

# SCIENTIFIC REPORTS



OPEN

## Selection for Oil Content During Soybean Domestication Revealed by X-Ray Tomography of Ancient Beans

Received: 25 November 2016

Accepted: 24 January 2017

Published: 27 February 2017

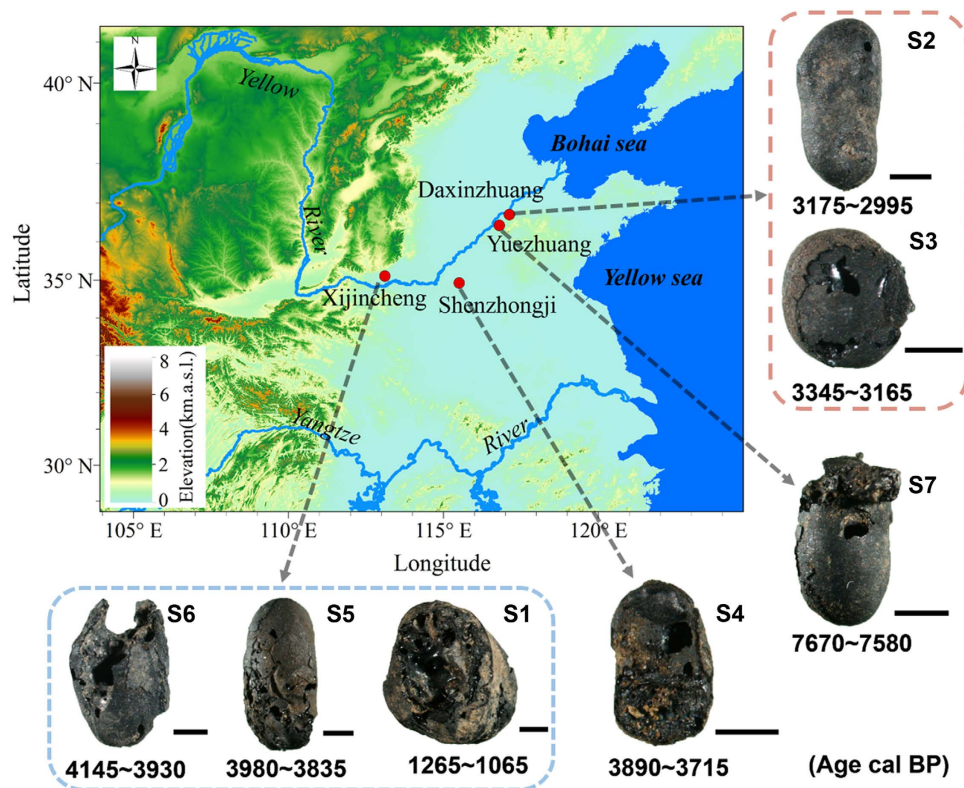
Yunbing Zong<sup>1</sup>, Shengkun Yao<sup>2</sup>, Gary W. Crawford<sup>3</sup>, Hui Fang<sup>4</sup>, Jianfeng Lang<sup>4</sup>, Jiadong Fan<sup>2</sup>, Zhibin Sun<sup>1</sup>, Yang Liu<sup>1</sup>, Jianhua Zhang<sup>1</sup>, Xiulan Duan<sup>1</sup>, Guangzhao Zhou<sup>5</sup>, Tiqiao Xiao<sup>5</sup>, Fengshi Luan<sup>4</sup>, Qing Wang<sup>4</sup>, Xuexiang Chen<sup>4</sup> & Huidong Jiang<sup>1,2</sup>

When and under what circumstances domestication related traits evolved in soybean (*Glycine max*) is not well understood. Seed size has been a focus of archaeological attention because increased soybean seed weight/size is a trait that distinguishes most modern soybeans from their ancestors; however, archaeological seed size analysis has had limited success. Modern domesticated soybean has a significantly higher oil content than its wild counterpart so oil content is potentially a source of new insight into soybean domestication. We investigated soybean oil content using X-ray computed tomography (CT; specifically, synchrotron radiation X-ray CT or SRX-CT) of charred, archaeological soybean seeds. CT identified holes in the specimens that are associated with oil content. A high oil content facilitates the development of small holes, whereas a high protein content results in larger holes. The volume of small holes increased slowly from 7,500 to 4,000 cal B.P. We infer that human selection for higher oil content began as early as 7,500 cal B.P. and that high oil content cultivars were well established by 4,000 cal B.P.

Soybean (*Glycine max*) is one of the world's most important crops largely because of its high protein (40%) and oil (20%) content<sup>1,2</sup>. Modern, domesticated soybean has significantly higher oil content than its wild/weedy counterpart (*G. max* subsp. *soja*, syn. *G. soja*)<sup>3–5</sup> so selection for oil content probably played a role in soybean domestication. What the role may have been is not known. The Yellow River is acknowledged as an important area of agricultural evolution as well as plant and animal domestication during the Late Pleistocene and Early Holocene<sup>6</sup> and appears to be one of at least three regions in East Asia where soybean was domesticated<sup>7</sup>. The earliest associations of people and soybean are in Early Holocene Neolithic communities in the region (9,000–7,500 cal B.P.)<sup>7–12</sup>. Cultivation also likely began at this time, when other crops such as millet and rice were being cultivated<sup>13</sup>. Domestication resulted in at least nine significant changes to the soybean phenotype and genotype<sup>7,14</sup>, among which seed size is the only trait so far accessible to archaeological examination. A study of archaeological soybean seed size points to a relatively late increase in seed size in China<sup>8</sup>. Thus, the selection processes that resulted in the traits that distinguish domesticated from wild soybean (*Glycine soja*, syn. *G. max* subsp. *soja*) today are poorly understood. For instance, soybean pods' reduced dehiscence permits seed retention for harvesting while still allowing the seeds to be easily released during post-harvest processing<sup>15–17</sup>. Only one pod has been reported from an archaeological context in China but it has yet to be described<sup>18</sup>. Oil content has potential as an archaeologically visible trait that may help resolve when selection for domestication related traits in soybean began.

