

Research Paper

Title: A tale of two rice varieties: Modelling the prehistoric dispersals of *japonica* and proto-*indica* rices

Authors: Fabio Silva^{1,2*}, Alison Weisskopf¹, Cristina Castillo¹, Charlene Murphy¹, Eleanor Kingwell-Banham¹, Ling Qin³ and Dorian Q. Fuller¹

Affiliations:

1. University College London, Institute of Archaeology 31-34 Gordon Square, London WC1H 0PY
2. Faculty of Humanities & Performing Arts, University of Wales Trinity Saint David, Lampeter Campus, Ceredigion SA48 7ED, United Kingdom
3. Peking University, School of Archaeology and Museology, Beijing, China

* Corresponding author, Email: f.silva@ucl.ac.uk

Abstract

We model the prehistoric dispersals of two rice varieties, *japonica* and *proto-indica*, across Asia using empirical evidence drawn from an archaeobotanical dataset of 400 sites from mainland East, Southeast and South Asia. The approach is based on regression modelling wherein goodness of fit is obtained from log-log quantile regressions of the archaeologically inferred age versus a least-cost distance from the origin(s) of dispersal. The Fast Marching method is used to estimate the least-cost distances based on simple geographical features. We explicitly test three hypotheses for the arrival of *japonica* rice to India where, it has been proposed, it hybridized with the indigenous *proto-indica*, subsequently spreading again throughout India. Model selection, based on information criteria, highlights the role of the Inner Asia Mountain Corridor in introducing *japonica* rice into northeast India, followed closely by a 'mixed route' model, where *japonica* was also almost simultaneously introduced via Assam, Bangladesh and Myanmar. Finally, we estimate the impact of future archaeological work on model selection, further strengthening the importance of the Inner Asia Mountain Corridor.

Keywords

Early agriculture, archaeobotany, South Asia, India, hybridization, model selection

Introduction

Rice (*Oryza sativa*) has played a central role in the agriculture of India, China and Southeast Asia since the Neolithic, and the possibility of more than one geographically distinct origin of rice-based agriculture, in India and China for example, has been debated for some time. There is strong support from archaeology and genetics that rice was domesticated in China, probably in the Yangtze basin, but is unlikely to be the only domestication process or the only initiation of rice cultivation (Fuller et al. 2010; Gross and Zhao 2014). A structure that is quite historically deep in this species has been recognized for some time in the division of Asian rice into two subspecies: *indica* and *japonica* (Grist 1965; Chang 1976; Oka 1988). More recently, genetic research has identified a group of *indica*-like rice lineages, referred to as *aus* rices, that appear at least as distinct as the taxonomically recognized subspecies *indica* and *japonica* (e.g. Garris et al. 2005; Kwon et al 2006; McNally et al 2009; Zhao et al 2011; Schatz et al. 2014; Civián et al. 2015; Travis et al. 2015; Choi et al. 2017), but the region of *aus* origins is poorly constrained by genetics or archaeology. Models of a single origin propose that rice differentiated into *indica* and *japonica* early on in the history of domestication (e.g. Chang 1976). Recent proponents of a single origin hypothesize that differentiation took place due to introgression of local wild rice genetics into domesticated rice as ancestral *japonica* spread from China across Southeast Asia and South Asia (e.g. Vaughan et al. 2008; Molina et al 2011; Huang et al 2012). One piece of evidence often cited in support of this is the presence of shared alleles involved with domestication traits, such as the *sh4* mutation involved in non-shattering (Li et al 2006), and other evidence for selective sweeps in the rice genomes inferred to be connected to domestication (Huang et al. 2012). On the other hand, data from chloroplast genomes, which are maternally inherited and not prone to recombination or introgression, have long indicated that *indica* and *japonica* derive from distinct wild source populations (e.g. Tang et al 2004; Takahashi 2008), while evidence across many parts of the nuclear genome also suggests differentiation in advance of domestication (e.g. Vitte et al 2004; Ohtsubo et al 2004; Londo et al 2006; Kovach et al 2007; McNally et al. 2009; He et al. 2011; Yang et al. 2012; Civián et al. 2015; Choi et al. 2017). Such evidence for multiple derivations from the wild must account for shared domestication loci through hybridization after they were selected for during the domestication process, and thus a history of dispersal of domesticated rice, presumably from China, that came into contact with populations that were ancestral to *indica* is posited (Fuller and Qin 2009; Choi et al. 2017). The aim of the present study is to constrain when and where domesticated *japonica* is most likely to have come

into contact with the maternal ancestor of *indica*, which we will refer to as *proto-indica*.

The *proto-indica* hypothesis proposes that domesticated *japonica* rice was introduced into northern India where *proto-indica* ancestors were under pre-domestication cultivation, and hybridization, followed by back-crossing with the *proto-indica* parent, transferred several genetic traits of domestication that improved the Indian rice, thus creating what is presently referred to as domesticated *indica* (Fuller et al 2010; Fuller 2011a; Fuller 2011b; Castillo et al. 2016). This hypothesis has been developed out of the “combination model” of Sang and Ge (2007) that proposed that two separately domesticated rice lineages later exchanged domestication genes. However, domestication-related genes that have been re-sequenced together with flanking DNA sequences appear to be derived from *japonica* or from perennial Chinese *O. rufiopogon* that was ancestral to *japonica*. This is clearly the case with the *rc* mutation that changes the rice pericarp color from red to white (or pale brown) and is widespread, but not universal in cultivated rice (Sweeney et al 2007). Among universal domestication genes we can include *prog1* which gives the rice plant a taller, straighter and more erect growth habit (Tan et al 2008). It also seems likely this was the case with *sh4*, involved additively in non-shattering but not the exclusive determining mutation (Ishikawa et al 2010; Zhu et al 2012), and perhaps *qsh3*, which appears to be an additional allele for reduction of shattering together with *sh4* (Htun et al 2014). Another key domestication trait with a shared mutation in *indica* and *japonica* is the closed panicle trait, encoded by *OsLG1*, which may have evolved before non-shattering as it increases harvest returns even in wild-type shattering rice (Ishii et al. 2013). More recently *LABA1*, a gene involved in reducing awn length and awn bristles has been identified as probably originating in *japonica* rice and being transferred to produce awnless *indica* varieties (Hua et al 2015). A mutation which increases grain size, *GS3*, was also transferred after selection in *japonica* to some *indica* varieties (Takano-Kai et al 2009). What these genetic factors in domestication indicate is that prior to the hybridization event (or events), *proto-indica* would have been a much less productive crop. For this reason, the main dispersal of rice throughout India (after 1000 BC) and the rise in sedentary farming villages with rice in their agriculture mix throughout the Ganges plains (mostly 1600-1000 BC), has been postulated to be a consequence of hybridization and the development of fully domesticated *indica* rice (Fuller et al 2010; Fuller 2011a). This would place the hybridization near the start of the second millennium BC. Interestingly, this is quite close to the estimated coalescence age of ca. 3900 BP for a bottleneck deriving *indica* rice from *japonica* in a coalescence model based on diversity in a single nuclear chromosome (Molina et al 2011). Also in

support of an early hybridization is ancient DNA from charred rice grains that indicates a mixture of *indica* and *japonica* chloroplast lineages in single rice populations from two sites in India over 2000 years old, while all ancient samples in Southeast Asia of similar age were purely *japonica* (Castillo et al. 2016).

In the present paper we report how well the proto-*indica* hypothesis fits the archaeological record of rice and assess alternative dispersal routes by which *japonica* could have first come into contact with proto-*indica* rices. This builds on previous spatial modelling of rice dispersal used to define the most probable centres of origin from which Asian rice dispersed (Silva et al 2015). This used the Fast Marching method (Sethian 1996; Silva and Steele 2012; 2014) to estimate the least-cost distances to all archaeological rice occurrences from any of 121 possible points of origin in mainland Asia where rice cultivation is possible. Using quantile regression between the archaeologically-inferred ages and the costs distances from the sites to the possible point of origin, a best-fitting centre of domestication, falling between the Lower and Middle Yangtze valley, was identified. Further modelling allowed for dual origins and arrived at a stronger fit for a model with two distinct centres of domestication of rice in the Middle and Lower Yangtze regions. This is therefore in agreement with the inferred dual origins of *japonica* rice posited by a number of archaeologists (e.g. Fuller and Qin 2009; Crawford 2012; Gross and Zhao 2014; Deng et al. 2015). Nevertheless, the best fitting regression also revealed clusters of early rice finds that fit poorly with this model, including a number of anomalously early finds of rice in northern India and Pakistan, which could be suggested to fit the hypothesis of a separate proto-*indica* dispersal of unimproved, cultivated but not domesticated, rice (Silva et al 2015). The present study explicitly models a separate proto-*indica* origin and explores the best fitting route and timing for *japonica* originating in the Yangtze to reach the proto-*indica* region.

Materials and methods

Materials

Our empirical evidence comes from the Rice Archaeological Database (RAD). The first version of this database was used for a synthesis of rice dispersal by Fuller et al (2010), a slightly expanded dataset (version 1.1) was used to model the dispersal of rice, land area under wet rice cultivation and associated methane emissions from 5000-1000 BP (Fuller et al 2011). The present dataset (version 2) was used in a previous analysis of the origins of rice domestication (Silva et al 2015). The database records sites

and chronological phases within sites where rice has been reported, including whether rice was identified from plant macro-remains, phytoliths or impressions in ceramics. Ages are recorded as the start and end date of each phase and a median age of the phase is then used for analysis. Dating is based on radiocarbon evidence: when available, direct AMS radiocarbon dates on rice are used (53 dates), otherwise associated bulk charcoal radiocarbon dates (158 dates) or, in some cases, based on a cultural association (material culture) that is cross-dated from other sites (251 dates). In addition to recording the presence of rice, evidence for domesticated or wild status is recorded as well as an inference of whether wet or dry cultivation ecologies of rice were employed. Version 2.0 includes 400 sites and a total of 470 phases (some sites have multiple phases). Its spatial distribution, as well as the database itself, are available from a previous publication (Silva et al 2015).

