

PLANT SCIENCES

A two-step adaptive walk rewires nutrient transport in a challenging edaphic environment

Emmanuel Tergemina^{1*}, Ahmed F. Elfarargi¹, Paulina Flis², Andrea Fulgione¹, Mehmet Göktay¹, Célia Neto^{1†}, Marleen Scholle³, Pádraic J. Flood^{1‡}, Sophie-Asako Xerri¹, Johan Zicola^{1§}, Nina Döring¹, Herculano Dinis^{4,5}, Ute Krämer³, David E. Salt², Angela M. Hancock^{1*}

Most well-characterized cases of adaptation involve single genetic loci. Theory suggests that multilocus adaptive walks should be common, but these are challenging to identify in natural populations. Here, we combine trait mapping with population genetic modeling to show that a two-step process rewired nutrient homeostasis in a population of *Arabidopsis* as it colonized the base of an active stratovolcano characterized by extremely low soil manganese (Mn). First, a variant that disrupted the primary iron (Fe) uptake transporter gene (*IRT1*) swept quickly to fixation in a hard selective sweep, increasing Mn but limiting Fe in the leaves. Second, multiple independent tandem duplications occurred at *NRAMP1* and together rose to near fixation in the island population, compensating the loss of *IRT1* by improving Fe homeostasis. This study provides a clear case of a multilocus adaptive walk and reveals how genetic variants reshaped a phenotype and spread over space and time.

INTRODUCTION

Determining the genetic mechanisms underlying adaptation to challenging environments is a central goal in biology; it also has practical applications for a broad range of critical issues in agriculture, conservation, and medicine. Potential uses include developing sustainable crops tailored to local environments, managing vulnerable populations and species, and mitigating emerging risks from pathogens.

While several cases of single-gene effects on putatively adaptive phenotypes have been described, e.g., (1–5), theory and data suggest that most traits are polygenic. Thus, after colonization of a novel environment, adaptation is likely to involve changes at multiple loci, producing multistep walks toward higher fitness (6–11). However, we lack concrete examples of multilocus processes in wild populations. As a result, there are many open questions. For example, are multilocus adaptive processes likely to involve proteins that interact at the molecular level or those that have distinct cellular functions? To what extent does the first evolutionary step limit subsequent steps? Do epistasis and pleiotropy limit evolutionary paths? Does the mode of adaptation fit a “hard selective sweep” model (12), where a single mutation rises rapidly to fixation in the population, or does it occur through “soft selective sweeps” (13, 14), where multiple mutations rise in frequency as a group? To address these questions in the context of multistep adaptive processes, we need detailed information about specific cases, including the precise functional variants involved,

their physiological consequences, and reconstructions of their spread over time and space.

We capitalize on a case where *Arabidopsis* recently colonized an active stratovolcano with a harsh edaphic environment to examine the genetic steps that led to adaptation. *Arabidopsis* colonized Fogo ~3 to 5 ka from Santo Antão, an island ~250 km to the north (Fig. 1A) (15). Fogo and Santo Antão are both volcanic islands in the Cape Verde archipelago, but the two islands have different pedogenic histories. While Fogo is an active stratovolcano, which experienced several eruptions over the past few thousand years, including the most recent in 2015, the last eruption on Santo Antão was much more ancient (~90 ka) (16). Today, the two island populations are genetically distinct from one another and from the continental populations due to severe colonization bottlenecks, which eliminated 99.4% of all shared variation, resulting in a phylogenetic speciation event (15).

RESULTS

Fogo plants have a distinct leaf ionome compared to Santo Antão and Morocco

When we propagated Fogo accessions in standard growth conditions, we found that they exhibited varying degrees of chlorosis, a phenotype we had not observed in native habitats (Fig. 1B). Quantitative image analysis showed that Fogo accessions were less green and smaller than their closest relatives in Santo Antão [Kruskal-Wallis (KW), greenness: $P = 3.72 \times 10^{-30}$, segmented area: $P = 3.98 \times 10^{-29}$] and Morocco (KW, greenness: $P = 1.49 \times 10^{-18}$, segmented area: $P = 1.15 \times 10^{-18}$) (Fig. 1C), and both traits were strongly correlated with relative chlorophyll content (greenness: Pearson's $R^2 = 0.56$, $n = 31$, $P = 1.38 \times 10^{-6}$; segmented area: Pearson's $R^2 = 0.64$, $n = 31$, $P = 6.72 \times 10^{-8}$) (fig. S1). In the most affected accessions, photosynthetic efficiency (Phi2) was severely reduced (fig. S2). Because chlorosis is often caused by mineral nutrient imbalance, we investigated the leaf ionome to determine whether the Fogo population showed differences relative to Santo Antão and Morocco. The concentrations of several mineral nutrients and trace elements were markedly shifted in Fogo accessions (fig. S3), resulting in a clear separation in

Copyright © 2022 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution License 4.0 (CC BY).

¹Department of Plant Developmental Biology, Max Planck Institute for Plant Breeding Research, 50829 Cologne, Germany. ²Future Food Beacon of Excellence and the School of Biosciences, University of Nottingham, Sutton Bonington Campus, Nr Loughborough, LE12 5RD Nottingham, UK. ³Faculty of Biology and Biotechnology, Ruhr University Bochum, 44801 Bochum, Germany. ⁴Parque Natural do Fogo, Direção Nacional do Ambiente, 115 Chã d'Areia, Praia, Santiago, Cabo Verde, Africa. ⁵Associação Projecto Vító, 8234, Xaguate, Cidade de São Filipe, Fogo, Cabo Verde, Africa. *Corresponding author. Email: hancock@mpipz.mpg.de (A.M.H.); tergemina@mpipz.mpg.de (E.T.)

†Present address: Department of Plant and Environmental Sciences, University of Copenhagen, Højbakkegård Allé 30, 2630 Taastrup, Copenhagen, Denmark.

‡Present address: Infarm—Indoor Urban Farming GmbH, Colditzstraße 30, 12099 Berlin, Germany.

§Present address: Georg-August-University Göttingen, Department of Crop Sciences, 37075 Göttingen, Germany.

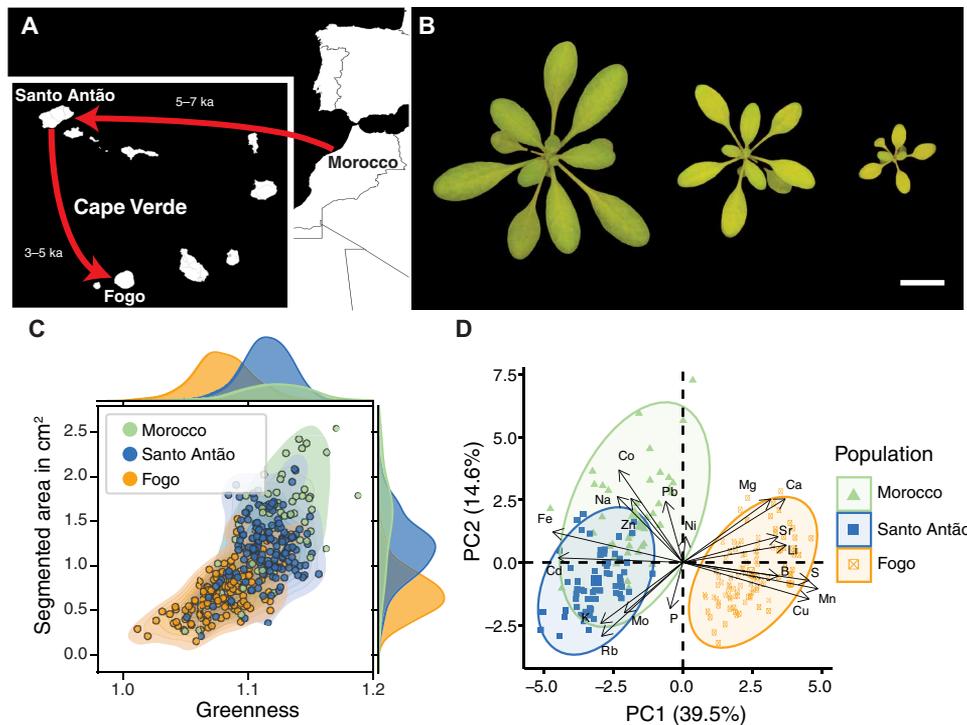


Fig. 1. Fogo plants have a distinct leaf ionome, characterized by Mn enrichment and Fe deficiency. (A) Schematic of the colonization of Fogo from Santo Antão. (B) Representative lines from Fogo showing high variation in size and greenness on standard potting mix. Scale bar, 1 cm. (C) Bivariate distribution of greenness (green/red) and segmented area (cm²) by population [Fogo (orange), Santo Antão (blue), and Morocco (green)]. Phenotype distributions are shown on the axes. (D) Fogo samples separate from Moroccan and Santo Antão samples on the first principal component of a matrix of leaf elemental concentrations, color-coded as in (C). In (C) and (D), each dot corresponds to the median value across replicates.

multivariate clustering [principal components analysis (PCA)] (Fig. 1D). Mn and Fe were the major contributors to this separation, with loadings of 11.53 and 10.73% on PC1, which explained 39.5% of the variation (fig. S4 and table S1). Mn levels were ~4-fold higher in leaves of Fogo plants compared to those from Santo Antão and Morocco [Fogo mean = 501.27 $\mu\text{g g}^{-1}$ dry weight (DW) \pm 5.62 standard error (SE); Santo Antão mean = 133.37 $\mu\text{g g}^{-1}$ DW \pm 1.47 SE; Morocco mean = 133.81 $\mu\text{g g}^{-1}$ DW \pm 2.20 SE]. Consistent with the observed chlorosis, Fe levels were significantly lower in Fogo leaves compared to plants from Santo Antão and Morocco (Fogo mean = 52.73 $\mu\text{g g}^{-1}$ DW \pm 0.45 SE; Santo Antão mean = 79.90 $\mu\text{g g}^{-1}$ DW \pm 0.84 SE; Morocco mean = 79.11 $\mu\text{g g}^{-1}$ DW \pm 0.79 SE) (fig. S3 and table S2).

The parent material from which volcanic soils develop results from magmatism and often has chemical characteristics that limit plant growth. These include high phosphorous retention and micronutrient insufficiencies (17). To characterize the nutrient composition of the native soil, we collected topsoil samples from *Arabidopsis* stands across Fogo and Santo Antão. Topsoil analyses indicated that, in comparison to Santo Antão, Fogo soils are more alkaline (Fogo pH mean = 7.7; Santo Antão pH mean = 7) and impoverished in several mineral nutrients (tables S3 and S4). Among the 22 water-extractable elements tested, Fogo and Santo Antão soils differed most significantly in Co, Fe, K/Rb, Ni, Mn, and P (fig. S5 and table S4). To more directly estimate the availability of elements in the native soil, we characterized leaf ionomes of Santo Antão (S1-1) and Fogo (F13-8) accessions grown on Fogo and Santo Antão soils in comparison to

standard potting mix. Mn levels were very low for plants grown on Fogo soil (S1-1 mean = 44.18 $\mu\text{g g}^{-1}$ DW \pm 9.82 SE, F13-8 mean = 72.30 $\mu\text{g g}^{-1}$ DW \pm 3.73 SE) compared to Santo Antão soil (S1-1 mean = 101.27 $\mu\text{g g}^{-1}$ DW \pm 13.93 SE, F13-8 mean = 114.98 $\mu\text{g g}^{-1}$ DW \pm 8.63 SE) and standard potting mix (S1-1 mean = 155.34 $\mu\text{g g}^{-1}$ DW \pm 7.64 SE, F13-8 mean = 208.10 $\mu\text{g g}^{-1}$ DW \pm 6.11 SE) (fig. S6) and were near the critical Mn deficiency concentration observed in plants (20 $\mu\text{g g}^{-1}$ DW) (18). Our results show that Fogo *Arabidopsis* grows on soils with low bioavailable Mn. Further, the enrichment of Mn in the leaves of Fogo plants suggests that the Fogo plants have adapted to enhance Mn transport into leaf tissues under Mn limitation imposed by the Fogo soil.

