



The spread of domesticated rice in eastern and southeastern Asia was mainly demic[☆]

José M. Cobo^a, Joaquim Fort^{a,b,*}, Neus Isern^a

^a Complex Systems Laboratory, University of Girona, C/. Maria Aurèlia Capmany 61, Girona, Catalonia, 17003, Spain

^b Catalan Institution for Research and Advanced Studies (ICREA), C/. Lluís Companys 23, Barcelona, Catalonia, 08010, Spain

ARTICLE INFO

Keywords:

Domesticated rice
Spread rate
Demic diffusion
Cultural diffusion

ABSTRACT

The Neolithic transition, i.e., the shift from hunting and gathering into farming, had a major impact in many aspects of human societies, from economics to demography and from health to ideology. There are two main models of Neolithic spread. The demic model assumes that the Neolithic spread mainly due to the diffusion of farming populations, whereas the cultural model considers that it was essentially due to transmission of cultural traits (domesticates and knowledge) from farmers to hunter-gatherers (without substantial diffusion of farmers themselves). Here we estimate the spread rate of the Neolithic transition in eastern and southeastern Asia, using Early Neolithic dates of 201 archaeological sites with domesticated rice (*Oryza sativa*). We show that domesticated rice, a staple Neolithic crop in eastern and southeastern Asia, spread at a rate of 0.72–0.92 km/yr (95% confidence level). Comparing these results to the predictions of a demic-cultural model implies that demic diffusion explains more than 76% of the spread observed rate, whereas cultural diffusion played a secondary role.

1. Introduction

In eastern, southeastern and southern Asia, rice is a staple cereal and its cultivation began (probably in several regions) about 8000–6000 yr BC, with fixation of domestication traits (non-shattering spikelet bases, large grain sizes, etc.) closer to 4000 yr BC (Fuller, 2011). At the latter time, rice field systems were established in the lower and middle Yangtze river basins (China), from where domesticated rice spread to the whole area. A difference with other Neolithic transitions is that rice was the main staple cereal that transformed most of this huge region from a land of hunter-gatherers into a continent of farmers. An exception is northern China, where millet arrived before rice (Fuller, 2011). In this paper we focus on the spread of rice, for which a detailed database exists.

A longstanding question is whether the spread of domesticated rice was mainly demic (i.e., due to the dispersal of rice cultivators) or primarily cultural (i.e., due to the incorporation of hunter-gatherers into the populations of rice cultivators). Some scholars have argued that the diffusion of domesticated rice was mainly demic (Higham, 2002; Bellwood, 2005), others have advocated for the importance of cultural diffusion (Kudo, 2000; Takamiya, 2007; Barton, 2012), and it is also

possible that both demic and cultural diffusion had some importance (Fuller, 2011). However, the relative importance of demic and cultural diffusion has been never estimated quantitatively for the spread of domesticated rice in southeastern Asia. Here we tackle this question, at the continental level, by performing a quantitative analysis of the dates of the Rice Archaeological Database by Silva et al. (2015) (see Supp. Info., S1 Table to the present paper). As we shall discuss (Sec. 4), ancient human genetic data (Lipson et al., 2018; McColl et al., 2018) are not yet numerous enough to provide an independent, quantitative answer to this question.

In Sec. 3.1 we will estimate the spread rate of domesticated rice. In Sec. 3.2 we will resort to a two-dimensional mathematical model (Fort, 2012) that combines cultural and demic diffusion and is based on cultural transmission (Cavalli-Sforza and Feldman, 1981) and front propagation (Fort and Pujol, 2008) theories. In this model, the cultural transmission intensity can be defined as the number of hunter-gatherers becoming farmers, per pioneering farmer and generation (see Supp. Info., Sec. S3 for details). The higher this cultural transmission intensity, the faster the spread rate predicted by the model (for given demic parameter values). This is very reasonable intuitively, because it is expected that if there are more farmers, their wave of advance will

[☆] This paper is dedicated to the memory of Luigi Luca Cavalli-Sforza (1922–2018), whose pioneering work on Neolithic wave-of-advance models was the starting point of the research reported in our previous works and this paper.

* Corresponding author. Complex Systems Laboratory, University of Girona, C/. Maria Aurèlia Capmany 61, Girona, Catalonia, 17003, Spain.

E-mail address: joaquim.fort@udg.edu (J. Fort).

spread faster. Therefore, for given demic diffusion parameter values, a slower wave of advance corresponds to a smaller importance of cultural diffusion. This point will be useful to interpret several results in the present paper.

2. Materials and methods

2.1. Database

In this study we use the dates of archaeological sites in China, Myanmar, Thailand, Vietnam, Bangladesh, Taiwan, South Korea and Japan reported in the Rice Archaeological Database (Silva et al., 2015), see [Supp. Info. S2 Table](#) to the present paper.

We do not include India, Pakistan and Sri Lanka because rice genomic data (Wang et al., 2018) support an independent domestication of rice in this region (*O. sativa* Indica) rather than just introgression from Chinese rice (*O. sativa* Japonica). This is also consistent with the fact that including those regions would lead to a poor value of the correlation coefficient ($r = 0.55$), so that assuming a linear dependence (to estimate an average spread rate) would not be justified ([Supp. Info., Sec. S1a](#)). The reason for this low value of r is the presence of old regions in the Indian subcontinent ([Fig. S1a](#)), which suggest independent domestication processes (in agreement with the genomic results summarized above). It would be interesting to estimate the spread rates of those processes, but unfortunately this is not possible with statistical confidence using the database available at present because the values of r would be even poorer, namely $r < 0.5$ ([Supp. Info., Sec. S1b](#)).

We do not include the sites in the Mariana Islands (U.S.), the Philippines, Indonesia and Malaysia either, because they are separated from the rest by huge areas of ocean. This implies long sea travels, and modelling such a process requires different dispersal equations and parameter values ([Isern et al., 2017b](#)).

The area of the Himalayas cannot be included, because of the lack of data in that region. Similarly, in Laos and Cambodia there are only 9 sites. They lead to anomalously early regions, which will probably disappear when more sites are discovered and dated. None of these 9 dates were obtained by direct AMS dating on rice itself (2 were dated by associated radiocarbon dates from the site considered, and 7 by cultural association). We do not include these 9 sites either. However, if we included them, our conclusions would not change (although the statistical confidence would be lower, see [Supp. Info., Sec. S1c](#)).