Modelling framework

Our approach expands on previous efforts to model the geographical origins, and subsequent spread, of *japonica* rice (Silva et al 2015). The methodology is based on the explicit modelling of dispersal hypotheses using the Fast Marching algorithm, which computes the cost-distance of an expanding front at each point of a discrete lattice or raster from the source(s) of diffusion (Sethian 1996; Silva and Steele 2012, 2014). Sites in the RAD database are then queried for their cost-distance, the distance from the source(s) of dispersal along the cost-surface that represents the hypothesis being modelled (see Douglas 1994; Connolly and Lake 2006; Silva and Steele 2014; Silva et al 2015 for more on this approach) and, together with the site's dating, used for regression analysis.

As an archaeological dataset such as RAD includes sites and phases which do not correspond to the first arrival of the dispersing element being studied, regressing to the conditional mean (as in an ordinary least squares regression) is not the most useful or valid approach, particularly when used for inference (see debate in Steele 2010 and Silva and Steele 2014). Quantile regression provides a reassuringly robust alternative as it allows for regression to a low quantile, and hence to a better identification of first arrival, without the need to sub-set the data or use an *ad hoc* weight function (Koenker 2005). Throughout this paper we have therefore used quantile regressions to the 10th-percentile.

Domestication is a long process, taking two to three thousand years to complete (e.g. Fuller et al 2009; Fuller et al 2014). It is unlikely that, during this period, rice cultivation would have spread that

far beyond its centre(s) of innovation, and indeed there is little to no evidence for this. Rather, one would expect its dispersal to start very slowly, but quickly pick up pace once rice nears full domestication. We therefore expect a non-linear relationship between the age of the archaeological site/phase and its distance from the supposed centre of innovation. Scattergrams show the data to be consistent with a log-log relationship between age and distance (see Figure 1 and discussion in Silva et al 2015), which involves calculating the natural logarithm of the age and cost distance for each site/phase and conducting the quantile regression on them. When the resulting regressed line is converted back into untransformed space it produces a logarithmic curve of best fit to the data, such as the blue line in Figure 1.

We have implemented a simple scenario where the domain available for dispersal is restricted only by two ecological factors: regions with a total number of degree-days in the year (e.g. Guedes and Butler 2014; Guedes et al 2015) lower than 2,500, as well as desert regions, where rice cannot grow. In addition, coastal transport was included by buffering out 40 km off-coast, allowing for land separated by a maximum of 80km to be accessible for dispersal. This is the same scenario and domain used in Silva et al (2015), where the subset of the Rice Archaeological Dataset that falls within this domain was published as supporting information.

The model previously identified as that which best-fits RAD v2 data is one where rice domestication originated from two independent centres, one in the Middle and another in the Lower Yangtze valley (Fuller 2011a; Silva et al 2015). This approach did not explicitly model the spread of *proto-indica* in the Indo-Gangetic plain, however the model can be used to predict which database entries would correspond to *proto-indica* rice: they would be the ones that are older than the expected arrival of *japonica* in the Indian subcontinent (highlighted by the yellow square in Figure 1).

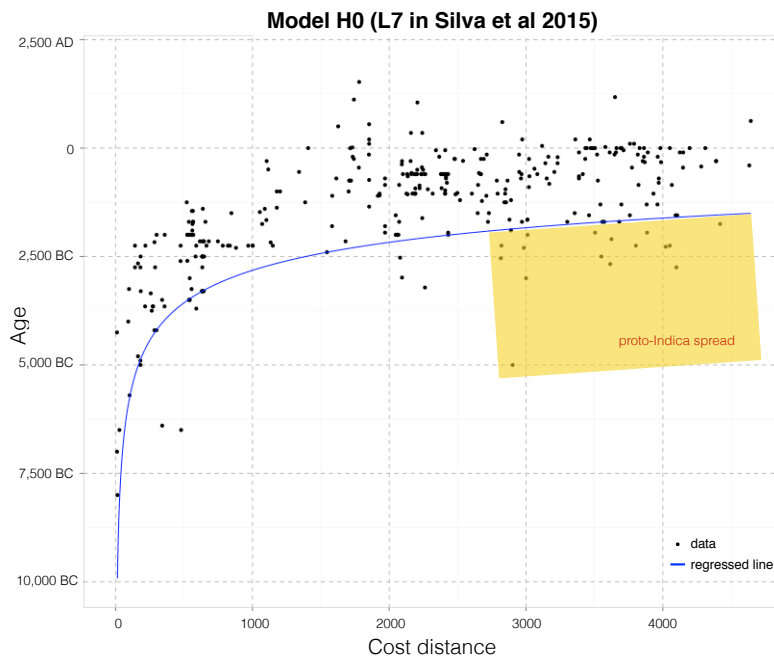


Figure 1 – Scatterplot of the dataset for the model of dispersal of *japonica* rice from two centres of innovation in the Yangtze basin (adapted from Silva et al 2015). The blue line corresponds to the best-fitting model, whereas the dots within the yellow box correspond to those Indian sites that are much earlier than predicted by the model.

Modelling the spread of proto-indica rice

To model the spread of *proto-indica* we first constructed its domain by identifying the ecoregions where such early instances of rice are found. This was done using the present potential ecoregions within the Terrestrial Ecoregions of the World dataset mentioned above (Olson et al 2001). The identified regions correspond to the Indo-Gangetic plain but also extend further to the west, where some instances of early archaeological rice are found, as well as to Bangladesh in the east, where there is no such evidence (Figure 2).

The spread of *proto-indica* was modelled on similar lines to that described above but limited to this considerably smaller domain. Its geographical origin was taken as the site of Lahuradewa which has been dated to 7000-3000 BC (Tewari et al. 2006; 2008) and is, therefore, the oldest in this region.

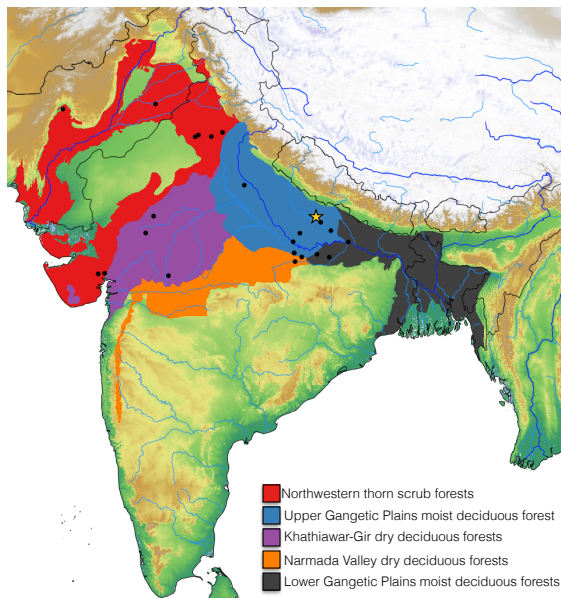


Figure 2 – Domain of spread for proto-*indica* rice. The corresponding present potential ecoregions are shown as coloured regions (see legend), whereas sites with early instances of rice, potentially corresponding to the proto-*indica* subspecies, are represented as black dots. The yellow star marks the site of Lahuradewa.

Our starting point was the regression model for *japonica* rice previously developed and described above (Figure 1). The process then was as follows:

- (1) identify the sites that fall in the Indo-Gangetic domain described above and that are older than the dates predicted by the *japonica* model;
- (2) do a log-log quantile regression of these sites against a simple cost-distance model for the spread of proto-*indica*;
- (3) redo the *japonica* log-log quantile regression without the sites identified in (1) above;
- (4) repeat steps (1) through (3) until the sum of the log-likelihood of the two regressions is maximized.

This ensures that the split of the RAD data into its proto-*indica* and *japonica* components, with associated dispersals, is done in such a way that best-fits both processes (i.e. by maximising the total likelihood). Each model has a minimum of eight parameters: the three associated with each regression (slope, y-intercept and variance), the speed ratio between the dispersal processes out of the Middle

Yangtze and Lower Yangtze, and the speed ratio between the proto-*indica* spread and one of the others, for which the Lower Yangtze process was chosen. However, once *japonica* and proto-*indica* met they hybridized into *indica* which could then have spread through the Indo-Gangetic plain faster since it was already a region of rice farmers. To account for this, we have added another parameter that controls the speed with which the hybridized variety spread through the Indo-Gangetic plain. A final parameter was added to the models that included an extra corridor (see below) to account for the fact that dispersal through this corridor could be faster (or slower) than in the rest of the domain. These parameters were allowed to fully vary so that, for instance, the speed values, and thus whether the processes were fast or slow, were entirely dictated by the statistical analysis of the data. The best-fitting values for these parameters were estimated using a Genetic Algorithm which was run for 300 generations, with a population of 50 and a mutation rate of 20% (Silva and Steele 2014; Haupt and Haupt 2004).

Comparing different dispersal routes

We have tested models based on three hypotheses for the arrival of non-shattering *japonica* rice in the Indian subcontinent and its subsequent hybridisation with native proto-*indica* rice (table 1). A “southern route” hypothesis (H1) would have seen the non-shattering variety introduced to northeast India from southwest China via Myanmar, Assam and Bangladesh. This was implicit in our previous modelling efforts (Fuller et al 2011; Silva et al 2015).

Another hypothesis was formulated based on the suggestion of trade routes connecting Central Asia to western China. If trade was the driving force for the introduction of wheat into western China via one of the routes suggested by Barton and An (2014), then it is feasible that rice could have been traded back westwards, and eventually introduced to northwest India (Fuller 2011a; Stevens et al 2016). Although Barton and An suggested three possible trade routes that circumvent the Tibetan plateau, considering that there is no archaeological evidence for rice in these regions, our modelling and statistical approach would be unable to differentiate between the different routes. Given this, we opted to implement the simplest one in schematic form only, as the others would yield the same results. This corridor was manually drawn and added to the dispersal domain mentioned above, which is available for the Fast Marching algorithm. The “Inner Asia Mountain Corridor” hypothesis (H2) therefore predicts *japonica* rice to arrive first in northwest India via a route that starts in the Yellow river valley, travels

west via the well-known Hexi corridor, then just south of the Inner Asian Mountains and thence to India (see Figure 3).