Mapping reveals a loss-of-function variant of *IRT1* and a segmental tandem triplication at *NRAMP1*

To investigate the genetic basis of the Fogo chlorosis/small-size phenotype in standard potting mix, we examined segregation of the trait in an F2 population created from a cross between a Fogo accession (F13-8) and the *Arabidopsis thaliana* reference accession (Col-0). We found that roughly 1 of 16 individuals were highly chlorotic [χ^2 (1, n = 454) = 0.41, P = 0.52] (Fig. 2, A and B), suggesting that a two-locus model could explain the chlorotic phenotype. We conducted bulk segregant analysis (BSA) by sequencing five pools of Col-0 \times F13-8 F2s with varying degrees of chlorosis (fig. S7). Sequencing of the most chlorotic pool revealed large deviations from the expected frequencies at two genomic regions. Severe chlorosis was linked to the combination of the Col-0 region at the end of

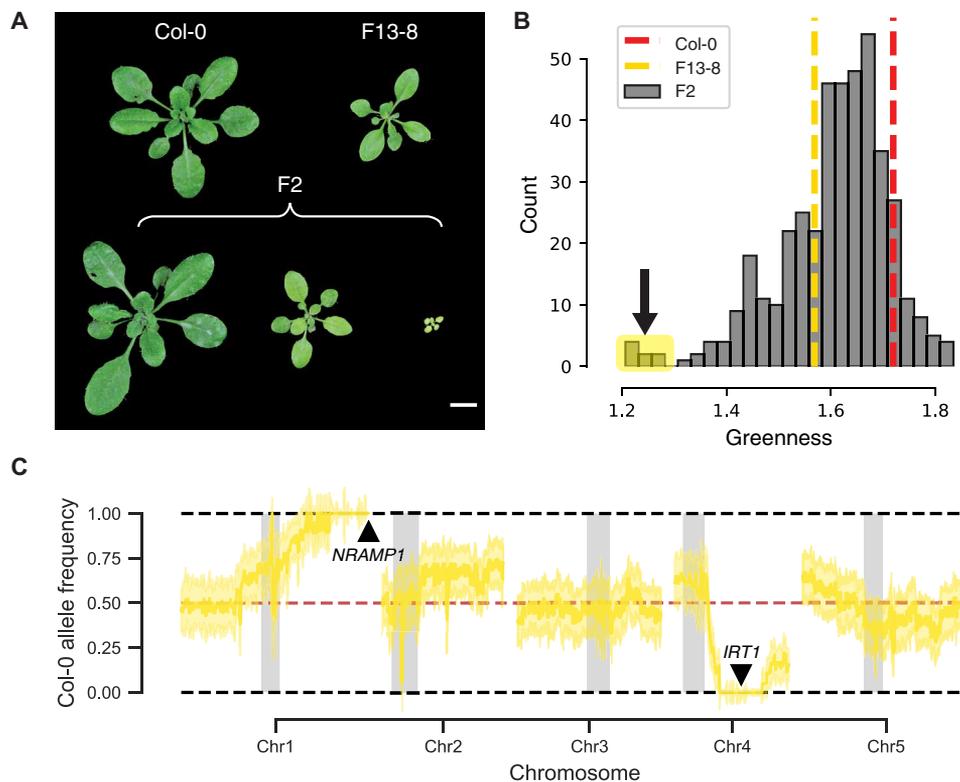


Fig. 2. Mapping reveals a loss-of-function variant of *IRT1* and a segmental tandem triplication at *NRAMP1*. (A) F2 progeny derived from a Col-0 × F13-8 cross vary in size and greenness on standard potting mix. Scale bar, 1 cm. (B) Quantitative measure of greenness in 454 F2 progeny reveals a subset of extremely chlorotic segregants. Red and yellow dashed lines represent the averages for replicates of Col-0 and F13-8, respectively. The arrow and yellow rectangle indicate the subpopulation of highly chlorotic plants used for BSA in (C). (C) Chlorotic individuals are fixed for the reference (Col-0) *NRAMP1* allele and the Fogo (F13-8) *IRT1* allele. Median allele frequency using 200-kb sliding windows with a step size of 5 kb (yellow line) and corresponding SD (yellow shade). Red dashed line indicates 0.5 allele frequency. Gray rectangles specify centromeric regions.

chromosome 1 and the F13-8 region on chromosome 4 (Fig. 2C and fig. S7). We next examined the BSA peaks for potential causative variants. No single-nucleotide polymorphism (SNP) variant stood out as a strong functional candidate in the chromosome 1 region, but in the chromosome 4 region, we identified an SNP in the first exon of *IRON REGULATED TRANSPORTER 1* (*IRT1*), leading to a premature stop codon at position 130 (*IRT1* G130X) (fig. S8 and tables S5 and S6). *IRT1* encodes a divalent cation transporter that primarily contributes to Fe uptake at the root surface in *A. thaliana* and therefore represents an excellent candidate. Indeed, *irt1* mutants exhibit severe chlorosis (19–21), reminiscent of the phenotype we observed in the most chlorotic pool of Col-0 × F13-8 segregants, supporting a causative role of *IRT1* 130X in the chlorotic phenotype.

Next, we more closely examined the BSA peak in the chromosome 1 region and found a ~15-kb region surrounding *NATURAL RESISTANCE-ASSOCIATED MACROPHAGE PROTEIN 1* (*NRAMP1*), where read coverage was ~3-fold higher than the surrounding genomic region, suggesting gene copy number variation (CNV) (fig. S9 and table S7). To confirm this, we generated a whole-genome de novo assembly with Nanopore sequencing technology for F13-8 and identified a segmental tandem triplication at the *NRAMP1* region (fig. S9B). We asked whether the CNV was associated with increased *NRAMP1* expression in root tissue and found (~2.2-fold) higher *NRAMP1* mRNA levels in F13-8 compared to Col-0 (*t* test, $P = 3.0 \times 10^{-3}$) (fig. S9C). *NRAMP1* encodes a metal transporter crucial for growth in

Mn-deficient conditions that cooperates with *IRT1* for Fe transport (22–24). To determine whether the *NRAMP1* protein in Fogo may show a different substrate spectrum compared to the Santo Antão outgroup population, we compared *NRAMP1* protein sequences in Fogo and Santo Antão plants to the Col-0 reference. We found no segregating amino acid variation at *NRAMP1* within Cape Verde and only a single amino acid change between the Cape Verdean accessions and Col-0 (L202I), which is unlikely to change its transport activity (fig. S10). Overall, our results show that the *NRAMP1* CNV is a strong candidate to explain the chromosome 1 peak.

We further validated that these two loci could largely explain the observed phenotypic variation in the Col-0 × F13-8 F2 population by genotyping each F2 individual at *IRT1* and *NRAMP1* [analysis of variance (ANOVA), $F = 23.5031$, $P < 2.2 \times 10^{-16}$] (fig. S11 and table S8). Further, the *IRT1* and *NRAMP1* mutations exhibited epistasis such that the effect of the *NRAMP1* CNV was only apparent in *IRT1* 130X homozygotes (table S9). Together, the BSA results provided compelling explanations for the Fe deficiency and the Mn enrichment observed in the leaves of the Fogo plants.

Loss of *IRT1* exerts strong pleiotropic effects on nutrient transport in the Fogo natural population, increasing Mn in the leaves but inducing Fe deficiency

Mapping in the recombinant population provided functional loci in a single Fogo individual (F13-8). We next examined the frequencies

and effects of these loci in the broader context of the population and species. We found that *IRT1* 130X is fixed in Fogo and absent in the closest outgroup populations (Santo Antão and Morocco). Moreover, we found no evidence for any other *IRT1* loss-of-function variant in worldwide polymorphism data (table S10) (25–28). Given that plant growth in *irt1* mutants is severely altered (19–21), the fixation of the *IRT1* truncation in Fogo was unexpected and led us to investigate its physiological effect as it occurred in Fogo.

To achieve this, we used the CRISPR-Cas9 system to generate a knockout allele of *IRT1* in S1-1, an accession from Santo Antão and the closest outgroup, which carries both the ancestral *IRT1* (functional) and *NRAMP1* (single-copy) haplotype (Fig. 3A and fig. S12). On standard potting mix, S1-1-*irt1* plants were chlorotic, but the severity of this phenotype was milder than the Col-0-*irt1* mutant, suggesting that other loci contribute to the difference between Col-0-*irt1* and S1-1-*irt1* (figs. S13 and S14). This genetic background effect may not be unexpected given the high divergence between the Cape Verde Islands (CVI) accessions and Eurasians, including the Col-0 accession. While outside of the scope of this study, further examination of variation in the impact of *IRT1* loss in diverse *A. thaliana* accessions would be warranted. We found that supplementing S1-1-*irt1* with Fe reduces chlorosis in standard potting soil (fig. S14). As expected, loss of *IRT1* in a Santo Antão background caused a significant reduction in leaf Fe [Mann-Whitney-Wilcoxon (MMW) test, $P = 4.06 \times 10^{-3}$] (Fig. 3B and figs. S13, S15, and S16), but it also caused

an unexpected ~3-fold increase in leaf levels of Mn (MMW test, $P = 4.06 \times 10^{-3}$) (Fig. 3B), an element that is limited in native Fogo soils (fig. S5). A PCA based on 19 elements in the leaves revealed that the loss of *IRT1* in a Santo Antão background strongly contributes to the multivariate divergence of the Fogo leaf ionome profile (Fig. 3C). Together, our results suggest that loss of *IRT1* was advantageous when it arose in Fogo because it increased the level of Mn in the leaves, which was severely limiting; however, *IRT1* exerts strong pleiotropic effects, and its loss of function reduced the potential for Fe transport.

Multiple independent segmental duplications arose at *NRAMP1*, reestablishing Fe homeostasis

At *NRAMP1*, we found evidence for three independent tandem duplication (TD) events segregating in the Fogo population, each with distinct and nonoverlapping breakpoints in the region (Fig. 4A and S17). These TDs result from two different microhomology-mediated junctions (CT and ATG) and one 10-bp (base pair) templated junction (AAGACATAA) and are referred to below as “CT-TD,” “ATG-TD,” and “AAGACATAA-TD” (Fig. 4A and figs. S18 to S21). This large number of independent TD events was striking based on the short time since colonization of Fogo (~3 to 5 ka) and low SNP diversity [$\theta_{w(\text{Fogo})} = 8.93 \times 10^{-5}$]. The genomic location of *NRAMP1* may contribute to the high TD rate in the region as this gene is located in the subtelomeric region (~50 kb from the telomere), where

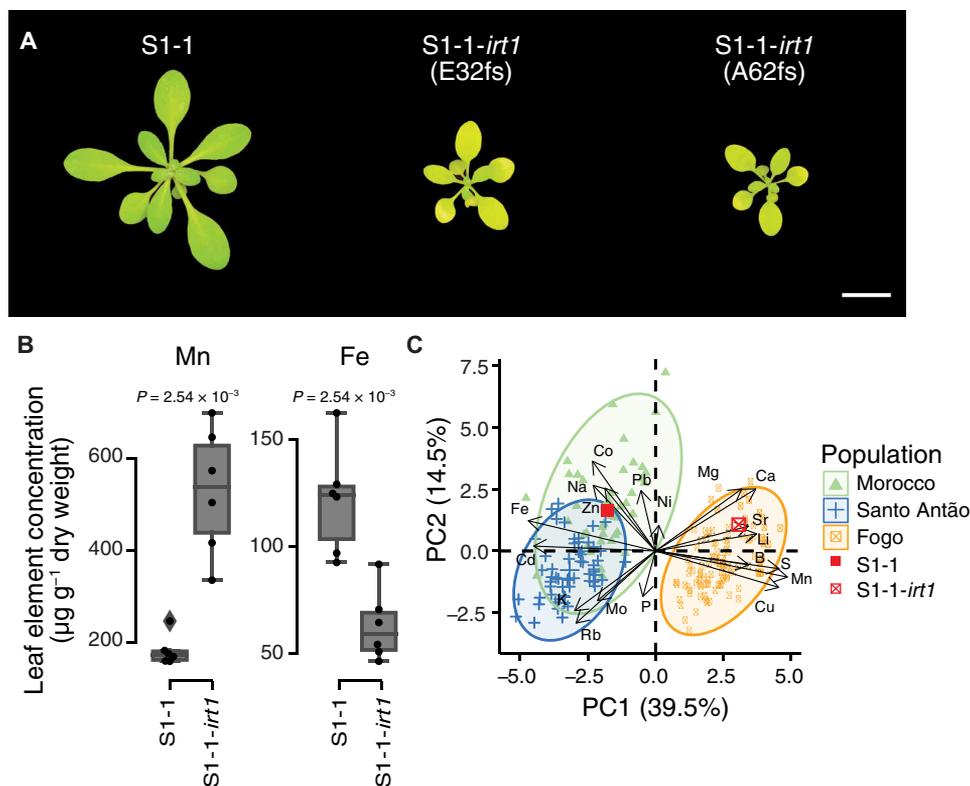


Fig. 3. The loss of *IRT1* exerts strong pleiotropic effects on nutrient transport in Fogo, increasing Mn in the leaves but inducing Fe deficiency. CRISPR to disrupt *IRT1* function in a Santo Antão individual (S1-1) exhibits (A) chlorosis and (B) increased Mn and reduced Fe in leaves, consistent with a strong role for the loss of *IRT1* in overall pattern of differences observed between Santo Antão and Fogo populations. In (B), each dot represents one replicate per genotype. $P = P$ value for MMW test. (C) CRISPR-induced loss of *IRT1* in a Santo Antão individual (S1-1) results in a shift into the Fogo cloud. Each dot represents the median value across replicates. In (B) and (C), S1-1-*irt1* corresponds to S1-1-*irt1*(E32fs).

