1   **Genomic history and ecology of the geographic spread of rice**

2

3   **Rafal M. Gutaker[1], Simon C. Groen[1], Emily S. Bellis[2], Jae Y. Choi[1], Inês S. Pires[1,3], R.**

4   **Kyle Bocinsky[4], Emma R. Slayton[5], Olivia Wilkins[1,6], Cristina C. Castillo[7,8], Sónia**

5   **Negrão[9], M. Margarida Oliveira[3], Dorian Q. Fuller[7,8], Jade A. d'Alpoim Guedes[10], Jesse**

6   **R. Lasky[2]*, Michael D. Purugganan[1,11,12]***

7

8   [1]Center for Genomics and Systems Biology, New York University, New York, NY 10003

9   USA; [2]Department of Biology, Pennsylvania State University, University Park, PA 16802

10   USA; [3]Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de

11   Lisboa, Av. da República, 2780-157 Oeiras, Portugal; [4]Crow Canyon Archaeological Center,

12   Cortez, CO 81321, USA; [5]Carnegie Mellon University Libraries, Pittsburgh, PA 15213-3890;

13   [6]Department of Biological Sciences, University of Manitoba, Winnipeg, MB R3T 2N2,

14   Canada; [7]Institute of Archaeology, University College London, London, WC1H 0PY, United

15   Kingdom; [8]School of Cultural Heritage, North-West University, Xi'an, Shanxi 710069,

16   China; [9]School of Biology and Environmental Science, University College Dublin, Belfield,

17   Dublin 4, Ireland; [10]Department of Anthropology and Scripps Institution of Oceanography,

18   University of California San Diego, La Jolla, CA 92093, USA; [11]Center for Genomics and

19   Systems Biology, New York University Abu Dhabi, Saadiyat Island, Abu Dhabi, United

20   Arab Emirates. [12]Institute for the Study of the Ancient World, New York University, New

21   York, NY 10028.

22

23   *Corresponding author

24

25   **Keywords**: domestication, crop evolution, diversification, paleoclimate, archaeobotany,

26   admixture analysis, coalescence, *Oryza sativa*

27

28   **ABSTRACT**

29   Rice (*Oryza sativa*) is one of the world's most important food crops. We reconstruct

30   the history of rice dispersal in Asia using whole-genome sequences of >1,400 landraces,

31   coupled with geographic, environmental, archaeobotanical and paleoclimate data. We also

32   identify extrinsic factors that impact genome diversity, with temperature a leading abiotic

33   factor. Originating ~9,000 years ago in the Yangtze Valley, rice diversified into temperate

34   and tropical japonica during a global cooling event ~4,200 years ago. Soon after, tropical rice

35   reached Southeast Asia, where it rapidly diversified starting ~2,500 yBP. The history of

36   indica rice dispersal appears more complicated, moving into China ~2,000 yBP.

37   Reconstructing the dispersal history of rice and its climatic correlates may help identify

38   genetic adaptation associated with the spread of a key domesticated species.

39

40   **One sentence summary**

41   We reconstructed the ancient dispersal of rice in Asia and identified extrinsic factors that

42   impact its genomic diversity.

43

44   Rice (*Oryza sativa* L.) is a major staple crop, providing > 20% of calories for more

45   than half of the human population. Domesticated rice encompasses genetically distinct

46   populations grown in sympatry, including major subgroups japonica and indica (sometimes

47   recognized as subspecies), as well as geographically more restricted *circum*-aus, and *circum*-

48   basmati rices (*1*, *2*). It is mainly cultivated in monsoon Asia, but rice is distributed across a

49   wide latitudinal range, spanning tropical and temperate zones of Asia, likely requiring local

50   water, temperature and photoperiod adaptation. Rice is grown in lowland ecosystems under

51   paddy, deepwater, or seasonal flood conditions, as well as in upland rainfed areas (*3*).

52   Archaeological evidence (*4–6*) indicates that cultivation of japonica rice began ~9,000

53   years before present (yBP) in the lower Yangtze Valley, while proto-indica rice cultivation

54   started >5,000 yBP in the lower Ganges valley (*7*). Archaeological (*8*) and most population

55   genetic analyses (*9–11*) suggest that important domestication alleles have a single origin in

56   japonica rice in East Asia. The spread of japonica to South Asia ~4,000 years ago led to

57   introgression of domestication alleles into proto-indica or local *O. nivara* populations and the

58   emergence of indica rice (*9–11*). From the Yangtze and Ganges Valleys, respectively,

59   japonica and indica dispersed across much of Asia over the last 5 millennia, providing

60   sustenance for emerging Neolithic communities in East, Southeast and South Asia (*12*).

61   Archaeological data shows the general directionality of rice dispersal (*7*, *13*); the

62   details of dispersal routes, times, and the environmental forces that shaped dispersal patterns,

63   however, remain unknown. Here, we undertake population genomic analyses to examine

64   environmental factors associated with the geographic distribution of rice diversity, and

65   reconstruct the ancient dispersal of rice in Asia. Together with archaeobotanical,

66   paleoclimatic and historical data, genomic data allows a robust reconstruction of the dispersal

67   history of *Oryza sativa*.

68

2

69   ***Structure of rice genomic diversity.*** We obtained whole genome re-sequencing data from

70   rice landraces/traditional varieties across a wide geographical distribution in Asia. Our

71   sample set includes 1,265 samples from the Rice 3K Genome Project (*1*) and additional 178

72   landraces sequenced for this study (Supplementary Table 1); the panel consists of 833 indica,

73   372 japonica, 165 *circum*-aus, 42 *circum*-basmati, and 31 unclassified samples. We identified

74   ~9.78 million single nucleotide polymorphisms (SNPs) with 9.63x mean coverage (s.d. =

75   5.03), which we used in subsequent analyses (Supplementary Fig. 1).

76   Analysis of molecular variance (AMOVA) indicated that subspecies affiliation

77   explained >36% of the total variation [AMOVA, permutation $P < 0.001$](*14*), congruent with

78   results from multidimensional scaling (MDS) of genomic distances (Supplementary Fig. 2a).

79   Only japonica and indica have wide geographic distributions (Fig. 1 a and b; Supplementary

80   Fig. 3), and AMOVA of these two subspecies (n=1,205) revealed that genomic variance is

81   explained by subspecies ($r^2 = 0.32$, permutation $P < 0.001$), country of origin ($r^2 = 0.11$, $P <$

82   0.001) and their interaction ($r^2 = 0.06$, $P < 0.001$). Landraces with mixed ancestry (n=154)

83   were excluded using silhouette scores (Supplementary Fig. 2b); henceforth, we analysed

84   these two subspecies independently.

