nature plants

LETTERS

https://doi.org/10.1038/s41477-021-00848-7

Check for updates

Ancient genomes reveal early Andean farmers selected common beans while preserving diversity

Emiliano Trucchi[1,2,18] ✉, Andrea Benazzo[2,18], Martina Lari[3,18], Alice Iob[2], Stefania Vai[3], Laura Nanni[4], Elisa Bellucci[4], Elena Bitocchi[4], Francesca Raffini[2], Chunming Xu[5], Scott A. Jackson[5], Verónica Lema[6,7], Pilar Babot[8,9], Nurit Oliszewski[8,10], Adolfo Gil[11,12], Gustavo Neme[11,12], Catalina Teresa Michieli[13], Monica De Lorenzi[14,19], Lucio Calcagnile[15], David Caramelli[3], Bastiaan Star[16], Hugo de Boer[17], Sanne Boessenkool[16,18], Roberto Papa[4,18] and Giorgio Bertorelle[2,18] ✉

All crops are the product of a domestication process that started less than 12,000 years ago from one or more wild populations[1,2]. Farmers selected desirable phenotypic traits (such as improved energy accumulation, palatability of seeds and reduced natural shattering[3]) while leading domesticated populations through several more or less gradual demographic contractions[2,4]. As a consequence, the erosion of wild genetic variation[5] is typical of modern cultivars, making them highly susceptible to pathogens, pests and environmental change[6,7]. The loss of genetic diversity hampers further crop improvement programmes to increase food production in a changing world, posing serious threats to food security[8,9]. Using both ancient and modern seeds, we analysed the temporal dynamics of genetic variation and selection during the domestication process of the common bean (*Phaseolus vulgaris*) in the southern Andes. Here, we show that most domestic traits were selected for before 2,500 years ago, with no or only minor loss of whole-genome heterozygosity. In fact, most of the changes at coding genes and linked regions that differentiate wild and domestic genomes are already present in the ancient genomes analysed here, and all ancient domestic genomes dated between 600 and 2,500 years ago are highly variable (at least as variable as modern genomes from the wild). Single seeds from modern cultivars show reduced variation when compared with ancient seeds, indicating that intensive selection within cultivars in the past few centuries probably partitioned ancestral variation within different genetically homogenous cultivars. When cultivars from different Andean regions are pooled, the genomic variation of the pool is higher than that observed in the pool of ancient seeds from north and central western Argentina. Considering that most desirable phenotypic traits are probably controlled by multiple polymorphic genes[10], a plausible explanation of this decoupling of selection and genetic erosion is that early farmers applied a relatively weak selection pressure[2] by using many phenotypically similar but genetically diverse individuals as parents. Our results imply that selection strategies during the past few centuries, as compared with earlier times, more intensively reduced genetic variation within cultivars and produced further improvements by focusing on a few plants carrying the traits of interest, at the cost of marked genetic erosion within Andean landraces.

The onset of domestication by early farmers has been suggested as the period during which the most intense genetic bottleneck affecting genome-wide diversity occurred[4,11]. Yet, artificial selection was probably a continuous process (possibly at different rates in different times) that produced both landraces and, more recently under modern breeding programmes, high-yielding and more resistant elite cultivars[2,12,13]. Understanding when the majority of genetic diversity was lost and how such loss is related to the intensity of the selection process is not only relevant from an evolutionary perspective, but it will also help in planning more sustainable breeding options to revert or mitigate crop genetic erosion[14,15]. We directly address these questions by analysing both modern and ancient genomes of common bean from South America.

Common bean constitutes one of the major sources of vegetable proteins worldwide. The wild progenitor of common bean originated in Mesoamerica and colonized South America along the Andes[16,17]. This natural colonization process, dated by whole-genome analysis between 146,000 and 184,000 years ago[17], was accompanied by an intense and long bottleneck, as supported by the higher genetic diversity observed in the Mesoamerican than in the Andean gene pool[18]. The domestication of common bean occurred independently more or less simultaneously in both Mesoamerica and the Andes about 8,000 years ago, followed by divergence into distinct landraces due to drift, local selection and adaptation[16].

We here obtained the first whole-genome shotgun data from 30 ancient common bean seeds representing nine archaeological sites

[1]Department of Life and Environmental Sciences, Marche Polytechnic University, Ancona, Italy. [2]Department of Life Sciences and Biotechnology, University of Ferrara, Ferrara, Italy. [3]Department of Biology, University of Florence, Firenze, Italy. [4]Department of Agricultural, Food, and Environmental Sciences, Marche Polytechnic University, Ancona, Italy. [5]Center for Applied Genetic Technologies, University of Georgia, Athens, GA, USA. [6]Universidad Nacional de Córdoba, Córdoba, Argentina. [7]Conicet, Consejo Nacional de Investigaciones Científicas y Técnicas, Córdoba, Argentina. [8]ISES, Instituto Superior de Estudios Sociales, CONICET, Consejo Nacional de Investigaciones Científicas y Técnicas, Tucumán, Argentina. [9]Instituto de Arqueología y Museo, Universidad Nacional de Tucumán, Tucumán, Argentina. [10]Universidad Nacional de Tucumán, Tucumán, Argentina. [11]Instituto de Evolución, Ecología Histórica y Ambiente (CONICET & UTN FRSR), San Rafael, Argentina. [12]Museo de Historia Natural de San Rafael, San Rafael, Argentina. [13]Instituto de Investigaciones Arqueológicas y Museo "Prof. Mariano Gambier", Universidad Nacional de San Juan, San Juan, Argentina. [14]Museo Arqueológico de Cachi, Cachi, Argentina. [15]CEDAD (Centre of Applied Physics, Dating and Diagnostics), Department of Mathematics and Physics "Ennio De Giorgi", University of Salento, Lecce, Italy. [16]Centre for Ecological and Evolutionary Synthesis, Department of Biosciences, University of Oslo, Oslo, Norway. [17]Natural History Museum, University of Oslo, Oslo, Norway. [18]These authors contributed equally: Emiliano Trucchi, Andrea Benazzo, Martina Lari, Sanne Boessenkool, Roberto Papa, Giorgio Bertorelle. [19]Deceased: Monica De Lorenzi. ✉e-mail: e.trucchi@univpm.it; ggb@unife.it

