HAL
open science

Yam genomics supports West Africa as a major cradle of crop domestication

Nora Scarcelli, Philippe Cubry, Roland Akakpo, Anne-Celine Thuillet, Jude Obidiegwu, Mohamed N. Baco, Emmanuel Otoo, Bonaventure Sonké, Alexandre Dansi, Gustave Djedatin, et al.

AGRICULTURE

Yam genomics supports West Africa as a major cradle of crop domestication

Nora Scarcelli[1]*, Philippe Cubry[1], Roland Akakpo[1], Anne-Céline Thuillet[1], Jude Obidiegwu[2],
Mohamed N. Baco[3], Emmanuel Otoo[4], Bonaventure Sonké[5], Alexandre Dansi[6],
Gustave Djedatin[6], Cédric Mariac[1], Marie Couderc[1], Sandrine Causse[7,8], Karine Alix[9],
Hâna Chaïr[7,8], Olivier François[10], Yves Vigouroux[1]

While there has been progress in our understanding of the origin and history of agriculture in sub-Saharan Africa, a unified perspective is still lacking on where and how major crops were domesticated in the region. Here, we investigated the domestication of African yam (*Dioscorea rotundata*), a key crop in early African agriculture. Using whole-genome resequencing and statistical models, we show that cultivated yam was domesticated from a forest species. We infer that the expansion of African yam agriculture started in the Niger River basin. This result, alongside with the origins of African rice and pearl millet, supports the hypothesis that the vicinity of the Niger River was a major cradle of African agriculture.

INTRODUCTION

The emergence of agricultural societies was associated with hotspots of plant domestication (*1*), often described as domestication centers (*2*). One of the best known hotspots is the Fertile Crescent in the Middle East, where wheat, barley, oat, lentil, and chickpea, among others, first appeared in the archaeological records (*3*). The history of crop domestication is much less documented in sub-Saharan Africa, probably because archaeological studies are largely fragmentary (*4*). One hypothesis about crop domestication in Africa suggests an origin encompassing a large area from Senegal to Somalia (*2*). This Sahel-wide hypothesis was mainly based on distributions of wild and cultivated African cereals, such as pearl millet (*Cenchrus americanus*), sorghum (*Sorghum bicolor*), fonio (*Digitaria exilis*), and African rice (*Oryza glaberrima*). Recent studies have challenged this hypothesis and proposed a more restricted area of origin in the western Sahel, near the Niger River basin. Pearl millet was domesticated in a region corresponding today to northern Mali and Mauritania (*5*), and African rice was also domesticated in Mali (*6*). To assess whether the vicinity of the Niger River basin could be identified as a major hotspot of domestication, we investigated the domestication of yam, another major staple crop originating from Africa.

Yams (*Dioscorea* spp.) were domesticated independently at least three times in three different continents: in Asia (*Dioscorea alata*), in America (*Dioscorea trifida*), and in Africa (*Dioscorea rotundata*) (*2*). In Africa, yam starchy tubers are mainly produced in the "yam belt," a region including the Republic of Côte d'Ivoire, Ghana, Togo, Benin, Nigeria, and Cameroon (*7*), which accounts for 97% of African yam production (www.fao.org/faostat). Yam production in West Africa is second only to that of cassava, surpassing that of maize, rice, and sorghum. It is therefore a key crop for African food security. The main cultivated yam, *D. rotundata*, has two close wild relatives (*8*), the savannah species *Dioscorea abyssinica* and the forest species *Dioscorea praehensilis*. Domesticated yam is likely derived from one of these two species (*9*) or from hybridization between them (*7*).

In this study, we used whole-genome resequencing of 167 wild and cultivated yams to clarify where and from which species yam was domesticated. Our findings in combination with recent results on pearl millet and African rice (*5*, *6*) suggest that the Niger River vicinity played a major role in the domestication of African crops.

RESULTS

Genetic diversity and structure

We sampled wild and cultivated yams in the main area of production, in Ghana, Benin, Nigeria, and Cameroon (Fig. 1 and table S1). A total of 86 *D. rotundata*, 34 *D. abyssinica*, and 47 *D. praehensilis* were fully resequenced an average of seven times and produced 3,570,940 single-nucleotide polymorphisms (SNPs).

Principal components analysis (PCA) highlighted four genetic groups. The wild forest species, *D. praehensilis*, was split into two major groups, one found in West Africa and the other found in Cameroon (Fig. 2A and fig. S1A). The third group corresponded to *D. abyssinica* and the fourth to *D. rotundata*, with all cultivated individuals (Fig. 2A and fig. S1A). Analysis of population structure showed these same four groups (Fig. 2B). The three wild genetic groups showed clear geographic patterns of ancestry (Fig. 2C), corresponding to the savannah (*D. abyssinica*) and to the upper and lower Guinean forests for Western and Cameroonian *D. praehensilis*, respectively.

The population of *D. praehensilis* from Cameroonian forests had the highest nucleotide diversity (11.9×10^{-4}; table S2). Nucleotide diversities for the three other groups were similar (mean, 9.7×10^{-4}), with those for Western *D. praehensilis* slightly lower than the others (8.6×10^{-4}; table S2). Genetic diversity of *D. rotundata*, estimated as the number of singletons (*10*), was reduced by two-third compared to *D. abyssinica* and Cameroonian *D. praehensilis* but only by half compared to Western *D. praehensilis* (table S1). For the

