nature ecology & evolution

BRIEF COMMUNICATION

https://doi.org/10.1038/s41559-018-0643-y

A western Sahara centre of domestication inferred from pearl millet genomes

Concetta Burgarella[1,2,10*], Philippe Cubry[1,10*], Ndjido A. Kane[3,4], Rajeev K. Varshney[5], Cedric Mariac[1], Xin Liu[6], Chengcheng Shi[6], Mahendar Thudi[5], Marie Couderc[1], Xun Xu[6], Annapurna Chitikineni[5], Nora Scarcelli[1], Adeline Barnaud[1,4], Bénédicte Rhoné[1,7], Christian Dupuy[8], Olivier François[9*], Cécile Berthouly-Salazar[1,3*] and Yves Vigouroux[1*]

There have been intense debates over the geographic origin of African crops and agriculture. Here, we used whole-genome sequencing data to infer the domestication origin of pearl millet (*Cenchrus americanus*). Our results supported an origin in western Sahara, and we dated the onset of cultivated pearl millet expansion in Africa to 4,900 years ago. We provided evidence that wild-to-crop gene flow increased cultivated genetic diversity leading to diversity hotspots in western and eastern Sahel and adaptive introgression of 15 genomic regions. Our study reconciled genetic and archaeological data for one of the oldest African crops.

The shift from a nomadic way of life to settlement based on agriculture marked a turning point for human civilization. A crucial step in this change was the domestication of crops and animals. In the first half of the twentieth century, Vavilov[1] identified eight geographic regions with high varietal diversity and wild relative species richness. This observation led him to propose the existence of eight 'centres of origin', where the wild species underwent domestication[1]. Contradicting Vavilov's vision, Harlan[2,3] proposed a 'non-centric origin hypothesis' for African crops, in which the domestication process was distributed over the Sahel, a region up to 1,000 km wide that spans the 5,400 km from the Altlantic Ocean to the Red Sea. Today, these hypotheses about the origin of African crops and agriculture are still debated[4–7]. Pearl millet (*Cenchrus americanus* (L.) Morrone syn. *Pennisetum glaucum* (L.) R. Br.) is the oldest African cereal in the archeological records, and it is associated with the dawn of agriculture in West Africa[8,9]. Pearl millet is also the most nutritious low input staple cereal cultivated in the arid and low fertility soils of Africa and Asia. Its wild progenitor, *P. glaucum monodii* (L.) R.Br., has a natural distribution that spans the Sahel region. The domestication origin of pearl millet has been hypothesized to occur in Senegal[5], the eastern Sahel[6]or even further north in the Sahara[7]. Recent sequencing of pearl millet genomes[4] unlocks the possibility to assess the origin of pearl millet domestication.

We analysed whole-genome sequences of 221 accessions of wild forms and traditional varieties representative of the geographical diversity of pearl millet[4] (Supplementary Table 1; Supplementary Fig. 1). Genetic diversity of wild pearl millet is structured in three major geographic groups[10,11]: a western group including individuals from Senegal and Mauritania, a central group including individuals from Mali and Niger and an eastern group including individuals from Chad and Sudan (Supplementary Fig. 2). Cultivated pearl millet was highly differentiated from the wild groups, but the traditional varieties did not show strong genetic stratification[4] (Supplementary Fig. 4). In further analyses, we split cultivated pearl millet into five geographic groups: western Sahel, central Sahel, eastern Sahel, southern Africa and India (Supplementary Table 1). Using coalescent models[12], we found support for a common origin for all cultivated groups. Each group, from West Africa to India, originated from a single wild population from the central Sahel (Supplementary Tables 2,3; Supplementary Figs. 5,6). More complex scenarios, including a domestication bottleneck, population growth and wild-crop gene flow, also pointed to a unique origin from the central Sahel (Supplementary Table 3).