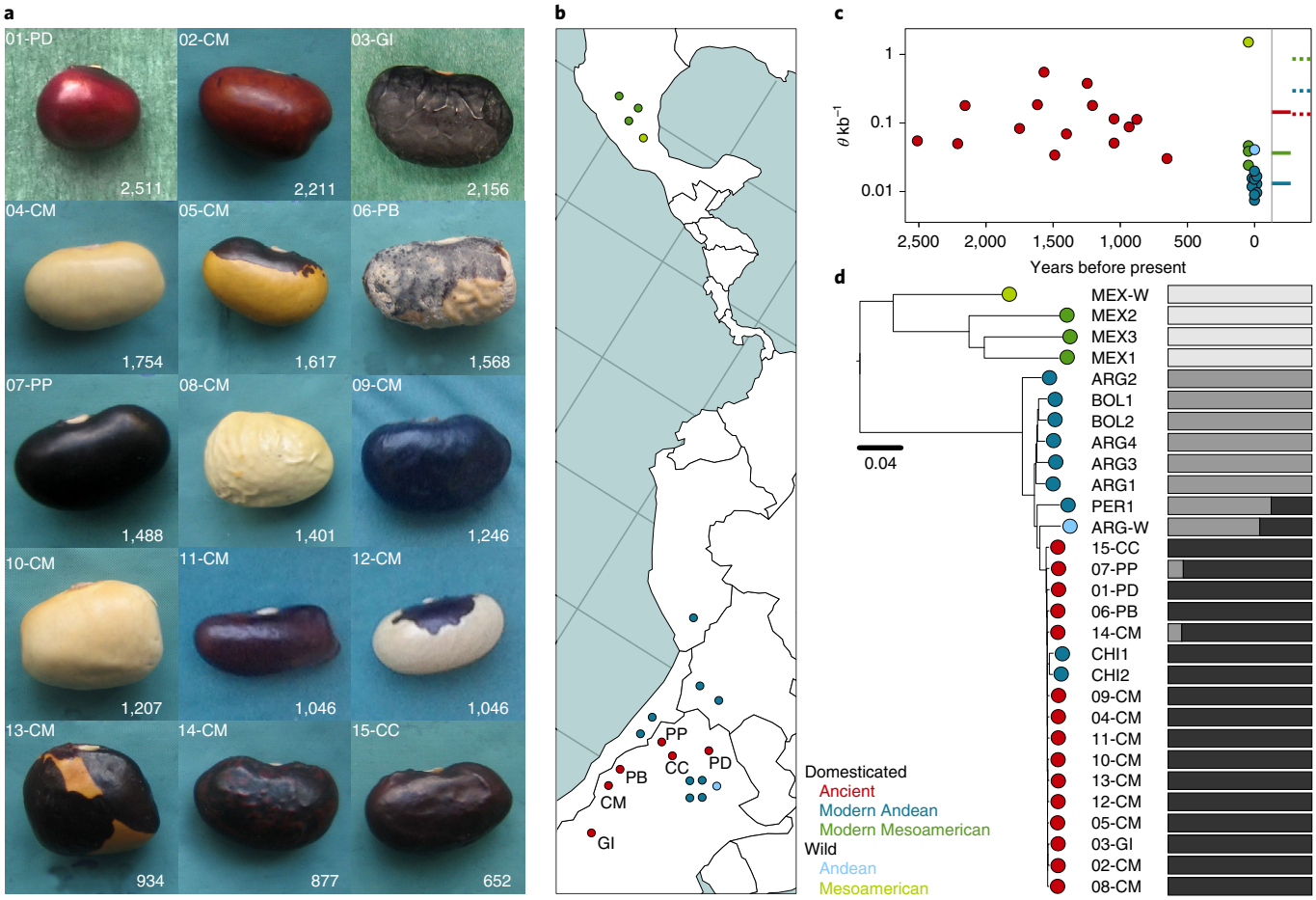


Fig. 1 | Diversity in ancient and modern common bean. a, Pictures of the 15 common bean seeds sequenced for the whole-genome analyses. The sample ID with two letters referring to the archaeological site of origin as shown in **b** is reported in the top left corner of each picture (Supplementary Section 1), and the AMS radiocarbon age in years before the present is reported in the bottom right corner. **b**, Map showing the locations of the archaeological sites where the seeds were retrieved (red) and the countries of origin of the modern domesticated cultivars from South and Central America (blue and green, respectively) and the wild specimens (pale blue and pale green). The modern seed locations are indicative of only the country of origin. **c**, Whole-genome diversity estimated as $\theta\,\mathrm{kb}^{-1}$ in individual seeds (circles), the average of individual estimates per group (solid lines) and in groups of seeds (dashed lines) using only transversions in callable regions (see Supplementary Section 5 for the details). Only domesticated cultivars are included in the group estimates. **d**, Neighbour-joining tree based on genetic distances and admixture analysis both based on transversions only (one single nucleotide polymorphism (SNP) every 50 kb was used in the admixture analysis) in callable regions. Note that modern seeds from Chile cluster with the ancient seeds.

in north and central western Argentina (Supplementary Sections 1, 2 and 4, Supplementary Fig. 1 and Supplementary Table 1). All 30 seeds were dated between 2,500 and 600 years before the present by accelerator mass spectrometry (AMS) radiocarbon dating (Supplementary Section 3). Initial screening showed remarkable DNA preservation, with an average of 44% (s.d. = 12%) endogenous DNA mapped to the common bean reference genome[17] (Supplementary Table 1). Unimodal length distributions with short fragment lengths (average, 65 base pairs; s.d., 22 base pairs) and purine enrichment upstream of the break points were consistent with those expected for degraded DNA (Supplementary Table 1 and Supplementary Figs. 3 and 4). The presence of deamination patterns at the ends of the reads varied from 0.76% to 32.75% C to T and G to A misincorporations, with half of the seeds showing percentages below 10% (Supplementary Table 1 and Supplementary Fig. 5). The damage patterns were partially explained by the locations of the archaeological sites: all seeds from sites at >2,500 meters above sea level showed less than 10% base misincorporations (with many lower than 5%), whereas most seeds from lower altitudes exhibited levels above 15% (Supplementary Fig. 6). The observed pattern shows that both favourable environmental conditions and the structure of the seed (for example, the presence of an external cuticle) probably resulted in good DNA preservation, as previously suggested for quinoa seeds from the same sites[19]. A subset of the specimens (Fig. 1a,b) was selected for further sequencing on the basis of DNA preservation and the representation of ages and locations, resulting in a final dataset of 15 ancient bean genomes with coverage from 4.2× to 23.2× in the non-repeated regions (Supplementary Table 1). Whole-genome data from modern common bean accessions (domesticated landraces/cultivars: nine Andean, three Mesoamerican; wild: one Andean, one Mesoamerican) and a closely related species (*Phaseolus hintonii*, one accession) were also included in the analyses (Supplementary Section 6 and Supplementary Table 3). All statistical analyses (excluding the selection test, where only mutations observed in all ancient seeds are considered) were performed on transversions to avoid the bias introduced by post-mortem deamination[20].