[1]DIADE, Univ Montpellier, IRD, Montpellier, France. [2]National Root Crops Research Institute, Umudike, PMB 7006, Umuahia, Abia State, Nigeria. [3]University of Parakou, BP 123 Parakou, Bénin. [4]CSIR-Crops Research Institute, P.O. Box 3785, Fumesua-Kumasi, Ghana. [5]University of Yaoundé I, Laboratory of Plant Systematics and Ecology, P.O. Box 047, Yaoundé, Cameroon. [6]National University of Sciences, Technologies, Engineering and Mathematics of Abomey, Laboratory BIORAVE, Dassa-Zoumè, Benin. [7]Cirad UMR AGAP, F-34398 Montpellier, France. [8]University Montpellier, CIRAD, INRA, Montpellier SupAgro, Montpellier, France. [9]GQE–Le Moulon, INRA, Univ Paris-Sud, CNRS, AgroParisTech, Université Paris-Saclay, 91190 Gif-sur-Yvette, France. [10]University Grenoble-Alpes, Grenoble INP, TIMC-IMAG CNRS UMR 5525, 38042 Grenoble Cedex, France.
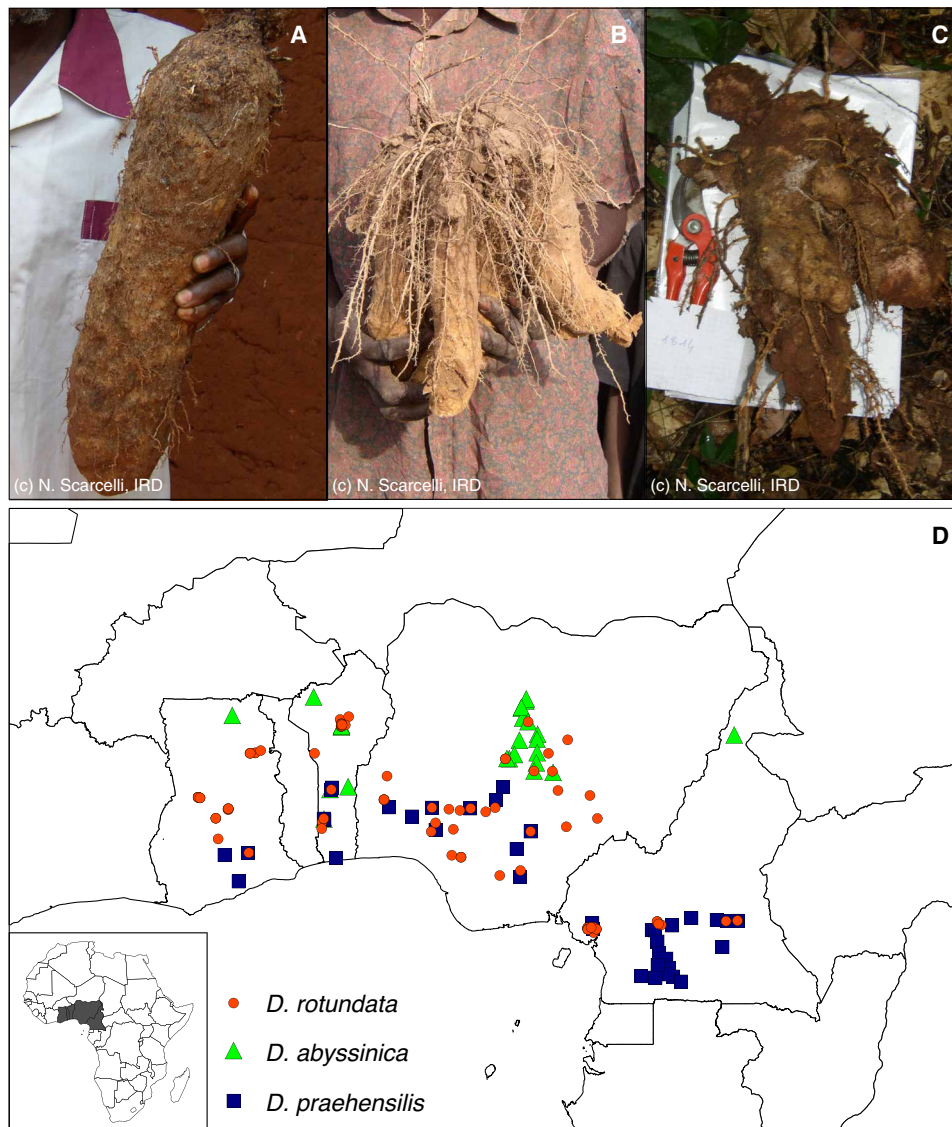*Corresponding author. Email: nora.scarcelli@ird.fr

Fig. 1. The three yam species analyzed in this study and corresponding sampling. (**A**) Tuber of *D. rotundata*. (**B**) Tuber of *D. abyssinic*a. (**C**) Tuber of *D. praehensilis*. (**D**) Map representing the geographical coordinates of each analyzed yam individual. Orange circle, *D. rotundata*; green triangle, *D. abyssinica*; blue square, *D. praehensilis*. Photo credits: Nora Scarcelli, IRD.

cultivated population, the maximum diversity was found in Ghana and Nigeria (fig. S1B). The linkage disequilibrium (LD) decay was smaller for populations of cultivated yam (ca. 0.16 at 250 kb) than for those of the wild species (ca. 0.12 at 250 kb fig. S2).

Origin of cultivated yam

Using a coalescent-based evolutionary approach (*11*), we modeled and tested the origin of cultivated yam from the three wild genetic groups identified. First, we investigated the relationships between the three wild populations: *D. abyssinica*, Cameroonian *D. praehensilis*, and Western *D. praehensilis*. In the most likely model, *D. abyssinica* diverged first, followed by the divergence of the two *D. praehensilis* populations (fig. S3A). Using this model, we tested a total of six divergence scenarios, assuming that the cultivated population derived from (i) only one of the three wild populations and (ii) from early hybridization of two of the three wild populations. The model with

the highest likelihood suggested that the cultivated population diverged from Western *D. praehensilis* (Fig. 3A and fig. S3B).

We then assessed the geographic origin of the expansion of cultivated yam by developing an approximate Bayesian computation (ABC) approach (*12*) with a spatially explicit diffusion model (*13*). The analysis pinpointed a region located between the eastern Ghana and western Nigeria and ranging from the Gulf of Guinea to the southern Niger (Fig. 3B). The best posterior estimate of location of origin was in northern Benin (mean, latitude 11.1° ± 4.4°N and longitude 2.8° ± 8.9°E). Posterior predictive checks showed that the predictions of the model were consistent with the empirical data (fig. S4).

