

Supporting Information for

Diversification, niche adaptation and evolution of a candidate phylum thriving in the deep Critical Zone

Wenlu Feng^{a,b,c,1}, Xiaonan Wan^{a,b,c,1}, Yiran Zhang^{a,b,c,1}, John Quensen^d, Tom A. Williams^e, Michael Thompson^f, Matthew Streeter^g, Yang Zhang^{a,b,c}, Shuo Jiao^{a,b}, Gehong Wei^{a,h}, Yuanjun Zhu^{a,i,2}, Jie Gu^{a,b,c,2}, James M. Tiedje^{a,b,d,2}, and Xun Qian^{a,b,c,2}

^aInterdisciplinary Research Center for Soil Microbial Ecology and Land Sustainable Productivity in Dry Areas, Northwest A&F University, Shaanxi 712100, China

^bDepartment of Environmental Science, College of Natural Resources and Environment, Northwest A&F University, Shaanxi 712100, China

^cKey Laboratory of Plant Nutrition and the Agri-environment in Northwest China, Ministry of Agriculture, Northwest A&F University, Shaanxi 712100, China

^dCenter for Microbial Ecology, Michigan State University, East Lansing, MI 48824

^eSchool of Biological Sciences, University of Bristol, Bristol BS8 1TH, United Kingdom

^fAgronomy Department, Iowa State University, Ames, IA 50011

^gIowa Geological Survey, The University of Iowa, Iowa City, IA 52242

^hDepartment of Microbiology, College of Life Sciences, Northwest A&F University, Shaanxi 712100, China

ⁱState Key Laboratory of Soil Erosion and Dryland Farming on the Loess Plateau, Institute of Soil and Water Conservation, Northwest A&F University, Shaanxi 712100, China

¹W.F., X.W., and Yiran Zhang contributed equally to this work.

²**Corresponding author:** Xun Qian, James M. Tiedje, Yuanjun Zhu, Jie Gu

Email: qianxun@nwfau.edu.cn, tiedjej@msu.edu, zhuyj@nwsuaf.edu.cn, gujie205@sina.com

This PDF file includes:

Note S1 and S2

Figures S1 to S11

Legends for Datasets S1 to S13

Legends for videos S1 and S2

SI References

Other supporting materials for this manuscript include the following:

Datasets S1 to S13

Supplementary Notes

Note S1: Soil physicochemical property analyses

Soil moisture was determined by oven-drying fresh samples at 105 °C until constant weight was achieved. Particle size analysis was performed using the pipette method (1). Soil pH was measured in a 1:1 soil:water mixture (2). Bulk density was measured by taking core samples of a specific volume, oven-drying the samples, and then comparing the oven-dry mass to the core volume (2). Soil organic carbon concentration (OC) was measured by dry combustion (Costech elemental analyzer) after fumigation of the samples with HCl to remove inorganic C. Total nitrogen (TN) was also measured by dry combustion using chromatography to separate and thermal conductivity to quantify oxidized N(ECS 4010 CHNSO Analyzer, Costech Analytical Technologies, Valencia, USA). Water-extractable organic carbon (WEOC) was determined in a 10:1 (volume:mass) soil extract with a high-temperature organic carbon analyzer after acidification of the extract with phosphoric acid to convert inorganic carbon to CO₂ (g) and sparging. The settings for X-ray CT scan were 75 kV and 450 μamps (microamps) with 2880 projections. The scan resolution was 35 μm (micron). The framerate was 12.5 fps and 4 frames were averaged per projection. The images were reconstructed with efX software (North Star Imaging, Rogers, MN).

Note S2: Description of the geological formations and paleosols of Iowa sites

The two sampling sites in Iowa, United States, were located at the Hitchcock Nature Center and the Loess Hills State Forest (hereafter abbreviated HNC and LHSF, respectively). Both sites are in the Loess Hills region of Iowa, where Pleistocene-age loess blankets the landscape. Some of the loess deposits are clearly derived from glacial meltwaters during periods in which continental glaciers advanced and retreated multiple times. The present-day Missouri River was the source of much of these glaciogenic wind-blown deposits, although the sources probably also included loess and microorganisms derived from nonglacial sources to the west of Iowa (3).

Our interpretations of the dates of loess deposition and soil formation at our sampling sites are based on those of Muhs et al. and Brown and Forman (3,4), who used optically stimulated thermoluminescence of quartz grains to date the sediments at an iconic nearby section, the Loveland Loess Paratype section (N41.50052°, W95.88934°). Our HNC site was about 10 km south and our LHSF site was about 40 km north of the Loveland section. We correlated the lithostratigraphy and pedostratigraphy at our sites with the chronostratigraphy established at the Loveland section.

At all LHSF sites and at two of the HNC sites, the entire depths of the cores we collected were composed of Peoria Loess. Muhs et al. (3) reported that Peoria Loess deposition in this part of Iowa began about 27,000 years before present and continued until about 17,000 years before present. They reported a date of ~17,100 years before present for sediment about 1 m below the present-day surface at the Loveland section. We cannot unequivocally correlate the top of our cores with the top of the Loveland section because erosion has likely continued since loess deposition ended. For example, the modern soil, at the top of our HNC-2 core (Fig.1), is a Typic Udorthent, i.e., a minimally developed soil with a thin surface horizon where organic matter has accumulated to a depth of 22 cm, but no B horizon has formed. The site is strongly sloping (~40%), and erosion has limited soil development to ~22 cm. Stable landscape positions are usually required for thicker A horizons or B horizons to develop.

