

# African genomes illuminate the early history and transition to selfing in *Arabidopsis thaliana*

Arun Durvasula<sup>a,b,c,1</sup>, Andrea Fulgione<sup>a,b,c,1</sup>, Rafal M. Gutaker<sup>d</sup>, Selen Irez Alacakaptan<sup>b,c</sup>, Pádraic J. Flood<sup>a</sup>, Célia Neto<sup>a</sup>, Takashi Tsuchimatsu<sup>e</sup>, Hernán A. Burbano<sup>d</sup>, F. Xavier Picó<sup>f</sup>, Carlos Alonso-Blanco<sup>g</sup>, and Angela M. Hancock<sup>a,b,c,2</sup>

<sup>a</sup>Department of Plant Developmental Biology, Max Planck Institute for Plant Breeding Research, 50829 Cologne, Germany; <sup>b</sup>Department of Structural and Computational Biology, University of Vienna, 1010 Vienna, Austria; <sup>c</sup>Vienna Biocenter, 1030 Vienna, Austria; <sup>d</sup>Research Group for Ancient Genomics and Evolution, Department of Molecular Biology, Max Planck Institute for Developmental Biology, 72076 Tübingen, Germany; <sup>e</sup>Department of Biology, Chiba University, Chiba 263-8522 Japan; <sup>f</sup>Departamento de Ecología Integrativa, Estación Biológica de Doñana, Consejo Superior de Investigaciones Científicas, 41092 Seville, Spain; and <sup>g</sup>Centro Nacional de Biotecnología, Consejo Superior de Investigaciones Científicas, 28049 Madrid, Spain

Edited by Johanna Schmitt, University of California Davis, CA, and approved April 11, 2017 (received for review October 13, 2016)

Over the past 20 y, many studies have examined the history of the plant ecological and molecular model, *Arabidopsis thaliana*, in Europe and North America. Although these studies informed us about the recent history of the species, the early history has remained elusive. In a large-scale genomic analysis of African *A. thaliana*, we sequenced the genomes of 78 modern and herbarium samples from Africa and analyzed these together with over 1,000 previously sequenced Eurasian samples. In striking contrast to expectations, we find that all African individuals sampled are native to this continent, including those from sub-Saharan Africa. Moreover, we show that Africa harbors the greatest variation and represents the deepest history in the *A. thaliana* lineage. Our results also reveal evidence that selfing, a major defining characteristic of the species, evolved in a single geographic region, best represented today within Africa. Demographic inference supports a model in which the ancestral *A. thaliana* population began to split by 120–90 kya, during the last interglacial and Abbassia pluvial, and Eurasian populations subsequently separated from one another at around 40 kya. This bears striking similarities to the patterns observed for diverse species, including humans, implying a key role for climatic events during interglacial and pluvial periods in shaping the histories and current distributions of a wide range of species.

evolution | population history | self-compatibility | climate | migration

The plant *Arabidopsis thaliana* is the principal plant model species, and as such has been useful not only to examine basic biological mechanisms but also to elucidate evolutionary processes. The exceptional resources available in this species, including seed stocks collected from throughout Eurasia for over 75 y, have been a valuable tool for learning about the natural history of *A. thaliana* on this continent (1, 2). Previous studies have shown that current variation in Eurasia is mainly a result of expansions and mixing from refugia in Iberia, Central Asia, and Italy/Balkans after the end of the last glacial period ~10 kya (3–8). The main finding of the recent analysis of 1,135 sequenced genomes was that a few Eurasian samples represent divergent relict lineages, whereas the vast majority derived from the recent expansion of a single clade (4). Given the large number of studies that examine the natural history of *A. thaliana*, one would expect that this history would by now be described rather completely and there would be no major surprises left to uncover. However, there are still many open questions about the ancient history of the species.

Several features differentiate *A. thaliana* from its closest relatives. Although most members of the *Arabidopsis* genus are obligate out-crossing perennials with large flowers and genome sizes of over 230 Mb and 8 chromosomes, *A. thaliana* is a predominantly selfing annual with reduced floral morphology and a reduced genome size of ~150 Mb and 5 chromosomes. The transition to predominant selfing in *A. thaliana* was likely the catalyst for these derived morphological and genomic features (9–13). These changes, in particular the rearranged and shrunken genome, created a strong reproductive barrier between *A. thaliana* and its closest relatives (14).

Although the genetic basis of self-compatibility in *A. thaliana* is known, the specific events that occurred during the transition to predominant selfing are still unclear. In obligate out-crossing *Arabidopsis* species, many highly divergent S-locus haplogroups (S-haplogroups) are maintained by balancing selection, providing a mechanism for inbreeding avoidance. In *A. thaliana*, three S-haplogroups are found, and each contains mutations that obliterate function of the S-locus genes (15–17). Loss-of-function occurred independently in each S-haplogroup (18–21), but because these three S-haplogroups were never found together in the same geographic region, self-compatibility is inferred to have evolved separately in multiple locations (16, 21, 22). However, the hypothesis of geographically distinct origins is difficult to reconcile with the major genomic and phenotypic changes that render *A. thaliana* incompatible with its out-crossing congeners (9–13). Shifts from out-crossing to predominant selfing are common and have been considered the most prevalent evolutionary transitions in flowering plants (23). Reconstructing the evolutionary history of the transition to selfing in *A. thaliana* could provide general insights into this common evolutionary

## Significance

The principal plant model species, *Arabidopsis thaliana*, is central to our understanding of how molecular variants lead to phenotypic change. In this genome-sequencing effort focused on accessions from Africa, we show that African populations represent the most ancient lineages and provide new clues about the origin of selfing and the species itself. Population history in Africa contrasts sharply with the pattern in Eurasia, where the vast majority of samples result from the recent expansion of a single clade. This previously unexplored reservoir of variation is remarkable given the large number of genomic studies conducted previously in this well-studied species and implies that assaying variation in Africa may often be necessary for understanding population history in diverse species.