Temporal dynamics of effective population size

Modeling the temporal dynamics of effective population size (*14*) indicated that both cultivated and wild populations strongly declined approximately 15,000 generations ago (Fig. 3C and fig. S5).
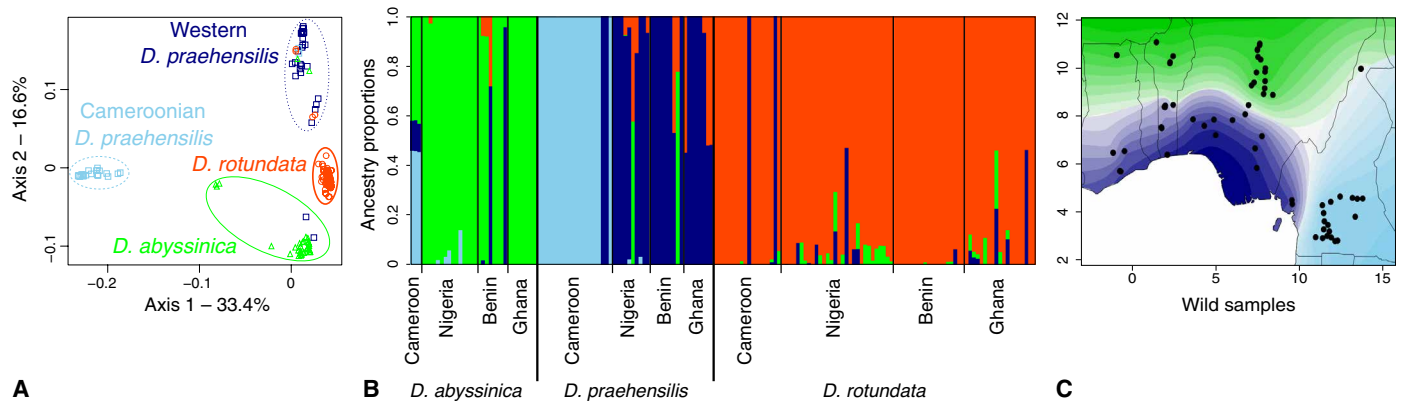
Fig. 2. Genetic structure of the three species. (**A**) PCA based on SNPs remaining after a 5% minimum allele frequency (MAF) cutoff. (**B**) Ancestry proportions of each sample estimated by admixture for $K = 4$ genetic groups. (**C**) Geographic distribution of ancestry proportions of wild population samples obtained from the same analysis. Orange, *D. rotundata*; green, *D. abyssinica*; dark blue, Western *D. praehensilis*; light blue, Cameroonian *D. praehensilis*.
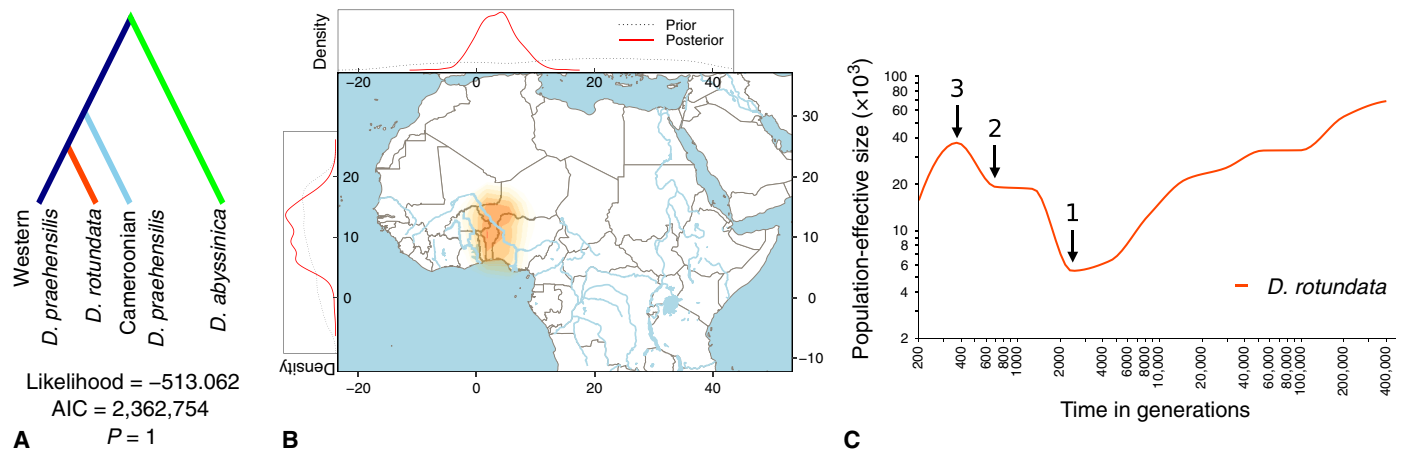

Fig. 3. Inference of African yam domestication history. (**A**) Yam species relationship. The best model inferred by coalescent-based analysis is presented here. AIC, Akaike information criterion. (**B**) Inferred area of geographic origin of cultivated yam based on an approximate Bayesian spatial model. (**C**) Demographic history of cultivated yam populations (effective size, *Ne*). The three arrows represent the following: (i) the first expansion of yam agriculture, ca. 2500 generations ago; (ii) the second expansion, ca. 700 generations ago; and (iii) the recent decrease, ca. 400 generations ago. Orange, *D. rotundata*; green, *D. abyssinica*; dark blue, Western *D. praehensilis*; light blue, Cameroonian *D. praehensilis*.

The cultivated population reached a minimum ca. 2500 generations ago. During this period, the cultivated population size was divided by four compared to the ancestral population size. The cultivated population size then increased to its maximum ca. 400 generations ago in two phases. During the first phase, a strong increase occurred between ca. 2500 and ca. 1500 generations ago, and the cultivated population size was increased fourfold. In the second phase, between ca. 700 and ca. 400 generations ago, the cultivated population size was then doubled. We observed a strong decline in the last 400 generations, during which the cultivated population size was more than halved. Western *D. praehensilis* and cultivated populations showed similar dynamics in the last ca. 3000 generations.