Rice domestication came to an end (full domestication) after a millennium or more of pre-domestication cultivation, and presumably a much longer period of wild rice use by foragers (Fuller et al., 2007). Domestication is a series of changes in the plant, relative to its wild type, whereas cultivation is a human activity that includes sowing and harvesting the plant before it has been fully domesticated. The three main traits that evolved under cultivation and finally led to domesticated rice are: (i) reduction in awns and hairs (they help the shed spikelet grip the soil, but this becomes less necessary as humans begin to plant seeds); (ii) increase in grain size; and (iii) evolution of a non-shattering spikelet base, which makes it possible for domesticated rice to retain its grains at maturity. In contrast, non-domesticated rice gradually loses part of its grains as it comes into maturity, and for this reason foragers harvest it early. Thus cultivated rice contains high proportions of immature spikelets (which can be identified due to their larger length-to-width ratios). This differentiates cultivated rice from wild rice. Another difference is that, as explained above, sowing leads to a reduction of awn hairs in cultivated rice, as compared to wild rice. The densities of awn hairs are still lower in domesticated rice (Fuller et al., 2007).

Since we are interested in the arrival of rice, we consider only the oldest phase for each archaeological site. We use only those phases for which both cultivation and domestication were reported. In this way, we observed that some extremely old sites became excluded, because no domesticated rice has been found in them. This is one reason why we

do not attempt to analyze the spread of cultivated rice before domestication. A second reason is that there are less data than for domesticated rice. A third reason is based on the fact that there were likely multiple origins of rice cultivation (Fuller, 2011). Fourthly, archaeological data suggest little or no spread of rice before domestication (Silva et al., 2015, 2018). All of these arguments make it reasonable to focus on the spread of domesticated (not just cultivated) rice. This is one of the differences with the approach by Silva et al. (2015, 2018); other differences are included in our [Supp. Info., Sec. S4](#).

The process described in the previous paragraphs reduced the number of sites (or phases) to 201. The resulting database, which has been analyzed in the present paper, is included as [Supp. Info. S2 Table](#). It contains 201 archaeological sites, their mean dates, coordinates, distances to the two oldest sites, namely Chengtoushan and Hemudu, both of them in China, and additional information for each site (for details on the database, see [Supp. Info., Sec. S1](#)). We mention that if we excluded all phases dated by cultural association, the main conclusion of this paper would not change ([Sec. S1f](#)), although the statistical validity and level of detail in the geostatistical maps would decrease.

2.2. Geostatistical analyses

We visualize the spread of domesticated rice by plotting the 201 sites on a map and interpolating their archaeological dates using Geographic Information System (GIS) software. We report results obtained using the natural neighbor interpolation technique (Mitas and Mitasova, 1999), but other techniques yield similar results. As in previous work on the spread of the Neolithic in Europe (Fort, 2015), we also perform a smoothing technique (see [Supp. Info. Sec. S1](#) to the present paper). Computing the local slope and orientation (downslope direction) of the interpolated surface at different locations, we find out the vectors giving the spread directions of the wave of advance of domesticated rice, which we represent on the map as arrows.

2.3. Front speed

For the spread of agriculture in Europe, estimating rates in different regions has been very useful. For example, the faster spread implied by the archaeological data along the Mediterranean has made it possible to estimate the distances of sea travel covered per generation by the first farmers (Isern et al., 2017b). Similarly, for the spread of domesticated rice in eastern and southeastern Asia it would be interesting to estimate local rates, or at least rates in several regions. Unfortunately this is not possible with the data available at present ([Supp. Info., Secs. S1d-B and S1b](#), respectively). Nevertheless, the estimation of an average rate is of importance for two reasons. First, it provides a methodology that will be useful to perform regional analyses when the number of data becomes large enough. And second, we shall use the average rate to analyze quantitatively the relative importance of demic and cultural diffusion at the continental level.

We perform time-space linear regressions with distances being measured as great circles from one of the oldest dates in the database, as in previous work (Fort et al., 2004b; Jerardino et al., 2014). We also need a distance origin. For the Neolithic spread across Europe, it has been found that using as distance origin any of the oldest sites yields almost the same results than using the node in a lattice that yields the highest value of r (Ammerman and Cavalli-Sforza, 1971; Pinhasi et al., 2005). Thus, as in recent work on Europe (Fort et al., 2012; Isern et al., 2017b) and Africa (Jerardino et al., 2014), here we will report results using one of the two oldest sites (Chengtoushan in the middle Yangtze and Hemudu in the lower Yangtze) as possible origins. We checked that using other old sites (or nearby points) as origins yields the same conclusions.

As usual (Fort et al., 2004b; Jerardino et al., 2014; Isern et al., 2017b), each regression is calculated by plotting the calibrated dates of sites versus their distances from the presumed source, i.e. we plot times

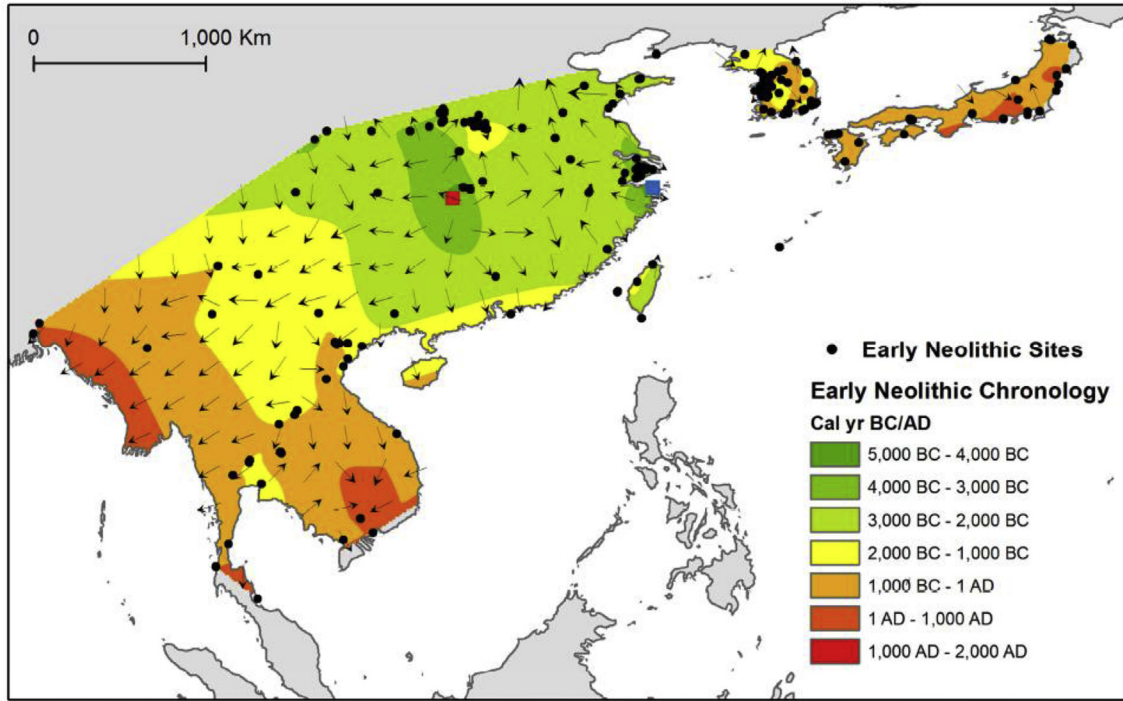


Fig. 1. Map of arrival times of the Neolithic transition in southeastern Asia. Map obtained from a natural neighbor interpolation of 201 early Neolithic sites (symbols) in southeastern Asia (see S2 Table) and smoothing 20 times. We also show the orientations of the local speed vectors (front propagation directions). Color regions correspond to the area advanced by the Neolithic wave every 1000 years. The red square is the oldest site in the Middle Yangtze River basin (Chengtoushan, China, dated 4250 BC), which is the origin of distances in the regression in Fig. 2. The blue square is the oldest site in the Lower Yangtze (Hemudu, China, dated 4800 BC). All dates are in calibrated years BC/AD. Map prepared with ESRI ArcGIS version 10 (<http://www.esri.com/>) using ArcMap and the ArcGIS Spatial Analyst extension.