Author contributions: A.D., A.F., and A.M.H. designed research; A.D., A.F., R.M.G., S.I.A., P.J.F., C.N., T.T., H.A.B., and A.M.H. performed research; A.D., A.F., F.X.P., C.A.-B., and A.M.H. contributed new reagents/analytic tools; A.D., A.F., and A.M.H. analyzed data; and A.D., A.F., and A.M.H. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The sequences reported in this paper has been deposited in the European Nucleotide Archive/Sequence Read Archive database, study PRJEB19780 (accession nos. ERS1575066–ERS1575147). Analysis scripts are available at [https://github.com/HancockLab/African\\_A.thaliana](https://github.com/HancockLab/African_A.thaliana).

<sup>1</sup>A.D. and A.F. contributed equally to this work.

<sup>2</sup>To whom correspondence should be addressed. Email: [hancock@mpipz.mpg.de](mailto:hancock@mpipz.mpg.de).

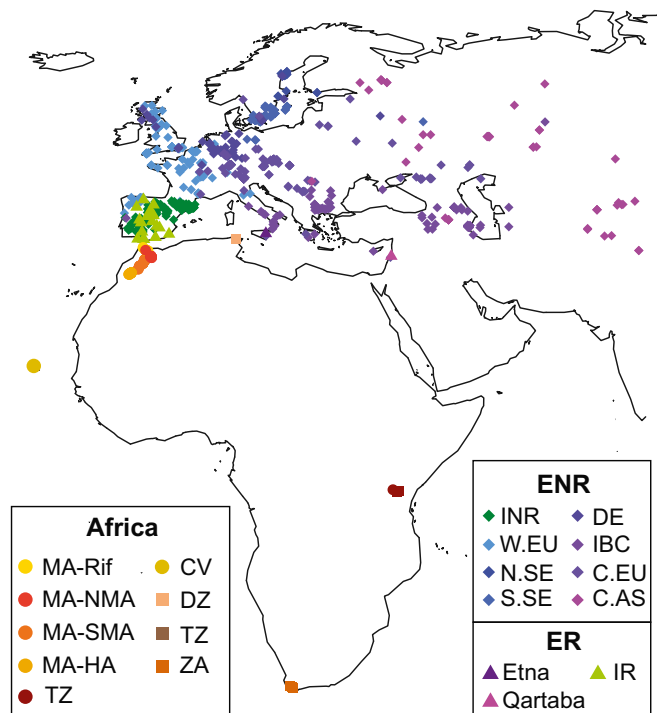
This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1616736114/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1616736114/-DCSupplemental).

transition. However, because substantial time has passed since this transition (24–26) and no intermediate forms have been found between *A. thaliana* and its obligate out-crossing relatives, this reconstruction is challenging.

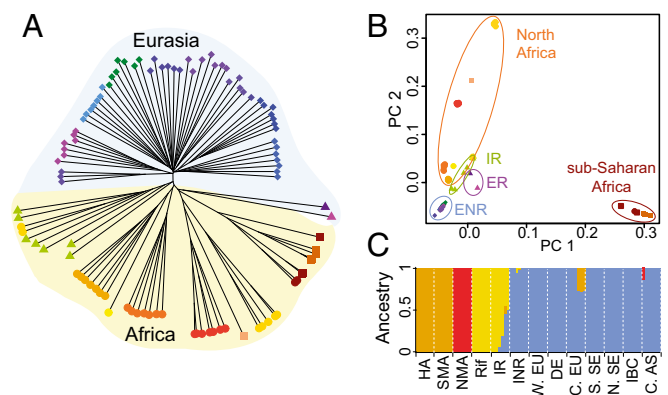
We sequenced the genomes of 78 African samples and analyze these in combination with 1,135 previously sequenced samples (4) (Fig. 1 and *SI Appendix, Table S1*). We find that African variation reveals the ancient history of the species and clarifies details concerning the transition to selfing. Congruence of *A. thaliana* population history with major climatic events and paleontological observations illustrates the relevance of population genetic studies for understanding climate-mediated demography more generally.

## Results

To examine the relationship between African individuals and other worldwide samples, we used three complementary clustering approaches. Distance-based clustering by neighbor-joining reveals a clear split between Eurasian and African samples, indicating deep divergence between the continents (Fig. 2A and *SI Appendix, Fig. S1*). The majority of Eurasian samples form a nearly star-shaped phylogeny, consistent with recent expansion of these lineages. Conversely, longer more bifurcated branches separate African subclusters and previously identified Eurasian relicts from each other and from the nonrelict clade. In general, the Eurasian clades cluster consistently with the nine groups defined previously (4). Exceptions are the Central Europe clade, which separates into two clusters, and the Iberian relicts, which cluster with the Moroccan Rif-Zin population. Moroccan samples separate into four clades, reflecting their geographic distribution and South Africa and Tanzania cluster together in a single clade. The results for South Africa and Tanzania are



**Fig. 1.** Sample map of accessions included in this study. Herbarium samples are shown as squares. Abbreviations are as follows: Algeria (DZ), Cape Verde (CV), Central Asia (C.AS), Central Europe (C.EU), Eurasian nonrelicts (ENR), Eurasian relicts (ER), Germany (DE), Italy, Balkans, and Caucasus (IBC), Iberian nonrelicts (INR), Iberian relicts (IR), Morocco (MA), North Sweden (N.SE), South Africa (ZA), South Sweden (S.SE), Tanzania (TZ), Western Europe (W.EU).