Detection of selection

We looked for genomic signatures of selection by comparing the nucleotide diversity π ratio (*15*) and the F_{ST} (*16*) of cultivated population and its wild relative, the Western *D. praehensilis* population. We also investigated complete selective sweeps in the genomes of the cultivated population (*17*). Overall, 294 annotated genes were

retrieved in regions detected by at least one test (table S3 and fig. S6). We retrieved six candidate genes previously reported in an analysis based only on Benin populations (*18*). Two of these genes are involved in root development [Scarecrow-like (*SCL*) gene Dr11126 and argininosuccinate lyase (*ASL*) gene Dr04385], one is involved in starch formation and storage [sucrose synthase (*SUS*) gene Dr18284], and three genes are involved in stomata regulation and osmotic stress [cellulose synthase–like (*CSL*) genes Dr13651, Dr13652, and Dr13653]. In addition, enrichment tests for gene ontology (GO) terms reveal a substantial number of genes in our dataset involved in the cellulose biosynthetic pathway and in auxin transport (table S4).

DISCUSSION

Before our study, two species were hypothesized to be the progenitors of cultivated yam: the savannah species *D. abyssinica* and the forest species *D. praehensilis* (*8, 9, 19*). Our genomic study provides statistically supported evidence that the forest species *D. praehensilis* is the most likely progenitor of African cultivated yam. As yam is

now cultivated in open fields, in contrast to the forest habitat of its progenitor, we expect that domestication triggered adaptations to the open environment (*18*). Accordingly, signatures of selection on genes involved in the regulation of stress were highlighted in our results. Domestication was also associated with the transformation of the fibrous tuber of *D. praehensilis* into a large, starchy tuber. We identified putative candidate genes directly associated with root development, starch formation, and storage. Selection on similar functions has been previously shown in other root and tuber crops such as cassava (*20*) and potato (*21*). This has been interpreted as the effect of human selection for increased tuber size and starch content, as well as adaptation to a dryer and brighter environment under cultivation. These interpretations are in line with our hypotheses on the origin and use of African yam.

The forest origin of cultivated yam challenges the hypothesis that domestication of sub-Saharan African plants arose mostly in tropical savannahs (*2*). *D. praehensilis* is a forest species, but it is more commonly found in disturbed forests up to the forest/savannah ecotone, where access to light is facilitated (*22*). Domestication might have occurred at the forest/savannah ecotone, as previously suggested (*2*, *7*).

The lack of archaeological remains precludes dating or tracing the history of cultivated yam (*7*), but our genomic study allows us to make inferences about the post-domestication diffusion of yams. The strong expansion of cultivated yam population ca. 2000 generations ago likely corresponds to the expansion of yam cultivation after its domestication. More precise estimates for the date of origin and for the rate of expansion would require a better estimate of the mutation rate (*11*). Unexpectedly, our analysis revealed a recent decrease in population size, ca. 400 generations ago. A similar decrease was observed in African rice, ca. 500 years ago (*6*), and is explained by the introduction of Asian rice in Western Africa (*6*). Before European colonization, yam was the most important starchy food crop in the yam belt region (*7*). Several non-African crops, notably maize and cassava, were introduced after European colonization and directly competed with yam cultivation. In 2016, the yam belt from the Republic of Côte d'Ivoire to Cameroon was dominated by cultivation of cassava and maize; yam-growing areas were restricted to 6.9 Mha, compared to 8.6 and 10.7 Mha for cassava and maize, respectively (www.fao.org/faostat). Overall, our results suggest that yam cultivation was affected early on, starting 400 generations ago, and the introduction of non-African crops was certainly a driver of its cultivation decline.

The estimated geographical origin of cultivated yam expansion was located within the basin of the Niger River, between eastern Ghana and western Nigeria. This presumed area of the origin of yam cultivation is now dry and unsuitable for *D. praehensilis*, the likely progenitor of yam. However, in the early Holocene, during the "green Sahara" period, this region was much wetter than today and it was covered by wooded savannah (*23*). It has been suggested that the intensification of Sahara desertification may have triggered the domestication of African rice and pearl millet (*5*, *6*). Similarly, environmental changes could also have contributed to the initiation of yam domestication. These changes in environmental conditions might have particularly affected the disturbed forest/savannah ecotone where we hypothesized the domestication to have occurred.

As for yams, more Northern origins than previously postulated (*2*, *24*) were found for both pearl millet (*5*) and African rice (*6*). The basin of the Niger River was presumably a hotspot of cultivation, as several archaeological sites with remains of cultivated crops are located in this region (*25–28*). Among the five crops of African origin that are most produced in Africa today (yam, African rice, sorghum, pearl millet, and cowpea; www.fao.org/faostat), four presumably originated in a restricted area: African yam expanded from the Niger River basin (present study), African rice was domesticated in the region of the Inner Niger Delta in Mali (*6*), pearl millet in northern Mali and Mauritania (*5*), while cowpea is posited to have originated in northern Ghana (*29*). Together, these results greatly refine our understanding of West African crops domestication history. They help identify a major cradle of domestication in West Africa, geographically localized around the Niger River (Fig. 4), comparable to the Fertile Crescent in the Near East.
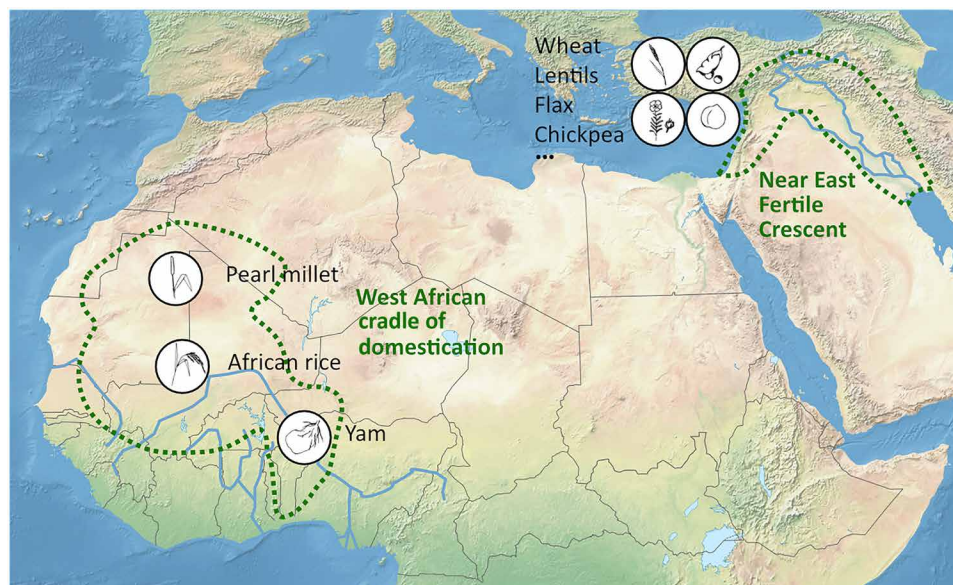

