Archaeogenetic study of prehistoric rice remains from Thailand and India: Evidence of early japonica in South and Southeast Asia

Cristina Cobo Castillo[1], Katsunori Tanaka[2], Yo-Ichiro Sato[3], Ryuji Ishikawa[2], Bérénice Bellina[4,] Charles Higham[5], Nigel Chang[6], Rabi Mohanty[7], Mukund Kajale[7], Dorian Q Fuller[1*]

1. Institute of Archaeology, University College London, 31-34 Gordon Square, London WC1H 0PY, U.K

2. Hirosaki University, 1 Bunkyo, Hirosaki, Aomori, 036-8560, Japan

3. Kyoto Sangyo University, Motoyama, Kamigamo, Kita, Kyoto, 603-8555, Japan

4. C.N.R.S (National Centre for Scientific Research) - UMR 7055 (Préhistoire et Technologie), 21, allée de l'Université, F-92023 Nanterre, France

5. Department of Anthropology and Archaeology, University of Otago

6. James Cook University, Townsville 4811, Queensland, Australia

7. Deccan College Post-Graduate and Research Institute, Pune, Maharashtra, India

* Corresponding author

Keywords: *Oryza sativa*, rice, domestication, archaeobotany, aDNA, charred grains

Abstract: We report successful extraction and sequencing of ancient DNA from carbonised rice grains (*Oryza sativa*) from six archaeological sites, including two from India and four from Thailand, ranging in age from *ca.* 2500 to 1500 BP. In total, 221 archaeological grains were processed by PCR amplification and primary-targeted fragments were sequenced for comparison with modern sequences generated from 112 modern rice populations, including crop and wild varieties. Our results include the genetic sequences from both the chloroplast and the nuclear genomes, based on 4 markers from the chloroplast and 6 from the nuclear genome. These markers allow differentiation of *indica* rice from *japonica* rice, the two major subspecies of Asian rice (*Oryza sativa*) considered to have separate geographical origins. One nuclear marker differentiates tropical and temperate forms of subspecies *japonica*. Other markers relate to phenotypic variation selected for under domestication, such as non-shattering, grain stickiness (waxy starch), and pericarp colour. Recovery and identification of sequences from nuclear markers was generally poor, whereas recovery of chloroplast sequences was successful, with at least one of four markers recovered in 61% of archaeological grains. This allowed for successful differentiation of *indica* or *japonica* subspecies variety, with *japonica* identified in all the Thai material and a mixture of *indica* and *japonica* chloroplasts in the two Indian assemblages. Rice subspecies was also assessed through conventional archaeobotanical methods relying on grain metrics, based on measurements from 13 modern populations and

499 archaeological grains. Grain metrics also suggest a predominance of *japonica* type grains in the Southeast Asian sites and a mixture of *japonica* and *indica* in the Indian sites with *indica* in the minority. The similar results of grain metrics and aDNA affirms grain measurements have some degree of reliability in rice subspecies identification. The study also highlights the great potential of ancient DNA recovery from archaeological rice. The data generated in the present study adds support to the model of rice evolution that includes hybridization between *japonica* and proto-*indica*.

Introduction

This paper provides the results from archaeobotanical, genetic and morphometric analysis conducted on selected rice remains from archaeological sites in South and Southeast Asia. To date, there have been no published ancient DNA (aDNA) studies on rice from prehistoric sites in Thailand and India. This paper provides the first genetic study of rice from four prehistoric sites in Thailand and two in India, totalling eleven archaeological rice assemblages with aDNA data, consisting of 211 rice grains. Morphometric analyses from the same assemblages are considered in order to compare morphological and genetic indications of rice subspecies from archaeological remains. The archaeological sites are from Bronze and Iron Age Thailand, two in Nakhon Ratchasima Province (Ban Non Wat and Noen U-Loke) and two in the Thai-Malay Peninsula (Khao Sam Kaeo, Chumphon and Phu Khao Thong, Ranong). The Indian Early Historic archaeological sites are Ter in Maharashtra and Balathal in Rajasthan [Figure 1].

Ban Non Wat [BNW] was excavated by Higham and Chang (Boyd and Chang 2010; Duke *et al.* 2010; Higham and Kijngam 2009, 2011, 2012a, 2012b; Kanthilatha et al. 2014); Noen U-Loke [NUL] was excavated as part of the 'Origins of Angkor' project spearheaded by Higham (Higham *et al.* 2007); and Khao Sam Kaeo [KSK] and Phu Khao Thong [PKT] by The Thai-French archaeological mission led by Bellina (Bellina *et al.* 2014). Extensive archaeobotanical research was undertaken in all these four sites in Thailand which gives a more robust interpretative framework of overall results. The working hypothesis is that the rice remains from all the prehistoric Thai sites used in the aDNA study were *Oryza sativa* subspecies *japonica* rather than subspecies *indica.* This hypothesis was based on archaeobotanical analysis, including morphometrics, and the archaeological geography of early rice (Castillo and Fuller 2010; Fuller *et al.* 2010; Castillo 2011). Whilst *indica* rice is the Indian originated subspecies (Cheng *et al.* 2003; Kovach *et al.* 2007) and could have reached Thailand during the initial contact between India and Thailand ~400 - 100 BC, this does not appear to be the case.

The Indian samples came from Balathal and Ter. The former is an important site of Chalcolithic date (*ca.* 3500-1500 BC) with later Early Historic occupation (*ca.* 300 BC-AD 300). Several seasons of excavation by Deccan College produced a rich archaeological record (e.g. Misra and Mohanty 2001; Shinde *et al.* 2004), including systematic flotation for archaeobotanical remains. These data indicate the dominance of winter cereals (wheat and barley) in the earlier prehistoric period, together with some indigenous Indian crops such as pulses, while rice is present in the Early Historic period contexts (Kajale 1996). Hand-collected samples of charred seed lenses from the 1996 excavation season included kodo millet (*Paspalum scrobiculatum*) and rice, the latter being reported on here. Ter is a well-known Early Historic (Satavahana to Indo-Roman cultural phases) settlement mound, identified with *Tagara* of ancient Roman sources (Casson 1989), where archaeological investigations were carried out in the 1960s and 1970s. Archaeobotanical studies include those of Vishnu-Mittre *et al.* (1971), based on hand-collected samples, and Kajale (1975) based on small-scale flotation. While recording a ~5 metre high exposed stratigraphic section in 1997, samples from a large charcoal lens were collected; these were found to consist of a sub-lens of kodo millet and rice.

This paper provides a brief summary on the background of rice with particular reference to how the markers used in the present study relate to current hypotheses on the origins, domestication and diversification of Asian rice. This is followed by the methodologies used in the genetic study and the morphometric analysis. The results from the archaeogenetic and morphometric studies are then presented, followed by a discussion of their implications.

Background on rice

There are two main cultivated subspecies of *O. sativa,* namely *indica* and *japonica*, and these have been taxonomically recognized as distinct since the 1930s (Kato 1930; Oka 1988). Their distinct chloroplast genomes indicate a multiple origins model (Chen *et al.* 1993; Nakamura *et al.* 1998; Takahashi *et al.* 2008). Geneticists have established that *indica* is derived from an extinct *Oryza nivara*-like ancestor, while *japonica's* wild progenitor is from a subset of *Oryza rufipogon (sensu stricto)* also extirpated from a likely domestication area in central China (Fuller *et al.* 2010). Some geneticists refer to *O. nivara* as annual *O. rufipogon* and therefore the literature often refers to only *O. rufipogon (sensu lato)* but with two ecotypes (Cheng *et al.* 2003). The genetic study by Garris *et al.* (2005) maintains distinct and separate domestication events for *indica* and *japonica* (temperate), with the divergence between the groups from the ancestral *O. rufipogon* estimated in 100,000s of years ago (Ma and Bennetzen 2004; Tang *et*

al. 2004; Tian *et al.* 2006; Wei *et al.* 2012). Garris *et al.* (2005) also suggest that temperate and tropical *japonica* have a deep divergence and that aromatic and *aus* clades can be separated among modern landraces. Thus many recent genetic studies recognize five groups of Asian rice: *indica, aus*, aromatic, temperate *japonica* and tropical *japonica* (the last including the subspecies *javanica*.

Figure 2 expresses the multiple origins hypothesis from Fuller *et al.* (2010), starting with the divergence of the wild progenitor into annual (*O. nivara*) and perennial (*O. rufipogon sensu stricto*) species and further into extinct (extirpated) populations within each of these that gave rise to the two major subspecies of *O. sativa*. This diagram incorporates introgression, which has been cited as an additional source of increased diversity in rice by proponents of both single and multiple origins (Fuller *et al.* 2010; Kovach *et al.* 2007; Molina *et al.* 2011; Sang and Ge 2007; Vaughan *et al.* 2008). The *japonica* varieties/lineages include temperate *japonica,* tropical *japonica* and aromatic rice. These are regarded to have differentiated after the domestication and early dispersal of a primitive *japonica* rice, which was probably closer to tropical *japonica*. By contrast the "proto-*indica*" cultivars (*sensu* Fuller and Qin 2009; Fuller *et al.* 2010; Fuller 2011a; 2011b) lacked several key domestication mutations that would have been acquired through a hybridisation process with domesticated *japonica*. As straight crossing of *indica* x *japonica* leads to poor seed yields due to incompatibility (Wan and Ikehashi 1997), it is inferred that backcrossing to the *indica* parent would have been necessary to produce early productive *indica* varieties (Fuller *et al.* 2010; Sato 1996). Back-crossing to the *indica* mother population is implied by the many studies that have identified a deep whole genome and chloroplast divergence between *indica* and *japonica* (Ma and Bennetzen 2004; Tang *et al.* 2004; Vitte *et al.* 2004).

Plant domestication is defined as the process whereby a plant is genetically modified from its wild predecessors and adapted to cultivation as a result of human manipulation. The increase in grain size, loss of seed dispersal, lack of awns, and synchronous ripening are the phenotypic expressions of these genetic mutations selected probably unconsciously during the domestication process (Fuller et al. 2014), while some traits like white pericarp or fragrance are post-domestication improvements which might have been actively sought by farmers (Purugganan and Fuller 2009; Meyer and Purugganan 2013). Genes responsible for domestication traits have been identified for many plants and contribute to our understanding

of the trajectory of domestication (Fuller and Allaby 2009; Meyer and Purugganan 2013) though this is often not a clear-cut path. For example, there are several quantitative trait loci (QTLs) identified as affecting non-dehiscence, of which only two have been located and sequenced in detail, the non-shattering genes *sh4* and *qsh1* (Konishi *et al*. 2006; Li *et al.* 2006). Nevertheless, on its own, the evolution of the *sh4* mutation would not have prevented shattering, as it must interact with other factors (Ishii *et al.* 2013). While *sh4* is regarded as universal in both *indica* and *japonica* rice, *qSH1* is restricted to a more limited subset of mainly temperate *japonica,* suggesting more recent evolution (Fuller and Sato 2008; Fuller *et al*. 2010; Kovach *et al.* 2007; Purugganan 2010; Sang and Ge 2007; Zhang *et al.* 2009). The archaeologically supported proto-*indica* hypothesis posits that *sh4* was introduced into proto-*indica* cultivars via introgression with *japonica* resulting in non-shattering *indica* (Fuller and Qin 2009; Fuller *et al.* 2010; Fuller 2011a), but this took place with *japonica* rice that lacked or pre-dated the *qSH1* non-shattering mutation. The loci *sh4* and *qSH1* were recovered in some of the archaeological rice grains used in our aDNA study.

Introgression has played an important role in the history of rice and shows the close contact between *indica* and *japonica. Rc* is a post-domestication gene, one that is neither necessary nor universal in cultivars, that is found in both *japonica* and *indica.* It is responsible for the white or brown pericarp, another single origin mutation originating from *japonica* and flowing to *indica* (Purugganan 2010; Sweeney *et al.* 2007). That this intraspecific introgression may have taken place intentionally by farmers cannot be discounted because it is often the case that certain traits are desired by groups of people, for example those involving colour. Black rices represent an alternative cultural selection on the *Pb* gene (Wang and Shu 2007). Cultural selection is also the case for the main genetic determinant responsible for fragrance *BADH2* (*frg),* which is found only in some rices (the aromatic group including basmati and jasmine) and was actively selected for independently by ancestral farmers on multiple occasions in different geographic regions (Kovach *et al.* 2009; Prathepha 2009). However, other factors besides cultural selection, such as climate, soil, seasonality and geography are also responsible for the emergence of aromatic rices. The *waxy* gene is responsible for sticky rice varieties in *japonica* rice although the degree of stickiness is variable. The origins of sticky rice probably lie in China along with sticky millets and travelled south into Southeast Asia (Fuller and Castillo in press) although Purugganan (2010) argues in favour of peninsular Southeast Asia as the origin. Sticky or glutinous rices are popular in East and Southeast Asia but are not found beyond the Assam region (Sharma *et al.* 1971). The sticky/non-sticky rice geographical

division is clearly caused by cultural food preference, which correlates with a broader set of deep historical differences in cooking (Fuller and Rowlands 2011). The same areas where one finds sticky rice, one will also find sticky maize and millets (Sakamoto 1996). We do not know when rice was made sticky. The aDNA study presented below attempts to find the *Wx* gene in prehistoric rice but unfortunately yielded no results. Thus, because of introgressive hybridisation, many genes such as *rc* (white pericarp) and *badh2.1* (fragrance) that have been demonstrated to have a single origin can be found in both *japonica* and *indica* (Kovach *et al.* 2009). The aDNA study presented here illustrates the presence of shared genes in archaeological, modern domesticated and modern wild rice which points towards introgression occurrences in the history of rice.

Materials and Methods

The present study reports aDNA results which drew on both modern reference samples of genetic sequence data and archaeological rice remains. The results of aDNA are considered in relation to grain morphometrics, since these have been more widely reported archaeologically and used to infer likely subspecies affiliation, i.e. *indica* versus *japonica.*

The morphometric analysis involved straightforward measurements of the length, width and thickness of rice grains. The samples measured comprised all the archaeological rice remains used in the aDNA analysis, and additional rice grains from these sites (Tables 1 and S1). Modern comparative rice morphometrics draws upon a substantial reference collection at UCL measured previously for archaeobotanical research (Harvey 2006; Fuller *et al.* 2007), in addition to measurements taken on the 35 modern populations used for comparative genetic data (Tables 2 and S2). Measurements for the archaeological rice remains were made with the low-powered microscope Leica EZ4D and for the modern rice remains with Image J I.45s.