Research directed at soybean improvement that compares domesticated soybean variation with that of wild soybean to uncover the genetic foundation of productivity and nutritional content<sup>19,20</sup> can inform the early domestication process. In particular, advances in omics, such as quantitative trait locus (QTL) mapping and

<sup>1</sup>State Key Laboratory of Crystal Materials, Shandong University, Jinan, Shandong 250100, China. <sup>2</sup>School of Physical Science and Technology, ShanghaiTech University, Shanghai 201210, China. <sup>3</sup>Department of Anthropology, University of Toronto Mississauga, Mississauga, Ontario, L5L 1C6, Canada. <sup>4</sup>Department of Archaeology, Shandong University, Jinan, Shandong 250100, China. <sup>5</sup>Shanghai Synchrotron Radiation Facility, Shanghai Institute of Applied Physics, Chinese Academy of Sciences, Shanghai 201800, China. Correspondence and requests for materials should be addressed to X.C. (email: xxchen@sdu.edu.cn) or H.J. (email: jianghd@shanghaitech.edu.cn)



**Figure 1.** Location of sites from which CT-scanned archaeological soybeans were recovered. The numbers below each soybean image indicate AMS dates (cal B.P., 2- sigma range). Scale bar: 1 mm. The map was created with ArcGIS v.10.1 (ESRI, <http://www.esri.com/software/arcgis>) and annotated by the authors.

genome-wide association studies (GWASs) are defining the genetic basis for soybean oil and protein variation<sup>21–31</sup>. In this report, we explore (1) X-ray computed tomography (CT) as a tool to examine a charred seed structure related to oil content and (2) how the results inform soybean domestication. CT was used to non-destructively reveal the interior three-dimensional structure of charred soybean seeds. CT scans of archaeological and modern control specimens indicate that a unique, porous structure in soybean comprised of non-randomly organized spaces that we call “holes” is related to oil content. The relationship also holds for protein because the oil and protein content of soybean seeds are negatively correlated. Archaeological, charred soybean seeds dating from 7,700 to 1,065 cal B.P. from the middle and lower Yellow River valley provide evidence that changes in oil content over time can be detected (Fig. 1). Furthermore, selection for higher levels of oil in soybean through time is evident.

The archaeological soybeans used in this study are from four sites on the North China Plain: Yuezhuang (YZ)<sup>32</sup>, Shenzhongji (SZJ)<sup>33</sup>, Daxinzhuang (DXZ)<sup>34</sup> and Xijincheng (BXJ)<sup>35</sup> (Fig. 1 and Supplementary Table S1), representing the Houli (B.P. 8,500–7,500), late Longshan (B.P. 4,200–3,900), Shang (B.P. 3,600–3,000) and Tang (B.P. 1,400–1,100) cultures. Soybean compositional traits are affected by environmental factors in addition to genetics so limiting this study to a narrow geographic region should minimize the impact of environmental differences on our study<sup>25,36,37</sup>.

A control sample of twenty-one seeds, seven from each of three land races of modern soybeans with known oil and protein content were CT-scanned to examine the relationship between the holes in the charred soybean seeds and their oil and protein content. The control samples were charred at 275 °C and 300 °C for 3 hours in order to better understand the impact of temperature and the resulting physical and chemical transformations affecting the structures of the archaeological specimens<sup>38</sup>. During the carbonization process, physical structure and chemical composition of seeds change as a function of temperature and heat treatment time<sup>39</sup>. Additionally, two charred, modern cultivated soybeans and one charred modern wild soybean were CT-scanned. Four other taxa with oil-rich seeds were also examined in order to determine whether oil content influences their structures after charring: hemp (*Cannabis sativa*), castor (*Ricinus communis*), perilla (*Perilla frutescens*), rapeseed (*Brassica napus*) and safflower (*Carthamus tinctorius*). These specimens were compared to archaeological seeds of four crops with low to no oil content but rich in starch or polysaccharides: bread wheat (*Triticum aestivum*), rice (*Oryza sativa*), buckwheat (*Fagopyrum esculentum*) and foxtail millet (*Setaria italica* subsp. *italica*). Finally, adzuki (*Vigna angularis*), a legume (Fabaceae) with low oil, high protein, and high starch content, rounded out the control sample.

Synchrotron radiation X-ray CT (SRX-CT), which was used in this study, is a non-destructive, three-dimensional imaging method used to obtain the three-dimensional or volumetric representation of an object. Synchrotron X-rays have higher flux and brightness than a conventional X-ray tube source, permitting

| Sample number | Site name   | Culture/Period | Radiocarbon Lab number (all Beta) | Conventional $^{14}\text{C}$ age B.P. | 2 Sigma calibrated (cal B.P.) |
|---------------|-------------|----------------|-----------------------------------|---------------------------------------|-------------------------------|
| S1            | Xijincheng  | Tang           | 418894                            | 1,230 ± 30                            | 1,265–1,065                   |
| S2            | Daxinzhuang | Shang          | 418904                            | 2,940 ± 30                            | 3,175–2,995                   |
| S3            | Daxinzhuang | Shang          | 418900                            | 3,040 ± 30                            | 3,345–3,165                   |
| S4            | Shenzhongji | Longshan       | 418895                            | 3,530 ± 30                            | 3,890–3,715                   |
| S5            | Xijincheng  | Longshan       | 418903                            | 3,600 ± 30                            | 3,980–3,835                   |
| S6            | Xijincheng  | Longshan       | 418898                            | 3,690 ± 30                            | 4,145–3,930                   |
| S7            | Yuezhuang   | Houli          | 418897                            | 6,820 ± 40                            | 7,700–7,590                   |

**Table 1. Direct AMS radiocarbon dates of archaeological charred soybean seeds.** The half-life of  $^{14}\text{C}$  is 5,568 y. CT-scanned S1, S3, and S5–6 were dated. S2, S4 and S7 used both the scanned and associated soybean seeds from the same archaeological feature to enable sufficient carbon for dating.

higher resolution data to be collected. This method has proven effective in resolving problems with large palaeontological and archaeological specimens, such as fossil embryos<sup>40–43</sup>, Neanderthal molars<sup>44</sup>, and papyrus rolls recovered at Herculaneum<sup>45</sup>. To the best of our knowledge, this study represents the first time that valuable compositional information from charred seeds has been obtained using CT. Although the beam time limits the number of specimens that can be examined, we selected as many archaeological and modern soybeans and other crops as time permitted. Overall, thirty modern seeds, seven archaeological seeds, and ten other crop seeds were studied.

## Results

**Radiocarbon Dating.** Accelerator mass spectrometer (AMS) radiocarbon dates were obtained for soybeans from each assemblage to confirm the ages of the archaeological beans (Table 1). The AMS dates confirm the chronological sequence; in addition, the soybeans are not intrusive into their respective contexts. The dates span a range of approximately 6,000 years.