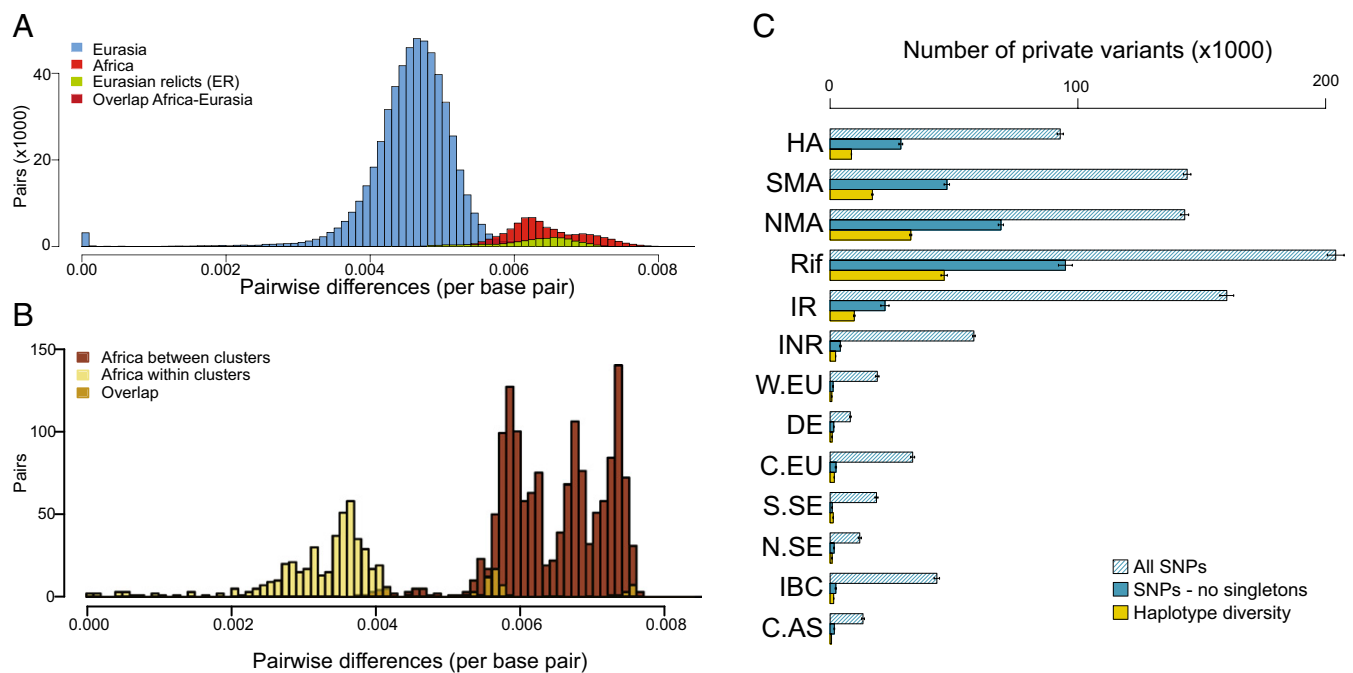
**Fig. 2.** Global population structure. (A) Unrooted neighbor-joining tree, (B) PCA, (C) ADMIXTURE results for  $K = 4$ .

striking because *A. thaliana* populations outside of Eurasia and North Africa were previously thought to be recently introduced by humans (27).

Similarly, principal component analysis (PCA) separates African populations from each other and from Eurasian populations (Fig. 1B). The first PC distinguishes sub-Saharan Africa, and the second separates the four Moroccan clusters from Eurasians. Subsequent PCs mainly discriminate populations within Africa, whereas Eurasian populations remain tightly clustered (*SI Appendix, Fig. S3*). Results from ADMIXTURE (28) reinforce this finding. Moroccan populations separate into three clusters and are distinguished from a single cluster of Eurasian samples (Fig. 1C, and *SI Appendix, Figs. S4 and S5, and Table S2*). PCA and ADMIXTURE results also suggest a Moroccan origin of the relicts in Iberia, which are spread between Rif-Zin and Iberian nonrelicts in PCA, and sizable portions of the Iberian relict genomes match the Moroccan clusters in ADMIXTURE. This finding is consistent with previous work (29) and with the accepted phylogeographical history of Mediterranean and North African flora characterized by a complex history of expansions and contractions driven by important climatic changes experienced in this vast region, particularly since the Pliocene (30, 31).

Furthermore, from pairwise differences, we recover the previously reported difference between Eurasian relicts and nonrelicts (4) and find that all African accessions are at least as divergent as samples previously classified as relicts (Fig. 3A and *SI Appendix, Fig. S6*). Therefore, in contrast to Eurasia, where most samples represent a single recently spread clade, all African individuals represent relict samples. The distribution of pairwise differences within Africa (Fig. 3B) further demonstrates the high diversity in these samples.

If populations in Africa truly are more ancient than the Eurasian clusters, we should also expect higher numbers of private variants in Africa. Indeed, we find that the Moroccan clusters and Iberian relicts, which appear likely derived from Morocco, harbor the highest numbers of private SNPs (Fig. 3C and *SI Appendix, Fig. S8 and Table S3*). This signal intensifies when we exclude recently arisen variants, which we find constitute the majority of the private variation in Eurasia. First, we excluded singletons, the class of SNPs most influenced by recent population growth, and found a 7.0- to 23.2-fold enrichment in Morocco compared with the top nonrelict cluster (Fig. 3C). Next, we considered the spatial distribution of private variants across the genome. Because novel private variants are unlikely to be tightly linked, clustering of several, contiguous private SNPs indicates old haplotypes. At the haplotype level, we again found extremely high enrichment (3.9- to 20.9-fold) for Moroccan clusters and Iberian relicts (4.5-fold) relative to the top Eurasian cluster (Fig. 3C). Notably, the Eastern



**Fig. 3.** Patterns of diversity across geographic regions. Distributions of genome-wide pairwise differences per base pair in: (A) worldwide comparison and (B) within and between African populations, where overlap between distributions is shown as described in legends. (C) Numbers of private SNPs and haplotypes found in each cluster. Error bars denote 95% confidence intervals.