versus distances (and not the other way round) because dates are affected by several errors, especially those that have been obtained by cultural association, whereas distances between sites are in principle known with more accuracy (Fort et al., 2004b). The great-circle distance between two locations i and j can be obtained from their geographical coordinates (latitude ϕ and longitude λ) using the Haversine equation (Sinnott, 1984),

$$d_{ij} = 2R \arcsin \left(\sqrt{\sin^2 \left(\frac{\phi_i - \phi_j}{2} \right) + \sin^2 \left(\frac{\lambda_i - \lambda_j}{2} \right) \cos(\phi_i) \cos(\phi_j)} \right) \quad (1)$$

where R is the average value of the Earth radius ($R=6371$ km).

As in previous work (Fort et al., 2004b; Jerardino et al., 2014), the speed is obtained from the slope of the time-space regression as follows

$$speed = \frac{1}{slope}, \quad (2)$$

and the standard error of this speed is obtained by applying error propagation theory (Taylor, 1997). This yields

$$\sigma_{speed} = \frac{\sigma_{slope}}{(slope)^2}, \quad (3)$$

where σ_{slope} is standard error of the slope. We have used Eqs. (1)–(3) to compute the 95% confidence-level (CL) interval for the speed of the spread of domesticated rice, i.e. the range ($speed - t\sigma_{speed}$, $speed + t\sigma_{speed}$), where t is Student's t-distribution for a 95% CL and $N - 2$ degrees of freedom, and N is the number of sites. As we shall see below, in the present paper $N \geq 185$, so $t \approx 2$ (Spiegel et al., 1975; Draper and Smith, 1981).

As an alternative approach, instead of using great-circle distances (Eq. (2)), we estimated cost distances that take into account the effects of topography, vegetation, etc., on the local direction of the wave of advance. The conclusions do not change (see Supp. Info., Sec. S1g for

details).

2.4. Demic-cultural diffusion model

The model that we shall apply to estimate the relative importance of demic and cultural diffusion is based on the following two equations for the population densities of farmers N (rice agriculturalists in our case) and hunter-gatherers P at position (x, y) and time t (Fort, 2012),

$$\begin{cases} N(x, y, t + T) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \tilde{N}(x + \Delta_x, y + \Delta_y, t) \phi_N(\Delta_x, \Delta_y) d\Delta_x d\Delta_y \\ P(x, y, t + T) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \tilde{P}(x + \Delta_x, y + \Delta_y, t) \phi_P(\Delta_x, \Delta_y) d\Delta_x d\Delta_y \end{cases} \quad (4a)$$

where T is the generation time, defined as the mean age difference between a parent and her/his children (and assumed to be approximately the same for both populations). $\phi_l(\Delta_x, \Delta_y)$ is the dispersal kernel of population $l=N, P$, defined as the probability to move distances (Δ_x, Δ_y) per generation. We have also introduced

$$\begin{aligned} \tilde{N}(x, y, t) &\equiv R_T [N(x, y, t)] + f \frac{R_T [N(x, y, t)] R_T [P(x, y, t)]}{R_T [N(x, y, t)] + \gamma R_T [P(x, y, t)]} \\ \tilde{P}(x, y, t) &\equiv R_T [P(x, y, t)] - f \frac{R_T [N(x, y, t)] R_T [P(x, y, t)]}{R_T [N(x, y, t)] + \gamma R_T [P(x, y, t)]} \end{aligned} \quad (4b)$$

where $R_T [N(x, y, t)] = \frac{e^{a_N T} K_N N(x, y, t)}{K_N + (e^{a_N T} - 1) N(x, y, t)}$ and $R_T [P(x, y, t)] = \frac{e^{a_P T} K_P P(x, y, t)}{K_P + (e^{a_P T} - 1) P(x, y, t)}$ are the new population densities due to logistic net reproduction (with initial growth rates a_l and carrying capacities K_l) during the time interval T . The last term in Eq. (4b) corresponds to cultural transmission, and is driven by parameters f and γ (for a detailed derivation of Eq. (4b) and the interpretation of these parameters, see the Supp. Info., Sec. S3). Purely demic diffusion corresponds to $f = 0$. This framework is the result of work on wave-of-advance models

during the last two decades, and it improves the classical approach due to Fisher (Ammerman and Cavalli-Sforza, 1984) in three main ways: (i) Eqs. (4a) take into account that newborn humans spend some time with their parents before they can survive on their own and reproduce (Fort and Méndez, 1999; Fort et al., 2007); (ii) Eqs. (4a) also take into account the detailed dispersal kernel (probability versus dispersal distance) rather than only an approximation given by the diffusion coefficient (Isern et al., 2008); and (iii) Eqs. (4b) include cultural transmission in addition to demic diffusion (Fort, 2012). Equations (4a–b) are valid in general, not only for specific crops or Neolithic transitions. They can be applied to model the spread of any cultural trait due to demic and, possibly, cultural diffusion.

3. Results

3.1. Geostatistical analyses and front speed

Fig. 1 is an interpolation map of the dates of 201 sites (S2 Table). Each color corresponds to a 1000-yr interval for the estimated arrival time of domesticated rice. We see that the oldest sites (darker green areas) are located at the northeast, in China. In particular, the two oldest regions are located at the Lower and Middle Yangtze River basins. This agrees with the two most likely regions of origin of cultivated rice previously identified by other authors (Fuller and Qin, 2009; Silva et al., 2015; Fuller et al., 2016).

3.1.1. Database of 201 sites

As mentioned above, there are two very ancient areas in China (darker green areas in Fig. 1). One is located at the Middle Yangtze River valley and its oldest site is Chengtoushan (4250 BC, red square in Fig. 1). The other one is located at the Lower Yangtze River basin and its oldest site is Hemudu (4800 BC, blue square in Fig. 1). In spite of the fact that Hemudu is the oldest site, we shall use Chengtoushan as the origin of distances, since it yields a much better correlation coefficient, namely $r = 0.76$ (versus $r = 0.50$ if assuming Hemudu as origin). This is reasonable for two reasons. First, $r = 0.50$ is such a poor value that the assumption of linearity is not justified, so an approximately constant spread rate cannot be estimated with confidence. Second, visual examination of the routes of spread indicated by the arrows in the interpolation map in Fig. 1 also suggests that the region of the site of Chengtoushan (red square in Fig. 1) is a likely area of origin of the main expansion of domesticated rice over the continent.