85   We find support for isolation-by-distance (IBD) in japonica ($r^2 = 0.294$, $P < 0.001$)

86   and indica ($r^2 = 0.265$, $P < 0.001$) [Supplementary Fig. 4]. Geographic distance explains

87   genetic distance much less in the Malay Archipelago (*i.e.* islands SE Asia) compared to

88   mainland Asia, suggesting a stronger effect of local migration barriers on island IBD

89   (Supplementary Fig. 5). Effective migration surfaces (*15*) identified geographic barriers for

90   dispersal over the Hengduan Mountains which separate China from South/Southeast Asia,

91   and the South China Sea which reduces movement between Borneo/Philippines and mainland

92   Southeast Asia (Fig. 1c and d; Supplementary Fig. 6). For human-dispersed species such as

93   crops, genetic distances may correlate better with travel resistance, meant to capture cost-

3

94   effective migration by humans. An isolation-by-resistance (IBR) model, using estimated

95   human-associated land and marine travel times (*16*), is a better explanation than the IBD

96   model for japonica landrace genetic distances based on Akaike Information Criterion (island

97   ΔAIC = -34, mainland ΔAIC = -17), but not for indica (island ΔAIC = +51, mainland ΔAIC =

98   +611)[Supplementary Fig. 5)].

99

100   ***Factors associated with spatial genomic structure.***

101   We used redundancy analysis (RDA) to partition genomic variance (*17*) associated with 22

102   different variables that include climatic and edaphic conditions, as well as interactions with

103   humans and wild relatives (Supplementary Table 1). We assume that while environments in

104   localities fluctuate over time, current genome diversity may be determined both by current

105   environment as well as long-term evolutionary history. SNP variation is better explained by

106   our predictors for japonica (adjusted r^2 = 0.363; Fig. 1e) than indica (adjusted r^2 = 0.164; Fig.

107   1f). Associations between predictor sets and SNPs are substantially collinear with each other.

108   For japonica and indica, travel time and geographic distance, respectively, explain most SNP

109   variation (adjusted r^2 = 0.326 and r^2 = 0.146), followed by abiotic conditions, language

110   groups, culinary properties (*i.e.*, cooked grain stickiness), and genetic composition of

111   proximal wild rice populations (Figs. 1e and f; Supplementary Fig. 7). Among abiotic

112   variables for japonica, temperature explains the greatest portion of SNP variation (adjusted r^2

113   = 0.180), followed by moisture (r^2 = 0.086) and soil characteristics (r^2 = 0.081). Similarly,

114   temperature explains the most SNP variation in indica (r^2 = 0.064), followed by soil

115   characteristics (r^2 = 0.038) and moisture (r^2 = 0.036) (Supplementary Fig. 7), although these

116   factors have weaker explanatory power in indica compared to japonica.

117   The first two RDA axes of environment-associated SNP variation (*18*) separated

118   japonica landraces consistent with geography (Fig. 1g), recapitulating results using total SNP

119   variation (Supplementary Fig. 8a). Temperate japonica landraces from northern latitudes are

120   most strongly identified by alleles associated with high coefficient of inter-annual variation in

121   growing degree days, and low minimum temperatures early in the growing season (Fig. 1g;

122   Supplementary Fig. 9a). Temperate landraces from upland rainfed ecosystems are further

123   characterized by alleles associated with inter-annual variation in precipitation.

124   For indica, the first two axes also grouped individuals by their geographic origins

125   (Fig. 1h; Supplementary Fig. 8b). Similar to japonica, indica Malay Archipelago genotypes

126   are characterized by alleles associated with high precipitation prior to the growing season.

127   Mainland Southeast Asian genotypes are characterized by alleles associated with warm

128   minimum growing season temperatures and presence of nearby freshwater sources (Fig. 1h;

129   Supplementary Fig. 9b). The latter contrasts with indica from China and most of India, where

130   irrigation is common and there is less reliance on natural water sources (19)(Supplementary

131   Table 1). Finally, genotypes in South India are identified by alleles associated with inter-

132   annual variation in precipitation.

133

134   ***Discrete subpopulations within japonica and indica.*** We clustered landraces based on

135   genomic distances by partitioning-around-medoids [PAM](20), identifying the number of

136   discrete clusters (K) using silhouette scores (21) [see Methods]. This discretization procedure

137   removed genetic gradients between subpopulations (Fig. 2a and 2d; Supplementary Figs. 10

138   and 11). We compared PAM clusters to those from the ADMIXTURE algorithm (22).

139   Silhouette filtering removed individuals with spurious subpopulation assignments

140   (Supplementary Figs. 12 and 13). In general, the clustering fit using silhouette scores is

141   greater for japonica than indica (Supplementary Fig. 14). We find consistently higher F_{ST}

142   values among japonica subpopulations (Supplementary Fig. 15), suggesting less migration

143   compared to indica. Finally, subpopulations of both subspecies clearly correspond with

144   geography (Fig. 2b and 2e; Supplementary Figs. 10 and 11), suggesting that contemporary

145   rice landraces retain genomic signals of past dispersal across Asia.

146

147   ***Relationships between japonica subpopulations***. We modelled subpopulation relationships

148   separately for japonica and indica using the admixture graph framework (*23*). We

149   reconstructed relationships between japonica subpopulations at k = 2 to 9 considering graphs

150   with population f-statistic z-scores <3. Throughout all k levels, we find two similar and

151   consistent graph topologies (Fig. 2c; Supplementary Fig. 16), which we used to infer

152   dispersal routes of japonica.

153   As expected (*2*), at k = 2 we observe divergence between lowland temperate varieties

154   in Northeast Asia (Korea, Japan, China and Taiwan) and tropical varieties from the Malay

155   Archipelago (Malaysia, Philippines and Indonesia). At k = 3, we find a major lineage of

156   tropical upland japonica in mainland Southeast Asia as sister group to Malay Archipelago

157   landraces or from admixture with an ancestral temperate lineage (Supplementary Figs. 10 and

158   16). At higher k, these mainland Southeast Asian upland landraces always incorporates

159   admixture from an ancestral temperate japonica population (see below).

160   At k = 4 we observe separation of primarily Indonesian from Philippine and Bornean

161   landraces. Subsequently, at k = 5, upland temperate japonica in Northeast Asia emerges as an

162   admixture between lowland temperate and upland tropical varieties. Further increase of k

163   allows separation of distinct Malay Archipelago subpopulations: a small subpopulation

164   associated with the Philippines splits first, followed by a subpopulation in the Indonesian

165   island of Java. Subsequent divisions among Malay Archipelago subpopulations are not fully

166   resolved (Supplementary Fig. 16). Nevertheless, at k = 8, we identify a Bhutanese

167   subpopulation closely related to upland Laotian landraces, and may represent a relict

168   descendant population of the first early split in tropical japonica.

169

170   ***The rise of temperate japonica***. Combining genomic, geographic, archaeological and

171   paleoenvironmental data, we reconstructed routes and timing of the ancient dispersal of rice

172   in Asia. Japonica represents the first domesticated *O. sativa* (*9–11*), and its tropical form was

173   cultivated in eastern China between the Yangtze and the Huang He (Yellow) river valleys

174   (*13*). This occurred during the Holocene Climate Optimum (HCO), a period of increased

175   monsoon activity and warmer temperatures between ~9,000 and 4,000 yBP (*24, 25*); this

176   coincides with the rise in frequency of non-shattering rice from ~20% just after 8,000 yBP to

177   fixation at ~5,000 yBP (*5, 6*).

