



Genomic adaptations to cereal-based diets contribute to mitigate metabolic risk in some human populations of East Asian ancestry

Running title: East Asian adaptations to cereal foods

Arianna Landini^{*1,2}, Shaobo Yu^{*1}, Guido Alberto Gnechi-Ruscione³, Paolo Abondio¹, Claudia Ojeda-Granados^{1,4}, Stefania Sarno¹, Sara De Fanti¹, Davide Gentilini^{5,6}, Anna M. Di Blasio⁶, Hanjun Jin⁷, Thanh Tin Nguyen⁸, Giovanni Romeo^{9,10}, Cecilia Prata¹¹, Eugenio Bortolini¹², Donata Luiselli¹², Davide Pettener¹, Marco Sazzini^{†,1,13}

¹Laboratory of Molecular Anthropology & Centre for Genome Biology, Department of Biological, Geological and Environmental Sciences, University of Bologna, Bologna, Italy

²Current address: Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, United Kingdom

³Department of Archaeogenetics, Max Planck Institute for the Science of Human History, Jena, Germany

⁴Department of Molecular Biology in Medicine, Civil Hospital of Guadalajara “Fray Antonio Alcalde” and Health Sciences Center, University of Guadalajara, Guadalajara, Jalisco, Mexico

⁵Department of Brain and Behavioral Sciences, University of Pavia, Pavia, Italy

⁶Italian Auxologic Institute IRCCS, Cusano Milanino, Milan, Italy

⁷Department of Biological Sciences, College of Natural Science, Dankook University, Cheonan, South Korea

⁸Hue University of Medicine and Pharmacy, Hue, Vietnam

⁹Medical Genetics Unit, S. Orsola Hospital, University of Bologna, Bologna, Italy

¹⁰European School of Genetic Medicine, Italy

¹¹Department of Pharmacy and Biotechnology, University of Bologna, Bologna, Italy

¹²Department of Cultural Heritage, University of Bologna, Ravenna, Italy

¹³Interdepartmental Centre Alma Mater Research Institute on Global Challenges and Climate Change, University of Bologna, Bologna, Italy

* These authors contributed equally to this work

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1111/EVA.13090](https://doi.org/10.1111/EVA.13090)

This article is protected by copyright. All rights reserved

† Corresponding author: E-mail: marco.sazzini2@unibo.it

Abstract

Adoption of diets based on some cereals, especially on rice, signified an iconic change in nutritional habits for many Asian populations and a relevant challenge for their capability to maintain glucose homeostasis. Indeed, rice shows the highest carbohydrates content and glycaemic index among the domesticated cereals and its usual ingestion represents a potential risk factor for developing insulin resistance and related metabolic diseases. Nevertheless, type 2 diabetes and obesity epidemiological patterns differ among Asian populations that rely on rice as a staple food, with higher diabetes prevalence and increased levels of central adiposity observed in people of South Asian ancestry rather than in East Asians. This may be at least partly due to the fact that populations from East Asian regions where wild rice or other cereals such as millet have been already consumed before their cultivation and/or were early domesticated have relied on these nutritional resources for a period long enough to have possibly evolved biological adaptations that counteract their detrimental side effects. To test such a hypothesis, we compared adaptive evolution of these populations with that of control groups from regions where the adoption of cereal-based diets occurred many thousand years later and which were identified from a genome-wide dataset including 2,379 individuals from 124 East Asian and South Asian populations. This revealed selective sweeps and polygenic adaptive mechanisms affecting functional pathways involved in fatty acids metabolism, cholesterol/triglycerides biosynthesis from carbohydrates, regulation of glucose homeostasis and production of retinoic acid in Chinese Han and Tujia ethnic groups, as well as in people of Korean and Japanese ancestry. Accordingly, long-standing rice- and/or millet-based diets have possibly contributed to trigger the evolution of such biological adaptations, which might represent one of the factors that play a role in mitigating the metabolic risk of these East Asian populations.

Keywords

human Asian populations, dietary selective pressures, genomic adaptation, metabolic risk, evolutionary medicine

Acknowledgments

We would like to thank all donors who participated to the study without whom this work would have not been possible. We are also grateful to Pier Massimo Zambonelli (CESIA, University of Bologna) for his IT assistance. S.S. and S.D.F. were supported by the ERC AdG n. 295733 to D.P.

CONFLICT OF INTEREST

None declared.

DATA AVAILABILITY STATEMENT

The genotype data generated during the current study will be made available at the figshare account of the corresponding author (https://figshare.com/authors/Marco_Sazzini/6292466).

DR. MARCO SAZZINI (Orcid ID : 0000-0001-5382-7827)

Article type : Original Article

Genomic adaptations to cereal-based diets contribute to mitigate metabolic risk in some human populations of East Asian ancestry

Running title: East Asian adaptations to cereal foods

Arianna Landini^{*1,2}, Shaobo Yu^{*1}, Guido Alberto Gnechi-Ruscione³, Paolo Abondio¹, Claudia Ojeda-Granados^{1,4}, Stefania Sarno¹, Sara De Fanti⁵, Davide Gentilini^{6,7}, Anna M. Di Blasio⁷, Hanjun Jin⁸, Thanh Tin Nguyen⁹, Giovanni Romeo^{10,11}, Cecilia Prata¹², Eugenio Bortolini¹³, Donata Luiselli¹³, Davide Pettener¹, Marco Sazzini^{†,1,5}

¹Laboratory of Molecular Anthropology & Centre for Genome Biology, Department of Biological, Geological and Environmental Sciences, University of Bologna, Bologna, Italy

²Current address: Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, United Kingdom

³Department of Archaeogenetics, Max Planck Institute for the Science of Human History, Jena, Germany

⁴Department of Molecular Biology in Medicine, Civil Hospital of Guadalajara “Fray Antonio Alcalde” and Health Sciences Center, University of Guadalajara, Guadalajara, Jalisco, Mexico

⁵Interdepartmental Centre Alma Mater Research Institute on Global Challenges and Climate Change, University of Bologna, Bologna, Italy

⁶Department of Brain and Behavioral Sciences, University of Pavia, Pavia, Italy

⁷Italian Auxologic Institute IRCCS, Cusano Milanino, Milan, Italy

⁸Department of Biological Sciences, College of Natural Science, Dankook University, Cheonan, South Korea

⁹Hue University of Medicine and Pharmacy, Hue, Vietnam

¹⁰Medical Genetics Unit, S. Orsola Hospital, University of Bologna, Bologna, Italy

¹¹European School of Genetic Medicine, Italy

¹²Department of Pharmacy and Biotechnology, University of Bologna, Bologna, Italy

¹³Department of Cultural Heritage, University of Bologna, Ravenna, Italy

* These authors contributed equally to this work

† Corresponding author: E-mail: marco.sazzini2@unibo.it

ABSTRACT

Adoption of diets based on some cereals, especially on rice, signified an iconic change in nutritional habits for many Asian populations and a relevant challenge for their capability to maintain glucose homeostasis. Indeed, rice shows the highest carbohydrates content and glycaemic index among the domesticated cereals and its usual ingestion represents a potential risk factor for developing insulin resistance and related metabolic diseases. Nevertheless, type 2 diabetes and obesity epidemiological patterns differ among Asian populations that rely on rice as a staple food, with higher diabetes prevalence and increased levels of central adiposity observed in people of South Asian ancestry rather than in East Asians. This may be at least partly due to the fact that populations from East Asian regions where wild rice or other cereals such as millet have been already consumed before their cultivation and/or were early domesticated have relied on these nutritional resources for a period long enough to have possibly evolved biological adaptations that counteract their detrimental side effects. To test such a hypothesis, we compared adaptive evolution of these populations with that of control groups from regions where the adoption of cereal-based diets occurred many thousand years later and which were identified from a genome-wide dataset including 2,379 individuals from 124 East Asian and South Asian populations. This revealed selective sweeps and polygenic adaptive mechanisms affecting functional pathways involved in fatty acids metabolism, cholesterol/triglycerides biosynthesis from carbohydrates, regulation of glucose homeostasis and production of retinoic acid in Chinese Han and Tujia ethnic groups, as well as in people of Korean and Japanese ancestry. Accordingly, long-standing rice- and/or millet-based diets have possibly contributed to trigger the evolution of such biological adaptations, which might

represent one of the factors that play a role in mitigating the metabolic risk of these East Asian populations.

KEYWORDS

human Asian populations, dietary selective pressures, genomic adaptation, metabolic risk, evolutionary medicine

INTRODUCTION

During the recent evolutionary history of *H. sapiens*, several dietary shifts are supposed to have introduced new selective pressures on human groups from different ecoregions, especially due to domestication of a variety of plants and animals (Lucock, 2007; Luca et al., 2010; Brown, 2012; Sazzini et al., 2016; Bayoumi et al., 2016; Sazzini et al., 2020). For instance, the adoption of a diet based on cereals, especially on rice, represented an iconic change in nutritional habits for many Asian populations and one of the most relevant challenges experienced by the human metabolism. In fact, this determined a remarkable increase in the intake of carbohydrate-rich foods and a shift towards a high glycaemic load (GL) diet, which is expected to have affected the capability of these populations to maintain a physiological glucose homeostasis. To this end, unpolished brown rice can be considered as a reliable proxy for the rice varieties long consumed by the ancestors of Asian people and both its traditional and modern cultivars present some of the highest carbohydrates content and glycaemic index (GI) among the domesticated cereals (Fitzgerald et al., 2011). Indeed, by considering different varieties it shows on average $GI = 88 \pm 5$ and $GL = 20$, while for example wheat GI and GL are attested respectively to 30 ± 9 and 11 (Atkinson, Foster-Powell, & Brand-Miller, 2008). This GI/GL difference between rice and other grains is even more striking when considering the polished white rice ($GI = 103 \pm 11$, $GL = 29$), which is consumed as a staple food by modern Asian populations. Therefore, rice has long acted and is still acting as the primary contributor to the dietary GL of Asian groups. By prompting rapid postprandial increase in glycaemia its usual and massive ingestion represents one of the potential risk factors for developing insulin resistance and related metabolic diseases, such as type 2 diabetes (T2D) and obesity (Hu et al., 2012; Boers, Seijen Ten Hoorn, & Mela, 2015). This is consistent with the fact that after the changes in dietary habits and lifestyle experienced by most Asian populations in the last decades, the number of individuals affected by T2D in these human groups has rapidly increased, until representing around 60% of the

world's diabetic subjects (Hu, 2011). The fast socio-economic transitions that have interested Asian countries have indeed led to a further increment in the intake of calories and high-GI foods due to augmented consumption of oils, fats, refined grains and sugars, dairy and convenience foods (Kelly, 2016). This was coupled also with the decrease in physical activity due to urbanization, adoption of sedentary jobs and motorised transports (Ramachandran et al., 2004; Wu, 2006; Yoon et al., 2006; Ramachandran & Snehalatha 2010; Ramachandran et al., 2012). In particular, substantial shifts from traditional to even more caloric and high-GL Westernized-type dietary patterns are expected to have exacerbated metabolic risk in East Asian groups such as those from China and Japan. Indeed, these populations considerably increased their utilisation of dairy, butter, and meat, along with the consumption of high-GI foods, such as confectionaries, jam, fruit, sugars, bread, noodles and potatoes (Popkin et al., 2001; Tomisaka et al., 2002; Murakami et al., 2006; Gabriel et al., 2018). For instance, Chinese and Japanese populations present a daily energy intake ranging from around 2,400 to 2,800 kcal. This is largely ascribable to an average consumption of respectively 450 g and 286 g of cereals per day (~ 62% of which represented by rice) and to a remarkable increase in the proportion of dietary fats (on average 26.4% of the total calorie entry). Moreover, as many other East Asian groups they have recently gone through an escalation of the per capita consumption of sugars (on average ~ 50 g/day) (Murakami et al., 2006; Food and Agriculture Organization of the United Nations, 2008; Zhai et al., 2009; Fujiwara et al., 2018). A similar, but quantitatively less pronounced nutrition transition has been experienced also by populations from the Indian subcontinent. In fact, they show lower daily energy intake (2,020-2,047 kcal), cereals consumption (368 g per day, ~ 52% of which represented by rice) and, percentage of dietary fats (16-21%) than Chinese and Japanese groups (Radhika et al. 2009; Dixit et al., 2011; Misra et al., 2011; Bharati et al., 2020). Conversely, they consume more fiber (13.3 g/day) with respect to Chinese and Japanese people (9.7 g/day and 7.6 g/day, respectively) and a slightly higher amount of sugars (55.3 g/day) (Dasgupta et al., 2015). Nevertheless, per capita sugar consumption was found to be significantly correlated with T2D prevalence in Central and South East Asian populations, but not in South Asian and East Asian ones, suggesting that other factors may contribute mostly to the high-GL diet of these human groups (Praveen et al., 2014). Interestingly, in contrast to what expected according to the described dietary shifts, increasing rates in the prevalence of T2D were found to be substantially higher in South Asian populations rather than in East Asians (e.g. 12.1% and 10.6% in India and Pakistan versus 6.1% and 7.6% in China and Korea) (Yoon et al., 2006; Chan et al., 2009; Ramachandran et