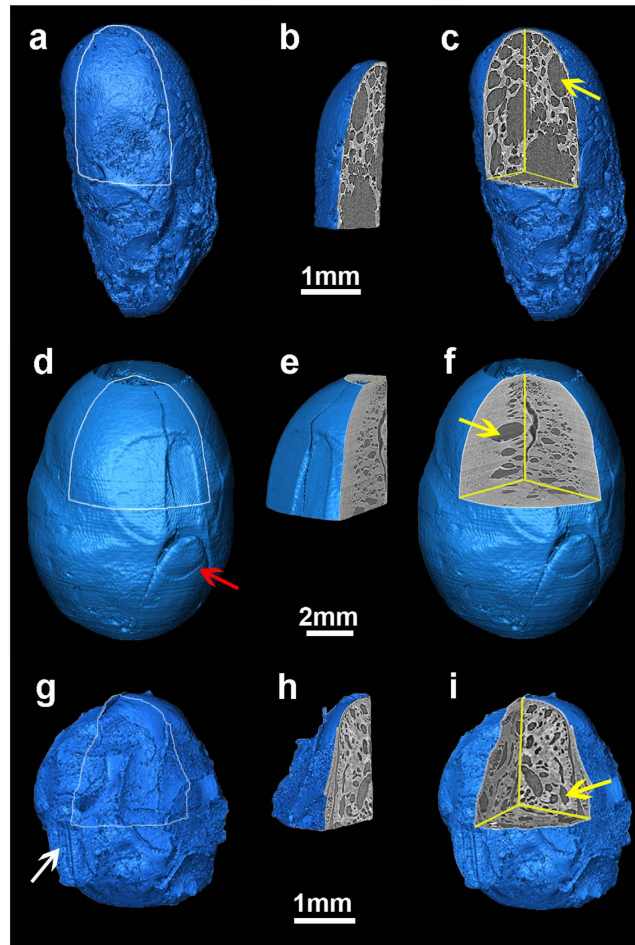
**Holes.** Charred soybean cotyledons almost never split apart and normally appear quite porous with numerous spaces or holes, rendering them unlike other charred bean family (Fabaceae) seeds recovered from archaeological sites in East Asia, such as adzuki (*V. angularis*). These spaces or holes are present in the cotyledons of each of the charred ancient, modern domesticated, and wild soybean specimens (Fig. 2a, d, and g, respectively). One spheroidal or ellipsoidal “hole” structure was discovered (yellow arrows in Fig. 2c, f and i). The sizes of the holes range from several tens of micrometers to hundreds of micrometers, which is substantially larger than the normal cotyledon cell size (Supplementary Fig. S1). The holes, therefore, are not specific cells.

**Hole Quantification.** The seeds were virtually segmented, and the hole volumes were calculated using Amira, a 3D visualization and analysis application (Mercury Computer Systems, Berlin, Germany). The three-dimensional segmentation result is displayed in Fig. 3. Holes that were too small to be differentiated were considered to be outliers and not included in the analysis. Large breaks that are a result of cotyledon separation were excluded. A few very small cracks are included because they could not be isolated from holes in the analysis. They represent less than 1 percent of the holes. The process of hole segmentation is shown in Supplementary Fig. S2 and indicates that holes are arranged throughout the cotyledons (Fig. 2), likely because the cotyledon cell structure is homogenous (Supplementary Fig. S3).

**Hole Formation.** Tissue paraffin-sections were examined with a visible light microscope (Supplementary Fig. S1d) and transmission electron microscopy (Supplementary Fig. S3). Charring at high temperatures (~275 °C) clearly leads to hole generation in soybean seeds (Supplementary Fig. S1a–c). Holes are present in the five other oil-rich crops but they are more irregularly distributed than in soybean (Supplementary Fig. S4a–e). Holes were not detected in the charred, archaeological wheat, rice, millet and buckwheat seeds (Supplementary Fig. S5). Adzuki, with a low oil content, has only uniformly distributed holes (Supplementary Fig. S4f). The six control samples confirm that oil (very little to no oil is present in wheat, rice, millet and buckwheat) has a significant role to play in soybean hole formation. TEM imaging shows the distribution of protein and oil in the cells (Supplementary Fig. S3).

The process of hole formation appears to vary among the high oil content seeds, with soybean having a unique pattern, likely resulting from its particular oil and protein content. High temperatures cause complex chemical and physicochemical processes that ultimately cause the small oil droplets that are distributed throughout the seed to coalesce into larger drops<sup>46</sup>. Heating also changes the protein structure so that the seed is permeable to oil; that is, the oil can move through the cotyledons to coalesce and form pockets of oil that form the holes in the charred seeds. Once carbonized, the large oil droplets are eliminated from the charred soybeans<sup>46</sup>. Furthermore, the irregular distribution of protein bodies in soybean seeds may help explain the irregular structure noted in charred soybean cotyledons and not in crops rich in starch or polysaccharides<sup>47</sup>.

To further assess how protein and oil affect the generation of holes, cultivated soybean seeds were examined. First, the protein and oil content was measured<sup>48–50</sup>. Then, the specimens were carbonized at 275 °C for 3 hours. We selected three groups of soybeans whose protein and oil content varied: Meng (protein 48.7% and oil 19.2%, P<sub>48.7</sub>O<sub>19.2</sub>), Fu (protein 42.5% and oil 15%, P<sub>42.5</sub>O<sub>15</sub>) and Tiefeng (protein 33.9% and oil 23.6%, P<sub>33.9</sub>O<sub>23.6</sub>). Seven seeds in each group were CT-scanned. The normalized hole numbers and normalized hole volumes (absolute



