Borrowed alleles and convergence in serpentine adaptation

Brian J. Arnold[a,b], Brett Lahner[c], Jeffrey M. DaCosta[a], Caroline M. Weisman[a], Jesse D. Hollister[d], David E. Salt[c,e], Kirsten Bomblies[a,f], and Levi Yant[a,f,1]

[a]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138; [b]Center for Communicable Disease Dynamics, Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, MA 02115; [c]Department of Horticulture and Landscape Architecture, Purdue University, West Lafayette, IN 47907; [d]Department of Ecology and Evolution, Stony Brook University, Stony Brook, NY 11794-5245; [e]Institute of Biological and Environmental Science, University of Aberdeen, Aberdeen, Scotland AB24 3UU, United Kingdom; and [f]Department of Cell and Developmental Biology, John Innes Centre, Norwich Research Park, Norwich, NR4 7UH, United Kingdom

Edited by Johanna Schmitt, University of California, Davis, CA, and approved May 25, 2016 (received for review January 9, 2016)

Serpentine barrens represent extreme hazards for plant colonists. These sites are characterized by high porosity leading to drought, lack of essential mineral nutrients, and phytotoxic levels of metals. Nevertheless, nature forged populations adapted to these challenges. Here, we use a population-based evolutionary genomic approach coupled with elemental profiling to assess how autotetraploid *Arabidopsis arenosa* adapted to a multichallenge serpentine habitat in the Austrian Alps. We first demonstrate that serpentine-adapted plants exhibit dramatically altered elemental accumulation levels in common conditions, and then resequence 24 autotetraploid individuals from three populations to perform a genome scan. We find evidence for highly localized selective sweeps that point to a polygenic, multitrait basis for serpentine adaptation. Comparing our results to a previous study of independent serpentine colonizations in the closely related diploid *Arabidopsis lyrata* in the United Kingdom and United States, we find the highest levels of differentiation in 11 of the same loci, providing candidate alleles for mediating convergent evolution. This overlap between independent colonizations in different species suggests that a limited number of evolutionary strategies are suited to overcome the multiple challenges of serpentine adaptation. Interestingly, we detect footprints of selection in *A. arenosa* in the context of substantial gene flow from nearby off-serpentine populations of *A. arenosa*, as well as from *A. lyrata*. In several cases, quantitative tests of introgression indicate that some alleles exhibiting strong selective sweep signatures appear to have been introgressed from *A. lyrata*. This finding suggests that migrant alleles may have facilitated adaptation of *A. arenosa* to this multihazard environment.

adaptation | plant | gene flow | population genomics

Serpentine barrens offer powerful venues for the study of multitrait adaptations. Soils at these sites feature dramatically skewed elemental contents, phytotoxic levels of heavy metals, drought risk, and very poor mineral nutrition (1–3). A defining characteristic of serpentine soils is a greatly reduced Ca:Mg ratio along with low K, N, and P, resulting in severe ion homeostasis challenges for plant colonists (4–6). Serpentine soils are also highly porous and thus chronically drought prone. As a result of these challenges, serpentine barrens are characterized by minimal ecosystem productivity and high rates of endemism (reviewed in refs. 2 and 3). Evolution has nevertheless repeatedly forged plant populations that overcome these hazards, making serpentine sites an important natural model for ecology, evolution, and physiology. Given the quantifiable challenges of serpentine adaptation presented by strongly skewed elemental levels and dehydration risk, adapted populations present a valuable opportunity to identify loci underlying adaptations important for understanding basic evolutionary processes, as well as candidate genes for rational crop design for tolerance of challenging growth conditions such as low nutrient soils, metal, or drought.