178   The first major population divergence in japonica separates temperate from tropical

179   landraces (Supplementary Figs. 10 and 16). Using sequentially Markovian coalescent

180   (SMC++), we estimated a cross-coalescence split time between temperate and tropical

181   japonica at ~5,000 to 1,500 years ago, with 75% of estimates between ~4,100 to 2,500 years

182   ago (Fig. 3a; Supplementary Fig. 17). Using dated archaeobotanical rice remains (*13*), we

183   note that rice agriculture spread north- and eastward along the Huang He river (*26*) and

184   westward into the Chengdu Plains and the Southwest China Highlands between ~5,000 to

185   4,000 yBP (*27–29*)(Fig. 3b; Supplementary Fig. 18). During a minor climatic cooling event at

186   ~5,000 yBP, rice appears maladapted in parts of eastern China (*30*). In the Shandong

187   Peninsula, rice disappeared by 5,000 yBP and briefly re-emerged 4,500 yBP as a short-

188   grained variety similar to contemporary temperate japonicas (*31*). A global temperature

189   decrease that followed the HCO at ~4,200 years ago, the '4.2k event' (*24, 25*), resulted in

190   waning rice agriculture in East China and strong pressure for japonica to adapt to a temperate

191   environment (*31*). Congruent with this, we observe that the highest density of estimated

192   temperate japonica split times start at ~4,100 years ago (Fig. 3a; Supplementary Fig. 17).

7

193   Temperate adaptation created opportunity for northeastern dispersal of japonica in

194   Asia. From our demographic analysis of temperate japonica we note a ~5-10-fold Ne

195   reduction between ~3,500 to 3,000 yBP (Fig. 3c; Supplementary Fig. 19), which we interpret

196   as a founder bottleneck during expansion to its new temperate niche. Indeed, this is consistent

197   with archaeological dates for the introduction of rice agriculture to Korea (*32*, *33*) and Japan

198   following decrease in rice remains in Eastern China (Supplementary Fig. 18).

199

200   ***The southward spread of japonica***. Throughout the HCO, tropical japonica was cultivated in

201   eastern China; its contemporary descendants however, are grown predominantly in Southeast

202   Asia (*1*), and we indeed find that Southeast Asian subpopulations descend from the tropical

203   lineage that diverged from temperate japonica. Demography reconstruction at k = 2-4 shows

204   that tropical japonica lineage experienced a ~50-100-fold population (Ne) contraction

205   between ~4,500 to 4,000 yBP, and partial Ne recovery starting ~2,500 yBP (Fig. 3d,

206   Supplementary Fig. 19). The population contraction in tropical japonica is contemporaneous

207   with the 4.2k event, raising the possibility that cooling explains the collapse of tropical rice

208   cultivation in East Asia and its southern relocation. This coincides with the arrival of rice in

209   the far south of China ~4,500 yBP and a shift to rainfed, upland cultivation (*34*).

210   Given the importance of temperature in shaping japonica genomic diversity across its

211   geographic distribution (Fig. 1g), we used a thermal niche model (*35*) based on

212   reconstruction of Holocene temperatures (*36*) to estimate the probability of tropical rice

213   cultivation in different areas during the post-HCO period (Fig. 3e; Supplementary Fig. 20).

214   Survival probabilities of tropical japonica between ~4,400 and 3,500 yBP dropped

215   dramatically in eastern China and high-altitude South China (survival probability < 50%)

216   compared to Southeast Asia [survival probability > 90%](Fig. 3e; Supplementary Video 1).

8

217   Indeed, after the cooling period we observe high densities of archaeological rice remains in

218   Southeast Asia (Fig. 3b; Supplementary Fig. 18).

219   After the HCO, rice dispersed from China to Southeast Asia into Laos and Bhutan,

220   and through maritime routes to the Philippines, Malaysia and Indonesia. In our admixture

221   graph analysis, we find an early split in the tropical lineage that separates Bhutan and Laos

222   upland rice from rice in the Malay Archipelago (Fig. 2c). From coalescence analyses we

223   observe a ~50-100-fold population contraction in the remote upland (Bhutan) rice population

224   between ~4,000 and 3,000 yBP (Fig. 4; Supplementary Fig. 19), which may arise from a

225   bottleneck associated with population movements into these new areas. Emergence of upland

226   rice in Laos and Bhutan coincides in time and space with widespread establishment of rainfed

227   rice agriculture in mainland Southeast Asia, ~4,000 yBP (*12*, *37*) and dispersal of metallurgy

228   traditions from Bronze Age Yunnan, ~3,500 yBP southwards to Thailand by ~3,000 yBP (*38*,

229   *39*). Subsequent agricultural intensification of rice production took place from ~2,500 to

230   1,500 yBP and included evolution of irrigation systems in present-day Thailand (*40*).

231   Consistent with these, ancient human DNA studies in Southeast Asia report two farmer-

232   associated migration events from East Asia, one at least 4,000 years ago and a second before

233   2,000 yBP (*41*, *42*).

234   Our analysis also shows an ~5-10-fold Ne decrease in the Malay archipelago between

235   ~3,000 and 2,500 yBP, and based on cross-coalescence analyses, divergence between

236   mainland and Malay Archipelago rice occurred between ~3,000 to 1,500 years ago (75% of

237   estimates in ~2,500 to 1,600 yBP) [Fig. 4; Supplementary Fig. 19]. Distinct island

238   populations in the Malay Archipelago diverged at around a similar timeframe, in an interval

239   from ~3,000 to 1,000 years ago (75% estimates fall between ~2,500 and 1,500 yBP). This

240   period coincides with dispersal of Dong Son drums in the Malay Archipelago (~2,400 years

241   ago) (*38*, *43*), and suggests maritime dispersal of rice from a North Vietnam hub within the

242   Austronesian Trading Sphere, which stretched between Taiwan and the Malay Peninsula (*44*,

243   *45*). Ancient DNA studies also suggest a wave of Austronesian human expansion into island

244   Southeast Asia ~2,000 years ago (*41*), which agrees with our estimates of japonica movement

245   into the area. Interestingly, upland temperate japonica in Japan appears to be an admixed

246   population of local lowland temperate rice and upland tropical rice from the Malay

247   Archipelago which may have moved northwards through Taiwan and perhaps the Ryukyu

248   Islands ~1,200 yBP (*46*).

249

250   ***Relationships and dispersal of indica subpopulations.*** We reconstructed relationships

251   between indica subpopulations with k = 2 to 7. Divergence between Sino-Indian and

252   Southeast Asian indica is present in all graph topologies beginning at k = 2 (Supplementary

253   Fig. 21). At k = 3 we observe separation of mainland and island Southeast Asian

254   subpopulations, while at k = 4 we observe separation of Indian from Chinese landraces. With

255   k = 5 and k = 7 we note differentiation of mainland Southeast Asian landraces into

256   subpopulations associated with Laos, Thailand and Cambodia (Fig. 2f). Interestingly, a

257   subpopulation associated primarily with Cambodia, and another in Indonesia, share ancestry

258   with the main Laos/Thailand Southeast Asian lineage as well as an early ancestral indica

259   population. Increasing to k = 8 also increases the number of admixture events in the model to

260   four, which renders further exhaustive graph topology searches unfeasible.