Muhs et al. (3) used particle size and geochemical data to differentiate three units of Peoria Loess (Lower, Middle, and Upper) at the Loveland site. In the Peoria Loess at HNC-3, we identified two weakly developed B horizons, separated by about 3 m, probably representing two periods in which the landscape stabilized enough for clay to be translocated and soil structure to form. We hypothesize that these paleosols correlate with a stable period before deposition of the Middle Peoria Loess ~23,800 years ago and with another, short, stable period before deposition of the Upper Peoria Loess that started about 21,200 years ago, according to

Muhs et al. (3).

Even without the usual markers of pedogenesis, the sediments of the Peoria Loess were affected by hydrological conditions. We observed common redoximorphic features in the Peoria Loess, down to about 10 m below the modern surface. Such features are usually interpreted to indicate a perched water table at some time in the past. For redoximorphic depletions and accumulations to form, a source of organic C is usually needed in addition to water saturation. Fe reduction is not likely to occur if temperatures are too low or if there is insufficient organic matter for microorganisms to access for energy and growth. For these reasons, redoximorphic zones in the Peoria Loess are usually interpreted to be relicts of periods when each zone was closer to the land surface and receiving dissolved organic matter from an overlying A horizon.

Below the Peoria Loess at site HNC-2 (Fig. 1), we encountered an earlier loess deposit, the Pisgah Loess, in which a 1-m-thick paleosol, the Farmdale Soil, was formed. Muhs et al. (3) date the Pisgah Loess accumulation from ~46,000 to ~31,000 years before present. The Farmdale Soil is a widely recognized marker of soil formation during the late Wisconsinan glacial episode in the Midwest. Here, as elsewhere, there was little evidence of a B horizon. The primary evidence of soil formation was the darker color of the soil material, usually interpreted to reflect an increase in organic C, even though the difference in organic C concentration was not detectable by dry combustion techniques that are complicated by the need to remove inorganic C first. In our core, we also recognized weak, platy and subangular blocky soil structure as well as clay coatings in this buried A horizon. The morphology would be consistent with a poorly drained soil in a low-lying landscape position, perhaps near a stream.

Below the Farmdale Soil at site HNC-2, an earlier deposit of loess, the Loveland Loess, occurred. This deposit extended to a depth of about 18 m from the modern surface, and it included about 3 m of the Sangamon Paleosol, another widely recognized pedostratigraphic unit that represents soil development during the interglacial period between the Illinoian glacial episode and the Wisconsinan episode. Brown and Forman (4) have dated the deposition of the Loveland Loess at the Loveland section to 192,000 – 133,000 years before present, depending on the method used to obtain the OSL dates. Our core did not reach the bottom of the Loveland Loess. At the top of the Loveland Loess, we described a well-developed soil with a 67-cm thick A horizon and a B horizon of about 180 cm thickness. The B horizon was an argillic horizon, probably reflecting clay accumulation due to both weathering and translocation of clay particles (Fig. 1).

It is unlikely that climatic conditions during the long Sangamon interglacial period were constant. Where the Sangamon Soil is recognized in other parts of the Midwest, the B horizon usually has reddish colors,

reflecting the weathering release of Fe from primary minerals and formation of Fe oxides. These features have long been interpreted to mean that the interglacial climate may have been warmer and wetter than it is now in the contemporary Midwest. We did not observe these colors in the Sangamon Soil in our core, although other exposures of the soil in western Iowa and across the rest of the state are usually described with 7.5 YR hues or redder.

The microbial habitats in the Sangamon Soil may have been quite different from those in the overlying paleosols and sediments. One reason is the added clay lining of pores in the B horizon would have held water at higher tension than in materials with less clay. In addition, it is worth noting that the dispersion and translocation of clay does not appreciably occur in calcareous sediments. This means that, while the pH of the Sangamon Soil at HNC-2 is now near 8, it was probably much lower than 8 before burial by the Pisgah Loess ~46,000 years before present. The Sangamon Soil has been resaturated with dissolved carbonate translocated from overlying horizons after burial.