A genomic understanding of adaptation to serpentine soils and their diverse challenges remains in its infancy. Within the molecularly tractable *Arabidopsis* genus, at least two species have been reported to have independently colonized serpentine barrens: diploid *Arabidopsis lyrata* (7) and autotetraploid *Arabidopsis arenosa*. As an obligate outcrosser, *A. arenosa* exhibits very high genetic diversity, a small (~200 Mb) genome, and very large effective population sizes (8, 9), enabling fruitful population genomic analysis (9–12). Tetraploid *A. arenosa* populations have colonized diverse habitats throughout central and northern Europe (13, 14). There is also evidence that hybridization of *A. arenosa* with *A. lyrata* resulted in a hybrid that escaped the ecological niche of its progenitors (15). Here, we focus on an *A. arenosa* population reported in a 1955 botanical survey of a serpentine barren on Gulsen Mountain in Austria (16).

We returned to Gulsen in 2010 and found an extant *A. arenosa* population on the serpentine site and also collected from 28 other sites across Europe. We first used quantitative elemental profiling of soil from *A. arenosa* sites, as well as leaves grown from plants in common gardens, to find that serpentine plants show a constitutively altered ability to control accumulation of elements in their leaves that matches the elemental challenges of their native soils. We performed a detailed demographic analysis that revealed gene flow into the serpentine population from both a nearby *A. arenosa* population, as well as from *A. lyrata*. This introgression signal from *A. lyrata* is specific to the serpentine population and not evident in other *A. arenosa* populations we sampled. We resequenced

Significance

Serpentine barrens are enormously hostile to plant life. Understanding how plants survive such a perfect storm of low mineral nutrient, drought prone, and toxic metal rich conditions offers a powerful model of adaptation and may help design resilient crops. Advances in genomics enable population-wide views of selection and deep insight into demographic histories. These approaches are agnostic to phenotype and can indicate which traits were most important in complex adaptations and, at the same time, provide novel candidate genes. Here, we identified candidate genes for serpentine adaptation and provide evidence that some selected alleles were borrowed from a related species, whereas others were independently involved in separate adaptation events in different species.

Author contributions: B.J.A., D.E.S., K.B., and L.Y. designed research; B.J.A., B.L., J.M.D., C.M.W., D.E.S., and L.Y. performed research; B.J.A., J.M.D., J.D.H., D.E.S., and L.Y. contributed new reagents/analytic tools; B.J.A., B.L., J.M.D., D.E.S., and L.Y. analyzed data; L.Y. conceived and headed the project; and B.J.A., D.E.S., K.B., and L.Y. 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 sequence reported in this paper have been deposited into the NCBI Sequence Read Archive (SRA) (BioProject PRJNA325082).

[1]To whom correspondence should be addressed. Email: levi.yant@jic.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1600405113/-/DCSupplemental.