Fig. 2 is the linear regression fit for the 201 sites (S2 Table and Fig. 1), using distances from Chengtoushan. The spread rate implied by the slope is 0.74 ± 0.09 km/yr at the 95% CL, i.e. a speed range of 0.65–0.83 km/yr (but this result is refined in the next paragraph) and $r = 0.76$.

3.1.2. Database of 185 sites

Taking into consideration that the estimated rate of spread can be affected by including early dates from regions in China where the domestication of rice may have arisen independently, we can improve (and check) the estimate in the previous subsection. Consider excluding Hemudu (blue square in Fig. 1), as well as other sites near it. In this way, we expect to obtain a still better estimate of the speed of the main spread coming from the region of Chengtoushan (red square in Fig. 1). Hemudu and the 15 sites nearby are shown as blue squares in Fig. 2. Note that they are all below the regression to the complete set of 201 sites, which suggests that they correspond to an old, independent spread of domesticated rice. The linear regression using only these 185 sites (S3 Table), again with distance origin in Chengtoushan, yields a still higher correlation coefficient ($r = 0.78$) and a spread rate of 0.82 ± 0.10 km/yr (95% CL), i.e., a speed range of 0.72–0.92 km/yr (Fig. S4). Thus we conclude that, according to the evidence available at present, domesticated rice spread across southeastern Asia from the Middle Yangtze River basin with an average speed in the range

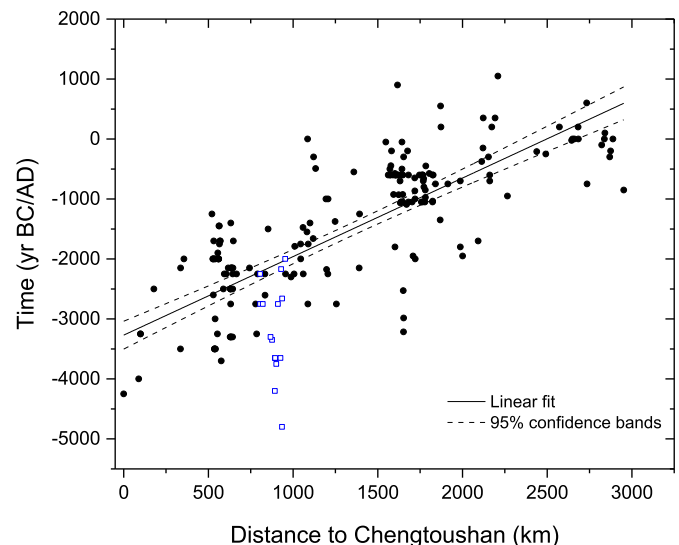


Fig. 2. Linear regression fit to the 201 sites in Fig. 1 and Fig. S2 Table. Blue squares correspond to 16 sites around Hemudu (China, dated 4800 BC). In this figure, there are 16 blue squares but a few of them cannot be distinguished. Distances have been computed as great-circle distances from Chengtoushan (China, dated 4250 BC). The solid line corresponds to the linear regression fit to all 201 sites, and the dashed lines to the 95% confidence bands. The front speed (inverse of the slope) is 0.65–0.83 km/yr (95% CL interval) and $r = 0.76$. If the blue squares are excluded ($N = 185$ sites), the corresponding linear regression (Fig. S4) yields $r = 0.78$ and 0.72–0.92 km/yr (95% CL interval), which we call the 'observed spread rate'.

0.72–0.92 km/yr (95% CL). We shall refer to this range as the 'observed spread rate'. In fact, the exclusion of the 16 sites in the Hemudu region is not necessary, because including them would lead (as seen above) to a speed about 10% slower, and this would only strengthen the final conclusion of this paper, namely that demic diffusion was more important than cultural diffusion (see the end of our Introduction). Thus, it is not really necessary to exclude the 16 sites close to Hemudu. But the analysis without these 16 sites is interesting for two reasons: (i) it has allowed us to check that their effect is small; and (ii) it yields a higher value of r , which increases the statistical significance of the results.

3.1.3. Dual-origin model

As a further check of our results, we consider an alternative approach based on assuming two spatial origins, i.e. by taking both old regions in Fig. 1 explicitly into account. A crucial problem with dual-origin models is how to assign each site to one of the two origins or sources. We cannot be absolutely sure on all such assignments, but we can use reasonable criteria to find out, for each site, from which of the two sources it is more likely that domesticated rice arrived. Obviously, if domesticated rice spread as two fronts, one from each source, the source of the front that arrived earlier at each site will be that from where domesticated rice arrived at the site considered. And if both sources have similar ages, as it is indeed our case (Fig. 1), and the speeds of both fronts are similar (which seems very reasonable because both populations are farmers, their staple food is rice, and we are not aware of important differences in climate, demography, etc. between both regions), then the source from where domesticated rice reached each site will presumably be the nearest of both sources to the site considered. We admit that we cannot assure that this procedure will identify the right source for all sites, but we expect that it will do so for most sites. Thus we assign each site (S2 Table) to the source that is nearest to it (Chengtoushan or Hemudu).

Fig. 3a is a regression for the sites assigned to Chengtoushan (i.e., those with distance to Chengtoushan smaller than to Hemudu). Its slope