Fig. 4. Near East and West African major cradles of domestication.

The West African cradle of domestication defined here is much restricted in size than the domestication noncenter postulated by Harlan (*1*). In this sense, it is closer to the definition of domestication centers that were defined all over the world (*2*, *24*, *30*).

MATERIALS AND METHODS
Sampling and sequencing
We collected a total of 167 samples in the main growing area of yam in Ghana, Benin, Nigeria, and Cameroon (Fig. 1 and table S1). They corresponded to 86 *D. rotundata*, 34 *D. abyssinica*, and 47 *D. praehensilis*. For each individual, we collected fresh young leaves and immediately dried them in silica gel. We extracted DNA using Matab (*9*). We performed paired-end whole-genome resequencing using Illumina HiSeq 2500 and 3000 technologies in Toulouse (Genotoul), France. We constructed 400–base paired (bp) libraries using a homemade library construction described in (*31*) using internal barcodes.

Bioinformatic treatment
We first demultiplexed raw reads, based on internal barcodes, using demultadapt (https://github.com/Maillol/demultadapt). We removed adapters and low-quality bases using cutadapt 1.2.1 (*32*). We then discarded reads with a mean quality lower than 30 using a Perl script (https://github.com/SouthGreenPlatform/arcad-hts/blob/master/scripts/arcad_hts_2_Filter_Fastq_On_Mean_Quality.pl). We mapped reads onto the *D. rotundata* reference genome (genome, BDMI01000001.1 to BDMI01000021.1; mitochondrial genome, LC219374.1; plastidal genome, KJ490011) using BWA-MEM software 0.7.12-r1039 (*33*). The yam genome assembly we used was 594 Mb, with an N50 (a measure of assembly quality) of 2.12 Mb (*34*). We performed SNP calling using GATK UnifiedGenotyper 3.6-0-g89b7209 (*35*). We mapped reads onto the organelle genome together with the nuclear genome to prevent calling of polymorphisms resulting from the integration of organelle DNA into the nuclear DNA. We then filtered the 66,567,139 raw SNPs obtained according to several criteria: low quality (<200), low-mapping quality (MQ0 ≥ 4), and low- and high-mean depth (10 < DP < 20,000). SNPs were then thinned to allow no more than three SNPs into a 10-bp window. We calculated percentage of missing data per SNP and individual and mean depth per individual using VCFtools (*36*). We kept only biallelic SNPs with less than 25% of missing data, resulting in a total of 3,570,940 retained SNPs.

Genetic diversity and structure
We assessed the genetic relationships between individuals by PCA using the SNPRelate (*37*) R package. We did the PCA (i) using all SNPs and (ii) after a cutoff for minimum allele frequency (MAF) of 5%. We investigated population genetic structure using a model-based approach implemented in admixture (*38*) using a MAF cutoff of 5%. We considered *K* (the number of putative ancestral populations) values ranging from 1 to 10, with 10 runs for each *K*. We geographically plotted the highest likelihood run for each *K* using the tess3r (*39*) R package. We represented the geographical distribution of singletons by using a kriging approach function of the fields (*40*) R package. For subsequent analyses, we considered the structure obtained with *K* = 4 ancestral populations. Considering an ancestry threshold of 80%, we discarded the 14 samples with mixed ancestries (table S1). We reclassified five misclassified samples to the genetic population for which they had the highest ancestry

(table S1). Therefore, the subsequent analyses were done using 29 *D. abyssinica*, 18 Cameroonian *D. praehensilis*, 26 Western *D. praehensilis*, and 80 *D. rotundata*.

We used VCFtools (*36*) for calculating (i) π, the nucleotide diversity (*16*), in a 1000-bp window; (ii) the number of singletons per individual, as an estimation of the genetic diversity (*11*); and (iii) the genotypic LD. For LD calculation, we considered only polymorphic SNPs with a MAF cutoff of 10% in a given species, i.e., 822,720 for *D. abyssinica*, 737,627 for Western *D. praehensilis*, 1,094,239 for Cameroonian *D. praehensilis*, and 799,350 for *D. rotundata*. We calculated LD between all pairs of loci found in 1-Mb windows. We then plotted smoothed LD against distance using the R package ggplot2 (*41*).