During the course of this research it was found that without imposing a barrier to movement somewhere between southwest China and Assam, a mixed scenario where rice was introduced to India both via the Inner Asia Mountain Corridor and the southern route entry points, roughly at the same time, was often recovered. The presence of such a barrier is suggested by both geographical and ecological features. Firstly, northwestern Yunnan is a region where three of Asia's greatest rivers run parallel to each other and are separated by the Hengduan mountain range, with peaks over six thousand metres high. The succession of high peaks and deep gorges creates isolated pockets which have certainly fostered biological and cultural diversity. The Three Parallel Rivers of Yunnan Protected Areas include about 1.7 million hectares (less than 0.2% of China's land area) but contain 20% of China's higher plant species (McGinley 2008). This and the broader Yunnan region support high ethnolinguistic diversity (Erard 2009). Similarly, the hills of northeastern India that separate the Brahmaputra valley from the plains of Myanmar, i.e. the "Indo-Burma hotspot" host high biotic and ethnolinguistic diversity (Gorenflo et al 2012). This region is recognized as a major frontier in both human and biotic diversity, and is postulated to have been a major barrier to human dispersal since the Palaeolithic and to have supported lower and more dispersed population densities (e.g. Boivin et al 2013). This borderland area constituted a major break in human genetic diversity (Metspalu et al 2004), and food related traditions: from Assam westwards dairy products are commonly used (Simoons 1970) while from Assam eastwards sticky varieties of rice and millets (with *wx* mutations) are popular but largely unknown to the west (Sakamoto 1996; Fuller and Rowlands 2009; Fuller and Castillo 2016). This constituted a major fracture zone in the genetic diversity of other major domesticates, including pigs (Larson et al 2010) and water buffalo (Zhang et al 2011; Mishra et al 2015). The highly variegated topography and dense, highly biodiverse forests can be expected to support a high degree of local adaptations and make traversing the region in an east to west direction difficult. Thus the main axis of rice dispersal has been from China southwards to Southeast Asia, as long recognized in the patterns of archaeology and historical linguistics (e.g. Bellwood 2011; Fuller 2011a; Sagart 2011; Higham 2014).

A second, complementary point, is the presence of *aus* rice varieties focused in Bangladesh and Assam. Genetically, *aus* is as distinct from *indica* and *japonica* as they are from each other (Garris et al 2005; Zhao et al 2011; Schatz et al 2014; Civan et al 2015; Travis et al 2015). Unlike *indica*, with

which *aus* varieties were traditionally classified, *aus* does not share key domesticated-related traits with *japonica*. For example, white pericarp varieties of *aus* have a unique *rc-s* mutation, although many varieties have the ancestral red trait (Sweeney et al 2007; Kovach et al 2007). The diversity of *aus* varieties comes from Bangladesh and India (Garris et al 2005; Zhao et al 2011), but where more detailed geolocations are available, these are specially from northeastern India, the states of Assam and West Bengal, although extending westwards as rainfed, short growing season rice through the Himalayan foothills. The term “*aus*” is traditionally used (in the Bengali language) to refer to the short season early monsoon rainfed rice, in contrast to irrigated winter *boro* rice (also genetically *aus* group), or the deep water *aswina* and *rayada* rices (as *aus* group) grown in the Ganges-Brahmaputra delta (Travis et al 2015). There are also very few genotypic *japonica* varieties found in Assam or West Bengal (Travis et al 2015). In contrast, less than 3% of rice landraces in Yunnan are attributed to *aus/boro* varieties whereas >75% of landraces are *japonica* (Zeng et al 2007). Thus it seems unlikely that early domesticated *japonica* would have dispersed through this region to give domestication related mutations to *indica*, such as *rc*, while different mutations in *aus* were not transmitted westwards. Varieties of *aus* found further west in India are likely to be younger, as are those in mainland Southeast Asia. Historical records indicate that *aus* rices came to China via Vietnam in the early Second Millennium AD, known as the “champa” rices (Barker 2011). We therefore conclude that it is unlikely that rice cultivation would have dispersed through this region, or at least not quickly and easily. To represent this in hypothesis H2, we have imposed a simple geometrical barrier that stops potential east-to-west movement along this region and which is represented in Figure 3.

However, we have also included the emergent “mixed route” hypothesis (H3) in the modelling to see whether it is supported at all by the archaeological data. This hypothesis, original to this paper, suggests two concurrent introductions of *japonica* rice into the Indian subcontinent: one into the northwest and another into the northeast (Figure 3), as discussed above.

Model		References
H1	Southern route	Kovach et al. (2007); Vaughan et al (2008); Ikehashi (2014)
H2	Inner Asia Mountain Corridor	Fuller (2011); Boivin et al (2012)

H3 | Mixed route

This paper

Table 1 – Literature-based models tested in this paper, with associated references.

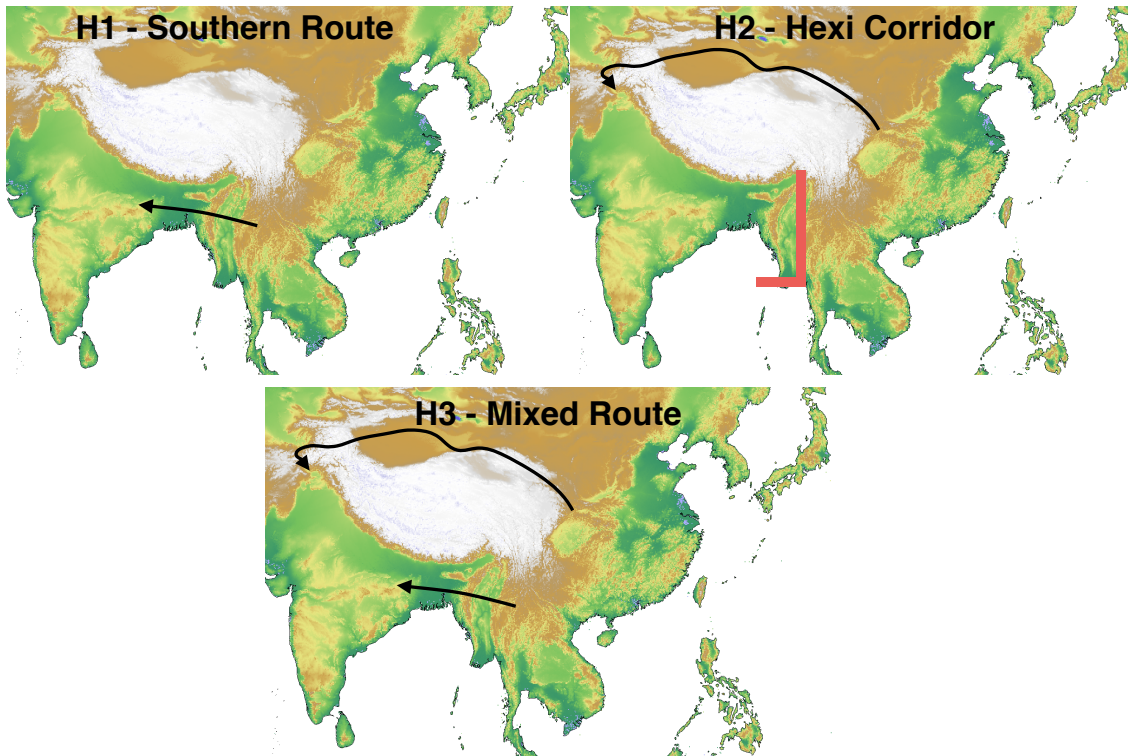


Figure 3 – The three hypotheses used for the spread of rice into India.

As models H2 and H3 include a trade route – the corridor – our modelling efforts must acknowledge that dispersal along such a route could occur at a different rate from that in the rest of our domain. To implement this, we have included an extra parameter in these models that effectively controls the friction of these corridors: a low friction yielding a higher spread rate or, alternatively, a higher friction yielding a lower spread rate (see Silva and Steele 2014 for more on this approach).

Given this, one needs to be careful with our model selection methodology, as models with more parameters will almost always fit the data better. The Akaike Information Criterion (AIC) has been widely applied for model selection and inference and has a solid foundation in information theory and likelihood statistics (Akaike 1973, Burnham and Anderson 2002). It penalises less parsimonious models so that it prefers a best-fitting model that does not do so at the cost of adding extra parameters. The best model is the one with the lowest AIC value and it is therefore useful to calculate a Δ value by subtracting this lowest value from the AIC values of all models, $\Delta = AIC - AIC_{min}$. Burnham and Anderson (2002: 70–

71) have argued that models with Δ values higher than four have very little empirical support in favour of them (see also Edwards 1992). However, this threshold is not theoretically driven and rather arbitrary so one should treat it as less of an automatic cut-off and more as a value judgment (Anderson 2008: 90). For this it helps to calculate w_i , the Akaike weights or model probabilities, which provide a measure of support for the different models (Burnham and Anderson 2002: 74–80). We have used the small sample corrected version for the AIC, often denominated AICc, throughout but, for simplicity, we have followed Burnham and Anderson (2002) in simply naming it AIC.

Estimating impact of future archaeological work on model selection

To estimate how the model selection results would be affected by future targeted archaeological work in the Myanmar/Assam region (see Discussion for the importance of this), we have simulated a number of extra site/phases and recalculated the corresponding AIC values for each model. The location of these spurious sites was randomly sampled from a domain falling between latitudes 15.5°–30.5°N and longitudes 87°–101°E. The simulated age was a combination of the expected age at that site taken from the model corresponding to the underlying assumption with random noise sampled from a Rayleigh distribution with a parameter of 0.05 (on this choice of distribution see Silva and Steele 2014). The AIC values of each model, assuming the same parameter values, as well as their Δ values were then calculated. This was done 1000 times for each extra site/phase, and the distributions of the resulting Δ values were then plotted as boxplots.