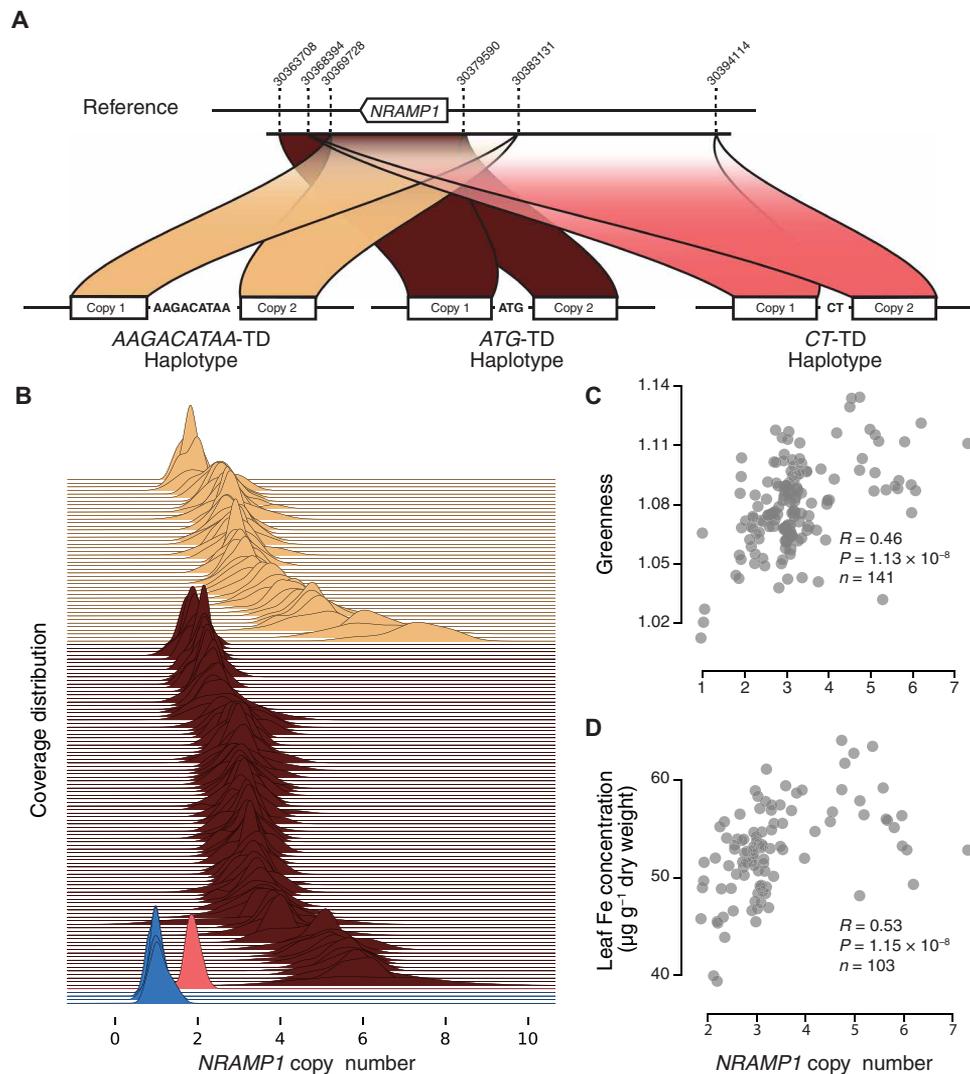


Fig. 4. Multiple independent TDs arise at *NRAMP1*, reestablishing Fe homeostasis. (A) Diagram of the three independent TD events at *NRAMP1* in Fogo. Breakpoint junctions are indicated in red. Dashed lines correspond to the boundaries for each TD. The numbers above the dashed line indicate the position on chromosome 1 of the TAIR10 reference genome. (B) Both major *NRAMP1* haplotypes exhibit CNV across lines. Coverage distribution between the breakpoint junctions using 800-bp sliding windows with a step size of 400 normalized to the 3-kb surrounding regions. Haplotypes are color-coded on the basis of the schematic in (A); blue denotes plants with only a single *NRAMP1* copy. *NRAMP1* CN is positively correlated with (C) greenness and (D) leaf Fe concentration in Fogo plants. *R*, Spearman's rho; *n*, number of measurements.

recombination rate and thus the probability of TD events may be increased (29). However, we found no TDs in or around *NRAMP1* in Santo Antão accessions (15) and very little evidence in previously sequenced worldwide *A. thaliana* genomes (25–28). Only two accessions harbored any evidence for TDs in the region, and only one of these (in Oua-0) shows what appears to be an intact TD containing the entire *NRAMP1* coding sequence (fig. S22), indicating that if TD rates in this subtelomeric region are high in general, these are normally subject to strong purifying selection. Overall, our results indicate that the presence of three independent TD events at *NRAMP1* in Fogo is extremely unusual.

Several studies have found evidence that gene CNV plays an important role in evolution (30–32). To estimate the *NRAMP1* copy number (CN) across Fogo accessions, we first analyzed the coverage between the breakpoints compared to the surrounding region and

then confirmed a subset of these using quantitative digital polymerase chain reaction (PCR) analysis (Fig. 4B, fig. S23, and table S11). We also generated whole-genome de novo assemblies for five additional Fogo genomes to further validate the breakpoint junctions and CN estimates from short-read sequencing data (fig. S24). Together, our results showed that the common ATG-TD and AAGACATAA-TD haplotypes varied in *NRAMP1* CN from two to seven, and the single CT-TD haplotype had two *NRAMP1* copies (fig. S23 and table S12). We also identified four Fogo accessions, all from a single stand (F10-13), with the ancestral *NRAMP1* haplotype (single *NRAMP1* copy).

We next investigated the impacts of *NRAMP1* CN on expression of the gene and physiological traits. We grew seedlings on medium and collected roots to assay *NRAMP1* expression using probe-based quantitative digital PCR (dPCR) analyses. Using multiple linear regression to account for treatment and IRT1 status, we found that

NRAMP1 CN largely explains variation in expression ($\beta_{\text{CNV}} = 1.06$, $t = 11.2$, $P < 1.83 \times 10^{-14}$; model adjusted $R^2 = 0.819$, $P < 2.2 \times 10^{-16}$) (fig. S25 and table S13). This experiment further showed that the loss of *IRT1* causes an increase in *NRAMP1* expression in the presence of Mn ($\beta_{\text{IRT1}} = 0.823$, $t = 3.343$, $P < 1.7 \times 10^{-3}$). We further assessed whether *NRAMP1* CN also affects physiological traits. On standard potting mix, *NRAMP1* CN in Fogo plants correlated strongly with greenness (Spearman's $\rho = 0.46$, $n = 141$, $P = 1.13 \times 10^{-8}$) (Fig. 4C), segmented rosette area (Spearman's $\rho = 0.44$, $n = 141$, $P = 5.50 \times 10^{-8}$), and relative chlorophyll content (Spearman's $\rho = 0.78$, $n = 31$, $P = 2.96 \times 10^{-7}$) (fig. S26). Among the 19 elements tested in the leaves, *NRAMP1* CN was most strongly correlated with Fe (Spearman's $\rho = 0.53$, $n = 103$, $P = 1.15 \times 10^{-8}$) (Fig. 4D and fig. S27). Further, using a mixed model association analysis (33) in the population sample, we found that *NRAMP1* CN explained 87.2% of the heritable variance in leaf Fe concentration (table S14). In contrast, *NRAMP1* CN had no effect on the heritable variance in leaf Mn concentration. In addition, F10-13-1, which carries only one *NRAMP1* copy, accumulated more Mn in the leaves compared to S1-1, further supporting the indirect causal effect of the loss of *IRT1* on the Mn enrichment observed in the leaves of the Fogo plants (fig. S13). Together, these results indicate that CN expansion at *NRAMP1* partially compensates for Fe deficiency resulting from *IRT1* 130X in

Fogo, consistent with the previously demonstrated role of *NRAMP1* on Fe uptake (22, 23).

The Fogo population rapidly expanded and spread across the island

To reconstruct the emergence and evolutionary dynamics of *IRT1* 130X and the *NRAMP1* TDs in Fogo, it is first necessary to understand the genome-wide demographic history of *Arabidopsis* in Fogo. Previously, we showed that Fogo was colonized from Santo Antão through a strong bottleneck that removed nearly all genetic variation, and that the Fogo population remained constrained in size and panmictic for ~900 years (15). Here, we examined relationships between subpopulations within Fogo. In PCA, we found that the genetic structure of the three major subpopulations (Monte Velha, Inferno, and Lava) closely recapitulated the geographic structure (Fig. 5, A and B, and fig. S28). To determine the order and timing of splits among the three subpopulations, we inferred their phylogenetic relationships across the genome by topology weighting (34). The dominant tree topology indicated that Monte Velha represents the most ancestral subpopulation, from which migrants spread into the Inferno and Lava regions (Fig. 5C and fig. S29). However, the two other possible tree topologies were nearly as common, creating a near polytomy and suggesting abrupt dispersal and divergence of the lineages. We next

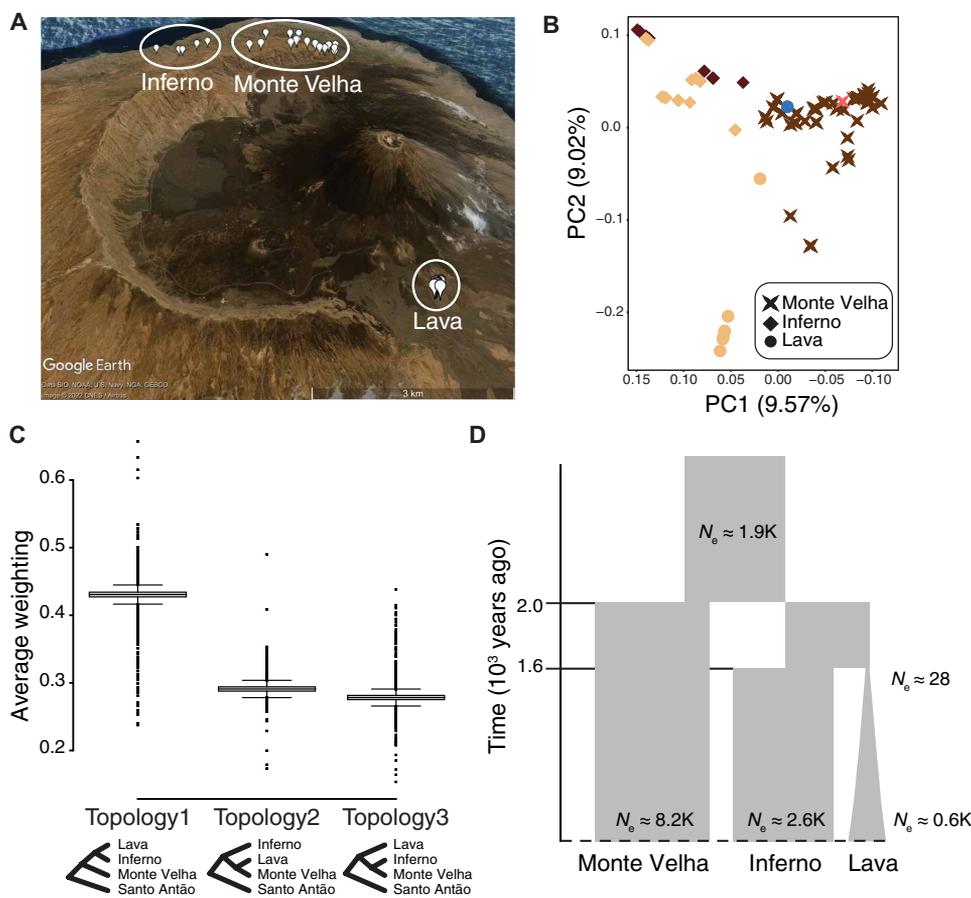


Fig. 5. The Fogo population is strongly structured because of rapid spread across the island. (A) *Arabidopsis* collection locations overlaid on a satellite image of Fogo. The three major geographic regions are shown: Monte Velha, Inferno, and Lava. (B) Genetic structure based on the first two principal components of a linkage disequilibrium (LD)-pruned SNP matrix; color-coded with the *NRAMP1* haplotypes. (C) Genome-wide variation in tree topologies using Santo Antão as outgroup. Boxplots show relative weighting across the genomes. (D) Schematic of the inferred history of *Arabidopsis* in Fogo. N_e corresponds to effective population size.

conducted modeling based on joint site frequency spectra (JSFS) between subpopulations, which also showed evidence for similar split times across the three subpopulation pairs (~1.6 to 2.0 ka) (Fig. 5D). However, the JSFS showed much higher shared variation between Monte Velha and Inferno compared to the geographically isolated Lava population, indicating a larger founding population size of the Inferno population or gene flow since the initial Monte Velha–Inferno split (fig. S30). Together, these results indicate that after remaining a geographically connected panmictic population for ~900 years, *Arabidopsis* spread rapidly across the island, fragmenting into distinct subpopulations by ~2 ka.