al., 2010; Ma & Chan, 2013). These peculiar epidemiological patterns may suggest that when coupled with recent transitions in lifestyle and dietary habits, and with subsequent increases in life expectancy, cereal-based diets, particularly those relying on rice as a staple food, might have further increased T2D/obesity susceptibility especially in populations from the Indian subcontinent. In fact, it seems that an analogous increase in metabolic risk is somehow mitigated at least in some East Asian groups, suggesting that their ancestors might have evolved adaptations triggered by several millennia of abundant rice (or other cereals) consumption that contribute to reduce the potential side effects of medium-to-high GI foods. In particular, we propose that populations from East Asian regions where wild rice (*Oryza rufipogon*) originated and was early domesticated have heavily relied on such a nutritional resource for a sufficient time (i.e. around 10,000 years) to have evolved genetic adaptations to it. In this regard, the Yangtze River valley in Eastern China is the geographical area where the oldest archaeological evidence of habitual wild rice consumption (dated to 12,000-11,000 years ago) was found (Zhao, 1998; Jiang & Liu, 2006) and where the modern rice subspecies *O. sativa japonica* was domesticated (Gross & Zhao, 2014; Silva et al., 2015; Silva et al., 2018). Accordingly, human groups from this region are supposed to have introduced rice in their diet long before its domestication and the development of cultivation techniques, which have been established 7,000-6,000 years ago after an extended period of pre-domesticated cultivation and a much longer period of wild rice use by foragers (Cao et al., 2006; Fuller et al., 2007). Afterwards, rice agriculture diffused from China to the Korean peninsula and Japan, replacing the indigenous cultivations of millet and beans (Gross & Zhao, 2014). A similar pattern might be hypothesized also for broomcorn millet (*Panicum miliaceum*) and foxtail millet (*Setaria italica*), whose domestication was completed 8,000-6,000 years ago in the Hebei and Manchuria provinces of Northern China (Zhao, 2011; Stevens et al., 2016; Deng et al., 2018), and whose cultivation was in parallel established in nearby Korean and Japanese regions (Gross & Zhao, 2014). Conversely, wheat consumption has played a marginal role in the diet of ancestral East Asian populations, having been introduced only 4,000 years ago and having remained circumscribed to few Chinese regions at least up to the 6th century AD (Zhou & Garvie-Lok, 2015). Moreover, it long represented a prerogative of a small fraction of these populations given that wheat was generally not cultivated as a high production subsistence but was instead consumed as a rare exotic good by social elites (Liu & Jones, 2014; Long et al., 2018).

An independent centre of rice domestication but involving a different subspecies (i.e. *O. sativa indica*) has been identified also in the Indo-Gangetic Plain crossing northern regions of

the Indian subcontinent (Molina et al., 2011). However, such a domestication process was completed only several thousand years later with respect to what occurred in the Yangtze River Valley, when the subspecies *O. sativa japonica* was introduced from China and hybridized with the local proto-*indica* rice (Gross & Zhao, 2014; Choi et al., 2017; Silva et al., 2018). Furthermore, *O. sativa japonica* was selected by East Asian farmers in response to specific dietary habits and this resulted in its stickier grains and glutinous phenotype, which are associated also to low amylose levels (0-20%) (Yamanaka et al., 2004; Olsen et al., 2006). Conversely, the preference for distinct, non-cohesive rice grains in South Asian cultures prevented the introgression of these traits into the *indica* variety (Fuller et al., 2016), which is indeed characterized by a higher amount of amylose (23-31%) than *O. sativa japonica* (Kaur, Ranawana, & Henry, 2016). Interestingly, inverse correlation between rice amylose content and GI was observed, with East Asian cultivars showing remarkably higher GI (on average ~100) with respect to South Asian ones (average GI = 60) (Fitzgerald et al., 2011), thus having potentially represented a more challenging selective pressure for the human metabolism.

Overall, these evidences suggest that the ancestors of modern Chinese, Korean and Japanese populations have adopted high GL cereal-based diets for a period of time considerably longer than other East Asian and South Asian human groups. Therefore, they might have had the opportunity to evolve genetic adaptations in response to such a peculiar dietary regimen. To test such a hypothesis, we assembled a genome-wide “Pan-Asian” dataset including 2,379 individuals from 124 East Asian and South Asian populations selected to be representative of human genetic variation observable at geographical areas where wild rice or millet originated and/or were early domesticated (i.e. candidate populations) or where the adoption of cereal-based diets occurred only several thousand years later (i.e. control populations) (Table S1). We then identified genetically homogenous population clusters and we investigated their adaptive evolution by searching for genomic signatures ascribable to the action of natural selection according to models of hard/soft selective sweeps (i.e. events of strong natural selection on single/few loci with large effects on a given phenotypic trait) and polygenic adaptation (i.e. events of weak, but pervasive natural selection on many loci functionally related with each other, but individually showing small effects on a given phenotypic trait). This enabled us to shortlist adaptive events at several genes playing a role primarily in fatty acids metabolism, regulation of glucose homeostasis and production of retinoic acid, which have been possibly evolved in response to rice- and/or millet-based diets by Chinese populations of Han and Tujia ancestry, Korean and Japanese people. Therefore, we provided

evidence for genetic adaptations of these human groups that might contribute to mitigate the side effects of their diets, thus shedding new light on the possible evolutionary causes having partly influenced the present-day epidemiological patterns of metabolic diseases observed across the Asian continent.

METHODS

Samples collection and genotyping

Saliva from 85 individuals representative of Korean (N = 36), Bangladeshi (N = 29) and Vietnamese (N = 20) populations was sampled using the Oragene DNA OG-500 kit (DNA Genotek, Ottawa, Ontario, Canada) and according to ethnographic information. Informed consent was also obtained from all individuals and on 04/08/2013 the University of Bologna ethics committee released approval for the present study (within the framework of the ERC-2011-AdG 295733 project), which was designed and conducted in accordance with relevant guidelines and regulations and according to ethical principles for medical research involving human subjects stated by the WMA Declaration of Helsinki. DNA was extracted from saliva samples using the *prepit-L2P* protocol (DNA Genotek, Ottawa, Ontario, Canada) and quantified with the Quant-iT dsDNA Broad-Range Assay Kit (Invitrogen Life Technologies, Carlsbad, CA, USA). DNA samples were genotyped for ~720,000 genome-wide SNPs by means of the HumanOmniExpress v 1.1 chip (Illumina, San Diego, CA, USA) at the Center for Biomedical Research & Technologies of the Italian Auxologic Institute (Milan, Italy).

Data curation and assembly of a Pan-Asian dataset

Quality control (QC) procedures were performed on the generated genome-wide data using PLINK v.1.9 (Purcell et al., 2007). Only autosomal loci with a genotyping success rate higher than 95% and no significant deviations from the Hardy-Weinberg equilibrium ($p > 1.41 \times 10^{-8}$) were retained. Individuals showing more than 5% of missing data were excluded, together with ambiguous strand single nucleotide polymorphisms (SNPs) with A/T or G/C alleles. The obtained dataset was then submitted to linkage-disequilibrium (LD) pruning and the identified SNPs in approximate LD with each other ($r^2 < 0.1$) were used to estimate the degree of recent shared ancestry (IBD) for each pair of subjects by calculating the genome-wide proportion of shared alleles (i.e. identity by state, IBS). Individuals showing IBD kinship coefficient higher than 0.1 were removed. A “high-quality” dataset of 82 samples typed for 688,987 SNPs was thus generated. An “extended” dataset of 4,356 samples

belonging to 162 worldwide populations typed for 231,947 SNPs was also obtained by merging the “high-quality” dataset with publicly available data retrieved from the 1000 Genomes Project phase 3 (1000 Genomes Project Consortium et al., 2015), the HGDP project (Li et al., 2008), as well as literature studies focused on South Asian and East Asian groups (Table S1). Such an “extended” dataset was prepared for population structure and admixture analyses by LD-pruning ($r^2 < 0.2$) and by removing sites with a minor allele frequency (MAF) below 0.01. Haplotypes phasing was performed with SHAPEIT2 v2.r790 (Delaneau et al., 2008) using default parameters and HapMap phase 3 recombination maps.

Population structure analyses

Principal Component Analysis (PCA) was applied first to the “extended” dataset to check for the presence of potential genotypes inconsistency due to errors occurred in the merging procedure and then to a “Pan-Asian” subset of 2,379 individuals belonging to 124 Asian populations. To compute PCA, the *smartpca* method implemented in the EIGENSOFT package v6.0.1 (Patterson et al., 2006) was used. To depict an overall picture of ancestry proportions for each subject belonging to a further representative subset of 1,171 samples from 57 Asian populations selected according to PCA results, and to test for the presence of genetically homogeneous population clusters, we applied the model-based clustering algorithm implemented in ADMIXTURE v1.22 (Alexander et al., 2009) assuming $K = 2$ to $K = 10$ clusters. This infers the relative proportions of the different genetic components observable within each genome and which are due to the gene flow that generally occurred between human groups. Therefore, although being formally a clustering analysis, it is commonly used to explore patterns of shared genetic ancestry among populations. We ran 50 replicates with different random seeds for each K to monitor for convergence and only those presenting the highest log-likelihood values were plotted. Concurrently, we calculated cross-validation (CV) errors for each K to identify the most plausible number of ancestry components.

Selection scans on the identified population clusters

We computed the Number of Segregating sites by Length (nSL) statistics (Ferrer-Admetlla et al., 2014) to investigate both hard and soft selective sweeps occurred in the genomes ascribable to the clusters pointed out by population structure analyses. The nSL test quantifies the extension of haplotype homozygosity around each of the considered SNPs and compares the patterns obtained for haplotypes carrying the ancestral allele or the derived allele to

pinpoint genomic regions that have been plausibly subjected to natural selection. Information about the ancestral and derived alleles at each SNPs was retrieved by considering the human ancestor alignment, which was obtained through multiple alignment of six primate genomes to the human reference sequence build 37 (UCSC hg19). We then used algorithms implemented in Selscan v1.1.0b (Szpiech & Hernandez 2014) to compute nSL scores for each SNP by considering a 20,000 bp threshold for gap scale and 200,000 bp as the maximum length a gap can reach, otherwise the calculation was aborted. Since the nSL statistics measures the distance between SNPs in terms of number of polymorphic sites across the considered genomic region, we set a maximum extension of 25 SNPs according to the SNPs density of our dataset. Unstandardized nSL scores were normalized by classifying SNPs in frequency bins across the genome and by subtracting to each value the mean nSL score in that bin and by dividing by the associated standard deviation.

Shortlisting of the most informative candidate adaptive genes

To minimize false positive results (i.e. isolated SNPs showing unusual nSL scores) that might be ascribable to stochastic variation in allele frequency due to genetic drift, we applied a 200,000 bp sliding windows procedure on the obtained genome-wide nSL distribution. Derived allele frequency (DAF) of each SNP was computed to enable their classification into bins based on DAF (Piras et al. 2012). We retained outlier SNPs in a given genomic window as those showing absolute nSL values falling within the 99th percentile of the related DAF-based bin distribution and we further classified windows into bins according to their total number of SNPs. Consecutive windows were merged together and the top 1% of windows for each bin of regions was extracted according to their proportion of outlier SNPs with respect to the overall number of SNPs in such a window. This enabled us to detect large genomic segments enriched for unusual nSL scores. The HumanOmniExpress bead chip annotation file (Illumina, San Diego, CA, USA) was used to assign genomic regions enriched for outlier SNPs to known genes. For each population clusters, we then investigated known functional relationships between genes included in the top 1% windows by using information from the STRING v10.5 protein-protein interactions database (Jensen et al., 2009). By assigning to each SNP the ranking value previously calculated for the respective window, a final ranking of merged chromosomal intervals was obtained according to the computation of the arithmetical mean of scores of the windows involved. Finally, the most informative candidate adaptive genes for each population cluster were identified as those included in the top 1% merged windows showing a proportion of outlier SNPs above 0.5.

Gene network analyses aimed at testing for polygenic adaptation

Population clusters pointed out by the performed selection scan as the most plausible groups having evolved adaptations in response to rice or millet-based diets were further investigated to explicitly test the occurrence of selective events according to a polygenic adaptation model. This assumes that adaptive traits have been modulated by weak positive selection acting concurrently on multiple loci belonging to specific gene networks rather than on single genes (Gnecchi-Ruscione et al., 2018), thus representing a more realistic description of the potential action of natural selection on the human genome. For this purpose, we computed the nSL statistics on whole genome sequence data generated by the 1000 Genomes Project for Han Chinese (CHB) and Japanese (JPT) populations (1000 Genomes Project Consortium et al., 2015) and we submitted the obtained genome-wide distributions to the gene network algorithm implemented in the *signet* R package (Gouy et al., 2017), as detailed in Gnecchi-Ruscione et al. (2018). Assignment of genes to the related functional pathways was performed according to the Kyoto Encyclopedia of Genes and Genomes (KEGG) database and significant shifts ($p < 0.05$) towards extreme *signet* values were searched for in the distribution of scores observed within annotated pathways to identified gene networks pervasively subjected to positive selection. Networks of candidate adaptive genes were finally plotted using Cytoscape v3.6.0 (Shannon et al., 2003).