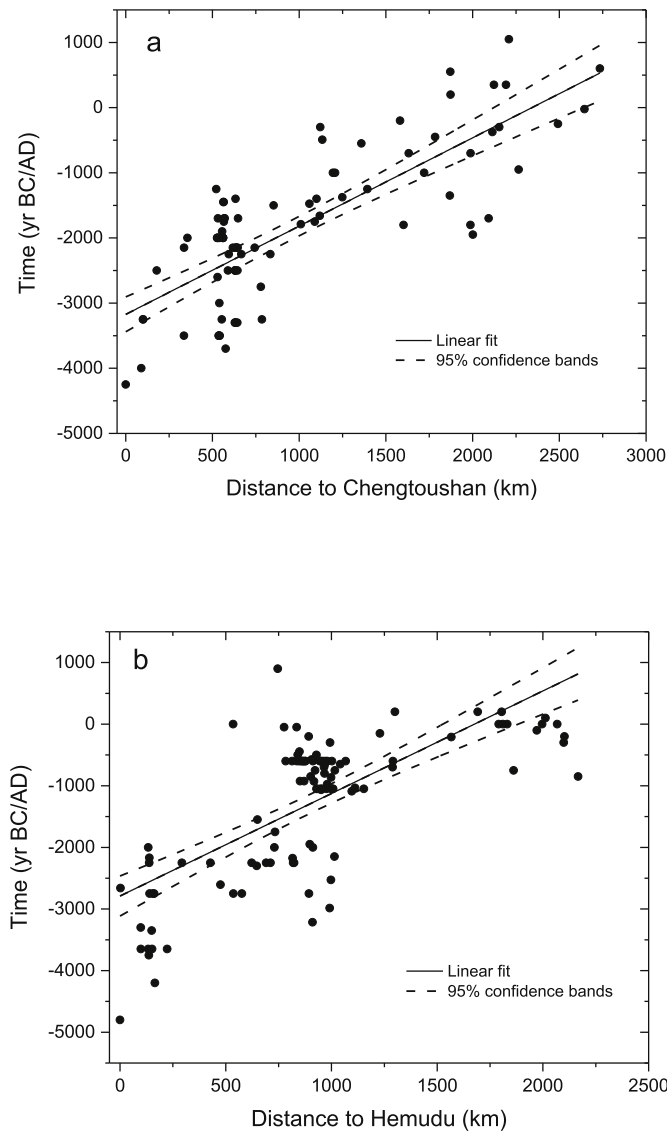


Fig. 3. Dual-source model (Sec. 3.1.3). (a), Regression of sites nearer to Chengtoushan than to Hemudu. (b), Regression of sites nearer to Hemudu than to Chengtoushan. The corresponding speed ranges are 0.62–0.86 km/yr ($r = 0.78$) and 0.48–0.72 km/yr ($r = 0.71$), respectively (95% confidence-level intervals).

implies a speed of 0.74 ± 0.12 km/yr or 0.62–0.86 km/yr (95% CL), with $N = 90$ sites and $r = 0.78$, which is a rather high correlation. This speed range is similar to the speed obtained in Sec. 3.1.2, using also Chengtoushan as distance origin (namely 0.72–0.92 km/yr, with $r = 0.78$). This gives confidence to the results.

Fig. 3b is a time-distance regression for the sites assigned to Hemudu (i.e., those with distance to Hemudu smaller than to Chengtoushan). The corresponding slope implies a speed of 0.60 ± 0.12 km/yr or 0.48–0.72 km/yr (95% CL), with $N = 111$ sites and $r = 0.71$. This is a lower correlation than that for the 90 sites assigned to Chengtoushan (namely, $r = 0.78$), but it is quite high.

We note that the two speed ranges above (0.62–0.86 km/yr or 0.48–0.72 km/yr) overlap, in agreement with the assumption above that both fronts had similar speeds. The intercepts (3442–2904 yr BC and 3117–2461 yr BC, 95% CL) also overlap, again in agreement with the assumption above that both waves of advance left each source at similar times.

The most important point is that both of the two speed ranges above (0.62–0.86 km/yr or 0.48–0.72 km/yr) are similar but slower than the

'observed spread range' obtained in Sec. 3.1.2 (0.72–0.92 km/yr). This is relevant because, as mentioned at the end of our Introduction, a slower wave of advance corresponds to a weaker cultural diffusion effect (Fort, 2012). Thus, the cultural effect will be surely smaller for the two slower speeds (obtained from the two-source model) than for the 'observed spread range' of 0.72–0.92 km/yr (obtained from the single-source model in Sec. 3.1.2). Therefore, using the dual-origin would not change the main conclusion which we shall obtain in Sec. 3.2, i.e. that demic diffusion was more important than cultural diffusion.

3.2. Demic versus cultural diffusion

Using the observed spread rate (Sec. 3.1.2), we can tackle the question of whether the spread of domesticated rice was mainly demic or mainly cultural. It has been shown that the model given by Eqs. (4a)–(4b) leads to the following wave-of-advance speed (Fort, 2012)

$$s = \min_{\lambda > 0} \frac{a_N T + \ln \left[(1 + C) \left(\sum_{j=1}^M p_j I_0(\lambda r_j) \right) \right]}{T \lambda}, \quad (5)$$

where $C = \frac{f}{\gamma}$ is the intensity of cultural transmission. It is equal to the mean number of hunter-gatherers converted into rice farmers by each pioneering rice farmer (Supp. Info., Sec. S3). The other quantities have been defined below Eqs. (4a–b). The summation in Eq. (5) takes into account the dependence of the front speed on the migration probability of farmers as a function of distance, i.e. the dispersal kernel ϕ_N in Eq. (4a). In the ethnographic literature, such dependences are usually reported using histograms, and for this reason p_j is defined as the probability for rice cultivators to disperse a distance r_j , which is the mean distance of the corresponding histogram bin, and each bin is identified by the index $j = 1, 2, \dots, M$ (see Supp. Info., Sec. S2). Finally, in Eq. (5) the following function appears,

$$I_0(\lambda r_j) = \frac{1}{2\pi} \int_0^{2\pi} d\theta \exp[-\lambda r_j \cos \theta], \quad (6)$$

which is the modified Bessel function of the first kind and order zero. For a detailed derivation of Eq. (5), see Fort (2012). We use realistic values of a_N , T , p_j and r_j (see below) and plot (for several values of C) the quotient in Eq. (5) as a function of λ . The speed of the front is the quotient in Eq. (5) for the value of λ corresponding to the minimum in the plot (Fort, 2012).

The most direct way to estimate the values of p_j and r_j is to consider distances between the birthplaces of parents and their children. Such distances were measured for rice farmers in rural Asian populations by Mehrai (1984). The corresponding kernel, computed from 714 distances, is given by the probabilities $\{p_j\} = \{0.803, 0.040, 0.022, 0.025, 0.063, 0.005, 0.009, 0.019, 0.014\}$ for their respective distances $\{r_j\} = \{0.5, 5.5, 15, 25, 35, 50.03, 57.2, 60.51, 97.65\}$ km (see Supp. Info., Sec S2a). Importantly, if we used other observed kernels for rice farmers, the conclusions would not change (Supp. Info., Sec S2). As in previous work (Fort, 2012; Jerardino et al., 2014), we use the range $0.023 \text{ yr}^{-1} \leq a_N \leq 0.033 \text{ yr}^{-1}$ for the initial growth rate (as estimated from archaeological and ethnographic observations of preindustrial farming populations, see Isern et al., 2008) and $29 \text{ yr} \leq T \leq 35 \text{ yr}$ for the generation time (as estimated from ethnographic data of preindustrial farming populations, see Fort et al., 2004a). Unfortunately, it does not seem possible to estimate a range for C from ethnographic data because, although there are some quantitative examples of hunter-gatherers learning farming (Fort, 2012), it would be also necessary to know the percentage of cases in which such a cultural transmission takes place (and this seems extremely difficult). Thus, we shall estimate a range for C from the spread rate. For each value of C , using the maximum value above for the reproduction rate of farmers ($a_N = 0.033 \text{ yr}^{-1}$) and the minimum one for their generation time ($T = 29 \text{ yr}$), we obtain