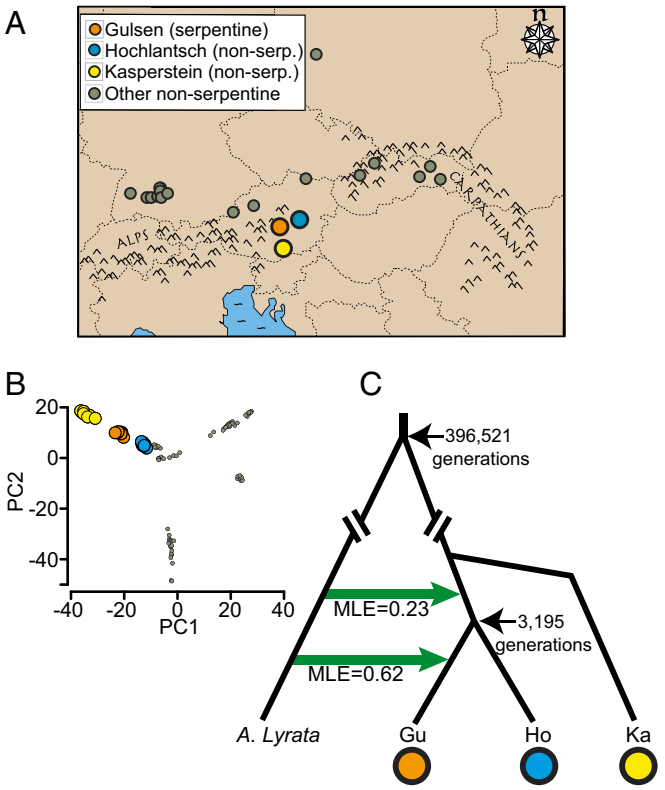

Fig. 1. *A. arenosa* populations sampled for this study. (*A*) Locations of the 29 *A. arenosa* populations sampled. Orange dot gives the location of the focal Gulsen (GU) serpentine population along with other highlighted populations at Hochlantsch (HO) and Kasparstein (KA). Note: one Swedish location is not pictured (see *SI Appendix*, Table S1 for global positioning system locations). (*B*) PCA of *A. arenosa* range-wide showing relatedness between highlighted populations. (*C*) Lineage topology highlighting the major introgression events (green arrows), with MLEs for introgression in lineages per generation and MLEs for divergence times in generations.

individuals from the serpentine *A. arenosa* population as well as the two most closely related nonserpentine populations and found the strongest signatures of selection in many genes with functions relevant to serpentine challenges. Interestingly, in several cases, selection acted on alleles that also show evidence of introgression from *A. lyrata* according to genome-wide quantitative tests. Our results highlight the important role that introgression may have played in these adaptations. Finally, we compare our findings to a previous study of an independent serpentine adaptation in *A. lyrata* to assess the degree of convergent evolution and find that some of the same genes were targeted by selection in these independent events.

Results and Discussion

Elemental Accumulation Profiles Are Highly Altered in Serpentine *A. arenosa*. As noted above, soils at serpentine sites are characterized by extremely low Ca:Mg ratios, low macronutrient (e.g., K and S) availability, high levels of particular metals, and high risk of dehydration due to porosity and low plant cover (4, 5). We noted that *A. arenosa* was listed in a botanical survey of a serpentine site on Gulsen Mountain, near Kraubath an der Mur, Austria (16) (Fig. 1*A*). To understand whether and how *A. arenosa* adapted to this challenging environment, we first analyzed the mineral nutrient and trace element composition of soil samples we collected from Gulsen and other *A. arenosa* sites. Relative to other *A. arenosa* sites, soil from Gulsen had the lowest levels of macronutrients K and S, very low Ca:Mg ratios, the highest levels of the heavy metal Ni, but very low levels of Cu, Zn, and Cd (Fig. 2, orange dots in brown soil distributions, and Dataset S1). These soil characteristics are consistent with Gulsen being a serpentine site (1–6).

To test the mineral nutrient uptake characteristics of these plants, we then analyzed leaf tissue of plants grown in common conditions in fertile artificial soil from seeds collected at Gulsen and 28 nonserpentine *A. arenosa* sites, including all of the sites from which we also sampled soils (*SI Appendix*, Table S1). Elemental analysis showed that Gulsen plants are similarly extreme outliers for the same elements as the serpentine soil, but in the opposite direction (Fig. 2, orange dots in green plant distributions, and Dataset S2). Relative to plants sampled from the other 28 populations, Gulsen plants accumulated the highest levels of K and S, excluded Ni and Mg, exhibited the highest Ca:Mg ratios, and took up comparatively high levels of Cu, Zn, and Cd. These findings indicate that, relative to other *A. arenosa* populations, the plants from Gulsen have genetically adapted to the challenging mineral composition of the serpentine site by a complex suite of adaptations, including exclusion or accumulation of different elements in accordance with local soil concentrations. These patterns are consistent with data from other serpentine adapted species (reviewed in refs. 2 and 3).

Demographic Analysis. To confirm the genetic placement of Gulsen among range-wide *A. arenosa* populations, we used a restriction site-associated DNA sequencing (RAD-seq) dataset from ref. 12 that surveyed 20 broadly distributed *A. arenosa* populations. We found that Gulsen is positioned neatly between Hochlantsch and Kasparstein in a principal component analysis (PCA) (Fig. 1*B*), consistent with ref. 12, but is most closely related to Hochlantsch in a simple phylogenetic analysis (*SI Appendix*, Section S1 and Table S2). This finding confirms that Gulsen is a member of the alpine lineage of *A. arenosa* and that, of all populations sampled across the *A. arenosa* range, the geographically most proximal populations (Hochlantsch and Kasparstein) provide the most closely related nonserpentine populations to Gulsen. Therefore, we chose Hochlantsch and Kasparstein as comparison groups for population resequencing.