Results

Table 2 shows the statistical results for the hypotheses considered. Included are all quantities used to calculate the AIC values, as well as the Δ values and model probabilities w_i , which provide a measure of support for the model, given the data. We also present the previously found best-fitting model, which did not explicitly model the spread of *proto-indica*, for comparison (here model H0, corresponding to model L7 of Silva et al 2015). The estimated values for the dispersal parameters – i.e. the speed ratios on the different domains – are also given in table 2.

Model	Estimated Dispersal Parameters	N	K	Log-likelihood	AIC	Δ	w_i
H0	{0.784}	330	4	-111.505	231.134	141.233	<i>inf</i>
H1	{0.66, 16.49, 3.41}	330	9	-45.510	109.582	19.682	0.005 %
H2	{0.67, 25.42, 1.11, 2.23}	330	10	-34.605	89.901	0	94.94 %
H3	{0.66, 13.96, 0.83, 3.54}	330	10	-37.539	95.767	5.866	5.05 %

Table 2 – Results for all considered models. Estimated parameters are speed ratios with relation to the Lower Yangtze process. N stands for sample size, K for the total number of parameters of each model (i.e. the dispersal parameters plus the regression parameters), *inf* stands for an infinitesimally small value. The best-fitting model (H2) is highlighted in bold.

Model H2, the Inner Asia Mountain Corridor hypothesis, is the preferred model with a 94.9% likelihood of it being the best model out of this set and given the present data. We therefore present also the scatterplot of the RAD data for this model (Figure 4).

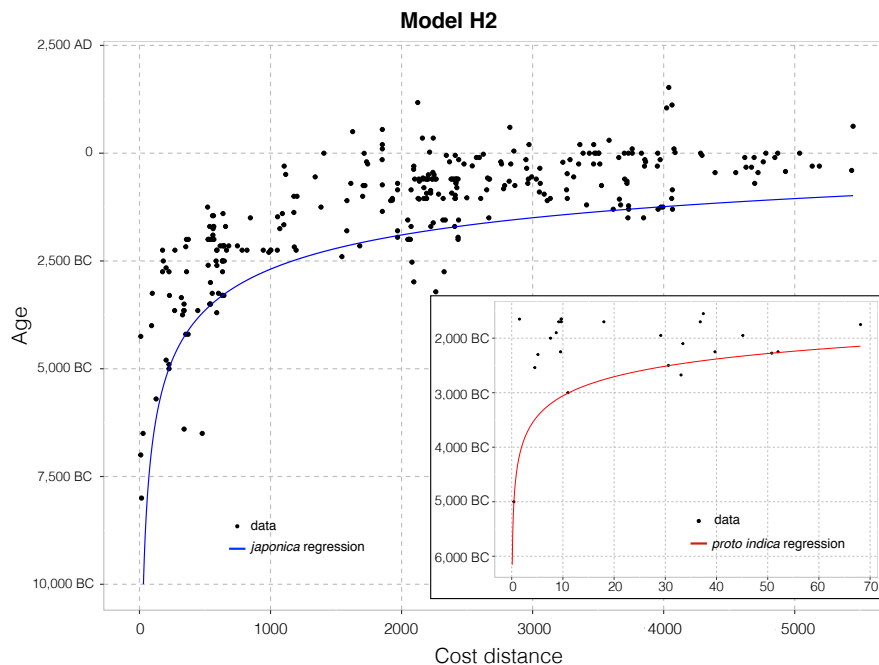


Figure 4 – Scatterplot of the best-fitting model H2 separated into its *japonica* (main plot) and *proto-indica* components (inset).

Discussion

The first result is that the explicit modelling of *proto-indica* spread (which was only done for models H1, H2 and H3) significantly increases the fit of the models to the archaeological dataset. This was to be expected by the addition of the extra parameters. However, the increase in log-likelihood is accompanied by an increase in the penalty given by the AIC to models with extra parameters, meaning that this addition is justified by the data.

The results also show that the addition of the Inner Asia Mountain Corridor significantly improves the model's fit to the data, particularly model H2 where rice is introduced to the Indian subcontinent exclusively via a trade route that circumvents the Tibetan plateau (see Figure 5). This agrees with independent archaeological evidence that sees millets spread westwards along this corridor perhaps as early as 3,000 BC (e.g. Rassamakin 1999; Kohler-Schneider and Canepelle 2009; Boivin et al 2012) and certainly by 2,500-2,000 BC (Frachetti et al. 2010; Spengler 2015; Stevens et al. 2016), i.e. in the same time frame as that predicted for rice in model H2. The arrival of western livestock (sheep, cattle) into central China, 2,500-2,000 BC (Yuan and Campbell 2009; Fuller et al. 2011), and wheat, ca. 2,000 BC (Flad et al 2010; Betts et al 2014; Zhao 2015; Stevens et al. 2106) add evidence for the role of the Inner Asia Mountain Corridor for domesticated species dispersal in this period.

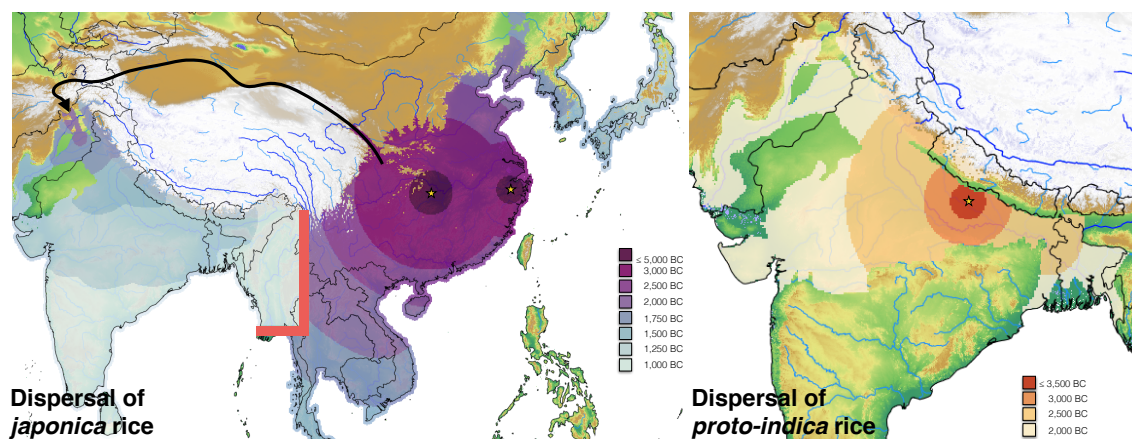


Figure 5 – Predicted arrival times of the two rice varieties across eastern and southeastern Asia, based on best-fitting model H2. The sources of dispersal are identified by yellow stars and the colours indicate arrival time in 500-year intervals.

This modelling approach, albeit entirely based on spatio-temporal data and dispersal

hypotheses, predicts which archaeological rice entries should be of the proto-*indica* variety. Table 3 presents the list of sites that are predicted to have had rice of this variety before the arrival of the hybrid non-shattering variety (*indica*). Also included is the median for the earliest date associated with rice at the site (more details can be found in RAD), as well as the predicted arrival time for *indica* at the site. The discrepancy between the two dates shows how long proto-*indica* would have been in use at the site. These are predictions that are testable in the future with recourse to morphometric and/or archaeogenetic analysis of the macroremains of these sites (e.g. Castillo et al 2016). In fact, archaeogenetic analysis of rice from the Historic period in the Indian site of Balathal, shows the presence of both *indica* and *japonica* varieties. So, it would be possible to test for the presence of proto-*indica* rice using aDNA in the Chalcolithic level at Balathal. Such analyses can, subsequently, be fed back into the model to further improve it. It should be noted that at least one or two sites that the model predicts were part of the proto-*indica* spread have produced evidence for domesticated or possible *japonica* type rices, although in this case the site occupation period, at one standard deviation, included the predicted arrival date. The first is Mahagara, where a small sample of spikelet bases are predominantly non-shattering, implying that hybridization between proto-*indica* and *japonica* had already occurred by the time this site was occupied (ca. 1650 BC), although predicted arrival date for domesticated *indica*, 1395 BC, nearly falls within the 1-sigma range of summed radiocarbon dates from the site (1700-1400 BC). The second site is Pirak (ca. 1750 BC), in western Pakistan, which produced grains that were argued to look like *japonica* on the basis of grain shape (Costantini 1979), although this could perhaps be re-assessed using more recent methods. These occurrences are around 250 years earlier than the expected arrival of domesticated *indica* and may indicate that after hybridization this species spread more quickly than predicted by the present model. This makes sense as rather than the spread being by demic diffusion, it could have been by direct transfer along trade routes between communities that were already farmers. In this sense the initial spread of domesticated *indica* through the territory of proto-*indica* can be expected to be much faster as it is a case of substituting an improved seed grain by those who already knew rice or similar crops.

RAD ID code	Site Name	Earliest Attested Rice Date (cal BC)			Expected <i>indica</i> arrival (cal BC)
		Start	Median	End	

266	Lahuradewa	7000	5000	3000	1488
264	Kunjhun	3500	3000	2500	1379
263	Kunal	2850	2675	2500	1774
261	Khairadih	2880	2540	2200	1444
260	Balu	2800	2500	2200	1747
258	Damdama	2600	2300	2000	1445
249	Lothal	2350	2275	2200	1303
250	Rangpur	2500	2250	2000	1295
256	Balathal	3000	2250	1500	1430
255	Senuwar	2500	2250	2000	1385
253	Banawali	2300	2100	1900	1777
262	Jhusi	2500	2000	1500	1421
248	Daulatpur	2200	1950	1700	1730
317	Harappa	2200	1950	1700	1929
247	Malhar	2200	1900	1600	1393
315	Pirak	1950	1750	1550	1556
242	Dangwada	2000	1700	1400	1364
244	Tokwa	2000	1700	1400	1389
239	Chirand	2000	1700	1400	1406
245	Atranjikhera	2000	1700	1400	1555
236	Koldihwa	1900	1650	1400	1395
237	Mahagara	1900	1650	1400	1395
238	Imlidih-Khurd	2000	1650	1300	1468
252	Ojyana	1700	1550	1400	1478

Table 3 – List of sites whose rice is predicted to be of the proto-*indica* variety by the best-fitting model, with its RAD code, dates and predicted arrival date for *indica* rice. The discrepancy between the dates shows how long proto-*indica* would have been in use at the site.

