

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- | | |
|-------------------------------------|--|
| n/a | Confirmed |
| ☐ | ☒ The exact sample size (<i>n</i>) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☒ | ☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| ☒ | ☐ A description of all covariates tested |
| ☒ | ☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☒ | ☐ For null hypothesis testing, the test statistic (e.g. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted
<i>Give P values as exact values whenever suitable.</i> |
| ☒ | ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☐ | ☒ Estimates of effect sizes (e.g. Cohen's <i>d</i> , Pearson's <i>r</i>), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	No software was used to collect the data
Data analysis	Data preparation was performed using ArcGIS 10.8; ArcGIS Pro – ESRI. Data analysis was performed using R version 4.1.2 and R Studio 2021.09.1 Build 372. The packages used are: caret v.6.0-90; sf v.1.0-4; srtingr v.1.4.0; dplyr v.1.0.7; tidyr v.1.1.4; raster v.3.5-2; terra v.1.5-17; ggplot2 v.3.4.3; ggridges v.0.5.3; ggthemes v.0.1.1; ggalluvial v.0.12.5; corrplot v.0.92; dendextend v.1.15.2; hrbrthemes v.0.8.0; viridis v.0.6.2; RColorBrewer v.1.1-3; tmap v.3.3-3; patchwork v.1.2.0; blockCV v.2.1.4; maxnet v.0.1.4; randomForest v.4.6-14; plotROC v.2.2.1. Further details are provided at the following public repository : https://github.com/MaCasti21/Nat-Comm_Castiello_2024 .

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The source archaeological data used in this study are available at: DOI: 10.5334/joad.72;

The paleoclimatic dataset used in this study was published by:

Karger, D. N., Nobis, M. P., Normand, S., Graham, C. H., Zimmermann, N. E. (2020). CHLSA-TraCE21k: Downscaled transient temperature and precipitation data since the last glacial maximum. EnviDat. <https://www.doi.org/10.16904/envodat.211>

and is available at: https://envicloud.wsl.ch/#/?bucket=https%3A%2F%2Ffos.zhdh.cloud.switch.ch%2Fchelsav1%2F&prefix=chelsa_trace%2F.

The computed analyses, archaeological and archaeobotanical (crop) datasets produced in this study are available on the public corresponding author's GitHub at https://github.com/MaCasti21/Nat-Comm_Castiello_2024.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender

Use the terms sex (biological attribute) and gender (shaped by social and cultural circumstances) carefully in order to avoid confusing both terms. Indicate if findings apply to only one sex or gender; describe whether sex and gender were considered in study design; whether sex and/or gender was determined based on self-reporting or assigned and methods used. Provide in the source data disaggregated sex and gender data, where this information has been collected, and if consent has been obtained for sharing of individual-level data; provide overall numbers in this Reporting Summary. Please state if this information has not been collected. Report sex- and gender-based analyses where performed, justify reasons for lack of sex- and gender-based analysis.

Reporting on race, ethnicity, or other socially relevant groupings

Please specify the socially constructed or socially relevant categorization variable(s) used in your manuscript and explain why they were used. Please note that such variables should not be used as proxies for other socially constructed/relevant variables (for example, race or ethnicity should not be used as a proxy for socioeconomic status). Provide clear definitions of the relevant terms used, how they were provided (by the participants/respondents, the researchers, or third parties), and the method(s) used to classify people into the different categories (e.g. self-report, census or administrative data, social media data, etc.) Please provide details about how you controlled for confounding variables in your analyses.

Population characteristics

Describe the covariate-relevant population characteristics of the human research participants (e.g. age, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."

Recruitment

Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.

Ethics oversight

Identify the organization(s) that approved the study protocol.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description

The study combines and integrates the archaeological, archaeobotanical information collected by Antolin F. et al. (2021) with the reconstructed paleoclimate information available in Karger et al 2020, for the western Mediterranean area during the 6th-3rd millennia BC using a statistical and machine learning based approach to investigate the impact of ecological and climatic constraint on the first Neolithic niches of humans and crops and to provide more exhaustive information on the initial spread of agriculture, while shedding light on sustainable practices and guidelines for future more responsible human actions.

Research sample	The study used open access datasets: 1) Archaeological sites and associated radiocarbon dates from Martinez-Grau et al. 2021. https://doi.org/10.5334/joad.72 ; 2) Archaeobotanical information (crops dataset) from Antolin et al. 2021 and Jesus and Antolin 2022. 3) Paleoclimatic variables from Karger et al. 2020 https://doi.org/10.16904/envidat.211 .
Sampling strategy	We used all available data within the open access dataset available at: https://doi.org/10.5334/joad.72 corresponding to the period 5900-2300 BC and covering the extent of the study area.
Data collection	No data collection was performed for this study. We used existing datasets from Martinez-Grau et al. 2021, Antolin et al. 2021, Jesus and Antolin 2022, Karger et al. 2020.
Timing and spatial scale	For timing and spatial scale please refer to Martinez-Grau et al. 2021, Antolin et al. 2021, Jesus and Antolin 2022, Karger et al. 2020.
Data exclusions	No data were excluded from the analysis
Reproducibility	The procedure is fully reproducible with the data and codes provided under https://github.com/MaCasti21/Nat-Comm_Castiello_2024 .
Randomization	Randomization is not relevant to our study as we focused on a specific time period (6th -3rd millennia BC) and geographical extent (Western mediterranean) and used all the data collected and available within this framework (archaeological sites and archaeobotanical information).
Blinding	Blinding is not relevant in this study as no experimental research was carried out.

Did the study involve field work? ☐ Yes ☒ No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks	Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.
Novel plant genotypes	Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.
Authentication	Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosaicism, off-target gene editing) were examined.