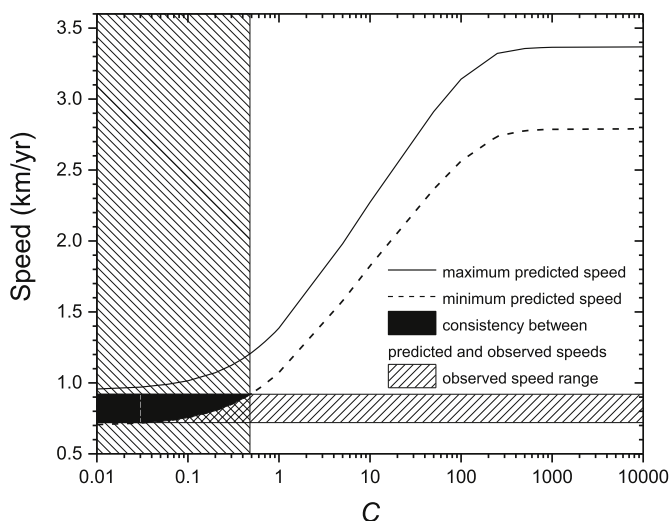


Fig. 4. Predicted Neolithic front speeds using the dispersal kernel reported by Mehrai (1984). The full curve is the maximum speed predicted by the two-dimensional demic-cultural model (Eq (5)), and the dashed curve is the minimum speed predicted by the same model. Thus the model agrees with all pairs of values (of the speed and C) in the area between both curves. The area within the horizontal rectangle agrees with the observed speed range (0.72–0.92 km/yr). The black area is the intersection between both areas, i.e. the pairs of values for which the model is consistent with the observed speed. The vertical rectangle gives the implied range of the cultural transmission intensity C . The purely demic model corresponds to $C = 0$ (i.e., no incorporation of hunter-gatherers into the populations of rice farmers).

the maximum predicted Neolithic front speed. The reason why the maximum speed corresponds to the minimum generation time T is that a lower value of T is equivalent to a lower time interval between two successive dispersal events, which will obviously lead to a faster front (Fort, 2012). Similarly, using the minimum reproduction rate ($a_N=0.023 \text{ yr}^{-1}$) and the maximum generation time ($T=35 \text{ yr}$), we obtain the minimum predicted Neolithic front speed. In this way we can plot the maximum and minimum speeds (predicted by the model) as a function of the intensity of cultural transmission C (full and dashed curves in Fig. 4, respectively).

The horizontal rectangle in Fig. 4 corresponds to the observed speed range for the spread of rice in eastern and southeastern Asia, estimated above from archaeological data (namely 0.72–0.92 km/yr, from the linear regression in Fig. 2). In Fig. 4, from the region of consistency (black area) between the archaeological data (horizontal rectangle) and the model (area between the full and dashed curves) we conclude that $C < 0.5$. According to the demic-cultural model (Fort, 2012) this means that, on average, less than 0.5 hunter-gatherers were converted into rice farmers by cultural transmission per rice farmer, i.e. by teaching (Fort, 2012) and/or interbreeding (Fort, 2011), when the first rice farmers arrived to regions previously occupied by hunter-gatherers (Fort, 2012). In other words, on average, at most one hunter-gatherer was converted into farmer by each couple of rice farmers.

Finally, in order to estimate the magnitude of cultural versus demic diffusion, we plot the cultural effect (in %) in Fig. 5. For a given value of C , this effect is defined as (Fort, 2012)

$$\text{cultural effect (in \%)} = 100 \frac{s - s_{C=0}}{s} \tag{7}$$

where s is the speed with cultural transmission (i.e. the speed obtained from Eq. (5) and shown in Fig. 4 for the value of C considered) and $s_{C=0}$ is the speed without cultural transmission (i.e. the speed obtained from Eq. (5) and shown in Fig. 4 for $C = 0$).

In Fig. 5 we show the results, obtained from the maximum (solid line) and minimum (dashed line) speeds in Fig. 4.

From the results in Fig. 5, we can estimate that the cultural effect on

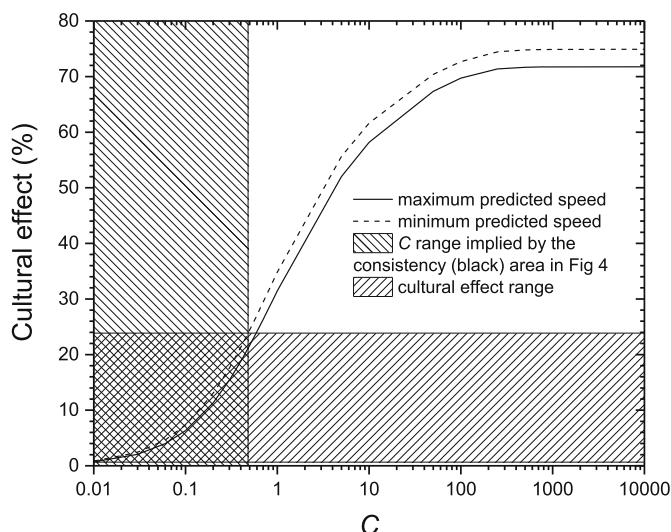


Fig. 5. Effect of cultural transmission on the front speed using the dispersal kernel reported by Mehrai (1984). Each curve gives the difference between the corresponding curve (speed) predicted by the demic-cultural model in Fig. 4 and the speed in the purely demic case (i.e., the speed for the corresponding curve without cultural transmission, $C = 0$ in Fig. 4), divided by the former and multiplied by 100 (see Eq. (7)). The vertical rectangle is the range of cultural transmission intensity C obtained from Fig. 4. The horizontal rectangle is the range of cultural transmission (%) consistent with the vertical rectangle. It indicates that cultural diffusion was responsible for less than 24% of the spread rate, so that demic diffusion was responsible for at least 76%.

the Neolithic wave of advance in southeastern Asia was less than 24% (horizontal rectangle). Accordingly, demic diffusion explains at least 76% of the spread rate and we conclude that, in this sense, the Neolithic front of domesticated rice in southeastern Asia was mainly demic. Moreover, it is noteworthy that using the kernels of other populations of rice cultivators, we find similar results (see Supp. Info., Sec S2). Thus our conclusions are robust to changes in the parameter values, as long as they are realistic (we stress that all parameter values used here have been obtained from empirical data). Moreover, a sensitivity analysis of hypothetical dispersal kernels shows that, for our model to be inconsistent with the observed spread rate, it would be necessary to assume kernels that are not reasonable for human populations (Sec. S2e).