A key expectation of the proto-*indica* hypothesis is that fully domesticated rice with non-shattering spikelet bases would appear in India first in the northwest and only after indirect contacts

with China had been established. Recent archaeobotanical data on rice from Haryana fits and strengthens this hypothesis (Bates et al 2017). Rice spikelet bases from Mature Harappan Masudpur VII (2500-2100 BC) and Masudpur I (2300-2000 BC) are both predominantly wild type. While the significant minorities of non-shattering (11.2% and 20%) suggest that some selection for this domestication trait had taken place in India, thus confirming proto-*indica* cultivation (Bates et al 2017), they do not suggest a fully indigenous domestication process. By contrast the later site of Bahola (1900-1700 BC) has more or less fully domesticated types (>76% non-shattering) comparable to sites near the end of domestication processes in Chinese rice or near eastern cereals (see Fuller et al 2014). This shift would either require local evolution at an order of magnitude faster than any previously documented domestication in the archaeobotanical record (see Fuller et al 2014: Table S1) or indicates that domesticated *japonica* or hybrid *indica* had arrived by this later period. In other words, these new data suggest a time for the arrival of *japonica* or *japonica*-derived hybrid *indica*. As these data were published after this article was initially submitted and after our modelling had been completed, we simply compare the predictions of the best-fitting model with the dates of these more recent studies, showing good agreement (Figure 6).

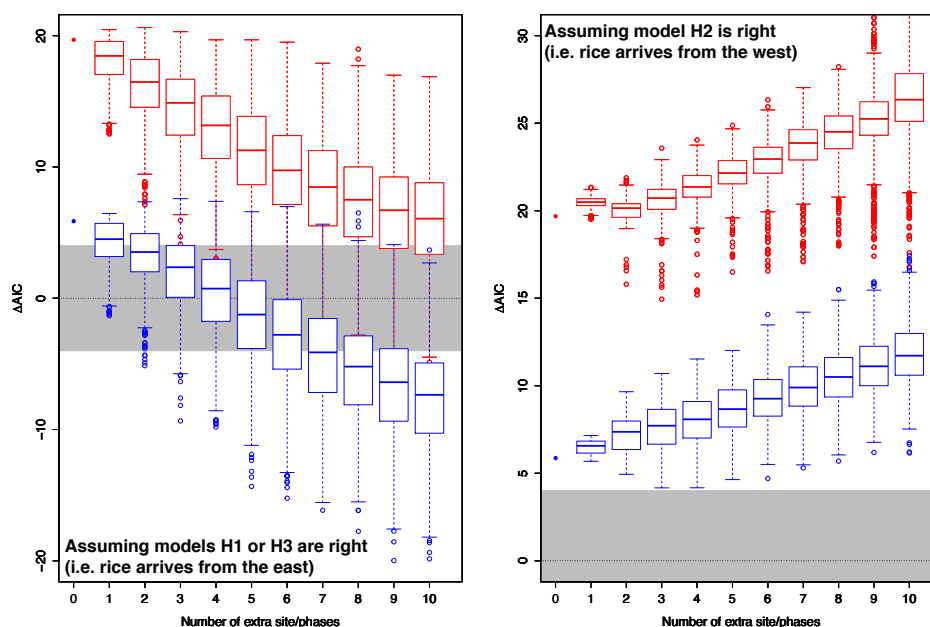


Figure 6 – Comparison of the best-fitting model's predictions for the arrival of non-shattering rice (blue-shaded normal distributions) with the latest dates from the Indian subcontinent. Red-shaded areas indicate the model's predicted time range for proto-*indica*, whereas green-shaded areas indicate that for *indica*.

Future refinements will be welcomed, not only on the modelling but also on the data collection front. Myanmar (Burma) and Assam are clear gaps in our dataset of particular importance to the present question as they would form the entry point of Chinese rice into northeastern India in models H1 and H3 although, in the latter, the influence of this in the spread of rice in the Indian subcontinent would have been limited. Targeted archaeobotanical work in these regions would considerably help further understand the dynamics of rice dispersal, and the timing of its arrival, in this region.

From a purely statistical point of view, the Δ values obtained (table 2) can be considered low: they are certainly lower than the ones obtained in our previous analysis of different literature-based hypotheses for the geographical origins of *japonica* using the same dataset (Silva et al 2015). The Δ between models H2 and H3 ($\Delta = 5.866$), is not that much greater than 4, suggesting that there is still some, albeit small, empirical support in favour of model H3. It is therefore valid to ask whether with an improved dataset this value could be reduced sufficiently for the two models to be statistically indistinguishable (i.e. Δ between -4 and 4) or whether the mixed route model would in fact become the best-fitting model (i.e. Δ values lower than -4).

To predict how model selection would be affected by future archaeological work we have simulated the presence of a number of early archaeological sites in this region. The methodology developed for this was described above, whereas the results are shown in figure 7. The figure shows the current Δ values as filled dots, and their expected distributions for increasing number of extra site/phases. The boxplots represent the expected variation in the age determination, hence the (relatively) large spreads. Nevertheless, the trends are quite clear. Firstly, assuming rice cultivation was introduced to this region from the east, as in models H1 and H3 (left figure), model H3 (in blue) will always outperform model H1 (in red), with an almost constant difference of about 14 in median Δ values. This gives further strength to the importance of the northern trade route in the narrative of rice dispersals into India, even as future empirical data might suggest a level of input from the southern route. Furthermore, one can see that one would need four or more early site/phases for model H3 (in blue) to outperform model H2 (viz. negative Δ value with respect to model H2). Model H3, however, would only become the best-fitting model when this Δ value would drop below -4, which occurs for seven or more early site/phases. That information criteria would not be affected by the inclusion of a small number of extra site/phases is to be expected from the statistical nature of the analysis, however that with only

seven more site/phases one would have to select a different model is a welcome surprise. On the right-hand side of figure 7, we have the converse hypothesis: that rice was introduced to the Myanmar/Assam region from India, as in model H2. Under these circumstances model H2 would be increasingly better than the others and, as expected, the Δ values have a steady increase.

Considering these two figures together, and under the assumption that their underlying hypotheses are equiprobable, one can say that, with four to ten extra dated early site/phases in the Myanmar/Assam region one would be able to more securely differentiate between the Inner Asia Mountain Corridor and the mixed route hypotheses. However, given our modelling assumptions and present data, an exclusively southern route is essentially out of the question.

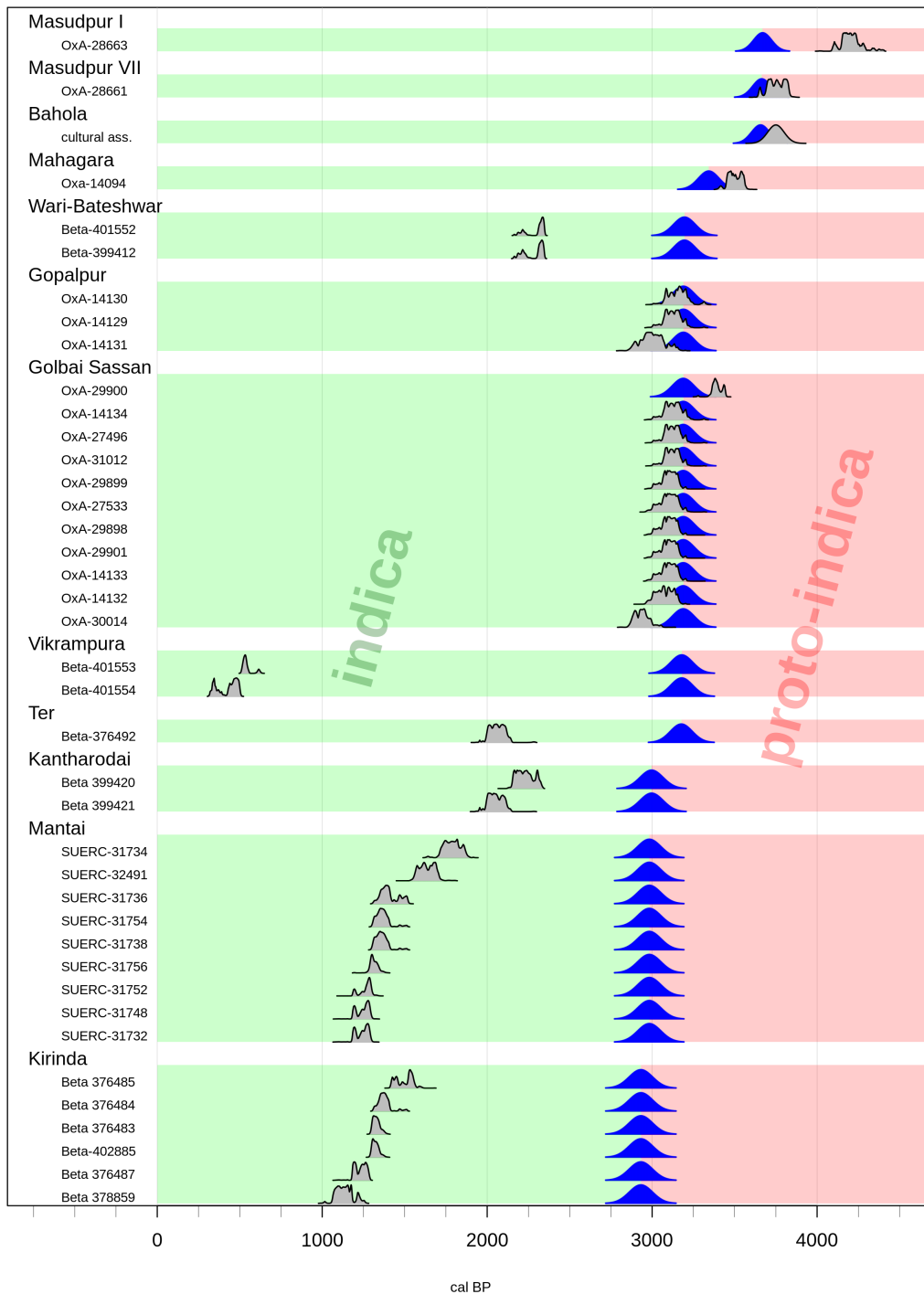


Figure 7 – Impact of future targeted archaeological work on model selection. The left figure assumes that rice arrived in the Assam/Myanmar region from the east (that is, as in models H1 and H3), whereas the right figure assumes that it arrives from the west (as in model H2). The boxplots show the spread in Δ values between the three models (vertical axis) for an increasing number of dated early archaeological site/phases (horizontal axis). Red boxplots for the Δ value of model H1 and blue for H3,

both with respect to model H2. The grey-shaded area represents the region with Δ values between -4 and 4, i.e. models that fall in this region have empirical support close to that of model H2.