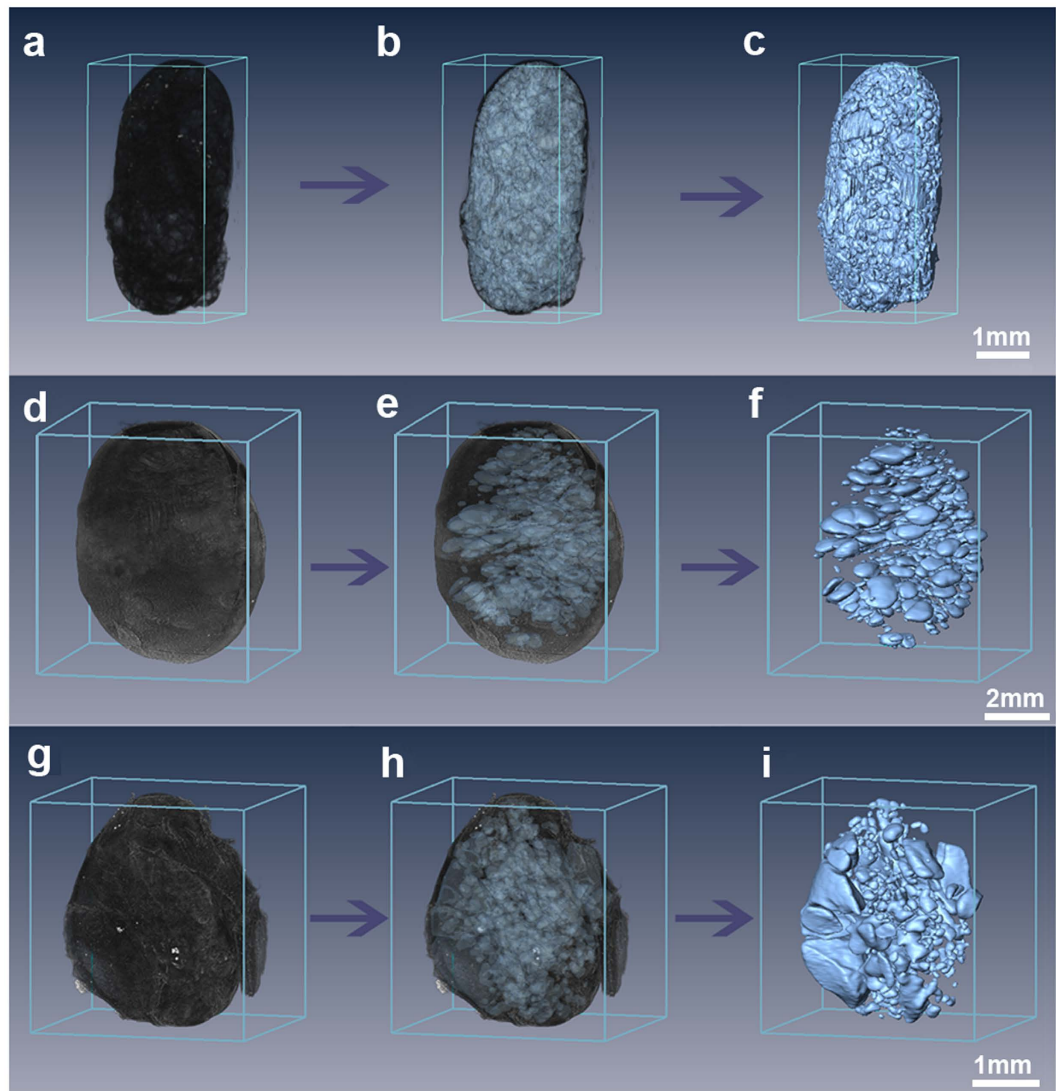
**Figure 2. Three-dimensional reconstructed images of charred soybeans and their internal structures.** (a,d,g) are the surface rendering results of archaeological soybean, modern cultivated soybean and modern wild soybean, respectively. (b,e,h) are segments cut from (a,d,g). (c,f,i) are the remaining portions. Little morphological information could be drawn from (a) because of the complicated physical, chemical and microorganism processes that resulted in significant degradation of the specimen. An embryo can be clearly recognized in (d), indicated by the red arrow. The hilum in (g) is indicated by the white arrow. (g) is a modern wild soybean collected locally. Holes in (c), (f) and (i) are indicated by the yellow arrows.

hole counts and volumes were divided by the soybean volume to calculate normalized values and are hereafter referred to simply as “volume” and “number.”) were used for statistics (Fig. 4). We then split the volumes into three categories using Amira: Small ( $1E-7$  order), Medium ( $1E-6$  order) and Large ( $1E-5$  order). There were more holes in  $P_{33.9}O_{23.6}$  than in  $P_{48.7}O_{19.2}$  or  $P_{42.5}O_{15}$  in the Small category ( $P = 0.0002$ , one-way ANOVA and Dunnett’s T3 analysis for both, Fig. 4). In the Medium category,  $P_{33.9}O_{23.6}$  and  $P_{48.7}O_{19.2}$  had more holes than  $P_{42.5}O_{15}$  that had the lowest oil content ( $P < 0.0001$  and  $P = 0.002$ , respectively, one-way ANOVA and Dunnett’s T3 analysis for both). However, in the Large category,  $P_{48.7}O_{19.2}$  had a greater mean than  $P_{42.5}O_{15}$  or  $P_{33.9}O_{23.6}$  due to its higher protein content ( $P < 0.0001$  for both, one-way ANOVA and Fisher’s LSD *post hoc* analysis for both). Regardless, this was not the case for the medium protein content of  $P_{42.5}O_{15}$  compared with  $P_{33.9}O_{23.6}$  ( $P = 0.002$ , one-way ANOVA and Fisher’s LSD *post hoc* analysis). In summary, the comparison of  $P_{48.7}O_{19.2}$  with  $P_{42.5}O_{15}$  indicates that the overall protein and oil content is directly correlated with hole number. Considering the extremely high oil content of  $P_{33.9}O_{23.6}$ , we conclude that oil facilitates the generation of smaller holes. In turn, a higher protein content leads to the generation of larger holes.

**Temperature Gradient Experiment.** A sample of six of the  $P_{33.9}O_{23.6}$  seeds were charred at  $300\text{ }^{\circ}\text{C}$  for 3 hours. The number of holes in the Small category ( $P = 0.0004$ , two-tailed t-test, Fig. 5), was less at  $300\text{ }^{\circ}\text{C}$  than at  $275\text{ }^{\circ}\text{C}$  and little difference was noted in the Medium and Large categories ( $P = 0.51$  and  $P = 0.07$ , two-tailed t-test for both). In summary, hole number in the Small category clearly reduced when the charring temperature increased from  $275\text{ }^{\circ}\text{C}$  to  $300\text{ }^{\circ}\text{C}$ .

**Water Saturation Experiment.** Soybean seeds readily absorb water so we investigated the potential role of water in hole formation. Modern seeds of the unnamed cultivar (shown in Fig. 2d) were soaked for 10 hours and then charred at  $275\text{ }^{\circ}\text{C}$  for 3 hours. After soaking, the beans enlarged approximately 1.4 times their original





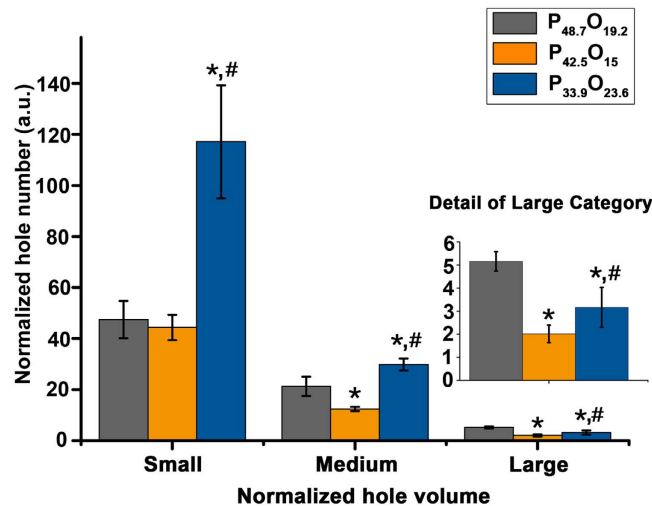
**Figure 3. Segmentation of holes in 3D reconstructed images of soybeans.** (a,d,g) are the volume rendering results of archaeological soybean, modern cultivated soybean and modern wild soybean, respectively, of the same seeds illustrated in Fig. 2. (b,e,h) are the transparent images showing holes in the soybeans. (c,f,i) are the surface results of the extracted holes.

size. After charring, the soaked beans were less dense and more brittle than the unsoaked beans. One charred seed was selected for CT scanning. We calculated the total hole volume, the post-charring mass density and other relevant parameters (Supplementary Table S2). The holes became larger compared with the dry seed due to soaking in water and smaller holes (volume level below  $1E-5$ ) were not evident (Fig. 6). The hole structure of charred, water-saturated seeds does not resemble any of the archaeological specimens, so water can be eliminated as a significant factor impacting the hole structure of the archaeological samples in this study.