261   Higher diversity of graph topologies in indica, likely due to weaker population

262   structure and elevated gene flow (Supplementary Figs. 14 and 15), makes it difficult to

263   reconstruct indica dispersal routes. Moreover, given the complexity in multiple reconstructed

264   admixture graph topologies, we can only confidently date separation of Chinese and Indian

265   indica, which is unaffected by admixture. Our analysis estimates this divergence at ~2,500

266   and 1,100 yBP (75% of estimates between ~2,000 and 1,400 yBP)[Fig. 5; Supplementary Fig.

267   17]. Possible routes for indica dispersal from India to China could be the Silk Road or more

268   direct passage to Southwest China across the Hengduan mountains. The timing agrees with

269   written reports of the introduction of Buddhism from India to China at ~1,950 yBP (*47*), but

270   is later than the earliest putative finds of indica rice in China (*48*). The close relationship

271   between Indian and Chinese subpopulations is mirrored by higher proportions of irrigated

272   varieties in both regions; in contrast, Southeast Asian varieties are more often rainfed (*19*).

273   Indica dispersal to Southeast Asia (e.g., Thailand and Cambodia) were either from

274   India or China (Fig.5; Supplementary Fig. 22). From archaeobotanical studies, indica arrived

275   in Central Thailand at ~1,800 years ago (*40*), at a time when Asian trade routes were well

276   established (*12*). Late adoption of indica in Southeast Asia is hypothesized to be due to early

277   availability of japonica in this region (*12*). There is no earlier archaeological evidence for

278   indica cultivation in Southeast Asia, and hence it comes as a surprise that indica mainland

279   subpopulations suffered dramatic population size reduction between ~5,000 and 3,500 yBP

280   (Supplementary Fig. 23). It is even more puzzling that a bottleneck in indica subpopulation in

281   Indonesia occurred between ~6,000 and 5,000 yBP, suggesting complex origins, perhaps

282   partially from local wild ancestors or managed pre-domesticated varieties (Supplementary

283   Fig. 22).

284

285   *Summary*. Rice domestication in the Yangtze Valley had an enormous impact on the peoples

286   of East, Southeast and South Asia. In the first ~4,000 years of its history, Japonica rice

287   cultivation was largely confined to China, and its dispersal and diversification did not occur

288   until the global 4.2k cooling event. This abrupt climate change event, which was

289   characterized by a global reduction in humidity and temperature, had widespread

290   consequences: it is believed to have caused the breakdown of rice agriculture in East Asia

291   (*24*, *31*), turnover of cattle ancestry in the Near East (*49*), and the collapse of civilizations

292   from Mesopotamia (*50*) to China (*51*). We find from our genomic and paleoclimate

293   modelling that the 4.2 k event coincides with the rise of temperate japonica and the dispersal

294   of rice agriculture southwards into Southeast Asia. Moreover, indica began to be

295   domesticated in South Asia at around this period, and spread later into China and Southeast

296   Asia. Correlation between changing climate and rice distribution raises the possibility for a

297   causal relationship, and indeed we find temperature is a key environmental factor patterning

298   contemporary rice genomic diversity.

299   The ability to infer dispersal patterns of rice arises from the availability of extensive

300   landrace populations, whole genome sequences and population genomic approaches, as well

301   as environmental, archaeobotanical and paleoclimate data. Reconstructing the history of

302   domesticated species provides insight into the evolutionary process, nature of human/plant

303   co-evolutionary dynamics, and extrinsic landscape, environmental, and cultural factors that

304   drive crop dispersal. Armed with knowledge of the pattern of rice dispersal and

305   environmental features that influenced this migration, it may be possible to examine the

306   evolutionary adaptations of rice as it spread to new environments, which could allow us to

307   identify traits and genes to help future breeding efforts.

308

References

1. W. Wang, R. Mauleon, Z. Hu, D. Chebotarov, S. Tai, Z. Wu, M. Li, T. Zheng, R. R. Fuentes, F. Zhang, L. Mansueto, D. Copetti, M. Sanciangco, K. C. Palis, J. Xu, C. Sun, B. Fu, H. Zhang, Y. Gao, X. Zhao, F. Shen, X. Cui, H. Yu, Z. Li, M. Chen, J. Detras, Y. Zhou, X. Zhang, Y. Zhao, D. Kudrna, C. Wang, R. Li, B. Jia, J. Lu, X. He, Z. Dong, J. Xu, Y. Li, M. Wang, J. Shi, J. Li, D. Zhang, S. Lee, W. Hu, A. Poliakov, I. Dubchak, V. J. Ulat, F. N. Borja, J. R. Mendoza, J. Ali, J. Li, Q. Gao, Y. Niu, Z. Yue, M. E. B. Naredo, J. Talag, X. Wang, J. Li, X. Fang, Y. Yin, J.-C. Glaszmann, J. Zhang, J. Li, R. S. Hamilton, R. A. Wing, J. Ruan, G. Zhang, C. Wei, N. Alexandrov, K. L. McNally, Z. Li, H. Leung, Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature*. **557**, 43–49 (2018).

2. J. C. Glaszmann, Isozymes and classification of Asian rice varieties. *Theor. Appl. Genet.* **74**, 21–30 (1987).

3. D. Q. Fuller, J. van Etten, K. Manning, C. Castillo, E. Kingwell-Banham, A. Weisskopf, L. Qin, Y.-I. Sato, R. J. Hijmans, The contribution of rice agriculture and livestock pastoralism to prehistoric methane levels: an archaeological assessment. *Holocene*. **21**, 743–759 (2011).

4. D. Q. Fuller, L. Qin, Water management and labour in the origins and dispersal of Asian rice. *World Archaeol.* **41**, 88–111 (2009).

5. D. Q. Fuller, L. Qin, Y. Zheng, Z. Zhao, X. Chen, L. A. Hosoya, G.-P. Sun, The domestication process and domestication rate in rice: spikelet bases from the Lower Yangtze. *Science*. **323**, 1607–1610 (2009).

6. R. G. Allaby, C. Stevens, L. Lucas, O. Maeda, D. Q. Fuller, Geographic mosaics and changing rates of cereal domestication. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **372** (2017), doi:10.1098/rstb.2016.0429.

7. F. Silva, A. Weisskopf, C. Castillo, C. Murphy, E. Kingwell-Banham, L. Qin, D. Q. Fuller, A tale of two rice varieties: modelling the prehistoric dispersals of japonica and proto-indica rices. *Holocene*. **28**, 1745–1758 (2018).

8. D. Q. Fuller, Pathways to Asian civilizations: tracing the origins and spread of rice and rice cultures. *Rice*. **4**, 78–92 (2011).

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

10. J. Y. Choi, M. D. Purugganan, Multiple origin but single domestication led to Oryza sativa. *G3*. **8**, 797–803 (2018).

11. J. Y. Choi, A. E. Platts, D. Q. Fuller, Y.-L. Hsing, R. A. Wing, Purugganan, The rice paradox: multiple origins but single domestication in Asian rice. *Mol. Biol. Evol.* **34**, 11

348   (2017).