RESULTS

Exploring population structure in the assembled Pan-Asian dataset

By applying stringent QC procedures to the generated genome-wide data and by merging them with publicly available genotypes (see Methods), we obtained an “extended” dataset of 4,356 unrelated samples belonging to 162 worldwide populations characterized for 231,947 SNPs. When we submitted it to PCA, we observed the expected cline “V-shaped” distribution of samples that reflects the main differentiation between populations of sub-Saharan African ancestry (e.g. Esan and Yoruba from Nigeria, Bantu, Mbuti and Biaka hunter-gatherers, Mandenka, Luhya from Kenya) and non-African groups (Figure S1), as pointed out by previous studies (Li et al., 2008; Mallick et al., 2016).

To focus on the overall genomic landscape of long-term cereal-eating populations, we repeated PCA by retaining a “Pan-Asian” subset of 2,379 individuals belonging to 124 populations from South and East Asia. Accordingly, PC1 was found to explain 4.3% of the

investigated genomic variation, enabling to distinguish between populations from the Indian subcontinent and groups from East Asian/South East Asian regions (Figure 1a and Figure S2). PC2 instead accounted for 0.61% of variance and described two latitudinal clines of diversity within both of these macro-geographical areas. On one side, South Asian populations ranged from Pakistani ethnic groups and Northern/Central Indian people speaking Indo-European languages to Southern Indian tribes speaking Dravidian languages and Austro-Asiatic speaking populations from Northeast India (Figure 1a), as previously described (Reich et al., 2009; Metspalu et al., 2011; Basu et al., 2016). On the other side, an East Asian gradient of variation was delimited by the opposite ends represented by high-altitude Himalayan groups and the Malay and other Austronesian populations (Figure 1a), reflecting the occurrence of complex demographic and adaptive processes more than a simple distribution along a latitudinal cline, as suggested by extensive literature (HUGO Pan-Asian SNP Consortium, 2009; Gneccchi-Ruscione et al., 2017; Gneccchi-Ruscione et al., 2018). Uygurs, Hazaras, and groups from Northeast India and Nepal that speak Tibeto-Burman languages instead occupied an intermediate position in the PCA space between the two previously described clines of variation (Figure 1a) in accordance to the admixture events they have experienced in historical times (Li et al., 2008; Gneccchi-Ruscione et al., 2017). Finally, individuals from the Andaman and Nicobar Islands and belonging to the Onge and Jarawa populations represented genetic outliers with respect to the bulk of the examined South Asian and East Asian groups mainly due to their shared ancestry with Oceanic Pacific Islanders and prolonged isolation, as already proposed by Basu et al. (2016).

Identification of genetically homogeneous population clusters

PCA results were used to refine the assembled “Pan-Asian” dataset by removing populations representing unusual outliers within the considered genomic landscape. We thus selected 1,171 samples belonging to 57 populations well representative of the described gradients of South Asian and East Asian variation and we submitted them to ADMIXTURE analyses to identify population clusters showing high internal genetic homogeneity, and thus having presumably shared a common genetic history. We tested $K = 2$ to $K = 10$ potential ancestral groups (Supplementary Results and Figure S3) and the best predictive accuracy was achieved by the model when eight ancestry components ($K = 8$) were assumed (Figure S4). The overall picture of genetic components recognized within each population at $K = 8$ (Figure 1b) was concordant with patterns of population structure pointed out by PCA and with findings from previous studies. In particular, Pakistani people were found to be almost entirely

characterized by the “Ancestral North Indian” component already identified by Reich et al. (2009), which was highly represented also in Indian Indo-European speaking populations, while an “Ancestral South Indian” component was predominant in Indian groups speaking Dravidian languages (Metspalu et al., 2011) (Figure 1b). Two additional ancestry fractions characterized Indian Austro-Asiatic speaking tribes and speakers of Tibeto-Burman languages (Figure 1b), as previously observed by Basu et al. (2016) and Gneccchi-Ruscione et al. (2017). Some Tibeto-Burman groups (e.g. Burmese and Lahu) also showed a minor proportion of the genome ascribable to the “Ancestral South Indian” component, which was evident even in the Malays (Figure 1b) and pointed to a reduced Indian gene flow to South East Asia in contrast to a documented cultural influence since at least 2,500 years ago (Mörseburg et al., 2016). A further component was found to be associated mainly to northern East Asian populations, representing a unique genetic signature for Japanese people and reaching high percentages in Koreans. On the contrary, a different component appeared to be related to southern East Asian populations, being overwhelming among Cambodian, Dai, Lahu, Vietnamese and Austronesian groups. A specific ancestry fraction plausibly associated to the expansion of Austronesian languages and the related demographic processes (Mörseburg et al., 2016) was observed to be predominant in the Igorots and was significantly represented in the Dusuns and Muruts from Borneo, as well as in the Malays and some Filipino groups (e.g. Luz and Vizaya) (Figure 1b). Finally, Birhor people showed an almost fixed ancestry component, which was widespread, although at low percentages, also in all South Asian and in many South East Asian populations (Figure 1b).

Distribution of ancestry components depicted by ADMIXTURE analysis at $K = 8$ enabled us to identify 10 different population clusters characterized by high internal genetic homogeneity (i.e. made up by individuals presenting the same ancestry components admixed at approximately the same proportions) and that grouped together populations showing also close geographical proximity, similar cultural patterns, and largely shared demographic histories (Figure 2). In particular, we identified the *Pakistani cluster* (Makrani, Balochi and Brahui individuals; $N = 74$); the *North West Indian cluster* (Pathan, Khatri, Sindhi, Punjabi people and Gujarati Brahmins; $N = 141$); the *Central South Indian cluster* (Gujarati people, inhabitants of the Uttar Pradesh, Telegu and other speakers of Dravidian languages, Sri Lankan Tamil, Maratha and Pallan castes; $N = 410$); the *Bangladeshi cluster* ($N = 114$); the *Tibeto-Burman cluster* (Jamatia, Tripuri, Lahu and people from Myanmar; $N = 80$); the *Dai-Vietnamese cluster* ($N = 239$); the *Japanese cluster* ($N = 132$); the *Han-Tujia cluster* ($N =$

262); the *Korean cluster* (N = 36); and the *South East Asian cluster* (Malays, Murut, Dusun, Vizaya, and Luz; N = 101).

Detection of genomic regions enriched for selective sweeps in each population cluster

By assuming that the ancestors of populations belonging to the same cluster have been plausibly subjected to similar selective pressures due to their largely shared environmental and cultural settings, we calculated genome-wide distributions of the nSL statistics for each group of populations to detect the main selective sweeps having mediated their adaptive histories. To minimize false positive results, we searched for the most plausible candidate chromosomal intervals to have undergone positive selection by retaining top 1% genomic regions in the distribution of 200 kb windows ranked according to their proportion of outlier SNPs showing unusual nSL scores. We finally investigated functional relationships among genes laying in these top 1% windows according to information retrieved from the STRING protein-protein interactions database (see Methods).

Large genomic intervals enriched for signatures of positive selection thus turned out to be differentiated especially between clusters from the Indian subcontinent or South East Asia and groups from the northernmost East Asian regions covered by the dataset.

In particular, all South Asian clusters presented candidate adaptive genes that play a role in the ubiquitination process (Figure S5-S9), a reversible post-translational modification of cellular proteins able to regulate a broad set of mechanisms including cell division, differentiation, signal transduction and protein trafficking (Mukhopadhyay & Riezman 2007). In detail, the *FBXW2* ring finger proteins class (RNF), *LTN1*, *SPSB4*, *ZBTB16* and ubiquitin-conjugating enzyme class (UBE) genes were some of the several targets of selection shared among the majority of South Asian and South East Asian clusters (Figure S5-S10). The Pakistani, Tibeto-Buman, South East Asian and Dai-Vietnamese groups also shared selection signatures on many genes functionally related to the *PIK3CA* locus (Figure S1, S9-S11). This contributes to the activation of signalling cascades involved in cell growth, survival, proliferation, motility and morphology, indirectly through the generation of phosphatidylinositol 3,4,5 trisphosphate. Moreover, it participates in cellular signalling in response to various growth factors and in the promotion of developmental processes in diverse tissues (Yamaguchi et al., 2011). Among loci functionally related to *PIK3CA*, *ERBB4* is known to contribute to regulation of the development of heart, central nervous system and mammary gland, as well as to specific gene transcription, cell proliferation, differentiation, migration and apoptosis (Sweeney et al., 2000; Muraoka-Cook et al., 2006; Iwamoto &

Mekada 2006). *BLK* instead encodes for a non-receptor tyrosine kinase involved in B-lymphocyte development, differentiation and signalling (Borowiec et al., 2009). Finally, *EPHB1* is responsible of cell migration and adhesion, especially during the nervous system development, when it regulates retinal axon guidance.

Selection signatures at genomic intervals associated to completely different biological functions with respect to those described above were instead observed for the Han-Tujia, Japanese and Korean clusters. The highest number of genes belonging to the top 1% genomic windows putatively subjected to selection and functionally related to each other (i.e. involved primarily in fat tissue metabolism) was observed in people of Han or Tujia ancestry (Figure S12). For instance, *AKR1C2* is a member of the aldo/keto reductase superfamily that catalyzes the conversion of aldehydes and ketones into their corresponding alcohols. It is mainly expressed in fat tissue (Fagerberg et al., 2014) and bind bile acid with high affinity to convert steroid hormones into the 3-alpha/5-alpha and 3-alpha/5-beta-tetrahydrosteroids (Hara et al., 1996; Couture et al., 2005; Faucher et al., 2007). The *ALDH9A1* gene instead encodes for an enzyme that belongs to the aldehyde dehydrogenase protein family and has a high activity for oxidation of gamma-aminobutyraldehyde and other amino aldehydes (Vaz et al., 2000). Likewise *AKR1C2*, it is expressed mainly in fat tissue (Fagerberg et al., 2014). The *ACSF2* locus codes for a protein that catalyzes the initial reaction in fatty acid metabolism and plays a role in adipocyte differentiation (Watkins et al., 2007). *CYP11B1* is a member of the cytochrome P450 superfamily of enzymes, monooxygenases that catalyze many reactions involved in drug metabolism and synthesis of cholesterol, steroids, retinoids and xenobiotics (Shimada et al., 1999). *ACSL3* encodes for an isozyme of the long-chain fatty-acid-coenzyme A ligase family that converts free long-chain fatty acids into fatty acyl-CoA esters and thereby play a key role in lipid biosynthesis and fatty acid degradation (Yao & Ye 2008). *ABHD5* codes for a lysophosphatidic acid acyltransferase that contributes to phosphatidic acid biosynthesis. Mutations at this gene have been associated with Chanarin-Dorfman syndrome, a triglyceride storage disease with impaired long-chain fatty acid oxidation (Ghosh et al., 2008). Finally, *GPX3* and *MGST3* code for proteins that protect cells and enzymes from oxidative damage by catalyzing the reduction of hydrogen peroxide, lipid peroxides and organic hydro peroxide.

Putative adaptive genes that showed functional relationships with each other were identified also for the Japanese cluster and involved again *ALDH9A1*, *GPX3* and *MGST3* (Figure S13). Moreover, a beta-adducin (*ADD2*), a TNFAIP3-interacting protein 3 (*TNIP3*), whose overexpression inhibits NF-kappa-B dependent gene expression in response to

lipopolysaccharide (Wullaert et al., 2007), as well as a chromodomain Y-like protein (*CDYL*), a glyoxylate reductase 1 homolog (*GLYR1*) and *ADH7* (alcohol dehydrogenase 7) were found to participate to the same biological processes. Even if it is a member of the alcohol dehydrogenase family, the enzyme encoded by *ADH7* is most active as a retinol dehydrogenase. Its expression is greatly concentrated in the esophagus (Fagerberg et al., 2014), while being almost absent in other tissues, and its variants have been linked to alcohol dependence (Park et al., 2013) and oral cavity or pharyngeal cancer (McKay et al., 2011). Some of these genes were pointed out as candidate targets of positive selection also in the Korean cluster (Figure S14), as is the case of *ALDH9A1*, *ACSF2* and *ADD2*, which were previously observed in the Han-Tujia and Japanese clusters, respectively. However, in people of Korean ancestry these loci were found to be functionally related also to two additional candidate adaptive genes (i.e. *ALDH1A2* and *AOX1*). The *ALDH1A2* locus encodes for an enzyme that catalyzes the synthesis of retinoic acid from retinaldehyde. Similarly to *ADH7*, it has been associated to Barrett's esophagus, a premalignant precursor of esophageal adenocarcinoma (Levine et al., 2013; Gharahkhani et al., 2016). The *AOX1* gene codes instead for an oxidase with broad substrate specificity (Beedham et al., 1995). Furthermore, the Korean cluster presented a second large signature ascribable to genes relevant in the lipid metabolism (Figure S14). It included acetyl and acyl transferases (*NAA20*, *AGPAT1* and *MOGAT1*) and *LPIN1*, which play important roles in controlling fatty acids metabolism at different levels.