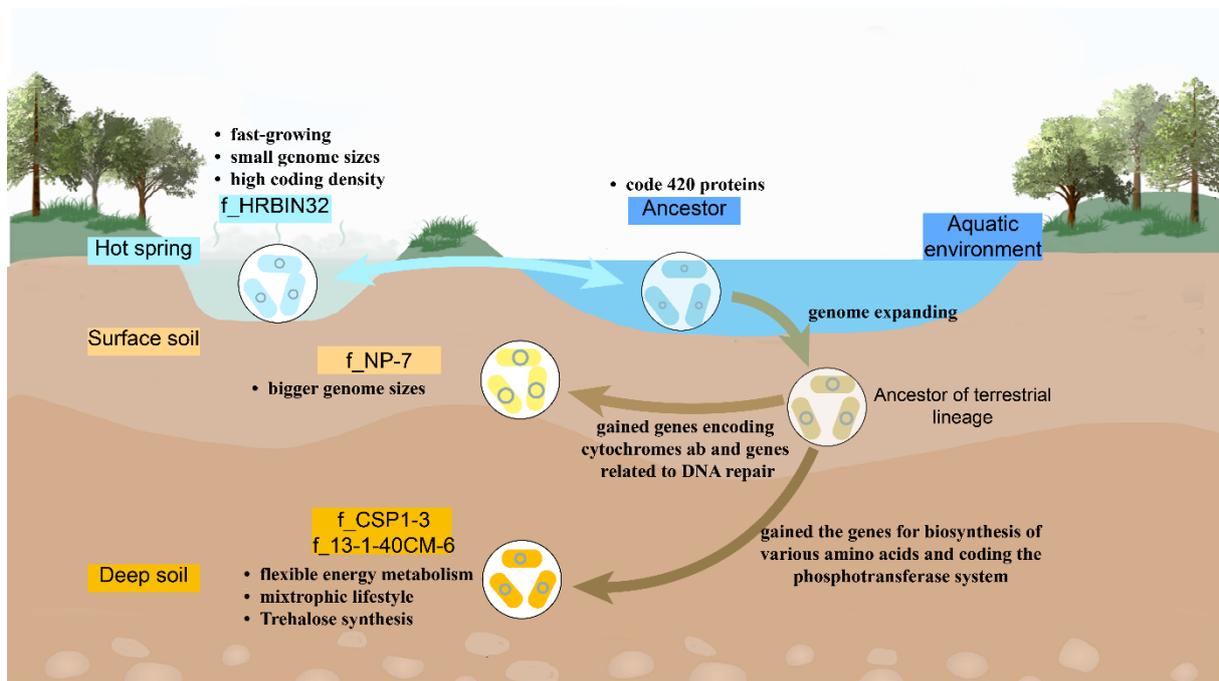


Fig. S1 A diagram showing the evolutionary history from an aquatic organism and adaptive traits of CSP1-3 phylum for each habitat.

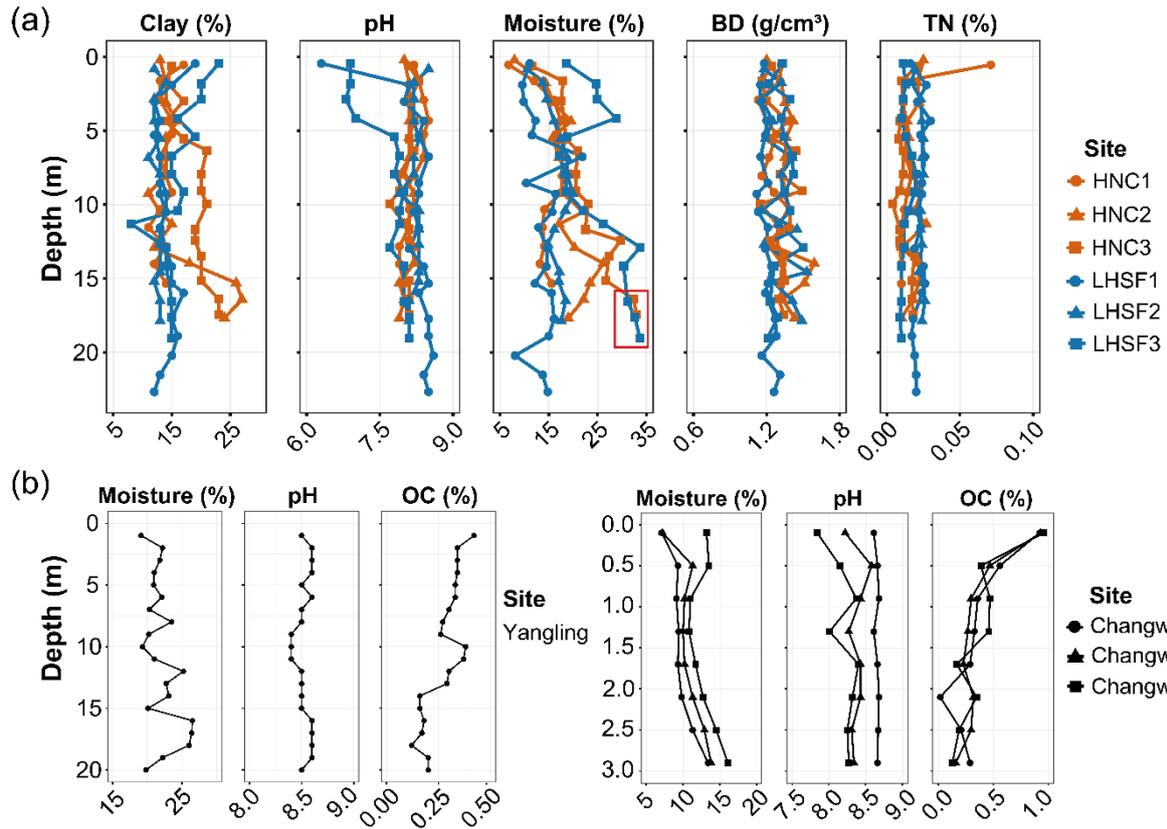


Fig. S2 Physicochemical properties of soils. a Six soil profiles collected at Hitchcock Nature Center (HNC, orange) and Loess Hills State Forest (LHSF, blue) in USA. BD: Bulk Density, TN: Total Nitrogen. Circle, triangle, and squares represent top, middle and bottom slope of the sampling site respectively. The red box indicates saturated moisture content due to proximity to the water table. b A soil profile collected in Yangling, China. OC: organic carbon. c Three soil profiles collected in Changwu, China.