Individual heterozygosity (probabilistically estimated as the variation parameter θ every thousand nucleotides (kilobase, kb); Supplementary Section 7) is higher in all ancient seeds than in each

of nine modern landraces/cultivars from the Andean gene pool (Fig. 1c and Supplementary Table 4), and the average heterozygosity is more than ten times higher in ancient than in modern seeds (0.144 versus 0.013 θkb^{-1}, Mann–Whitney U test, $P < 0.001$). This difference remains high (0.110 versus 0.013 θkb^{-1}) and highly significant when the ancient seeds with the lowest coverage are excluded (Supplementary Section 7). The effect of decreased heterozygosity due to the selfing procedure applied to modern accessions in seed banks was tested in two ways: by pooling pairs of modern seeds and estimating θkb^{-1} in different seed pairs, and by increasing the heterozygosity values in modern seeds as a function of the number of selfing cycles to which they were subjected (Supplementary Section 8). As expected, the variation in modern pairs increased compared with single seeds, but it is on average still significantly smaller (at least 2.5 times) than that observed in ancient seeds (Mann–Whitney U test, $P < 0.005$; Supplementary Fig. 9 and Supplementary Table 5). It is possible that our pairing strategy did not entirely recover the heterozygosity just before the implementation of the seed bank protocol. A more exact estimate cannot be obtained since the detailed protocols applied to each modern Andean cultivar in seed banks are not known, and some level of outcrossing during this procedure cannot be excluded. Furthermore, the average θkb^{-1} is probably increased in modern seed pairs because our paired seeds belong to different landraces (that is, the increase in the seed pairs is inflated by the genetic structure among landraces). When single-seed heterozygosity was conservatively doubled or quadrupled in modern seeds subjected to one or two cycles of selfing, respectively, the average variation remained significantly higher in the ancient seeds (Supplementary Section 8). We therefore conclude that single modern seeds representative of modern landraces have lower genetic variation than single ancient seeds.

The modern genome available from a wild Andean seed[21] and two publicly released but unpublished genomes from wild Argentinean seeds (Supplementary Section 7) have a heterozygosity lower than that of the wild Mesoamerican seed in our dataset but comparable to that observed in the ancient Argentinean seeds (Fig. 1c and Supplementary Fig. 7). This result is consistent with the bottleneck that occurred during the natural colonization of the Andes from Mesoamerica[18], and it indirectly supports the finding that modern Andean domestic seeds have lower diversity than ancient domestic seeds, because otherwise we should exclude any decrease of heterozygosity throughout the whole domestication process up to the modern cultivars. In addition, this result seems compatible with a minor effect of the initial Andean domestication on bean genetic diversity. With the current dataset, it cannot be ruled out that heterozygosity was initially reduced following domestication but subsequently restored by mutation. Such an effect would, however, require a mutation rate ten times higher than that estimated on the basis of the divergence time between *P. vulgaris* and *P. hintonii* (Supplementary Section 10). These findings challenge the conventional view of the domestication bottleneck resulting in severe genetic erosion[3] and is consistent with recent observations in maize, sorghum and barley[22].