A two-step adaptive walk rewired nutrient transport

We next examined the specific evolutionary histories of the *IRT1* and *NRAMP1* mutations. For *IRT1* 130X, fixation could have occurred by random genetic drift as a by-product of the bottleneck in Fogo. However, given the strong physiological effect on Mn levels and the very low available Mn in Fogo soils, it is also possible that the variant was fixed due to positive selection. To differentiate between selective and neutral models at *IRT1*, we applied a test for hard selective sweeps. Specifically, we used a statistic that compares the ratio of the likelihood of a sweep versus the likelihood of neutrality across the genome while accounting for local recombination rate (35). Overall, we found unexpectedly little evidence for hard selective sweeps across the genome (Fig. 6A), especially given the large proportion of variants inferred to be fixed by selection during the

initial phase after colonization (15). This discrepancy may be due to the relatively short time a hard sweep signal is expected to persist in a population after fixation (15, 36). The highest composite likelihood ratio (CLR) peak genome-wide localized to the *IRT1* region, where the likelihood of a sweep was ~47× greater than neutrality (Fig. 6A, fig. S31, and table S15). To infer the timing and strength of selection on *IRT1*, we conducted coalescent reconstruction of the region (37). We inferred that *IRT1* 130X arose ~2.76 ka [time to most recent ancestor (tMRCA); 95% confidence interval (CI): 2.09 to 3.29 ka], before the Monte Velha, Inferno, and Lava subpopulations split, and rapidly rose to fixation in Fogo (Fig. 6B and table S16). The coalescent tree for the region also provides information about the expected rate of the allele frequency change over time. From this, we estimated a selection coefficient of 13.1% for *IRT1* 130X (Fig. 6B and table S17) (38). Overall, the historical reconstructions at *IRT1* are consistent with a hard selective sweep acting before the Fogo population split into subpopulations.

As a group, the *NRAMP1* TD haplotypes are nearly fixed in Fogo, with a combined frequency of 98%, and they exhibit strong structure across subpopulations. Two haplotypes are common (ATG-TD and AAGACATAA-TD), while the CT-TD haplotype is present in a single Monte Velha individual. The single-copy *NRAMP1* haplotype is exclusive to F10-13, a population collected in the Lava region but clustering closely with Monte Velha (Fig. 5B and fig. S28). The two common haplotypes exhibit clear geographic patterns, with ATG-TD fixed in Monte Velha (except for the single CT-TD

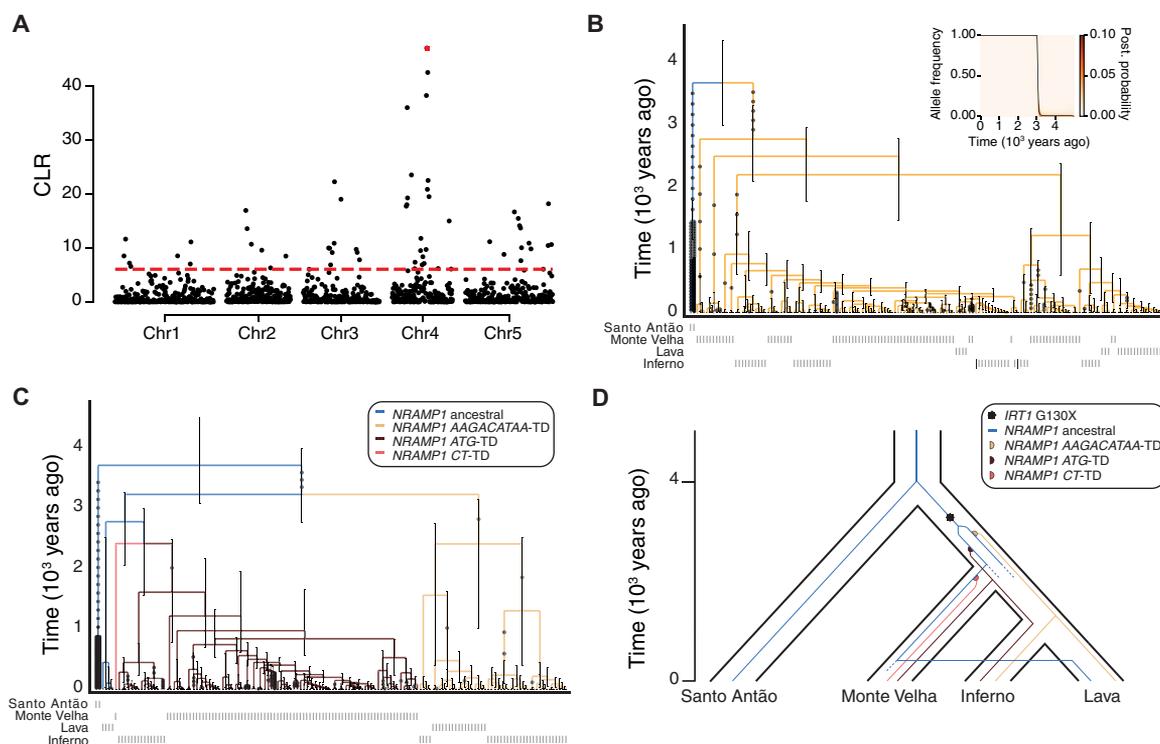


Fig. 6. A two-step adaptive walk rewired nutrient transport. (A) The highest peak in a scan to identify selective sweeps in Fogo (CLR) pinpoints the *IRT1* region. CLR with nonoverlapping window of 100 kb. The red dashed line indicates the 1% tail of genome-wide CLR. The *IRT1* region is highlighted in red. (B) Evolutionary history of *IRT1* 130X in Fogo. In (B) to (D), the ancestral allele is colored in black. Inset: Inferred allele frequency trajectory of the derived *IRT1* 130X variant in Fogo shows evidence for rapid fixation, consistent with a hard selective sweep. The black line corresponds to the inferred allele frequency change over time and the color to the posterior probability. (C) Evolutionary history of *NRAMP1* TDs in Fogo. (D) Model of the evolutionary history of *IRT1* and *NRAMP1* in Fogo. The blue line from Monte Velha to Lava represents an apparent recent migration event.

individual), AAGACATAA-TD fixed in Lava accessions (except the single F10-13 stand), and both haplotypes segregating at intermediate frequency in Inferno (fig. S32). Coalescent reconstruction (37) indicates that the two common *NRAMP1* TD haplotypes likely arose after *IRT1* 130X was fixed and after Monte Velha and Inferno subpopulations split, with estimates of ~2.38 ka (tMRCA; 95% CI: 1.07 to 3.16 ka) for AAGACATAA-TD and ~1.60 ka (tMRCA; 95% CI: 0.78 to 2.48 ka) for ATG-TD (Fig. 6C and table S16). We estimated selection coefficients of the two major *NRAMP1* haplotypes based on the trajectories of allele frequencies in the early stages (1.5 to 3.0 ka) to be 2.5 and 6.5% for the AAGACATAA-TD and ATG-TD haplotypes, respectively (table S18). Figure 6D shows a historical reconstruction of the appearance and spread of the *IRT1* and *NRAMP1* variants in the context of the population history. Multiple independent *NRAMP1* TDs arose and, as a group, reached near fixation within Fogo, likely as a result of their ability to compensate the Fe deficiency caused by the loss of *IRT1*. At the subpopulation scale, these appear to represent hard selective sweeps combined with some admixture, while at the island-wide scale, they represent a soft sweep, consistent with theoretical predictions (39, 40).

DISCUSSION

We showed that a two-step process rewired mineral nutrient transport in *Arabidopsis* as it colonized the volcanic island of Fogo. In the first step, an SNP that caused the truncation of the major Fe transporter, *IRT1*, exerted a strong multivariate effect on the ionome and swept to fixation in the population. In the second step, a set of independent TDs arose and reestablished Fe homeostasis by increasing CN and expression level of *NRAMP1*, which increased to near fixation by a soft selective sweep. Our results have several implications regarding evolutionary mechanisms in wild populations.

First, our results relate to the controversy over the roles of hard versus soft sweeps in adaptation (39–42). In this case, a hard selective sweep at *IRT1* appears to have occurred before the Fogo population expanded and structure developed, whereas a soft selective sweep at *NRAMP1* arose more recently, as substructure was beginning to develop on the island. This population structure would have reduced gene flow across the metapopulation and provided time for the accumulation of multiple mutations at *NRAMP1*. This is consistent with previous results showing the probability of a soft sweep from recurrent mutation centers on the chance of a second mutation arising at a locus before the initial beneficial mutation fixes in the population (39, 41–43). Our results thus support theoretical predictions that the extent of population structure should be an important determinant of the mode of adaptation and that the definition of a hard or soft sweep may depend on the scale over which the analysis is conducted (40–43).

Second, our results provide insights regarding the architecture of adaptation. In equilibrium populations, theory predicts that most traits are complex and impacted by many small effect genetic variants (6, 44). However, it has been argued that the first steps of adaptation to a novel environment should tend to have large effects (45). This pattern has been found in microbial populations in artificial selection experiments (11), but it has been less clear whether it holds in populations of multicellular organisms in the wild. This is because identifying the precise mutations that underlie adaptation and reconstructing their histories in natural populations is challenging. The pattern in Fogo is consistent with a model of rapid adaptation,

where an initial large and pleiotropic evolutionary step was followed by a more moderate step, consistent with theoretical expectations of adaptation after a sudden change in the optimum (8).

Third, our results further support the theoretical prediction that epistasis could constrain evolution (46). Although *IRT1* and *NRAMP1* do not physically interact, they have similar functions in metal transport, and there is a delicate balance between the metal ions they transport (Fe and Mn), which are required for core photosynthetic processes (47, 48). An imbalance in these metals can result in chlorosis. This could help to explain the epistatic interaction we observed with respect to the chlorotic phenotype (fig. S11). Once *IRT1* 130X was fixed in the Fogo population, the next evolutionary step was apparently severely constrained to increase CN and thus expression of *NRAMP1*.

Last, our results also have implications for sustainable agriculture. Advances in gene editing will provide new opportunities to design cost-effective specialized crops for local environments, potentially improving global food security and dietary diversity (49). However, to realize this potential, it is necessary to identify the genetic steps required to tailor plants to specific local conditions, including the soil environment. Here, we show that mutations in two transition metal transporters with related functions and substrates reshaped nutrient homeostasis. Overall, our results demonstrate that genetic changes at individual loci can produce substantial changes in nutrient levels in plants and thus reveal the high potential of leveraging nutrient transporters for crop improvement.