Assuming that the central Sahel is the origin of domestication, this location should correspond to the area with the greatest genetic diversity[1]. Serial founder events during crop diffusion are expected to decrease cultivated genetic diversity from the centre of origin to the edge of diffusion. Investigating the geographic distribution of rare alleles[13], we observed that pearl millet hotspots of diversity were located in the western and eastern Sahel, outside the central region (Fig. 1). These two regions correspond to "the most conspicuous areas of interactions among cultivated and wild and weedy races"[3]. Our explanation for those observed patterns is that they arose after secondary contact with local wild relatives. To support our argument, we evaluated correlations between cultivated diversity and genetic proximity with wild populations. We found a weak but significant correlation for all genetic measures tested (Fig. 1b,c; Supplementary Figs. 7–9). To confirm gene flow and rule out the confounding effect of shared ancestral polymorphisms, we used *f*-statistics[14]. We found clear signatures of gene flow with sympatric wild populations (Supplementary Table 4) for the cultivated group in both the western Sahel ($f_4 = -0.0032$, $Z = -39.0$, $P < 10^{-10}$) and the eastern Sahel ($f_4 = -0.0066$, $Z = -87.3$, $P < 10^{-10}$). Next, we used TreeMix[15] to explore the genealogy of pearl millet populations. Model fit was equal to 99.7% when gene flow between sympatric wild and cultivated populations from the eastern and western Sahel

[1]IRD, Université Montpellier, DIADE, Montpellier, France. [2]AGAP, Université Montpellier, CIRAD, INRA, Montpellier SupAgro, Montpellier, France. [3]Senegalese Agricultural Research Institute, Dakar, Senegal. [4]Laboratoire Mixte International LAPSE, Dakar, Senegal. [5]Center of Excellence in Genomics and Systems Biology, International Crops Research Institute for the Semi-Arid Tropics, Hyderabad, Telangana, India. [6]BGI-Shenzhen, Shenzhen, China. [7]Université Lyon 1, CNRS, Laboratoire de Biométrie et Biologie Evolutive, Lyon, France. [8]Institut des Mondes Africains, Paris, France. [9]Université Grenoble-Alpes, Centre National de la Recherche Scientifique, Grenoble, France. [10]These authors contributed equally to this work: Concetta Burgarella, Philippe Cubry. *e-mail: concetta.burgarella@gmail.com; philippe.cubry@hotmail.fr; olivier.francois@imag.fr; cecile.berthouly@ird.fr; yves.vigouroux@ird.fr