Genetic origin of cultivated yam
We used fastsimcoal 2.5.2.21 (*12*) to evaluate from which wild relative the cultivated yam likely originated. We used a subset of 100,000 randomly sampled SNPs with observed heterozygosity lower than 0.8 for this analysis. We estimated a joint multidimensional SFS (site frequency spectrum) for all pairs of populations using a customized R script. To reduce potential bias due to missing data, we first computed realized allele frequencies, i.e., allele frequencies for each SNP were calculated excluding missing data. We then rescaled the realized allele frequencies to $(1 - X)N$ to get the final SFS, with X as the mean frequency of missing data and N as total number of genotypes. Models implemented in fastsimcoal (*12*) included three or four populations, with size ranging between 100 and 1,000,000 haploid individuals, time of divergence ranging from 100 to 1,000,000 generations, and mutation rate ranging from 1×10^{-7} to 1×10^{-9}. The models assumed no migration and no change in growth rate. All the models tested (3 + 6) are provided in fig. S3. For each model, we estimated demographic parameters using 100 runs, each run corresponding to 250,000 to 1,000,000 simulations and 40 to 150 expectation-conditional maximization loops with a convergence criterion of 1%. As fastsimcoal implements a composite-likelihood approach, one needs to refine likelihood estimation (*12*). For each model, we used estimated parameters resulting in the best likelihood value for the first round of simulations to refine likelihood estimation. We reestimated model likelihoods using 100 additional runs of 250,000 simulations. We kept the best likelihood value to estimate the Akaike information criterion (AIC) of each model as AIC = $2k - 2\log_{10}(L)$ with k as the number of estimated parameters and L as the likelihood. The AIC served as basis for model choice within the first (three models to compare) and second (six models to compare) sets.

Geographic origin of the expansion of cultivated yam
We performed spatially explicit simulations of genetic polymorphisms using SPLATCHE2 (*14*). For each simulation, we randomly picked the origin coordinates within a rectangle of side lengths 16°W to 40°E and 5°S to 20°N. This region includes both savannahs and tropical forests from West, Central, and East Africa, as well as the South Sahara. In the model, the map of Africa was discretized as a grid with 7221 cells (~90 km × 90 km). Positions corresponding to water cells were withdrawn from the model. We used uniform prior distributions for most model parameters. The pre-bottleneck-effective population size ranged from 500 to 100,000, the bottleneck-effective size ranged from 100 to 10,000, the duration of bottleneck ranged from 0 to 5000 generations, the duration of post-bottleneck expansion ranged from 50 to 2000 generations,

the growth rate ranged from 0.1 to 0.6, and the migration rate ranged from 0.15 to 0.7. To decrease time required for simulations, we fixed the mutation rate at 1×10^{-6}, meaning that the time scale was not calibrated. For each cell, we defined the carrying capacity (C) according to the vegetation type; C was fixed to 10 for desert, while another C value was assigned for other vegetation types, with a uniform distribution from 30 to 150. We used a single friction map to constrain migration by elevation, i.e., migration was more difficult in mountains than in plains. We simulated a total of 100,000 independent sites per haploid sample. We sampled simulated data at the same coordinates as the observed data. We performed a total of 1,000,000 independent simulations.

We computed two groups of summary statistics on observed and simulated data. The first group was derived from the number of singletons per individual, a statistic representing a local measure of genetic diversity, useful for the inference of the geographical origin of an expansion (11). The second group of statistics was derived from the SFS of the genetic sample, a statistic related to the demographic history of the population. We estimated the relative proportion of singletons for 11 groups (fig. S4) defined by using the k-means clustering method of the R stats package (42). We summarized the SFS by using eight bins: (i) SNPs found only once (singletons), (ii) SNPs found twice, (iii) SNPs found three to four times, (iv) SNPs found 5 to 11 times, (v) SNPs found 12 to 21 times, (vi) SNPs found 22 to 35 times, (vii) SNPs found 36 to 55 times, and (viii) SNPs found more than 55 times. We compared observed and simulated summary statistics by using an ABC analysis. We performed ABC using the ABC (13) R package with 500 independent neural networks and a tolerance rate of 0.001 (1000 simulations retained). To assess the goodness of fit of our model, we performed posterior predictive checks, i.e., we used the parameters of the 1000 retained simulations to perform 1000 extra simulations. We then checked whether these new simulations were able to recover the values of the summary statistics observed in the data.

Temporal dynamics of effective population size
We estimated past changes in effective size, Ne, using SMC++ (15). We used MSMC tools (https://github.com/stschiff/msmc-tools) to create masks for mappable positions (https://github.com/popgen-methods/smcpp/blob/master/README.rst). To get more reliable estimates of effective population size, we reestimated the frequencies of derived alleles five times, using five randomly chosen individuals as distinct lineages. We set the maximum number of generations to 100,000 and let SMC++ (15) define the lower bound based on a heuristic approach. We fixed the mutation rate to 6.5×10^{-9}, corresponding to the mutation rate estimated for monocotyledons (rice, maize, and barley) (43), and the generation time to 1 year.

Detection of selection
We investigated the existence of selection signatures using three complementary methods: (i) detection of selective sweeps for the cultivated individuals using the likelihood-ratio test of SweeD 3.3.2 (18), (ii) calculation of π ratio as $\pi_{wild}/\pi_{cultivated}$, and (iii) calculation of Weir and Cockerham's (44) F_{ST} between wild and cultivated populations. For composite-likelihood ratio (CLR) computation, we adapted the grid of each chromosome to calculate a CLR peak every 100 bp. We considered CLR within 100-kb windows as potentially resulting from similar selection events. We thus grouped them into common windows of 50 kb on each side of the most extreme peak. We calculated π ratio and F_{ST} using sliding windows of 50 kb every 10 kb with VCFtools (36). For the three methods, we reported genome annotations of the 1% most extreme values using BEDTools 2.2 (44) (windowBed command, default parameters) based on *D. rotundata* genome annotations (34). We only considered genes with a meaningful annotation. We associated GO terms with loci using a customized R script. We performed enrichment tests for GO biological processes, cellular components, and molecular function terms using the Fisher exact test implemented in the R package topGO (45).