Modern rice accessions were analysed for comparative modern gene sequences, including 20 subspecies *indica* and 15 subspecies *japonica* - (Table S3), and 77 accessions of wild rices (Table S4), including mostly those of the AA genomes (*O. barthii* (n=6), *Oryza glumaepatula* (n=4), *Oryza longistaminata* (n=3), *Oryza merodionalis* (n=17), and *O. rufipogon sensu lato* (n=40), CC genomes (*Oryza officinalis*, n=3), and the BB or BBCC genomes (*Oryza punctata*, n=4). Among *O. rufipogon sensu lato*, are annual types, *O. nivara*, and perennial types (*O. rufipogon sensu stricto*). *O. nivara* accessions, in addition to having an annual ecology usually had chloroplast DNA (cpDNA) deletions at the *orf100* locus (Table S4). These modern rice

samples were classified into their respective varieties by Oka (1958). Of the 20 *indica* accessions, one is a modern cultivar type and 19 are landraces whereas all the *japonica* accessions are landraces. The chloroplast sequences of NC_008155 (*indica*) and NC_001320 (*japonica*) and the nuclear sequences of chromosomes 6 of NC_008399 (*Wx*, DJ6) and DQ280635 (*wx*) and chromosome 7 of NC_008400 (*rc*) and AB247503 (*Rc*) which are registered in the National Center for Biotechnology Information (NCBI) in the US were also compared with the prehistoric rice samples (Table S3, S6 and S7). Seeds of these accessions were provided by the National Institute of Genetics (NIG) in Japan, Genebank, National Institute of Agrobiological Sciences (NIAS) in Japan, and the International Rice Research Institute (IRRI) in the Philippines. These accessions were cultivated in the field of Hirosaki University, Japan, and DNA extraction of subspecies *indica* and *japonica* was carried out at Hirosaki University and DNA of wild species was supplied by NIG.

A total of 211 archaeological whole rice grains and fragments were used in the aDNA study: Balathal (n=40), Ter (n=40), Noen U-Loke (n=20), Ban Non Wat (n=78), Phu Khao Thong (n=15) and Khao Sam Kaeo (n=18). The archaeological rice remains from BNW, KSK and PKT were recovered by flotation by Castillo. The stratigraphic layers from which the rice grains came are considered secure after employing a methodology to assess context security (Castillo 2013). Furthermore, rice grains from BNW, KSK and PKT belonging to these same assemblages used in the aDNA analysis presented here were sent for AMS radiocarbon dating. The results verify their antiquity and place BNW samples in the Late Bronze to Early Iron Age, and the KSK and PKT samples in the Metal Age (Table 3). The samples from NUL were handpicked by the excavators during fieldwork and also came from a secure context. The site has been dated and belongs to the Iron Age, from *ca.* 200-300 BC up to the middle of the first millennium AD (Higham *et al.* 2007). The NUL samples come from a pit in layer 4, which was full of charred rice from which three rice grains were AMS radiocarbon dated corresponding to the Iron Age (Table 3). The samples from the Indian sites of Ter and Balathal were hand collected from rice-rich charred deposits taken from early historic levels at both sites. The general age range at both sites date to 300 BC - 300 AD. Examples of typical archaeological rice grains from these sites are illustrated in Figure 3.

Archaeogenetic methodology

Ancient DNA (aDNA) was extracted using the procedure found in Tanaka *et al.* (2010) with some minor modifications. This section provides the methods employed to amplify the DNA

extracted from archaeological rice remains. Grains from all archaeological sites were processed for aDNA at Research Institute for Humanity and Nature (RIHN), Kyoto in a dedicated aDNA laboratory. In addition, half of each grain from five sites were extracted separately at Hirosaki University (BNW K500 4:2 GEN, BNW V200 7:4 Δ27, Noen-U-Loke, Ter, and Balathal). This gives us confidence that the aDNA results are authentic.

DNA extraction

For DNA extraction of modern rice accessions, seeds were sown on filter paper and were grown at 30ºC in a '16 h light-8 h dark' cycle at light intensity 46.5 $\mu Ms^{-1}m^{-2}$. Ten-day-old seedlings were individually ground in liquid nitrogen, and total DNA was extracted using the procedure of Murray and Thompson (1980) with minor modifications.

To prevent DNA contamination by modern rice, extraction and analysis of ancient DNA was done in the Ancient DNA laboratory room in the Research Institute for Humanity and Nature, Kyoto, Japan. After removing fine debris on the grains with tweezers, the grains were sterilized in 1.0 % NaClO for one min, cleaned with Ultrapure Water for Molecular Biology (EMD Millipore Co., USA) and dried. The aDNA was extracted using the procedure of Mutou *et al.* (2014) with minor modifications, purified with an Illustra MicroSpin G-25 Column (GE Healthcare, USA) and modified with a PreCR Repair Mix enzyme cocktail (New England Biolabs, USA). To check DNA contamination, a sample without seeds in ultrapure water was used as a negative control for DNA extraction in the aDNA analysis.

Primer design

A total of ten markers of nine regions were analysed, including four markers from the chloroplast genome and six from the nuclear genome (Table 4, Figure 4). Four chloroplast markers included one insertion and deletion (In/Dels) in each of the intergenic regions *orf100* and *petN-trnC* (I-32 region), one simple sequence repeat (SSR) in *rpl14-rpl16*, and one set of neighbouring single nucleotide polymorphisms (SNPs) in the gene region of *rpl16* to distinguish the chloroplast genome types of modern *indica* and *japonica* subspecies (Nakamura *et al.* 1997; Takahashi *et al.* 2008; Tang *et al.* 2004). Six nuclear markers included SNPs in each of the gene regions *qSh1* and *Sh4*, In/Del in *Waxy* and *Rc*, In/Del in the intron2 of *Acp1* (purple acid phosphatase, Fukuda *et al.* 2001), and one set of In/Dels of 4 bp and 221 bp in the intergenic region of rice chromosome number 6 (DJ6 region). DJ6 region was designated to dominantly classify tropical and temperate *japonica* (Hanamori *et al.* 2011). The basic helix-

loop-helix (bHLH) motif in the *Rc* gene (Sweeney *et al*., 2007; Furukawa *et al*., 2007) is used to distinguish between white/brown and red rice pericarp. The waxy gene in rice, *Waxy,* encodes a granule-bound starch synthase (Wanchana *et al*. 2003) that distinguishes between glutinous and non-glutinous rice. Two genes relate to the key domestication trait of non-shattering, including *sh4* which is regarded as universal in domesticated rice and which was selected early in the domestication process (Li *et al.* 2006; Zhang *et al.* 2009; cf. Ishikawa *et al.* 2010), and *qSH1* which is found in a smaller subset of temperate *japonica* (Konishi *et al.* 2006). Of these 10 markers, primer sets A and B in the study of Nakamura *et al.* (1997) and 2F and 2R in the study of Tanaka *et al.* (2010) were used to analyse modern DNA and aDNA in the *rpl14-rpl16* region, respectively. Glu-23F and Glu-23R in the study of Wanchana *et al*. (2003) and primers F1, R1 and R2 in the study of Hanamori *et al.* (2011) were used in both DNA analyses in the *Waxy* and DJ6 regions, respectively. The primer set in regions *petN-trnC* and *Rc* were used in the study by Tanaka *et al.* (submitting). For the remaining five markers, five primer sets were developed using Primer 3 (Untergasser *et al.* 2012) from rice chloroplast genome sequences, NC_001320 (*japonica*) and JN861109 (*indica*), and of chromosome number 1, NC_008394 (*qsh1*) and EU999846 (*qSH1*); chromosome number 4, NC_008397 (*sh4*) and EU999926 (*Sh4*); chromosome number 6, NC_008399 (*Wx*) and DQ280650 (*wx*); and chromosome number 12, NC_008405 (*japonica*) and AAAA02035407 (*indica*) in the National Centre for Biotechnology Information (NCBI, USA) and the DNA Data Bank of Japan (DDBJ, Japan). To recognise the 69 bp insertion in the *orf100* region and In/Dels of 4 bp and 221 bp in the DJ6 region, *orf100* R1 primer and DJ6 R2 primer were constructed in the insertion region of 69 bp and 221 bp, respectively, and were amplified with *orf100* F1 primer and Ch6 F1 primer, respectively. All of the markers used in the aDNA analysis are reasonably short, with expected products between 70-200 bp, within the range that is expected to preserve in ancient DNA based on previous studies (e.g. Palmer *et al.* 2009; 2011).

DNA analysis

The aDNA was amplified using the first PCR product as a template. The same primer set was used in the first and second PCR. PCR amplification was carried out twice on all ancient DNA extracts. In addition, a negative control amplification was carried out to assess possible contamination. Amplification, except for regions *qSH1, sh4* and *Waxy,* was done in a 20μl mixture including 2.0 μl aDNA, 1×Ex*Taq*^{TM} buffer (10 mM Tris-HCl [pH 8.3], 50 mM KCl), 0.25 U Ex*Taq*^{TM} polymerase (TAKARA, Japan), 0.1 mM dNTPs, 2.0 mM $MgCl_2$ and 0.25 μM of each primer by using an Mastercycler Ep Gradient (Eppendorf, Germany). Initial denaturing

was at 95°C for 3 min, 35 PCR cycles at 95°C for 30 min, annealing for 30 min, and 72°C for 30 min, and then a final extension was done at 72°C for 3 min. Annealing was done at the temperature indicated in Table 4. Amplification of regions *qSH1*, *sh4* and *Waxy*, which are GC rich regions, was done in a 20μl mixture including 2.0 μl aDNA, 2×GC buffer I, 0.50 U LA*Taq*^{TM} polymerase (TAKARA, Japan), 0.1 mM dNTPs and 0.25 μM of each primer by using a Mastercycler Ep Gradient (Eppendorf, Germany). The PCR reaction was basically the same as the analysis above, except for the extension reaction of region *Waxy* which was done at 72°C for 1 min. The PCR product was electrophoresed on 3.0 % agarose gel (NuSieve® 3:1 agarose gel, CAMBREX, USA) at a constant voltage of 100V (Mupid-2, Cosmo Bio, Japan) and visualized with ethidium bromide staining. The experiment was replicated two times. To confirm the sequence of the amplicon in rice remains, the amplicon was resequenced using the specific primer set.

Modern DNA analyses were performed for regions *orf100, petN-trnC, rpl14-rpl16, Waxy,* DJ6, *Rc* and *Acp1*. The PCR amplification and electrophoresis were basically the same as for aDNA analysis, except in region *rpl14-rpl16*. For the analysis of *rpl14-rpl16* including SNPs of *rpl16*, PCR amplification and electrophoresis was done according to Nakamura *et al.* (1997), and the amplicon underwent direct sequencing.

Sequence analysis of aDNA

For direct sequencing in the 10 regions, the aDNA fragments which were similar in size to modern rice, were excised from the agarose gel and were purified with a WizardR SV Gel and PCR Clean-UP System (Promega, USA). Sequencing reaction was run with a BigDye^{TM}Cycle Sequence Ready Reaction DNA (Applied Biosystems, USA) and a specific primer for each region. The nucleotide sequence was determined by an ABI PRISM^{R} 3730 DNA Analyzer (Applied Biosystems, USA). This was done for both strands except in the PS-ID region, whose sequence for the reverse strand could not be determined due to two SSRs located near the annealing site of a reverse primer.

Results and Discussion

Ancient DNA recovery and sequence identification

DNA amplification remains an issue for ancient DNA studies because of preservation variables. DNA degradation is especially high in hot and wet climates, or in acidic soils. DNA survives best in cool, dry, dark, anaerobic and slightly alkaline conditions (Bollongino *et al.* 2008; Schlumbaum *et al.* 2008). Desiccated plant materials make excellent samples (Palmer *et al.* 2009; Bunning *et al.* 2012). On the other hand, it has been observed that charred remains are not conducive to archaeogenetics, success depending on the extent of charring (Palmer *et al.* 2011; Wales et al 2014). All the specimens studied here were charred, and success implies that small pockets on uncharred grain tissues were preserved, and within these pockets fragmented aDNA remained. Although the ancient charring conditions for the archaeological rice remains are unknown, success rates might reflect different charring conditions, with temperatures less than 200°C under anaerobic conditions being preferable (Threadgold and Brown 2003).

The sample sizes of the rice caryopses and fragments sent for aDNA analysis are found in Table 5, with an indication of the recovery of each marker for each site. Extraction was attempted from 211 grains, of which 82 (~39%) produced no DNA products. This means that the majority of grains (~61%) produced at least one recognisable marker. Most remarkably, all 20 grains from NUL produced some aDNA results, but there is no single marker recovered from every grain. In general, chloroplast DNA results were more often recovered than nuclear markers (Figure 5). This is expected since chloroplast DNA has a greater number of genome copies compared to nDNA. Extraction of aDNA was variable and the samples from sites located in the south of Thailand, KSK and PKT, had much lower success rates than those from the northeast, BNW and NUL. Indian samples generally produced similar recovery rates to the southern Thai sites, although southern Ter had a lower success rate than that of Balathal, which is from a drier northwestern location. This could mean that more southerly, tropical conditions reduced aDNA preservation. Using the naked eye, the samples from BNW and NUL were well preserved, showing an intact pericarp surrounding the whole grain whereas the whole caryopses from PKT and KSK had heavily degraded pericarps (Figure 3).

Fragmentation may be another factor affecting preservation. The two samples with the lowest success rates were composed fully or partially of grain fragments. DNA has higher chances of preservation if it remains protected by a sturdy or lignified exocarp (Schlumbaum *et al.* 2008). Therefore, a ruptured pericarp, the protective bran layer of a rice caryopsis, may result in lower

degrees of DNA preservation in the cellular chambers. The fracture would allow for more degradation caused by carbonisation or environmental factors.