Mediterranean and Caucasus regions, which were previously favored as points of origin and major centers of diversity of the species (3), do not exhibit a striking pattern for any of the metrics examined (see IBC in Fig. 3C). These findings evoke a model in which polymorphism in Africa is because of ancient variation and Eurasian polymorphism is mainly because of recent expansion.

To better understand the relevance of variation in Africa for the early history of the *A. thaliana* lineage, we examined variation at the locus that confers self-compatibility. S-locus variation in Africa differs in several ways from what is found in Eurasia (Fig. 4 and *SI Appendix, Table S4*). We found all three S-haplogroups together in a single geographic region, with S-haplogroup B private to Africa. In addition, the A-C recombinant is also present at low frequency in Morocco, in contrast to what is found in Eurasia (32). Finally, we discovered deletion haplotypes in haplogroups A and C in Morocco (*SI Appendix, Fig. S9*). The finding that all S-locus haplogroups are present together implies that selfing evolved in a single geographic region.

Taken together, the patterns in population structure, and levels of variation across the genome and at the S-locus, specifically, imply a deep history in Africa. To clarify the details of past demographic events, we inferred historical effective population sizes ( $N_e$ ) and split times among populations based on cross-coalescent rates (CCR) using a multiple sequentially Markovian coalescent approach (MSMC) (33).

In ancient times, we find the highest  $N_e$  is in Africa, peaking at around 500–400 kya (Figs. 5 and *SI Appendix, Fig. S10*). All Eurasian populations (including Iberian relicts) show the same trajectories as the Africans, but with lower amplitudes. Given that these curves are in phase with one another and we do not see evidence for a population split until 120 kya (Fig. 6A), we interpret this as population structure in the ancestral population combined with bottlenecks as more derived populations migrated away from this ancestral population. This finding is consistent with our finding that variation in Eurasia is often a subset of variation present within Africa and is similar to the situation in humans (34). Notably, the IBC cluster, which includes previously hypothesized

*A. thaliana* origins and refugia (3, 5, 7), exhibits a much lower ancient population size than Africa (*SI Appendix, Fig. S10*).

At 120–90 kya, there are bottlenecks in all populations (Fig. 5) and a split among the major clades (Moroccan, Tanzanian, and Levant) (Fig. 6 and *SI Appendix, Fig. S11*). This roughly corresponds to the Abbassia Pluvial, a period when migration corridors were open because of high precipitation and humidity in Africa (35) and also marks the last interglacial at Marine Isotope Stage 5e (130–116 kya), when temperatures were 1–2° warmer than present-day conditions, providing favorable conditions in Eurasia. As this interglacial period came to a close, there was a worldwide shift toward cooler, drier conditions as the most recent and severe Pleistocene glaciation phase began (36). Beginning at this time, CCR implies a progressive decline in population connectivity, consistent with decreasing temperature and increasing aridity (Fig. 6A). We checked for consistency using a complementary method that relies on the joint site frequency spectrum between populations (*δ<sub>adi</sub>*) (37) and found a slightly older estimate for the split and overlapping confidence intervals (141–116 kya) (*SI Appendix, Table S5*). We propose that the most likely scenario is that *A. thaliana* was colonizing broadly within Africa as well as in the Levant during the last interglacial (130–116 kya), but that connections between populations began to break down as populations spread and as climate became cooler and drier at the end of this period.

MSMC based on eight haplotypes detects several changes in  $N_e$  in more recent times (Fig. 5B). Since ~120 kya, the population size changes in Europe and Asia are often out of phase with those in Africa, consistent with geographical separation and exposure to different climatic regimes. Maxima in African populations occur at around 60–40 kya and 11–5 kya, corresponding to orbital-scale climate shifts from arid to moist conditions (i.e., pluvial periods) that occurred at 11–5 kya and 59–47 kya in North Africa (38, 39). We find relatively recent split times between sub-Saharan African clades (South Africa and Tanzania) and between Western Europe and Central Asia, at around 40 kya (Fig. 6).

There are a few caveats to consider regarding the demographic inference. Because MSMC can spread instantaneous population

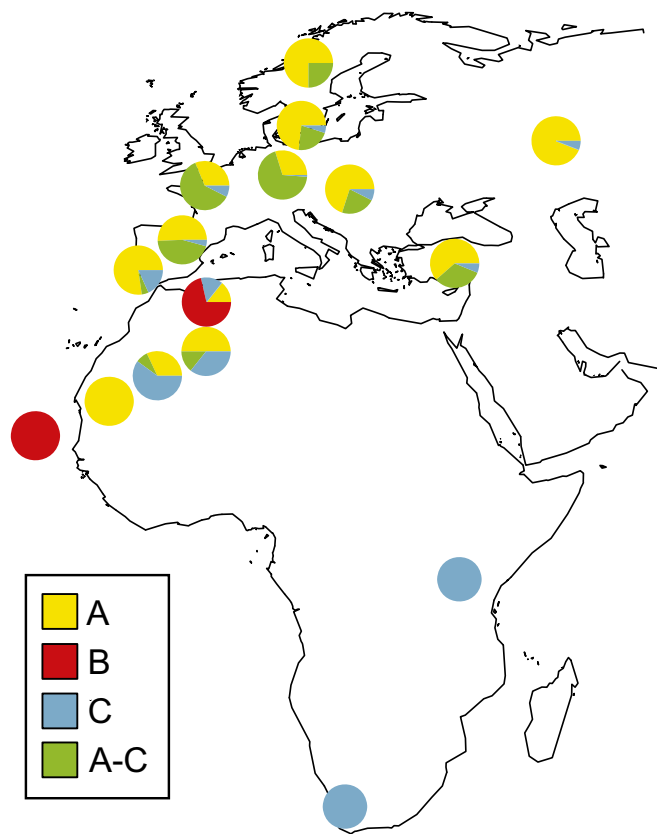


Fig. 4. Map of S-locus haplogroup diversity.