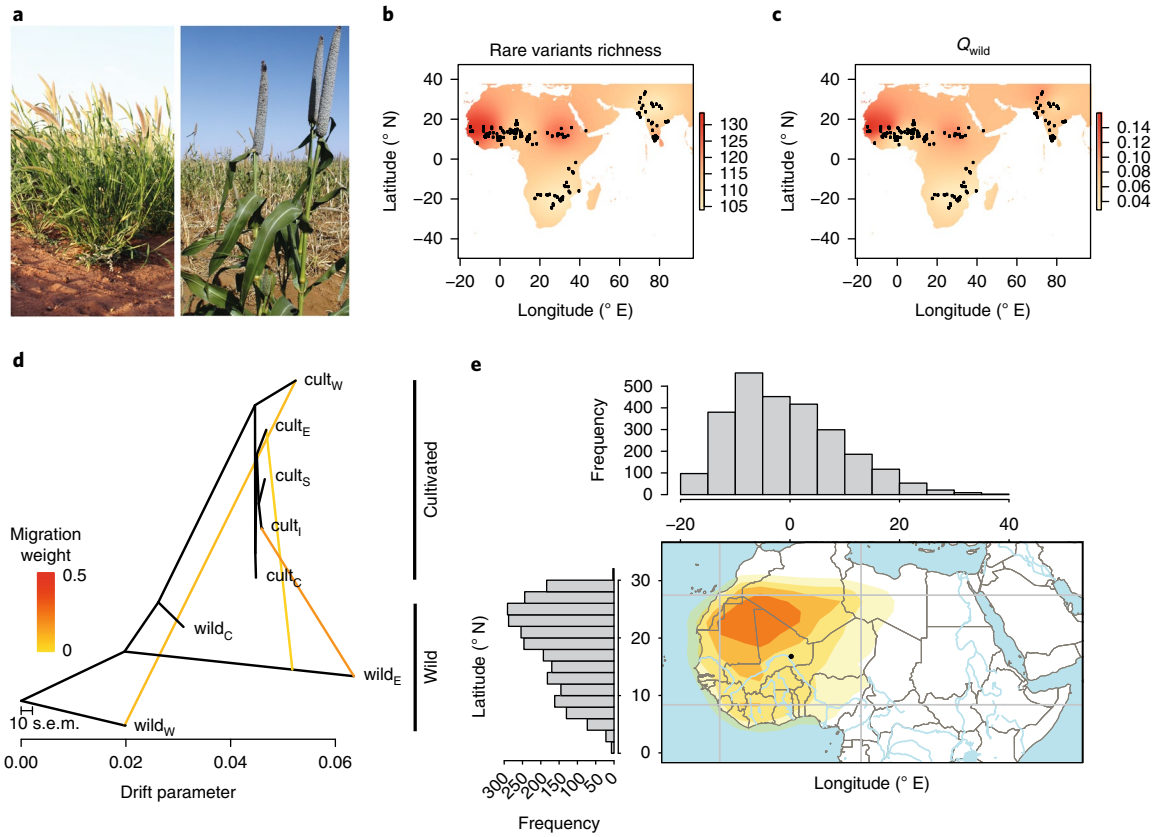

Fig. 1 | Genetic diversity and origin of pearl millet domestication. a, Wild (left) and cultivated (right) plants of pearl millet. **b**, Genetic diversity of cultivated pearl millet accessions. The diversity is assessed through the rare variants statistic, that is, we computed the individual number of singletons (variants present in only one individual) and used a kriging method to spatially represent the distribution of this statistic[13]. **c**, Geographic representation of wild ancestry (Q_{wild}) in the cultivated accessions obtained from sNMF analysis. **d**, TreeMix analysis of the relationships between cultivated geographic groups and wild genetic groups. The groups $cult_W$, $cult_C$, $cult_E$, $cult_S$ and $cult_I$ correspond to cultivated varieties accessions from the western Sahel, central Sahel, eastern Sahel, southern Africa and India, respectively. The groups $wild_W$, $wild_C$ and $wild_E$ correspond to wild populations found in the western, central and eastern Sahel, respectively. The population tree was built assuming three migration events. This analysis pinpointed gene flow between sympatric wild and cultivated populations in the eastern and western Sahel. **e**, Inferred geographical origin of pearl millet domestication. Posterior predictions for latitude gave a mode of 23.57° N (95% CI 5.34–26.17) and for longitude gave a mode of −6.61° E (95% CI −15.58 to 21.63). Black dots show the location of the most ancient archaeological remains known to date.

was included (Fig. 1d; Supplementary Fig. 10). The wild populations from the central Sahel were the closest populations to all cultivated groups (Fig. 1d). Put together, our results provided evidence that worldwide cultivated pearl millet varieties were derived from a common ancestor of wild populations found today in the central Sahel. During its agricultural diffusion, cultivated pearl millet was introgressed by wild relatives in the western and eastern Sahel, leading to hotspots of diversity outside the centre of origin.