In the case of NUL, the preservation of the rice grains may have been good because they were concentrated in a pit and when burning occurred, the surface grains probably smouldered while lower grains were toasted. NUL#105 belongs to the Iron Age 3 (IA3) period dating to *ca.* AD 200-400 (Higham 2011). Graves and pits of this period at NUL were sometimes lined and sealed in clay (*ibid.*). It is not clear how these rice grains preserved and although unconfirmed, perhaps the pit they came from was similarly lined with clay keeping the prehistoric rice grains cool, dry and dark after charring. At BNW, it is possible that the archaeological rice remains were preserved in close to anaerobic and permanently wet conditions sealed by surrounding and overlying hard floors. BNW V200 7:3 Δ2 constitutes the fill of a narrow trench cut into one of the many hard floors that overlie each other in this excavation unit. These hard floors may relate to industrial activity associated with iron working and/or salt production (Duke *et al.* 2010). BNW V200 7:4 Δ27 appears to be one of a series of post-holes in the base of BNW V200 7:3 Δ2. Together, these may represent a wall or palisade structure associated with the floors. BNW K500 4:2 GEN is the top 10 cm spit (arbitrary excavation unit) of organic muddy sediment that fills a wide cut into the natural clays at what was likely the NE edge of the mounded site at the time. BNW K500 4:5 GEN lies deeper in this same wide-cut feature. This may have been one of the first moats around the site that was subsequently filled in as the site expanded laterally through the Iron Age. The flat base on this feature led excavators to hypothesise that it may have been an early rice padi cut at the edge of the site. However, analysis of plant remains from the Bronze to early Iron Age suggests dryland systems of cultivation at BNW (Castillo 2011; 2013).

By summarizing sequence polymorphisms of four chloroplast genome markers, eight chloroplast genome types were observed in 35 modern rice cultivars (Table S5). *Japonica* and *indica* chloroplast types were distinguished by combining these markers. In the aDNA analysis, the *indica*-specific chloroplast sequence was only recognized in six samples from India, with two from Balathal producing the *orf100* deletion while four from Ter had the *petN-trnC* non-deletion sequence (Table S6). Indicators of the *japonica* chloroplast sequence were detected in rice grains from all four sites in Thailand and in India from fourteen grains at Balathal and seven grains at Ter (Figure 6). The most successful marker for detecting *japonica* cpDNA

across the whole dataset was the *rpl14-rpl16* (in the PS-ID region), but no cases of *indica* sequence were detected with this marker.

As already noted, the nuclear markers targeted were on the whole less often recovered than cpDNA markers (Figures 5 and 7). Despite our attempt to extract markers of interest to the cultural history of rice, including those involved with non-shattering (*sh4, qsh1*), success rates were too low to provide informative results. Where detected, the *sh4* SNP was identical to modern domesticated types (Figure S1), implying the presence of non-shattering rice at both Indian and Thai sites. Indeed, archaeological spikelet bases from these sites indicate a high dominance of non-shattering abscission scar morphology (Table 6) and demonstrates domestication status. Mutant-types of *qsh1* confer a much higher degree of non-shattering, and are today restricted to a subset of *japonica* rice, primarily in temperate *japonica* (Konishi *et al.* 2006; Zhang *et al.* 2009). It is unclear how early this mutation arose or became fixed in the non-shattering homozygous condition in rice populations. This allele was recovered from seven ancient grains from BNW and NUL, the sites with the best aDNA recovery overall. Six of these produced the functional (abscission-forming) *qsh1* allele found in wild rice and across a wide cross-section of cultivated *indica* and *japonica* (Figure S2).

When present, monomorphic sequences were recognized in the amplicons of, *Waxy* and *Rc* genes, these indicate the typical non-sticky (wild type) starch and white pericarp, respectively (Figures S3, S4). White pericarp was detected in one from seven grains from BNW and one from NUL (Table 5). Given that all wild rices have a red pericarp *Rc* and that most white pericarp rices share an identical recessive mutation that leads to a loss of pericarp pigmentation (*rc*) (Sweeney *et al.* 2007), these data are consistent with the evolution of white pericarp rice prior to the dispersal of *japonica* from China into mainland Southeast Asia. In the case of non-glutinous (*Wx*) rice, this was detected in only one grain from KSK and five from BNW (Table 5). Given that most East and Southeast Asian glutinous rices share a mutation at this locus (*wx*), it has similarly been postulated that this form had a single origin, perhaps in northern Southeast Asia (Olsen and Purugganan 2002) or in central China, alongside sticky millet types (Fuller and Castillo, in press). The absence in the material in this study would be consistent with such types having been rare in earlier prehistory or having spread later than the initial dispersal of domesticated, white-grain rices in mainland Southeast Asia.

By contrast nuclear genome sequences from chromosomes 6 and 7 indicated the presence of some heterozygosity in the remains. DJ6 was one of the most successfully identified aDNA markers in the study overall. But *Acp1* was much less successfully identified (Figure 7; Table 5). In both markers a few alternative sequences were recovered, such as the 98 bp and 315 bp types in the DJ6 region. The 315 bp type was recognised as a 102 bp fragment which was part of the 315 bp fragment in the analysis of the DJ6 region using F1 and R2 primers in modern and ancient DNA (Tables S3, S6); and the 122 bp type and 146 bp type in *Acp1*. While none of these is exclusive to any wild rice group, in modern domesticated subspecies *indica*, DJ6 315 bp type appears to be exclusive and *Acp1* 122 bp type is common. Also modern *japonica* *Acp1* 146 bp type is most common, and DJ6 315 and -98 are both common (Tables 7-8). In DJ6, both of the recurrent markers occur among Indian and Thai archaeological rice samples, and heterozygous grains were detected at KSK, BNW, and NUL (Table 7). The chromosome 7 marker, *Acp1*, indicated two alternative markers present in archaeological samples from Thailand, with both (146, 122) present at BNW (Table 8). In the few instances in which both *Acp1* and DJ6 could be amplified from the same grains, three different nuclear genome combinations were detected, including *Acp1* 122/ DJ6 98 and *Acp1* 146/ DJ6 98 at BNW; and *Acp1* 146/ DJ6 315 at PKT (Table 9). While this indicates genetic diversity within ancient Thai rices, the data are too limited to offer much in the way of historical interpretation of these patterns. Nevertheless, the *indica* associated combination (*Acp1* 122/ DJ6 315) is absent from the Thai samples. The combinations present are all associated with *japonica* rice, including two temperate *japonica* associated genotypes at BNW, while others could all represent tropical *japonica* genomes. Of note is the presence of DJ6 98 in two grains from BNW. As in our modern samples, this type is only known from perennial wild rice (*O. rufipogon* with *orf100* =162) from Laos and New Guinea. This suggests that this allele had been transferred by introgression from wild *O. rufipogon* in Southeast Asia after domesticated *japonica* rice was introduced.

The sequence analysis revealed that the same sequences found in modern *japonica* and *indica*, and some wild populations in the chloroplast and nuclear genome, are also detected in archaeological rice remains. Taking the results as a whole, it is clear that a significant proportion (58%) share sequences with modern domesticated *japonica* populations, and in only a few cases (2.8%) from Indian archaeological sites for *indica* specific sequences detected (Figure 6; Table 5).

Morphometric analysis in relation to the aDNA study

Until the application of aDNA extraction becomes more widespread, it is worth considering how our aDNA results compare to those of conventional grain morphometrics, as these have been widely applied in the past to archaeobotanical remains to distinguish the presence of *indica* or *japonica* (e.g. Castillo 2011, 2013; Fuller *et al.* 2007; Oka 1988). The use of morphometrics as an identification ratio does not work to discern whether rice is domesticated or wild because of the wide variation in the size and proportions of wild rice Fuller *et al.* (2007, 2008). This problem is exacerbated by the inclusion of immature grain measurements. Immature rice is expected to have been common during the early stages of rice domestication (see also, Fuller *et al.* 2009). However, if the domestication status of rice is already established, then morphometric studies can provide some useful indications to whether the archaeological assemblage lean towards the *japonica* or *indica* tendencies. In this study, the length-width (L/W) ratios of the archaeological rice grains were compared with those of modern populations of domesticated and of wild rice. According to Ahn (1993), L/W ratios are not affected by charring so ancient and modern rice should therefore be comparable.

In past studies, it has been shown that *indica* rice normally has a L/W ratio of >2.5 whereas *japonica* rice is <2.3 (Fuller *et al.* 2007). In the present study, this ratio has been recalculated using the latest IRRI data, used to reassess the *indica*/*japonica* assignments of modern populations. The L/W ratios of >2.2 indicate *indica*-type rice and ratios <2.0 are *japonica*-type (Figure 8). However, it must be noted that such tendencies are complicated and by no means absolute. Some variation in rice grain length is environmental or climatic (Kitano *et al.* 1993; Oka 1988). More northerly temperate *japonica* landraces are short-grained, whilst tropical varieties (the *javanica* race rices) are very long. Complicating this trend is that of longer-grained upland rices versus shorter-grained lowland rices in East Asia generally (Nitsuma 1993), whilst some varieties at high elevations in the tropical mountains (e.g. Yunnan, Nepal) may be very short-grained. This is illustrated in Figure 8 where the L/W ratios of *indica* and *japonica* overlap mostly in the 2.0-2.4 L/W ratio. Nevertheless, it remains the case that most *indica* are longer and thinner, whereas many *japonica* are shorter and plump. Accepting that there is no definitive division we have nevertheless applied ratios as a guide to more likely subspecies assignment.

Since the domestication status of rice in four of the sites was established through the analysis of spikelet bases (Table 6), morphometric analysis was then used to indicate to which domesticated subspecies the archaeological rice approximates. In the case of the rice samples coming from three Thai and one Indian archaeological site, the rice spikelet bases were of the domesticated-type as stated earlier. The samples from NUL and Balathal did not contain spikelet bases. The morphometric analyses of rice grains from the four Thai archaeological sites dating to the Bronze and Iron Ages (BNW, KSK, NUL and PKT) suggest that rice in prehistoric Thailand was *O. sativa japonica;* whereas the morphometrics from the two Indian sites dating to the historic period show a mixed ratio distributions of both *indica* and *japonica* (Figures 9 and 10).

Conclusions

We have successfully extracted ancient DNA from charred archaeological rice grains, both from India and Thailand. In particular, the recovery of cpDNA has been very successful in distinguishing archaeological rices affiliated with subspecies *japonica* or subspecies *indica*. In addition, recovery of selected nuclear markers, highlight the potential to look more broadly at rice genetic diversity over time and potentially to reconstruct histories of hybridisation between lineages. Nuclear markers also hold the potential to assess directly the presence or absence of key functional alleles, for agronomic traits or domestication traits, such as seed shattering (*qsh1*, *sh4*), seed width (*qsw5*), seed pericarp colour (*rc*), waxy (*wx*) and plant height (*sd1*) (Konishi *et al*.2006; Lin *et al*. 2007; Li *et al.* 2007; Furukawa *et al*. 2007; Sweeney *et al*. 2006; Wanchana *et al*. 2003; Ashikari *et al*. 2005; Nagano *et al*. 2005). Although the results from such markers in this study have been too limited in the current material to contribute new insights into the history of selection in rice, they nevertheless fit with the present view (Kovach *et al.* 2007) of an early evolution of white pericarp, probably prior to the dispersal of domesticated *japonic*a from China to Southeast Asia, and perhaps a later evolution of sticky (waxy) rices.

Both aDNA and morphometric studies are able to differentiate the early historic Indian and prehistoric Thai rice populations studied. In the case of the results of the Indian samples there is some overlap with those of Thailand and the expected morphometrics of typical *japonica*, but many other grains fit our expected typical *indica* length:width ratios. The results of ancient DNA on these Indian populations indicate a mixture of *japonica* and *indica* associated genotypes. This fits with the inferred history of rice in India, in which an early hybridisation

with domesticated *japonica* was important in the improvement and dispersal of *indica* starting perhaps 2000 BC and certainly by *ca.* 1000 BC (Fuller *et al.* 2010).

The morphometric analysis and aDNA study presented above provide a strong case that the rice subspecies *japonica* was being consumed and cultivated in mainland Southeast Asia in the Bronze and Iron Ages. This includes evidence from *ca.* 1050-420 BC and up to *ca.* 200-400 AD in Northeast Thailand and *ca.* 400-100 BC in peninsular Thailand. Similar aDNA studies on Neolithic samples are clearly an imperative, but at this juncture, one can at least infer that rice found in Thailand in earlier periods, if domesticated, would also have been *japonica* derived from domestication in the Yangtze basin. There is no indication that early rice in Thailand was *indica,* which means that *indica* must have been brought into Thailand only after the initial period of Indian contact. KSK and PKT belong to this early period of Indian contact and both sites have *japonica*-type rice veering toward the *japonica* end of the spectrum as established by the morphometric analysis. The presence of awns has also been identified in rice plant parts from BNW, KSK and PKT (Castillo 2013). Although one could assume the rice to contain some wild stands, it may also signify an awned variety of domesticated rice, namely tropical *japonica* (*javanica* or *bulu*) which is often cultivated in rainfed systems today. Today *indica* is the dominant rice type and rainfed cultivation is the main agricultural system practiced in Thailand. The absence of the *indica* genotype from the studied remains fits with the hypothesis that the establishment of *indica* rices in mainland Southeast Asia took place only within the past 2000 years (Castillo and Fuller 2010), at a time still to be determined empirically.

This study highlights the potential for more archaeogenetic work on ancient rices. aDNA studies allow both archaeologists and geneticists working on modern material to prove or disprove theories on origins. For archaeologists, ancient DNA studies are needed in order to confirm hypotheses postulated that cannot be directly assessed from preserved grain or spikelet base morphology. For hypotheses derived from modern genetics, aDNA can confirm the antiquity of particular genotypes in particular places, and directly test molecular clock estimates of divergence times (e.g. Molina *et al.* 2011). This study shows that the aDNA study conducted on prehistoric rice in Thailand supports the morphometric analysis confirming the existence of *japonica* rice but not *indica* from the Bronze Age to the Iron Age in the northeast and southern Thailand. On the other hand, both *japonica* and *indica* co- existed in India during the early Historic period, as would be expected in light of the hypothesis for a prehistoric

introduction of *japonica* to India and its hybridisation with '*proto-indica*'. The results from this study endorse our understanding of the spread of rice cultivation in Southeast Asia, where *japonica* preceded *indica*, and *indica* was a late arrival in Thailand, possibly during the historic period. The two sites with early contact with India are KSK and PKT, which yielded ample evidence of Indian domesticates, in particular the pulses mungbean (*Vigna radiata*) and horsegram (*Macrotyloma uniflorum*) [Castillo and Fuller 2010; Castillo 2013]. Interestingly, we found no evidence for *indica* rice from these sites, even though it would have been available from the Indian source region of the introduced pulses, as indicated by our aDNA and morphometric results from Ter. We therefore conclude that the Indian subspecies *indica* arrived in Thailand at a later period in the first centuries AD, and it may be that this relates to the establishment of more intensive forms of wet rice cultivation.