MATERIALS AND METHODS

Plant and sequencing materials

We phenotyped in this study accessions from Cape Verde (15) and Morocco (50) together with Cvi-0 (CS76116), Col-0 (CS76113), Col-0-*irt1* (SALK_054554C), and S1-1-*irt1*. The Col-0-*irt1* mutant was genotyped for the T-DNA insertion using the primers P1, P2, and P3 described in table S19. The S1-1-*irt1* mutant was generated in a Santo Antão background (S1-1) using CRISPR-Cas9 following the method described in (51). Briefly, we defined the target sites with the webtool “Knockout guide design” developed by Synthego (<https://design.synthego.com/#/>). Oligonucleotides containing the first 19 nucleotides of the target sites (P4, P5, P6, and P7) were used in a PCR using the pCBC-DT1T2 module vector as template. The resulting PCR product was introduced into the pHEE401E binary vector via Golden Gate cloning. Stable transgenic bacteria were selected with kanamycin (50 $\mu\text{g ml}^{-1}$) and confirmed by colony PCR with the primers P8 and P9. The target sites were verified by Sanger sequencing using the primers P10 and P11. The resulting binary vectors were then transformed into *Agrobacterium tumefaciens* strain GV3101-pSOUP by electroporation. The stable transgenic bacteria were selected with rifampicin (50 $\mu\text{g ml}^{-1}$), gentamicin (50 $\mu\text{g ml}^{-1}$), tetracycline (5 $\mu\text{g ml}^{-1}$), and kanamycin (50 $\mu\text{g ml}^{-1}$). Before plant transformation, stable bacteria were grown to OD₆₀₀ (optical density at 600 nm) of 0.7 to 1.0. The growth medium was then supplemented with 5% sucrose and Silwet L-77 (500 $\mu\text{l liter}^{-1}$), and the S1-1 accession was transformed with the final binary vector following the floral dip method (52). We then grew the collected seeds on standard potting mix (Einheits erde Special Type Min Tray substrate) and selected a T1 chlorotic plant for further analysis. Sanger sequencing analysis at *IRT1* using the primers P10 and P11 indicated heterozygosity at the first target site (single-guide RNA1) with the absence of wild-type

allele and suggested that the T1 plant was a biallelic *irt1* mutant. We then grew 60 plants of the next generation and observed chlorosis in every plant. T2 plants lacking the Cas9 transgene were identified by PCR using the primers P12 and P13 and Sanger-sequenced at *IRT1* using the primers P10 and P11. Sanger sequencing analysis indicated that two disrupting alleles of *IRT1* (*IRT1* E32fs and *IRT1* A62fs) segregated in T2 generation. In every experiment, S1-1-*irt1* corresponds to T3 seeds homozygous for the *IRT1* E32fs allele and lacking the Cas9 transgene. Every primer described here are present in table S19.

We predicted off-targets with Cas-OFFinder (53) using a maximum of five mismatches. We used TAIR10 as target genome and SpCas9 from *Streptococcus pyogenes* (5'-NGG-3') as CRISPR-Cas-derived RNA-guided endonuclease. We performed whole-genome Illumina resequencing to exclude any off-targets in the two S1-1-*irt1* lines. DNA was extracted with the DNeasy Plant Mini Kit (Qiagen, catalog no. 69106). Libraries were prepared using NEBNext Ultra II FS DNA Library Prep [Illumina, New England Biolabs (NEB)]. Sequencing was performed on an Illumina HiSeq3000 platform (Illumina, NEB) with 150-bp paired-end reads to reach 50× coverage. The resulting reads were aligned to the *A. thaliana* reference genome (TAIR10) with bwa (version 0.7.15) using the mem algorithm (54). Duplicated reads were annotated and sorted with Picard (version 2.21.1). SNPs and indels were called with GATK HaplotypeCaller (version 4.2.3.0) using the -ERC GVCF mode (55), and the resulting gVCFs were merged with the parental line S1-1. We applied a filter at the position level and kept only segregating sites. Last, we examined evidence for overlap between predicted off-target sites and SNPs and indels with a custom python script (<https://github.com/HancockLab/Fogo-Edaphic>) and found no overlap. We used Illumina sequence data for 335 Cape Verdean accessions (15), 64 Moroccan accessions (25), 14 accessions from Madeira (26), 118 accessions from China (28), and 1135 Eurasian accessions (27) for multiple analyses involving population structure, selection scans, and structural variation analyses.

Characterization of chlorosis and leaf ionome profiling on standard potting mix

To quantify the chlorosis observed in the Fogo population, we grew in four replicates 141 accessions from Fogo together with 69 accessions from Santo Antão, including Cvi-0, 62 accessions from Morocco, as well as Col-0 and S1-1-*irt1* lines. First, the seeds were stratified for 7 days at 4°C in the dark on petri dishes supplemented with 800 µl of GA4/7 (100 µM), then sown on 54-well trays (60 cm by 40 cm by 6 cm) filled with standard potting mix (Einheits erde Special Type Min Tray substrate), and fertilized with Osmocote start 11-11-7+2MgO+TE fertilizer (1 g liter⁻¹). The plants were grown in controlled growth chamber conditions (8 hours of light, 21°C at day, 14°C at night). We took pictures 24 days after sowing using a Nikon D90 SLR digital camera with a Nikon DX AF-S NIKKOR 18-105 mm 1:3.5-5.6G ED lens. We next conducted image analysis using a custom Fiji (56) script (<https://github.com/HancockLab/Fogo-Edaphic>) to assess the greenness and the segmented area of each individual. Five to 6 weeks after sowing, we assessed relative chlorophyll content, quantum yield of Photosystem II, and nonphotochemical quenching using MultispeQ V2.0 following the manufacturer's instructions.

We then harvested two to three leaves per plant for ionomic analyses. Leaf elemental content analysis was conducted using inductively coupled plasma mass spectrometry (ICP-MS) as described in (57). To compare the leaf ionome of the three populations, we first computed in each population a *z* score for every element extracted and

removed plants with any element showing an absolute *z* score above 2. After filtering out genotypes with less than two replicates, we used the median across replicates for further ionomic analysis. Ti and Cr were used to monitor for soil contamination. As and Se were removed because the relative standard deviation (RSD) was too high. We finally combined the three populations together to conduct the PCA analysis. To add the S1-1-*irt1* on the PCA plot, we included the S1-1-*irt1* line in the Fogo population to compute the *z* score and filter out the outliers. The pipeline can be found in <https://github.com/HancockLab/Fogo-Edaphic>.

To characterize the effect of Fe supplementation on chlorosis in the Fogo natural accessions and the S1-1-*irt1* line, we watered the plants either with tap water only or with Fe-EDDHA (1 g liter⁻¹) (Basafer Plus, Compo Expert).

Soil characterization

We sampled soil from around *A. thaliana* plants growing on Santo Antão and Fogo islands using 20- and 50-ml Falcon tubes. Each soil sample was taken within a 20-cm radius and up to a 20-cm depth around focal plants. These samples were then subjected to ionomic analysis. We adapted a published method using water as the solvent for the elemental extraction (58). Soil samples were air-dried for 48 hours on weighing boats, and 5 g of each soil was weighed into 50-ml Falcon tubes. Soil samples were extracted with 25 ml of 18.2 megohm-cm of Milli-Q water (Merck Millipore) by shaking at 400 rpm in vertical position for 1 hour at room temperature. Samples were then centrifuged for 10 min at 10,000 rpm, and supernatant was collected and centrifuged again for 5 min at 14,000 rpm before sampling. The extracts were acidified with nitric acid (Primar Plus for trace metal analysis) to 2% (v/v) and analyzed with iCAP-Q ICP-MS (Thermo Fisher Scientific) equipped with collision cell technology with energy discrimination. Thirty-one elements were monitored using the following stable isotopes: ⁷Li, ⁹Be, ¹¹B, ²³Na, ²⁴Mg, ²⁷Al, ³¹P, ³⁴S, ³⁹K, ⁴⁴Ca, ⁴⁸Ti, ⁵⁴V, ⁵²Cr, ⁵⁵Mn, ⁵⁶Fe, ⁵⁹Co, ⁶⁰Ni, ⁶³Cu, ⁶⁶Zn, ⁷⁵As, ⁷⁸Se, ⁸⁵Rb, ⁸⁸Sr, ⁹⁵Mo, ¹⁰⁷Ag, ¹¹¹Cd, ¹³³Cs, ¹³⁷Ba, ²⁰⁵Tl, ²⁰⁸Pb, and ²³⁸U. To compare the soil composition of Fogo and Santo Antão, we used the median across extractions per field site. Sample concentrations were calculated using external calibration method within the instrument software. Further data processing was performed using Microsoft Excel.

For the ionomic experiment on native soils, we used natural soils Fogo (F4—Monte Velha) and Santo Antão (S3—Espongeiro) in comparison to standard potting mix (Einheits erde Special substrate with Osmocote start 11-11-7+2MgO+TE fertilizer). For this experiment, all the pots (height = 5.5 cm, diameter = 6 cm, volume = ~100 cm³; Teku) were filled with 4 cm of natural soil recovered with 1 cm of the respective natural soil previously heated at 65°C in an oven for 3 days. The plants were grown in growth chamber-controlled conditions (12 hours of light, 21°C at day, 14°C at night). Five weeks after sowing, two to three young leaves were collected and washed three times with 18.2 megohm-cm of Milli-Q water (Merck Millipore) before being transferred to 1.5-ml Eppendorf tubes. The samples were then dried overnight at 80°C in an oven. To compare the leaf ionome of S1-1 and F13-8 in the different soils, we removed samples with leaf Ti concentration above 20 in µg g⁻¹ of DW (ppm).

Bulk segregant analysis

To perform BSA, we generated an F2 population by crossing the Fogo accession F13-8 to the *A. thaliana* reference accession Col-0. The F2

seeds and their corresponding parents were first stratified for 7 days at 4°C in the dark on petri dishes supplemented with 800 μl of GA4/7 (100 μM), then sown on squared pots (7 cm by 7 cm, height = 8 cm; Göttinger) with standard potting mix, and transferred to controlled greenhouse conditions (16 hours of light, 8 hours of dark, 21°C). To examine the chlorosis segregation, we took pictures of each F2 individual's rosette 4 weeks after sowing using a Nikon D90 SLR digital camera with a Tamron SP AF 90 mm F/2.8 macro lens and ColorChecker (X-Rite, catalog MSCCC) for color control. We counted 24 highly chlorotic plants out of 454 in F2 individuals in total. DNA was extracted from two young rosette leaves of each F2 individual with the BioSprint 96 DNA Plant Kit (Qiagen, catalog no. 941557) and quantified using a Qubit fluorometer (Thermo Fisher Scientific, catalog no. Q33238). The DNA concentration was then adjusted to 1 ng μl^{-1} for each extraction, and the DNA integrity was checked on a 1% agarose gel. We next conducted image analysis using a custom Fiji (56) script (<https://github.com/HancockLab/Fogo-Edaphic>) to assess the greenness of each F2 individual. After examining the greenness distribution, we could recover nine highly chlorotic individuals for bulk sequencing and pooled 20 ng of DNA from each selected individual (Fig. 3B). Library preparation and Illumina HiSeq3000 sequencing were performed by the Max Planck Genome Center in Cologne (Germany). Briefly, the libraries were prepared using NEBNext Ultra II FS DNA Library Prep (Illumina, NEB) and sequenced with 150-bp paired-end reads to 36 \times coverage using the Illumina HiSeq3000 platform (Illumina, NEB).

The reads were aligned to the *A. thaliana* reference genome (TAIR10) with bwa (version 0.7.15) using the mem algorithm (54). Duplicated reads were annotated and sorted with picard (version 2.21.1). SNPs and short indels were called with GATK HaplotypeCaller (version 4.1.3.0) (55) using the reference confidence model (BP_resolution mode). We filtered the variants with a custom Python (version 2.7.13) script to keep only the positions with a genotype quality GQ > 25, a total depth DP > 3, and no missing allele in the corresponding parents. In addition, we filtered out variants with minor allele frequencies below 5% in the Fogo population. To identify candidate genes after BSA, we annotated with SnpEff (version 4.3r) (59) the variants present in windows where the Col-0 allele frequency was above 0.99 for chromosome 1 and below 0.01 for chromosome 4. We then extracted the predicted high-impact variants present in F13-8. For chromosome 1, since we did not identify strong candidates based on SNP and short indel variation, we examined the presence of structural variation in F13-8 using Delly (version 0.8.3) (60) and annotated them using SnpEff (version 4.3r) (59). The pipeline can be found in <https://github.com/HancockLab/Fogo-Edaphic>. Multiple alignments were produced with BoxShade (<https://sourceforge.net/projects/boxshade/>) based on the genomic variant calls at *IRT1* and *NRAMP1* (15).