We individually barcoded and sequenced a total of 24 autotetraploid individuals from Gulsen, Hochlantsch, and Kasparstein to an average depth of 21× aligned coverage per individual (*SI Appendix*, Table S3). Because all plants sequenced are autotetraploids, this approach samples 96 chromosomes at each site in the genome. Following a previously successful approach (9–12, 17), we aligned to

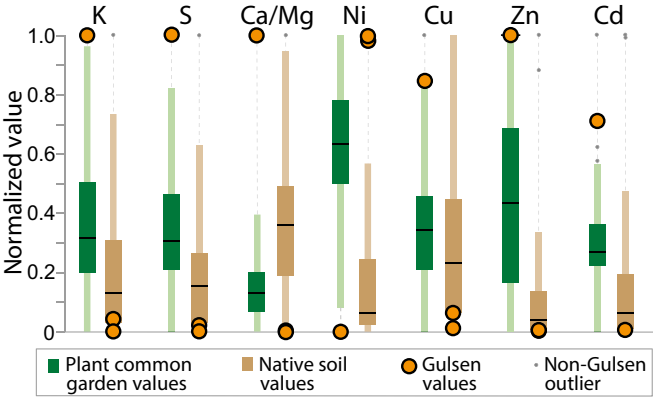

Fig. 2. Serpentine *A. arenosa* is an extreme outlier for the accumulation of many elements. Elemental profiling of 29 *A. arenosa* populations. Green distributions represent plant tissue data, brown distributions represent data from soil collected at plant sites. Orange dots indicate position in distribution where the serpentine autotetraploid Gulsen sample lies. (*A*) S, sulfur; K, potassium; Ca/Mg, calcium-to-magnesium ratio. (*B*) Ni, nickel; Cu, copper; Zn, zinc; Cd, cadmium. We normalized all values to the 0–1 range using feature scaling, where $x' = (x - x_{min})/(x_{max} - x_{min})$.