Acknowledgements:

This work was supported by the Sato Project of the Research Institute for Humanity and Nature (RIHN), Japan and the Special Research of Hirosaki University entitled 'Research Project for Promoting the Utilization of Archaeological Remains'. This is contribution number XXII from the Sato Project of the Research Institute for Humanity and Nature (RIHN), Japan. We would like to thank Robin Allaby, Diane Lister and Mark Thomas for their comments and Dr. Nori Kurata from the National Institute of Genetics, Japan, for kindly supplying seeds of modern rice and DNA of wild rice. The Thai-French archaeological mission in the Upper Thai-Malay peninsula was funded by the 'commission consultative des fouilles' of the French Ministry of Foreign Affairs. Further funding was obtained from the Institute of Archaeology, UCL and the Royal Thai Embassy in London for archaeobotanical fieldwork in Thailand. Radiocarbon dating, archaeobotanical analysis and writing by CCC and DQF was supported by a grant from Natural Environment Research Council (NE/K003402/1) on “the impact of evolving rice systems from China to Southeast Asia.”

BIBLIOGRAPHY

Ahn, S-M. 1993. *Origin and differentiation of domesticated rice in Asia.* Unpublished PhD dissertation, Institute of Archaeology, University of London.

Ashikari, M., Sakakibara, H., Lin, S., Yamamoto, T., Takashi, T., Nishimura, A., Angeles, E.R., Qian, Q., Kitano, H. and Matsuoka, M. 2005. Cytokinin oxidase regulates rice grain production. *Science* 309: 741-5.

Bellina, B., Silapanth, P., Chaisuwan, B., Thongcharoenchaikit, C., Allen, J., Bernard, V., Borell, B., Bouvet, P., Castillo, C., Dussubieux, L., Malakie, J., Srikanlaya, S., Peronnet, S. and Pryce, T.O. 2014. The development of coastal polities in the Upper Thai-Malay Peninsula, in *Before Siam: Essays in Art and Archaeology*. Revire, N. and Murphy, S.A. (eds), 69-89.

Bollongino, R., Tresset, A. and Vigne, J.-D. 2008. Environment and excavation: Prelab impacts on ancient DNA analyses. *Comptes Rendus Palevol* 7: 91-8.

Boyd, W. and Chang, N.J. 2010. Integrating social and environmental change in prehistory: a discussion of the role of landscape as a heuristic in defining prehistoric possibilities in northeast Thailand, in *Altered Ecologies: fire, climate and human influence on terrestrial landscapes*. Terra Australis, 32. ANU E Press, Canberra, ACT, Australia, 273-97.

Bunning, S.L., Jones, G. and Brown, T.A. 2012. Next generation sequencing of DNA in 3300-year-old charred cereal grains. *Journal of Archaeological Science* 39: 2780-4.

Casson, L. 1989. *The Periplus Maris Erythraei: Text with introduction, translation, and commentary*. Princeton: Princeton University Press.

Castillo, C. 2011. Rice in Thailand: the archaeobotanical contribution. *Rice* 4(3-4): 114-20.

Castillo, C. 2013. *The archaeobotany of Khao Sam Kaeo and Phu Khao Thong: The agriculture of late prehistoric Southern Thailand.* Unpublished PhD thesis, Institute of Archaeology, University College London.

Castillo, C. and D.Q. Fuller. 2010. Still too fragmentary and dependent upon chance? Advances in the study of early Southeast Asian archaeobotany, in *50 years of archaeology in Southeast Asia: Essays in honour of Ian Glover*, Bellina-Pryce, B., Pryce, T.O., Bacus, E. and Wisseman-Christie, J. (eds). River Books, Bangkok, 90-111.

Chen, W., Nakamura, I., Sato, Y. and Nakai, H. 1993. Distribution of deletion type in cpDNA of cultivated and wild rice. *Jpn J Genet* 68:597–603.

Cheng, C., Motohashi, R., Tsuchimoto, S., Fukuta, Y., Ohtsubo, H. and Ohtsubo, E. 2003. Polyphyletic origin of cultivated rice: based on the interspersion pattern of SINEs. *Mol. Biol. Evol*. 20: 67-75.

Duke, B.J., Carter, A.K., and Chang, N.J. 2010. *The excavation of iron age working floors and small-scale industry at Ban Non Wat, Thailand. Papers from the Institute of Archaeology,* 20: 123-30.

Fukuda, T., Osaki, M., Shinano, T. and Wasaki, J. 2001. Cloning and characterization of two secreted acid phosphatases from rice calli. *Plant Nutrition* 92: 34-5.

Fuller, D.Q. 2011a. Pathways to Asian civilizations: Tracing the origins and spread of rice and rice cultures. *Rice* 4(3-4): 78-92.

Fuller, D.Q. 2011b. Finding plant domestication in the Indian subcontinent. *Current Anthropology* 52: S347-62.

Fuller, D.Q. and Allaby, R., 2009. Seed dispersal and crop domestication: Shattering, germination and seasonality in evolution under cultivation. *Annual Plant Reviews* 38: 238-95.

Fuller, D.Q. and Castillo, C. (in press). Diversification and cultural construction of a

crop: the case of glutinous rice and waxy cereals in the food cultures of eastern Asia, in *The Oxford Handbook of the Archaeology of Food and Diet*. Lee-Thorp, J. (ed.). Oxford University Press.

Fuller, D.Q. and Harvey, E.L. 2006. The archaeobotany of Indian pulses: identification, processing and evidence for cultivation. *Environmental Archaeology* 11(2): 219-46.

Fuller, D.Q. and Qin, L. 2009. Water management and labour in the origins and dispersal of Asian rice. *World Archaeology* 41(1): 88-111.

Fuller, D.Q., Sato, Y.-I., Castillo, C., Qin, L., Weisskopf, A.R., Kingwell-Banham, E.J., Song, J., Ahn, S-M., van Etten, J. 2010. Consilience of genetics and archaeobotany in the entangled history of rice. *Archaeological and Anthropological Science* 2: 115–31.

Fuller, D.Q. and Rowlands, M. 2011. Ingestion and food technologies: Maintaining differences over the long-term in West, South and East Asia, in *Interweaving worlds - systematic interactions in Eurasia, 7th to 1st millennia BC. Essays from a conference in memory of Professor Andrew Sherratt.* Bennet, J., Sherratt, S. and Wilkinson, T.C. (eds). Oxbow Books Ltd., Oxford, 37-60.

Fuller, D.Q. and Sato, Y.-I. 2008. *Japonica* rice carried to, not from, Southeast Asia. *Nature Genetics* 40: 1264-5.

Fuller, D.Q., Harvey, E. and Qin, L. 2007. Presumed domestication? Evidence for wild rice cultivation and domestication in the fifth millennium BC of the Lower Yangtze region. *Antiquity* 81: 316-31.

Fuller, D.Q., Qin, L. and Harvey, E. 2008. Evidence for a late onset of agriculture in the Lower Yangzi region and challenges for an archaeobotany of rice, in *Past Human Migrations in East Asia. Matching Archaeology, Linguistics and Genetics.* Sanchez-Mazas, A., Blench, R., Ross, M.D., Peiros, I. and Lin, M. (eds). London: Routledge, 40-83.

Fuller, D.Q, Qin, L., Zheng, Y., Zhao, Z., Chen, X., Hosoya, L.A. and Sun, G. 2009. The Domestication Process and Domestication Rate in Rice: Spikelet bases from the Lower Yangtze. *Science* 323: 1607-10.

Fuller, D.Q., Denham, T., Arroyo-Kalin, M., Lucas, L., Stevens, C.J., Qin, L., Allaby, R. and Purugganan, M.D. 2014. Convergent Evolution and Parallelism in Plant Domestication Revealed by an Expanding Archaeological Record. *Proceedings of the National Association of Sciences* 111(17): 6147-52.

Furukawa, T., Maekawa, M., Oki, T., Suda, I., Iida, S., Shimada, H., Takamure, I. and Kadowaki, K. 2007. The Rc and Rd genes are involved in proanthocyanidin synthesis 288 in rice pericarp. *Plant Journal* 49: 91–102.

Garris, A.J., Tai, T.H., Coburn, J., Kresovich, S. and McCouch, S.R. 2005. Genetic structure and diversity in *Oryza sativa* L. *Genetics* 169: 1631–8.

Hanamori, K., Ishikawa, S., Saito, H., Tanaka, K., Sato, Y.-I. and Okada, Y. 2011. Making divergent marker of tropical and temperate types of *Oryza sativa* L. var. *japonica* based on

insertion/ deletion DNA region and its utilization of carbonated rice from Toro Ⅰ site. *J Sch Mar Sci Technol Tokai Univ* 9: 19-25 (in Japanese with English abstract).

Harvey, E.L. 2006. *Early agricultural communities in northern and eastern India: an archaeobotanical investigation.* Unpublished PhD dissertation, Institute of Archaeology, UCL.

Higham, C.F.W. 2011. The Iron Age of the Mun Valley, Thailand. *The Antiquaries Journal* 91: 101-44.

Higham, C.F.W. and Kijngam, A., (eds.) 2009. *The Origins of the Civilization of Angkor.* Volume III. *The Excavation Ban Non Wat, Introduction.* Bangkok, The Fine Arts Department of Thailand.

Higham, C.F.W. and Kijngam, A., (eds.) 2011. *The Origins of the Civilization of Angkor.* Volume IV. *The Excavation Ban Non Wat: the Neolithic Occupation.* Bangkok, The Fine Arts Department of Thailand.

Higham, C.F.W. and Kijngam, A., (eds.) 2012a. *The Origins of the Civilization of Angkor.* Volume V. *The Excavation Ban Non Wat: Bronze Age.* Bangkok, The Fine Arts Department of Thailand.

Higham, C.F.W. and Kijngam, A., (eds.) 2012b. *The Origins of the Civilization of Angkor.* Volume V. *The Excavation Ban Non Wat: The Iron Age, Summary and Conclusions.* Bangkok, The Fine Arts Department of Thailand.

Higham, C. and Higham, T. 2009. A new chronological framework for prehistoric Southeast Asia based on a Bayesian model from Ban Non Wat, *Antiquity* 83, 125-44.

Higham, C.F.W., Kijngam, A. and Talbot, S. 2007. *The origins of the civilization of Angkor, vol. 2: The excavation of Noen U-Loke and Non Muang Kao*. The Thai Fine Arts Department, Bangkok.

Ishii, T., Numaguchi, K., Miura, K., Yoshida, K., Thanh, P.T., Htun, T.M., Yamasaki, M., Komeda, N., Matsumoto, T., Terauchi, R., Ishikawa, R. and Ashikari, M. 2013. OsLG1 regulates a closed panicle trait in domesticated rice. *Nature genetics* 45(4): 462-5.

Ishikawa, R., Sato, Y.-I., Tang, T. and Nakamura, I. 2002a. Different maternal origins of Japanese lowland and upland rice populations. *Theoretical and Applied Genetics* 104: 976-80.

Ishikawa, R., Nakamura, I., Nishihara, T., Kikuchi, M., Senda, M., Akada, S., Harada, T. and Niizeki, M. 2002b. Origin of cytoplasm substituted rice cultivars found in Japan. *Theoretical and Applied Genetics* 105: 608-13.

Ishikawa, R., Thanh, P.T., Nimura, N., Htun, T.M., Yamasaki, M. and Ishii, T. 2010. Allelic interaction at seed-shattering loci in the genetic backgrounds of wild and cultivated rice species. *Genes and Genetic Systems* 85: 265-71.

Kajale, M.D. 1975. A Note on grain remains from excavations at Ter (District Osmanabad) in

1975. *Bulletin Deccan College Research Institute*, 35: 62-4.

Kajale, M.D. 1996. Palaeoethnobotanical investigations at Balathal: Preliminary results. *Man Environ*, 21(1): 99-102.

Kanthilatha, N., Boyd, W., Dowell, A., Mann, A., Chang, N.J., Wohlmuth, H., and Parr, J. 2014. Identification of preserved fatty acids in archaeological floor sediments from prehistoric sites at Ban Non Wat and Nong Hua Raet in northeast Thailand using gas chromatography. *Journal of Archaeological Science*, 46: 353-62.

Kato, S., and Kosaka, H. 1930. On the affinity of the cultivated varieties of rice plants, *Oryza sativa* L. *Journal of the Department of Agriculture Kyushu* 2: 241-76.

Kitano, H., Futsuhara, Y. and Satoh, H. 1993. Morphological variations in rice cultivars, in *Science of the rice plant. Volume One. Morphology.* Matsuo, T. and Hoshikawa, K. (eds). Food and Agriculture Policy Research Center, Tokyo, 79–88

Konishi, S., Izawa, T., Lin, S.Y., Ebana, K., Fukuta, Y., Sasaki, T. and Yano, M. 2006. An SNP caused loss of seed shattering during rice domestication. *Science* 312: 1392-6

Kovach, M.J., Sweeney, M.T. and McCouch, S.R., 2007. New insights into the history of rice domestication. *TRENDS in Genetics* 23(11): 578-87.

Kovach, M.J., Calingacion, M.N., Fitzgerald, M.A. and McCouch, S.R., 2009. The origin and evolution of fragrance in rice (*Oryza sativa* L.). *PNAS* 106: 14444-9.

Li, C., Zhou, A. and Sang, T. 2006. Rice domestication by reducing shattering. *Science* 311: 1936-9.

Lin, Z., Griffith, M.E., Li, X., Zhu, Z., Tan, L., Fu, Y., Zhang, W., Wang, X., Xie, D. and Sun, C. 2007. Origin of seed shattering in rice (*Oryza sativa* L.). *Planta* 226: 11-20.

Li, X., Dodson, J., Zhou, X., Zhang, H. and Masutomoto, R. 2007. Early cultivated wheat and broadening of agriculture in Neolithic China. *Holocene* 17: 555–60.

Ma, J. and Bennetzen, J.L. 2004. Rapid recent growth and divergence of rice nuclear genomes. *Proc Natl Acad Sci USA* 101:12404–10.

Meyer, R.S. and Purugganan M.D. 2013. Evolution of crop species: genetics of domestication and diversification. *Nature Reviews Genetics* 14: 840–52.

Misra, V.N., and Mohanty, R.K. 2001. A rare Chalcolithic pottery cache from Balathal, Rajasthan. *Man and Environment*, 26(2): 67–74.

Molina, J., Sikora, M., Garud, N., Flowers, J. M., Rubinstein, S., Reynolds, A., Huang, P., Jackson, S., Schaal, B.A., Bustamante, C.D., Boyko, A.R. and Purugganan, M.D. 2011. Molecular evidence for a single evolutionary origin of domesticated rice. *PNAS* 108(20): 8351-6.