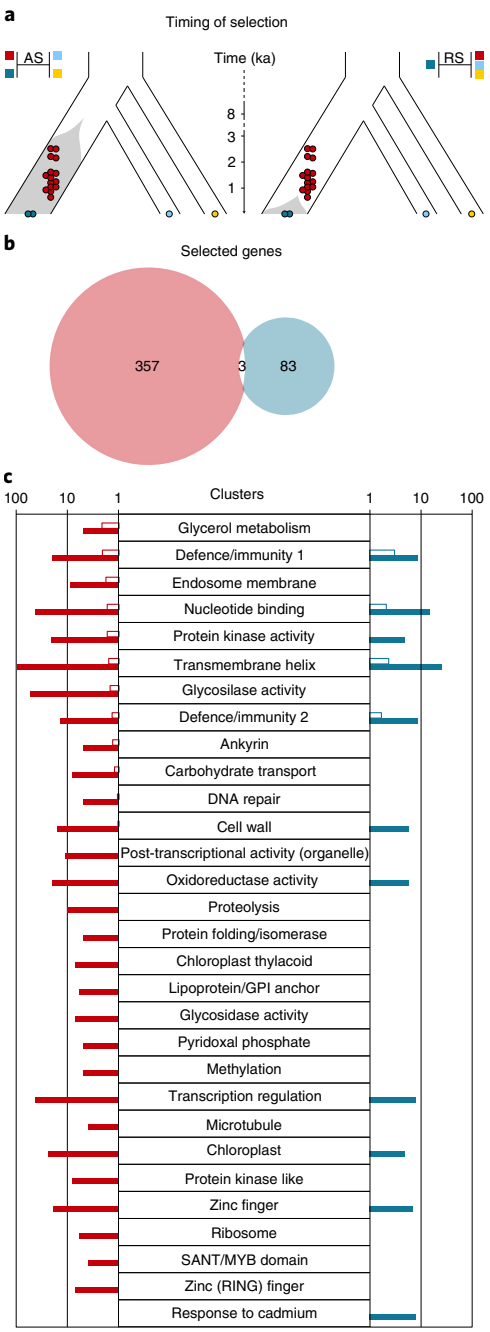


Fig. 2 | Timing of selection in common bean from South America. **a**, Alternative topologies for a SNP. On the left (AS, ancient selection), the SNP is fixed for an alternative variant in both the ancient seeds from Argentina (red circles) and the modern cultivars from Chile (blue circles) as compared with the wild Andean *P. vulgaris* seed (pale blue circle) and the outgroup (*P. hintonii*, yellow circle). On the right (RS, recent selection), the SNP is fixed for an alternative variant in only the modern cultivars from Chile as compared with the ancient seeds, the wild Andean and the outgroup. The grey shading represents the temporal fixation of an alternative allele. The ancient seeds' age as well as the putative divergence time between Andean domesticated and wild *P. vulgaris* is reported (ka, thousand years ago). Using the genome of the wild Mesoamerican *P. vulgaris* as the outgroup instead of *P. hintonii* produced largely similar results (Supplementary Table 7). **b**, Number of genes significantly enriched with SNPs in RS (blue) and/or AS (red) topologies (FDR < 0.001, top 1% for the proportion of SNPs within gene ±1 kb, and with an *Arabidopsis thaliana* orthologue). **c**, Functional gene clusters (Supplementary Section 11) in RS (blue) or AS (red) enriched groups (the complete list of genes can be found in Supplementary Table 7). The filled bars indicate the maximum gene count in the cluster; the empty bars indicate the enrichment score of the cluster (only scores >1 are reported). All genes (27,000) in the common bean genome were tested for a significant enrichment of fixed alternative alleles according to the topologies described above, taking into account the different levels of drift expected in periods of different length (see Supplementary Section 11 for the details).

The lack of a temporal trend across ancient seeds of different ages (Fig. 1c) indicates that genome-wide diversity was largely maintained between 2,500 and 600 years ago by agricultural practices of Andean societies. The loss of diversity within each cultivar therefore occurred more recently, certainly less than 600 years ago and probably in the past century[13]. A similar loss of diversity characterizes the recent domestication history of horses[20]. The absence of data on seeds from the past 600 years currently hinders a direct reconstruction of the temporal dynamics of genomic diversity during the past six centuries.

When estimating genetic variation in groups of modern Andean ($0.294\ \theta kb^{-1}$) or Mesoamerican seeds ($0.860\ \theta kb^{-1}$), we observe a substantial increase compared with the average individual estimates (Fig. 1c), indicating that modern genomic diversity is highly structured in different, highly homogeneous cultivars. The abandonment and extinction of any modern cultivar would therefore result in the substantial loss of its private fraction of the whole crop diversity[8,23]. In contrast, the estimate of genetic diversity obtained by pooling all ancient individuals is very similar to the average of the 15 individual estimates ($0.134\ \theta kb^{-1}$; Fig. 1c), implying that these ancient genomes represent a single genomic pool. Ancient seeds, as a group, do not reach the level of variation observed in the pooled modern cultivars. This is not unexpected considering that modern cultivars have a long, complex and partially independent history of recent landrace establishment, and the ancient domestic seeds were used during the Holocene by human populations with shared subsistence practices, which included the use of wild beans[24]. A large fraction of the ancient seeds could individually recover most of the ancient crop diversity in that area, while modern seeds never reach individual heterozygosity values close to the pooled group from Chile, Bolivia, Argentina and Peru, or the groups obtained by separately pooling the Chilean ($0.045\ \theta kb^{-1}$), Bolivian ($0.102\ \theta kb^{-1}$) or Argentinean ($0.198\ \theta kb^{-1}$) landraces. A more widespread sample of ancient beans might also reveal that ancient populations harboured more genetic diversity than pooled cultivars.

Three main genetic groups were revealed by different clustering methods (Supplementary Section 9) and a whole-genome neighbour-joining tree (Fig. 1b and Supplementary Figs. 10 and 11). The largest divergence is observed between Mesoamerican and Andean genomes, while the Andean seeds are further partitioned into two less-differentiated groups. All ancient seeds belong to the same genomic clade, indicating that the same or similar ancient landraces were in use in Argentina for about 2,000 years. This clade also includes modern landraces used today on the other side of the Andes, in Chile. The so-called Chilean race[25] therefore seems to be the direct descendants of the seeds used in Argentina before the Inca conquest of the area occurred at the end of the fifteenth century[26]. We interpret this result as a consequence of the trade of goods, including crops, between the eastern (in Argentina) and western (in Chile) slopes of the Andes, with early farmers relying on llama caravans for regular long-distance exchanges[27]. Modern Argentinian cultivars, in contrast, do not trace their ancestry to the local ancient cultivar but were probably introduced in the area some time during the past 600 years.

A gene-by-gene selection scan (Fig. 2a) was performed to discriminate early (>2,500 years ago) from late (<600 years ago) selection targets, by taking into account the different levels of genetic drift expected in the two periods of different lengths (Supplementary Section 11). We found that (1) a selection signature is present in 443 of 27,000 genes tested (false discovery rate (FDR), <0.001); (2) selection affected many more genes (~4.2 times more) in the early phase than in the late phase (Fig. 2b), suggesting that the recent improvement of cultivars strongly affected few genes, but the main genomic turnover occurred several thousand years ago (if adjacent genes are merged in 'sweep blocks', thus considering that genes under selection may produce selection signals also in adjacent genes, the ratio between ancient and recent sweeps remains larger than 3; see Supplementary Section 11); and (3) genes selected in the early phase belong to functional clusters mainly related to glycerol metabolism, carbohydrate and sugar transport and metabolism, intracellular transport (endosome membrane), regulatory elements (nucleotide binding), the modification of proteins, and glycosylation, whereas recent selection primarily targeted traits involved in immunity and defence, regulatory elements, and transmembrane transport (Fig. 2c and Supplementary Table 7). Evidence for selection on sugar biosynthesis genes at the early stages of domestication has also been found in ancient maize from 2,000 and 750 years ago[28]. One of the functional clusters identified in the early selected genes includes the MYB-domain genes that are important regulatory elements of development, metabolism and responses to biotic and abiotic stresses[29] and have been suggested as causative of different phenotypic changes associated with the shattering trait in common bean[30], maize and rice[4,31].