Fine mapping of the most informative candidate adaptive genes

After having identified large genomic regions plausibly subjected to positive selection, along with the associated biological functions, we prioritized the genes contained in them (see Methods) and we focused on the loci showing the strongest selective signatures in each cluster.

Again, we observed appreciably different adaptive profiles among the considered clusters, especially between South Asian/South East Asian groups and East Asian people from more northerly latitudes. In fact, only four genes (i.e. *DENNDIA*, *MCC*, *SCRNI* and *WIPF3*) were found to be included in the top 1% windows of all clusters (Table S2). Among them, *DENNDIA* showed the strongest and widest signature of selection as its variants were generally included in chromosomal intervals characterized by the highest-ranking positions in the majority of clusters and were distributed through different windows. Several genome-wide association studies suggested that variation at this gene likely influences polycystic

ovary syndrome risk and succeeded in identifying risk SNPs that showed remarkable association with the disease in most populations of Asian ancestry (Chen et al., 2011; Shi et al., 2012; Ha et al., 2015). In addition, the Pakistani and the North West Indian clusters shared five candidate adaptive genes (i.e. *FBXW2*, *LOC253039*, *PHF19*, *TRAF1-C5* and *NRG3*), with a limited overlap also with signatures observed in the Bangladeshi group (i.e. *FBXW2* and *PHF19*).

Conversely, the remaining genes included in the top 1% candidate windows were found to be differentiated among most of the examined clusters (Table S2). Interestingly, similar results were pointed out for the Han-Tujia and Japanese groups, which distinguished them also from most of the other East Asian clusters. In fact, they shared significant nSL scores at SNPs located in a wide genomic region including the *ERGIC3*, *FER1L4* and *CPNE1* genes.

Among them, *ERGIC3* represented the most interesting locus because genome-wide association studies have supported its implication in the modulation of total cholesterol levels (Teslovich et al., 2010; Willer et al., 2013). In particular, the derived allele (T) of rs2277862 was reported to significantly correlate with lower level of such a lipid trait and to present *cis*-acting associations with *ERGIC3* transcript levels in liver, omental and subcutaneous fat (Teslovich et al., 2010). We found this allele to be encompassed in the genomic window showing of one of the strongest signatures of positive selection in both the Han-Tujia and Japanese clusters, being also included in the top 1% windows identified in the Korean group. This region also harbors other six and seven SNPs presenting unusual nSL values respectively in the Han-Tujia and Japanese populations. Moreover, with the sole exception of rs6088887, which was found to be an outlier SNP only in people of Han or Tujia ancestry, exactly the same variants were pointed out as potential targets of selection in both the Han-Tujia and Japanese clusters.

Finally, *FOXP1* was included in the top 1% windows of both Han-Tujia, Japanese and Korean groups, but emerged among the top 10 candidate genes only in the Japanese population.

This gene belongs to the FOX family of transcription factors and is involved in a broad range of biological functions. It is indeed widely expressed and play a role in embryological, immunological and haematological processes, as well as in the development of speech and language (Shi et al., 2008; Hannenhalli & Kaestner, 2009; Benayoun et al., 2011).

Interestingly, in the context of a complex network of transcriptional factors and cofactors, it is also known to modulate hepatic glucose production in response to hormonal and nutrient stimuli (Nakae et al., 2002). In particular, we found several *FOXP1* variants contributing to

one of the strongest signatures of selection observed in the Han-Tujia, Japanese and Korean clusters.

Investigation of polygenic adaptive events in the Han-Tujia and Japanese clusters

Both windows-based and single gene level nSL analyses pointed to the Han-Tujia, Korean and Japanese clusters as the groups having most plausibly evolved adaptations triggered by dietary-related selective pressures. To validate these findings on independent cohorts of samples and to deepen the investigation of their adaptive history, we used whole genome sequence data generated by the 1000 Genomes Project (1000 Genomes Project Consortium et al., 2015) for the CHB and JPT populations and a gene-network approach to test for the occurrence of adaptive events also according to a polygenic adaptation model (see Methods). Gene networks belonging to the *Glycerolipid metabolism*, *AGE-RAGE signaling in diabetic complications* and *FoxO signaling* pathways were found to have undergone extensive positive selection in CHB (Figure 3, Table S3). Signatures at genes of the *Glycerolipid metabolism* pathway and involved in fatty acids metabolism, such as *LPIN1* and *MOGAT1*, were found to overlap with those detected by the windows-based selection scan performed on the Korean cluster. Most of the remaining loci instead belonged to the other two pathways and have been implicated in the modulation of pathological traits associated to insulin resistance and T2D, as is the case of several genes encoding for mitogen-activated proteins kinases (*MAPK*) and phospholipases (*PLCB*), or in the response to dietary stimuli, such as the insulin like growth factor 1 receptor (*IGF1R*) and *FOXO* genes (Figure 3, Table S3). Gene networks belonging to the *FoxO signaling*, *Glycolysis/Gluconeogenesis* and *Glucagon signaling* pathways were found to have been subjected to positive selection in JPT (Figure 3, Table S4). In details, multiple *MAPK* and *FOXO* genes were the same pointed out as candidate loci in CHB, while *ALDH9A1* and *ADH7* were suggested to have undergone selective sweeps by single gene level analyses performed on the Han-Tujia, Japanese and Korean clusters. On the contrary, genes belonging to the *Glucagon signaling* pathway, such as several cAMP responsive element binding proteins (*CREB*), *PPARGC1A* and *PRKACG*, were found to have played a putative adaptive role only in JPT (Figure 3, Table S4).

DISCUSSION

Introduction of massive cereals consumption, in particular of rice, in the diet of the ancestors of Asian populations has represented a substantial challenge for their metabolism due to the

outstanding carbohydrates content and glycemic index of such a nutritional resource with respect to other domesticated plants (Atkinson, Foster-Powell, & Brand-Miller, 2008). Daily and abundant rice intake is indeed considered a risk factor for the development of insulin resistance, thus potentially increasing susceptibility to T2D and obesity in populations that use it as a staple food (Hu et al., 2012; Boers, Seijen Ten Hoorn, & Mela, 2015).

Nevertheless, epidemiological patterns related to these metabolic diseases present appreciable variation across human groups of Asian ancestry. In fact, with respect to people from the Indian subcontinent, some East Asian populations show a lower increase in T2D prevalence subsequent to the shifts in dietary habits and lifestyles due to the social and economic transitions occurred in most Asian regions in the last decades (Yoon et al., 2006; Chan et al., 2009; Ramachandran et al., 2010). Accordingly, we hypothesized that genetic adaptations against the detrimental side effects of cereal-based diets evolved by populations that relied on these nutritional resources for a period long enough to enable natural selection to shape variation at disease-associated loci might contribute to such differential metabolic risk. In particular, human groups from Eastern China, whose ancestors have routinely consumed wild rice long before (i.e. since around 10,000 years ago) its domestication and the development of cultivation techniques (Zhao, 1998; Jiang & Liu, 2006; Cao et al., 2006), represent the most reliable candidate populations to be tested for the occurrence of these adaptive events.

Secondly, people from Korea and Japan, the regions to which millet and rice agriculture early diffused from China (Gross & Zhao, 2014), may have evolved comparable adaptive traits, contrary to populations of South East Asia or the Indian subcontinent who instead adopted usual cereals consumption only several thousand years later (Molina et al., 2011; Gross & Zhao, 2014; Choi et al., 2017).

To test this hypothesis, we investigated the adaptive evolution of populations whose ancestors inhabited the geographical areas where wild rice and/or millet originated and were early domesticated (i.e. Han and Tujia Chinese groups, Koreans and Japanese people) and compared it with that of control groups from South East Asian and South Asian regions where cereals agriculture spread only more recently. For this purpose, we took advantage from newly generated and literature genomic data for 2,379 individuals belonging to 124 Asian populations.

Population clusters within the South Asian and East Asian genomic landscapes

To frame patterns of genetic diversity observable at both candidate and control populations within the overall South Asian and East Asian genomic landscape and to test for

representativeness of the selected samples with regard to their ethnic group, we applied PCA and ADMIXTURE algorithms to the assembled “Pan-Asian” dataset. These population structure analyses confirmed the well-known genetic differentiation between groups from the Indian subcontinent and from East Asian or South East Asian regions, along with gradients of variation observable within these macro-geographical areas (Figure 1a).

In detail, genetic affinity among populations of South Asian ancestry was found to follow a cline that roughly reflects both their latitudinal dislocation across the Indian subcontinent and the distribution of different language groups, as previously attested by several studies (Reich et al., 2009; Majumder, 2010; Metspalu et al., 2011; Chaubey et al., 2011; Basu et al., 2016).

People speaking Indo-European languages and showing high proportions of the “Ancestral North Indian” ancestry component and groups speaking Dravidian languages, who are instead characterized by the “Ancestral South Indian” component, were located at the opposite ends of the South Asian gradient of diversity (Figure 1a). On the contrary, tribes speaking Austro-Asiatic languages and presenting the “Ancestral Austro-Asiatic” ancestry component (Figure 1b) were more scattered in the PCA space, with especially those from Eastern India

clustering close to Dravidian populations and being slightly shifted towards the East Asian cline (Figure 1a). As expected, both PCA and ADMIXTURE analyses pointed to populations characterized by the “Ancestral Tibeto-Burman” component to be more genetically similar to East Asians than to South Asians, despite being located mainly south of the Himalayas (Figure 1). This confirms the findings from studies that indicated groups speaking Tibeto-Burman languages as clear examples of gene flow from the Far East to the Indian subcontinent (Chaubey et al., 2011; Mörseburg et al., 2016; Gneccchi-Rusccone et al., 2017).

A complex gradient of genetic diversity was observed when considering people of East Asian and South East Asian ancestries, plausibly reflecting the tangled patterns of admixture events, isolation processes and adaptive dynamics experienced by these ethnic groups. Tibetans and Sherpa stood out at the top of this cline of variation and resulted relatively differentiated from the bulk of East Asian peoples (Figure 1a). This has been proved to be ascribable to their prolonged isolation on high-altitude Himalayan regions, which was made possible by the evolution of genetic adaptations to hypobaric hypoxia (Beall et al., 2010; Simonson et al., 2010; Yi et al., 2010; Gneccchi-Rusccone et al., 2018). Long-term isolation is also supposed to have driven the “Ancestral Tibeto-Burman” ancestry component near to fixation especially in Sherpa groups (Figure 1b) due to strong genetic drift (Gneccchi-Rusccone et al., 2017).

Conversely, populations from the Malay Peninsula, Borneo and Philippines were located at the opposite end of the East Asian/South East Asian cline of variation (Figure 1a). According

to ADMIXTURE analyses, they showed remarkable proportions of an ancestry fraction that previous studies proposed to have been spread by demographic processes related to the expansion of Austronesian languages (Figure 1b) (Mörseburg et al., 2016).

Given that ADMIXTURE runs testing eight potential ancestral groups turned out to be the model that best fitted with the data (Figure S4), profiles of ancestry components depicted at $K = 8$ were used to guide the identification of distinguishable population clusters. Accordingly, 10 clusters characterized by high internal genetic homogeneity and grouping populations with close geographical proximity, similar cultural patterns and shared demographic histories were described (Figure 2) and used to maximize the number of samples to be submitted to selection scans. Populations of Han-Tujia, Korean and Japanese ancestries, who archaeological evidence suggest to be the sole groups having relied on rice/millet consumption long before their cultivation or in conjunction with their very ancient domestication (Zhao, 1998; Fuller et al., 2010; Gross & Zhao, 2014), represented the best candidate clusters for testing the evolution of adaptive events triggered by cereal-based diets. Conversely, the remaining clusters were considered as control groups who have adopted such a typology of diet too recently to genetically adapt to it. For instance, the earliest evidence of usual rice consumption outside Eastern China, Korea and Japan was associated to villages with domestic rice livestock located in the Indo-Gangetic Plain, which have been proved to become widely established not earlier than 4,000 years ago (Fuller, 2006; Fuller et al., 2010).