Murray, M. G., and W. F. Thompson. **1980**. Rapid isolation of high molecular weight plant

DNA. *Nucleic acids research* 8.19: 4321-6.

Mutou, C., Tanaka, K., Ishikawa, R. 2014. DNA extraction from rice endosperm (Including a protocol for extraction of DNA from ancient Seed samples), in *Cereal Genomics: Methods and Protocols, Methods in Molecular Biology vol 1099*. Henry, R.J. and Furtado, A. (eds). Humana Press, New York, 7-15.

Nagano H., Onishi K., Ogasawara M. Horiuchi, Y. and Sano, Y. 2005. Genealogy of the "Green Revolution" in rice. *Genes and Genetic Systems* 80(5): 351–6.

Nakamura, I., Urairong, H., Kameya, N., Fukuta, Y., Chitrkon, S. and Sato, Y.-I. 1998. Six different plastid subtypes were found in the *O. sativa–O. rufipogon* complex. *Rice Genet Newslett* 15:80-2.

Nakamura, I., Kameya, N., Kato, K., Yamanaka, S.I., Jomori, H. and Sato, Y.-I. 1997. A proposal for identifying the short ID sequence which addresses the plastid subtype of higher plants. *Breeding Science* 47: 385-8.

Nitsuma, Y. 1993. Upland rice, in *Science of the rice plant, vol. 1. Morphology*. Matsuo, T. and Hoshikawa, K. (eds). Food and Agriculture Policy Research Center, Tokyo, 70–6.

Oka, H.I. 1958. Intervarietal variation and classification of cultivated rice. *Indian J Genet Plant Breed* 18(2): 79-89.

Oka, H. 1988. *Origins of cultivated rice*. Elsevier, Amsterdam.

Olsen, K.M. and Purugganan, M.D. 2002. Molecular evidence on the origin and evolution of glutinous rice. *Genetics* 162: 941-50.

Palmer, S.A., Moore, J.D., Clapham, A.J., Rose, P. and Allaby, R.G. 2009. Archaeogenetic evidence of ancient Nubian barley evolution from six to two-row indicates local adaptation. *PLoS ONE 4*(7): e6301.

Palmer, S.A., Smith, O. and Allaby, R.G., 2011. The blossoming of plant archaeogenetics. Annals of Anatomy 194: 146-56.

Prathepha, P. 2009. The fragrance (fgr) gene in natural populations of wild rice (*Oryza rufipogon* Griff.). *Genetic Resources and Crop Evolution* 56(1): 13-8.

Purugganan, M.D. 2010. The evolution of rice: molecular vignettes on its origins and spread. *Archaeological and Anthropological Sciences* 2: 61-8.

Purugganan, M.D., and Fuller, D.Q. 2009. The nature of selection during plant domestication." *Nature* 457(7231): 843-8.

Sakamoto, S. 1996. Glutinous-endosperm starch food culture specific to eastern and southeastern Asia, in *Redefining nature: Ecology, culture and domestication*. Ellen, R. and Fukui, K. (eds). Berg Publishers, Washington, DC, 215-31.

Sang, T. and S. Ge. 2007. The puzzle of rice domestication. *Journal of Integrative Plant*

Biology 49: 760-8.

Sato, Y.-I. 1996. *DNA-ga Kataru Inasaku Bunmaei Kingento Tenkai [The origins of rice culture inferred from DNA analysis]* (in Japanese). NHK Books, Tokyo.

Schlumbaum, A., Tensen, M. and Jaenicke-Després, V. 2008. Ancient plant DNA in archaeobotany. *Vegetation History and Archaeobotany* 17: 233-44.

Sharma, S.D., Vellanki, J., Hakim, K. and Singh, R. 1971. Primitive and current cultivars of rice in Assam—a rice source of valuable genes. *Current Science* 40: 126-8.

Shinde, V.S., Deshpande, S.S. and Yasuda, Y. 2004. Human responses to Holocene climate changes: A case study of Western India between 5th and 3rd Millennia BC, in *Monsoon and civilization.* Yasua, Y. and Shinde, V.S. (eds). New Delhi: Lustre Press/Roli Books, 383-406.

Sweeney, M.T., Thomson, M.J., Pfeil, B.E. and McCouch, S. 2006. Caught red-handed: Rc encodes a basic helix-loop-helix protein conditioning red pericarp in rice. *Plant Cell* 18: 283-94.

Sweeney, M.T., Thomson, M.J., Cho, Y.G., Park, Y.J., Williamson, S.H., Bustamante, C.D. and McCouch, S.R. 2007. Global dissemination of a single mutation conferring white pericarp in rice. *PLoS Genet* 3: e133.

Takahashi, H., Sato, Y.-I. and Nakamura, I. 2008. Evolutionary analysis for two plastid DNA sequences in cultivated and wild species of *Oryza*. *Breeding Science 58*: 225-33.

Tanaka, K., Honda, T. and Ishikawa, R. 2010. Rice archaeological remains and the possibility of DNA archaeology: examples from Yayoi and Heian periods of Northern Japan. *Archaeological and Anthropological Sciences* 2: 69-78.

Tanaka, K., Kamijo, N., Tabuchi, H., Hanamori, K., Matsuda, R., Suginomori, J., Sato, Y.-I., Udatsu, T. and Ishikawa, R. Submitting. Morphological and molecular genetics of ancient remains and modern rice confirm genetic conservation and erosion in Japan. *Genetic Resources and Crop Evolution.*

Tang, J., Xia, H., Cao, M., Zhang, X., Zeng, W., Hu, S., Tong, W., Wang, J., Wang, J., Yu, J., Yang, H. and Zhu, L. 2004. A comparison of rice chloroplast genomes. *Plant Physiol* 135: 412-20.

Tian, X., Zheng, J., Hu, S. and Yu, J. 2006. The rice mitochondrial genomes and their variations. *Plant Physiology* 140: 401-10.

Thompson, G.B. 1996. *The excavation of Khok Phanom Di, a prehistoric site in central Thailand. Vol. IV: subsistence and environment: the botanical evidence (the biological remains, part II).* The Society of Antiquaries of London, London.

Threadgold, J., and Brown, T.A. 2003. Degradation of DNA in artificially charred wheat seeds. *Journal of Archaeological Science* 30(8): 1067-76.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M. and Rozen, S.G. 2012. Primer3 - new capabilities and interfaces. *Nucleic Acids Res* 40: e115.

Vaughan, D.A., Lu, B.R. and Tomooka, N. 2008. The evolving story of rice evolution. *Plant Science* 174: 394-408.

Vishnu-Mittre, Prakash, U. and Awasthi, N. 1971. Ancient plant economy at Ter, Maharashtra. *Geophytology*, 1(2): 170–7.

Vitte, C., Ishii, T., Lamy, F., Brar, D. and Panaud, O. 2004. Genomic paleontology provides evidence for two distinct origins of Asian rice (*Oryza sativa* L.). *Mol Genet Genomics* 272:504–11.

Wales, N., Andersen, K., Cappellini, E., Ávila-Arcos, M. C., and Gilbert, M. T. P. (2014) Optimization of DNA recovery and amplification from non-carbonized archaeobotanical remains. *Plos One* 9:e86827

Wanchana, S., Toojinda, T., Tragoonrung, S. and Vanavichit, A. 2003. Duplicated coding sequence in the waxy allele of tropical glutinous rice (*Oryza sativa* L.). *Plant Science* 165: 1193–9.

Wan, J. and Ikehashi, H. 1997. Identification of two types of differentiation in cultivated rice (*Oryza sativa* L.) detected by polymorphism of isozymes and hybrid sterility. *Euphytica* 94(2): 151-61.

Wang, C. and Shu, Q. 2007. Fine mapping and candidate gene analysis of purple pericarp gene *Pb* in rice (*Oryza sativa* L.). *Chinese Science Bulletin* 52(22): 3097-104.

Wei, X., Wang, R., Cao, L., Yuan, N., Huang, J., Qiao, W., Zhang, W., Zeng, H. and Yang, H. 2012. Origin of *Oryza sativa* in China inferred by nucleotide polymorphisms of organelle DNA. *PLoS ONE* 7(11): e49546.

Zhang, L.B., Zhu, Q., Wu, Z.Q., Ross-Ibarra, J., Gaut, B.S., Ge, S. and Sang, a.T. 2009. Selection on grain shattering genes and rates of rice domestication. *New Phytologist* 184: 708-20.

Figures:

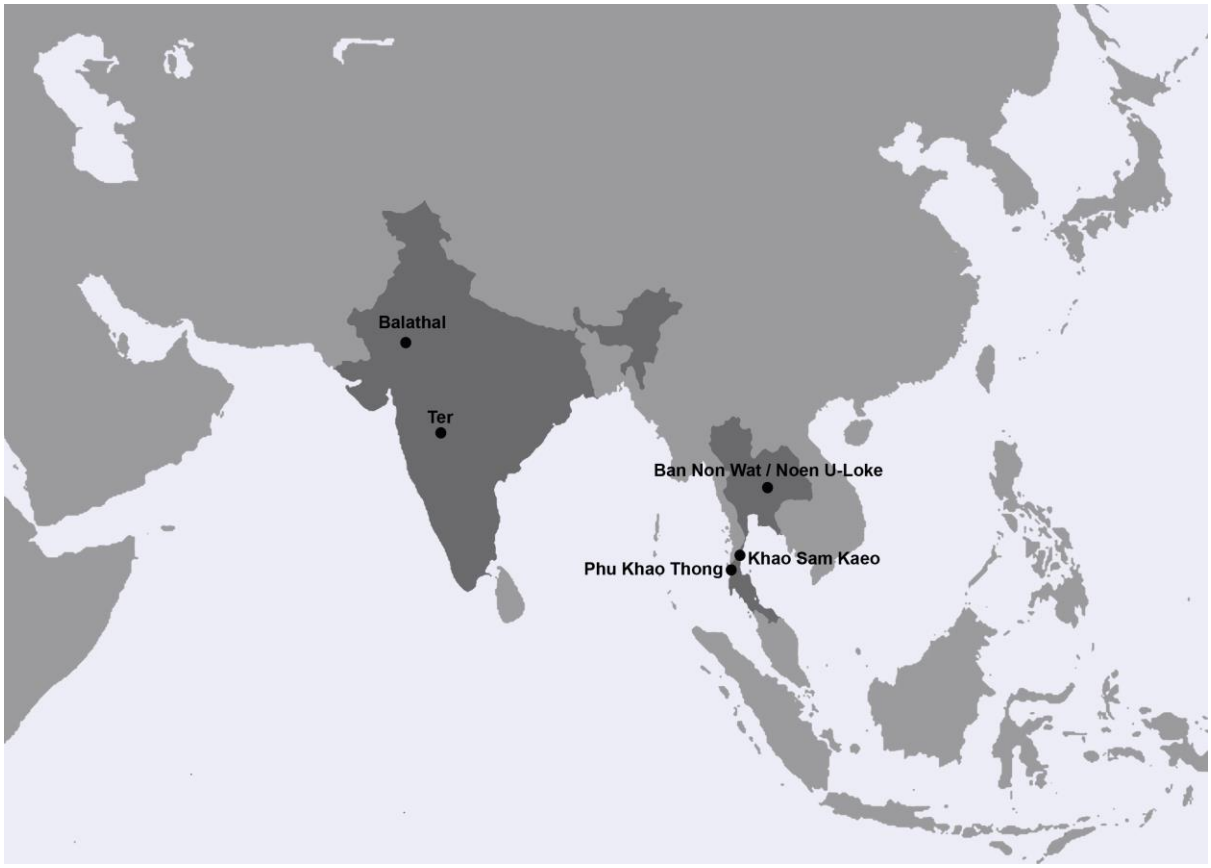

Figure 1. Map showing the location of the six archaeological sites where the rice samples for aDNA were taken from.

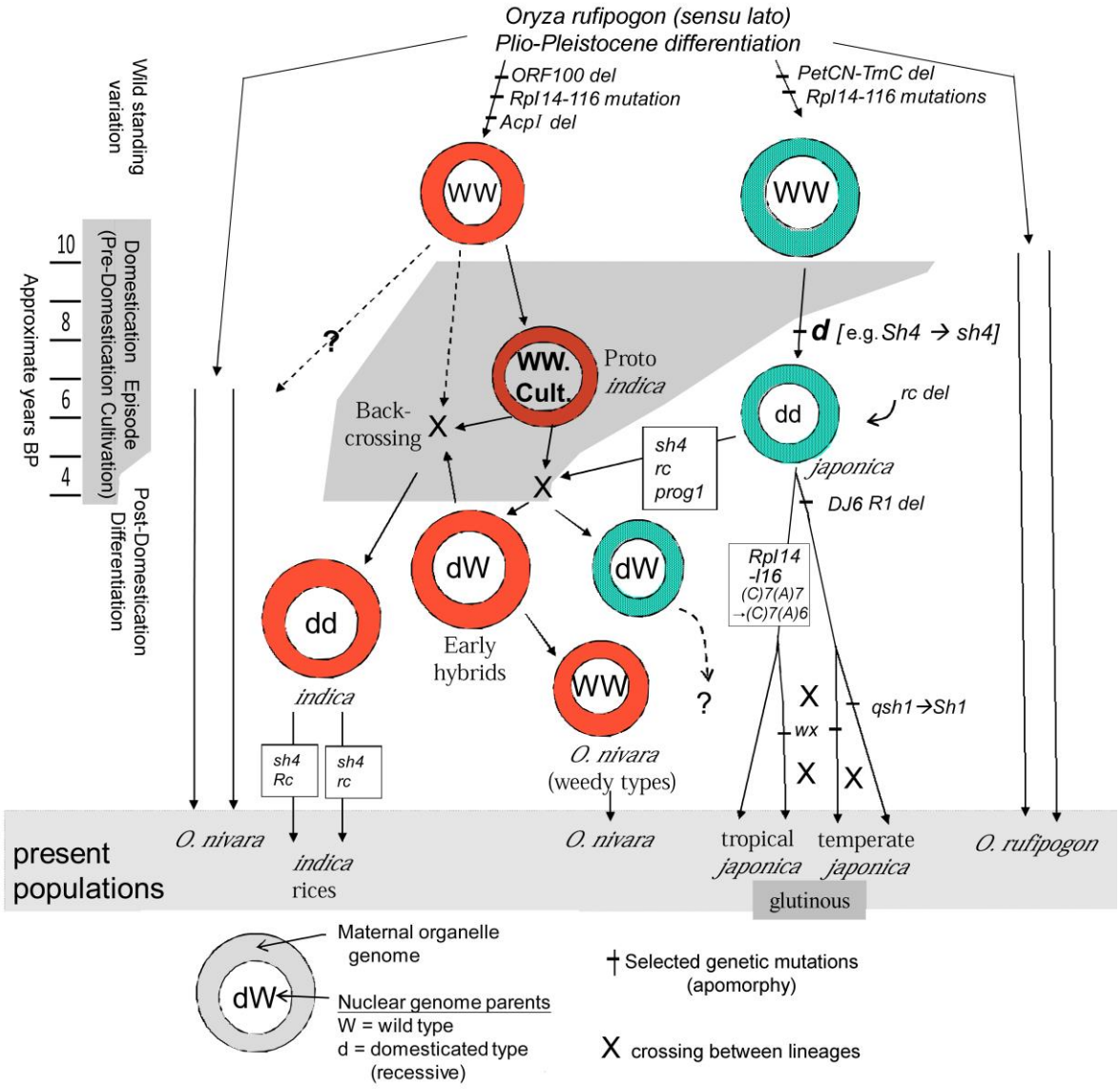

Figure 2. The schematic representation of the evolutionary and domestication history of rice, incorporating the proto-*indica* hypothesis, indicating the approximate phylogenetic position of DNA markers targeted in this study. This is revised from Fuller et al 2010.