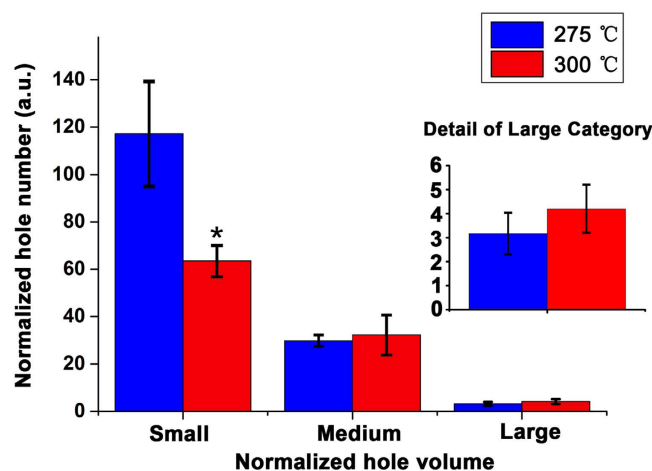
### Discussion

CT scanning revealed a particular hole structure in charred seeds that we linked to oil and protein content. The hole distribution for each bean is presented as the normalized hole volume plotted against numerical rank (Fig. 7). Numerical rank represents the holes in each bean sorted from large to small. Each curve represents one soybean, and each symbol on a curve indicates one hole. Seven archaeological soybeans and three modern soybeans are shown (Fig. 7). The number of holes may be less important than the size distribution of holes because the complete hole structure of many seeds is masked by their incompleteness.

Oil percentage has tended to increase through time (Fig. 7a). Three modern soybean seeds (two domesticated and one wild) and two archaeological soybean seeds from the Tang dynasty (S1) and Longshan period (S4) have significantly higher hole volumes than the others (Group 1). Beans from the Shang (S2, S3), Longshan (S5, S6) and Houli periods (S7) have a lower hole volume distribution (Group 2). The modern cultivated soybean has significantly more oil than modern wild soybean (Fig. 7b), so the earliest soybean seeds in the archaeological



**Figure 4. Statistics of normalized hole volume vs. normalized hole number of known protein and oil content soybeans.** P<sub>48.7</sub>O<sub>19.2</sub> (Meng), P<sub>42.5</sub>O<sub>15</sub> (Fu), and P<sub>33.9</sub>O<sub>23.6</sub> (Tiefeng) are three different cultivars of soybeans with varying protein and oil content. Seven seeds were averaged for each cultivar. The asterisk (\*) indicates significant differences compared with holes in P<sub>48.7</sub>O<sub>19.2</sub> ( $P < 0.05$ ; one way ANOVA test). The pound sign (#) indicates significant differences compared with holes in P<sub>42.5</sub>O<sub>15</sub> ( $P < 0.05$ ; one way ANOVA test). The bars indicate standard error. Statistically not significant differences with  $P > 0.05$  are not indicated.

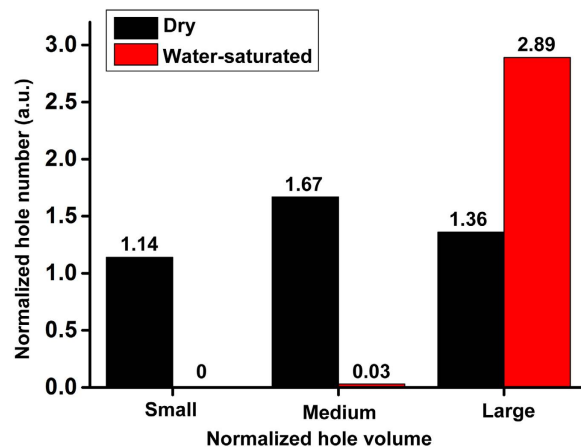


**Figure 5. Temperature gradient experiment.** Six P<sub>33.9</sub>O<sub>23.6</sub> soybean seeds were carbonized at 275 °C. In the Small category, seeds charred at 300 °C had about half the number of seeds with small holes then when fired at 275 °C. The Medium and Large categories were not affected by temperature. The asterisk (\*) indicates significant differences compared with holes in 275 °C ( $P < 0.01$ , two-tailed t-test). The bars indicate standard error. Statistically not significant differences with  $P > 0.05$  are not indicated.

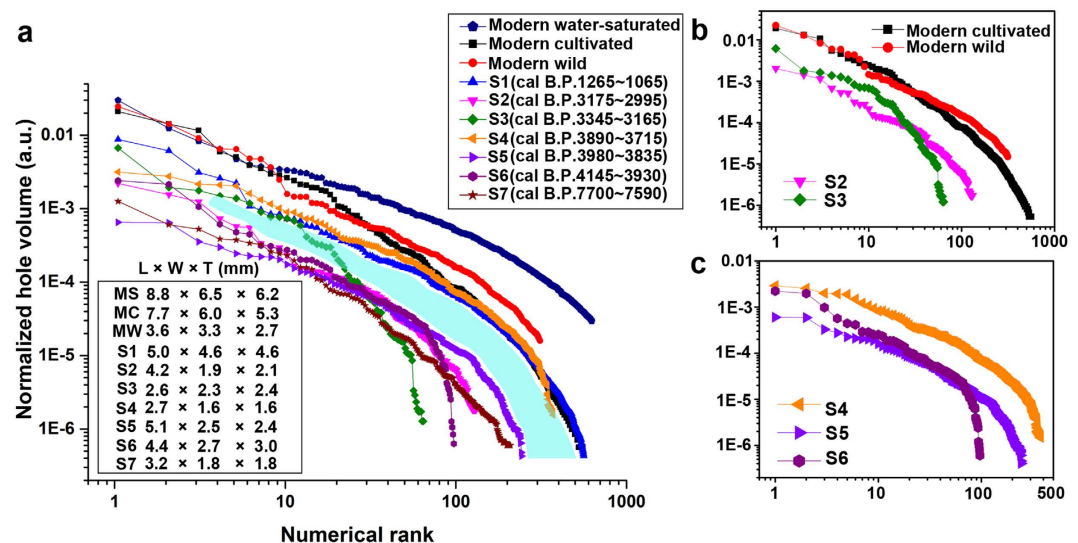
record, if wild, should have also had low oil content. All archaeological specimens in this study appear to have substantially higher oil content than that of the modern wild soybean seed.