Conclusions

Through a combination of explicit spatial modelling and simulation we have demonstrated the high likelihood that dispersal of rice via traders in Central Asia introduced *japonica* rice into South Asia. Only slightly less likely is a combination of introduction via two routes including a Central Asia to Pakistan/northwestern India route as well as introduction to northeastern India directly from China/Myanmar. However, there is a very low probability that current archaeological evidence for rice fits with a single introduction of *japonica* into India via the northeast. We have also simulated the minimum amount of archaeobotanical sampling from the Neolithic (to Bronze Age) period in the regions of northeastern India and Myanmar that will be necessary to strengthen support for the combined introduction (model H3) or a single Central Asian introduction (model H2). As few as four Neolithic data points, and certainly less than ten, ought to resolve this. This means that targeted archaeobotanical sampling in these regions should be a current research priority and will greatly refine our understanding of the early evolution and geographical distribution of cultivated rice subspecies with relatively little archaeological effort.

In addition, our model explicitly predicts the spread of non-domestication cultivation of proto-*indica* in India and Pakistan. Therefore, sites listed in Table 3, or sites of the same cultural affiliations (age and region) are predicted to have used proto-*indica*. This can be tested through future archaeobotanical studies. For example, such sites are predicted to produce rice remains with morphologies consistent with wild rice, such as shattering and immature rice spikelet bases (*sensu* Fuller et al 2009; Deng et al 2015), grain shapes consistent with wild rice or *indica* (*sensu* Castillo et al 2015), and ancient DNA signatures of the *indica* chloroplast haplotypes (Castillo et al 2015). Only after the arrival of *japonica*, with maximum ages for different regions predicted in Figure 8 we expect aDNA evidence for the presence of *japonica* chloroplast haplotypes, and mixtures of *japonica* and *indica* haplotypes in archaeological material, as found in currently available ancient DNA in India, from after 300 BC (Castillo et al 2015). In addition, we expect to find mixtures of *indica* and *japonica* grain morphology after this introduction (as reported for after 300 BC in Castillo et al 2015), and the presence

of non-shattering, domesticated spikelet bases. Presently such spikelet bases have been observed from Mahagara (ca. 1650 BC) (Fuller 2011b), Gopalpur and Golbai Sassan in eastern India (ca. 1250 BC) (Kingwell-Banham 2015), Wari-Bateshwar in Bangladesh after 400 BC (Rahman et al., n.d.), Paithan in south India after 300 BC (Fuller, n.d.) and in northern Sri Lanka at Mantai after 100 BC (Kingwell-Banham 2015) and Kantharodai after 150 BC (Murphy et al., n.d.) (Fig. 8). In the case of the eastern Indian site the occurrences of domesticated rice, and indeed the foundation of these sites based on available radiocarbon dates is fairly close to the predicted arrival date of domesticated *indica*, at most ca. 250 years earlier (ca. 1500 BC for the earliest range of their radiocarbon dates).

As noted earlier, spikelet bases from Mahagara are earlier than predicted by the present model, also by close to 250 years, and therefore even if this site is within the spread zone of proto-*indica*, this particular community appears to have been founded after domesticated rice had become established in the middle Ganges region. Also, grains from Pirak in Pakistan were inferred to be *japonica* (Costantini 1979), which would contradict the model prediction that this site was part of the proto-*indica* spread. There are likely two factors to contribute to these contradictions. The first is that the ecological zones that we used to define the geographical constraints of proto-*indica* may be imperfect, as in addition to bioclimatic conditions factors of cultural tradition may have limited the dispersal of rice. Second, once hybridization occurred the spread of domesticated *indica* was very fast within the region that already had agriculture and/or familiarity of proto-*indica*, which makes the initial adoption and substitution of this rice variety different from the spread of the crop into new regions, which was more likely driven by demic diffusion-like processes, for example with the spread of *japonica* through Southeast Asia, or the later spread of *indica* through southern India and Sri Lanka. Nevertheless, systematic collection of more such data from sites nearer to the predicted time of *japonica* arrival and region, or regions, of that arrival in northwestern and/or northeastern India can further serve to test the predictions of our alternative northwestern (H2) or combined introduction (H3) models. Such research indicates archaeobotanical evidence provides a fossil record of cultivated rice that can be quantitatively modelled to complement and improve inferences of rice's evolutionary history derived only from modern genetic evidence.

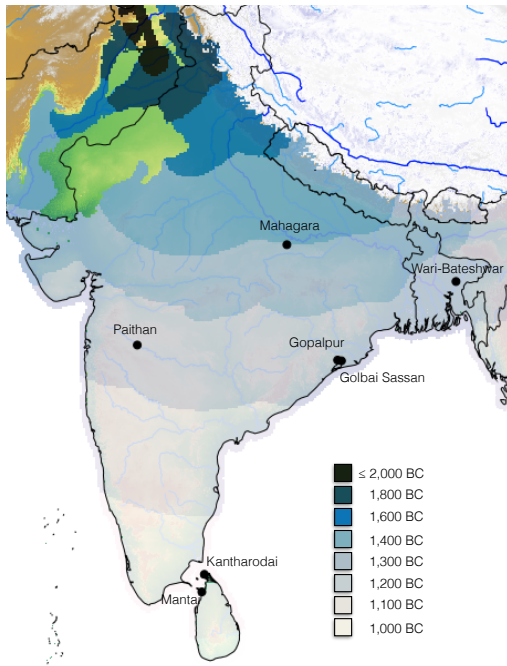


Figure 8 – Predicted arrival times of the non-shattering rice variety (*japonica* or the hybrid *indica*) across southern Asia, based on best-fitting model H2. Included are also sites with known presence of non-shattering spikelet bases (see text).

Acknowledgements

This research is part of the Natural Environment Research Council (UK) funded project “The impact of evolving rice systems from China to Southeast Asia” (NE/G005540/1) awarded to Dorian Q Fuller and Andrew Bevan. Some work on the database by EKB and CM was carried out as part of the European Research Council project “Comparative Pathways to Agriculture” (Advanced Grant no. 323842). We would also like to acknowledge work by Chris J. Stevens on the database, as well as the many discussions around the proto-*indica* hypothesis.

All rasters and map outputs were generated in GRASS GIS (version 7), and the off-coast buffering was done using its *r.buffer* routine. All modelling and other output generation was done in R (R Core Team 2014), in particular: the quantile regressions were done using the *quantreg* package (Koenker 2013) and parameter estimation was done using the genetic algorithm of package *GA* (Scrucca 2013). The modified Fast Marching implementation of Silva and Steele (2012, 2014) was recoded in R, based on previously written MATLAB code, and is available to download at <https://github.com/f-silva-archaeo/fastmaRching>

References

- Akaike H (1973) Information theory as an extension of the maximum likelihood principle. In: Petrov BN and Caski F (eds) *Second International Symposium on Information Theory*. Budapest: Akademiai Kiado, pp.267–281.
- Anderson DR (2008) *Model Based Inference in the Life Sciences: a primer on evidence*. New York: Springer.
- Barker R (2011) The Origin and Spread of Early-Ripening Champa Rice: It's Impact on Song Dynasty China. *Rice* 4: 184–186.
- Barton L and An C-B (2014). An evaluation of competing hypotheses for the early adoption of wheat in East Asia. *World Archaeology* 46(5): 775–798. Doi: 10.1080/00438243.2014.953703
- Bates J, Petrie CA and Singh RN (2017) Approaching rice domestication in South Asia: New evidence from Indus settlements in northern India. *Journal of Archaeological Science* 78: 193–201.
- Bellwood P (2011) The checkered prehistory of rice movement southwards as a domesticated cereal – from the Yangze to the equator. *Rice* 4: 93–103.
- Betts A, Jia PM and Dodson J (2014) The origins of wheat in China and potential pathways for its introduction: A review. *Quaternary International* 348: 158–168.
- Boivin N, Fuller DQ and Crowther A (2012) Old World globalization and the Columbian exchange: Comparison and contrast. *World Archaeology* 44: 452–469.
- Boivin N, Fuller DQ, Dennell R, Allaby R and Petraglia MD (2013) Human dispersal across diverse environments of Asia during the Upper Pleistocene. *Quaternary International* 300: 32–47. Doi: 10.1016/j.quaint.2013.01.008

Burnham KP and Anderson DR (2002) *Model Selection and Multimodel Inference: A Practical Information-Theoretic approach*, second ed. New York: Springer.

Castillo CC, Tanaka K, Sato Y-I, Ishikawa R, Bellina B, Higham C, Chang N, Mohanty R, Kajale M and Fuller DQ (2016) Archaeogenetic study of prehistoric rice remains from Thailand and India: evidence of early japonica in South and Southeast Asia. *Archaeological and Anthropological Sciences* 8(3): 523–543. Doi: 10.1007/s12520-015-0236-5

Chang TT (1976) Origin, evolution, cultivation, dissemination, and diversification of Asian and African rices. *Euphytica* 25: 425–441.