Genomic studies of ancient DNA are much less numerous in plants than in animal remains[32,33], but they have already provided fundamental insights into the process of domestication[28,34–38]. Our results show that common bean seeds, and probably legume seeds in general, can be an excellent source of high-quality ancient DNA, especially if the seeds are preserved in favourable climatic conditions (that is, cold and dry environments). The dynamics of domestication are still debated[22], but, as recently summarized[2], the consensus is that domestication was slow and gradual, with selected traits emerging in association with population decline and the loss of genetic variation. Here, we contribute to this debate with a study on ancient bean genomics and provide an alternative example in which selection and the loss of variation are decoupled during the process of domestication. Our data reveal that in common bean early agriculture in the Andes, genomic diversity was preserved within single seeds, whereas more recent breeding practices produced structured landraces that are internally more homogenous. At the same time, assuming that the genetic changes we infer correspond at least in part to phenotypic changes, early agriculture was also very efficient in selecting most of the desirable traits now typical of this crop. We hypothesize that these patterns are the result of the larger number of seeds used by early farmers as founders in each generation, all probably displaying the selected trait(s) but heterogeneous in the rest of the genome. The initial improvement of the common bean was therefore based on serial soft sweeps, as has also been suggested for the domestication of maize[10]. Encompassing several thousand years and probably assisted by cultivar exchanges and hybridization with wild plants[2,25,39,40], this breeding practice allowed many traits to be selected for, without the substantial loss of genomic variation that has probably occurred in more recent times within landraces and cultivars.

Methods

Ancient and modern bean samples. Ancient bean seeds were selected from five museum collections in Argentina: Museo de La Plata (La Plata, Buenos Aires, Argentina); Museo de Historia Natural de San Rafael (Parque Mariano Moreno, San Rafael, Argentina); Instituto de Investigaciones y Museo Arqueológico "Prof. Mariano Gambier", Universidad Nacional de San Juan (San Juan, Argentina); Instituto de Arqueología y Museo, Universidad Nacional de Tucumán (Tucumán, Argentina); and Museo Arqueológico Pio Pablo Díaz (Cachi, Salta, Argentina).

Seeds were originally collected from nine archaeological sites located in different geographical regions of north and central-western Argentina (Fig. 1a). A brief description of each site is reported in Supplementary Section 1. The codes for each seed, referring to Fig. 1b, are reported in Supplementary Table 1.

Thirty ancient bean seeds (Supplementary Fig. 1) were selected for preliminary screening on the basis of the site locations and archaeological context. All seeds had a clear domestic phenotype. After visual inspection, we selected seeds with different morphological features (for example, coat colour and dimensions) and good preservation (for example, less visible cracks or damages in the outer coat). The seeds were AMS radiocarbon dated at CEDAD, Center of Applied Physics, Dating and Diagnostics, Department of Mathematics and Physics "Ennio De Giorgi", University of Salento.

Genomic data already available from modern domesticated and wild common bean samples were included in our analyses (Supplementary Section 6 and Supplementary Table 3).

Ancient DNA extraction and sequencing. DNA extraction and library preparation were performed at the Molecular Anthropology and Paleogenetic Laboratory of the Department of Biology, University of Florence, using facilities that are exclusively dedicated to processing ancient DNA and following all the necessary precautions to minimize contamination from exogenous DNA (Supplementary Section 4). One sample (09-CM) was extracted and a library built at the ancient DNA laboratory of the University of Oslo, Norway.

The libraries were sequenced at the Norwegian Sequencing Centre, University of Oslo, Norway, where all samples were pooled in equimolar ratios and paired-end sequenced on an Illumina HiSeq 2500. First, the samples were screened for DNA preservation by low-coverage sequencing. On the basis of the results of this first sequencing run and sample characteristics (age, location, phenotype, mapping quality and age-related damage patterns), a subset of 18 samples was repooled and sequenced further on four HiSeq 2500 lanes. Additionally, one sample (O1-PD) was sequenced on one lane on the HiSeq 4000 platform. One sample (09-CM) was sequenced on the HiSeq 2500 in a different pool from all other beans. On the basis of the final results, we further selected 15 samples with coverage higher than 4× for downstream population genomic analyses.

Ancient DNA validation and mapping. Paleomix v.1.2.12 (ref. [41]), a pipeline specifically designed for ancient DNA, was employed for the filtering and mapping of the ancient DNA data (Supplementary Section 5). The filtered reads were aligned to the *P. vulgaris* reference genome v.2.1 (ref. [17]). Fragment length distributions, break points and misincorporation patterns were explored in each sample (Supplementary Figs. 3–5) using MapDamage2.0 (ref. [42]) as a function of sample age and site location (Supplementary Fig. 6). Downstream analyses including both ancient and modern samples were based on transversions, since transitions are more affected by post-mortem damages than transversions and could reflect DNA damage rather than true DNA variation. Only the selection test (Supplementary Section 11) considered both transitions and transversions, because in this test only changes observed in all the 15 ancient seeds are used. We estimate that these simultaneous changes are very unlikely to be related to post-mortem damages. To further support the absence of contamination in our ancient DNA data, we estimated the heterozygosity level in the chloroplast calls. As contamination can introduce additional chloroplast haplotypes in the original library, resulting in the presence of heterozygous sites, we performed a SNP calling analysis in chloroplast reads extracted from the ancient seeds.

Estimates of genomic diversity and inference of population structure. We compared levels of diversity between ancient and modern common bean seeds as both individual-level and population-level diversity. Genetic diversity was measured (using only transversions) by estimating Watterson's θ (θ_w) in non-overlapping windows of 10 kb using ANGSD v.0.916 (ref. [43]), while taking into account per-individual inbreeding coefficients and hence accounting for possible deviations from Hardy–Weinberg equilibrium, as proposed by Vieira et al.[44] (Supplementary Section 7). As the genomic diversity of each modern accession was probably reduced due to the selfing protocol applied after collecting the seeds from the seed bank, different strategies were implemented to exclude major effects of selfing on heterozygosity estimates of single modern seeds (Supplementary Section 8).

The main components of the genetic diversity and the eventual admixture proportion in all samples were inferred by admixture analyses[45], neighbour-joining trees on genome-wide distance matrices and FineStructure[46] (Supplementary Section 9).

Inference of ancient and modern selection. A gene-by-gene selection scan was performed to identify genes targeted by artificial selection during the early and/or late stages of domestication. A gene underlying a desirable trait is expected to have undergone one or more selective sweeps targeting specific alleles coupled with recurrent hitchhiking at linked neutral sites. Selected genes will then accumulate variants, which either are directly advantageous or got fixed because of locally low effective population size due to linked selection[47–49]. We then expect to find an excess of fixed differences at genes under selection when samples from pre-domestication (that is, wild) and post-domestication populations are compared.

