixed flour sourdough from the earthenware IIIM27:9 raw or processed (baked)? To address this question we performed microscopic analysis of its starch grains. Their morphology is usually preserved if the material was stored under dry conditions, however food processing changes it remarkably. In the sample IIIM27:9 individual grains show clearly visible boundaries and looked damaged (gelatinized), having unclear or absent extinction cross – these notable morphological changes indicating thermal treatment [4,7] (Fig. 6). We therefore concluded that the sourdough from Subexi was baked. Since IIIM27:9 does not bear clear traces of fire the sourdough was likely first baked, then cut in pieces and placed into the serving bowl.

4. Conclusions and perspectives

We applied geLC–MS/MS proteomics to elucidate the composition of the ancient cereals food discovered during archeological excavation at Subexi. The conventional database searching method (MASCOT) was extended by homology driven identifications by MS BLAST.

Taken together, proteomics evidences indicated that the food specimen from the Subeixi tomb was a sourdough bread made from milled barley with addition of broomcorn millet by leavening with a renewable microbiota starter comprising baker's yeast and lactic acid bacteria. Whereas *P. miliaceum* is indigenous to northern China since the 8th millennium BC, barley originated from Near East and is believed to be less common to the food in Xinjiang region 2500 years ago. The

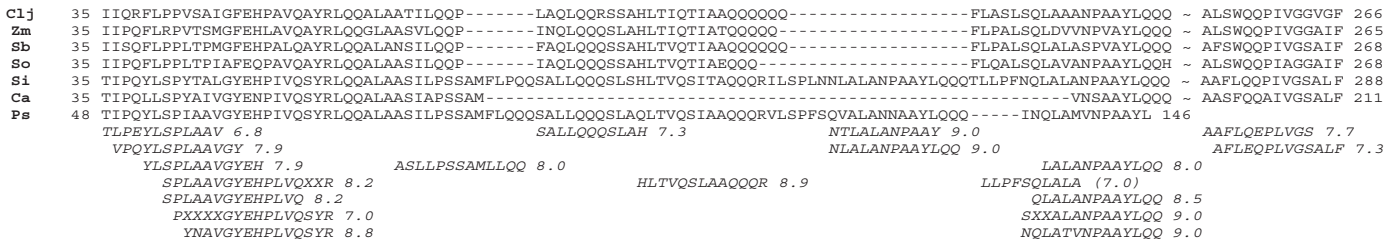

Fig. 5 – Identification of major seeds storage protein in M27:9 by sequence similarity searches. MS BLAST matched peptide sequences to seed storage proteins from seven *Panicodeae* plants: Clj — *Coix lacrima-jobi*, gi296508; Zm — *Zea mays*, gi1182065; Sb — *Sorhum bicolor*, gi168805246; So — *Saccharum officinarum*, gi145337077; Si — *Setaria italica*, gi514809774; Ca — *Cenchrus americanus*, gi148748936; Ps — *Panicum sumatrense*, gi58760523. The predicted sequences are shown below the corresponding alignments; numbers next to peptide sequences are PepNovo sequence quality scores showing the expected number of amino acid residues correctly called by automated de novo sequencing.

advanced food processing revealed by proteomics indicates that barley was not only cultivated in Turpan at 500–300 BC [26], but already was a common staple food component. The inhabitants of Subeixi were able to ferment cereals grains and sourdough bread-like products belonged to their quotidian diet. Using barley together with millet could also improve nutritional quality of the cereals food and not only provide the ancient inhabitants of Subeixi with rich source of fiber and proteins, but also B-vitamins, essential minerals and antioxidants [34].

Barley was first domesticated in the Near East and our findings suggest that Turpan basin was one of the main routes for its spreading toward Eastern China and, in antiquity, played an important role in the cultural communication between the Western and Eastern Eurasia.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jprot.2013.11.016.

Acknowledgments

This work was supported by Max Planck Society funding and grants from the National Science Foundation in China (41172164) and the CAS Strategic Priority Research Program (XDA05130303).

The authors are grateful to Dr. Marc Gentzel (MPI CBG) for the expert experimental support and useful discussions.