Next, we developed a spatially explicit model including wild-to-crop gene flow in order to estimate the geographic origin and the timing of the expansion of pearl millet cultivation (Supplementary Table 5; Supplementary Fig. 11). The spatial model identified an origin at latitude higher than the current range of wild populations in the central Sahel (Fig. 1e). The identified region corresponded to the Taoudeni Basin in the western Sahara (−6.61° E, 23.58° N), matching the wetter climate that prevailed in the Sahara 6,000 years BP. During that period, the Sahara hosted a rich Poaceae community and it was characterized by a widespread system of lakes[16]. Drying of the Sahara led the plant communities to move south to their current distribution about 3,200 years ago[16]. Thus a northern distribution of the ancestors of current wild central Sahel populations was expected. Using our spatial model and time calibration based on archaeological remains (Supplementary Table 6; Supplementary Figs. 12,13), we dated the onset of diffusion of pearl millet agriculture around 4,892 years ago (95% CI 3,685–5,889). Our estimate corresponded to the intensification of drying in the Sahara[16]. The estimate might predate the date of domestication, as the timing of diffusion of a fully domesticated plant differs from the duration of the domestication process per se. The observed genetic patterns fit well with the archaeological remains of pearl millet seeds and pottery found in the northern inner Niger delta region in approximately the same time frame[8,9].

Emerging evidence for a key evolutionary role of adaptive introgression is increasing in humans[17], animals[18] and plants[19]. In crops, wild populations have been shown to be a reservoir of adaptation[20]. Based on the assumption that the diversity of cultivated pearl millet was shaped by differential gene flow from sympatric wild populations, we questioned whether wild introgression could have facilitated the local adaptation of cultivated varieties. To test the hypothesis of adaptive introgression, we first performed a whole-genome scan for selection in each cultivated group using a composite likelihood ratio test approach (CLR)[21]. We identified 254 selective sweeps across all 5 geographical cultivated groups (Fig. 2a,b; Supplementary Table 7; Supplementary Fig. 14). Gene ontology annotation of the 254 regions revealed an enrichment of genes involved in pollen–pistil interaction and fundamental metabolic