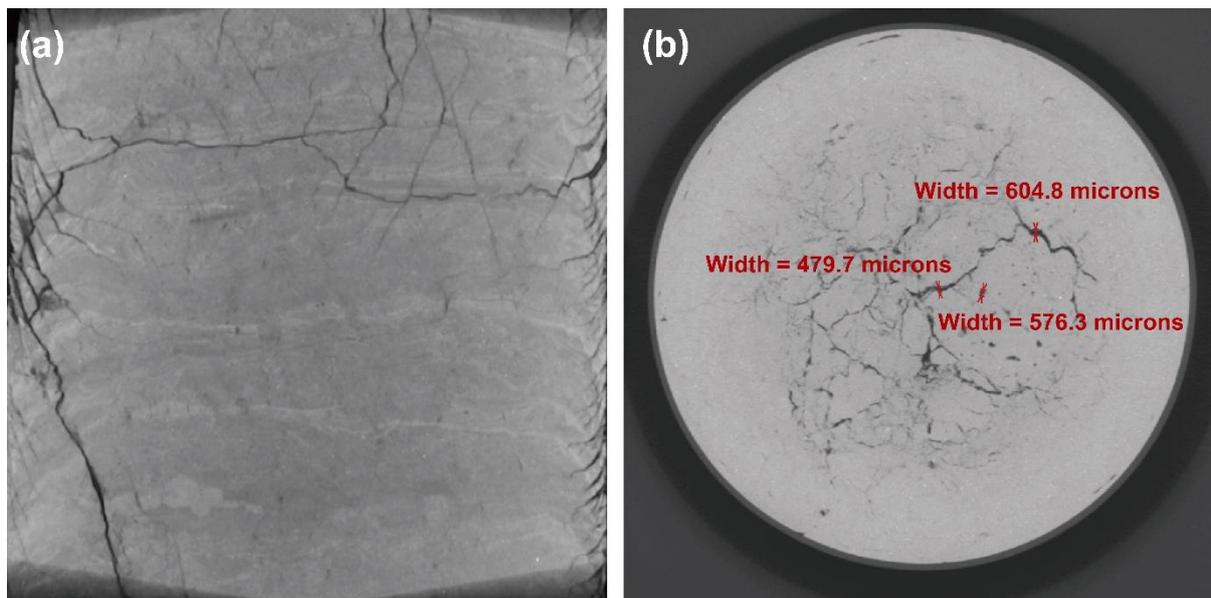


Fig. S3 X-ray Computed Tomography images of soil cores. a Vertical mage showing fine layering of aeolian deposited particles and vertical structure of an intact soil core (19.8–22.0 m) from Loess Hills State Forest. Fractures at edges are from disruption from coring. b Image of the horizontal dimension showing pore distribution and pore dimensions of an intact soil core (15.6–15.8 m) from Hitchcock Nature Center. 3-D scans of the entire cores can be seen in the Supplementary videos.