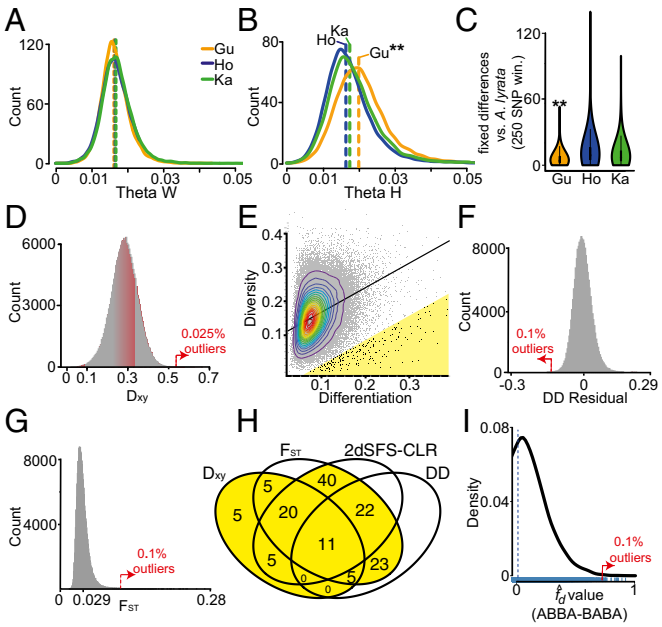

Fig. 3. Measures of differentiation. (*A*) Watterson estimator θ_W diversity in resequenced populations over genome windows. The vertical dashed line for each population gives the mean. (*B*) θ_H, a diversity metric sensitive to extreme frequency SNPs (double asterisk signifies that Gulsen distribution is highly significantly different ($P < 2.2e\text{-}16$) from Hochlantsch or Kasparstein populations). (*C*) Mean number of fixed differences relative to Austrian *A. lyrata* in windows across the genome in each population (double asterisk signifies that Gulsen distribution is highly significantly different [$P < 2.2e\text{-}16$] from HO or KA populations). (*D*) D_{xy}, absolute net divergence between Gulsen and nonserpentine *A. arenosa* over genomic windows. (*E*) Relationship of diversity and differentiation in windows, indicating 0.1% empirical outliers in yellow. (*F*) DD residual values, indicating outliers with lower diversity for their given level of differentiation, a classic selective sweep signature. (*G*) F_{ST} distribution with outliers marked. (*H*) Overlap of outlier gene loci by all tests. (*I*) Positive f_d values from four taxon ABBA-BABA test with outliers marked and blue rug indicating each window value.

the closely related *A. lyrata* genome (18) to call SNPs. We obtained information for 52 million nucleotide positions, of which 4.9 million are polymorphic with confident SNP calls following all filtering steps (*SI Appendix*, Sections S2 and S3). We detected extensive shared variation between serpentine and nonserpentine populations (2.7 million sites). These patterns are consistent with recent colonization by multiple individuals and/or substantial levels of gene flow between populations.

Because serpentine barrens present a broad array of challenges to colonizers, we expected the Gulsen population to exhibit very low effective population sizes, resulting from a hypothetical bottleneck upon colonization. Surprisingly, however, the Gulsen population had normal diversity levels (Watterson estimator θ_W) (19) compared with nearby nonserpentine populations (Fig. 3*A*) and similar estimates of Tajima's *D* in comparison with 13 other autotetraploid *A. arenosa* populations sampled from across the *A. arenosa* range (*SI Appendix*, Fig. S1), indicating the lack of an extreme bottleneck and/or gene flow. Intriguingly, Gulsen had significantly higher ($P < 2.2e\text{-}16$) values of θ_H (Fig. 3*B*) (20), a diversity metric sensitive to high-frequency polymorphisms. We hypothesized that interspecies admixture between the Gulsen population and *A. lyrata*—the species used as the reference to align sequence data and polarize mutations—drives this excess of high-frequency–derived mutations. These sites, derived with respect to the reference sequence, are fixed in the other *A. arenosa* population samples but are not fixed for the derived allele in Gulsen, due to an influx of reference-like (ancestral) polymorphism from *A. lyrata* (Fig. 3*C*).

To better understand the demographic history of the Gulsen population and its relatedness to the nearest nonserpentine *A. arenosa* population we sampled, Hochlantsch, we explicitly modeled population histories using the coalescent, which we adapted for autotetraploids (21). Because we detected hints of admixture between the Gulsen population of *A. arenosa* and *A. lyrata*, we included an Austrian *A. lyrata* genome sequence (from ref. 12) as an outgroup to quantify possible interspecies gene flow. With these three populations and five possible migration parameters (including migration between Gulsen, Hochlantsch, and *A. lyrata*) (*SI Appendix*, Table S4), we used a model selection approach (22) to determine which of these migration parameters were statistically supported by the data and thus potentially biologically meaningful. We constructed 32 different migration models, each a distinct permutation of the five possible migration rates. We fit each model to fourfold degenerate SNP data using fastsimcoal2 (23) and used the model likelihoods to calculate an Akaike weight for each or the probability a particular model is best among all candidates (*SI Appendix*, Sections S1 and S4). The Pearson correlation between the number of migration parameters and the model likelihood was 0.61, suggesting not all migration parameters explain the data.