349   12. D. Q. Fuller, C. C. Castillo, C. Murphy, in *The Routledge handbook of archaeology and*
350   *globalization*, T. Hodos, Ed. (Routledge, New York, 2016), pp. 711–729.

351   13. F. Silva, C. J. Stevens, A. Weisskopf, C. Castillo, L. Qin, A. Bevan, D. Q. Fuller,
352   Modelling the geographical origin of rice cultivation in Asia using the rice
353   archaeological database. *PLoS One*. **10**, e0137024 (2015).

354   14. L. Excoffier, P. E. Smouse, J. M. Quattro, Analysis of molecular variance inferred from
355   metric distances among DNA haplotypes: application to human mitochondrial DNA
356   restriction data. *Genetics*. **131**, 479–491 (1992).

357   15. D. Petkova, J. Novembre, M. Stephens, Visualizing spatial population structure with
358   estimated effective migration surfaces. *Nat. Genet.* **48**, 94–100 (2016).

359   16. E. R. Slayton, *Seascape corridors: modeling routes to connect communities across the*
360   *Caribbean Sea* (Sidestone Press, 2018).

361   17. P. R. Peres-Neto, P. Legendre, S. Dray, D. Borcard, Variation partitioning of species
362   data matrices: estimation and comparison of fractions. *Ecology*. **87**, 2614–2625 (2006).

363   18. J. R. Lasky, D. L. Des Marais, J. K. McKay, J. H. Richards, T. E. Juenger, T. H. Keitt,
364   Characterizing genomic variation of Arabidopsis thaliana: the roles of geography and
365   climate. *Mol. Ecol.* **21**, 5512–5529 (2012).

366   19. S. M. Haefele, A. Nelson, R. J. Hijmans, Soil quality and constraints in global rice
367   production. *Geoderma*. **235-236**, 250–259 (2014).

368   20. L. Kaufmann, Clustering by means of medoids. *Proc. Statistical Data Analysis Based on*
369   *the L1 Norm* (1987) (available at https://ci.nii.ac.jp/naid/10027761751/).

370   21. P. J. Rousseeuw, Silhouettes: a graphical aid to the interpretation and validation of
371   cluster analysis. *J. Comput. Appl. Math.* (1987) (available at
372   https://www.sciencedirect.com/science/article/pii/0377042787901257).

373   22. D. H. Alexander, J. Novembre, K. Lange, Fast model-based estimation of ancestry in
374   unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).

375   23. N. Patterson, P. Moorjani, Y. Luo, S. Mallick, N. Rohland, Y. Zhan, T. Genschoreck, T.
376   Webster, D. Reich, Ancient admixture in human history. *Genetics*. **192**, 1065–1093
377   (2012).

378   24. C.-B. An, L. Tang, L. Barton, F.-H. Chen, Climate change and cultural response around
379   4000 cal yr B.P. in the western part of Chinese Loess Plateau. *Quat. Res.* **63**, 347–352
380   (2005).

381   25. M. J. C. Walker, M. Berkelhammer, S. Björck, L. C. Cwynar, D. A. Fisher, A. J. Long,
382   J. J. Lowe, R. M. Newnham, S. O. Rasmussen, H. Weiss, Formal subdivision of the
383   Holocene series/epoch: a discussion paper by a working group of INTIMATE
384   (integration of ice-core, marine and terrestrial records) and the subcommission on
385   Quaternary stratigraphy (International Commission on Stratigraphy). *J. Quat. Sci.* **27**,

14

386   649–659 (2012).

387   26.   R. E. Lanehart, R. H. Tykot, A. P. Underhill, F. Luan, H. Yu, H. Fang, C. Fengshu, G.
388   Feinman, L. Nicholas, Dietary adaptation during the Longshan period in China: stable
389   isotope analyses at Liangchengzhen (southeastern Shandong). *J. Archaeol. Sci.* **38**,
390   2171–2181 (2011).

391   27.   J. D. Guedes, M. Jiang, K. He, X. Wu, Z. Jiang, Site of Baodun yields earliest evidence
392   for the spread of rice and foxtail millet agriculture to south-west China. *Antiquity*. **87**,
393   758–771 (2013).

394   28.   J. D. Guedes, E. E. Butler, Modeling constraints on the spread of agriculture to
395   Southwest China with thermal niche models. *Quat. Int.* **349**, 29–41 (2014).

396   29.   R. Dal Martello, R. Min, C. Stevens, C. Higham, T. Higham, L. Qin, D. Q. Fuller, Early
397   agriculture at the crossroads of China and Southeast Asia: archaeobotanical evidence
398   and radiocarbon dates from Baiyangcun, Yunnan. *Journal of Archaeological Science:*
399   *Reports*. **20**, 711–721 (2018).

400   30.   D. Q. Fuller, A. R. Weisskopf, C. Castillo, Pathways of rice diversification across Asia.
401   *Archaeology International*. **19**, 84–96 (2016).

402   31.   J. d'Alpoim Guedes, G. Jin, R. K. Bocinsky, The impact of climate on the spread of rice
403   to north-eastern China: a new look at the data from Shandong province. *PLoS One*. **10**,
404   e0130430 (2015).

405   32.   G. W. Crawford, G.-A. Lee, Agricultural origins in the Korean Peninsula. *Antiquity*. **77**,
406   87–95 (2003).

407   33.   S.-M. Ahn, The emergence of rice agriculture in Korea: archaeobotanical perspectives.
408   *Archaeol. Anthropol. Sci.* **2**, 89–98 (2010).

409   34.   X. Yang, W. Wang, Y. Zhuang, Z. Li, Z. Ma, Y. Ma, Y. Cui, J. Wei, D. Q. Fuller, New
410   radiocarbon evidence on early rice consumption and farming in South China. *Holocene*.
411   **27**, 1045–1051 (2017).

412   35.   J. d'Alpoim Guedes, R. K. Bocinsky, Climate change stimulated agricultural innovation
413   and exchange across Asia. *Sci Adv*. **4**, eaar4491 (2018).

414   36.   S. A. Marcott, J. D. Shakun, P. U. Clark, A. C. Mix, A reconstruction of regional and
415   global temperature for the past 11,300 years. *Science*. **339**, 1198–1201 (2013).

416   37.   C. C. Castillo, D. Q. Fuller, P. J. Piper, P. Bellwood, M. Oxenham, Hunter-gatherer
417   specialization in the late Neolithic of southern Vietnam – the case of Rach Nui. *Quat.*
418   *Int*. **489**, 63–79 (2018).

419   38.   C. Higham, *The Bronze Age of Southeast Asia* (Cambridge University Press, 1996).

420   39.   C. F. W. Higham, Debating a great site: Ban Non Wat and the wider prehistory of
421   Southeast Asia. *Antiquity*. **89**, 1211–1220 (2015).

422   40.   C. C. Castillo, C. F. W. Higham, K. Miller, N. Chang, K. Douka, T. F. G. Higham, D. Q.