We tested each of the 27,000 genes for a significant enrichment of fixed alternative alleles in two possible configurations. The first configuration is AS, indicated by the topology in the left panel in Fig. 2a, where all the ancient and modern Chilean samples have the same allele, and the wild samples (one wild Andean *P. vulgaris* and one *P. hintonii* as a wild outgroup) have the alternative allele. We expected AS to be frequent in genes that experienced changes before 2,500 years ago. The second configuration is RS, indicated by the topology in the right panel in Fig. 2b, where all modern Chilean samples have the same allele, and all the other samples (ancient and wild samples) have the alternative alleles. We expected RS to be frequent in genes that experienced changes in the past 600 years—that is, more recently than the estimated age interval for the ancient seeds.

Clearly, the enrichment of AS or RS allelic configurations (corresponding to the different topologies based on fixed differences) depends on many factors besides the selection dynamic, such as the sample sizes and the drift intensity in different time intervals. We therefore used a randomization procedure that keeps the groups, their sample sizes and the total number of SNPs constant, but randomly assigns the SNPs to the genes. The relative frequencies of SNPs with fixed differences in AS and RS topologies, computed after this randomization, becomes our null hypothesis, which is used to test whether the observed frequencies of these topologies in each gene are higher than expected by chance. The procedure we applied is highly conservative, and we probably missed genes where the selection process may have modified the allelic compositions. It allows us, however, to significantly reduce the rate of false positives and to correctly estimate the relative proportion of genes with relevant modifications before and after the time interval defined by the ages of the seeds we analysed.

The two final lists of genes showing signatures of AS and/or RS were grouped into functional clusters using the web-based software David[50]. For this analysis, we could use only genes with a known orthologue in *A. thaliana*, using the functional enrichment score estimated by David[50] to rank the functional clusters found in the two lists of genes without excluding any cluster. The analyses were also replicated using the genome of a wild Mesoamerican *P. vulgaris* as the outgroup instead of *P. hintonii* (Supplementary Table 7).

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The sequenced raw reads from the ancient samples are publicly available as NCBI Bioproject (ID: PRJNA574560). The modern sample genomic data are available with the following NCBI accession codes: SRR10161640, SRR10161629, SRR10161767, SRR10161647, SRR10161646, SRR10161645, SRR10161601, SRR10161592, SRR10161584, SRR10161684, SRR10161683, SRR10161697, SRR10161690, SRR10161688, SRR10161651, SRR10161745 and SRR10161723.

Code availability

The custom scripts developed in this study are publicly available in the GitHub repository at https://github.com/anbena/ancient-beans.

Received: 19 October 2019; Accepted: 4 January 2021;
Published online: 8 February 2021

References

1. Diamond, J. Evolution, consequences and future of plant and animal domestication. *Nature* **418**, 700–707 (2002).
2. Purugganan, M. D. Evolutionary insights into the nature of plant domestication. *Curr. Biol.* **29**, 705–714 (2019).
3. Meyer, R. S. & Purugganan, M. D. Evolution of crop species: genetics of domestication and diversification. *Nat. Rev. Genet.* **14**, 840–852 (2013).
4. Doebley, J. F., Gaut, B. S. & Smith, B. D. The molecular genetics of crop domestication. *Cell* **127**, 1309–1321 (2006).
5. Van de Wouw, M., Kik, C., van Hintum, T., van Treuren, R. & Visser, B. Genetic erosion in crops: concept, research results and challenges. *Plant Genet. Resour.* **8**, 1–15 (2010).
6. Babiker, E. M. et al. Mapping resistance to the Ug99 race group of the stem rust pathogen in a spring wheat landrace. *Theor. Appl. Genet.* **128**, 605–612 (2015).
7. Dale, J. et al. Transgenic Cavendish bananas with resistance to *Fusarium* wilt tropical race 4. *Nat. Commun.* **8**, 1496 (2017).
8. Esquinas-Alcázar, J. Protecting crop genetic diversity for food security: political, ethical and technical challenges. *Nat. Rev. Genet.* **6**, 946–953 (2005).
9. Gepts, P. Plant genetic resources conservation and utilization. *Crop Sci.* **46**, 2278–2292 (2006).
10. Beissinger, T. M. et al. Recent demography drives changes in linked selection across the maize genome. *Nat. Plants* **2**, 16084 (2016).
11. Hyten, D. L. et al. Impacts of genetic bottlenecks on soybean genome diversity. *Proc. Natl Acad. Sci. USA* **103**, 16666–16671 (2006).
12. Fuller, D. Q. et al. Convergent evolution and parallelism in plant domestication revealed by an expanding archaeological record. *Proc. Natl Acad. Sci. USA* **111**, 6147–6152 (2014).
13. Khush, G. S. Green revolution: the way forward. *Nat. Rev. Genet.* **2**, 815–822 (2001).
14. Fu, Y. B. Understanding crop genetic diversity under modern plant breeding. *Theor. Appl. Genet.* **128**, 2131–2142 (2015).
15. Bevan, M. W. et al. Genomic innovation for crop improvement. *Nature* **543**, 346–354 (2017).
16. Bitocchi, E. et al. Mesoamerican origin of the common bean (*Phaseolus vulgaris* L.) is revealed by sequence data. *Proc. Natl Acad. Sci. USA* **109**, E788–E796 (2012).