A comprehensive analysis of archaeological soybean seed size indicated that both wild/weedy and domesticated soybean seeds are represented in the archaeological record at sites dating to the Longshan and Shang periods in China<sup>7</sup>. We compared the protein and oil content of the two size modes at the Shang period Daxinzhuan site (groups S2 and S3; Fig. 7b) in order to test whether there are significant differences in the seed composition of the two size categories. Their size difference is distinct, with S2 averaging (L, W, T) 4.2 by 1.9 by 2.1 mm and S3 2.6 by 2.3 by 2.4 mm. The previous study suggested that S3 was wild/weedy and that S2 was cultivated/domesticated<sup>7</sup>. S3 (hypothetically wild/weedy) has a lower percentage of oil but higher percentage of protein than S2. S2 (hypothetically domesticated soybean) has a higher percentage of oil and lower protein percentage than S3. The results are consistent with S2 being cultivated/domesticated and S3 being wild/weedy.

S5 and S6 are from the Longshan period Xijincheng site. The specimens both fall within the large mode although some size differentiation is evident. S5 is longer than S6 (5.1 by 2.5 by 2.4, product = 30.6) vs. (4.4 by 2.7



**Figure 6. Water saturation experiment.** A water-saturated soybean seed was charred and CT-scanned. Normalized hole volumes were split into three categories: Small ( $1E-7$  order), Medium ( $1E-6$  order) and Large ( $1E-5$  order). Compared with one dry seed of the same cultivar, few holes (below  $1E-5$ ) were detected in the water-saturated seed.



**Figure 7. Quantification analysis of soybean holes.** (a) Seven archaeological soybeans from different periods and three modern soybeans are compared with each other and with modern water-saturated (MS), modern cultivated (MC) and modern wild (MW) soybeans. The blue shading separates the data into two: modern cultivated and wild soybean and archaeological soybeans S1 and S4 in the upper region, and others in the lower region. (b) Compares modern cultivated and wild soybeans, S2 and S3. The negative correlation between the oil and protein content of S2 and S3 is indicated. Modern cultivated soybean has a higher oil content than the modern wild seed. (c) Three Longshan period soybeans: S4–6. A negative correlation between the oil and protein content is indicated in S5 and S6. However, S4 shows an unusual distribution.

by 3.0, product = 35.6). Though differentiation in the hole number range 20–60 is similar, there are more small holes after number 80 in S5 that has a higher oil percentage than S6 (Fig. 7c).

The archaeological record for soybean in China indicates that seed size increase was a late development and that the size increase when it did occur was not comparable to the changes in soybean size documented in the Japanese and Korean archaeological record<sup>7</sup>. Thus, in China, traits other than seed size were probably being selected during the Early through Late Neolithic. If high oil content was being selected, then oil content would increase and not necessarily be directly correlated with size during that period. This appears to be the case. The Longshan S4 sample is quite small (2.7 by 1.6 by 1.6 mm) compared to S5 and S6 yet has higher protein and oil content (Fig. 7c). However, due to the lack of large-sized seeds from the same context, the hole distribution in S4 is difficult to explain. The Houli seed (S7), the oldest specimen considered in this study, has a higher oil content than the modern wild soybean seeds. We suggest that the Houli soybeans had already undergone some selection for higher oil content by 7,500 B.P.

The distribution of smaller holes indicates that oil content significantly increased relative to protein content as soybean was undergoing domestication in China. The light-blue region (Fig. 7a) divides Group 1 from Group 2, indicating a wide chronological gap during which selection proceeded. The Houli period Yuezhuang sample (S7) suggests that selection under cultivation had begun at least 7,500 years ago and that soybean domestication spanned a period of at least 3,500 years and probably longer. A larger sample from the period spanning 8,000–4,000 B.P. is needed to fully test this hypothesis.

To compare the control soybeans (P<sub>48.7</sub>O<sub>19.2</sub>, P<sub>42.5</sub>O<sub>15</sub> and P<sub>33.9</sub>O<sub>23.6</sub>) with the archaeological seeds (S1–S7), a similar plot of hole distribution is displayed in Supplementary Fig. S6. Before hole number 2500, P<sub>48.7</sub>O<sub>19.2</sub> and P<sub>42.5</sub>O<sub>15</sub> are located an upper position due to their high content of protein (48.7% and 42.5%). After hole number 2500, P<sub>33.9</sub>O<sub>23.6</sub> surpasses P<sub>48.7</sub>O<sub>19.2</sub> and P<sub>42.5</sub>O<sub>15</sub> due to its high oil content.

## Conclusions

The effectiveness of CT scanning in investigating plant domestication was tested on a sample of archaeological soybeans from four archaeological sites in the Yellow River valley spanning the period from 7,500 to 1,065 cal B.P. A “hole” structure in charred soybean seeds appears to be linked to protein and oil content. We hypothesize that smaller holes are related to oil content and larger holes are related to protein content. Selection for seed composition appears to have occurred as early as 7,500 B.P., indicating that soybean domestication and cultivation were occurring by this time. Significant differentiation of soybean is not evident until 4,000 B.P., and by 3,500 B.P., during the Shang period, soybean seed size and oil content substantially increased. Soybean likely underwent further selection for large seed size and higher oil and protein content, distinguishing it further from its wild/weedy ancestor that had smaller seeds, less oil, and higher protein<sup>25</sup>. Importantly, the “hole” structure is a new clue that helps our understanding of soybean domestication; without CT, this structure could not have been assessed or quantified. CT offers significant potential for archaeologists and biologists exploring plant domestication.

## Methods

**Soybeans and other crop seeds.** All archaeological seeds were provided by the Department of Archaeology, Shandong University. The reference soybeans were provided by Institute of Crop Sciences at the Chinese Academy of Agricultural Sciences (ICS-CAAS). The wild soybeans were provided by the Shandong Academy of Agricultural Sciences. The other modern domesticated crop seeds were purchased from a local seed supplier.

**Carbonization experiment.** A laboratory muffle furnace was used to char the modern seeds. The seeds were baked for 3 hours under anoxic conditions.

**X-Ray computed tomography experiment and 3D reconstruction.** The tomography experiment was conducted on beamline 13W1 at the Shanghai Synchrotron Radiation Facility (SSRF) (Supplementary Fig. S7). The Wiggler system provides high monochromaticity of  $\Delta E/E < 5 \times 10^{-3}$ . For specimens of different sizes, an X-ray energy range of 12–15 keV was selected. A charge coupled device (CCD) detector of a pixel size of 6.5  $\mu\text{m}$ , with two interchangeable lenses -  $2 \times$  and  $4 \times$  magnification, were used for the experiment. Thus, the effective pixel resolution can be respectively reduced to 3.25  $\mu\text{m}$  and 1.625  $\mu\text{m}$  for tomographic images. To enhance the projection contrast, that is, to make full use of both absorption contrast and phase contrast, we set the distance between samples and CCD as 8 cm. The in-line phase contrast tomography improved contrast compared with conventional absorption contrast tomography<sup>51,52</sup>. Flat images were collected when the sample moved out of the beam path. Finally, when all of the projections and flat images had been collected, the shutter was switched off to shut down the X-ray beam, and five dark images were taken in order to subtract the dark noise. Then, parts of the images were used to reconstruct the three-dimensional result using equally sloped tomography (EST)<sup>53</sup>. The others were reconstructed using a filtered back projection (FBP) algorithm<sup>54</sup> with the PITRE software<sup>55</sup>. All details regarding the scanning parameters and radiation dose for each seed are listed in Supplementary Table S3.