F2 association

To follow up the results from BSA, we further examined segregation of specific candidate variants in the F13-8 \times Col-0 F2 population. We genotyped these by derived cleaved amplified polymorphic sequences (dCAPS) on each F2 individual and five replicates per parental line as controls. We genotyped SNPs lying within the coding region of *IRT1* (Chr4: 10 707 874) and *NRAMP1* (Chr1: 30 374 389) by using the restriction enzymes Dde I (NEB, catalog no. R0175S) and Sac I-HF (NEB, catalog no. R3156L), respectively. We used dCAPS finder 2.0 (<http://helix.wustl.edu/dcaps/dcaps.html>) to generate the

PCR primers (P14 and P15 for *IRT1*, P16 and P17 for *NRAMP1*). The polymorphic regions were PCR-amplified in a 25- μl reaction containing 15 μl of dH₂O, 0.5 μl of dNTPs (10 mM), 1 μl of forward and reverse primer (10 μM each), 2.5 μl of MgCl₂ (25 mM) (Bio-Budget, catalog no. 80-60010100), 2.5 μl of 10 \times BD buffer (Bio-Budget, catalog no. 80-60010100), 0.5 μl of Taq polymerase (Bio-Budget, catalog no. 80-60010100), and 2 μl of genomic DNA. Ten microliters of the corresponding PCR products was digested for 3 hours at 37°C with 2 μl of 10 \times CutSmart buffer, 0.5 μl of appropriate restriction enzyme, and 7.5 μl of H₂O. The resulting digested products were visualized on a 3.5 \times agarose gel. Primers described here are detailed in table S19.

Characterizing variation in the *NRAMP1* region from short-read data

We used Delly (version 0.8.3) (60) to genotype the region surrounding *NRAMP1* for the presence or absence of a TD in short-read-sequenced genomes from Morocco (25, 50), Eurasia (27), and Cape Verde (15). We confirmed every TD manually in the Integrative Genomics Viewer (IGV) (61). The pipeline can be found in <https://github.com/HancockLab/Fogo-Edaphic>.

Whole-genome assembly

The seeds intended for Nanopore sequencing were stratified for 7 days at 4°C in the dark on petri dishes before sowing on pots (7 cm by 7 cm, height = 8 cm; Göttinger) containing standard potting mix. Five seedlings per pot were left and grown under greenhouse conditions (16 hours of light, 8 hours of dark, 21°C) for 3 weeks. Library preparation and Nanopore sequencing were performed by the Max Planck Genome Center in Cologne (Germany). Briefly, the genomic DNA was extracted from pooled leaf tissue from the same genotype. DNA fragments above 30 kb were selected with BluePippin (Sage Science). The libraries were prepared using Ligation Sequencing Kit 1D (Oxford Nanopore Technologies, catalog no. SQK-LSK109) and sequenced using the GridION X5 platform (Oxford Nanopore Technologies).

De novo assemblies were generated with miniasm-0.3 (62) and minimap2-2.17 (63). The resulting draft assemblies were first polished twice with racon (version 1.4.10) (64) using the raw Nanopore reads and then 10 more times with pilon (version 1.23) (65) using the corresponding Illumina short reads. We oriented and scaffolded the contigs using the reference genome assembly (TAIR10). To prevent mapping of Illumina short reads to anchoring points of the contigs, we inserted 1000 N's between the contigs. The pipeline can be found in <https://github.com/HancockLab/Fogo-Edaphic>.

NRAMP1 CN estimation

To estimate *NRAMP1* CN based on short-read sequenced genomes, we first extracted the breakpoint positions using Delly (version 0.8.3) (60). We then obtained the coverage at each base pair between the positions 303,5000 and 304,0000 on chromosome 1 using the "--depth" command in VCFtools (version 0.1.16) (66). We next used a Python (version 3.8.3) custom script to calculate the coverage in 800-bp sliding windows every 400 bp and normalized it to the 3-kb regions upstream and downstream of the breakpoints. For the single *NRAMP1* copy, since no breakpoints were found, we use the most extreme breakpoint positions found at *NRAMP1* in Fogo (start: 30363708; end: 30394114). We then used the median of each ratio to test the association between *NRAMP1* CN and the different phenotypes.

The pipeline can be found in <https://github.com/HancockLab/Fogo-Edaphic>.

To assess *NRAMP1* CN by dPCR, we first extracted genomic DNA from leaf tissue with the DNeasy Plant Mini Kit (Qiagen, catalog no. 69106) and quantified it using the Qubit fluorometer (Thermo Fisher Scientific, catalog no. Q33238). The dPCRs were performed with the QIAcuity 5-plex device (Qiagen, catalog no. 911021) in 26K (24-well) Nanoplates (Qiagen, catalog no. 250001). We chose Eco RI–HF for DNA fragmentation after confirming the absence of Eco RI–HF site at the DNA targets. The dPCR mixtures were set up with the QIAcuity EG PCR Kit (Qiagen, catalog No. 250111) following the manufacturer's instructions with 5 ng of genomic DNA together with 0.2 U of Eco RI–HF (NEB) per reaction and left for 15 min at room temperature for DNA fragmentation. The thermal cycling conditions were as follows: initial heat activation, 2 min at 95°C; three-step cycling conditions (40 cycles), 15 s at 95°C, 15 s at 55°C, and 15 s at 72°C; cooling down, 5 min at 40°C. Absolute quantification and CNV were estimated with QIAcuity Software Suite 1.2.18 (Qiagen). EvaGreen detection was done using the green channel with 100-ms exposure and gain 6. We chose *PHOSPHATASE 2A* (*PP2A*) (AT1G13320) as a reference gene after excluding the presence of extra copies with IGV (61). We used the primers P18 and P19 for *NRAMP1* and P20 and P21 for *PP2A* for the dPCRs. Every primer described here is detailed in table S19.

To estimate the number of *NRAMP1* copies based on long-read sequenced genomes, we aligned all assembled genomes individually to the reference genome assembly (TAIR10) with the nucmer package from MUMmer (version 3.1) (67) with the following parameters: --maxmatch, -l 100, -c 500. We then identified structural variants using Assemblytics (version 1.2) (68) with the following parameters: unique sequence length required, 1000 bp; maximum variant size, 100,000 bp; and minimum variant size, 50 bp. The pipeline can be found in <https://github.com/HancockLab/Fogo-Edaphic>.

NRAMP1 expression analysis

To quantify mRNA levels in root tissues, seedlings were grown under sterile conditions on vertically oriented square 120-mm polystyrene petri dishes (Greiner). After surface sterilization using sodium hypochlorite solution (40%), the seeds were stratified for 7 days at 4°C in the dark and transferred to a growth chamber (Percival Scientific) in long-day condition (16 hours of light, 8 hours of dark, 21°C). Total RNA extraction was performed using TRIzol (Invitrogen, catalog no. 15596026). One microgram of total RNA was treated with the DNA-free DNA Removal Kit (Invitrogen, catalog no. AM1906) for 1 hour at 37°C following the manufacturer's instructions. Complementary DNA was generated using the SuperScript II reverse transcriptase (Invitrogen, catalog no. 18064022) together with oligo(dT)18 primer for 2 hours at 42°C.

To compare the expression of *NRAMP1* between F13-8 and Col-0, we used a (1×) Murashige and Skoog (MS) medium and harvested root tissue from 15 plants 12 days after the shift to long-day conditions. mRNA levels were assessed by quantitative PCR on a LightCycler 480 instrument (Roche) using the EvaGreen dye (Biotium, catalog no. 31000) and the my-Budget Taq DNA polymerase (Bio-Budget, catalog no. 80-60010100). We used the primers P22 and P23 to detect *NRAMP1* mRNA and the primers P20 and P21 to detect *PP2A* mRNA. We applied the $-\Delta\Delta C_t$ method with Pfaffl correction using *PP2A* (AT1G13320) as reference gene and S1-1 as control genotype.

To assess the expression of *NRAMP1* in response to different Mn concentrations, we cultivated seedlings for 7 days in a Hoagland

medium solidified with 0.8% (w/v) agar type M (Sigma-Aldrich) and containing 1% (w/v) sucrose and 10 μ M FeHBED (69). We then transferred the seedlings onto a fresh modified Hoagland medium, which was solidified with EDTA-washed 0.8% (w/v) agar type M to remove any contaminant metals, and which contained 10 μ M FeHBED and either 10 μ M Mn (+Mn) or no Mn (–Mn). We harvested root tissue from 15 14-day-old plants. mRNA levels were quantified by dPCR using the QIAcuity five-plex device (Qiagen, catalog no. 911001) in QIAcuity Nanoplates 26K 24-well plates (Qiagen, catalog no. 250001). The dPCR mixtures were set up in duplex with the QIAcuity Probe PCR Kit (Qiagen, catalog no. 250101) following the manufacturer's instructions. We used the primers P25 and P26 with the probe P26 (5' FAM, BHQ1 3') for *NRAMP1* and the primers P20 and P21 with the probe P27 (5' ROX, BHQ2 3') for *PP2A*. For normalization, we used *PP2A* as reference gene and Col-0 in +Mn as control genotype. Primers and probes used in this study are detailed in table S19.

Heritability

We estimated the heritable variance in Fe and Mn concentrations with GEMMA (version 0.94) (33). We applied a linear mixed model that accounts for population structure by including a kinship matrix in the model. As previously described (15), we used a VCF file generated with GATK. This VCF file was filtered with VCFtools (version 0.1.16) (66) to consider only biallelic positions and individuals having a genotype quality above 25 ($GQ > 25$) and a total depth above 3 ($DP > 3$). We then used *NRAMP1* CN estimates as a covariate.

Ionomics on plates

To quantify metal concentrations in leaf and root tissues, seedlings were cultivated under sterile conditions on vertically oriented square 120-mm polystyrene petri dishes (Greiner Bio-One) in 16-hour light (22°C, 120 μ mol $m^{-2} s^{-1}$, white light)/8-hour dark (18°C) cycles in a growth chamber (Percival CU-41L4; CLF Climatics). After surface sterilization for 1 min in 70% (v/v) EtOH, followed by 10 min in 1.4% (v/v) NaOCl and 0.02% (v/v) Triton X-100, and four final washing steps in sterile water, 40 seeds per plate were transferred onto modified Hoagland medium containing 1% (w/v) sucrose and solidified with 0.8% (w/v) agar type M (Sigma-Aldrich) (69). Col-0 and Col-0-*irt1* were stratified at 4°C in the dark for 2 days, whereas media for all other genotypes were additionally supplemented with 100 μ M GA4+7 (Duchefa), followed by stratification for 7 days. After precultivation for 10 days, seedlings were transferred onto fresh modified Hoagland medium containing 10 μ M FeHBED and solidified with EDTA-washed 0.8% (w/v) agar type M, from which contaminant metals had been removed by washing in EDTA solutions (69). Pools of leaf and root tissues of 15 to 30 17-day-old seedlings were harvested from one or two plates by cutting below the hypocotyl using a scalpel. Desorption to remove extracellularly bound metal cations was conducted in 50 ml of 5 mM CaSO₄, 1 mM MES-KOH (pH 5.7) for 10 min, 50 ml of 5 mM CaSO₄, 10 mM Na₂EDTA, and 1 mM MES-KOH (pH 5.7) for 5 min, followed by two washes in 50 ml of ultrapure water for 1 min, all conducted on ice with occasional stirring. Desorbed tissues were blotted dry using paper towels (Blauer Engel) and dried in paper bags at 60°C for 3 days. After 3 days of equilibration to room temperature, plant material was weighed and acid-digested as in (70). Multi-element analysis was conducted using inductively coupled plasma optical emission spectrometry (ICP-OES) in an iCAPDuo 6500 instrument (Thermo Fisher Scientific).

Calibration was done with a blank and a series of five multielement standards from single-element standard solutions (AAS Standards, Bernd Kraft), with QC (quality control) using a sample blank and an intermediate calibration standard solution, as well as digests of certified reference material (Virginia tobacco leaves, INCT-PVTL 6; Institute of Nuclear Chemistry and Technology, PL).

Population structure and migration history of the Fogo populations

To examine population structure, the original VCF (15) was first filtered using bcftools (version 1.9) (71) with the command “bcftools view -m2 -M2 -v snps -i ‘MIN(FMT/DP)>3 & MIN(FMT/GQ)>25 & F_MISSING = 0.’” The resulting biallelic SNPs were then pruned for short-range linkage disequilibrium (LD) by removing any SNP that showed correlation coefficients (r^2) greater than 0.1 in a window size of 50 kb and step size of 10 bp. We conducted PCA using the --pca option in PLINK (version 1.9) (72).

We quantified the evolutionary relationships among subtrees of the Fogo populations using Twisst (34) with the Santo Antão population as outgroup. Biallelic SNPs were first converted to “.geno” format using the script “parseVCF.py” (https://github.com/simonhmartin/genomics_general/tree/master/VCF_processing). Subsequently, we generated maximum-likelihood trees in sliding windows of 50 SNPs with the BIONJ (73) algorithm in PhyML (version 3.9_360-500M) (74) using the script “phyml_sliding_windows.py” (https://github.com/simonhmartin/genomics_general/tree/master/phylo). We ran Twisst on the inferred trees using the “--complete method” option. The pipeline can be found in <https://github.com/HancockLab/Fogo-Edaphic>.