size changes over time, maxima and minima are informative but the slope should be interpreted with caution (33). In addition, the precise timing associated with inferred population size changes and splits is dependent on parameters that are difficult to measure and may vary over space and time, including mutation rate, degree of purifying selection, and the possible input from a seed bank. We used the best available data for mutation rate [based on mutation accumulation experiments (40)] and made the usual simplifying assumptions for other parameters (one generation per year), but

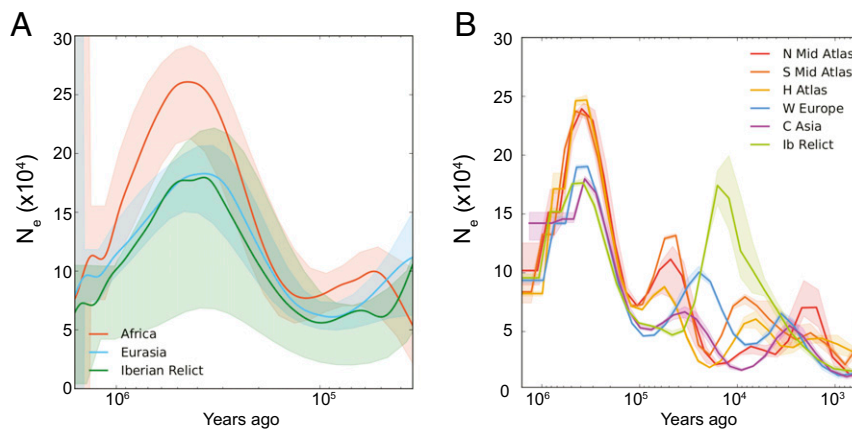


Fig. 5. Historical effective population size of *A. thaliana* inferred using MSMC. Although two-haplotype analysis provides more resolution in the distant past, eight-haplotype analysis provides better resolution in the recent past. (A) Inference using pairs of haplotypes, with lines representing medians and shading representing  $\pm 1$  SD calculated across pairs. This analysis is expected to produce unbiased estimates between 40 kya and 1.6 Mya (SI Appendix). (B) Inference based on sets of eight haplotypes with lines representing medians. This analysis is expected to produce unbiased estimates as recently as 1.6 kya.

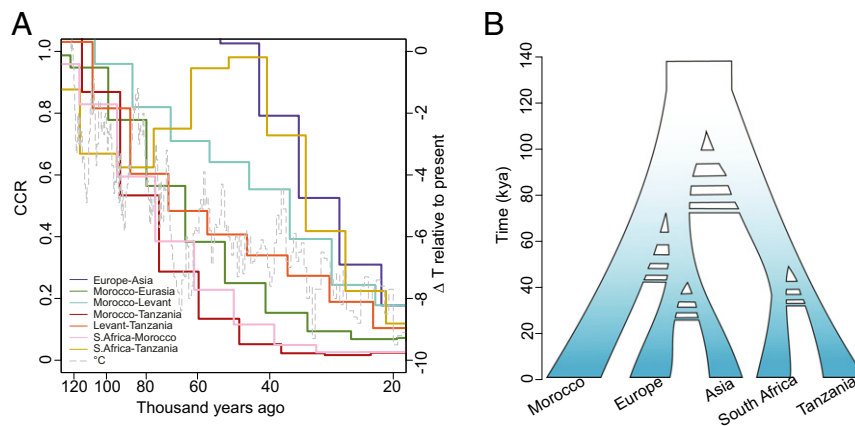
the timing we infer would need to be revised if these assumptions turned out to be incorrect.

## Discussion

Genomic studies thus far have amassed data for nearly 2,000 Eurasian *A. thaliana* accessions but were unable to provide insight into the early history of the species. Here, in a genome-scale sequencing effort focused on African accessions, we find clear evidence for a deep history of African *A. thaliana* populations, which harbor variation that was either lost or never present in Eurasia. Several specific results were unexpected based on current knowledge in this well-studied species. First, we discovered surprising and clear evidence that *A. thaliana* is native not only to North Africa but also to Afro-alpine regions of sub-Saharan Africa. Second, our results revealed that the deepest splits species-wide separate the African lineages from one another and that in ancient times, the effective population size was largest in Africa. Finally, we learned that variation at the S-locus is highest in Africa and that all three S-haplogroups are present there.

Based on our results, we can outline a model for the early history and transition to selfing in *A. thaliana* (detailed in SI Appendix, Fig. S12). In the first step, we infer that the population ancestral to *A. thaliana* became geographically separated from its parental out-crossing population. Our results suggest that this separation involved migration of the ancestral subpopulation into Africa by 1.2–0.8 Mya. This timing corresponds to the Middle Pleistocene Transition, a shift to drier more variable climate and more open habitats in Africa (i.e., grasslands versus woodlands), as evidenced by soil carbon analysis showing an increase in the ratio of C4 to C3 plants (41, 42).