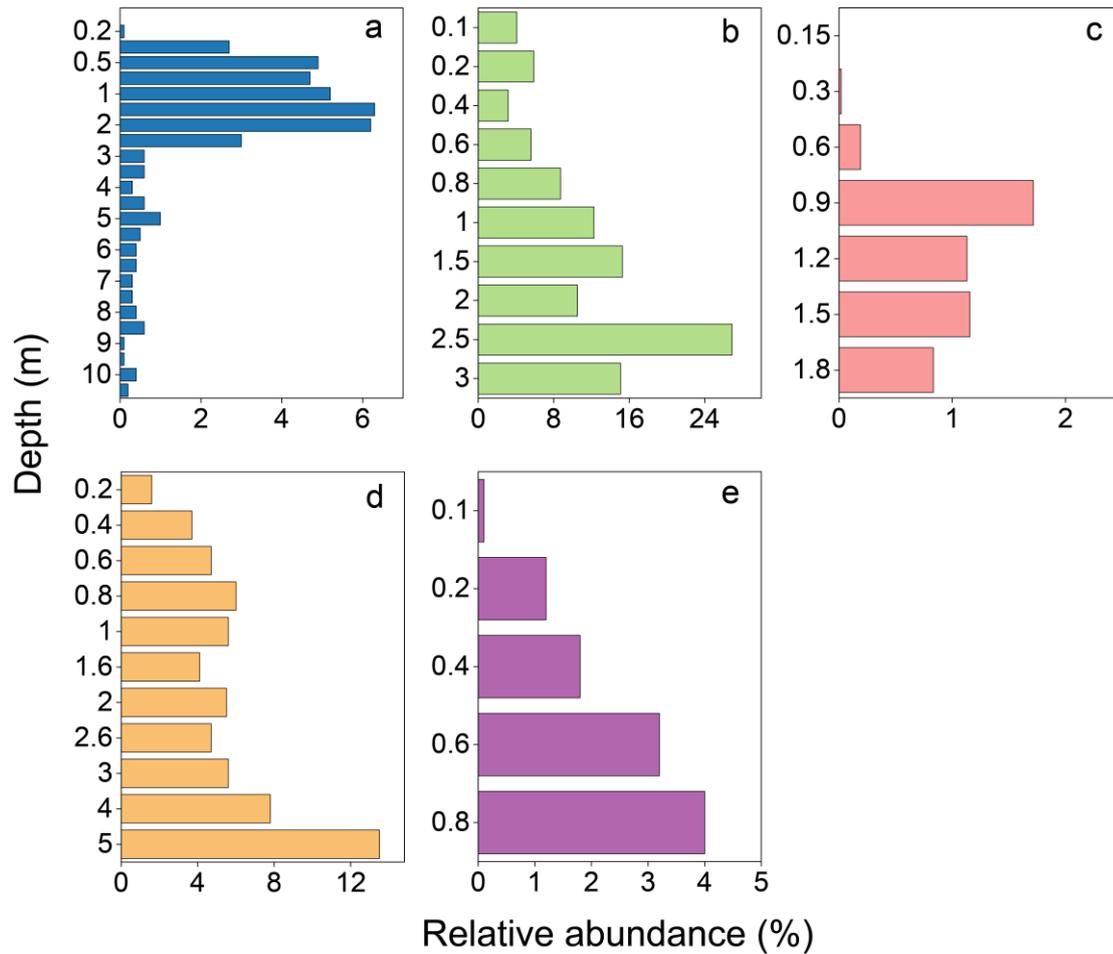


Fig. S4 Abundance of CSP1-3 phylum in published deep soil datasets. a Soil samples (0–10.5 m) from a farmland in Hebei, China (5). b Soil samples (0–3 m) from a forest in Baoji, China (6). c Soil samples (0–1.8 m) from a farmland in Iowa, USA (7). d Soil samples (0–5m) from a cropland, a grassland and a forest in Changwu, China (8). e Soil samples (0–0.8m) from a cropland in Jiangxi, China (9).