Distinctive patterns of adaptive evolution in candidate and control population clusters

Large chromosomal intervals enriched for footprints left by the action of positive selection were identified by means of a sliding windows approach based on the computation of genome-wide distributions of nSL scores for each population clusters (Figure S5-S14, Table S2). Interestingly, adaptive profiles depicted for the candidate and control population groups were found to be highly differentiated in terms of both broad biological functions and single genes having experienced selective sweeps. In particular, most of the adaptive events inferred for the majority of South Asian and South East Asian control clusters occurred at loci involved in ubiquitination or in the regulation of cell proliferation/differentiation, for instance in the neural tissue or in the immune system (Figure S5-S10). These signatures seem to concern essential developmental processes but are unlikely ascribable to adaptive evolution in response to dietary-related selective pressures, certainly not to the adoption of cereals consumption. Therefore, the discussion of these findings was beyond the scope of the present study. Even when the most remarkable footprints of positive selection at single genes were

shortlisted for each cluster, limited overlap between adaptive events inferred in candidate and control groups was observed (Table S2). Moreover, the few loci presenting evidence of selection in both South Asian and East/South East Asian clusters again seemed to have been not evolved in response to metabolic stresses. For instance, the gene showing the strongest and widest selection signature (i.e. *DENND1A*) plays a role in the development of polycystic ovaries, anovulation and hyperandrogenism that mainly characterize the polycystic ovary endocrine syndrome (Chen et al., 2011; Shi et al., 2012; Ha et al., 2015).

Conversely, despite presenting distinct genetic backgrounds according to ADMIXTURE analyses (Figure 1b) and being known to have experienced comparable but not identical demographic histories (Wang et al., 2018), people of Han and Tujia ancestries, along with populations from Korea and Japan, showed adaptive profiles very similar with each other and remarkably different with respect to those of control clusters. It is worth noting that these candidate groups overall showed evidence of adaptive evolution at genes implicated in biochemical processes occurring in the fat tissue, in particular at the level of fatty acids metabolism, cholesterol and other lipids biosynthesis, and adipocyte differentiation (Watkins et al., 2007; Yao & Ye, 2008; Ghosh et al., 2008; Fagerberg et al., 2014). For instance, the *LPIN1* gene was suggested to have experienced a selective sweep in Koreans according to windows-based nSL analyses (Figure S14) and to have contributed to polygenic adaptation of Han people according to the gene-network approach applied to CHB whole genome sequence data (Figure 3, Table S3). This locus acts as a key regulator of fatty acids metabolism in several model organisms (Harris & Finck, 2011) and variants reducing its expression in the human adipose tissue have been associated with insulin resistance and impaired glucose homeostasis (Suviolahti et al., 2006; Yao-Borengasser et al., 2006; Aulchenko et al., 2007; Loos et al., 2007; Bego et al., 2011; Zhang et al., 2013). In addition, the *ERGIC3* gene, which is involved in the modulation of cholesterol biosynthesis, has undergone selective sweeps in all the candidate clusters, showing a putative adaptive haplotype highly conserved especially between people of Han/Tujia and Japanese ancestries (Table S2). This haplotype carries derived alleles known to be associated to low cholesterol levels and reduced risk of cardiovascular and metabolic traits, including coronary artery disease, waist-hip ratio and BMI (Teslovich et al., 2010; Willer et al., 2013). This might suggest genetic adaptation resulting in the reduction of both the conversion of dietary carbohydrates into cholesterol and fatty acids and of the combination of the latter with glycerol to form triglycerides to be stored in the adipose tissue. Similarly, convergent selective sweeps at the *FOXP1* gene were inferred for both the Han-Tujia, Korean and Japanese clusters (Table S2), in addition to

pervasive selection at multiple genes functionally related to this locus (e.g. *FOXO1*) as pointed out by gene-network analyses performed on CHB and JPT genome sequences (Figure 3, Tables S3-S4). *FOXO1* is involved also in the regulation of hepatic synthesis of glucose in response to hormonal and nutrient stimuli (Nakae et al., 2002) by promoting its production to prevent life-threatening hypoglycemia during prolonged starvation (Matsumoto et al., 2007). In particular, decreased blood insulin levels represent a signal for *FOXO1* to stimulate the expression of key gluconeogenic enzymes via direct binding to insulin response elements (IREs) located in their promoter regions, while *FOXPI* acts as a transcriptional repressor that inhibits them by directly interacting with the *FOXO1* gene or by competing with it to bind IREs of target genes (Accili & Arden, 2004; Zou et al., 2015). Interestingly, abnormal elevation of hepatic glucose production is known to contribute to fasting hyperglycemia in T2D (Saltiel & Kahn, 2001), and in a diabetic murine model increased gluconeogenesis promoted by *FOXO1* seems to be ascribable to decreased *FOXPI* expression in the liver. Accordingly, induced *FOXPI* overexpression has been proved to suppress hepatic gluconeogenesis in these diabetic mice, thus contributing to reduce glycaemia and to re-establish glucose homeostasis (Zou et al., 2015). Therefore, we can speculate that adaptive evolution at *FOXPI* and *FOXO1* genes might play a relevant role in the fine-tuning of blood glucose levels in populations having long relied on a staple food characterized by medium-to-high glycaemic index, such as rice or millet (Atkinson, Foster-Powell, & Brand-Miller, 2008). An analogous explanation may be invoked also for selection signatures at loci belonging to the *Glucagon signaling* pathway identified by gene-network analyses in people of Japanese ancestry (Figure 3, Table S4). Among them, the *CREB1*, *ATF2*, *PPARGC1A*, and *CPTIC* genes are indeed known to modulate insulin resistance in the adipose tissue of obese individuals (McCarty, 2005; Qi et al., 2009; Gao et al., 2009; Maekawa et al., 2010). Finally, putative adaptive events shared by all candidate clusters were observed at loci encoding for alcohol and aldehyde dehydrogenases, particularly at genes (e.g. *ADH7*, *ALDH1A2*, and *ALDH9A1*) implicated in the oxidation of retinol to retinaldehyde and in the subsequent synthesis of retinoic acid from it (Figure 3, Figure S12-S14 and Table S4). This pattern might reflect adaptations of Han, Tujia, Korean and Japanese people aimed at optimizing the production of retinoic acid, which is the active metabolite of vitamin A and a powerful regulator of gene transcription that plays a crucial role in cellular proliferation and differentiation (Ross & Zolfaghari, 2011). In fact, vitamin A deficiency represents a major health issue for many developing societies (West & Darnton-Hill, 2001), being especially

associated in most Asian populations with the poverty-related predominant consumption of rice, which lacks pro-vitamin A in the edible part of the grain (Paine et al., 2005).

In conclusion, by using genome-wide SNP and whole genome sequence data to reconstruct the adaptive history of a large “Pan-Asian” set of populations, we provided evidence for selective sweeps and polygenic mechanisms having possibly contributed to the biological adaptation to rice- and/or millet-based diets of people of Han/Tujia, Korean and Japanese ancestries. Actually, inferring the specific environmental or cultural stresses having triggered a given biological adaptation is not straightforward and we cannot rule out the possibility that other selective pressures in addition to those represented by cereal-based diets have played a role in shaping the detected signatures of adaptive evolution. For instance, climate conditions differ between the geographical areas occupied by candidate and control populations, with temperate/continental climate being typical of northern China, northern Japan and Korea in contrast to the subtropical or tropical climates of most Indian regions. Accordingly, some metabolic adaptations observed in the East Asian candidate populations might have been evolved also in response to a more energy demanding environment than that in which South Asian people live, as previously suggested for other adaptive loci (Sazzini et al., 2014).

Indeed, these traits might be implicated in optimizing the use of high-calories nutrients such as carbohydrates as fuel for thermogenic processes rather than as sources for triglycerides storing in the adipose tissue, thus resulting in better glucose homeostasis and reduced adiposity. Nevertheless, the described findings are reasonably consistent also with archaeobotanical data indicating that the ancestors of Han/Tujia, Korean and Japanese populations have relied on rice or millet as a staple food since several thousand years before than other East Asian or South Asian populations did. In particular, our results suggested that the observed adaptations have been mediated by changes in fatty acids metabolism, in the biosynthesis of cholesterol and triglycerides from carbohydrates, as well as in the production of the vitamin A metabolite retinoic acid. Nowadays, these adaptive traits might play a role in mitigating some of the metabolic side effects of a usual massive intake of polished white rice, especially those related to the increased risk of developing overweight or obesity and impaired glucose homeostasis early in life, and in compensating the scarcity of vitamin A due to such a peculiar nutritional regimen. Therefore, the present study succeeded in pinpointing some of the deep causes that are contributing to influence present-day epidemiological patterns observable across the Asian continent for T2D and obesity, thus taking a step forward in demonstrating the usefulness of an evolutionary approach to the study of the differential susceptibility of human populations to modern complex diseases.

ACKNOWLEDGMENTS

We would like to thank all donors who participated to the study without whom this work would have not been possible. We are also grateful to Pier Massimo Zambonelli (CESIA, University of Bologna) for his IT assistance. S.S. and S.D.F. were supported by the ERC AdG n. 295733 to D.P.

DATA AVAILABILITY STATEMENT

The genotype data generated during the current study are available at the corresponding author page of the figshare repository (https://figshare.com/authors/Marco_Sazzini/6292466).

CONFLICT OF INTEREST

None declared.

REFERENCES

- 1000 Genomes Project Consortium, Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., ... Abecasis, G. R. (2015). A global reference for human genetic variation. *Nature*, 526, 68-74.
- Accili, D., & Arden, K. C. (2004). FoxOs at the crossroads of cellular metabolism, differentiation, and transformation. *Cell*, 117, 421-426.
- Alexander, D. H, Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19, 1655-1664.
- Atkinson, F. S., Foster-Powell, K., & Brand-Miller, J. C. (2008). International Tables of Glycemic Index and Glycemic Load Values: 2008. *Diabetes care*, 31, 2281-2283.
- Aulchenko, Y. S., Pullen, J., Kloosterman, W. P., Yazdanpanah, M., Hofman, A., Vaessen, N., ... van Duijn, C. M. (2007). LPIN2 is associated with type 2 diabetes, glucose metabolism, and body composition. *Diabetes*, 56, 3020-3026.
- Basu, A., Sarkar-Roy, N., & Majumder, P. P. (2016). Genomic reconstruction of the history of extant populations of India reveals five distinct ancestral components and a complex structure. *Proceedings of the National Academy of Sciences of the United States of America*, 113, 1594-1599.
- Bayoumi, R., De Fanti, S., Sazzini, M., Giuliani, C., Quagliariello, A., Bortolini, E., ... Luiselli, D. (2016). Positive selection of lactase persistence among people of Southern

Arabia. *American Journal of Physical Anthropology*, 161, 676-684.

- Beall, C. M., Cavalleri, G. L., Deng, L., Elston, R. C., Gao, Y., Knight, J., ... Zheng, Y. T. (2010). Natural selection on EPAS1 (HIF2alpha) associated with low hemoglobin concentration in Tibetan highlanders. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 11459-11464.
- Beedham, C., Critchley, D. J., & Rance, D. J. (1995). Substrate specificity of human liver aldehyde oxidase toward substituted quinazolines and phthalazines: a comparison with hepatic enzyme from guinea pig, rabbit, and baboon. *Archives of Biochemistry and Biophysics*, 319, 481-490.
- Benayoun, B. A., Caburet, S., & Veitia, R. A. (2011). Forkhead transcription factors: Key players in health and disease. *Trends in Genetics*, 27, 224-232.
- Bego, T., Dujic, T., Mlinar, B., Semiz, S., Malenica, M., Prnjavorac, B., ... Causević, A. (2011). Association of PPARG and LPIN1 gene polymorphisms with metabolic syndrome and type 2 diabetes. *Medicinski Glasnik*, 8, 76-83.
- Bharati, P., & Kulkarni, U.N. (2020). Nutritional sufficiency of traditional meal patterns. In: *Nutritional and Health Aspects of food in South Asian Countries*, ed. Prakash, J., Waisundara, V., & Prakash, V. Amsterdam: Elsevier.
- Boers, H. M., Seijen Ten Hoorn, J., & Mela, D. J. (2015). A systematic review of the influence of rice characteristics and processing methods on postprandial glycaemic and insulinaemic responses. *British Journal of Nutrition*, 114, 1035-1045.
- Borowiec, M., Liew, C. W., Thompson, R., Boonyasrisawat, W., Hu, J., Mlynarski, W. M., ... Doria, A. (2009). Mutations at the BLK locus linked to maturity onset diabetes of the young and beta-cell dysfunction. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 14460-14465.
- Brown, E. A. (2012). Genetic explorations of recent human metabolic adaptations: hypotheses and evidence. *Biological reviews of the Cambridge Philosophical Society*, 87(4), 838-855.
- Cao, Z. H., Ding, J. L., Hu, Z. Y., Knicker, H., Kögel-Knabner, I., Yang, L. Z., ... Dong, Y. H. (2006). Ancient paddy soils from the Neolithic age in China's Yangtze River Delta. *Naturwissenschaften*, 93, 232-236.
- Chan, J. C., Malik, V., Jia, W., Kadowaki, T., Yajnik, C. S., Yoon, K. H., Hu, F. B. (2009). Diabetes in Asia: epidemiology, risk factors, and pathophysiology. *Journal of the American Medical Association*, 301, 2129-2140.
- Chaubey, G., Metspalu, M., Choi, Y., Mägi, R., Romero, I. G., Soares, P., ... Kivisild, T.