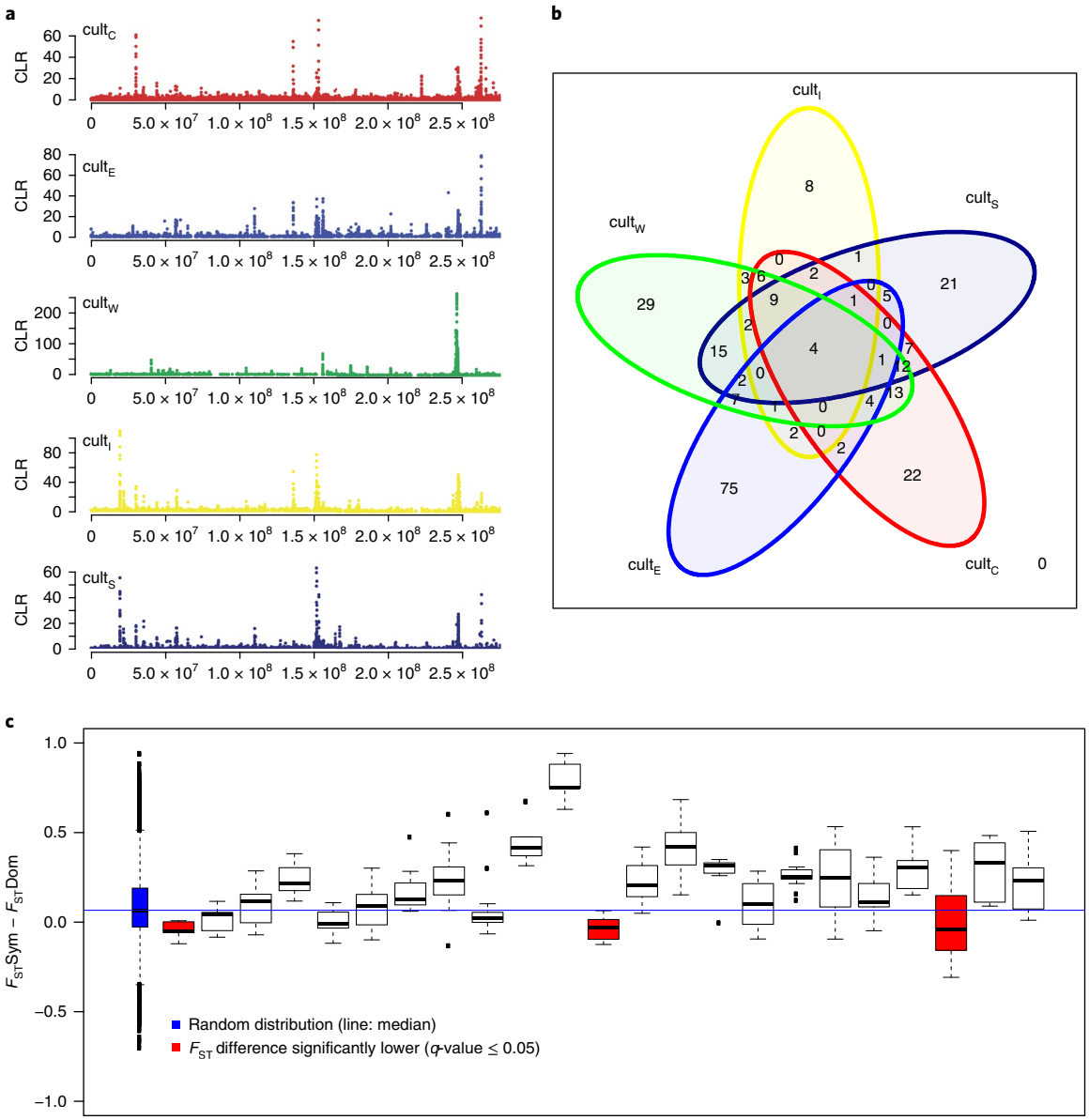

Fig. 2 | Selection and introgression in pearl millet. a, CLR of selection for each cultivated group from western ($cult_W$), central ($cult_C$) and eastern Sahel ($cult_E$), southern Africa ($cult_S$) and India ($cult_I$). The analysis here represents the CLR value calculated every 3 kb. **b**, Venn diagram of selected regions shared by the five cultivated groups. Four regions were common to all the cultivated groups. We found 75 regions specific to the western group, 29 to the central group, 8 to the eastern group, 22 to the southern African group and 21 to India. **c**, We investigated if these specific selected regions (white boxplots) showed introgression from local wild relatives in the eastern and in the western cultivated groups. These introgressions are identified by a lower differentiation of cultivated populations from their respective wild sympatric populations than from the wild population from the central Sahel (F_{ST}Sym – F_{ST}Dom). Genome wide expectation was calculated on 1,000 random fragments (blue boxplot). Red boxplots indicate selected regions with significantly lower relative differentiation than expected by chance (at 5% false discovery rate; *q*-value). In boxplots, the centre line corresponds to the median, the box limits correspond to upper and lower quartiles, whiskers correspond to 1.5 × inter-quartile range, outliers are illustrated with points. The results presented in **a** and **c** refer to chromosome 1.

processes such as lipid and cellulose synthesis (Supplementary Table 8). Four signatures of selection were shared by all cultivated groups and may pinpoint early selection, which is potentially associated with the domestication process. We also found signals of selection specific to each cultivated group. Focusing on local selective sweeps in the western Sahel (75 genomic regions) and the eastern Sahel (8 genomic regions), we asked whether the selective sweeps could harbour signatures of introgression from their respective wild sympatric populations. Assuming that adaptive introgression occurred in the western and eastern Sahel, we predicted that cultivated populations were less differentiated from their respective wild sympatric populations than from the wild population from the central Sahel in the selected genomic regions (Fig. 2c; Supplementary Fig. 15). A total of ten and five selection signatures fitted this pattern for two differentiation statistics in the western Sahel and eastern Sahel respectively (Supplementary Table 9; Supplementary Figs. 16,17). Among these 15 genomic regions, 2 were related to panicle number[4]. Our findings stressed the importance of wild-to-crop gene flow during and after crop domestication.