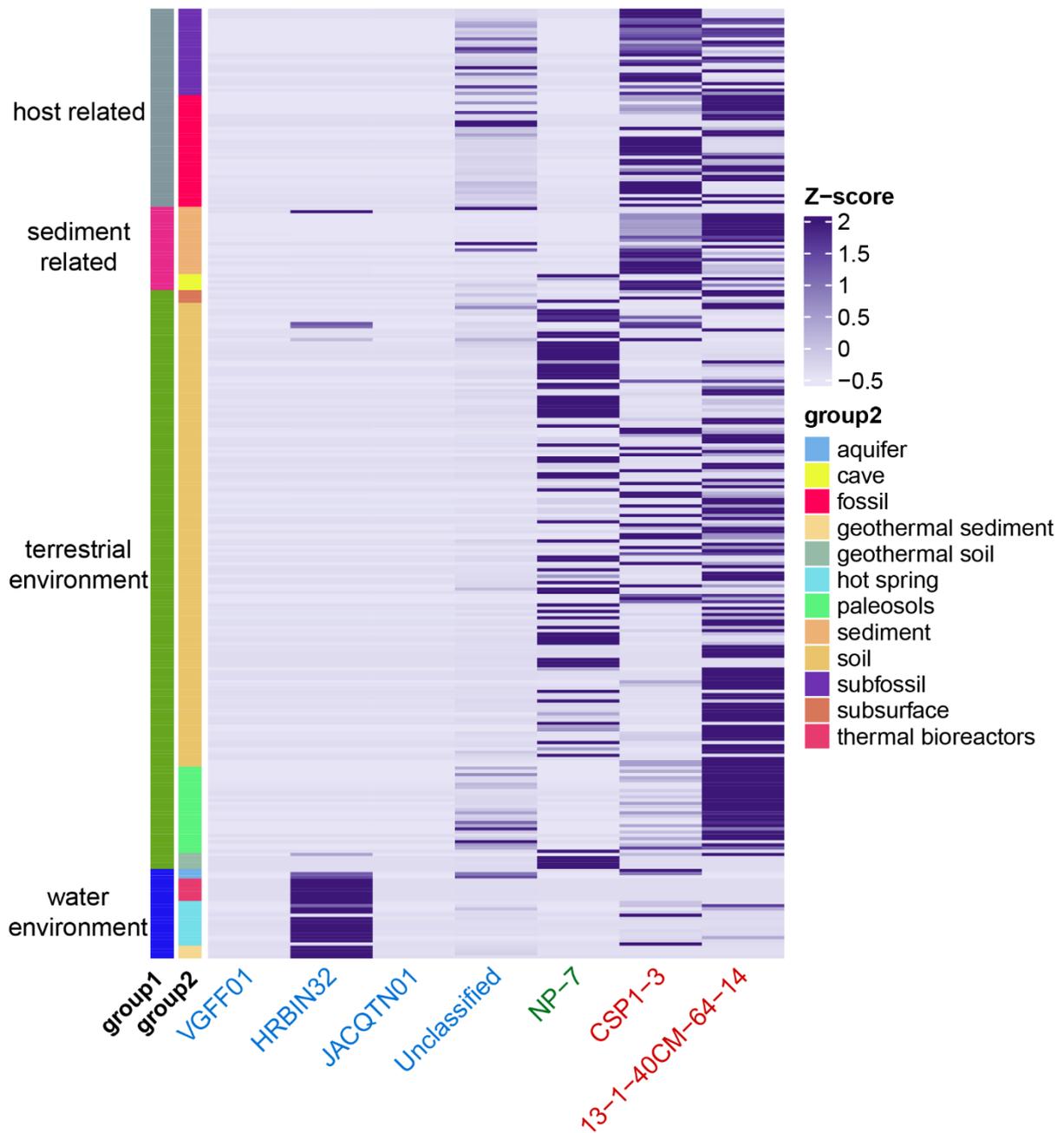


Fig. S6 Abundance of various families within CSP1-3 phylum in different biomes. The labels on the x-axis represent family names, with the text color distinguishing different lineages: blue for aquatic lineages, green for topsoil lineage, and red for deep soil lineages.

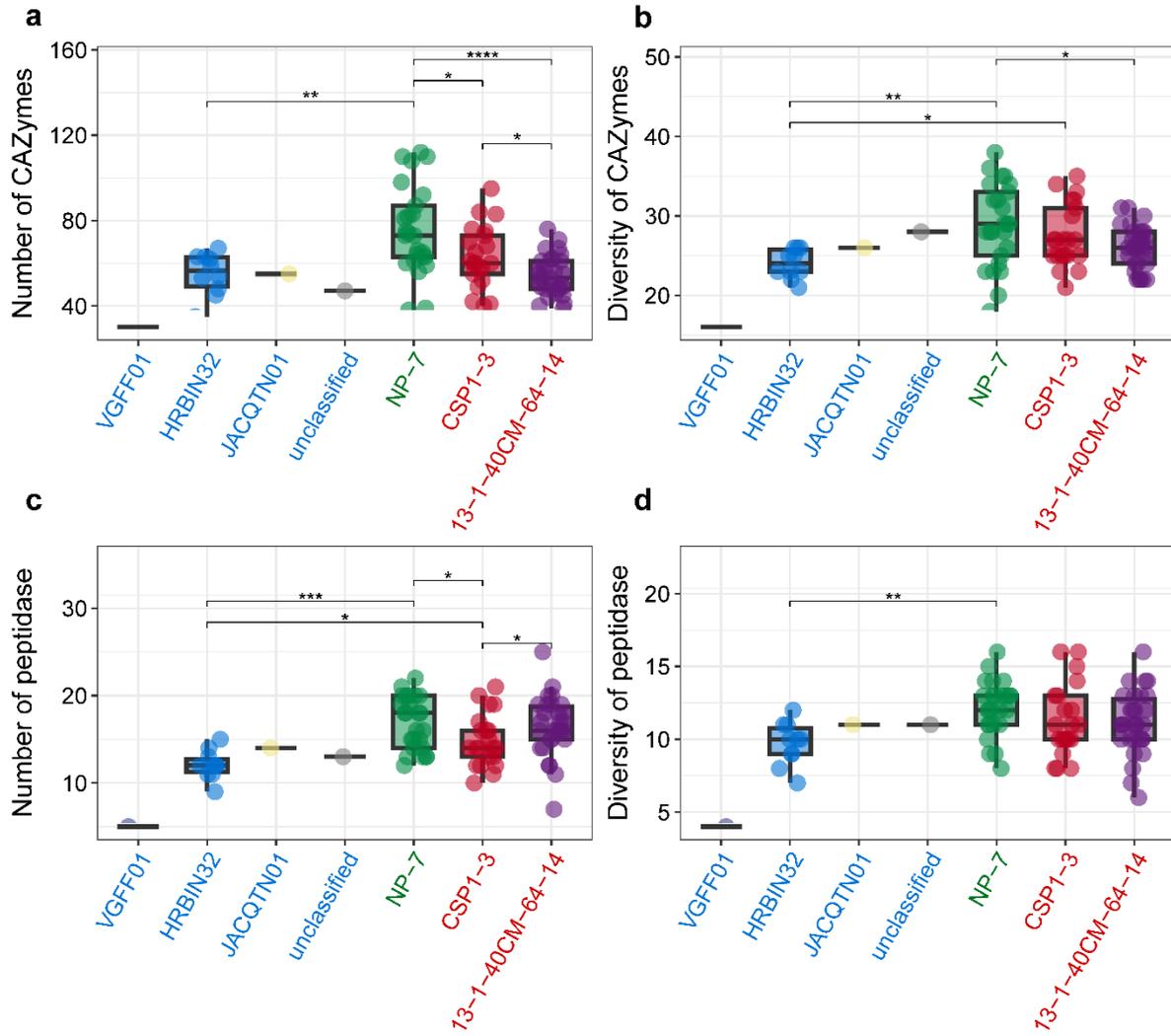


Fig. S7 Genomic differences between different lineages of CSP1-3 phylum. To reduce the estimation biases arising from incomplete genomes, only MAGs with $\geq 80\%$ completeness were retained for the calculation of genomic features. The labels on the x-axis represent family names, with the text color distinguishing different lineages: blue for aquatic lineages, green for topsoil lineage, and red for deep soil lineages.