- (2011). Population genetic structure in Indian Austroasiatic speakers: the role of landscape barriers and sex-specific admixture. *Molecular Biology and Evolution*, 28, 1013-1024.
- Chen, Z. J., Zhao, H., He, L., Shi, Y., Qin, Y., Shi, Y., ... Zhao, Y. (2011). Genome-wide association study identifies susceptibility loci for polycystic ovary syndrome on chromosome 2p16.3, 2p21 and 9q33.3. *Nature Genetics*, 43, 55-59.
- Choi, J. Y., Platts, A. E., Fuller, D. Q., Hsing Y. I., Wing, R. A., & Purugganan, M. D. (2017). The Rice Paradox: Multiple Origins but Single Domestication in Asian Rice. *Molecular Biology and Evolution*, 34, 969-979.
- Couture, J. F., de Jésus-Tran, K. P., Roy, A. M., Cantin, L., Côté, P. L., Legrand, P., ... Breton, R. (2005). Comparison of crystal structures of human type 3 3alpha-hydroxysteroid dehydrogenase reveals an "induced-fit" mechanism and a conserved basic motif involved in the binding of androgen. *Protein Science*, 14, 1485-1497.
- Dasgupta, R., Pillai, R., Kumar, R., & Arora, N. K. Sugar, salt, fat, and chronic disease epidemic in India: is there need for policy interventions? *Indian Journal of Community Medicine*, 2015, 40:71-74.
- Delaneau, O., Coulonges, C., & Zagury, J. F. (2008). Shape-IT: new rapid and accurate algorithm for haplotype inference. *BMC Bioinformatics*, 9, 540.
- Deng, Z., Hung, H. C., Fan, X., Huang, Y., & Lu, H. (2018). The ancient dispersal of millets in southern China: New archaeological evidence. *The Holocene*, 28, 34-43.
- Dixit, A. A., Azar, K. M., Gardner, C. D., Palaniappan, L. P. (2011). Incorporation of whole, ancient grains into a modern Asian Indian diet to reduce the burden of chronic disease. *Nutrition Reviews*, 69, 479-488.
- Fagerberg, L., Hallström, B. M., Oksvold, P., Kampf, C., Djureinovic, D., Odeberg, J., ... Uhlén, M. (2014). Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Molecular & Cellular Proteomics*, 13, 397-406.
- Food and Agriculture Organization of the United Nations. (2008). Food Consumption Nutrients spreadsheet. http://www.fao.org/fileadmin/templates/ess/documents/food_security_statistics/FoodConsumptionNutrients_en.xls. Accessed on 30/04/2020.
- Faucher, F., Cantin, L., Pereira de Jésus-Tran, K., Lemieux, M., Luu-The, V., Labrie, F., & Breton, R. (2007). Mouse 17alpha-hydroxysteroid dehydrogenase (AKR1C21) binds steroids differently from other aldo-keto reductases: identification and

characterization of amino acid residues critical for substrate binding. *Journal of Molecular Biology*, 369, 525-540.

Ferrer-Admetlla, A., Liang, M., Korneliusson, T., & Nielsen, R. (2014). On detecting incomplete soft or hard selective sweeps using haplotype structure. *Molecular Biology and Evolution*, 31, 1275-1291.

Fitzgerald, M. A., Rahman, S., Resurreccion, A. P., Concepcion, J., Daygon, V. D., Dipti, S. S., ... Bird, A.R. (2011). Identification of a major genetic determinant of glycaemic index in rice. *Rice*, 4, 66-74.

Fuller, D. Q., Harvey, E., & Qin, L. (2007). Presumed domestication? Evidence for wild rice cultivation and domestication in the 5th millennium BC of the lower Yangtze region. *Antiquity*, 81, 316-331.

Fuller, D. Q., Sato, Y. I., Castillo, C., Qin, L., Weisskopf, A. R., Kingwell-Banham, E. J., ... van Etten, J. (2010). Consilience of genetics and archaeobotany in the entangled history of rice. *Archaeological and Anthropological Sciences*, 2, 115–131.

Fuller, D. Q., & Castillo, C. (2016). Diversification and Cultural Construction of a Crop: The Case of Glutinous Rice and Waxy Cereals in the Food Cultures of Eastern Asia. In: *The Oxford Handbook of the Archaeology of Diet*, ed. J. Lee-Thorp, and M. Anne Katzenberg (online DOI: 10.1093/oxfordhb/9780199694013.013.8). Oxford: Oxford University Press.

Fujiwara, A., Murakami, K., Asakura, K., Uechi, K., Sugimoto, M., Wang, H.C., ... Sasaki, S. (2018). Estimation of Starch and Sugar Intake in a Japanese Population Based on a Newly Developed Food Composition Database. *Nutrients*, 10, 1474.

Gabriel, A. S., Ninomiya, K., & Uneyama, H. (2018). The role of the Japanese traditional diet in healthy and sustainable dietary patterns around the world. *Nutrients*, 10, E173.

Gao, X. F., Chen, W., Kong, X. P., Xu, A. M., Wang, Z. G., Sweeney, G., & Wu, D. (2009). Enhanced susceptibility of Cpt1c knockout mice to glucose intolerance induced by a high-fat diet involves elevated hepatic gluconeogenesis and decreased skeletal muscle glucose uptake. *Diabetologia*, 52, 912-920.

Gharahkhani, P., Fitzgerald, R. C., Vaughan, T. L., Palles, C., Gockel, I., Tomlinson, I., ... Schumacher, J. (2016). Genome-wide association studies in oesophageal adenocarcinoma and Barrett's oesophagus: a large-scale meta-analysis. *The Lancet Oncology*, 17, 1363-1373.

Ghosh, A. K., Ramakrishnan, G., Chandramohan, C., & Rajasekharan, R. (2008). CGI-58, the causative gene for Chanarin-Dorfman syndrome, mediates acylation of

lysophosphatidic acid. *Journal of Biological Chemistry*, 283, 24525-24533.

- Gnecchi-Ruscione, G. A., Jeong, C., De Fanti, S., Sarno, S., Trancucci, M., Gentilini, D., ... Sazzini, M. (2017). The genomic landscape of Nepalese Tibeto-Burmans reveals new insights into the recent peopling of Southern Himalayas. *Scientific Reports*, 7, 15512.
- Gnecchi-Ruscione, G. A., Abondio, P., De Fanti, S., Sarno, S., Sherpa, M. G., Sherpa, P. T., ... Sazzini, M. (2018). Evidence of Polygenic Adaptation to High Altitude from Tibetan and Sherpa Genomes. *Genome Biology and Evolution*, 10, 2919-2930.
- Gouy, A., Daub, J. T., & Excoffier, L. (2017). Detecting gene subnetworks under selection in biological pathways. *Nucleic Acids Research*, 45, e149.
- Gross, B. L., & Zhao, Z. (2014). Archaeological and genetic insights into the origins of domesticated rice. *Proceedings of the National Academy of Sciences of the United States of America*, 111, 6190-6197.
- Ha, L., Shi, Y., Zhao, J., Li, T., & Chen, Z. J. (2015). Association Study between Polycystic Ovarian Syndrome and the Susceptibility Genes Polymorphisms in Hui Chinese Women. *PLoS One*, 10, e0126505.
- Hannenhalli, S., & Kaestner, K. H. (2009). The evolution of Fox genes and their role in development and disease. *Nature Reviews Genetics*, 10, 233-240.
- Hara, A., Matsuura, K., Tamada, Y., Sato, K., Miyabe, Y., Deyashiki, Y., & Ishida, N. (1996). Relationship of human liver dihydrodiol dehydrogenases to hepatic bile-acid-binding protein and an oxidoreductase of human colon cells. *Biochemical Journal*, 313, 373-376.
- Harris, T. E., & Finck, B. N. (2011). Dual function lipin proteins and glycerolipid metabolism. *Trends in Endocrinology & Metabolism*, 22, 226-233.
- Hu, F. B. (2011). Globalization of diabetes: the role of diet, lifestyle, and genes. *Diabetes Care*, 34, 1249-1257.
- Hu, E.A., Pan, A., Malik, V., & Sun, Q. (2012). White rice consumption and risk of type 2 diabetes: meta-analysis and systematic review. *BMJ*, 344, e1454.
- HUGO Pan-Asian SNP Consortium, Abdulla, M. A., Ahmed, I., Assawamakin, A., Bhak, J., Brahmachari, S. K., ... Indian Genome Variation Consortium. (2009). Mapping human genetic diversity in Asia. *Science*, 326, 1541-1545.
- Iwamoto, R., & Mekada, E. (2006). ErbB and HB-EGF signaling in heart development and function. *Cell Structure and Function*, 31, 1-14.
- Jensen, L. J., Kuhn, M., Stark, M., Chaffron, S., Creevey, C., Muller, J., ... von Mering, C. (2009). STRING 8--a global view on proteins and their functional interactions in 630

- organisms. *Nucleic Acids Research*, 37, D412-416.
- Jiang, L., & Liu, L. (2006). New evidence for the origins of sedentism and rice domestication in the lower Yangzi river, China. *Antiquity*, 80, 355–361.
- Kaur, B., Ranawana, V., & Henry, J. (2016). The Glycemic Index of Rice and Rice Products: A Review, and Table of GI Values. *Critical Reviews in Food Science and Nutrition*, 56, 215-236.
- Kelly, M. (2016). The Nutrition Transition in Developing Asia: Dietary Change, Drivers and Health Impacts. In: *Eating, Drinking: Surviving. Briefs in Global Understanding*, ed. Jackson, P., Spiess, W., & Sultana, F. Cham: Springer.
- Levine, D. M., Ek, W. E., Zhang, R., Liu, X., Onstad, L., Sather, C., ... Vaughan, T. L. (2013). A genome-wide association study identifies new susceptibility loci for esophageal adenocarcinoma and Barrett's esophagus. *Nature Genetics*, 45, 1487-1493.
- Li, J. Z., Absher, D. M., Tang, H., Southwick, A. M., Casto, A. M., Ramachandran, S., ... Myers, R. M. (2008). Worldwide human relationships inferred from genome-wide patterns of variation. *Science*, 319, 1100-1104.
- Liu, X., & Jones, M. K. (2014). Food globalisation in prehistory: top down or bottom up? *Antiquity*, 88, 956-963.
- Long, T., Leipe, C., Jin, G., Wagner, M., Guo, R., Schröder, O., & Tarasov, P. E. (2018). The early history of wheat in China from 14C dating and Bayesian chronological modelling. *Nature Plants*, 4, 272-279.
- Loos, R. J., Rankinen, T., Pérusse, L., Tremblay, A., Després, J. P., & Bouchard, C. (2007). Association of lipin 1 gene polymorphisms with measures of energy and glucose metabolism. *Obesity (Silver Spring)*, 15, 2723-2732.
- Luca, F., Perry, G. H., & Di Rienzo, A. (2010). Evolutionary adaptations to dietary changes. *Annual Review of Nutrition*, 30, 291-314.
- Lucock, M. (2007). *Molecular Nutrition and Genomics: Nutrition and the Ascent of Humankind*. Hoboken: John Wiley and Sons.
- Ma, R. C. W., & Chan, J. C. N. (2013). Type 2 diabetes in East Asians: similarities and differences with populations in Europe and the United States. *Annals of the New York Academy of Sciences*, 1281, 64-91.
- Maekawa, T., Jin, W., & Ishii, S. (2010). The role of ATF-2 family transcription factors in adipocyte differentiation: antiobesity effects of p38 inhibitors. *Molecular and Cellular Biology*, 30, 613-625.
- McCarty, M. F. (2005). Up-regulation of PPARgamma coactivator-1alpha as a strategy for

preventing and reversing insulin resistance and obesity. *Medical Hypotheses*, 64, 399-407.

McKay, J. D., Truong, T., Gaborieau, V., Chabrier, A., Chuang, S. C., Byrnes, G., ... Brennan, P. (2011). A genome-wide association study of upper aerodigestive tract cancers conducted within the INHANCE consortium. *PLoS Genetics*, 7, e1001333.

Majumder, P. P. (2010). The human genetic history of South Asia. *Current Biology*, 20, R184-187.

Mallick, S., Li, H., Lipson, M., Mathieson, I., Gymrek, M., Racimo, F., ... Reich, D. (2016). The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature*, 538, 201-206.

Matsumoto, M., Poci, A., Rossetti, L., Depinho, R. A., & Accili, D. (2007). Impaired regulation of hepatic glucose production in mice lacking the forkhead transcription factor Foxo1 in liver. *Cell Metabolism*, 6, 208-216.

Metspalu, M., Romero, I. G., Yunusbayev, B., Chaubey, G., Mallick, C. B., Hudjashov, G., ... Kivisild, T. (2011). Shared and unique components of human population structure and genome-wide signals of positive selection in South Asia. *American Journal of Human Genetics*, 89, 731-744.

Misra, A., Singhal, N., Sivakumar, B., Bhagat, N., Jaiswal, A., & Khurana, L. (2011). Nutrition transition in India: Secular trends in dietary intake and their relationship to diet-related non-communicable diseases. *Journal of Diabetes*, 3, 278-292.