The model with the unambiguously highest Akaike weight (*SI Appendix*, Table S4) contained four of five possible migration parameters (Fig. 1*C*). Maximum likelihood estimates (MLEs) of migration probabilities were highest from *A. lyrata* into Gulsen (population migration rate $4N_em = 0.62$ migrant lineages per generation) and were significantly higher than those for interspecific introgression into Hochlantsch using 90% confidence intervals (CIs) (*SI Appendix*, Table S5). Whereas the model selection analysis suggests each of these migration parameters has statistical support, migration probability CIs contained very small values ($4N_em \sim 0$), except for *A. lyrata* to Gulsen (90% CI $4N_em = 0.43$–0.85). Parameter MLEs also indicate a divergence time of 3,195 generations (90% CI = 1,398–4,555) between Gulsen and Hochlantsch. Given that size estimates of these populations are on the order of $4N_e \sim 30{,}000$ haploid chromosomes, where N_e is the effective number of tetraploids, the divergence time MLE is $\sim 0.1 \times 4N_e$ generations, just a fraction of the average time it takes for rare or intermediate-frequency neutral mutations to fix in a finite population ($\sim 3 \times 4N_e$ to $\sim 4 \times 4N_e$ generations) (24). These findings suggest a very recent colonization of this serpentine barren with few detectable neutral changes between Gulsen and Hochlantsch, especially considering the potential for extensive gene flow between them (90% CIs for population migration rate from Hochlanstch to Gulsen $4N_em = 0.01$–1.57) (*SI Appendix*, Table S5).

Selective Sweeps Associated with Serpentine Adaptation. To identify loci under selection in the Gulsen population, we conscribed the genome into 25-SNP windows in which we characterized metrics of both absolute (D_{xy}) and relative (F_{ST}) divergence, as well as the site frequency spectrum (Fig. 3 and *SI Appendix*, Fig. S2). We chose 25-SNP windows (median width = 391 bp) because estimates of diversity between adjacent windows of this size were uncorrelated, consistent with low linkage disequilibrium in *A. arenosa* (*SI Appendix*, Fig. S3). To capture selection on regulatory changes, we included genes that either overlap with or lie within 2 kb of an outlier window.

To obtain top outliers exhibiting the most robust evidence of selective sweep, we retained the extreme outlier windows from four window-based differentiation and allele frequency spectrum metrics comparing Gulsen with Hochlantsch and Kasparstein: (*i*) maximum absolute net divergence (D_{xy}) (Fig. 3*D*) (25), (*ii*) maximum relative divergence (F_{ST}; Fig. 3G) (26), (*iii*) maximum negative residuals of a diversity/differentiation (DD) metric inspired by the Hudson–Kreitman–Aguade test (DD residual test) (Fig. 3 *E* and *F*) (10), and (*iv*) top scoring windows from a 2D site frequency spectrum composite likelihood test (2dSFS-CLR)