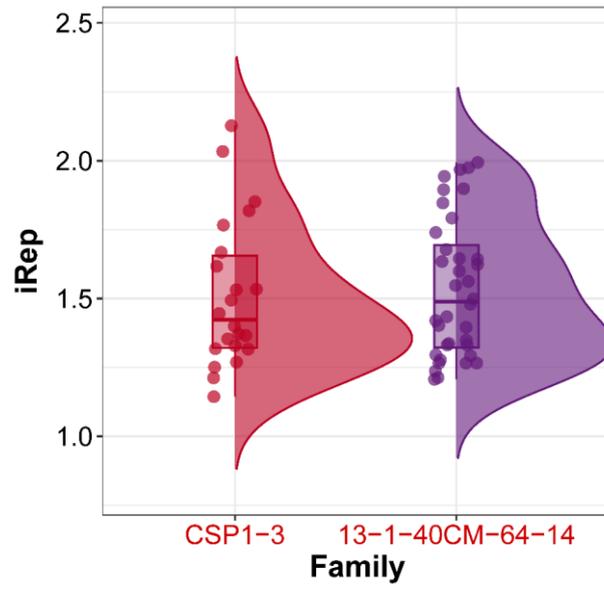


Fig. S9 Replication index (iRep) *in situ* of CSP1-3 and 13-1-40CM-64-14 families in deep soils.
iRep values greater than one indicates active genome replication at time of sampling.

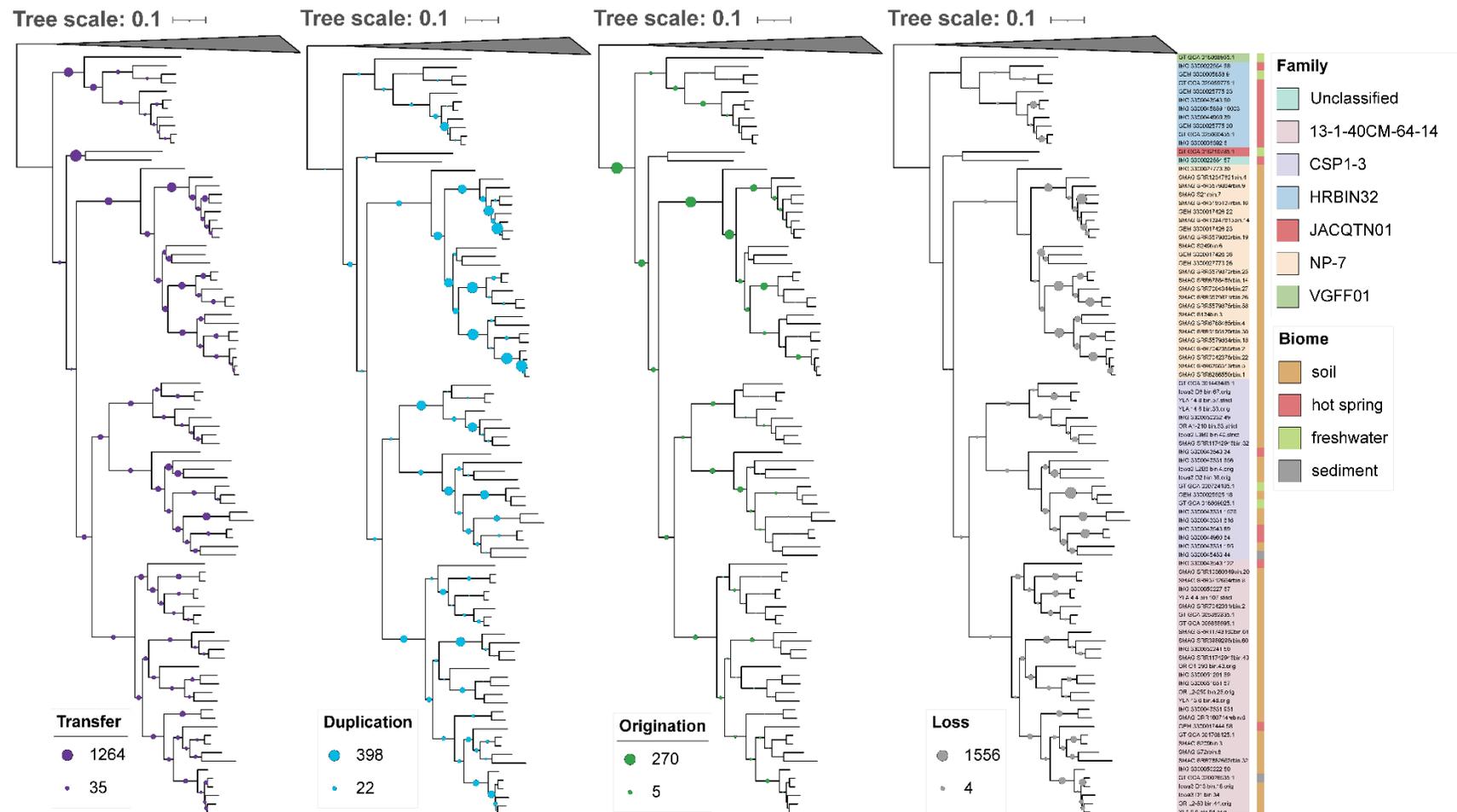


Fig. S10 Number of predicted events inferred for CSP1-3 ancestors. The scale number represents the range of the number of predicted events for different mechanisms, and the size of the circle is proportional to the number of events.