4. Discussion and conclusions

In this work we have used a database of domesticated rice to estimate its spread rate, and compared the latter to the predictions of a demic-cultural mathematical model to estimate the relative importance of demic and cultural diffusion. It is important to stress that these findings are not measures of the phenomenon itself, but estimates based on several assumptions. One reason is that more than half of the sites were dated by cultural association and, although excluding them leads to the same conclusions, it causes a decrease in the statistical confidence of the results (Supp. Info., Sec. S1f). Another reason is that the mathematical model is based on well-defined assumptions (Sec. S3) and parameter values, for which there is no direct proof of validity (for the times and places corresponding to this phenomenon), in spite of the fact that they are reasonable according to ethnographic and archaeological data (for a detailed discussion on the dispersal kernel, see Sec. S2).

We have estimated that the Neolithic transition in southeastern Asia spread at a rate of $0.82 \pm 0.10 \text{ km/yr}$ (Fig. 2). By comparing this observed speed range to the predictions of a demic-cultural wave-of-advance model (Fort, 2012), we have concluded that the spread of domesticated rice in eastern and southeastern Asia was mainly demic. This is the key conclusion of the present paper. More precisely, we have found that the cultural effect can explain at most 24% of the observed

speed for the Neolithic wave of advance (Fig. 5). From the comparison between archaeological data and the demic-cultural wave-of-advance model, we have also found that the cultural diffusion intensity was $C \leq 0.5$, which means that less than 0.5 hunter-gatherers (on average) were converted into rice farmers (per each rice farmer) by cultural transmission, i.e. by teaching (Fort, 2012) and/or interbreeding (Fort, 2011) when the first rice farmers arrived to regions previously occupied by hunter-gatherers. This means that, on average, at least two newcomers who farm rice were needed for a hunter-gatherer to adopt rice agriculture.

We would like to stress that the primacy of demic over cultural diffusion for the spread of domesticated rice in southeastern Asia has been established here at the continental level only, and that this does not preclude that cultural diffusion could have been more important in specific regions (Kudo, 2000; Takamiya, 2007; Fuller, 2011; Barton, 2012).

Our conclusion that the spread of the Neolithic in eastern and southeastern Asia was mainly demic agrees with the views of several authors. Indeed, Fuller (2011) has pointed out that intensive wet rice agriculture in the lower and middle Yangtze about 4.5 kyr BC must have supported growing populations, and this would have driven outward migrations. It has been also argued that the spread of rice agriculture played an important role in the propagation of languages (Van Driem, 2002; Starosta, 2005; Sagart, 2008; Fuller, 2011). The reconstruction of a proto-vocabulary for current southeastern Asian languages suggests an origin related to rice exploitation, and since current archaeological data suggest that rice domestication arose in the Yangtze River basin, this could be the origin of the spread of these languages (Bellwood, 2011). It has been also argued that, whereas cultural diffusion (e.g., processes of cultural shift and contact-induced change) often yields local (rather than continental) language expansions, it seems unlikely that it can explain the long-distance movements observed for the language families in southeastern Asia (Bellwood, 2006). The latter include, for example, the spread of languages from southern China, through Thailand and Indochina, into Malaysia (Higham, 2002) and Indonesia, and even reaching New Zealand and Madagascar (Diamond and Bellwood, 2003). Therefore, the observed linguistic distribution could be related to a process of strong demic diffusion of native speakers (rice farmers) during the Neolithic spread (Bellwood, 2006), which is compatible with some cultural diffusion, i.e. the incorporation of local hunter-gatherers into the populations of migrating farmers (Bellwood, 2018).

Very recently, the first ancient genomic data from the region have been published. Although they are still very few (and there are none yet from some key areas, e.g. southern China), they indicate demic diffusion of rice farmers from China, as well as some hunter-gather ancestry due to cultural diffusion (Lipson et al., 2018; McColl et al., 2018). This genomic picture, and the linguistic one summarized in the previous paragraph, do not provide any quantitative estimation for the cultural diffusion intensity C or the percentage of cultural diffusion, but they are qualitatively consistent with our results (which have been obtained from archaeological data). When more genetic data become available, it will be probably possible to identify ancient genetic clines and use them to estimate (more precisely than in the present paper) the value of the cultural diffusion intensity C and the percentage of the cultural effect, as has been recently done for the spread of the Neolithic in the Near East and Europe (Isern et al., 2017a).

It is also noteworthy that our conclusions do not change if we use other dispersal kernels for rice farmers that have been reported in the literature (Supp. Info. Sec. S2). This indicates that, although uncertainty in the dispersal kernel affects the spread rate predicted by the model (and thus the percentage of cultural diffusion), this uncertainty does not affect our main conclusion, namely the primacy of demic diffusion. In other words, the cultural effect is below 50% for all dispersal kernels and values of the other parameters, as long as they are realistic (i.e., that they are not chosen but based on independent observations, as in

this study).

Funding

This work was supported by Ministerio de Ciencia, Innovación y Universidades [Grant FIS-2016-80200-P], Fundación Banco Bilbao Vizcaya Argentaria [Grant NeoDigit-PIN2015E], and an Academia award in Humanities from the Catalan Institution for Research and Advanced Studies (to JF).

Acknowledgements

The authors are thankful to Peter Bellwood for comments. This paper is dedicated to the memory of Luigi Luca Cavalli-Sforza (1922–2018), whose pioneering work on Neolithic wave-of-advance models was the starting point of the research reported in our previous works and this paper.

Appendix A. Supplementary data

Supplementary files to this article (a file with additional text and figures, and a file with several lists of sites) can be found online at <https://doi.org/10.1016/j.jas.2018.12.001>.