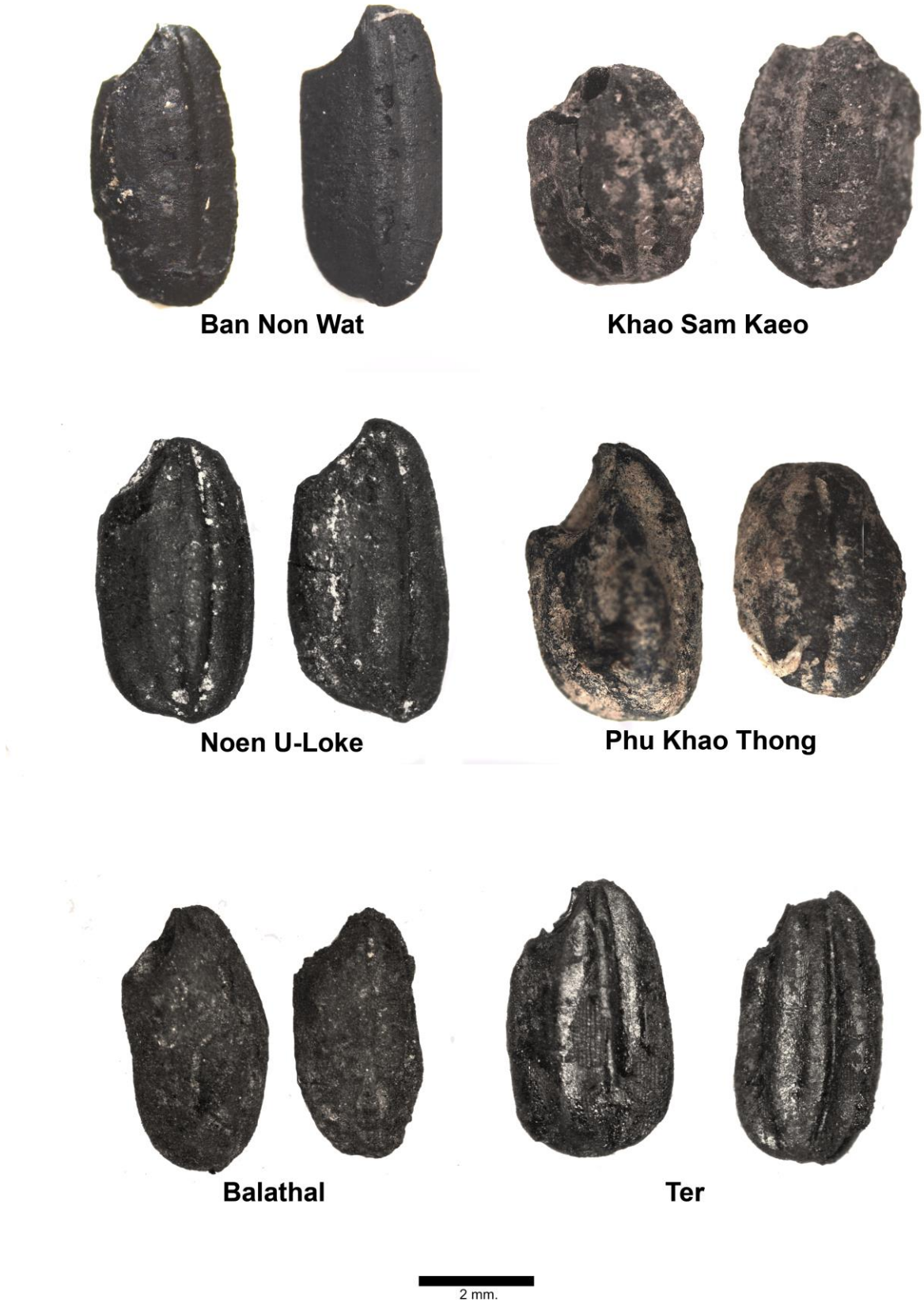

Figure 3. Images of representative rice grains from each of the archaeological populations tested for aDNA and subjected to metrical characterization.

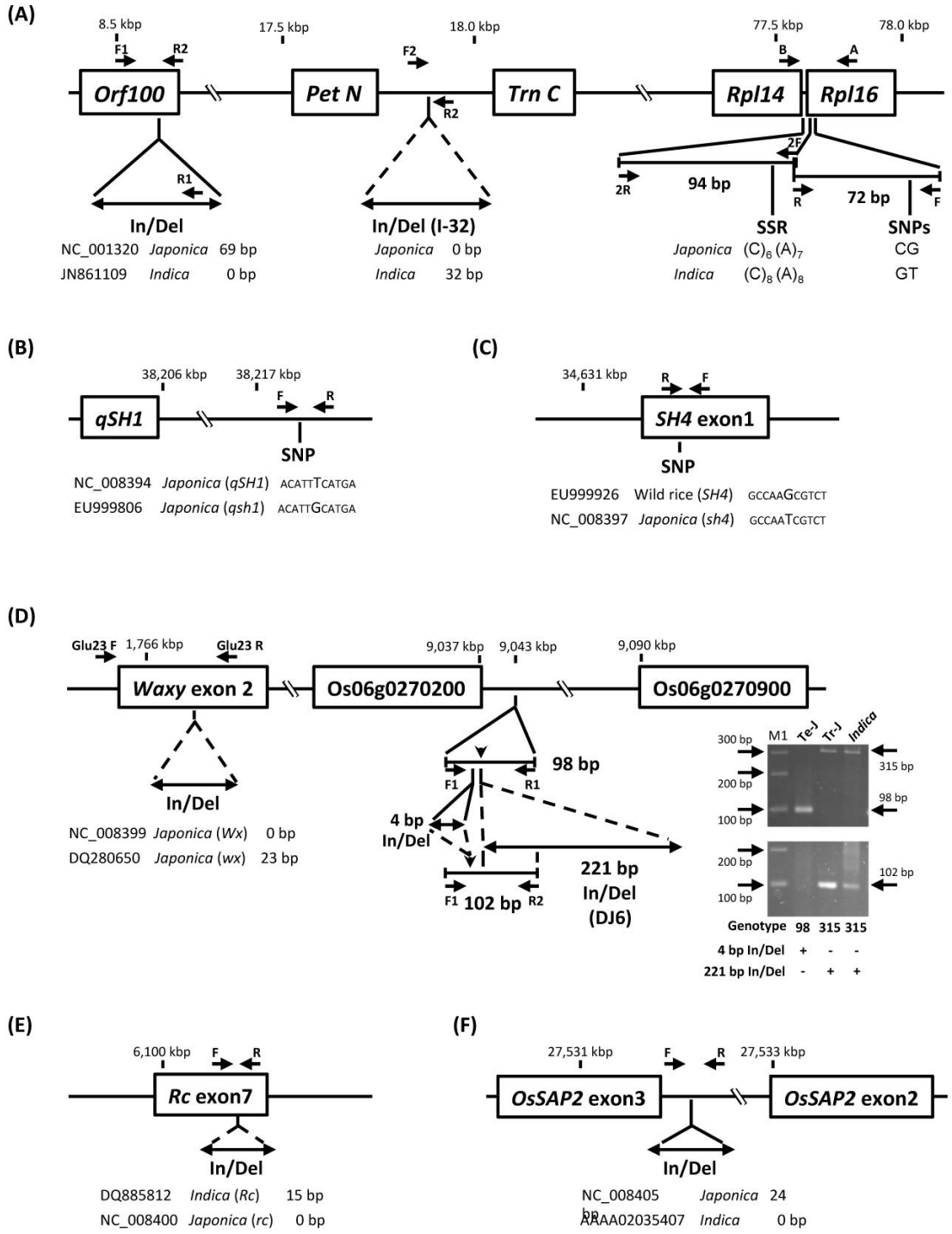

Figure 4. Location of primer pairs on the chloroplast genome and nuclear genome. Location of the chloroplast genome (**A**); and chromosome no. 1 (**B**), no. 4 (**C**), no. 6 (**D**), no. 7 (**E**), and no. 12 (**F**) are shown on the reference sequence of *O. sativa japonica* 'Nipponbare' were assigned NC_001320, NC_008394, NC_008397, NC_008399, NC_008400 and NC_008405, respectively, in the National Center for Biotechnology Information (NCBI, USA). Primer size and location are not to scale. Vertical arrowheads show insertion and/or deletion site (In/Del), and solid and broken line with bidirectional arrows are In/Del with and without 'Nipponbare'. Length of In/Del and sequence of single nucleotide polymorphism (SNP) are shown under their

sites with accession number in reference sequence registered in NCBI and the DNA Data Bank of Japan (DDBJ), except for DJ6 region. Numbers above the solid bar are the fragment size of PCR amplification with the specific primer set cited with the bar. Analysis of In/Del regions 4 bp and 221 bp in DJ6 are shown on **(D)** and three modern rice accessions were amplified with F1 and R1 primers and with F1 and R2 primers; agarose gel electrophoresis patterns of their PCR product are shown in the upper and lower panels, respectively. **M1** = 100bp DNA Ladder (TaKaRa, Japan), **Te-J** = Temperate *japonica* rice 'Nipponbare', **Tr-J** = Tropical *japonica* rice 'Malagkit sinaguing' (T0221). Genotype was estimated by presence and absence of In/Dels, which are indicated by '+' and '-', respectively. Further details of sequences of modern and ancient DNA and electrophoresis related to the analysis of their regions are in Figure S1-S9.

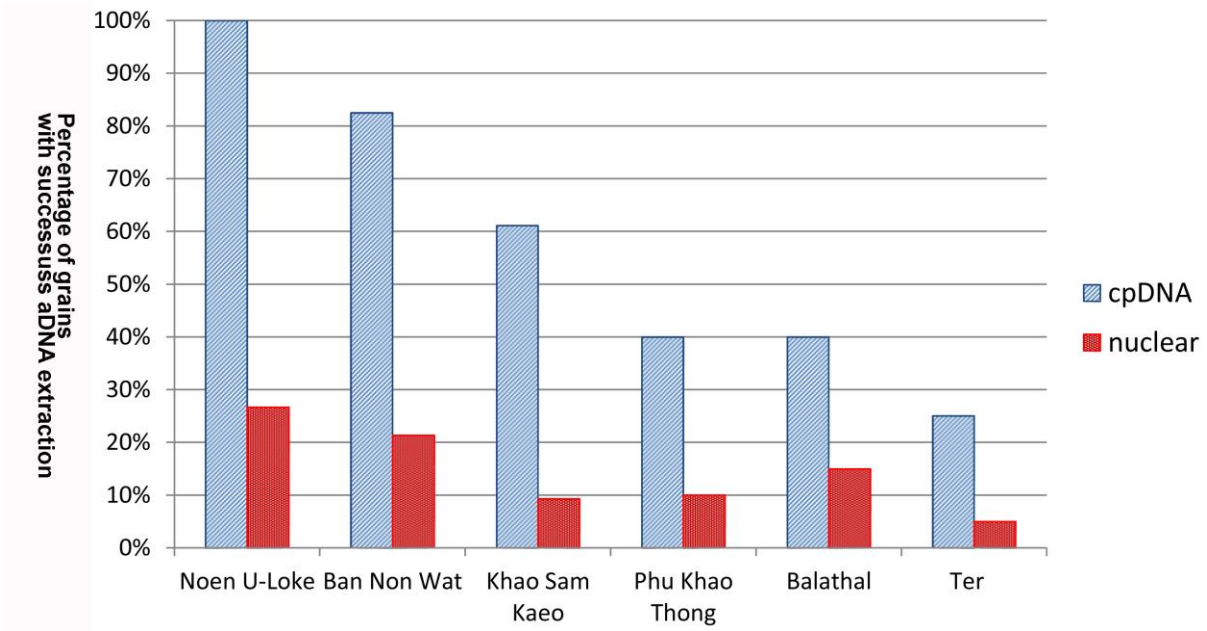

Figure 5. Rates of recovery of aDNA results comparing presence of chloroplast (cpDNA) markers and nuclear DNA (nDNA) markers. These represent percentage of tested grains on each site that produced a positive result of any cpDNA or nuclear primer.