We used custom scripts to compute folded JSFS (<https://github.com/HancockLab/Fogo-Edaphic>) across each pair of genetic groups in Fogo, namely, Monte Velha–Inferno, Monte Velha–Lava, and Inferno–Lava. The JSFS were computed on the entire genome after excluding sites with more than 5% missing data, CpG sites, due to their hypermutable nature, pericentromeric regions, which are rich in satellite repeats, and other repeat regions identified with Heng Li’s SNPable approach (<http://bit.ly/snpable>). For each SNP that passed quality filtering, and each genetic cluster, we took a random set of 90% of the accessions with no missing data at the focus SNP to generate the JSFS.

We inferred demographic histories of the three Fogo subpopulations (Monte Velha, Inferno, and Lava) by fitting models to the JSFS using ∂adi (version 2.1.0) (75). We used two sets of models: First, we optimized parameters for four two-population models applied to all pairwise combinations of the three populations. The models are as follows: (i) a population split with no migration and constant population size (N_e); (ii) a population split with a single change in N_e , which can converge toward a bottleneck; (iii) a split followed by exponential changes in N_e and no migration; and (iv) an isolation with migration model (IM): a split followed by exponential changes in N_e and asymmetric migration.

We ran ∂adi in 200 independent replicates with a maximum of 50 iterations, for each model and each combination of populations. In each run, we randomly drew starting values for parameters within standard parameter boundaries: ($10^{-3} \times N_{\text{ref}}$, $20 \times N_{\text{ref}}$) for effective population sizes (N_e), (0 , $20/N_{\text{ref}}$) for migration rate, and (0 , $10 \times N_{\text{ref}}$) for split times, where N_{ref} is the size of the ancestral population. Within each model, among the 200 independent runs, we selected the parameter set that resulted in the highest likelihood. Then, we

compared support across models with the Akaike information criterion (AIC). We assumed a generation time of 1 year and a mutation rate of 7.1×10^{-9} based on estimates from mutation accumulation lines (76).

Geographical maps

For the geographical distribution of the natural populations in Fogo, we used Google Earth: Image 2021 CNES/Airbus, Data SIO, NOAA, U.S. Navy, NGA, GEBCO. We used R (version 3.6.2) for the geographical distribution of the *NRAMP1* haplotypes in Fogo. The pipeline can be found in <https://github.com/HancockLab/Fogo-Edaphic>.

Inferences of selection

For the inference of genomic signatures of selective sweeps, we used SweepFinder2 (35) with a grid size of 100 kb. We used a recombination map inferred from crosses (77), and we corrected it for the rate of outcrossing inferred from natural populations (78). Because the Fogo population evolved in isolation from the continents after a colonization bottleneck, we polarized the site frequency spectrum to the reference genome, TAIR10. We considered genomic regions with the highest percentile of the CLR distribution as having the strongest signatures of selection. To define the borders of the selective sweep region overlapping with *IRT1*, we examined the successive 100-kb windows above the 1% percentile of the genome-wide CLR (18.28) and obtained a 502,269-bp region (Chr4: 10548683 and Chr4: 11050953). We then annotated variants present within this region and private to the Fogo population using SnpEff (version 4.3r) (59).

Reconstruction of evolutionary histories of *IRT1* and *NRAMP1* loci in Fogo

We inferred the genealogical trees for the derived alleles of *IRT1* and *NRAMP1* with Relate (version 1.1.4) (37). Then, we inferred historical frequency trajectories and selection coefficients for *IRT1* 130X (Chr4: 10707974) and the *NRAMP1* TDs (Chr1:30363708 and Chr1:30369728) using importance sampling over local genealogies in CLUES (38). To produce genealogies, we used as the outgroup S1-1 and S8-505, two accessions that originated from the Santo Antão subpopulation that is genetically closest to the Fogo population based on clustering in an NJ tree (15). We used bcftools (version 1.9) (71) to extract from the VCF file biallelic SNPs with the command “bcftools view -U -c 1 -m2 -M2 -v snps” and to remove any position where at least one individual did not pass the filter quality with the command “bcftools view DP > 3 MIN(FMT/DP)>3 & MIN(FMT/GQ)>25 & F_MISSING=0.” For *NRAMP1*, we introduced variants into the “hap” file corresponding to the break point junctions as shown in (Fig. 5A). We ran Relate (37) under a haploid model using a mutation rate parameter based on the spontaneous mutation rate previously estimated in *A. thaliana* (76), corrected for missing data for each region with the formula $7 \times 10^{-9} \times \left(1 - \frac{\text{missing variants}}{\text{total variants}}\right)$. We obtained 1.55189×10^{-9} for *IRT1* G130X (Chr4: 10707974) and 1.42794×10^{-9} for *NRAMP1* TDs (Chr1: 30363708 and Chr1: 30369728). For the “--map” parameter, we used a previously described recombination map (77) and divided the genetic distances by 20 to correct for the outcrossing rate of 5% estimated in natural populations (78). We set the generation time to 1 year “--year_per_gen 1,” sampled 200 branch lengths with the “SampleBranchLength.sh” script, and plotted the trees of interest using the “TreeViewSample.sh” script.

Selection coefficients were inferred by comparing the local genealogies against the genome-wide coalescence rates (--coal) inferred previously (15), which provides conservative estimates for the local trees against background population growth. We obtained estimates of the posterior distributions of allele frequencies over time using a recessive model for *IRT1* 130X (-- dom 0): <inference.py --popFreq 0.99 --tCutoff 5000 --coal relate.popsiz.coal --sMax 1 --df 100 --dom 0> and an additive model for the *NRAMP1* TDs (-- dom 0.5): <inference.py --popFreq 0.99 --tCutoff 5000 --coal relate.popsiz.coal --sMax 1 --df 100 --dom 0.5>. We inferred selection coefficients jointly across time bins of 1.5 ka between the present day and the time in the past when the variants arose. We used three-time bins (epochs) (0 to 1.5, 1.5 to 3, and 3 to 4.5 ka) for *IRT1* 130X (table S17) and two-time bins (0 to 1.5 and 1.5 to 3 ka) for *NRAMP1* TDs (table S18). The pipeline can be found in <https://github.com/HancockLab/Fogo-Edaphic>.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <https://science.org/doi/10.1126/sciadv.abm9385>

REFERENCES AND NOTES

- R. D. H. Barrett, S. Laurent, R. Mallarino, S. P. Pfeifer, C. C. Y. Xu, M. Foll, K. Wakamatsu, J. S. Duke-Cohan, J. D. Jensen, H. E. Hoekstra, Linking a mutation to survival in wild mice. *Science* **363**, 499–504 (2019).
- Y. F. Chan, M. E. Marks, F. C. Jones, G. Villarreal, M. D. Shapiro, S. D. Brady, A. M. Southwick, D. M. Absher, J. Grimwood, J. Schmutz, R. M. Myers, D. Petrov, B. Jónsson, D. Schluter, M. A. Bell, D. M. Kingsley, Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a *Pitx1* enhancer. *Science* **327**, 302–305 (2010).
- M. Karageorgi, S. C. Groen, F. Sumbul, J. N. Pelaez, K. I. Verster, J. M. Aguilar, A. P. Hastings, S. L. Bernstein, T. Matsunaga, M. Astourian, G. Guerra, F. Rico, S. Dobler, A. A. Agrawal, N. K. Whiteman, Genome editing retraces the evolution of toxin resistance in the monarch butterfly. *Nature* **574**, 409–412 (2019).
- S. Lamichhane, J. Berglund, M. S. Almén, K. Maqbool, M. Grabherr, A. Martinez-Barrio, M. Promerová, C.-J. Rubin, C. Wang, N. Zamani, B. R. Grant, P. R. Grant, M. T. Webster, L. Andersson, Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature* **518**, 371–375 (2015).
- D. L. Stern, V. Orgogozo, The loci of evolution: How predictable is genetic evolution? *Evolution* **62**, 2155–2177 (2008).
- N. H. Barton, P. D. Keightley, Understanding quantitative genetic variation. *Nat. Rev. Genet.* **3**, 11–21 (2002).
- R. A. Fisher, *The Genetical Theory of Natural Selection* (Clarendon Press, 1930).
- H. A. Orr, The population genetics of adaptation: The distribution of factors fixed during adaptive evolution. *Evolution* **52**, 935–949 (1998).
- S. Wright, Evolution in mendelian populations. *Genetics* **16**, 97–159 (1931).
- J. P. Castro, M. N. Yancoskie, M. Marchini, S. Belohlavý, L. Hiramatsu, M. Kučka, W. H. Beluch, R. Naumann, I. Skuplik, J. Cobb, N. H. Barton, C. Rolian, Y. F. Chan, An integrative genomic analysis of the Longshanks selection experiment for longer limbs in mice. *eLife* **8**, e42014 (2019).
- S. E. Schoustra, T. Bataillon, D. R. Gifford, R. Kassen, The properties of adaptive walks in evolving populations of fungus. *PLOS Biol.* **7**, e1000250 (2009).
- J. Maynard Smith, J. Haigh, The hitch-hiking effect of a favourable gene. *Genet. Res.* **23**, 23–35 (1974).
- J. Hermisson, P. S. Pennings, Soft sweeps: Molecular population genetics of adaptation from standing genetic variation. *Genetics* **169**, 2335–2352 (2005).
- T. Karasov, P. W. Messer, D. A. Petrov, Evidence that adaptation in *Drosophila* is not limited by mutation at single sites. *PLOS Genet.* **6**, e1000924 (2010).
- A. Fulgione, C. Neto, A. F. Elfarargi, E. Tergemina, S. Ansari, M. Göktay, H. Dinis, N. Döring, P. J. Flood, S. Rodriguez-Pacheco, N. Walden, M. A. Koch, F. Roux, J. Hermisson, A. M. Hancock, Parallel reduction in flowering time from de novo mutations enable evolutionary rescue in colonizing lineages. *Nat. Commun.* **13**, 1461 (2022).
- S. Plesner, P. M. Holm, J. R. Wilson, 40Ar–39Ar geochronology of Santo Antão, Cape Verde Islands. *J. Volcanol. Geotherm. Res.* **120**, 103–121 (2003).
- V. E. Neall, Volcanic soils. *Land Use Land Cover Soil Sci.* **7**, 23–45 (2009).
- M. Broadley, P. Brown, I. Cakmak, Z. Rengel, F. Zhao, in *Marschner's Mineral Nutrition of Higher Plants (Third Edition)*, P. Marschner, Ed. (Academic Press, 2012), pp. 191–248.
- G. Vert, N. Grotz, F. Dedaldecamp, F. Gaymard, M. L. Guerinet, J. F. Briat, C. Curie, IRT1, an Arabidopsis transporter essential for iron uptake from the soil and for plant growth. *Plant Cell* **14**, 1223–1233 (2002).
- R. Henriques, J. Jasik, M. Klein, E. Martinoia, U. Feller, J. Schell, M. S. Pais, C. Koncz, Knock-out of Arabidopsis metal transporter gene *IRT1* results in iron deficiency accompanied by cell differentiation defects. *Plant Mol. Biol.* **50**, 587–597 (2002).
- C. Varotto, D. Maiwald, P. Pesaresi, P. Jahns, F. Salamini, D. Leister, The metal ion transporter IRT1 is necessary for iron homeostasis and efficient photosynthesis in Arabidopsis thaliana. *Plant J.* **31**, 589–599 (2002).
- R. Cailliatte, A. Schikora, J.-F. Briat, S. Mari, C. Curie, High-affinity manganese uptake by the metal transporter NRAMP1 is essential for Arabidopsis growth in low manganese conditions. *Plant Cell* **22**, 904–917 (2010).
- L. Castaings, A. Caquot, S. Loubet, C. Curie, The high-affinity metal transporters NRAMP1 and IRT1 team up to take up iron under sufficient metal provision. *Sci. Rep.* **6**, 37222 (2016).
- C. Curie, J. M. Alonso, M. Le Jean, J. R. Ecker, J. F. Briat, Involvement of NRAMP1 from Arabidopsis thaliana in iron transport. *Biochem. J.* **347** (Pt. 3), 749–755 (2000).
- A. Durvasula, A. Fulgione, R. M. Gutaker, S. I. Alacakaptan, P. J. Flood, C. Neto, T. Tsuchimatsu, H. A. Burbano, F. X. Picó, C. Alonso-Blanco, A. M. Hancock, African genomes illuminate the early history and transition to selfing in Arabidopsis thaliana. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 5213–5218 (2017).
- A. Fulgione, M. Koornneef, F. Roux, J. Hermisson, A. M. Hancock, Madeiran Arabidopsis thaliana reveals ancient long-range colonization and clarifies demography in Eurasia. *Mol. Biol. Evol.* **35**, 564–574 (2018).
- 1001 Genomes Consortium, 1,135 genomes reveal the global pattern of polymorphism in Arabidopsis thaliana. *Cell* **166**, 481–491 (2016).
- Y. P. Zou, X. H. Hou, Q. Wu, J. F. Chen, Z. W. Li, T. S. Han, X. M. Niu, L. Yang, Y. C. Xu, J. Zhang, F. M. Zhang, D. Tan, Z. Tian, H. Gu, Y. L. Guo, Adaptation of Arabidopsis thaliana to the Yangtze River basin. *Genome Biol.* **18**, 239 (2017).
- L. Giraut, M. Falque, J. Drouaud, L. Pereira, O. C. Martin, C. Mézard, Genome-wide crossover distribution in Arabidopsis thaliana meiosis reveals sex-specific patterns along chromosomes. *PLOS Genet.* **7**, e1002354 (2011).
- M. Hanikenne, I. N. Talke, M. J. Haydon, C. Lanz, A. Nolte, P. Motte, J. Kroymann, D. Weigel, U. Kramer, Evolution of metal hyperaccumulation required cis-regulatory changes and triplication of HMA4. *Nature* **453**, 391–395 (2008).
- S. Nair, D. Nash, D. Sudimack, A. Jaidee, M. Barends, A.-C. Uhlemann, S. Krishna, F. Nosten, T. J. C. Anderson, Recurrent gene amplification and soft selective sweeps during evolution of multidrug resistance in malaria parasites. *Mol. Biol. Evol.* **24**, 562–573 (2007).
- G. H. Perry, N. J. Dominy, K. G. Claw, A. S. Lee, H. Fiegler, R. Redon, J. Werner, F. A. Villanea, J. L. Mountain, R. Misra, N. P. Carter, C. Lee, A. C. Stone, Diet and the evolution of human amylase gene copy number variation. *Nat. Genet.* **39**, 1256–1260 (2007).
- X. Zhou, M. Stephens, Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* **44**, 821–824 (2012).
- S. H. Martin, S. M. V. Belleghem, Exploring evolutionary relationships across the genome using topology weighting. *Genetics* **206**, 429–438 (2017).
- M. DeGiorgio, C. D. Huber, M. J. Hubisz, I. Hellmann, R. Nielsen, SweepFinder2: Increased sensitivity, robustness and flexibility. *Bioinformatics* **32**, 1895–1897 (2016).
- P. S. Pennings, J. Hermisson, Soft sweeps III: The signature of positive selection from recurrent mutation. *PLOS Genet.* **2**, e186 (2006).
- L. Speidel, M. Forest, S. Shi, S. R. Myers, A method for genome-wide genealogy estimation for thousands of samples. *Nat. Genet.* **51**, 1321–1329 (2019).
- A. J. Stern, P. R. Wilton, R. Nielsen, An approximate full-likelihood method for inferring selection and allele frequency trajectories from DNA sequence data. *PLOS Genet.* **15**, e1008384 (2019).
- J. Hermisson, P. S. Pennings, Soft sweeps and beyond: Understanding the patterns and probabilities of selection footprints under rapid adaptation. *Methods Ecol. Evol.* **8**, 700–716 (2017).
- R. B. Harris, A. Sackman, J. D. Jensen, On the unfounded enthusiasm for soft selective sweeps II: Examining recent evidence from humans, flies, and viruses. *PLOS Genet.* **14**, e1007859 (2018).
- J. D. Jensen, On the unfounded enthusiasm for soft selective sweeps. *Nat. Commun.* **5**, 5281 (2014).
- P. W. Messer, D. A. Petrov, Population genomics of rapid adaptation by soft selective sweeps. *Trends Ecol. Evol.* **28**, 659–669 (2013).
- P. S. Pennings, J. Hermisson, Soft sweeps II—Molecular population genetics of adaptation from recurrent mutation or migration. *Mol. Biol. Evol.* **23**, 1076–1084 (2006).
- R. A. Fisher, XV.—The correlation between relatives on the supposition of Mendelian inheritance. *Trans. R. Soc. Edinb.* **52**, 399–433 (1919).
- H. A. Orr, J. A. Coyne, The genetics of adaptation: A reassessment. *Am. Nat.* **140**, 725–742 (1992).
- D. M. Weinreich, N. F. Delaney, M. A. DePristo, D. L. Hartl, Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* **312**, 111–114 (2006).
- R. Therby-Vale, B. Lacombe, S. Y. Rhee, L. Nussaume, H. Rouached, Mineral nutrient signalling controls photosynthesis: Focus on iron deficiency-induced chlorosis. *Trends Plant Sci.*, S1360138521003113 (2021).