In conclusion, our results support a Saharan cradle of pearl millet domestication. This western Saharan origin fitted recent archaeological hypotheses[7,22] and reconciled them with genetic studies.

Methods

We used whole genome data for 221 geo-referenced cultivated and wild accessions of pearl millet available from ref. [4] (Supplementary Table 1). We assessed the genetic structure of our sample using two clustering approaches: sNMF v1.2[23], implemented in the package LEA[24] under the R environment[25], and TESS3[26], which makes explicit use of spatial information.

To investigate the history of domestication of pearl millet, we tested which of the three wild groups is genetically closest to each of the five cultivated groups with the model-based inference framework implemented in Fastsimcoal v2.5.2.21[12,27]. To assess and test the occurrence of wild-to-crop gene flow, we used two approaches, Treemix[15], and f_3 or f_4 statistics[14] (Treemix v1.12).

We implemented spatially explicit simulations using the SPLATCHE2 software[28] to infer the geographic origin of pearl millet domestication in Africa. We extended a previous demographic model[13] by adding the wild-to-crop gene flow. We defined three parameters ω, γ and ε, which reflect the intensity of gene flow between sympatric wild and cultivated populations in western, central and eastern Sahel, respectively. The posterior distributions of the demographic parameters were inferred by using an approximate Bayesian approach[29] with a neural network algorithm and a tolerance rate of 0.5% (ref. [30]).

We used the oldest known archeological remains of cultivated pearl millet to estimate a date for the dispersion of pearl millet agriculture (Supplementary Table 6, Supplementary Fig. 12). We performed 5,000 simulations by sampling each parameter of the spatial model from their posterior distribution. The arrival date at a specific location in the spatial model was then fit with the observed archeological dates. From these 5,000 regressions, we estimated the density distribution of the beginning of the expansion of pearl millet agriculture.

To detect signatures of positive selection in each cultivated group, we used the CLR test[31] implemented in the SweeD software[32]. Positions within the first 0.1 percentile of CLR values were identified as candidates for selection. Candidate positions less than 100 kb apart were considered as a single selective sweep and were combined. To assess whether selective sweeps co-localized with wild-to-crop genomic introgression, we evaluated the level of differentiation between cultivated and wild groups with the F_{ST} statistic[33,34] and a measure of absolute divergence, d_{xy} (refs. [35,36]). We calculated the difference F_{ST}Sym – F_{ST}Dom, where F_{ST}Sym is the differentiation between the sympatric cultivated and wild populations and F_{ST}Dom is the differentiation between the cultivated population and the wild population from central Sahel (Supplementary Figure 15). We tested whether F_{ST} difference was significantly more negative in selective sweeps than expected by chance by randomly resampling 1,000 regions of similar length from the genome. Significance was assessed with a Wilcoxon–Mann–Whitney test and a false discovery rate approach to correct for multiple testing (R package qvalue[37]).

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

Data availability. Pearl millet genome data are available from ref. [4] and genomic sequences from *Pennisetum pedicellatum* can be accessed under SRA accessions SAMN09499320 and SAMN09499321. Customized R-scripts are available at https://github.com/Africrop/pearl_millet.