**Paraffin sectioning and electron microscopy experiment.** The soybean tissue paraffin-sectioning was performed in reference to standard plant paraffin section flow<sup>56</sup>. Seeds underwent electron microscopy using a general laboratory biology electron microscope. Protein and oil were identified by staining.

**Statistical analysis.** All statistical analyses were performed using Predictive Analysis Software 18.0 (SPSS Inc., Chicago, IL, USA). Intergroup comparisons were performed using one-way ANOVA followed by the LSD test (with equal variances assumed) or Dunnett's T3 test (equal variances not assumed). One-way ANOVA was used to compare the three soybean cultivars. A two-tailed t-test was used to compare two groups of P<sub>33.9</sub>O<sub>23.6</sub> that were charred at different temperatures.

**Data availability.** The data that support the findings of this study are included in this article and in the Supplementary Information Files, or are available from the corresponding authors upon request.

## References

- Kim, M. Y., Van, K., Kang, Y. J., Kim, K. H. & Lee, S. H. Tracing soybean domestication history: From nucleotide to genome. *Breeding Sci* **61**, 445–452 (2012)
- Singh, R. J., Nelson, R. L. & Chung, G. H. Soybean (*Glycine max* (L.) Merr.). In *Genetic Resources, Chromosome Engineering, and Crop Improvement. Oilseed Crops* Vol. 4 (ed. Singh, R. J.) 13–50 (CRC Press, 2007).
- Caldwell, B. E. & Howell, R. W. *Soybeans: Improvement, Production, and Uses*. Vol. 6 (American Society of Agronomy, 1973).
- Chen, Y. W. & Nelson, R. L. Genetic variation and relationships among cultivated, wild, and semiwild soybean. *Crop Sci.* **44**, 316–325 (2004).



5. Pantalone, V. R., Rebetzke, G. J., Wilson, R. F. & Burton, J. W. Relationship between seed mass and linolenic acid in progeny of crosses between cultivated and wild soybean. *J Am Oil Chem Soc* **74**, 563–568 (1997).
6. Zhao, Z. New archaeobotanic data for the study of the origins of agriculture in China. *Curr Anthropol* **52**, S295–S306 (2011).
7. Lee, G. A., Crawford, G. W., Liu, L., Sasaki, Y. & Chen, X. Archaeological soybean (*Glycine max*) in East Asia: does size matter? *PLoS One* **6**, e26720 (2011).
8. Hymowitz, T. On the domestication of the soybean. *Economic Botany* **24**, 408–421 (1970).
9. Crawford, G. W. East Asian plant domestication. In *Archaeology of Asia* (ed. Stark, M. T.) 77–95 (Blackwell, Malden, MA, 2006).
10. Zohary, D. & Hopf, M. Domestication of Pulses in the Old World: Legumes were companions of wheat and barley when agriculture began in the Near East. *Science* **182**, 887–894 (1973).
11. Lee, G. A., Crawford, G. W., Liu, L. & Chen, X. Plants and people from the early Neolithic to Shang periods in North China. *Proc Natl Acad Sci USA* **104**, 1087–1092 (2007).
12. Crawford, G. W. *et al.* Late Neolithic plant remains from Northern China: Preliminary results from Liangchengzhen, Shandong. *Curr Anthropol* **46**, 309–317 (2005).
13. Zheng, Y., Crawford, G. W., Jiang, L. & Chen, X. Rice Domestication Revealed by Reduced Shattering of Archaeological rice from the Lower Yangtze valley. *Sci Rep* **6**, 28136 (2016).
14. Zeder, M. A., Emshwiller, E., Smith, B. D. & Bradley, D. G. Documenting domestication: The intersection of genetics and archaeology. *Trends Genet* **22**, 139–155 (2006).
15. Dong, Y. *et al.* Pod shattering resistance associated with domestication is mediated by a NAC gene in soybean. *Nat Commun* **5**, 3352 (2014).
16. Tang, H. *et al.* Seed shattering in a wild sorghum is conferred by a locus unrelated to domestication. *Proc Natl Acad Sci USA* **110**, 15824–15829 (2013).
17. Funatsuki, H. *et al.* Molecular basis of a shattering resistance boosting global dissemination of soybean. *Proc Natl Acad Sci USA* **111**, 17797–17802 (2014).
18. Fuller, D. Q., Harvey, E., Qin, L. & Zhang, H. A preliminary report of the survey archaeobotany of the upper Ying Valley (Henan Province). In *Archaeological Discovery and Research at Wangchenggang Site in Dengfeng (2002–2005)* (eds School of Archaeology and Museology of Peking University and Henan Provincial Institute of Cultural Relics and Archaeology) 916–958 (Great Elephant Publisher, 2007).
19. Larkin, P. J. & Scowcroft, W. R. Somaclonal variation - a novel source of variability from cell cultures for plant improvement. *Theor Appl Genet* **60**, 197–214 (1981).
20. Kaul, S. *et al.* Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796–815 (2000).
21. Kim, M. Y. *et al.* Whole-genome sequencing and intensive analysis of the undomesticated soybean (*Glycine soja* Sieb. and Zucc.) genome. *Proc Natl Acad Sci USA* **107**, 22032–22037 (2010).
22. Lam, H. M. *et al.* Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nat Genet* **42**, 1053–1059 (2010).
23. Li, Y. *et al.* Molecular footprints of domestication and improvement in soybean revealed by whole genome re-sequencing. *BMC Genom* **14**, 579 (2013).
24. Chung, W. H. *et al.* Population structure and domestication revealed by high-depth resequencing of Korean cultivated and wild soybean genomes. *DNA Res* **21**, 153–167 (2014).
25. Zhou, Z. *et al.* Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nat Biotechnol* **33**, 408–414 (2015).
26. Vaughn, J. N., Nelson, R. L., Song, Q., Cregan, P. B. & Li, Z. The genetic architecture of seed composition in soybean is refined by genome-wide association scans across multiple populations. *G3: Genes Genom Genet* **4**, 2283–2294 (2014).
27. Hwang, E. Y. *et al.* A genome-wide association study of seed protein and oil content in soybean. *BMC Genom* **15**, 1 (2014).
28. Sonah, H., O'Donoghue, L., Cober, E., Rajcan, I. & Belzile, F. Identification of loci governing eight agronomic traits using a GBS-GWAS approach and validation by QTL mapping in soya bean. *Plant. Biotechnol. J.* **13**, 211–221 (2014).
29. Wang, Y., Chen, P. & Zhang, B. Quantitative trait loci analysis of soluble sugar contents in soybean. *Plant Breeding* **133**, 493–498 (2014).
30. Pathan, S. M. *et al.* Genetic mapping and confirmation of quantitative trait loci for seed protein and oil contents and seed weight in soybean. *Crop Sci* **53**, 765–774 (2012).
31. Warrington, C. *et al.* QTL for seed protein and amino acids in the Benning × Danbaekkong soybean population. *Theor Appl Genet* **128**, 839–850 (2015).
32. Crawford, G. W., Chen, X., Luan, F. & Wang, J. People and plant interaction at the Houli Culture Yuezhuang site in Shandong Province, China. *The Holocene* **26**, 1594–1604 (2016).
33. Zhi, T. *et al.* Test excavation of Shenzhongji site, Caoxian County, Shandong Province (in Chinese). *Archaeology* **5**, 385–390 (1980).
34. Chen, X. & Fang, H. A case study of the agricultural economy in the Shang Dynasty: Macro-plant remains from the Daxinzhuang Site, Jinan, China (in Chinese). *East Asia Archaeol* **5**, 43–64 (2008).
35. Chen, X., Wang, L. & Wang, Q. Discussion on the plant remains of Xijincheng Site (2006–2007), Boai County, Henan Province (in Chinese). *Huaxia Archaeol* **3**, 67–76 (2010).
36. He, S., Wang, Y., Li, D. & Yi, T. Environmental and historical determinants of patterns of genetic differentiation in wild soybean (*Glycine soja* Sieb. et Zucc.). *Sci Rep* **6**, 22795 (2016).
37. Clemente, T. E. & Cahoon, E. B. Soybean oil: genetic approaches for modification of functionality and total content. *Plant Physiol* **151**, 1030–1040 (2009).
38. Renfrew, C. & Bahn, P. *Archaeology: theories, methods, and practice*. (Thames and Hudson, London, 1996).
39. Braadbaart, F., Wright, P. J., Vander Horst, J. & Boon, J. J. A laboratory simulation of the carbonization of sunflower achenes and seeds. *J Anal Appl Pyrolysis* **78**, 316–327 (2007).
40. Chen, L., Xiao, S., Pang, K., Zhou, C. & Yuan, X. Cell differentiation and germ-soma separation in Ediacaran animal embryo-like fossils. *Nature* **516**, 238–241 (2014).
41. Chen, J. *et al.* Phosphatized polar lobe-forming embryos from the Precambrian of southwest China. *Science* **312**, 1644–1646 (2006).
42. Donoghue, P. C. J. *et al.* Synchrotron X-ray tomographic microscopy of fossil embryos. *Nature* **442**, 680–683 (2006).
43. Hagadorn, J. W. *et al.* Cellular and subcellular structure of Neoproterozoic animal embryos. *Science* **314**, 291–294 (2006).
44. Macchiarelli, R. *et al.* How Neanderthal molar teeth grew. *Nature* **444**, 748–751 (2006).
45. Mocella, V., Brun, E., Ferrero, C. & Delattre, D. Revealing letters in rolled Herculaneum papyrus by X-ray phase-contrast imaging. *Nat Commun* **6**, 5895, doi:10.1038/ncomms6895 (2015).
46. Shahidi, F. *Bailey's Industrial Oil and Fat Products* Vol 5 (ed. Shahidi, F.) 109–118 (John Wiley & Sons, Inc., Publication, 2005).
47. Henry, A. G., Hudson, H. F. & Piperno, D. R. Changes in starch grain morphologies from cooking. *J Archaeol Sci* **36**, 915–922 (2009).
48. Qiu, L. *et al.* A platform for soybean molecular breeding: The utilization of core collections for food security. *Plant Mol Biol* **83**, 41–50 (2013).
49. Wang, L. *et al.* Establishment of Chinese soybean *Glycine max* core collections with agronomic traits and SSR markers. *Euphytica* **151**, 215–223 (2006).
50. Guo, Y., Li, Y., Hong, H. & Qiu, L. Establishment of the integrated applied core collection and its comparison with mini core collection in soybean (*Glycine max*). *Crop J* **2**, 38–45 (2014).