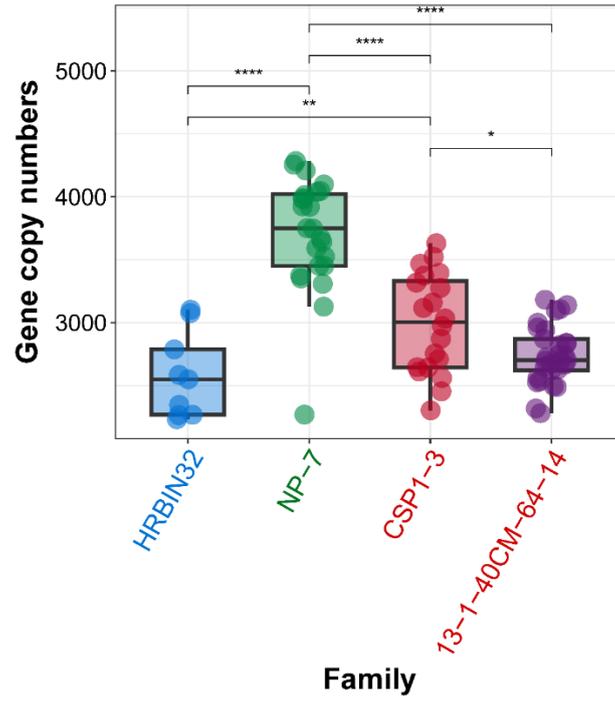


Fig. S11 The predicted proteome sizes at the internal ancestral nodes of major CSP1-3 clades.

Colors of family names distinguish different lineages as noted above.

Dataset S1 (separate file). A genomic catalog of the CSP1-3 phylum.

Dataset S2 (separate file). Information of representative MAGs of the CSP1-3 phylum.

Dataset S3 (separate file). Pairwise AAI of representative MAGs of the CSP1-3 phylum.

Dataset S4 (separate file). Annotation of central carbon metabolisms in genomes of the CSP1-3 phylum.

Dataset S5 (separate file). Annotation of genes encoding carbohydrate active enzymes (CAZymes) in genomes of the CSP1-3 phylum.

Dataset S6 (separate file). Annotation of genes encoding secreted peptidases in genomes of the CSP1-3 phylum.

Dataset S7 (separate file). Functional annotation of the CSP1-3 phylum genomes based on the KEGG database.

Dataset S8 (separate file). Predicted nitrogen metabolisms in genomes of the CSP1-3 phylum.

Dataset S9 (separate file). Predicted sulfur metabolisms in genomes of the CSP1-3 phylum.

Dataset S10 (separate file). Quantified mechanisms of proteome change during evolution of the CSP1-3 phylum.

Dataset S11 (separate file). Functional gains during major evolutionary transitions of the CSP1-3 phylum.

Dataset S12 (separate file). Information of soil samples used for sequencing.

Dataset S13 (separate file). Information on five published 16S rRNA gene amplicon datasets from deep soils.

Movie S1 3-D scans of the entire cores of an intact soil core (19.8–22.0 m) from Loess Hills State Forest.

Movie S2 3-D scans of the entire cores of an intact soil core (15.6–15.8 m) from Hitchcock Nature Center

SI References

1. J. M. Kimble, E. G. Knox, C. S. Holzhey, Soil Survey Laboratory Methods for Characterizing Physical and Chemical Properties and Mineralogy of Soils. (1993).
2. S. S. L. Staff, Soil Survey Laboratory Methods Manual. (2004).
3. D. R. Muhs, *et al.*, Chronology and provenance of last-glacial (Peoria) loess in western Iowa and paleoclimatic implications. *Quaternary Res.* **80**, 468–481 (2013).
4. N. D. Brown, S. L. Forman, Evaluating a SAR TT-OSL protocol for dating fine-grained quartz within Late Pleistocene loess deposits in the Missouri and Mississippi river valleys, United States. *Quaternary Geochronology* **12**, 87–97 (2012).
5. F. Wang *et al.*, Long-term nitrogen fertilization alters microbial community structure and denitrifier abundance in the deep vadose zone. *J. Soils Sed.* **21**, 2394-2403 (2021).
6. S. Jiao *et al.*, Soil microbiomes with distinct assemblies through vertical soil profiles drive the cycling of multiple nutrients in reforested ecosystems. *Microbiome* **6**, 146-146 (2018).
7. J. Hao *et al.*, The effects of soil depth on the structure of microbial communities in agricultural soils in Iowa, USA. *Appl. Environ. Microbiol.* **87**, e02673-02620 (2021).
8. F. Yang, M. Huang, C. Li, X. Wu, L. Fang, Vegetation restoration increases the diversity of bacterial communities in deep soils. *Appl. Soil Ecol.* **180**, 104631 (2022).
9. L. Luan *et al.*, Coupling bacterial community assembly to microbial metabolism across soil profiles. *mSystems* **5**, e00298-00220 (2020).
10. M. N. Price, A. M. Deutschbauer, A. P. Arkin, GapMind: Automated Annotation of Amino Acid Biosynthesis. *mSystems* **5**, 00291-20 (2020).