REFERENCES AND NOTES
1. J. F. Doebley, B. S. Gaut, B. D. Smith, The molecular genetics of crop domestication. *Cell* **127**, 1309–1321 (2006).
2. J. R. Harlan, *Crops and Man* (American Society of Agronomy, ed. 2, 1992).
3. S. Lev-Yadun, A. Gopher, S. Abbo, The cradle of agriculture. *Science* **288**, 1602–1603 (2000).
4. D. Q. Fuller, S. Nixon, C. J. Stevens, M. A. Murray, in *Archaeology of African Plant Use*, C. J. Stevens, S. Nixon, M. A. Murray, D. Q. Fuller, Eds. (Left Coast Press Inc., 2014), pp. 17–24.
5. C. Burgarella, P. Cubry, N. A. Kane, R. K. Varshney, C. Mariac, X. Liu, C. Shi, M. Thudi, M. Couderc, X. Xu, A. Chitikineni, N. Scarcelli, A. Barnaud, B. Rhoné, C. Dupuy, O. François, C. Berthouly-Salazar, Y. Vigouroux, A western Sahara centre of domestication inferred from pearl millet genomes. *Nat. Ecol. Evol.* **2**, 1377–1380 (2018).
6. P. Cubry, C. Tranchant-Dubreuil, A. C. Thuillet, C. Monat, M. N. Ndjiondjop, K. Labadie, C. Cruaud, S. Engelen, N. Scarcelli, B. Rhoné, C. Burgarella, C. Dupuy, P. Larmande, P. Wincker, O. François, F. Sabot, Y. Vigouroux, The rise and fall of African rice cultivation revealed by analysis of 246 new Genomes. *Curr. Biol.* **28**, 2274–2282.e6 (2018).
7. D. G. Coursey, in *Origins of African Plant Domestication*, J. R. Harlan, J. M. J. D. Wet, A. B. L. Stemler, Eds. (De Gruyter Mouton, 1976).
8. N. Scarcelli, S. Tostain, Y. Vigouroux, C. Agbangla, O. Daïnou, J.-L. Pham, Farmers' use of wild relative and sexual reproduction in a vegetatively propagated crop. The case of yam in Benin. *Mol. Ecol.* **15**, 2421–2431 (2006).
9. J. Magwé-Tindo, J. J. Wieringa, B. Sonké, L. Zapfack, Y. Vigouroux, T. L. P. Couvreur, N. Scarcelli, Guinea yam (*Dioscorea* spp., Dioscoreaceae) wild relatives identified using whole plastome phylogenetic analyses. *Taxon* **67**, 905–915 (2018).
10. P. Cubry, Y. Vigouroux, O. François, The empirical distribution of singletons for geographic samples of DNA sequences. *Front. Genet.* **8**, 139 (2017).
11. L. Excoffier, I. Dupanloup, E. Huerta-Sánchez, V. C. Sousa, M. Foll, Robust demographic Inference from genomic and SNP data. *PLOS Genet.* **9**, e1003905 (2013).
12. K. Csilléry, O. François, M. G. B. Blum, abc: An R package for approximate Bayesian computation (ABC). *Methods Ecol. Evol.* **3**, 475–479 (2012).
13. N. Ray, M. Currat, M. Foll, L. Excoffier, SPLATCHE2: A spatially explicit simulation framework for complex demography, genetic admixture and recombination. *Bioinformatics* **26**, 2993–2994 (2010).
14. J. Terhorst, J. A. Kamm, Y. S. Song, Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nat. Genet.* **49**, 303–309 (2017).
15. M. Nei, *Molecular Evolutionary Genetics* (Columbia Univ. Press, 1987).
16. B. S. Weir, C. C. Cockerham, Estimating F-Statistics for the analysis of population structure. *Evolution* **38**, 1358–1370 (1984).
17. P. Pavlidis, D. Živković, A. Stamatakis, N. Alachiotis, SweeD: Likelihood-based detection of selective sweeps in thousands of genomes. *Mol. Biol. Evol.* **30**, 2224–2234 (2013).
18. R. Akakpo, N. Scarcelli, H. Chaïr, A. Dansi, G. Djedatin, A.-C. Thuillet, B. Rhoné, O. François, K. Alix, Y. Vigouroux, Molecular basis of African yam domestication: Analyses of selection

point to root development, starch biosynthesis, and photosynthesis related genes. *BMC Genomics* **18**, 782 (2017).

19. R. Terauchi, V. A. Chikaleke, G. Thottappilly, S. K. Hahn, Origin and phylogeny of Guinea yams as revealed by RFLP analysis of chloroplast DNA and nuclear ribosomal DNA. *Theor. Appl. Genet.* **83**, 743–751 (1992).

20. W. Wang, B. Feng, J. Xiao, Z. Xia, X. Zhou, P. Li, W. Zhang, Y. Wang, B. L. Møller, P. Zhang, M.-C. Luo, G. Xiao, J. Liu, J. Yang, S. Chen, P. D. Rabinowicz, X. Chen, H.-B. Zhang, H. Ceballos, Q. Lou, M. Zou, L. J. C. B. Carvalho, C. Zeng, J. Xia, S. Sun, Y. Fu, H. Wang, C. Lu, M. Ruan, S. Zhou, Z. Wu, H. Liu, R. M. Kannangara, K. Jørgensen, R. L. Neale, M. Bonde, N. Heinz, W. Zhu, S. Wang, Y. Zhang, K. Pan, M. Wen, P.-A. Ma, Z. Li, M. Hu, W. Liao, W. Hu, S. Zhang, J. Pei, A. Guo, J. Guo, J. Zhang, Z. Zhang, J. Ye, W. Ou, Y. Ma, X. Liu, L. J. Tallon, K. Galens, S. Ott, J. Huang, J. Xue, F. An, Q. Yao, X. Lu, M. Fregene, L. A. B. López-Lavalle, J. Wu, F. M. You, M. Chen, S. Hu, G. Wu, S. Zhong, P. Ling, Y. Chen, Q. Wang, G. Liu, B. Liu, K. Li, M. Peng, Cassava genome from a wild ancestor to cultivated varieties. *Nat. Commun.* **5**, (2014).