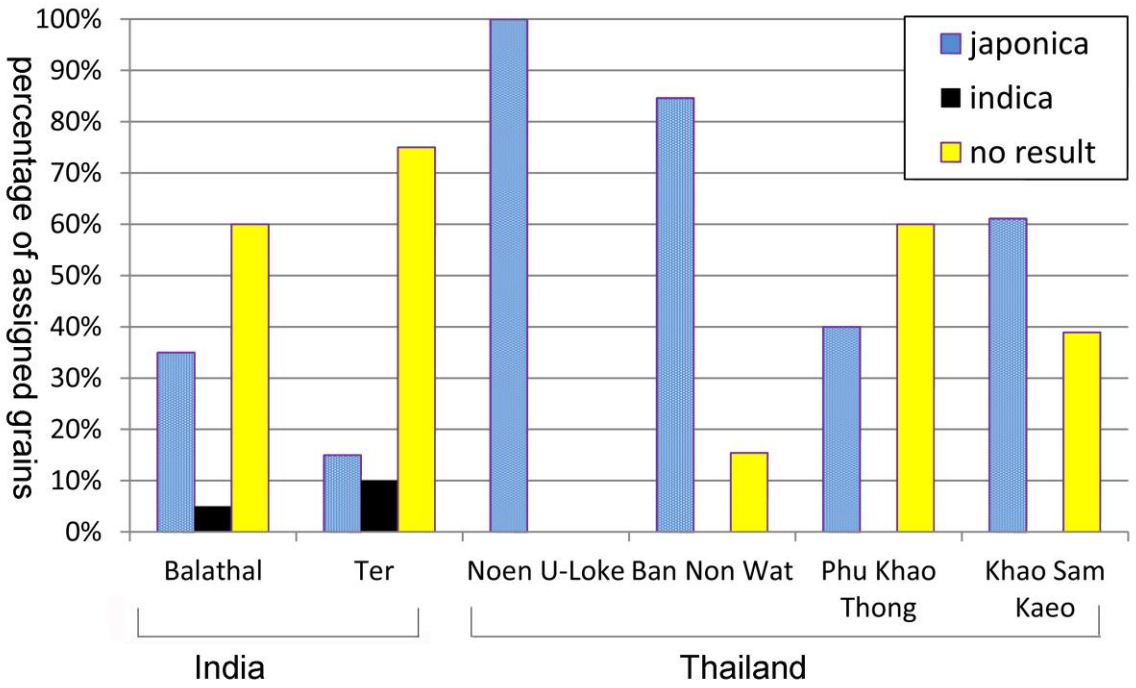

Figure 6. Identification of *indica* versus *japonica* subspecies of rice in six archaeological sites based on aDNA results.

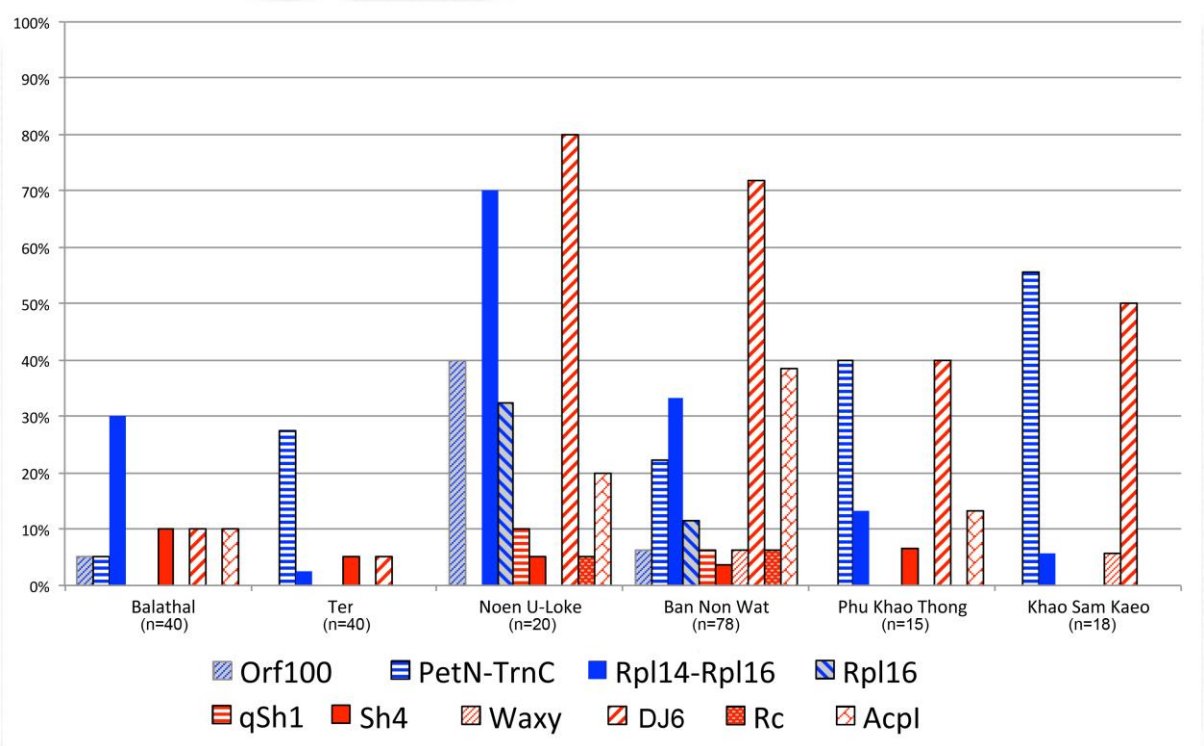

Figure 7. Percentage of recovery of various aDNA markers in the archaeological rice.

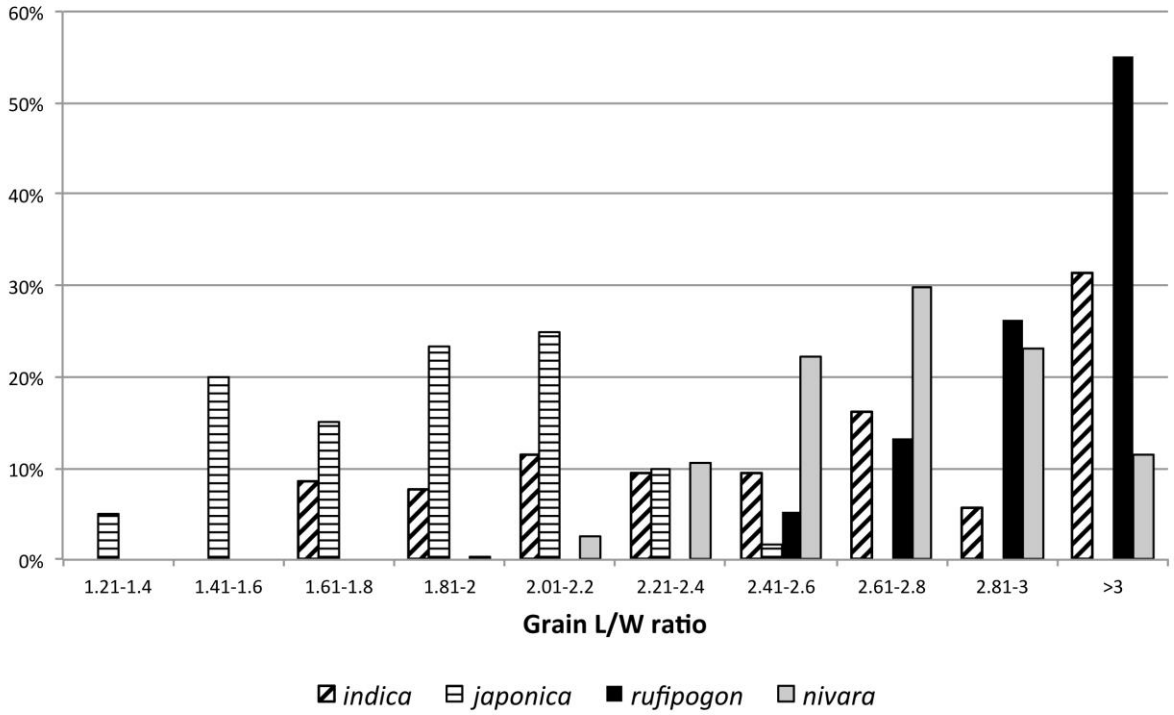

Figure 8. Comparison of L/W ratios of domesticated rice (*indica* and *japonica*) and wild rice (*rufipogon* and *nivara*). Modern reference material after Fuller et al. 2007, but *indica/japonica* assignment reassessed based on latest IRRI data; *aus* population excluded.

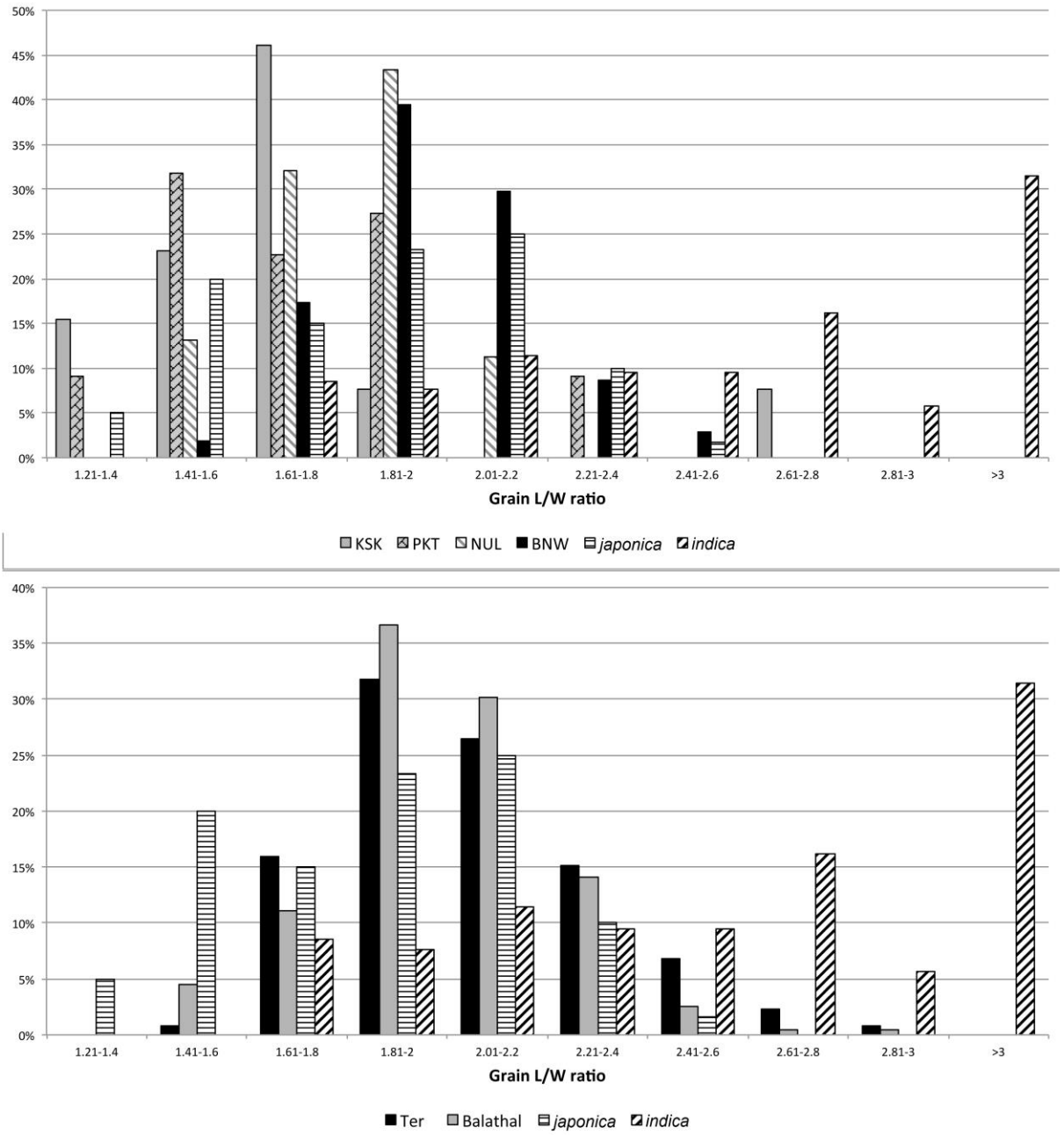

Figure 9. Archaeological rice grain L/W ratio distributions compared to modern subspecies comparisons. Archaeological rice: Terr, Balathal, Khao Sam Kaeo (KSK), Phu Khao Thong (PKT), Noen U-Loke (NUL), Ban Non Wat (BNW); Modern rice: *japonica*, *indica.*

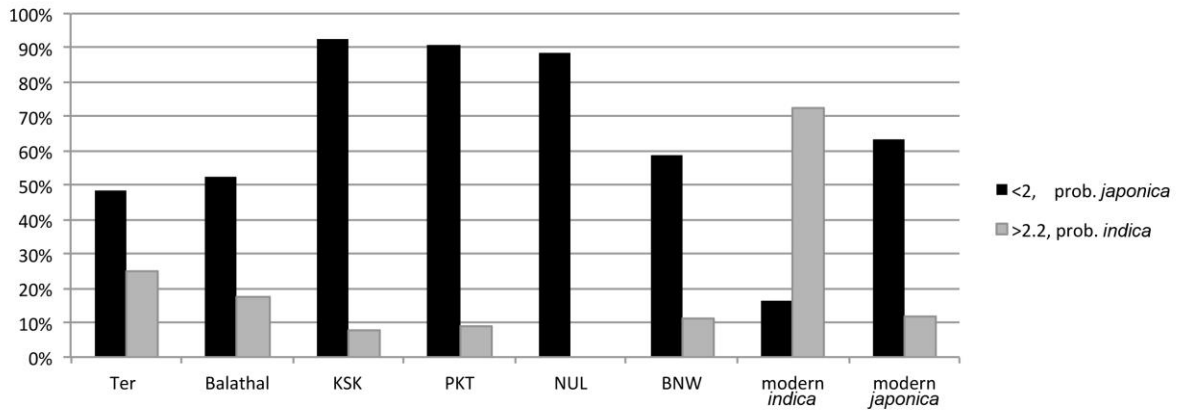

Figure 10. Bar chart indicating proportion of *indica* versus *japonica* assignment based on L/W ratio (using <2 to indicate *japonica*-type and >2.2 to indicate *indica*-type). Employing <2 and >2.2, the majority of modern reference populations are correctly assigned, but 10-15% are incorrectly assigned (and grains between 2-2.2 L/W have already been excluded). We therefore suggest an error margin of +/- 15%.

Tables:

Site	Seed size (mm) Length	Width	Thickness	L/W
Balathal				
AVE	4.3	2.1	2.0	2.0
STD	0.3	0.2	0.2	0.2
n=	199			
max	5.0	3.0	2.9	2.8
min	3.3	1.7	1.4	1.4
Ban Non Wat				
AVE	5.0	2.5	1.8	2.0
STD	0.5	0.3	0.2	0.2
n=	80			
max	6.1	3.2	2.6	2.9
min	3.8	1.7	1.2	1.6
Khao Sam Kaeo				
AVE	3.8	2.5	1.9	1.6
STD	0.5	0.5	0.2	0.2
n=	8			
max	4.5	3.1	2.3	1.8
min	3.3	1.9	1.7	1.3
Noen U-Loke				
AVE	5.1	2.8	2.0	1.8
STD	0.4	0.2	0.2	0.2
n=	50			
max	5.7	3.2	2.6	2.2
min	4.2	2.2	1.5	1.5
Phu Khao Thong				
AVE	4.7	2.7	2.1	1.7
STD	0.5	0.4	0.3	0.2
n=	30			
max	6.0	3.6	2.7	2.3
min	3.9	2.1	1.2	1.3
Ter				
AVE	5.1	2.5	1.6	2.1
STD	0.5	0.2	0.2	0.3
n=	132			
max	6.6	3.1	2.3	2.8
min	4.2	2.0	1.1	1.6

Table 1: Summary measurements of archaeological rice (from Supplemental table S1).