423   Fuller, Social responses to climate change in Iron Age north-east Thailand: new
424   archaeobotanical evidence. *Antiquity*. **92**, 1274–1291 (2018).

425 41. H. McColl, F. Racimo, L. Vinner, F. Demeter, T. Gakuhari, J. V. Moreno-Mayar, G. van
426   Driem, U. Gram Wilken, A. Seguin-Orlando, C. de la Fuente Castro, S. Wasef, R.
427   Shoocongdej, V. Souksavatdy, T. Sayavongkhamdy, M. M. Saidin, M. E. Allentoft, T.
428   Sato, A.-S. Malaspinas, F. A. Aghakhanian, T. Korneliussen, A. Prohaska, A.
429   Margaryan, P. de Barros Damgaard, S. Kaewsutthi, P. Lertrit, T. M. H. Nguyen, H.-C.
430   Hung, T. Minh Tran, H. Nghia Truong, G. H. Nguyen, S. Shahidan, K. Wiradnyana, H.
431   Matsumae, N. Shigehara, M. Yoneda, H. Ishida, T. Masuyama, Y. Yamada, A. Tajima,
432   H. Shibata, A. Toyoda, T. Hanihara, S. Nakagome, T. Deviese, A.-M. Bacon, P.
433   Duringer, J.-L. Ponche, L. Shackelford, E. Patole-Edoumba, A. T. Nguyen, B. Bellina-
434   Pryce, J.-C. Galipaud, R. Kinaston, H. Buckley, C. Pottier, S. Rasmussen, T. Higham, R.
435   A. Foley, M. M. Lahr, L. Orlando, M. Sikora, M. E. Phipps, H. Oota, C. Higham, D. M.
436   Lambert, E. Willerslev, The prehistoric peopling of Southeast Asia. *Science*. **361**, 88–92
437   (2018).

438 42. M. Lipson, O. Cheronet, S. Mallick, N. Rohland, M. Oxenham, M. Pietrusewsky, T. O.
439   Pryce, A. Willis, H. Matsumura, H. Buckley, K. Domett, G. H. Nguyen, H. H. Trinh, A.
440   A. Kyaw, T. T. Win, B. Pradier, N. Broomandkhoshbacht, F. Candilio, P. Changmai, D.
441   Fernandes, M. Ferry, B. Gamarra, E. Harney, J. Kampuansai, W. Kutanan, M. Michel,
442   M. Novak, J. Oppenheimer, K. Sirak, K. Stewardson, Z. Zhang, P. Flegontov, R.
443   Pinhasi, D. Reich, Ancient genomes document multiple waves of migration in Southeast
444   Asian prehistory. *Science*. **361**, 92–95 (2018).

445 43. A. Calò, *The distribution of bronze drums in early Southeast Asia: trade routes and
446   cultural spheres* (Archaeopress, 2009).

447 44. C. C. Castillo, B. Bellina, D. Q. Fuller, Rice, beans and trade crops on the early
448   maritime Silk Route in Southeast Asia. *Antiquity*. **90**, 1255–1269 (2016).

449 45. H.-C. Hung, Y. Iizuka, P. Bellwood, K. D. Nguyen, B. Bellina, P. Silapanth, E. Dizon,
450   R. Santiago, I. Datan, J. H. Manton, Ancient jades map 3,000 years of prehistoric
451   exchange in Southeast Asia. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 19745–19750 (2007).

452 46. H. Takamiya, M. J. Hudson, H. Yonenobu, T. Kurozumi, T. Toizumi, An extraordinary
453   case in human history: Prehistoric hunter-gatherer adaptation to the islands of the
454   Central Ryukyus (Amami and Okinawa archipelagos), Japan. *Holocene*. **26**, 408–422
455   (2016).

456 47. E. Zürcher, in *The Buddhist conquest of China* (Brill, 1972).

457 48. Z. Deng, L. Qin, Y. Gao, A. R. Weisskopf, C. Zhang, D. Q. Fuller, From early
458   domesticated rice of the middle Yangtze basin to millet, rice and wheat agriculture:
459   archaeobotanical macro-remains from Baligang, Nanyang basin, central China (6700-
460   500 BC). *PLoS One*. **10**, e0139885 (2015).

461 49. M. P. Verdugo, V. E. Mullin, A. Scheu, V. Mattiangeli, K. G. Daly, P. Maisano Delser,
462   A. J. Hare, J. Burger, M. J. Collins, R. Kehati, P. Hesse, D. Fulton, E. W. Sauer, F. A.
463   Mohaseb, H. Davoudi, R. Khazaeli, J. Lhuillier, C. Rapin, S. Ebrahimi, M. Khasanov, S.
464   M. F. Vahidi, D. E. MacHugh, O. Ertuğrul, C. Koukouli-Chrysanthaki, A. Sampson, G.

465   Kazantzis, I. Kontopoulos, J. Bulatovic, I. Stojanović, A. Mikdad, N. Benecke, J.
466   Linstädter, M. Sablin, R. Bendrey, L. Gourichon, B. S. Arbuckle, M. Mashkour, D.
467   Orton, L. K. Horwitz, M. D. Teasdale, D. G. Bradley, Ancient cattle genomics, origins,
468   and rapid turnover in the Fertile Crescent. *Science*. **365**, 173–176 (2019).

469   50. A. Gibbons, How the akkadian empire was hung out to dry. *Science*. **261**, 985 (1993).

470   51. J. Wang, L. Sun, L. Chen, L. Xu, Y. Wang, X. Wang, The abrupt climate change near
471   4,400 yr BP on the cultural transition in Yuchisi, China and its global linkage. *Sci. Rep.*
472   **6**, 27723 (2016).