Molina, J., Sikora, M., Garud, N., Flowers, J. M., Rubinstein, S., Reynolds, A., ... Purugganan, M. D. (2011). Molecular evidence for a single evolutionary origin of domesticated rice. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 8351-8356.

Mörseburg, A., Pagani, L., Ricaut, F. X., Yngvadottir, B., Harney, E., Castillo, C., ... Kivisild, T. (2016). Multi-layered population structure in Island Southeast Asians. *European Journal of Human Genetics*, 24, 1605-1611.

Mukhopadhyay, D., & Riezman, H. (2007). Proteasome-independent functions of ubiquitin in endocytosis and signaling. *Science*, 315, 201-205.

Muraoka-Cook, R. S., Sandahl, M., Husted, C., Hunter, D., Miraglia, L., Feng, S. M., ... Earp, H. S. 3rd. (2006). The intracellular domain of ErbB4 induces differentiation of mammary epithelial cells. *Molecular Biology of the Cell*, 17, 4118-4129.

Murakami, K., Sasaki, S., Takahashi, Y., Okubo, H., Hosoi, Y., Horiguchi, H., ... Kayama, F. (2006). Dietary glycemic index and load in relation to metabolic risk factors in

Japanese female farmers with traditional dietary habits. *American Journal of Clinical Nutrition*, 83, 1161-1169.

Nakae, J., Biggs, W. H. 3rd, Kitamura, T., Cavenee, W. K., Wright, C. V., Arden, K. C., & Accili, D. (2002). Regulation of insulin action and pancreatic beta-cell function by mutated alleles of the gene encoding forkhead transcription factor Foxo1. *Nature Genetics*, 32, 245-253.

Olsen, K. M., Caicedo, A. L., Polato, N., McClung, A., McCouch, S., & Purugganan, M. D. (2006). Selection under domestication: Evidence for a sweep in the rice waxy genomic region. *Genetics*, 173, 975-983.

Paine, J. A., Shipton, C. A., Chaggar, S., Howells, R. M., Kennedy, M. J., Vernon, G., ... Drake, R. (2005). Improving the nutritional value of Golden Rice through increased pro-vitamin A content. *Nature Biotechnology*, 23, 482-487.

Park, B. L., Kim, J. W., Cheong, H. S., Kim, L. H., Lee, B. C., Seo, C. H., ... Choi, I. G. (2013). Extended genetic effects of ADH cluster genes on the risk of alcohol dependence: from GWAS to replication. *Human Genetics*, 132, 657-668.

Patterson, N., Price, A. L., & Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genetics*, 2, e190.

Piras, I.S., De Montis, A., Calò, C.M., Marini, M., Atzori, M., Corrias, L., ... Contu, L. (2012). Genome-wide scan with nearly 700,000 SNPs in two Sardinian sub-populations suggests some regions as candidate targets for positive selection. *European Journal of Human Genetics*, 20, 1155-1161.

Popkin, B. M., Horton, S., Kim, S., Mahal, A., & Shuigao, J. (2001). Trends in diet, nutritional status, and diet-related noncommunicable diseases in China and India: the economic costs of the nutrition transition. *Nutrition Reviews*, 59, 379-390.

Praveen, W., Sayumi, J., Yashasvi, P., Ganga, J., & Saroj, J. (2014). Per capita sugar consumption and prevalence of diabetes mellitus - global and regional associations. *BMC Public Health*, 14, 186.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., ... Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, 81, 559-575.

Qi, L., Saberi, M., Zmuda, E., Wang, Y., Altarejos, J., Zhang, X., ... Montminy, M. (2009). Adipocyte CREB promotes insulin resistance in obesity. *Cell Metabolism*, 9, 277-286.

Radhika, G., Van Dam, R.M., Sudha, V., Ganesan, A., Mohan, V. (2009). Refined grain consumption and the metabolic syndrome in urban Asian Indians (Chennai Urban

- Rural Epidemiology Study 57). *Metabolism*, 58, 675-681.
- Ramachandran, A., Snehalatha, C., Baskar, A. D. S., Mary, S., Sathish Kumar, C. K., Selvam, S., ... Vijay, V. (2004). Temporal changes in prevalence of diabetes and impaired glucose tolerance associated with lifestyle transition occurring in the rural population in India. *Diabetologia*, 47, 860-865.
- Ramachandran A., & Snehalatha, C. (2010). Rising Burden of Obesity in Asia. *Journal of Obesity*, 2010, 868573.
- Ramachandran, A., Ma, R. C., & Snehalatha, C. (2010). Diabetes in Asia. *Lancet*, 375, 408-418.
- Ramachandran, A., Chamukuttan, S., Shetty, S. A., Arun, N., Susairaj, P. (2012). Obesity in Asia - is it different from rest of the world. *Diabetes/Metabolism Research and Reviews*, 28, 47-51.
- Reich, D., Thangaraj, K., Patterson, N., Price, A. L., & Singh, L. (2009). Reconstructing Indian population history. *Nature*, 461, 489-494.
- Ross, A. C., & Zolfaghari, R. (2011). Cytochrome P450s in the regulation of cellular retinoic acid metabolism. *Annual Review of Nutrition*, 31, 65-87.
- Saltiel A.R., & Kahn, C.R. (2001). Insulin signalling and the regulation of glucose and lipid metabolism. *Nature*, 414, 799–806.
- Sazzini, M., Schiavo, G., De Fanti, S., Martelli, P.L., Casadio, R., & Luiselli D. (2014). Searching for signatures of cold adaptations in modern and archaic humans: hints from the brown adipose tissue genes. *Heredity (Edinb)*, 113, 259-267.
- Sazzini, M., Gnecci-Ruscione, G. A., Giuliani, C., Sarno, S., Quagliariello, A., De Fanti, S., ... Luiselli, D. (2016). Complex interplay between neutral and adaptive evolution shaped differential genomic background and disease susceptibility along the Italian peninsula. *Scientific Reports*, 6, 32513.
- Sazzini, M., Abondio, P., Sarno, S., Gnecci-Ruscione, G. A., Ragno, M., Giuliani, C., ... Garagnani, P. (2020). Genomic history of the Italian population recapitulates key evolutionary dynamics of both Continental and Southern Europeans. *BMC Biology*, 18, 51.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., ... Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, 13, 2498-2504.
- Shi, C., Sakuma, M., Mooroka, T., Liscoe, A., Gao, H., Croce, K. J., ... Simon, D. I. (2008). Down- regulation of the forkhead transcription factor Foxp1 is required for monocyte

differentiation and macrophage function. *Blood*, 112, 4699-4711.

Shi, Y., Zhao, H., Shi, Y., Cao, Y., Yang, D., Li, Z., ... Chen, Z. J. (2012). Genome-wide association study identifies eight new risk loci for polycystic ovary syndrome. *Nature Genetics*, 44, 1020-1025.

Shimada, T., Watanabe, J., Kawajiri, K., Sutter, T. R., Guengerich, F. P., Gillam, E. M., & Inoue, K. (1999). Catalytic properties of polymorphic human cytochrome P450 1B1 variants. *Carcinogenesis*, 20, 1607-1613.

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. *The Holocene*, 28, 1745-1758.

Simonson, T. S., Yang, Y., Huff, C. D., Yun, H., Qin, G., Witherspoon, D. J., ... Ge, R. (2010). Genetic evidence for high-altitude adaptation in Tibet. *Science*, 329, 72-75.

Stevens, C. J., Murphy, C., Roberts, R., Lucas, L., Silva, F., & Fuller, D. Q. 2016. Between China and South Asia: A Middle Asian corridor of crop dispersal and agricultural innovation in the Bronze Age. *The Holocene*, 26, 1541-1555.

Suviolahti, E., Reue, K., Cantor, R. M., Phan, J., Gentile, M., Naukkarinen, J., ... Peltonen, L. (2006). Cross-species analyses implicate Lipin 1 involvement in human glucose metabolism. *Human Molecular Genetics*, 15, 377-386.

Sweeney, C., Lai, C., Riese, D. J. 2nd, Diamonti, A. J., Cantley, L. C., & Carraway KL 3rd. (2000). Ligand discrimination in signaling through an ErbB4 receptor homodimer. *Journal of Biological Chemistry*, 275, 19803-19807.

Szpiech, Z. A., & Hernandez, R. D. (2014). selscan: an efficient multithreaded program to perform EHH-based scans for positive selection. *Molecular Biology and Evolution*, 31, 2824-2827.

Teslovich, T. M., Musunuru, K., Smith, A. V., Edmondson, A. C., Stylianou, I. M., Koseki, M., ... Kathiresan, S. (2010). Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*, 466, 707-713.

Tomisaka, K., Lako, J., Maruyama, C., Anh, N., Lien, D., Khoi, H. H., Van Chuyen, N. (2002). Dietary patterns and risk factors for type 2 diabetes mellitus in Fijian, Japanese and Vietnamese populations. *Asia Pacific Journal of Clinical Nutrition*, 11, 8-12.

- Vaz, F. M., Fouchier, S. W., Ofman, R., Sommer, M., & Wanders, R. J. (2000). Molecular and biochemical characterization of rat gamma-trimethylaminobutyraldehyde dehydrogenase and evidence for the involvement of human aldehyde dehydrogenase 9 in carnitine biosynthesis. *Journal of Biological Chemistry*, 275, 7390-7394.
- Wang, Y., Lu, D., Chung, Y. J., & Xu, S. (2018). Genetic structure, divergence and admixture of Han Chinese, Japanese and Korean populations. *Hereditas*, 155, 19.
- Watkins, P. A., Maignel, D., Jia, Z., & Pevsner, J. (2007). Evidence for 26 distinct acyl-coenzyme A synthetase genes in the human genome. *The Journal of Lipid Research*, 48, 2736-2750.
- West, K. P. & Darnton-Hill, I. (2001). Vitamin A deficiency. In: *Nutrition and Health in Developing Countries* (pp. 267-306). Totowa: Humana Press.
- Willer, C. J., Schmidt, E. M., Sengupta, S., Peloso, G. M., Gustafsson, S., Kanoni, S., ... Global Lipids Genetics Consortium. (2013). Discovery and refinement of loci associated with lipid levels. *Nature Genetics*, 45, 1274-1283.
- Wu, Y. (2006). Overweight and obesity in China. *The British Medical Journal*, 333, 362-363.
- Wullaert, A., Verstrepen, L., Van Huffel, S., Adib-Conquy, M., Cornelis, S., Kreike, M., ... Beyaert, R. (2007). LIND/ABIN-3 is a novel lipopolysaccharide-inducible inhibitor of NF-kappaB activation. *The Journal of Biological Chemistry*, 282, 81-90.
- Yamaguchi, H., Kuboki, Y., Hatori, T., Yamamoto, M., Shiratori, K., Kawamura, S., ... Furukawa, T. (2011). Somatic mutations in PIK3CA and activation of AKT in intraductal tubulopapillary neoplasms of the pancreas. *The American Journal of Surgical Pathology*, 35, 1812-1817.
- Yamanaka, S., Nakamura, I., Watanabe, K. N., & Sato, Y. I. (2004). Identification of SNPs in the waxy gene among glutinous rice cultivars and their evolutionary significance during the domestication process of rice. *Theoretical and Applied Genetics*, 108, 1200-1204.
- Yao, H., & Ye, J. (2008). Long chain acyl-CoA synthetase 3-mediated phosphatidylcholine synthesis is required for assembly of very low density lipoproteins in human hepatoma Huh7 cells. *The Journal of Biological Chemistry*, 283, 849-854.
- Yao-Borengasser, A., Rasouli, N., Varma, V., Miles, L. M., Phanavanh, B., Starks, T. N., ... Kern, P. A. (2006). Lipin expression is attenuated in adipose tissue of insulin-resistant human subjects and increases with peroxisome proliferator-activated receptor gamma activation. *Diabetes*, 55, 2811-2818.
- Yi, X., Liang, Y., Huerta-Sanchez, E., Jin, X., Cuo, Z. X., Pool, J. E., ... Wang, J. (2010).

Sequencing of 50 human exomes reveals adaptation to high altitude. *Science*, 329, 75-78.

Yoon, K. H., Lee, J. H., Kim, J. W., Cho, J. H., Choi, Y. H., Ko, S. H., ... Son, H. Y. (2006). Epidemic obesity and type 2 diabetes in Asia, *Lancet*. 368, 1681-1688.

Zhai, F., Wang, H., Du, S., He, Y., Wang, Z., Ge, K., & Popkin, B. M. (2009). Prospective study on nutrition transition in China. *Nutrition Reviews*, 67, S56-S61.

Zhang, R., Jiang, F., Hu, C., Yu, W., Wang, J., Wang, C., ... Jia, W. (2013). Genetic variants of LPIN1 indicate an association with Type 2 diabetes mellitus in a Chinese population. *Diabetic Medicine*, 30, 118-122.