21. M. A. Hardigan, F. P. E. Laimbeer, L. Newton, E. Crisovan, J. P. Hamilton, B. Vaillancourt, K. Wiegert-Rininger, J. C. Wood, D. S. Douches, E. M. Farré, R. E. Veilleux, C. R. Buell, Genome diversity of tuber-bearing *Solanum* uncovers complex evolutionary history and targets of domestication in the cultivated potato. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E9999–E10008 (2017).

22. R. Dumont, A. Dansi, P. Vernier, J. Zoundjihèkpon, *Biodiversity and Domestication of Yams in West Africa: Traditional Practices Leading to Dioscorea rotundata Poir* (CIRAD-IPGR, Montpellier, Repères, 2006).

23. C. Hély, P. Braconnot, J. Watrin, W. Zheng, Climate and vegetation: Simulating the African humid period. *C. R. Geosci.* **341**, 671–688 (2009).

24. M. D. Purugganan, D. Q. Fuller, The nature of selection during plant domestication. *Nature* **457**, 843–848 (2009).

25. S. Ozainne, L. Lespez, A. Garnier, A. Ballouche, K. Neumann, O. Pays, E. Huysecom, A question of timing: Spatio-temporal structure and mechanisms of early agriculture expansion in West Africa. *J. Archaeol. Sci.* **50**, 359–368 (2014).

26. K. Manning, R. Pelling, T. Higham, J.-L. Schwenniger, D. Q. Fuller, 4500-year old domesticated pearl millet (*Pennisetum glaucum*) from the Tilemsi Valley, Mali: New insights into an alternative cereal domestication pathway. *J. Archaeol. Sci.* **38**, 312–322 (2011).

27. F. Marshall, E. Hildebrand, Cattle before crops: The beginnings of food production in Africa. *J. World Prehist.* **16**, 99–143 (2002).

28. S. Kahlheber, K. Neumann, in *Rethinking Agriculture: Archaeological and Ethnoarchaeological Perspectives*, T. Denham, J. Iriarte, L. Vrydaghs, Eds (Taylor and Francis Group, 2007), pp. 320–346.

29. A. C. D'Andrea, S. Kahlheber, A. L. Logan, D. J. Watson, Early domesticated cowpea (*Vigna unguiculata*) from Central Ghana. *Antiquity* **81**, 686–698 (2007).

30. D. R. Piperno, The origins of plant cultivation and domestication in the new world tropics: Patterns, process, and new developments. *Curr. Anthropol.* **52**, S453–S470 (2011).

31. C. Mariac, N. Scarcelli, J. Pouzadou, A. Barnaud, C. Billot, A. Faye, A. Kougbeadjo, V. Maillol, G. Martin, F. Sabot, S. Santoni, Y. Vigouroux, T. L. P. Couvreur, Cost-effective enrichment hybridization capture of chloroplast genomes at deep multiplexing levels for population genetics and phylogeography studies. *Mol. Ecol. Resour.* **14**, 1103–1113 (2014).

32. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10 (2011).

33. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

34. M. Tamiru, S. Natsume, H. Takagi, B. White, H. Yaegashi, M. Shimizu, K. Yoshida, A. Uemura, K. Oikawa, A. Abe, N. Urasaki, H. Matsumura, P. Babil, S. Yamanaka, R. Matsumoto, S. Muranaka, G. Girma, A. Lopez-Montes, M. Gedil, R. Bhattacharjee, M. Abberton, P. L. Kumar, I. Rabbi, M. Tsujimura, T. Terachi, W. Haerty, M. Corpas, S. Kamoun, G. Kahl, H. Takagi, R. Asiedu, R. Terauchi, Genome sequencing of the staple food crop white Guinea yam enables the development of a molecular marker for sex determination. *BMC Biol.* **15**, 86 (2017).

35. A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, M. A. DePristo, The genome analysis toolkit:

36. P. Danecek, A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, R. Durbin; 1000 Genomes Project Analysis Group, The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).

37. X. Zheng, D. Levine, J. Shen, S. M. Gogarten, C. Laurie, B. S. Weir, A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* **28**, 3326–3328 (2012).

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

39. K. Caye, T. M. Deist, H. Martins, O. Michel, O. François, TESS3: Fast inference of spatial population structure and genome scans for selection. *Mol. Ecol. Resour.* **16**, 540–548 (2016).

40. D. Nychka, *fields: Tools for Spatial Data* (UCAR/NCAR–Computational and Information Systems Laboratory, 2016); www.image.ucar.edu/fields.

41. H. Wickham, *ggplot2: Elegant Graphics for Data Analysis (Use R!)* (Springer, 2009).

42. R Core Team, *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2013); www.R-project.org.

43. B. S. Gaut, B. R. Morton, B. C. McCaig, M. T. Clegg, Substitution rate comparisons between grasses and palms: Synonymous rate differences at the nuclear gene Adh parallel rate differences at the plastid gene rbcL. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 10274–10279 (1996).

44. A. R. Quinlan, I. M. Hall, BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

45. A. Alexa, J. Rahnenfuhrer, topGO: Enrichment analysis for gene ontology. R package version 2.32.0 (2016).

A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

Science Advances

Yam genomics supports West Africa as a major cradle of crop domestication

Nora Scarcelli, Philippe Cubry, Roland Akakpo, Anne-Céline Thuillet, Jude Obidiegwu, Mohamed N. Baco, Emmanuel Otoo, Bonaventure Sonké, Alexandre Dansi, Gustave Djedatin, Cédric Mariac, Marie Couderc, Sandrine Causse, Karine Alix, Hâna Chaïr, Olivier François and Yves Vigouroux

ARTICLE TOOLS	http://advances.sciencemag.org/content/5/5/eaaw1947
SUPPLEMENTARY MATERIALS	http://advances.sciencemag.org/content/suppl/2019/04/29/5.5.eaaw1947.DC1
RELATED CONTENT	http://science.sciencemag.org/content/sci/364/6439/422.full
REFERENCES	This article cites 34 articles, 5 of which you can access for free http://advances.sciencemag.org/content/5/5/eaaw1947#BIBL
PERMISSIONS	http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the Terms of Service