473

474 *Acknowledgments*

Figure 1: Factors underlying geographic distribution of genomic diversity in japonica and indica. Maps of collection sites for (a) japonica and (b) indica landraces used in this study. Colors represent regions of origin. (c) Japonica and (d) indica effective migration surfaces representing migration barriers (orange) and channels (cyan). (e) Japonica and (f) indica genomic diversity is best explained by a combination of four factors represented in Euler plots: travel time (migration resistance) or geographic distance, abiotic variables (temperature, moisture and soil characteristics), linguistic group, and culinary properties (stickiness). Fields of squares represent total genomic variation, while elliptic shapes represent genomic variation explained by particular factor. (g) Japonica and (h) indica genotypes projected on the first two canonical axes of redundancy analysis. Arrows represent environmental predictors (acronyms explained in the legend) that strongly correlate with a maximal proportion of linear combinations of SNPs.
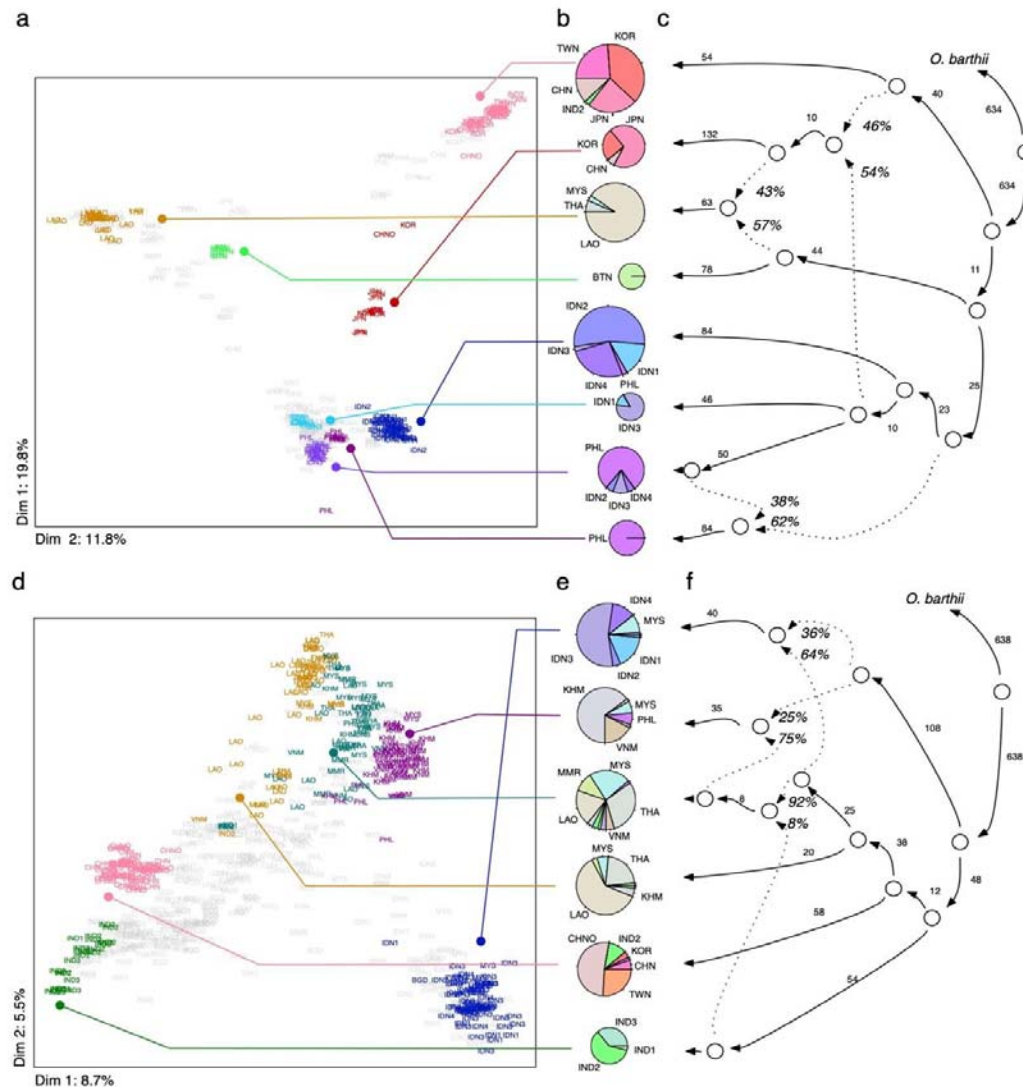
Figure 2: Japonica and indica rice subpopulations. (a) All japonica and (d) indica landraces projected onto first two dimensions after multidimensional scaling of genomic distances. (a) japonica genotypes were clustered using k-medoids (k = 9 subpopulations) and filtered using silhouette parameters, which resulted in K = 8 discrete subpopulations (colored labels). (d) indica genotypes were clustered using k-medoids (k = 7 subpopulations) and filtered resulting in K = 6 discrete subpopulations (colored labels). Pie charts representing the geographical composition of each discrete subpopulation of (b) japonica and (e) indica subgroups. Chart diameter is proportional to the number of individuals in each subpopulation. (c) Admixture graph for k = 9, K = 8 japonica subpopulations, rooted with *Oryza barthii* as an outgroup. This graph represents topology consistent between models for all lower k's. (f) Best admixture graph for k = 7, K = 6 indica subpopulations, rooted with *O. barthii* as an outgroup. Although this represents the best model, it is not consistent with other topologies at lower k's, likely due to complex history of indica. (c and f) Solid lines with arrowheads represent uniform ancestries (attached numbers show scaled drift parameter f_2), while dashed lines represent mixed ancestries (% values indicate estimated proportion of ancestry).
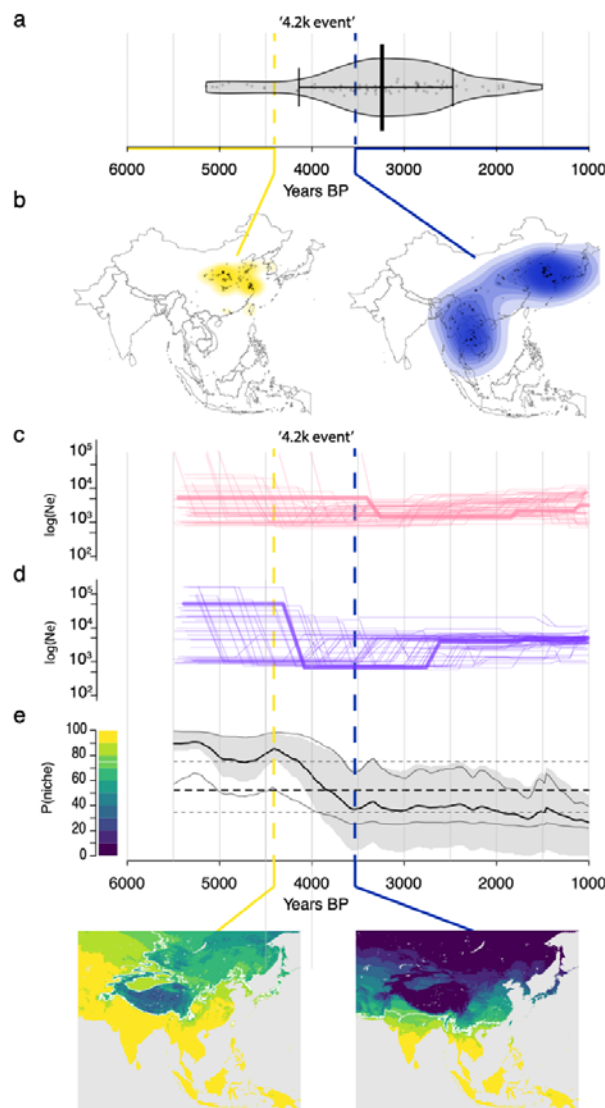
20