REFERENCES

[1] Blandino A, Al-Aseeri ME, Pandiella SS, Cantero D, Webb C. Cereal-based fermented foods and beverages. Food Res Int 2003;36:527–43.
[2] McGovern PE, Zhang J, Tang J, Zhang Z, Hall GR, Moreau RA, et al. Fermented beverages of pre- and proto-historic China. Proc Natl Acad Sci U S A 2004;101:17593–8.
[3] McGovern PE. Ancient Wine: The Search for the Origins of Viniculture. Princeton University Press; 2007.
[4] Samuel D. Investigation of ancient Egyptian baking and brewing methods by correlative microscopy. Science 1996;273:488–90.
[5] Samuel D. Bread in archeology. Bread ovens and hearths of the past. Bruxelles: Institut de sociologie de l'Université Libre de Bruxelles; 2002.
[6] Katz SH, Weaver WW. Encyclopedia of food and culture. New York: Charles Scribners and Sons; 2002.
[7] Henry AG, Hudson HF, Piperno DR. Changes in starch grain morphologies from cooking. J Archaeol Sci 2009;36:915–22.
[8] Shillito LM. Grains of truth or transparent blindfolds? A review of current debates in archaeological phytolith analysis. Veg Hist Archaeobot 2013;22:71–82.
[9] Cortella AR, Pochettino ML. Starch grain analysis as a microscopic diagnostic feature in the identification of plant material. Econ Bot 1994;48:171–81.
[10] Evershed RP, Payne S, Sherratt AG, Copley MS, Coolidge J, Urem-Kotsu D, et al. Earliest date for milk use in the Near East

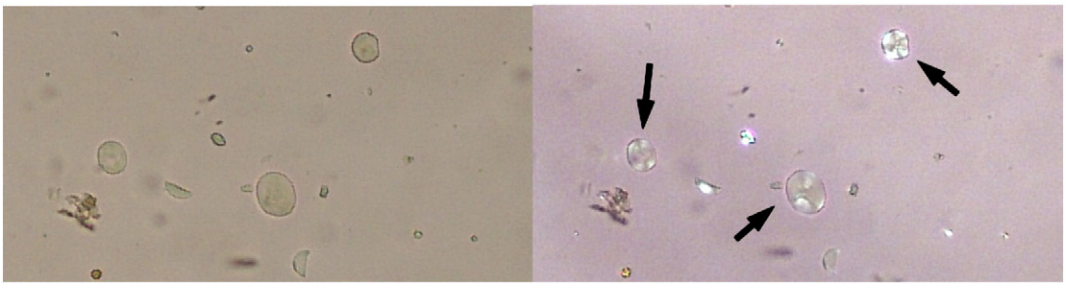

Fig. 6 – Starch grains analysis in III M27:9 food. Microscopic view with ×100 magnification of starch grains under transmitted (left) and polarized light (right). Starch grains are arrow-headed. They appear gelatinized and have diffused extinction cross, indicating thermal treatment (baking) of the material.