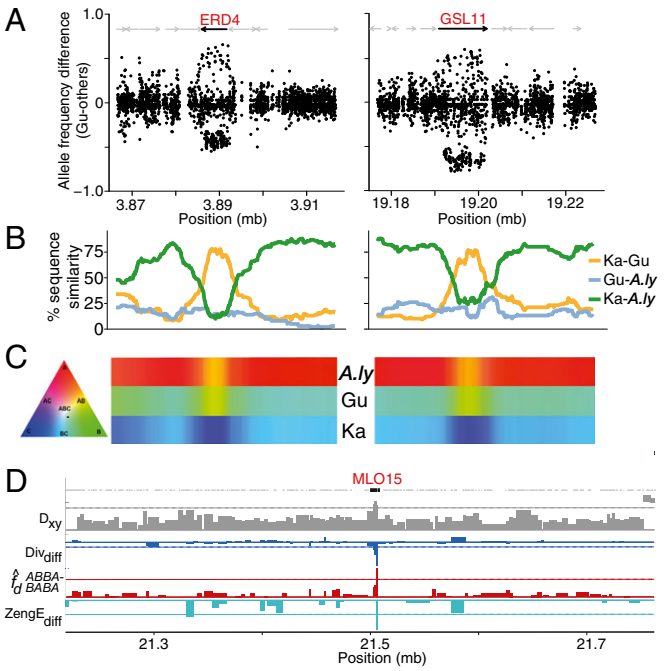

Fig. 4. Selective sweeps on *A. lyrata*-like alleles in serpentine *A. arenosa*. (*A*) Allele frequency differences in example differentiated regions. Dots represent polymorphic SNPs. The *x* axis gives chromosome location; *y* axis gives degree of differentiation calculated by plotting the difference in allele frequencies between serpentine and nonserpentine populations. Arrows indicate gene models. Black arrow indicates sweep candidate with localized differentiation. (*B*) Linear plot showing the proportion of SNPs shared between the three pairwise population comparisons in the same region as in *A*. (*C*) Sequence similarity at the same regions among *A. lyrata*, Gulsen, and Kasparstein visualized using a color triangle. Areas where two rows show the same color (yellow) indicate localized high similarity specifically between Gulsen and *A. lyrata*, but not Kasparstein. (*D*) Genomic view of divergence and gene flow metrics at a postive ABBA-BABA outlier and top sweep candidate locus. D_{xy} gives net divergence, Div_{diff}, a selective sweep signature (relatively reduced diversity specifically in Gulsen vs. other *A. arenosa*; more negative values indicate specifically low diversity in Gulsen), f_d gives ABBA-BABA outlier status, $ZengE_{diff}$, negative values give localized negative excesses of rare variants in Gulsen (also see *SI Appendix*, Section S5). Dashed lines represent 1% outlier levels.

following ref. 27. (For tests, see *SI Appendix*, Section S5.) Following inspection of allele frequency plots spanning all outlier windows at different cutoffs, we found that the top 0.025% outliers of the absolute net divergence metric D_{xy} yielded the most dramatic signatures of selective sweep (48 windows overlapping or within 2 kb of 51 gene-coding loci) (*SI Appendix*, Table S7). We also detected convincing outliers by the overlap of the other metrics: top candidates were retained from the F_{ST} 0.1% outlier list if they were also among the 0.1% DD residual outliers (DD outliers exhibit low diversity in sweep regions relative to differentiation, a classic sweep signature) (Fig. 3*E*) or 0.1% 2dSFS-CLR test outliers. Finally, to enable a direct comparison between this study and another study of serpentine adaptation (7), we retained loci within 2 kb of any SNP with >0.8 absolute allele frequency difference between serpentine and nonserpentine populations, resulting in a further 26 candidate loci at genome-wide maximally diverged SNPs. Together, these approaches yielded the highest proportions of loci with sharp peaks of differentiation that trailed off immediately flanking the peaks, as previously observed in *A. arenosa* (10, 12) (Fig. 4). Despite using hard cutoffs of extreme outliers, there was considerable overlap in our highest stringency lists (Fig. 3*H*), with only 162 loci represented on this final list of top sweep candidates (*SI Appendix*, Section S5 and Tables S6 and S8).

Serpentine-Specific Sweeps Represent Processes Involved in "Serpentine Syndrome." A broad range of processes is represented in our top sweep candidates list. Approximately half of the genes are documented to function in, or show altered expression as a result of, traits or stresses in *Arabidopsis thaliana* that are directly linked to the challenges of persisting on serpentine barrens (*SI Appendix*, Table S8), with each process represented by several genes. Many of these categories fit well with observed elemental challenges at Gulsen (e.g., low K^+ and S^{2-}, high Mg^{2+}, and low Ca:Mg ratios) (Fig. 2 and *SI Appendix*, Table S8). For example, the *A. thaliana* orthologs of many of these genes encode proteins involved in ion (particularly $S04^{2-}$, K^+, NO_3^+, Mg^{2+}, Ca^{2+}) transport or signaling, such as sulfate transporter 1;1 (SULTR1;1), K^+ uptake permease 9 (KUP9), and ammonium transporter 2;1 (AMT2;1), along with Casparian strip membrane domain protein 1 (CASP1), which is involved in the Casparian strip, a critical root component that broadly influences mineral nutrient uptake, water uptake, and stress resistance (28–30). CASP1 and AMT2;1 exhibit five and seven high-frequency amino acid substitutions differentiated between Gulsen and other *A. arenosa* populations, respectively.