527
528 **Figure 3: Demographic, paleoenvironmental and archaeological context of temperate
529 japonica rice emergence.** (a) The distribution of temperate-tropical split times estimated
530 from cross-coalescence analysis carried out for 50 pairs of temperate and tropical individuals.
531 (b) Maps indicating geographic locations and densities of archaeological sites with rice
532 macro-remains. To the left: cumulative archaeobotanical evidence from 9,000-4,400 years
533 BP, to the right: cumulative archaeobotanical evidence from 3,500-1,000 years BP. Effective
534 population sizes over time in (c) tropical and (d) temperate japonica subpopulations. Thin
535 lines represent demographic histories for 50 randomly sampled individuals, while bold lines
536 represent joint models. (e) Probability of tropical rice being in the thermal niche (assuming
537 requirement of 2900 growing degree days, at 10°C base) over time. The mean (thick black
538 line) and the interquartile range, 25% to 75% (gray shaded area) of probability of being in the
539 thermal niche. The thin black lines are the mean probabilities using the lower and upper
540 confidence intervals of the temperature reconstruction. The two inset maps show the
541 geographic distribution of niche probabilities; to the left: before climate cooling (4,400 years
542 BP), to the right: after climate cooling (3,500 years BP).
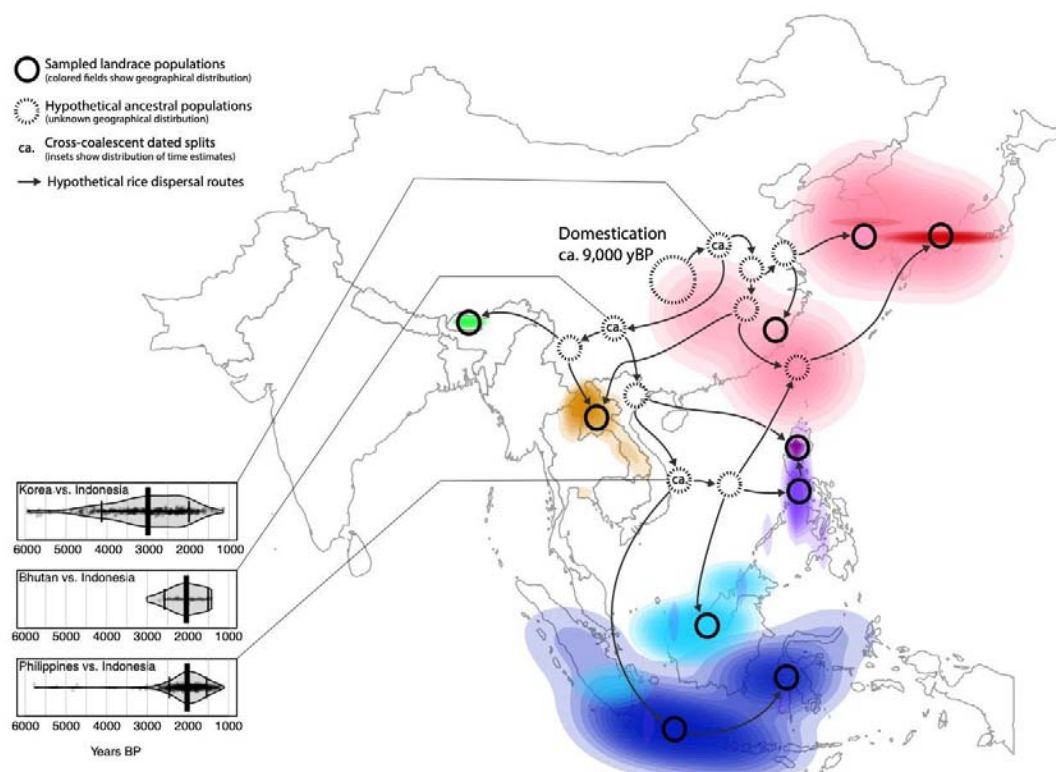
21

543
Figure 4: Proposed dispersal map of japonica rice in Asia. Map generated for japonica, K
= 8 discrete subpopulations. The geographic distributions of subpopulations were represented
as colored, two-dimensional Kernel density fields. Bold circles represent leaves in the
admixture graphs and are mapped close to the centers of subpopulation distributions. Dashed
circles represent hypothetical ancestral subpopulations inferred from splits in best-matching
admixture graphs; their precise geographic placement is uncertain. The distribution of split
times between non-admixed subpopulations was created from cross-coalescence estimates
summarized over all k levels and presented as violin plots. Arrows indicate hypothetical
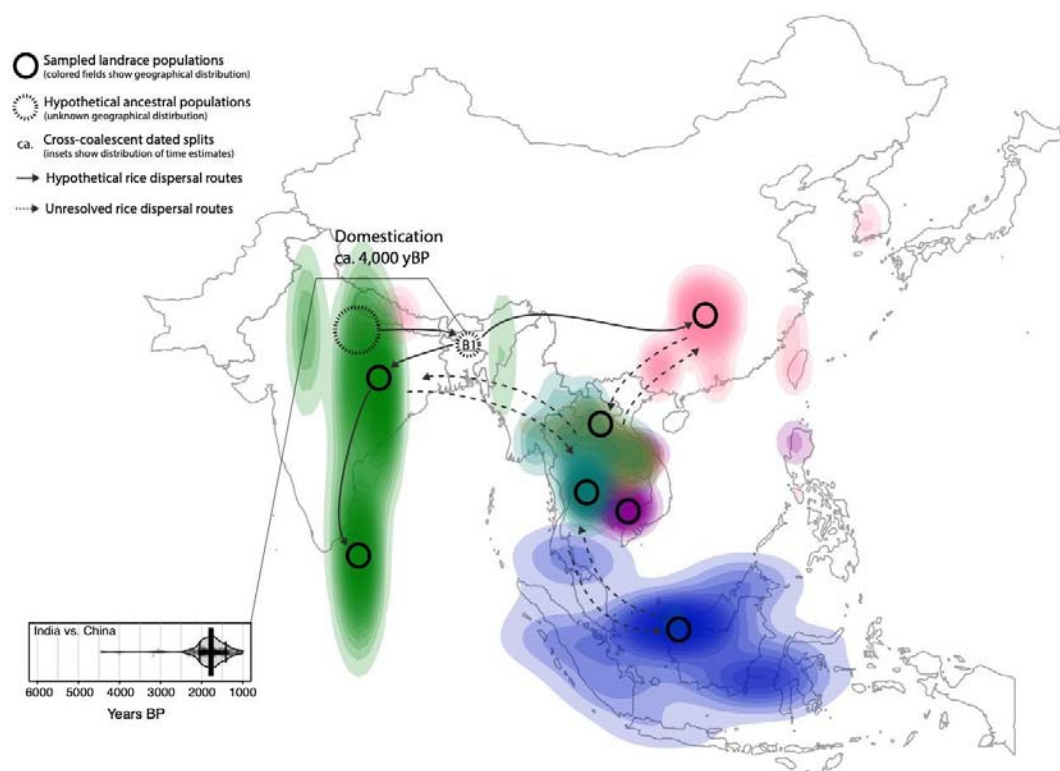routes of dispersal.

554
Figure 5: Proposed dispersal map of indica rice in Asia. Map generated for indica, K = 6
discrete subpopulations. The geographic distributions of subpopulations were represented as
colored, two-dimensional Kernel density fields. Bold circles represent leaves in the admixture
graphs and are mapped close to the centers of subpopulation distributions. Dashed circle
represents consistent split; its geographic position is uncertain. The distribution of split times
between non-admixed subpopulations was created from cross-coalescence estimates
summarized over all k levels and presented as violin plots. Solid arrows indicate hypothetical
routes of dispersal, while dotted arrows indicate possible routes that remain unresolved from
admixture graphs.

564