Although the estimated divergence times between *A. thaliana* and *Arabidopsis lyrata* center around 5–7 Mya (9, 43), the origin of *A. thaliana* itself appears to be much younger. Our model predicts that there was an initial bottleneck as the subpopulation that led to *A. thaliana* split from a *A. lyrata*-like ancestral population [similar to that observed in *Mimulus nasutus* (44) and *Capsella rubella* (45–47)], followed by an expansion in  $N_e$  as the selfing population began to spread. In this case, we could interpret the MSMC results to suggest that the transition to selfing occurred between 1 Mya and 500 kya, before the most ancient maximum in  $N_e$ . This finding is in line with an estimate based on the depth of the *A. thaliana* genealogy (0.84% maximum divergence among individuals sampled here) under a simple model ( $T \sim D/2 \mu \sim 598$  kya). Our estimated timing is also consistent



**Fig. 6.** Inferred timing of population splits. (A) Relative CCR between populations. Decreasing values from 1.0 indicate population separation. The dashed line represents historical temperature (63). (B) A schematic model for the demographic history of *A. thaliana* based on CCR results, with hashes to represent uncertainty regarding possible timing of gene flow events.

with previous estimates for the loss of self-incompatibility and origin of selfing (24–26).

Once selfing was established, traits associated with the “selfing syndrome” would have been favored, including reduced pollen number and petal size (48). Such phenotypic shifts are common in predominantly selfing species and have occurred in *A. thaliana* compared with its closest relatives (26). At the genomic level, *A. thaliana* exhibits major chromosomal rearrangements and a reduction in genome size and number of chromosomes (49). This genomic reduction is also likely a by-product of the shift to predominant selfing in *A. thaliana* (9–11), consistent with an observed link between reduced genome size and selfing in other plant species (11–13). These changes introduce a strong reproductive barrier as found in hybrids of *A. thaliana* and *A. lyrata*, which are infertile because of the chromosomal rearrangements that occurred in *A. thaliana* (14).

Given that all three S-haplogroups co-occur in Morocco, we hypothesize that the transition to predominant selfing occurred in a single region, best represented today in Morocco. This finding differs from previous assertions that these events likely happened separately in geographically distinct populations (15, 16). Moreover, it allows for the possibility that the transition to selfing was aided by a shared precursor mutation, a shared climate, and the bottleneck that occurred during the migration away from the ancestral population (22). Our proposed model parallels observations in partially selfing populations of *A. lyrata* (50). Here, self-compatibility is associated with two different S-haplogroups in Great Lakes populations and self-compatibility may have been favored because of the bottleneck that initially limited S-haplogroup diversity and thus mate availability.

After the origin and initial population size increase of *A. thaliana*, we infer several demographic changes that are congruent with known climatic shifts. At 120–90 kya, we find evidence from MSMC and *daði* for splitting among the major clades: Morocco, Levant and sub-Saharan Africa. This split corresponds to the Abbassia pluvial, which produced migration corridors within Africa (120–90 kya) (35, 39) as well as Marine Isotope Stage 5e (130–116 kya), the last interglacial period, when worldwide temperatures were 1–2° warmer than they are currently (51, 52). This is consistent with a model in which *A. thaliana* spread widely throughout Africa and into Eurasia when conditions were favorable (~120 kya), with isolation as gene flow was reduced (*SI Appendix*, Fig. S12). More recent major demographic events include the split between European and Asian populations at around 40 kya and the increase in  $N_e$  within Africa during the most recent pluvial.

The patterns we observe and their concordance with climatic events suggest that the transition to selfing and speciation occurred within Africa, with subsequent migration out of Africa into Eurasia. However, it is also possible that the initial transition to selfing occurred within Eurasia followed by migration into Africa and concomitant loss of variation in Eurasia. This alternative would require that the ancient variation in the *A. thaliana* lineage was either lost or has not been sampled in Eurasia and the bottleneck into Africa was mild enough to preserve high levels of genetic variation.

Overall, the patterns in *A. thaliana* bear striking similarities to those observed for human populations, particularly in the larger effective population size in Africa (34), the exodus from Africa approximately 120 kya (39, 53–55), and the splitting of major human populations in Europe and Asia (approximately 45–35 kya) (53, 54). Analogous to what we propose here, demographic events in human populations have been attributed to major climate transitions (35, 39, 56).

Moreover, the timing and types of demographic events we infer during the history of *A. thaliana* are consistent with previous observations in a broad range of other plant species. Specifically, the shift to predominance of C4 plants across Africa at 1.2–0.8 Mya and the intensification of glacial cycles worldwide (57) correspond with our estimated timing of the evolution of selfing in *A. thaliana* and a clustering of speciation events more generally (58). The geographic expansion approximately 120 kya corresponds to an African pluvial and worldwide interglacial, which resulted in expansion of forests across Africa (59) and Eurasia (51, 52). Finally, we see evidence of an increase in effective population size overlapping with the most recent and well-described African pluvial at 11–5 kya, when the Sahara was heavily vegetated and filled with lakes (38, 60, 61). The concordance between inferred population size changes, climate, and reports for other species implies that the patterns we observe in *A. thaliana* may be representative of climate-mediated population dynamics across diverse taxa.

## Materials and Methods

For full materials and methods, please see *SI Appendix*, *Supplementary Text*.

We sequenced the genomes of 79 *A. thaliana* individuals, including 70 fresh samples and 9 herbarium samples (*SI Appendix*, Table S1). For fresh leaf samples, sequencing libraries were prepared using Illumina TruSeq DNA sample prep kits (Illumina) and sequenced on Illumina Hi-Seq instruments. DNA from herbarium specimens was extracted, authenticated, and treated with uracil glycosylase to remove damaged nucleotides in a clean room facility at the University of Tübingen. To align the sequences to the TAIR10 reference genome and to call variants, we used two different pipelines: the MPI-SHORE pipeline (62) and a more conservative pipeline designed to reduce false positives resulting from indels.