51. Gureyev, T. E., Mayo, S., Wilkins, S. W., Paganin, D. & Stevenson, A. W. Quantitative in-line phase-contrast imaging with multienergy X rays. *Phys Rev Lett* **86**, 5827–5830 (2001).
52. Wu, X., Liu, H. & Yan, A. Phase-contrast X-ray tomography: Contrast mechanism and roles of phase retrieval. *Eur J Radiol* **68**, S8–S12 (2008).
53. Miao, J., Förster, F. & Levi, O. Equally sloped tomography with oversampling reconstruction. *Phys Rev B* **72**, 052103 (2005).
54. Flannery, B. P., Deckman, H. W., Roberge, W. G. & D'Amico, K. L. Three-dimensional X-ray microtomography. *Science* **237**, 1439–1444 (1987).
55. Chen, R. C. *et al.* PITRE: software for phase-sensitive X-ray image processing and tomography reconstruction. *J Synchrotron Radiat* **19**, 836–845 (2012).
56. Liu, Y., Xiu, Z., Meeley, R. & Tan, B. Empty pericarp5 encodes a pentatricopeptide repeat protein that is required for mitochondrial RNA editing and seed development in maize. *Plant Cell* **25**, 868–883 (2013).

## Acknowledgements

We thank the National Crop Germplasm Resources of China (2014–004, 2015–004) and Lifeng Zhang for providing the modern soybean samples. We also acknowledge the experimental assistance of Xiaoqiong Liu, Bo Liu, Shuo Li, Hongchun Wang, Qi Wang, Wei Yang, and Wei Gong. This work was supported by the National Natural Science Foundation of China (31430031, 21390414, U1332118), the National Social Science Foundation of China (14BKG009) and the Fundamental Research Funds of Shandong University (2014JC007, 2082012041).

## Author Contributions

H.J., X.C., and H.F. designed the study. Y.Z., S.Y., J.F., Y.L., X.D., J.Z., Z.S., G.Z. and T.X. performed the CT experiments. J.L., F.L., Q.W. collected soybean seeds. Y.Z. and H.J. analyzed majority of data and performed statistical analyses. Y.Z., G.W.C., H.J. and X.C. wrote the manuscript. All authors discussed the results and commented on the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Zong, Y. *et al.* Selection for Oil Content During Soybean Domestication Revealed by X-Ray Tomography of Ancient Beans. *Sci. Rep.* **7**, 43595; doi: 10.1038/srep43595 (2017).

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



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017