17. Schmutz, J. et al. A reference genome for common bean and genome-wide analysis of dual domestications. *Nat. Genet.* **46**, 707–713 (2014).
18. Bitocchi, E. et al. Beans (*Phaseolus* ssp.) as a model for understanding crop evolution. *Front. Plant Sci.* **8**, 722 (2017).
19. Winkel, T. et al. Discontinuities in quinoa biodiversity in the dry Andes: an 18-century perspective based on allelic genotyping. *PLoS ONE* **13**, e0207519 (2018).
20. Fages, A. et al. Tracking five millennia of horse management with extensive ancient genome time series. *Cell* **177**, 1419–1435 (2019).
21. Rendón-Anaya, M. et al. Genomic history of the origin and domestication of common bean unveils its closest sister species. *Genome Biol.* **18**, 60 (2017).
22. Allaby, R. G., Ware, R. L. & Kistler, L. A re-evaluation of the domestication bottleneck from archaeogenomic evidence. *Evol. Appl.* **12**, 29–37 (2019).
23. Castañeda-Álvarez, N. P. et al. Global conservation priorities for crop wild relatives. *Nat. Plants* **2**, 16022 (2016).
24. Pochettino, M. L. & Scattolin, M. C. Identificación y significado de frutos y semillas carbonizados de sitios arqueológicos de la ladera occidental del Aconquija, Prov. Catamarca, Rca. Argentina. *Rev. Mus. La Plata, Antropol.* **9**, 169–181 (1991).
25. Singh, S. P., Gepts, P. & Debouck, D. G. Races of common bean (*Phaseolus vulgaris*, Fabaceae). *Econ. Bot.* **45**, 379–396 (1991).
26. Williams, V. I. Formaciones sociales en el noroeste argentino: variabilidad prehispánica en el surandino durante el Periodo de Desarrollos Regionales y el estado Inca. *Rev. Haucaypata* **9**, 62–76 (2015).
27. Núñez, L. & Nielsen, A. E. *En Ruta: Arquelogía, Historia y Etnografía del Tráfico Surandino* (Encuentro Grupo Editor, 2011).
28. Da Fonseca, R. R. et al. The origin and evolution of maize in the southwestern United States. *Nat. Plants* **1**, 14003 (2015).
29. Dubos, C. et al. MYB transcription factors in *Arabidopsis*. *Trends Plant Sci.* **15**, 573–581 (2010).
30. Rau, D. et al. Genomic dissection of pod shattering in common bean: mutations at non-orthologous loci at the basis of convergent phenotypic evolution under domestication of leguminous species. *Plant J.* **97**, 693–714 (2019).
31. Saitoh, K., Onishi, K., Mikami, I., Thidar, K. & Sano, Y. Allelic diversification at the *C* (*OsC1*) locus of wild and cultivated rice: nucleotide changes associated with phenotypes. *Genetics* **168**, 997–1007 (2004).
32. Estrada, O., Breen, J., Richards, S. M. & Cooper, A. Ancient plant DNA in the genomic era. *Nat. Plants* **4**, 394–396 (2018).
33. Brunson, K. & Reich, D. The promise of paleogenomics beyond our own species. *Trends Genet.* **35**, 319–329 (2019).
34. Mascher, M. et al. Genomic analysis of 6,000-year-old cultivated grain illuminates the domestication history of barley. *Nat. Genet.* **48**, 1089–1093 (2016).
35. Ramos-Madrigal, J. et al. Genome sequence of a 5,310-year-old maize cob provides insights into the early stages of maize domestication. *Curr. Biol.* **26**, 3195–3201 (2016).
36. Wagner, S. et al. High throughput DNA sequencing of ancient wood. *Mol. Ecol.* **27**, 1138–1154 (2018).
37. Kistler, L. et al. Multiproxy evidence highlights a complex evolutionary legacy of maize in South America. *Science* **362**, 1309–1313 (2018).
38. Smith, O. et al. A domestication history of dynamic adaptation and genomic deterioration in sorghum. *Nat. Plants* **5**, 369–379 (2019).
39. Lema, V. Non-domesticated cultivation in the Andes: plant management and nurturing in the Argentine northwest. *Veg. Hist. Archaeobot.* **24**, 143–150 (2015).
40. Oliszewski, N. & Babot, P. in *Avances y Desafíos Metodológicos en Arqueobotánica: Miradas Consensuadas y Diálogos Compartidos desde Sudamérica* (eds Belmar, C. & Lema, V.) 301–324 (Facultad de Patrimonio Cultural y Educaciòn Universidad SEK Chile, 2015).
41. Schubert, M. et al. Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nat. Prot.* **9**, 1056–1082 (2014).
42. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. & Orlando, L. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**, 1682–1684 (2013).
43. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: analysis of next generation sequencing data. *BMC Bioinform.* **15**, 356 (2014).
44. Vieira, F. G., Fumagalli, M., Albrechtsen, A. & Nielsen, R. Estimating inbreeding coefficients from NGS data: impact on genotype calling and allele frequency estimation. *Genome Res.* **23**, 1852–1861 (2013).
45. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
46. Lawson, D. J., Hellenthal, G., Myers, S. & Falush, D. Inference of population structure using dense haplotype data. *PLoS Genet.* **8**, e1002453 (2012).
47. Slotte, T. The impact of linked selection on plant genomic variation. *Brief. Funct. Genomics* **13**, 268–275 (2014).
48. Renaut, S. & Rieseberg, L. H. The accumulation of deleterious mutations as a consequence of domestication and improvement in sunflowers and other Compositae crops. *Mol. Biol. Evol.* **32**, 2273–2283 (2015).
49. Beissinger, T. M. et al. Recent demography drives changes in linked selection across the maize genome. *Nat. Plants* **2**, 16084 (2016).
50. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).

Acknowledgements

This study was supported by the University of Ferrara, the CUIA (Consorzio Universitario Italiano per l'Argentina, 5^ Research Program), the ERA-CAPS project Bean_Adapt, internal grants from the Marche Polytechnic University and the University of Firenze, the Italian Ministry of Education, University and Research (project Dipartimenti di Eccellenza 2018–2022, PRIN2017 grant no. 20174BTC4R), the Swedish Research Council grant no. VR-UF E0347601 and the Norwegian Research Council grant no. 262777. We also thank CONICET (Consejo Nacional de Investigaciones Cientificas y Técnicas) and the Institutions and Museums in Argentina for their support in the fieldwork, and in particular for recovering the archaeo-botanical specimens and studying the archaeological sites. M.D.L. sadly passed away during the preparation of this paper. One of the seeds we sequenced was kindly donated by the Archaeological Museum Pío Pablo Díaz in Cachi (Salta, Argentina), where M.D.L. was Director for many years.

Author contributions

G.B. and R.P. conceived the project. G.B., E.T. and A.B. designed the research. M.L., S.V., L.C. and S.B. performed the experiments. E.T., A.B., M.L., S.B., A.I. and C.X. analysed the data. G.B., E.T., A.B., M.L., F.R., B.S. and S.B. contributed to the manuscript preparation. V.L., P.B., N.O., A.G., G.N., C.T.M. and M.D.L. collected the samples. G.B., E.T., A.B., R.P., M.L., S.B. and B.S. interpreted the data. G.B. and E.T. wrote the manuscript. All authors revised and approved the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41477-021-00848-7.

Correspondence and requests for materials should be addressed to E.T. or G.B.