and southeastern Europe linked to cattle herding. Nature 2008;455:528–31.
[11] Salque M, Bogucki PI, Pyzel J, Sobkowiak-Tabaka I, Grygiel R, Szmyt M, et al. Earliest evidence for cheese making in the sixth millennium BC in northern Europe. Nature 2013;493:522–5.
[12] Shanks OC, Kornfeld M, Ream W. DNA and protein recovery from washed experimental stone tools. Archaeometry 2004;46:663–72.
[13] Solazzo C, Fitzhugh WW, Rolando C, Tokarski C. Identification of protein remains in archaeological potsherds by proteomics. Anal Chem 2008;80:4590–7.
[14] Buckley M, Melton ND, Montgomery J. Proteomics analysis of ancient food vessel stitching reveals >4000-year-old milk protein. Rapid Commun Mass Spectrom 2013;27:531–8.
[15] Leo G, Bonaduce I, Andreotti A, Marino G, Pucci P, Colombini MP, et al. Deamidation at asparagine and glutamine as a major modification upon deterioration/aging of proteinaceous binders in mural paintings. Anal Chem 2011;83:2056–64.
[16] Leo G, Cartechini L, Pucci P, Sgamellotti A, Marino G, Birolo L. Proteomic strategies for the identification of proteinaceous binders in paintings. Anal Bioanal Chem 2009;395:2269–80.
[17] Fremout W, Dhaenens M, Saverwyns S, Sanyova J, Vandenabeele P, Deforce D, et al. Tryptic peptide analysis of protein binders in works of art by liquid chromatography–tandem mass spectrometry. Anal Chim Acta 2010;658:156–62.
[18] Liska AJ, Shevchenko A. Expanding organismal scope of proteomics: cross-species protein identification by mass spectrometry and its implications. Proteomics 2003;3:19–28.
[19] Shevchenko A, Valcu CM, Junqueira M. Tools for exploring the proteomosphere. J Proteomics 2009;72:137–44.
[20] Katz A, Waridel P, Shevchenko A, Pick U. Salt-induced changes in the plasma membrane proteome of the halotolerant alga *Dunaliella salina* as revealed by Blue-Native gel electrophoresis and nanoLC–MS/MS analysis. Mol Cell Proteomics 2007;6:1459–72.
[21] Waridel P, Frank A, Thomas H, Surendranath V, Sunyaev S, Pevzner P, et al. Sequence similarity-driven proteomics in organisms with unknown genomes by LC–MS/MS and automated de novo sequencing. Proteomics 2007;7:2318–29.
[22] Junqueira M, Spirin V, Balbuena TS, Thomas H, Adzhubei I, Sunyaev S, et al. Protein identification pipeline for the homology-driven proteomics. J Proteomics 2008;71:346–56.
[23] Habermann B, Oegema J, Sunyaev S, Shevchenko A. The power and the limitations of cross-species protein identification by mass spectrometry-driven sequence similarity searches. Mol Cell Proteomics 2004;3:238–49.
[24] Shevchenko A, Sunyaev S, Loboda A, Shevchenko A, Bork P, Ens W, et al. Charting the proteomes of organisms with unsequenced genomes by MALDI-quadrupole time-of-flight mass spectrometry and BLAST homology searching. Anal Chem 2001;73:1917–26.
[25] Xinjiang Institute of Archaeology. MoT, Subeixi site and cemeteries in Shanshan County, Xinjiang. Kaogu 2002;6:42–57.
[26] Gong Y, Yang Y, Ferguson DK, Tao D, Li W, Wang C, et al. Investigation of ancient noodles, cakes, and millet at the Subeixi Site, Xinjiang, China. J Archaeol Sci 2011;38:470–9.
[27] Knaust A, Shevchenko A, Shevchenko A. Horizontal carryover of proteins on one-dimensional polyacrylamide gels may jeopardize gel-enhanced liquid chromatography mass spectrometry proteomic interpretations. Anal Biochem 2012;421:779–81.
[28] Shevchenko A, Tomas H, Havlis J, Olsen JV, Mann M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. Nat Protoc 2006;1:2856–60.
[29] Shevchenko A, Loboda A, Ens W, Schraven B, Standing KG, Shevchenko A. Archived polyarylamide gels as a resource for proteome characterization by mass spectrometry. Electrophoresis 2001;22:1194–203.
[30] Vasilj A, Gentzel M, Ueberham E, Gebhardt R, Shevchenko A. Tissue proteomics by one-dimensional gel electrophoresis combined with label-free protein quantification. J Proteome Res 2012;11:3680–9.
[31] Reidel B, Thompson JW, Farsiu S, Moseley MA, Skiba NP, Arshavsky VY. Proteomic profiling of a layered tissue reveals unique glycolytic specializations of photoreceptor cells. Mol Cell Proteomics 2011;10:1–14.
[32] Lu EG. Main achievements of the excavation of Subeixi site. In: Yin Q, editor. New discussions on Turpan. Urumchi: Xinjiang People's Publishing House; 2006. p. 239–52.
[33] Hong C, Jiang H, Lu E, Wu Y, Guo L, Xie Y, et al. Identification of milk component in ancient food residue by proteomics. PLoS One 2012;7:e37053.
[34] Zannini E, Garofalo C, Aquilanti L, Santarelli S, Silvestri G, Clementi F. Microbiological and technological characterization of sourdoughs destined for bread-making with barley flour. Food Microbiol 2009;26:744–53.
[35] De Vuyst L, Vrancken G, Ravyts F, Rimaux T, Weckx S. Biodiversity, ecological determinants, and metabolic exploitation of sourdough microbiota. Food Microbiol 2009;26:666–75.
[36] Marklinder I, Johansson L. Sour dough fermentation of barley flours with varied content of mixed-linked (1 → 3), (1 → 4) β-d-glucans. Food Microbiol 1995;12:363–71.
[37] Lu H, Zhang J, Liu KB, Wu N, Li Y, Zhou K, et al. Earliest domestication of common millet (*Panicum miliaceum*) in East Asia extended to 10,000 years ago. Proc Natl Acad Sci U S A 2009;106:7367–72.
[38] Lee GA, Crawford GW, Liu L, Chen X. Plants and people from the Early Neolithic to Shang periods in North China. Proc Natl Acad Sci U S A 2007;104:1087–92.
[39] Shukla R, Cheryan M. Zein: the industrial protein from corn. Ind Crops Prod 2001;13:171–92.