References

- Ammerman, A.J., Cavalli-Sforza, L.L., 1971. Measuring the rate of spread of early farming in Europe. *Man* 6, 674–688.
- Ammerman, A.J., Cavalli-Sforza, L.L., 1984. *The Neolithic Transition and the Genetics of Populations in Europe*. Princeton University Press, Princeton.
- Barton, H., 2012. The reversed fortunes of sago and rice, *Oryza sativa*, in the rainforests of Sarawak, Borneo. *Quaternary Int.* 249, 96–104.
- Bellwood, P., 2005. *First Farmers: the Origins of Agricultural Societies*. Oxford University Press, Oxford.
- Bellwood, P., 2006. Asian farming diasporas? Agriculture, languages, and genes in China and Southeast Asia. In: Stark, M.T. (Ed.), *Archaeology of Asia*. Blackwell Publishing, Malden, USA, pp. 96–118.
- Bellwood, P., 2011. The checkered prehistory of rice movement southwards as a domesticated cereal—from the Yangzi to the equator. *Rice* 4, 93–103.
- Bellwood, P., 2018. The search for ancient DNA heads east. *Science* 361, 31–32.
- Cavalli-Sforza, L.L., Feldman, M.W., 1981. *Cultural Transmission and Evolution: a Quantitative Approach*. Princeton University Press, Princeton.
- Diamond, J., Bellwood, P., 2003. Farmers and their languages: the first expansions. *Science* 300, 597–603.
- Draper, N.R., Smith, H., 1981. *Applied Regression Analysis*, second ed. Wiley, New York.
- Fort, J., 2011. Vertical cultural transmission effects on demic front propagation: theory and application to the Neolithic transition in Europe. *Phys. Rev. E* 83, 056124.
- Fort, J., 2012. Synthesis between demic and cultural diffusion in the Neolithic transition in Europe. *Proc. Natl. Acad. Sci. U. S. A.* 109, 18669–18673.
- Fort, J., 2015. Demic and cultural diffusion propagated the Neolithic transition across different regions of Europe. *J. R. Soc. Interface* 12, 20150166.
- Fort, J., Jana, D., Humet, J.M., 2004a. Multidelayed random walks: theory and application to the Neolithic transition in Europe. *Phys. Rev. E* 70, 031913.
- Fort, J., Méndez, V., 1999. Time-delayed theory of the Neolithic transition in Europe. *Phys. Rev. Lett.* 82, 867–870.
- Fort, J., Pérez-Losada, J., Isern, N., 2007. Fronts from integrodifference equations and persistence effects on the Neolithic transition. *Phys. Rev. E* 76, 031913.
- Fort, J., Pujol, T., 2008. Progress in front propagation research. *Rep. Prog. Phys.* 71, 086001.
- Fort, J., Pujol, T., Cavalli-Sforza, L.L., 2004b. Palaeolithic populations and waves of advance. *Camb. Archaeol. J.* 14, 53–61.
- Fort, J., Pujol, T., vander Linden, M., 2012. Modelling the Neolithic transition in the Near East and Europe. *Am. Antiq.* 77, 203–220.
- Fuller, D.Q., 2011. Pathways to Asian civilizations: tracing the origins and spread of rice and rice cultures. *Rice* 4, 78–92.
- Fuller, D.Q., Harvey, E., Qin, L., 2007. Presumed domestication? Evidence for wild rice cultivation and domestication in the fifth millennium BC of the lower Yangtze region. *Antiquity* 81, 316–331.
- Fuller, D.Q., Qin, L., 2009. Water management and labour in the origins and dispersal of Asian rice. *World Archaeol.* 41, 88–111.
- Fuller, D.Q., Weisskopt, A.R., Castillo, C.C., 2016. Pathways to rice diversification across Asia. *Archaeol. Int.* 19, 84–96.
- Higham, C.F.W., 2002. Languages and farming dispersals: austroasiatic languages and rice cultivation. In: Bellwood, P., Renfrew, C. (Eds.), *Examining the Farming/language Dispersal Hypothesis*. McDonald Institute for Archaeological Research, Cambridge, pp. 223–232.
- Isern, N., Fort, J., de Rioja, V., 2017a. The ancient cline of haplogroup K implies that the Neolithic transition in Europe was mainly demic. *Sci. Rep.* 7, 11229.

- Isern, N., Fort, J., Pérez-Losada, J., 2008. Realistic dispersion kernels and cohabitation reaction-dispersion equations. *J. Stat. Mech. Theor. Exp.* 2008, P10012.
- Isern, N., Zilhão, J., Fort, J., Ammerman, A.J., 2017b. Modeling the role of voyaging in the coastal spread of the early neolithic in the west mediterranean. *Proc. Natl. Acad. Sci. U. S. A.* 114, 897–902.
- Jerardino, A., Fort, J., Isern, N., Rondelli, B., 2014. Cultural diffusion was the main driving mechanism of the Neolithic transition in southern Africa. *PLoS One* 9, e113672.
- Kudo, M., 2000. *Hodai Emishi (The Emishi of the Ancient Times)*. Yoshikawa Kobun Kan, Tokyo.
- Lipson, M., Cheronet, O., Mallick, S., Rohland, N., Oxenham, M., Pietrusewsky, M., et al., 2018. Ancient genomes document multiple waves of migration in Southeast Asian prehistory. *Science* 361, 92–95.
- McColl, H., Racimo, F., Vinner, L., Demeter, F., Gajuhari, T., Moreno-Mayar, J.C., et al., 2018. The prehistoric peopling of Southeast Asia. *Science* 361, 88–92.
- Mehrai, H., 1984. *A Demographic Study of Some Populations in Shahrestan Nowshahr, Mazandaran, Iran, with Reference to the Genetic Structure*. PhD Thesis. Durham University Available from: <http://etheses.dur.ac.uk/7169/>.
- Mitas, L., Mitasova, H., 1999. Spatial interpolation. In: Longley, P.A., Goodchild, M.F., Maguire, D.J., Rhind, D.W. (Eds.), *Geographical Information Systems: Principles, Techniques, Management and Applications*. Wiley, London, pp. 481–492.
- Pinhasi, R., Fort, J., Ammerman, A.J., 2005. Tracing the origin and spread of agriculture in Europe. *PLoS Biol.* 3, e410.
- Sagart, L., 2008. The expansion of *Setaria* farmers in East Asia: a linguistic and archaeological model. In: Sanchez-Mazas, A., Blench, R., Ross, M., Peiros, I., Lin, M. (Eds.), *Past Human Migrations in East Asia: Matching Archaeology, Linguistics and Genetics*. Routledge, New York, pp. 133–157.
- Silva, F., Stevens, C.J., Weisskopf, A., Castillo, C., Qin, L., Bevan, A., Fuller, D.Q., 2015. Modelling the geographical origin of rice cultivation in Asia using the Rice Archaeological Database. *PLoS One* 10, e0137024.
- Silva, F., Weisskopf, A., Castillo, C., Murphy, C., Kingwell-Banham, E., Qin, L., Fuller, D.Q., 2018. A tale of two rice varieties: modelling the prehistoric dispersals of japonica and proto-indica rices. *Holocene* 28, 1745–1758.
- Sinnott, R.W., 1984. Virtues of the haversine. *Sky Telescope* 68, 159.
- Spiegel, M.R., Schiller, J.J., Srinivasan, R.A., LeVan, M., 1975. *Probability and Statistics*, third ed. McGraw-Hill, New York.
- Starosta, S., 2005. Proto-East asian and the origin and dispersal of the languages of East and southeast Asia and the pacific. In: Sagart, L., Blench, R., Sanchez-Mazas, A. (Eds.), *The Peopling of East Asia: Putting Together Archaeology, Linguistics and Genetics*. Routledge Curzon, New York, pp. 182–198.
- Takamiya, H., 2007. How did agriculture spread? - case studies from Japan. In: Komoto, M. (Ed.), *A Study on the Environmental Change and Adaptation System in Prehistoric Northeast Asia*. Kumamoto University, Kumamoto, pp. 110–118.
- Taylor, J., 1997. *Introduction to Error Analysis, the Study of Uncertainties in Physical Measurements*, second ed. University Science Books, Sausalito.
- Van Driem, G., 2002. Tibeto-Burman phylogeny and prehistory: languages, material culture and genes. *Antiquity* 72, 885–897.
- Wang, W., et al., 2018. Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature* 557, 43–49.