Peer review information *Nature Plants* thanks Robin Allaby, Kelly Swarts and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2021, corrected publication 2021

nature research

Corresponding author(s): Giorgio Bertorelle
Emiliano Trucchi

Last updated by author(s): 2020/12/15

Reporting Summary

Nature Research 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 Research policies, see Authors & Referees 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 (*n*) 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 *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ 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. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒ 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 *d*, Pearson's *r*), 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	Genomic DNA was sequenced using Illumina HiSeq2500 and HiSeq4000 instruments and the base calling was performed using their proprietary workflow.
Data analysis	Bioinformatic, population genetic and statistical analysis were performed using Paleomix v1.2.12, AdapterRemoval v2.2.2, bwa 0.7.15, samtools v1.9, Picard tools v2.18, MapDamage v2.0, GATK v3.5, Trimmomatic v0.36, bedtools v2.19, vcftools v0.1.15, angsd v0.916, ngsFHMM, plink v1.90, Admixture v1.3, FigTree1.4.4, FineStructure v2.1.1, David v6.8, R package ape v5.3, python v2.7 function qvalue.py (https://github.com/nfusi/qvalue), custom code in awk and phyton (https://github.com/anbena/ancient-beans).

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/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research 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 list of figures that have associated raw data
- A description of any restrictions on data availability

Sequenced raw reads from ancient samples are publicly available in NCBI database (https://www.ncbi.nlm.nih.gov/bioproject/) with Bioproject ID: PRJNA574560. Modern sample genomic data are publicly available (SRR10161640; SRR10161629; SRR10161767; SRR10161647; SRR10161646; SRR10161645; SRR10161601; SRR10161592; SRR10161584; SRR10161684; SRR10161683; SRR10161697; SRR10161690; SRR10161688; SRR10161651; SRR10161745; SRR10161723) and upon request from ERA-CAPS funded project “BeanAdapt: The Genomics of Adaptation during Crop Expansion in Bean”, coordinated by Roberto Papa, Marche Polytechnic University.

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	In this study, we investigated the pattern of genome-wide diversity and the intensity and temporal dynamic of selection during domestication in the common bean (Phaseolus vulgaris). We sequenced complete genomes from ancient seeds dated between 2500 and 650 years ago, and we analyzed them as compared to modern and wild accessions of the same species.
Research sample	Thirty ancient seeds found in nine archaeological sites in north and central western Argentina were whole-genome sequenced. Such sample is representative of the domesticated genomic diversity of common bean (Phaseolus vulgaris) cultivated in that area between 2500 and 650 years ago.
Sampling strategy	Sampling was designed on the basis of available ancient material collected in archeaelogical sites and available in Argentinian museums, in order to cover the selected area and to guarantee an adequate statistical robustness in our population genetics analyses.
Data collection	Speciments were collected from Museum collections. Giorgio Bertorelle and Roberto Papa selected the specimens with the help of the researchers at the Argentinean Museums (Veronica Lema, Pilar Babot, Nurit Oliszewski, Adolfo Gil, Gustavo Neme, Teresa Michieli, Monica De Lorenzi). Martina Lari took pictures of each seed, sampled each seed in a dedicated ancient DNA facility at the Molecular Anthropology and Paleogenetic Laboratory of the Department of Biology, University of Florence, sent the required amount for AMS radiocarbon dating to CEnter of Applied Physics, DAting e Diagnostics, Department of Mathematics and Physics "Ennio De Giorgi", University of Salento, Italy (headed by Lucio Calcagnile). Total DNA extraction and library preparation were performed at the Molecular Anthropology and Paleogenetic Laboratory of the Department of Biology, University of Florence, by Martina Lari and Stefania Vai. Whole-genome sequencing was performed at the sequencing facility of the University of Oslo under the supervision of Sanne Boessenkool.
Timing and spatial scale	Ancient seeds were obtained from different museum in Argentina in 2014. Ancient seeds were were previously collected during archaeological excavations in north and central western Argentina, and were radiocarbon dated in this study between 2500 and 650 years ago. Modern seeds used here for comparison are from seed banks and refer to different South-American areas and Mexico.
Data exclusions	One seed gave clearly inconsistent dating results, probably due to a mislabeling of the sample. It was not possible to reconstruct precisely the origin of the mislabeling, and the seed was therefore excluded from any further analysis.
Reproducibility	Ancient samples used in this study for genomic DNA extraction and genomic sequencing libraries preparation were partially processed at two different ancient genomics facilities at the University of Florence, Italy, and at the University of Oslo, Norway, to assess for reproducibility. All statistical analyses in the study were replicated at least three times by different operators starting from raw genomic DNA data.
Randomization	The study is based on observational data and not on an experimental setup, therefore randomization is not applicable
Blinding	The study is based on observational data and not on an experimental setup, therefore blinding is not applicable

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
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

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

Palaeontology

Specimen provenance	Ancient bean seeds come from nine archaeological sites in north and central western Argentina: Los Morrilos, Rio Salado, Cerro Calvario and Punta del Barro in the province of San Juan; Pampa Grande and Puente del Diablo in the province of Salta; Gruta del Indio in the province of Mendoza; Cueva de los Corrales in the province of Tucuman; Punta de la Pena in the province of Catamarca. Permits for the work were obtained in 2014 from the following museum Institutions: Museo de La Plata (La Plata, Buenos Aires, Argentina); Museo de Historia Natural de San Rafael (Parque Mariano Moreno, San Rafael, Argentina); Instituto de Investigaciones y Museo Arqueológico "Prof. Mariano Gambier", Universidad Nacional de San Juan (San Juan, Argentina); Instituto de Arqueología y Museo, Universidad Nacional de Tucumán (Tucumán, Argentina); Museo Arqueológico Pio Pablo Díaz (Cachi, Salta, Argentina)
Specimen deposition	Leftover tissues of the specimen (i.e., ancient common bean seeds) used in this study are stored at the ancient DNA facility of the Molecular Anthropology and Paleogenetic Laboratory, Department of Biology, University of Florence, Italy. Additional specimen from the same populations (i.e., biological replicates) are available from the museum institutions listed above.
Dating methods	AMS radiocarbon dating of the ancient bean seeds was performed at CEDAD, CEnter of Applied Physics, DAting e Diagnostics, Department of Mathematics and Physics "Ennio De Giorgi", University of Salento, Italy. Radiocarbon ages were calibrated for calendar years applying the software OxCal Ver. 3.10 based upon atmospheric data [Reimer PJ, et al. 2013 Radiocarbon 55 No. 4-1869-1887].

☒ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.