Received: 10 May 2018; Accepted: 12 July 2018;
Published online: 06 August 2018

References

1. Vavilov, N. I. & Dorofeev, V. F. *Origin and Geography of Cultivated Plants* (Cambridge Univ. Press, Cambridge, 1992).
2. Harlan, J. R., De Wet, J. M. J. & Stemler, A. B. L. *Origins of African Plant Domestication* (De Gruyter, Berlin, 1976).
3. Harlan, J. R. *Science* **174**, 468–474 (1971).
4. Varshney, R. K. et al. *Nat. Biotechnol.* **35**, 969–976 (2017).
5. Hu, Z. et al. *BMC Genom.* **16**, 1048 (2015).
6. Dussert, Y., Snirc, A. & Robert, T. *Mol. Ecol.* **24**, 1387–1402 (2015).
7. Manning, K. in *West African Archaeology: New Developments, New Perspectives* (ed. Allsworth-Jones, P.) 43–52 (Archaeopress, Oxford, 2010).
8. Manning, K., Pelling, R., Higham, T., Schwenniger, J.-L. & Fuller, D. Q. *J. Archaeol. Sci.* **38**, 312–322 (2011).
9. Ozainne, S. et al. *J. Archaeol. Sci.* **50**, 359–368 (2014).
10. Oumar, I., Mariac, C., Pham, J.-L. & Vigouroux, Y. *Theor. Appl. Genet.* **117**, 489–497 (2008).
11. Berthouly-Salazar, C. et al. *Front. Plant Sci.* **7**, 777 (2016).
12. Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C. & Foll, M. *PLoS Genet.* **9**, e1003905 (2013).
13. Cubry, P., Vigouroux, Y. & François, O. *Front. Genet.* **8**, 139 (2017).
14. Reich, D., Thangaraj, K., Patterson, N., Price, A. L. & Singh, L. *Nature* **461**, 489–494 (2009).
15. Pickrell, J. K. & Pritchard, J. K. *PLOS Genet.* **8**, e1002967 (2012).
16. Kröpelin, S. et al. *Science* **320**, 765–768 (2008).
17. Huerta-Sánchez, E. et al. *Nature* **512**, 194–197 (2014).
18. Ai, H. et al. *Nat. Genet.* **47**, 217–225 (2015).
19. Arnold, B. J. et al. *Proc. Natl Acad. Sci. USA* **113**, 8320–8325 (2016).
20. Hufford, M. B. et al. *PLoS Genet.* **9**, e1003477 (2013).
21. Nielsen, R. et al. *Genome Res.* **15**, 1566–1575 (2005).
22. Dupuy, C. *Encycl. Berbère* **XXXIX**, 6529–6544 (2015).
23. Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G. & François, O. *Genetics* **196**, 973–983 (2014).
24. Frichot, E. & François, O. *Methods Ecol. Evol.* **6**, 925–929 (2015).
25. R Core Team *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, 2016).
26. Caye, K., Deist, T. M., Martins, H., Michel, O. & François, O. *Mol. Ecol. Resour.* **16**, 540–548 (2016).
27. Excoffier, L. & Foll, M. *Bioinformatics* **27**, 1332–1334 (2011).
28. Ray, N., Currat, M., Foll, M. & Excoffier, L. *Bioinformatics* **26**, 2993–2994 (2010).
29. Csilléry, K., Blum, M. G. B., Gaggiotti, O. E. & François, O. *Trends Ecol. Evol.* **25**, 410–418 (2010).
30. Csilléry, K., François, O. & Blum, M. G. B. *Methods Ecol. Evol.* **3**, 475–479 (2012).
31. Nielsen, R. et al. *PLoS Biol.* **3**, e170 (2005).
32. Pavlidis, P., Živković, D., Stamatakis, A. & Alachiotis, N. *Mol. Biol. Evol.* **30**, 2224–2234 (2013).
33. Weir, B. S. *Genetic Data Analysis II: Methods for Discrete Population Genetic Data* (Oxford Univ. Press, Oxford, 1996).
34. Weir, B. S. & Cockerham, C. C. *Evolution* **38**, 1358–1370 (1984).
35. Nei, M. & Li, W. H. *Proc. Natl Acad. Sci. USA* **76**, 5269–5273 (1979).
36. Cruickshank, T. E. & Hahn, M. W. *Mol. Ecol.* **23**, 3133–3157 (2014).
37. Bass, J. D., Dabney, A. & Robinson, D. *qvalue: Q-value Estimation for False Discovery Rate Control* R package version 2.12.0 (2015).