48. S. B. Schmidt, M. Eisenhut, A. Schneider, Chloroplast transition metal regulation for efficient photosynthesis. *Trends Plant Sci.* **25**, 817–828 (2020).
49. E. T. Wurtzel, C. E. Vickers, A. D. Hanson, A. H. Millar, M. Cooper, K. P. Voss-Fels, P. I. Nikel, T. J. Erb, Revolutionizing agriculture with synthetic biology. *Nat. Plants* **5**, 1207–1210 (2019).
50. A. C. Brennan, B. Méndez-Vigo, A. Haddioui, J. M. Martínez-Zapater, F. X. Picó, C. Alonso-Blanco, The genetic structure of *Arabidopsis thaliana* in the south-western Mediterranean range reveals a shared history between North Africa and southern Europe. *BMC Plant Biol.* **14**, 17 (2014).
51. Z. P. Wang, H. L. Xing, L. Dong, H. Y. Zhang, C. Y. Han, X. C. Wang, Q. J. Chen, Egg cell-specific promoter-controlled CRISPR/Cas9 efficiently generates homozygous mutants for multiple target genes in *Arabidopsis* in a single generation. *Genome Biol.* **16**, 144 (2015).
52. S. J. Clough, A. F. Bent, Floral dip: A simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16**, 735–743 (1998).
53. S. Bae, J. Park, J.-S. Kim, Cas-OFFinder: A fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* **30**, 1473–1475 (2014).
54. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
55. G. Van der Auwera, B. O'Connor, *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra* (O'Reilly Media Inc., ed. 1, 2020).
56. J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, A. Cardona, Fiji: An open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
57. I. Salas-González, G. Reyt, P. Flis, V. Custódio, D. Gopaulchan, N. Bakhroum, T. P. Dew, K. Suresh, R. B. Franke, J. L. Dangl, D. E. Salt, G. Castrillo, Coordination between microbiota and root endodermis supports plant mineral nutrient homeostasis. *Science* **372**, eabd0695 (2020).
58. B. J. Arnold, B. Lahner, J. M. Da Costa, C. M. Weisman, J. D. Hollister, D. E. Salt, K. Bomblies, L. Yant, Borrowed alleles and convergence in serpentine adaptation. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 8320–8325 (2016).
59. P. Cingolani, A. Platts, L. L. Wang, M. Coon, T. Nguyen, L. Wang, S. J. Land, X. Lu, D. M. Ruden, A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain *w¹¹¹⁸; iso-2; iso-3*. *Fly* **6**, 80–92 (2012).
60. T. Rausch, T. Zichner, A. Schlattl, A. M. Stütz, V. Benes, J. O. Korbel, DELLY: Structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* **28**, i333–i339 (2012).
61. H. Thorvaldsdottir, J. T. Robinson, J. P. Mesirov, Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–192 (2013).
62. H. Li, Minimap and miniasm: Fast mapping and de novo assembly for noisy long sequences. *Bioinformatics* **32**, 2103–2110 (2016).
63. H. Li, Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
64. R. Vaser, I. Sović, N. Nagarajan, M. Šikić, Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* **27**, 737–746 (2017).
65. B. J. Walker, T. Abeel, T. Shea, M. Priest, A. Abouelliel, S. Sakthikumar, C. A. Cuomo, Q. Zeng, J. Wortman, S. K. Young, A. M. Earl, Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLOS ONE* **9**, e112963 (2014).
66. 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, The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
67. S. Kurtz, A. Phillippy, A. L. Delcher, M. Smoot, M. Shumway, C. Antonescu, S. L. Salzberg, Versatile and open software for comparing large genomes. *Genome Biol.* **5**, R12 (2004).
68. M. Nattestad, M. C. Schatz, Assemblytics: A web analytics tool for the detection of variants from an assembly. *Bioinformatics* **32**, 3021–3023 (2016).
69. M. J. Haydon, M. Kawachi, M. Wirtz, S. Hillmer, R. Hell, U. Krämer, Vacuolar nicotianamine has critical and distinct roles under iron deficiency and for zinc sequestration in *Arabidopsis*. *Plant Cell* **24**, 724–737 (2012).
70. S. A. Sinclair, C. Larue, L. Bonk, A. Khan, H. Castillo-Michel, R. J. Stein, D. Grolmund, D. Begerow, U. Neumann, M. J. Haydon, U. Krämer, Etiolated seedling development requires repression of photomorphogenesis by a small cell-wall-derived dark signal. *Curr. Biol.* **27**, 3403–3418.e7 (2017).
71. H. Li, A statistical framework for SNP calling, mutation discovery, association mapping and population genetic parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).
72. S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, P. C. Sham, PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
73. O. Gascuel, BIONJ: An improved version of the NJ algorithm based on a simple model of sequence data. *Mol. Biol. Evol.* **14**, 685–695 (1997).
74. S. Guindon, J.-F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, O. Gascuel, New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).
75. R. N. Gutenkunst, R. D. Hernandez, S. H. Williamson, C. D. Bustamante, Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet.* **5**, e1000695 (2009).
76. S. Ossowski, K. Schneeberger, J. I. Lucas-Lledó, N. Warthmann, R. M. Clark, R. G. Shaw, D. Weigel, M. Lynch, The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science* **327**, 92–94 (2010).
77. P. A. Salomé, K. Bomblies, J. Fitz, R. A. E. Laitinen, N. Warthmann, L. Yant, D. Weigel, The recombination landscape in *Arabidopsis thaliana* F₂ populations. *Heredity* **108**, 447–455 (2012).
78. K. Bomblies, L. Yant, R. A. Laitinen, S.-T. Kim, J. D. Hollister, N. Warthmann, J. Fitz, D. Weigel, Local-scale patterns of genetic variability, outcrossing, and spatial structure in natural stands of *Arabidopsis thaliana*. *PLOS Genet.* **6**, e1000890 (2010).

Acknowledgments: We thank N. Barton, J. Hermisson, T. Karasov, and S. Laurent for comments on the manuscript. We thank the Instituto Nacional de Investigación e Desenvolvimento Agrário (INIDA) in Cape Verde, in particular, Â. Moreno, and the Max Planck Genome Center for performing the Nanopore sequencing and the digital PCR in this study. We thank B. Bizuayehu, B. Frein von Loë, and P. Unger for technical assistance. We thank S. Young and S. Vazquez Reina from the Division of Agricultural and Environmental Sciences at the University of Nottingham for performing the ICP-MS analysis of the soil extracts. We thank P. Düchting and the Ruhr-Universität Bochum for the elemental analysis on plates. **Funding:** This work was supported by Max Planck Society funds, ERC-StG CVI_ADAPT 638810, and EPPN2020 CVI_IONOMICS 114 to A.M.H. and by ERC-AdG 788380 LEAP-EXTREME to U.K. We thank the Max Planck Society for financial support and use of facilities. **Author contributions:** E.T. and A.M.H. conceived and designed the study. E.T. and N.D. performed the laboratory and greenhouse work. E.T., A.M.H., H.D., A.F., C.N., and A.F.E. collected samples. P.F. and D.E.S. conducted ionomics analysis of soil and plant material and contributed to its interpretation. P.J.F. generated the F2 population derived from Col-0 and F13-8. E.T. and J.Z. contributed to the image analysis. M.G. assembled de novo genomes. M.S. and E.T. performed experiments on plates; M.S. and U.K. contributed to its interpretation. E.T., A.F.E., A.M.H., A.F., and S.-A.X. contributed to analyses and interpretation of results. E.T. and A.M.H. wrote the paper. All authors read and edited the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Nanopore and short-read data generated here have been deposited in NCBI SRA under the project number PRJNA802114. Source data files have been deposited on dryad (<https://doi.org/10.5061/dryad.2jm63xsr6>).

Submitted 22 October 2021

Accepted 1 April 2022

Published 18 May 2022

10.1126/sciadv.abm9385