For population structure analyses, we subsampled the complete dataset to match sample sizes across clusters as some Eurasian geographic regions are heavily oversampled, which could cause biases in some analyses, and we pruned SNPs based on LD to select a representative set. For ADMIXTURE, the number of clusters ( $K$ ) was determined based on the outcome of cross-validation analyses.

To infer patterns of effective population size and population separations over historical time, we used a MSMC v2 (33). Because *A. thaliana* accessions are inbred, we created pseudodiploids by combining chromosomes from pairs of individuals from the same populations and ran MSMC in the two- and eight-haplotype configurations (Fig. 5). We assumed a mutation rate of  $7.1 \times 10^{-9}$  based on results of mutation accumulation experiments (40) and a generation time of 1 y. We confirmed inferences using  $\delta a \delta i$  (37) on joint site frequency spectra from pairs of populations.

- Provart NJ, et al. (2016) 50 years of *Arabidopsis* research: Highlights and future directions. *New Phytol* 209:921–944.
- Somerville C, Koornneef M (2002) A fortunate choice: The history of *Arabidopsis* as a model plant. *Nat Rev Genet* 3:883–889.
- Beck JB, Schmuths H, Schaal BA (2008) Native range genetic variation in *Arabidopsis thaliana* is strongly geographically structured and reflects Pleistocene glacial dynamics. *Mol Ecol* 17:902–915.
- Consortium G; 1001 Genomes Consortium. Electronic address: magnus.nordborg@gmi.oavv.ac.at; 1001 Genomes Consortium (2016) 1135 genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*. *Cell* 166:481–491.
- François O, Blum MG, Jakobsson M, Rosenberg NA (2008) Demographic history of European populations of *Arabidopsis thaliana*. *PLoS Genet* 4:e1000075.
- Picó FX, Méndez-Vigo B, Martínez-Zapater JM, Alonso-Blanco C (2008) Natural genetic variation of *Arabidopsis thaliana* is geographically structured in the Iberian peninsula. *Genetics* 180:1009–1021.
- Sharbel TF, Haubold B, Mitchell-Olds T (2000) Genetic isolation by distance in *Arabidopsis thaliana*: Biogeography and postglacial colonization of Europe. *Mol Ecol* 9:2109–2118.
- Lee C-R, et al. (2017) On the post-glacial spread of human commensal *Arabidopsis thaliana*. *Nat Commun* 8:14458.
- Hu TT, et al. (2011) The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. *Nat Genet* 43:476–481.
- Oyama RK, et al. (2008) The shrunken genome of *Arabidopsis thaliana*. *Plant Syst Evol* 273:257–271.
- Wright SI, Ness RW, Foxe JP, Barrett SCH (2008) Genomic consequences of outcrossing and selfing in plants. *Int J Plant Sci* 169:105–118.
- Albach DC, Greilhuber J (2004) Genome size variation and evolution in *Veronica*. *Ann Bot (Lond)* 94:897–911.
- Trivers R, Burt A, Palestis BG (2004) B chromosomes and genome size in flowering plants. *Genome* 47:1–8.
- Nasrallah ME, Yogeewaran K, Snyder S, Nasrallah JB (2000) *Arabidopsis* species hybrids in the study of species differences and evolution of amphiploidy in plants. *Plant Physiol* 124:1605–1614.
- Sherman-Broyles S, et al. (2007) S locus genes and the evolution of self-fertility in *Arabidopsis thaliana*. *Plant Cell* 19:94–106.
- Shimizu KK, Shimizu-Inatsugi R, Tsuchimatsu T, Purugganan MD (2008) Independent origins of self-compatibility in *Arabidopsis thaliana*. *Mol Ecol* 17:704–714.
- Tsuchimatsu T, et al. (2010) Evolution of self-compatibility in *Arabidopsis* by a mutation in the male specificity gene. *Nature* 464:1342–1346.
- Dwyer KG, et al. (2013) Molecular characterization and evolution of self-incompatibility genes in *Arabidopsis thaliana*: The case of the Sc haplotype. *Genetics* 193:985–994.
- Liu P, Sherman-Broyles S, Nasrallah ME, Nasrallah JB (2007) A cryptic modifier causing transient self-incompatibility in *Arabidopsis thaliana*. *Curr Biol* 17:734–740.
- Nasrallah JB (2002) Recognition and rejection of self in plant reproduction. *Science* 296:305–308.
- Boggs NA, Nasrallah JB, Nasrallah ME (2009) Independent S-locus mutations caused self-fertility in *Arabidopsis thaliana*. *PLoS Genet* 5:e1000426.
- Vekemans X, Poux C, Goubet PM, Castric V (2014) The evolution of selfing from outcrossing ancestors in Brassicaceae: What have we learned from variation at the S-locus? *J Evol Biol* 27:1372–1385.
- Stebbins GL (1974) *Flowering Plants: Evolution Above the Species Level* (Harvard Univ Press, Cambridge, MA).
- Bechsgaard JS, Castric V, Charlesworth D, Vekemans X, Schierup MH (2006) The transition to self-compatibility in *Arabidopsis thaliana* and evolution within S-haplotypes over 10 Myr. *Mol Biol Evol* 23:1741–1750.
- Tang C, et al. (2007) The evolution of selfing in *Arabidopsis thaliana*. *Science* 317:1070–1072.
- Shimizu KK, Tsuchimatsu T (2015) Evolution of selfing: Recurrent patterns in molecular adaptation. *Annu Rev Ecol Syst* 46:593–622.