Acknowledgements

We are grateful to the Genotoul sequencing platform Toulouse Midi-Pyrenees for help during sequencing. We also want to thank N. Tando and the IRD itrop Plantes Santé bioinformatic platform for providing HPC resources and support for our research project. Y.V. acknowledges support from the Agence Nationale de la Recherche (ANR-13-BSV7-0017). C.B.-S. acknowledges support from Agropolis Fondation under the reference ID 1403-057 through the Investissements d'avenir programme (Labex Agro: ANR-10-LABX-0001-01).

Author contributions

C.B., P.C., N.A.K., A.B., O.F., C.B.-S. and Y.V. designed the analysis. C.B. and P.C. performed statistical analyses. C.M. and M.C. performed additional experimental work. B.R., N.S., C.D., M.T., C.S. and O.F. contributed to analytic tools, data, methods and participated in data analysis; X.L., X.X., R.K.V. and Y.V. managed and designed the pearl millet genome project. O.F., C.B.-S. and Y.V. managed this genomic diversity study. C.B., P.C., C.B.-S. and Y.V. wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41559-018-0643-y.

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

Correspondence and requests for materials should be addressed to C.B. or P.C. or O.F. or C.B. or Y.V.

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

nature research

Corresponding author(s): Yves Vigouroux

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.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. 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
☐	☒	An indication of 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 statistics 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
☒	☐	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	no software was used
Data analysis	SNMF v.1.2, TESS3 (genetic clustering analysis), Fastsimcoal v2.5.2.21, SPLATCHE2 (coalescent simulations), customized R-scripts (including spatial model, ABC simulaiton, fastsimcoal simulation and site frequency spectrum generation; https://github.com/Africrop/pearl_millet), DAPC implemented in adegenet 2.1.1 (assignment analysis), Treemix v1.12 (genealogical tree with migration events; statistically testing for gene flow evidence), SweeD (detection of selection), TopGo (enrichment analysis on gene ontology), vcftools and Popgenome R package (for differenciation statistics)

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 upon request. 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

Pearl millet genome data are available from (4:Nature Biotechnology volume 35, pages 969–976 (2017) doi:10.1038/nbt.3943) and genomic sequences from Pennisetum pedicellatum can be access under SRA accession: SAMN09499320, SAMN09499321

Field-specific reporting

Please select 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/authors/policies/ReportingSummary-flat.pdf

Ecological, evolutionary & environmental sciences study design

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

Study description	Population genetics study on real and simulated data
Research sample	Wild and cultivated accessions of the cereal pearl millet (Cenchrus americanus (L.) Morrone syn. Pennisetum glaucum (L.) R.Br.) from Africa and India
Sampling strategy	Samples include wild accessions (31) and cultivated varieties (190) i.e. local landraces (no breeding material) that are representative of the geographical diversity of pearl millet. We used all geo-references cultivated traditionnal landraces available. The focus of the study is the domestication history of cultivated pearl millet, for which the relative sample sizes of wild and cultivated accessions are adequate
Data collection	Samples are maintained at the ICRISAT genebank (Hyderabad, Telangana State, India) and at The IRD genebank (Montpellier, France)
Timing and spatial scale	Information about data collection can be obtain from Table S40 and S42 from Nature Biotechnology volume 35, pages 969–976 (2017) doi:10.1038/nbt.3943
Data exclusions	No exclusion was done priors to the study. If some samples were excluded in a particular analysis it is clearly state in the text, which one and why.
Reproducibility	Reproducibility was not assessed as it would imply to repeat the sampling and whole genome sequencing
Randomization	Samples were not randomized. They were classed as "wild" or "cultivated" based on information associated to each accession in seed banks (passeport data). For different analyses, grouping based on genetic homogeneity or geographical position was used.
Blinding	Blinding was applied in clustering analysis, when samples were assigned to genetic groups withouth taking into account their wild or cultivated origin

Did the study involve field work? ☐ Yes ☒ No

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants

Methods

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