				Seed size (mm)			
Accession	Origin	Subspecies	Variety	L	W	Th	L/W
Fuller/Orissa 2004	Orissa, India	*indica*					
AVE				6.0	2.3	1.7	2.66
STD				0.4	0.1	0.1	0.08
n=				15			
max				6.6	2.6	1.8	2.83
min				5.2	2.0	1.4	2.54
PI 584609 / IRGC-12894	Sikkim, India	*indica*	Gompa 2				
AVE				4.9	2.8	1.8	1.80
STD				0.2	0.2	0.1	0.12
n=				15			
max				5.2	3.0	2.0	2.00
min				4.6	2.4	1.6	1.63
PI 161082 / IRGC-1608	China	*indica*	White Balga breed				
AVE				5.4	2.5	1.9	2.16
STD				0.4	0.1	0.1	0.12
n=				15			
max				5.9	2.7	2.0	2.48
min				4.6	2.2	1.8	1.96
PI 38755 / IRGC-3638	Andhra Pradesh, India	*indica*	Kamod				
AVE				7.2	2.4	1.9	3.09
STD				0.9	0.1	0.1	0.42
n=				15			
max				8.1	2.6	2.1	3.52
min				4.8	2.1	1.6	1.92
PI 67125 / IRGC-3643	Bihar, India	*indica*	Ramgarh				
AVE				6.3	2.1	1.6	3.05
STD				0.2	0.2	0.1	0.28
n=				15			
max				6.7	2.4	1.8	3.59
min				6.0	1.7	1.4	2.63
IRGC 8906	Sri Lanka	*indica*	Kotta Hanthirang				
AVE				5.8	2.4	1.9	2.40
STD				0.2	0.2	0.2	0.19
n=				15			
max				6.2	2.7	2.1	2.76
min				5.4	2.1	1.2	2.12
Thompson	Thailand	*indica*					
AVE				7.4	2.2	1.8	3.43
STD				0.4	0.1	0.1	0.24
n=				15			
max				8.1	2.4	1.9	3.86
min				6.5	2.0	1.5	2.83
Castillo	Thailand		Hom Mali				
AVE				7.2	2.1	1.6	3.40
STD				0.5	0.1	0.1	0.31
n=				114			
max				8.2	2.4	1.9	4.54
min				5.8	1.6	1.4	2.57
PI 584555 / IRGC-27630	Bagmati, Nepal	*japonica* (temp)	Darmali				
AVE				5.1	2.6	1.8	1.93
STD				0.1	0.1	0.1	0.10
n=				15			
max				5.2	2.8	2.0	2.08
min				4.9	2.4	1.5	1.75
PI 434623 / IRGC-36807	Bhutan	*japonica*	Thimphu local				
AVE				5.8	2.7	2.0	2.15
STD				0.2	0.1	0.1	0.14
n=				15			
max				6.2	2.9	2.2	2.54
min				5.4	2.4	1.9	2.00
PI 133005	Tamil Nadu, India		Vulgaris				
AVE				6.0	3.0	2.1	2.03
STD				0.5	0.2	0.1	0.16
n=				15			
max				6.6	3.3	2.3	2.32
min				5.3	2.6	1.9	1.76
PI 431084	Myanmar	*indica*	Dsi Sel Dangar Sing				
AVE				6.1	2.6	2.1	1.97
STD				0.2	0.2	0.2	0.02
n=				15			
max				6.8	2.8	2.3	2.00
min				5.7	2.2	1.6	1.92
PI 584567 / IRGC 33188	Myanmar	*japonica* (trop)	Kaukkyi Ani				
AVE				4.1	2.7	1.8	1.49
STD				0.2	0.2	0.2	0.08
n=				15			
max				4.5	3.0	2.0	1.60
min				3.8	2.5	1.5	1.33

Table 2: Measurements of modern rice. (Sources: Castillo 2013; Fuller & Harvey 2006; Thompson 1996).

Sample	Laboratory Reference No.	Radiocarbon Age BP	Calibrated Age (2σ)	δ13C (‰)	Period
Noen U-Loke 105	Wk-562	1650±70	237-562 AD	-	Iron Age
Noen U-Loke 105	Beta 376490	1690±30	255-295 AD and 320-415 AD	-25.2	Iron Age
Phu Khao Thong S7 US5	Beta 376491	1950±30	20-10 BC; 0-90 AD; 100-125 AD	-25.5	Metal Age
Khao Sam Kaeo TP57 US16	Beta 378858	1980±30	45 BC-75 AD	-28.8	Metal Age
Ter	Beta 376492	2090±30	195-40 BC	-24.7	Historic
Balathal	Beta 376482	2250±30	395-345 BC and 320-205 BC	-23.3	Historic
Ban Non Wat K500 4:2 GEN Δ	BA121030	2290±45	441-203 BC	-29.66796	Iron Age 1*
Ban Non Wat K500 4:5 GEN Δ	BA121028	2510±40	795-421 BC	-18.68128	Bronze Age 4 - Iron Age 1*
Ban Non Wat V200 7:∑3 Δ2	BA121029	2330±35	515-235 BC	-28.64504	Bronze Age 5 - Iron Age 1*
Ban Non Wat V200 7:∑4 Δ27	BA121031	2375±30	705-389 BC	-24.75305	Bronze Age 5 - Iron Age 1*

* after Higham & Higham 2009 chronology for Ban Non Wat

Table 3: AMS dates of rice grains.

Genome		Gene name[a]	Application		Forward (F), Reverse (R) Primer (5' to 3')[b]	Position (bp)[a]	Polymorphism type[c]	Product size (bp)[ad]	Anneling Temparature (°C)
Chloroplast		*orf100*	Modern rice Ancient rice	F1	TGGATTTCGAAAGTCAATTTT	8,500-8,520	In/Del	100	55.0
				R1	CCTTTTCCCACTCGCTCTCTA	8,599-8,579			
				R2	TCCATGATTCCTATTTCCAAG	8,661-8,641		162	
		petN-trnC I-32 region	Modern rice Ancient rice	F2	ATCAGTTCAAAGAATTTACTC	17,758-17,778	In/Del	75	55.0
				R2	TATTTATACTTAATGCTCCCC	17,832-17,812			
		rpl14-rp116	Modern rice	A	AAAGATCTAGATTCCGTAAACAACATAGAGGAAGAA	78,118-78,091	SSR	527	58.0
				B	ATCTGCAGCATTTAAAAGGGTCTGAGGTTGAATCAT	77,592-77,619			
			Ancient rice	2F	TCAATTTCTTCGGTTAGAAATA	77,732-77,753		94	58.0
				2R	GAAAGAAATATTGTCTTTCCAG	77,681-77,660			
			Ancient rice	F	GGCGGAGTATCCGAAACTGTA	77,821-77,801	SNPs	72	58.0
				R	TTGACTTCGTATGGGCATTTT	77,770-77,750			
Nuclear	chr. 1	*qSH1* promoter	Ancient rice	F	ATGGTATTGATGTATACTGGA	38,217,614-38,217,634	SNP	71	58.0
				R	CATCTCGTCCAAAGATCCTTA	38,217,684-38,217,664			
	chr. 4	*sh4* exon1	Ancient rice	F	AGACGCTCATCCTCATCACC	34,631,533-34,631,552	SNP	142	63.0
				R	TAGTTCTCCACCCACTTCCAC	34,631,431-34,631,411			
	chr. 6	*Waxy* exon2	Modern rice Ancient rice	23F	TGCAGAGATCTTCCACAGCA	1,765,885-1,765,904	In/Del	196	60.0
				23R	GCTGGTCGTCACGCTGAG	1,766,080-1,766,063			
		Non-cording region: DJ6 region	Modern rice Ancient rice	F1	TGACCGGTTCTGTAGCAGTG	9,043,039-9,043,058	In/Del	98	57.0
				R1	CCAGTTTAATGTTTTYTCATTGCC	9,043,136-9,043,113			
				R2	GATTTTCCGTTTTCCGTGCC	-		NA	
	chr. 7	*Rc* exon7	Modern rice Ancient rice	F	CAGAAACACCTGAATCAAGGG	6,100,183-6,100,203	In/Del	85	55.0
				R	TCTCTTTCAGCACATGGTTG	6,100,267-6,100,248			
	chr. 12	*Acp1* intron 2	Modern rice Ancient rice	F	ATCTTTTAAGCCTAATCGCGT	27,531,803-27,531,823	In/Del	146	54.0
				R	CTGTCAGAAAACTGCGACATG	27,531,911-27,531,891			

a; Gene name, position and product size were referred to chloroplast genome (accession No. NC_001320) and nuclear genome of chromosome no.1, 4, 6, 7 and 12 (accession No. NC_008394, NC_008397, NC_008399, NC_008400 and NC_008405) in *japonica* 'Nipponbare'.
b; The specific forward primer was F1 for specific reverse primers, R1 and R2, in *orf100* region and DJ6 region. Y = C or T.
c; SNP sequence and sequence length of insertion and/or deletion region are indicated in parenthesis, respectively.
d; "NA" indicated PCR amplification was not successful with specific forward and reverse primer set.

Table 4: Primer information in this study for modern and aDNA extractions.

			Site (no. of grains)					
marker	size (bp)	Total (n=211)	Balathal (n=40)	Ter (n=40)	Noen U-Loke (n=20)	Ban Non Wat (n=78)	Phu Khao Thong (n=15)	Khao Sam Kaeo (n=18)
orf100 [cp]	100-162	7.1%	5.0%	0.0%	40.0%	6.4%	0.0%	0.0%
petN-trnC	75	32.7%	5.0%	27.5%	0.0%	22.1%	40.0%	55.6%
rpl14-	94	31.8%	30.0%	2.5%	70.0%	33.3%	13.3%	5.6%
rpl16	72	16.8%	N/A	N/A	32.5%	11.5%	0.0%	0.0%
qSH1	71	1.9%	0.0%	0.0%	10.0%	6.4%	0.0%	0.0%
sh4	69-142	5.2%	10.0%	5.0%	5.0%	3.8%	6.7%	0.0%
Waxy	124-196	4.6%	N/A	N/A	0.0%	6.4%	0.0%	5.6%
DJ6	98, 315	43.6%	10.0%	5.0%	80.0%	71.8%	40.0%	50.0%
Rc	85	4.6%	N/A	N/A	5.0%	6.4%	0.0%	0.0%
Acp1	146	22.7%	10.0%	0.0%	20.0%	38.5%	13.3%	0.0%
Subspecies identification								
	Japonica	58.3%	35.0%	15.0%	100.0%	84.6%	40.0%	61.1%
	Indica	2.8%	5.0%	10.0%	0.0%	0.0%	0.0%	0.0%
	no result	38.9%	60.0%	75.0%	0.0%	15.4%	60.0%	38.9%

Table 5: Numbers of specimens processed for aDNA and recovery rates for each marker.

site	domesticated	wild	immature	indeterminate	total
Ban Non Wat	53 (96%)	1 (2%)	1 (2%)	14	69
Khao Sam Kaeo	836 (86%)	115 (12%)	23 (2%)	241	1215
Phu Khao Thong	493 (91%)	35 (6%)	9 (2%)	98	637
Ter	114 (89%)	13 (10%)	1 (1%)	-	128

Table 6: Archeological rice spikelet bases from the sites studies, classified according to morphological examination of rachilla scar following the criteria of Fuller et al. (2009). Ban Non Wat, Khao Sam Kaeo and Phu Khao Thong data from Castillo 2013; Ter data from Fuller (unpublished). There were no spikelet bases from Noen U-Loke and Balathal.

	Balathal	Ter	Noen U-Loke	Ban Non Wat	Phu Khao Thong	Khao Sam Kaeo	modern *indica*	*O. nivara* type*	*O. rufipogon* type*	modern *japonica*
DJ6 98	-	1	7	26	1	1	-	2	3	6
DJ6 94	-	-	-	-	-	-	-	-	2	-
DJ6 285	-	-	-	-	-	-	-	2	-	-
DJ6 315	4	1	4	13	5	6	20	21	9	9
DJ6 98/315	-	-	5	16	-	2	-	-	2	-
DJ6 Null	-	-	-	-	-	-	-	3	1	-
DJ6 failed	36	38	4	23	9	9	-	-	-	-

Table 7: Recovery of variant Chromosome 6 markers (DJ6). 98/315 = Heterozygous. For modern accessions, heterozygotes have also been scored for both the genotypes present. *For this table, O. nivara was assigned on the basis of possessing orf100 162 bp type, while O. rufipogon was assigned based on orf100 93 bp type. For details on the geographical distribution and modern germplasm accessions see Supplementary Tables S3, S4 and S6. PCR prodcut size is used as genotype name.

	Noen U-Loke	Ban Non Wat	Phu Khao Thong	Khao Sam Kaeo	modern *indica*	*O. nivara* type*	*O. rufipogon* type*	modern *japonica*
AcpI 122	-	4	-	-	18	9	2	1
AcpI 146	-	5	2	-	2	18	10	13
AcpI 290	-	-	-	-	-	-	1	-
AcpI failed	20	69	13	18	-	-	1	-

Table 8: Recovery of variant Acp1 markers. This was not attempted for the Indian samples. Heterozygous modern examples have been excluded. *For this table O. nivara was assigned on the basis of possessing orf100 162 bp type, while O. rufipogon was assigned based on orf100 93 bp type. For details on the geographical distribution and modern germplasm accessions see Supplementary Tables S3, S4 and S6. PCR product size is used as genotype name.

	Ban Non Wat	Phu Khao Thong	*indica*	Temperate *japonica*	Tropical *japonica*
Acp1 122/ DJ6 98	3	-	-	-	1
Acp1 122/ DJ6 315	-	-	18	-	-
Acp1 146/ DJ6 98	2	-	-	5	-
Acp1 146/ DJ6 315	-	2	2	3	5

Table 9: Combined DJ6/Acp1 genotypes, when both were recovered. Modern occurrences within our landrace set are shown for comparison. PCR product size is used as genotype name.