Zhao, Z. (1998). The middle Yangtze region in China is one place where rice was domesticated: phytolith evidence from Tehe Diaotonghuan Cave, northern Jiangxi. *Antiquity*, 72, 885-897.

Zhao, Z. (2011). New archaeobotanical data for the study of the origins of agriculture in China. *Current Anthropology*, 52, S295-S306.

Zhou, L., & Garvie-Lok, S. J. (2015). Isotopic evidence for the expansion of wheat consumption in northern China. *Archaeological Research in Asia*, 4, 25-35.

Zou, Y., Gong, N., Cui, Y., Wang, X., Cui, A., Chen, Q., ... Chang, Y. (2015). Forkhead Box P1 (FOXP1) Transcription Factor Regulates Hepatic Glucose Homeostasis. *The Journal of Biological Chemistry*, 290, 30607-30615.

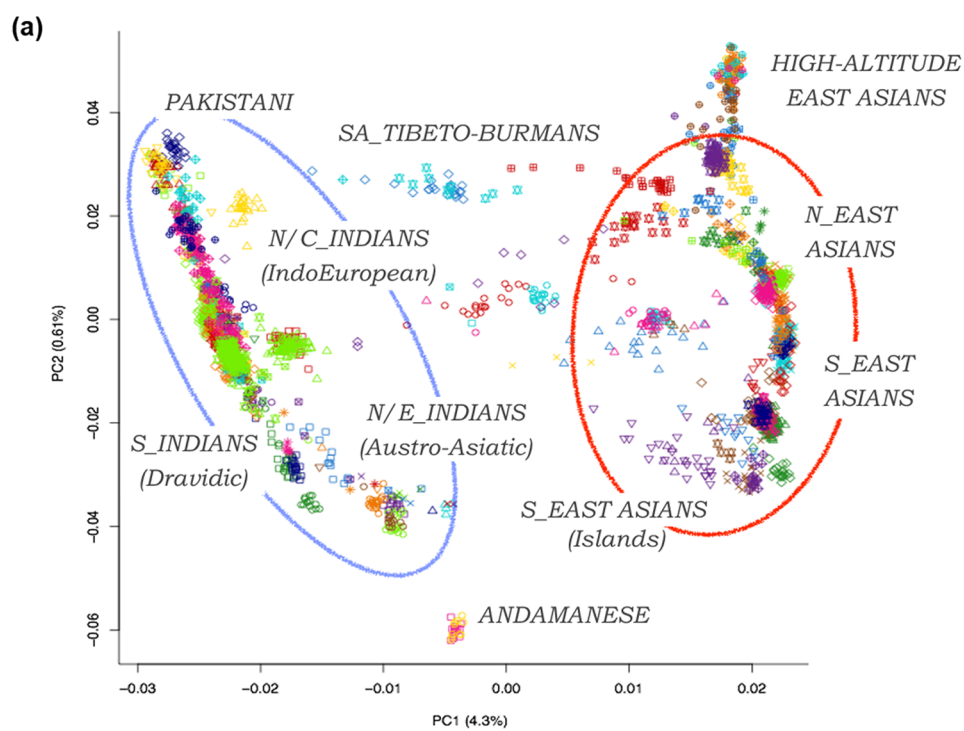
FIGURE LEGENDS

FIGURE 1 Principal component and ADMIXTURE analyses performed on the “Pan-Asian” dataset. (a) PCA was applied to 2,379 individuals belonging to 124 populations from South and East Asia. Plot of PC1 vs PC2 pointed out the distinction between populations from the Indian subcontinent (blue ellipse) and groups from East Asian/South East Asian regions (red ellipse), along with gradients of genetic variation within both of these macro-geographical areas. Individuals are color-coded according to their population of origin, as described in the legend of Figure S2. (b) Results of ADMIXTURE clustering analysis at $K = 8$ performed on a refined dataset of 1,171 samples from 57 populations representative of the gradients of South Asian and East Asian variation (see Figures S3 and S4, for the full set of K s tested and CV-errors). Labels on top of the graph represent the main clusters identified: Pak, Pakistani; NW_Ind, North West Indians; CS_Ind, Central South Indians; Ban, Bangladeshis; Tib-Bur, Tibeto-Bumans; Dai-Viet, Dai-Vietnamese; Japan, Japanese; Han-Tuj, Han-Tujia Chinese; K,

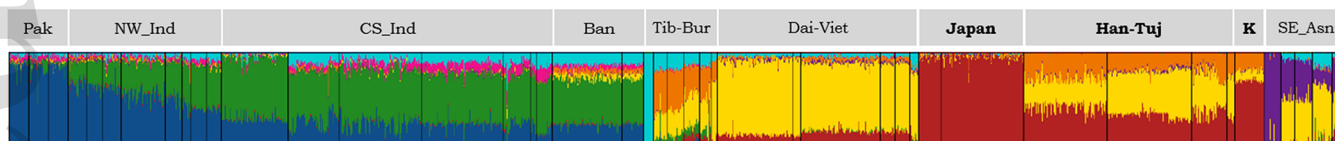
Koreans cluster; SE_Asn, South East Asians. In bold are reported the populations whose ancestors inhabited the regions where wild rice/millet originated and were early domesticated.

FIGURE 2 Approximate geographical distribution of Asian population clusters pointed out by ADMIXTURE analysis and their relative position with respect to known centers of rice/millet domestication. Blue clusters showed predominant South Asian ancestry, while red ones are enriched for East and South East Asian ancestry. Labels of each cluster are reported as described in the legend of Figure 1b. Red concentric circles indicate archaeological sites along the Yangtze River valley in Eastern China where remains suggesting usual consumption of wild rice have been dated to at least 12,000 years ago and where *O. sativa japonica* was early domesticated. From there, rice agriculture diffused primarily to the Korean peninsula and the Japanese archipelago. Green concentric circles indicate archaeological sites in the Hebei and Manchuria provinces of Northern China where remains suggesting early cultivation of broomcorn millet and foxtail millet were found. Populations native from these and the nearby regions thus represent the candidate clusters tested for adaptations to cereal-based diets and are highlighted by bold circles. Conversely, all the remaining clusters were used as control groups (i.e. populations not expected to have evolved adaptations to cereal-based diets despite using rice as a staple food). Blue concentric circles indicate archaeological sites across the Indo-Gangetic Plain where evidence for more recent domestication of *O. sativa indica* were found. Such a domestication process was likely enabled by hybridization of *O. sativa japonica* from China with the local proto-indica rice.

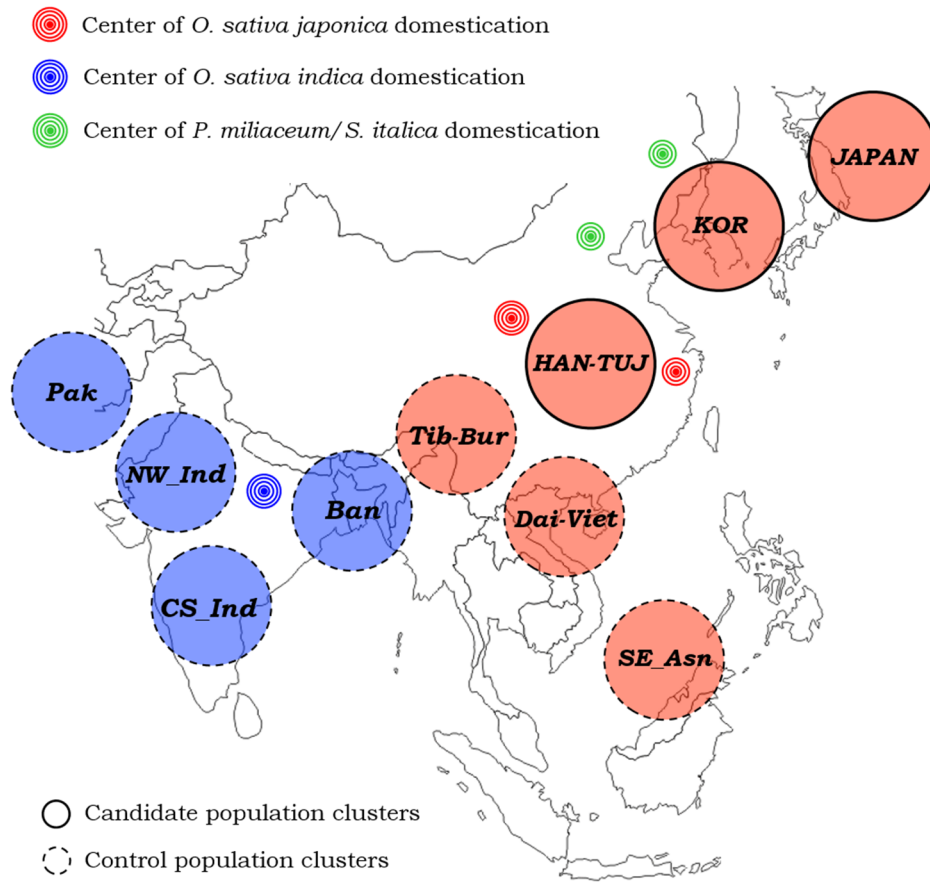
FIGURE 3 Gene networks showing signatures of pervasive positive selection in Han Chinese and/or Japanese populations according to *signet* analysis performed on whole genome distribution of nSL scores and aimed at testing for polygenic adaptation. On the left, genes belonging to the *Glycerolipid metabolism*, *AGE-RAGE signaling in diabetic complications* and *FoxO signaling* pathways pinpointed as candidate adaptive loci in the CHB population. On the right, genes belonging to the *FoxO signaling*, *Glycolysis/Gluconeogenesis* and *Glucagon signaling* pathways pinpointed as candidate adaptive loci in the JPT population. In the centre of the plot are reported the genes belonging to the *FoxO signaling* pathway that showed signatures of positive selection in both the considered populations. The size and colour of the circles are proportional to the nSL peak value associated to each gene according to the displayed colour-scale.



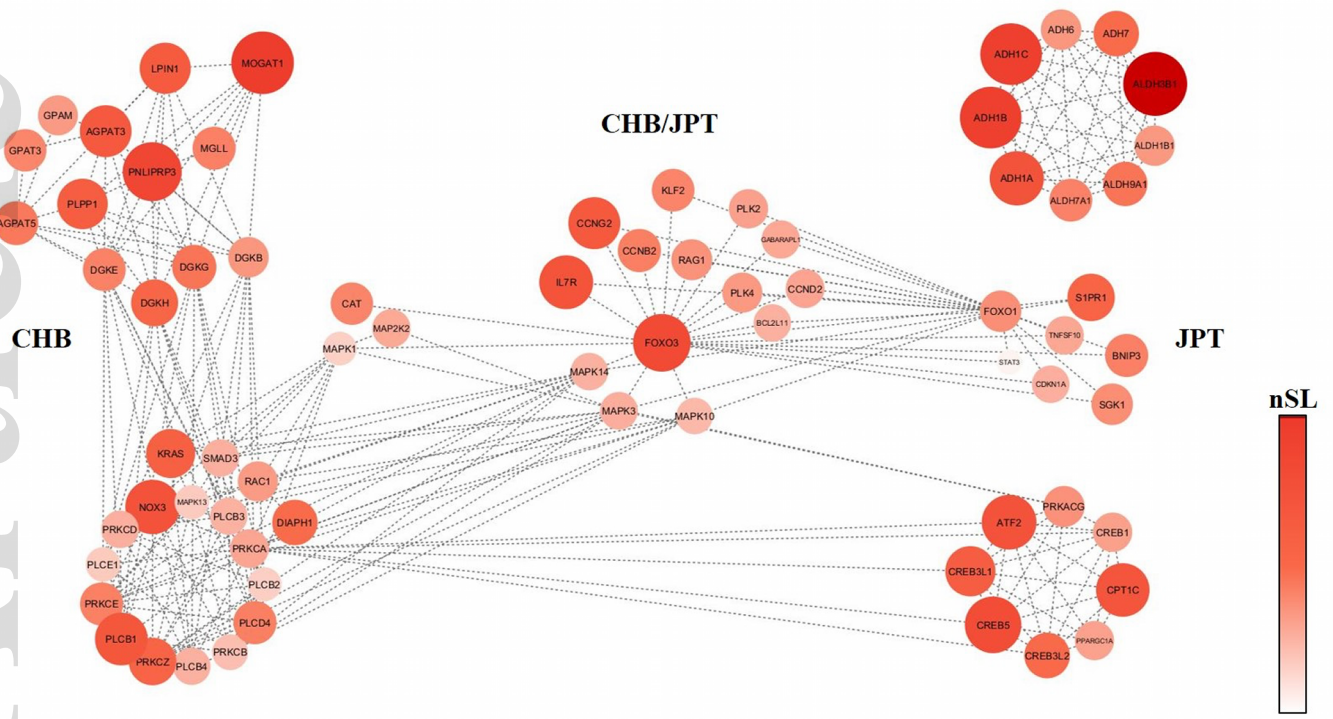
(b)



eva_13090_f1.tif



eva_13090_f2.tif



eva_13090_f3.tif