Choi JY, Platts AE, Fuller DQ, Hsing Y, Wing RA and Purugganan MD (2017) The rice paradox: Multiple origins but single domestication in Asian rice. *Molecular Biology and Evolution* 34 (4): 969–979. Doi: 10.1093/molbev/msx049

Civáň P, Craig H, Cox CJ and Brown TA (2015) Three geographically separate domestications of Asian rice. *Nature Plants* 1: 15164. Doi: 10.1038/nplants.2015.164

Connolly J and Lake M (2006) *Geographic Information Systems in Archaeology*. Cambridge: Cambridge University Press.

Costantini L (1979) Plant remains at Pirak. In: Jarrige J-F and Santoni M (eds) *Fouilles de Pirak*. Paris: Diffusion de Boccard, pp.326–33.

Crawford GW (2012) Early rice exploitation in the lower Yangzi valley: What are we missing? *The Holocene* 22(6): 613–621.

Crawford GW and Shen C (1998) The origins of rice agriculture: recent progress in East Asia. *Antiquity* 72(278): 858–866.

- Deng Z, Qin L, Gao Y, Weisskopf AR, Zhang C and Fuller DQ (2015) From Early Domesticated Rice of the Middle Yangtze Basin to Millet, Rice and Wheat Agriculture: Archaeobotanical Macro-Remains from Baligang, Nanyang Basin, Central China (6700–500 BC). *PLoS ONE* 10(10): e0139885. Doi: 10.1371/journal.pone.0139885
- Douglas DH (1994) Least cost path in GIS using an accumulated cost surface and slope lines. *Cartography* 31(3): 27–51. Doi: 10.3138/D327-0323-2JUT-016M
- Edwards AWF (1992) *Likelihood*. Expanded ed. Baltimore: Johns Hopkins University Press.
- Erard M (2009) How many languages? Linguists discover new tongues in China. *Science* 324 (5925): 332–333.
- Flad R, Li S, Wu X and Zhao Z (2010) Early wheat in China: Results from new studies at Donghuishan in the Hexi Corridor. *The Holocene* 20: 955–965.
- Frachetti MD, Spengler RN, Fritz GJ and Mar'yashev AN (2010) Earliest direct evidence for broomcorn millet and wheat in the central Eurasian steppe region. *Antiquity* 84(326): 993–1010.
- Fuller DQ (2011a) Pathways to Asian civilizations: Tracing the origins and spread of rice and rice cultures. *Rice* 4(3-4): 78–92. Doi: 10.1007/s12284-011-9078-7
- Fuller DQ (2011b) Finding Plant Domestication in the Indian Subcontinent. *Current Anthropology* 52(S4): S347–S362.
- Fuller DQ (n.d.) Archaeobotany. In: Kennet D, Bai MK and VaraprasadaRao J (eds) *Transformations in early historic and early medieval India: excavations at Paithan, Maharashtra 1996–1999*. Delhi: Archaeological Survey of India, pp.289-327. [in press]

Fuller DQ and Castillo C (2016) Diversification and cultural construction of a crop: the case glutinous rice and waxy cereals in the food cultures of eastern Asia. In: J. Lee-Thorp and M. A. Katzenberg (eds) *The Oxford Handbook of the Archaeology of Diet*. Oxford University Press. doi: 10.1093/oxfordhb/9780199694013.013.8

Fuller DQ, Denham T, Arroyo-Kalin M, Lucas L, Stevens CJ, Qin L, Allaby RG and Purugganan MD (2014) Convergent evolution and parallelism in plant domestication revealed by an expanding archaeological record. *Proceedings of the National Academy of Sciences* 111(17): 6147–6152. Doi: 10.1073/pnas.1308937110

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

Fuller DQ, Qin L, Zheng Y, Zhao Z, Chen X, Hosoya LA and Sun G-p (2009) The Domestication Process and Domestication Rate in Rice: Spikelet bases from the Lower Yangtze. *Science* 323: 1607–1610. Doi: 10.1126/science.1166605

Fuller DQ, Rowlands M (2009) Towards a Long-Term Macro-Geography of Cultural Substances: Food and Sacrifice Tradition in East, West and South Asia. *Chinese Review of Anthropology* 12: 1–37.

Fuller DQ, Sato Y-I, Castillo C, Qin L, Weisskopf AR, Kingwell-Banham EJ, Song J, Ahn S-M and van Etten J (2010) Consilience of genetics and archaeobotany in the entangled history of rice. *Archaeological and Anthropological Sciences* 2 (2): 115–131. Doi: 10.1007/s12520-010-0035-y

Fuller DQ, van Etten J, Manning K, Castillo C, Kingwell-Banham EJ, Weisskopf AR, Qin L, Sato Y and Hijmans R (2011) The contribution of rice agriculture and livestock pastoralism to prehistoric methane levels An archaeological assessment. *The Holocene* 21: 743–759. Doi: 10.1177/0959683611398052

Garris AJ, Tai TH, Coburn J, Kresovich S and McCouch SR (2005) Genetic structure and diversity in

Oryza sativa L. *Genetics* 169: 1631–1638

Gorenflo LJ, Romaine S, Mittermeier RA, Walker-Painemilla K (2012) Co-occurrence of linguistic and biological diversity in biodiversity hotspots and high biodiversity wilderness areas. *Proceedings of the National Academy of Sciences* 109 (21): 8032–8037.

Grist DH (1965) *Rice*, fourth ed. London: Longman.

Gross BL and Zhao Z (2014) Archaeological and genetic insights into the origins of domesticated rice. *Proceedings of the National Academy of Sciences* 111: 6190–6197.

Guedes JDA and Butler EE (2014) Modeling constraints on the spread of agriculture to Southwest China with thermal niche models. *Quaternary International* 349: 29–41. Doi: 10.1016/j.quaint.2014.08.003

Guedes JDA, Lu H, Hein AM, Schmidt AH (2015) Early evidence for the use of wheat and barley as staple crops on the margins of the Tibetan Plateau. *Proceedings of the National Academy of Sciences* 112(18): 5625–5630. Doi: 10.1073/pnas.1423708112

Haupt RL, Haupt SE (2004) *Practical Genetic Algorithms*. New York: John Wiley and Sons.

He Z, Zhai W, Wen H, Tang T, Wang Y, Lu X, Greenberg AJ, Hudson RR, Wu C-I and Shim S (2011) Two evolutionary histories in the genome of rice: the roles of domestication genes. *PLoS Genetics* 7(6): e1002100. Doi: 10.1371/journal.pgen.1002100

Higham C (2014) *Early Mainland Southeast Asia: From First Humans to Angkor*. Bangkok: River Books.

Htun TM, Inoue C, Chhoun O, Ishii T and Ishikawa R (2014) Effect of quantitative trait loci for seed shattering on abscission layer formation in Asian wild rice *Oryza rufipogon*. *Breeding Science* 64(3): 199–205.

Hua L, Wang DR, Tan L, Fu Y, Liu F, Xiao L, Zhu Z, Fu Q, Sun X, Gu P, Cai, H, McCouch SR and Sun C (2015) LABA1, a domestication gene associated with long, barbed awns in wild rice. *The Plant Cell* 27: 1875–1888.

Huang X, Kurata N, Wei X, Wang Z-X, Wang A, Zhao Q, Zhao Y, Liu K, Lu H, Li W, Guo Y, Lu Y, Zhou C, Fan D, Weng Q, Zhu C, Huang T, Zhang L, Wang Y, Feng L, Furuumi H, Kubo T, Miyabayashi T, Yuan X, Xu Q, Dong G, Zhan Q, Li C, Fujiyama A, Toyoda A, Lu T, Feng Q, Qian Q, Li J and Han B (2012). A map of rice genome variation reveals the origin of cultivated rice. *Nature* 490: 497–501. Doi: 10.1038/nature11532

Ikehashi H (2014) Domestication and Long-Distance Dissemination of Rice: A Revised Version. *Journal Rice Research* 3: 128. Doi: 10.4172/2375-4338.1000128

Ishii T, Numaguchi K, Miura K, Yoshida K, Thanh PT, Htun TM, 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: 462–465.

Ishikawa R, Thanh PT, Nimura N, Htun TM, Yamasaki M and Ishii T (2010) Allelic interaction at seed-shattering loci in the genetic backgrounds of wild and cultivated rice species. *Genes & Genetic Systems* 85: 265–271.

Kingwell-Banham E (2015). *Early rice agriculture in South Asia: identifying cultivation systems using archaeobotany*. PhD Thesis, University College London, UK.

Koenker R (2005) *Quantile Regression*. Cambridge: Cambridge University Press.

Koenker R (2013) quantreg: Quantile Regression. R package version 5.05. Available at: <http://CRAN.R-project.org/package=quantreg> (accessed 16 February 2016).

- Kohler-Schneider M and Canepelle A (2009) Late Neolithic agriculture in eastern Austria: archaeobotanical results from sites of the Baden and Jevišovice cultures (3600–2800 B.C.). *Vegetation History and Archaeobotany* 18: 61–74.
- Kovach MJ, Sweeney MT and McCouch SR (2007) New insights into the history of rice domestication. *Trends in Genetics* 23(11): 578–587.
- Kwon S-J, Lee JK, Hong S-W, Park Y-J, McNally KL and Kim N-S (2006) Genetic diversity and phylogenetic relationship in AA *Oryza* species as revealed by Rim2/Hipa CACTA transposon display. *Genes & Genetic Systems* 81: 93–101.
- Larson G, Liu R, Zhao X, Yuan J, Fuller DQ, Barton L, Dobney K, Fan Q, Gu Z, Liu Y, Luo X, Lu P, Andersson L and Li N (2010) Patterns of East Asian pig domestication, migration, and turnover revealed by modern and ancient DNA. *Proceedings of the National Academy of Sciences* 107 (17): 7686–7691.
- Li C, Zhou A and Sang T (2006) Rice domestication by reducing shattering. *Science* 311: 1936–1939
- Londo JP, Chiang Y-C, Hung K-H, Chiang T-Y and Schaal BA (2006) Phylogeography of Asian wild rice, *Oryza rufipogon*, reveals multiple independent domestications of cultivated rice, *Oryza sativa*. *Proceedings of the National Academy of Sciences* 103(25): 9578–9583.
- McGinley M (2008) Three Parallel Rivers of Yunnan Protected Areas, China. In: *The Encyclopedia of Earth*. Available at: <http://www.eoearth.org/view/article/156617> (accessed 29 December 2015).
- McNally KL, Childs KL, Bohnert R, Davidson RM, Zhao K, Ulat VJ, Zeller G, Clark RM, Hoen DR, Bureau TE, Stokowski R, Ballinger DG, Frazer KA, Cox DR, Padhukasahasram B, Bustamante CD, Weigel D, Mackill DJ, Bruskiewich RM, Ratsch G, Buell CR, Leung H, Leach JE (2009) Genomewide SNP variation reveals relationships among landraces and modern varieties of rice. *Proceedings of the National Academy of Sciences* 106 (30): 12273–12278.