- Hoffmann MH (2002) Biogeography of *Arabidopsis thaliana* (L.) Heynh. (Brassicaceae). *J Biogeogr* 29:125–134.
- Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 19:1655–1664.
- Brennan AC, et al. (2014) The genetic structure of *Arabidopsis thaliana* in the southwestern Mediterranean range reveals a shared history between North Africa and southern Europe. *BMC Plant Biol* 14:17.
- Désamolé A, et al. (2011) Out of Africa: Northwestwards Pleistocene expansions of the heather *Erica arborea*. *J Biogeogr* 38:164–176.
- Quézel P (1978) Analysis of the flora of Mediterranean and Saharan Africa. *Ann Mo Bot Gard* 65:479–534.
- Tsuchimatsu T, et al. (April 4, 2017) Patterns of polymorphism at the self-incompatibility locus in 1,083 *Arabidopsis thaliana* genomes. *Mol Biol Evol*, 10.1093/molbev/msx122.
- Schiffels S, Durbin R (2014) Inferring human population size and separation history from multiple genome sequences. *Nat Genet* 46:919–925.
- Auton A, et al.; 1000 Genomes Project Consortium (2015) A global reference for human genetic variation. *Nature* 526:68–74.
- Osborne AH, et al. (2008) A humid corridor across the Sahara for the migration of early modern humans out of Africa 120,000 years ago. *Proc Natl Acad Sci USA* 105:16444–16447.
- Quante M (2010) The changing climate: Past, present, future. *Relict Species: Phylogeography and Conservation Biology*, eds Habel JC, Assmann T (Springer, Berlin).
- Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD (2009) Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet* 5:e1000695.
- Tierney JE, deMenocal PB (2013) Abrupt shifts in Horn of Africa hydroclimate since the Last Glacial Maximum. *Science* 342:843–846.
- Timmermann A, Friedrich T (2016) Late Pleistocene climate drivers of early human migration. *Nature* 538:92–95.
- Ossowski S, et al. (2010) The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science* 327:92–94.
- Cerling TE, Hay RL (1986) An isotopic study of paleosol carbonates from Olduvai Gorge. *Quat Res* 25:63–78.
- deMenocal PB (2004) African climate change and faunal evolution during the Pliocene-Pleistocene. *Earth Planet Sci Lett* 220:3–24.
- Hohmann N, Wolf EM, Lysak MA, Koch MA (2015) A time-calibrated road map of Brassicaceae species radiation and evolutionary history. *Plant Cell* 27:2770–2784.
- Brandvain Y, Kenney AM, Flagel L, Coop G, Sweigart AL (2014) Speciation and introgression between *Mimulus nasutus* and *Mimulus guttatus*. *PLoS Genet* 10:e1004410.
- Brandvain Y, Slotte T, Hazzouri KM, Wright SI, Coop G (2013) Genomic identification of founding haplotypes reveals the history of the selfing species *Capsella rubella*. *PLoS Genet* 9:e1003754.
- Foxe JP, et al. (2009) Recent speciation associated with the evolution of selfing in *Capsella*. *Proc Natl Acad Sci USA* 106:5241–5245.
- Guo YL, et al. (2009) Recent speciation of *Capsella rubella* from *Capsella grandiflora*, associated with loss of self-incompatibility and an extreme bottleneck. *Proc Natl Acad Sci USA* 106:5246–5251.
- Sicard A, Lenhard M (2011) The selfing syndrome: A model for studying the genetic and evolutionary basis of morphological adaptation in plants. *Ann Bot (Lond)* 107:1433–1443.
- Johnston JS, et al. (2005) Evolution of genome size in Brassicaceae. *Ann Bot (Lond)* 95:229–235.
- Mable BK, et al. (2017) What causes mating system shifts in plants? *Arabidopsis lyrata* as a case study. *Heredity (Edinb)* 118:110.
- Kaspar F, Kühl N, Cubasch U, Litt T (2005) A model-data comparison of European temperatures in the Eemian interglacial. *Geophys Res Lett* 32:L11703.
- Kukla GJ, et al. (2002) Last interglacial climates. *Quat Res* 58:2–13.
- Henn BM, Cavalli-Sforza LL, Feldman MW (2012) The great human expansion. *Proc Natl Acad Sci USA* 109:17758–17764.
- Mallick S, et al. (2016) The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* 538:201–206.
- Pagani L, et al. (2016) Genomic analyses inform on migration events during the peopling of Eurasia. *Nature* 538:238–242.
- deMenocal PB, Stringer C (2016) Human migration: Climate and the peopling of the world. *Nature* 538:49–50.
- Tzedakis PC, Crucifix M, Mitsui T, Wolff EW (2017) A simple rule to determine which insolation cycles lead to interglacials. *Nature* 542:427–432.
- deMenocal PB (1995) Plio-Pleistocene African climate. *Science* 270:53–59.
- Dupont L (2011) Orbital scale vegetation change in Africa. *Quat Sci Rev* 30:3589–3602.
- Cohmap M; COHMAP MEMBERS (1988) Climatic changes of the last 18,000 years: Observations and model simulations. *Science* 241:1043–1052.
- Kuper R, Kröpelin S (2006) Climate-controlled Holocene occupation in the Sahara: Motor of Africa's evolution. *Science* 313:803–807.
- Ossowski S, et al. (2008) Sequencing of natural strains of *Arabidopsis thaliana* with short reads. *Genome Res* 18:2024–2033.
- Kawamura K, et al. (2007) Northern Hemisphere forcing of climatic cycles in Antarctica over the past 360,000 years. *Nature* 448:912–916.