Metspalu M, Kivisild T, Metspalu E, Parik J, Hudjashov G, Kaldma K, Serk P, Karmin M, Behar DM, Gilbert MTP, Endicott P, Mastana S, Papiha SS, Skorecki K, Torroni A, VILLEMS R (2004) Most of the extant mtDNA boundaries in South and Southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans. *BMC Genetics* 5: 26.

Mishra BP, Dubey PK, Prakash B, Kathiravan P, Goyal S, Sadana DK, Das GC, Goswami RN, Bhasin V, Joshi BK, Kataria RS (2015) Genetic analysis of river, swamp and hybrid buffaloes of north-east India throw new light on phylogeography of water buffalo (*Bubalus bubalis*). *Journal of Animal Breeding and Genetics* 132: 454–466.

Molina J, Sikora M, Garud N, Flowers JM, Rubinstein S, Reynolds A, Huang O, Jackson S, Schall BA, Bustamante CD, Boyko AR and Puruggana MD (2011) Molecular evidence for a single evolutionary origin of domesticated rice. *Proceedings of the National Academy of Sciences* 108: 8351–6.

Murphy C, Weisskopf, A, Bohingamuwa W, Adikari G, Perera N, Blinkhorn J, Horton M, Fuller DQ and Boivin N (In Review) Early Agriculture in Sri Lanka: New Archaeobotanical Analyses and Radiocarbon Dates from the Early Historic Sites of Kirinda and Kantharodai. *Archaeological Research in Asia* [submitted]

Ohtsubo H, Cheng C, Ohsawa I, Tsuchimoto S and Ohtsubo E (2004) Rice retroposon p-SINE1 and origin of cultivated rice. *Breeding Science* 54: 1–11.

Oka H (1988) *Origins of cultivated rice*. Amsterdam: Elsevier.

Olson DM, Dinerstein E, Wikramanayake ED, Burgess ND, Powell GV, Underwood EC, D'Amico JA, Itoua I, Strand HE, Morrison JC, Loucks CL, Allnutt TF, Ricketts TH, Kura Y, Lamoreux JF, Wettengel WW, Hedao P and Kaseem KR (2011) Terrestrial ecoregions of the world: a new map of life on earth. *BioScience* 51: 933–938.

R Core Team (2014) R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Available: <http://www.R-project.org/>.

Rahman M, Castillo CC, Murphy C, Rahman SM and Fuller DQ (n.d.) Early agricultural systems in Bangladesh: archaeobotanical flotation results from Early Historic Wari-Bateshwar and Early Medieval Vikrampura. *Archaeological and Anthropological Sciences* [submitted]

Rassamakin Y (1999) The Eneolithic of the Black Sea Steppe: Dynamics of Cultural and Economic Development 4500-2300 BC. In: Levine M, Rassamakin Y, Kislenko A and Tatarintseva N (eds) *Late Prehistoric exploitation of the Eurasian Steppe*. Cambridge: McDonald Institute, pp. 59–182.

Sagart L (2011) How many independent rice vocabularies in Asia?. *Rice* 4(3): 121–133.

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

Sang T and Ge S (2007) The puzzle of rice domestication. *Journal of Integrative Plant Biology* 49(6): 760–768

Schatz MC, Maron LG, Stein JC, Wences AH, Gurtowski J, Biggers E, Lee H, Kramer M, Antoniou E, Ghiban E, Wright MH, Chia J-m, Ware D, McCouch SR and McCombie WR (2014) Whole genome de novo assemblies of three divergent strains of rice, *Oryza sativa*, document novel gene space of aus and indica. *Genome Biology* 15: 506.

Scrucca L (2013) GA: A Package for Genetic Algorithms in R. *Journal of Statistical Software* 53(4): 1–37. Available: <http://www.jstatsoft.org/v53/i04/>.

- Sethian JA (1996) A fast marching level set method for monotonically advancing fronts. *Proceedings of the National Academy of Sciences* 93(4): 1591–1595.
- Silva F and Steele J (2012) Modeling boundaries between converging fronts in prehistory. *Advances in Complex Systems* 15(01n02): 1150005. Doi: 10.1142/S0219525911003293
- Silva F and Steele J (2014) New methods for reconstructing geographical effects on dispersal rates and routes from large-scale radiocarbon databases. *Journal of Archaeological Science* 52: 609–620. Doi: 10.1016/j.jas.2014.04.021
- Silva F, Stevens CJ, Weisskopf A, Castillo C, Qin L, Bevan A and Fuller DQ (2015) Modelling the geographical origin of rice cultivation in Asia using the Rice Archaeological Database. *PLoS ONE* 10(9): e0137024. Doi: 10.1371/journal.pone.0137024
- Simoons F (1970) The traditional limits of milking and milk use in southern Asia. *Anthropos* 65: 547–593.
- Spengler RN (2015) Agriculture in the Central Asian Bronze Age. *Journal of World Prehistory* 28: 215–253
- Steele J (2010) Radiocarbon dates as data: quantitative strategies for estimating colonization front speeds and event densities. *Journal of Archaeological Science* 37(8): 2017–2030. Doi: 10.1016/j.jas.2010.03.007
- Stevens CJ, Murphy C, Roberts R, Lucas L, Silva F and Fuller DQ (2016) Between China and South Asia: A Middle Asian corridor of crop dispersal and agricultural innovation in the Bronze Age. *The Holocene* 26(10): 1541–1555. Doi: 10.1177/0959683616650268.
- Sweeney MT, Thomson MJ, Cho YG, Park YJ, Williamson SH, Bustamante CD and McCouch SR (2007) Global Dissemination of a Single Mutation Conferring White Pericarp in Rice. *PLoS*

Genetics 3(8): e133. Doi: 10.1371/journal.pgen.0030133

Takano-Kai N, Jiang H, Kubo T, Sweeney M, Matsumoto T, Kanamori H, Padhukasahasram B, Bustamante C, Yoshimura A, Doi K and McCouch S (2009) Evolutionary history of GS3, a gene conferring grain length in rice. *Genetics* 182: 1323–1334. Doi: 10.1534/genetics.109.103002

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

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 Physiology* 135: 412–20

Tan L, Li X, Liu F, Sun X, Li C, Zhu Z, Fu Y, Cai H, Wang X, Xie D and Sun C (2008) Control of a key transition from prostrate to erect growth in rice domestication. *Nature Genetics* 40: 1360–1364.

Tewari R, Srivastava RL, Saraswat KS, Singh IB and Singh KK (2008) Early Farming at Lahuradewa. *Pradghara* 18: 347–373.

Tewari R, Srivastava RL, Singh KK, Saraswat KS, Singh IB, Chauhan MS, Pokharia AK, Saxena A, Prasad V, Sharma M (2006) Second Preliminary Report of the excavations at Lahuradewa District Sant Kabir Nagar, U.P. 2002-2003-2004 & 2005-06. *Pragdhara* 16: 35–68.

Travis AJ, Norton GJ, Datta S, Samar R, Dasgupta T, Savio FL, Macaulay M, Hedley PE, McNally KL, Sumon MH, Islam MR and Price AH (2015) Assessing the genetic diversity of rice originating from Bangladesh, Assam and West Bengal. *Rice* 8: 35.

Vaughan DA, Lu B and Tomooka N (2008) Was Asian rice (*Oryza sativa*) domesticated more than once? *Rice* 1: 16–24.

Vitte C, Ishii T, Lamy F, Brar D, Panaud O (2004) Genomic paleontology provides evidence for two

- distinct origins of Asian rice (*Oryza sativa* L.). *Molecular Genetics and Genomics* 272: 504–511.
- Yang C-C, Kawahara Y, Mizuno H, Wu J, Matsumoto T and Itoh T (2012) Independent domestication of Asian rice followed by gene flow from japonica to indica. *Molecular Biology and Evolution* 29(5): 1471–1479. Doi: 10.1093/molbev/msr315
- Yuan J and Campbell R (2009) Recent archaeometric research on ‘the origins of Chinese civilisation’. *Antiquity* 83 (319): 96-109.
- Zeng Y, Zhang H, Li Z, Shen S, Sun J, Wang M, Liao D, Liu X, Wang X, Xiao F and Wen G (2007) Evaluation of genetic diversity of rice landraces (*Oryza sativa* L.) in Yunnan, China. *Breeding Science* 57: 91–99.
- Zhang Y, Vankan D, Zhang Y and Barker JSF (2011) Genetic differentiation of water buffalo (*Bubalus bubalis*) populations in China, Nepal and south-east Asia: inference on the region of domestication of swamp buffalo. *Animal Genetics* 42: 366–377.
- Zhao Z (2015) Study of how wheat was transmitted into China—based on archaeobotanic data. *Nanfang Wenwu [Relics from the South]* 2015(3): 44-52. [in Chinese].
- Zhao K, Tung C-W, Eizenga GC, Wright MH, Ali ML, Price AH, Norton GJ, Islam MR, Reynolds A, Mezey J, McClung AM, Bustamante CD and McCouch SR (2011) Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nature Communications* 2: 467. Doi: 10.1038/ncomms1467
- Zhu Y, Ellstrand NC, Lu B-R (2012) Sequence polymorphisms in wild, weedy, and cultivated rice suggest seed-shattering locus *sh4* played a minor role in Asian rice domestication. *Ecology and Evolution* 2 (9): 2106–2113.