Whereas many of the identified sweep candidates have orthologs in *A. thaliana* that are root expressed or play roles in root architecture and elemental challenges, others have been demonstrated to play roles in intracellular ion dynamics, including proteins involved in Ca^{2+} signaling and transport, Ca^{2+}-modulated signaling networks, and cellular stress responses (31) (*SI Appendix*, Table S8 and Datasets S3–S7), indicating adaptation to changes in intracellular physiology. It is interesting that the primary Ca^{2+} channel in the vacuole, two pore channel (TPC1) (32) contains high-frequency–derived changes in Gulsen and is also a 0.1% DD residual outlier, along with many Ca^{2+}-related genes. *TPC1* levels directly modulate salt tolerance and control the Ca^{2+}-mediated root-to-shoot stress signal (31). Indeed, many of the top loci are implicated in stress signaling and tolerance, such as *early responsive to dehydration stress protein 4* (*ERD4*) and *high expression of osmotically responsive genes 2* (*HOS2*) (references to functional assessments in *SI Appendix*, Table S8).

Early flowering is a common drought escape mechanism and the Gulsen population is no exception. Gulsen plants flower much earlier than their closest relatives (days to open flower: Gulsen = 49 ± 1.3, Hochlantsch = 100 ± 12, Kasparstein = 105 ± 14; *SI Appendix*, Fig. S4). It is interesting to note that in addition to stress signaling and tolerance, *HOS2* also controls flowering time (33). We also see other genes controlling flowering time in the top sweep candidates, including *LACCASE 8* (34), among others in each 0.1% outlier list. This finding, combined with diverse loci controlling ion transport, signaling, intracellular ion dynamics, and stress signaling, indicates that a spectrum of functionally diverse loci underlies serpentine adaptation, rather than a small number of "master regulators."

Introgression and Selection on *A. lyrata* Alleles Among Top Sweep Candidates. We observed localized high similarity to Austrian *A. lyrata* in regions overlapping several top sweep candidates specifically in the Gulsen population. This pattern is maintained across entire gene-coding regions, directly overlapping selective sweep signatures (compare Fig. 4*A* with 4*B* and 4*C*). To understand these signals in a genomic context, we constructed a window-based four-taxon analysis following ref. 35 that tests for an excess of shared variants between Gulsen and *A. lyrata*, using *A. thaliana* as the outgroup (*SI Appendix*, Section S4). For biallelic sites with alleles A and B, ABBA and BABA patterns are equally likely if incomplete lineage sorting is the sole cause of paraphyly, with gene flow driving these patterns to diverge in frequency. Extreme ABBA patterns (top f_d values) indicate increased allele sharing between Gulsen and *A. lyrata*. Consistent with the demographic and allele frequency spectrum results above (genome-wide *A. lyrata*-like SNPs and high θ_H in Gulsen specifically; Fig. 3 *B* and

C), the mean f_d value was positive, indicating gene flow from *A. lyrata*, but the genome-wide distribution was not significantly positive by bootstrap or jackknife resampling (*SI Appendix*, Section S6).

Of the f_d outliers in the top 99th percentile (24 windows genome-wide; values >0.73; Fig. 3*I*), six loci are present among our 162 top sweep candidates (*SI Appendix*, Table S9) [less than one expected by chance ($P < 2.2e-05$); hypergeometric test, nearby gene loci collapsed into single observations to ensure independence] (*SI Appendix*, Section S5). We note that *A. lyrata* alleles are not retained genome-wide, which suggests that after hybridization with *A. lyrata*, selection favored increases in the abundance of *A. lyrata* alleles at only a few loci (Fig. 4 and *SI Appendix*, Table S9), whereas signals of introgression in the rest of the genome largely eroded. Thus, this subset of sweep loci are candidates for interspecies adaptive gene flow. Interspecific hybridization has been noted in other systems (reviewed in ref. 36) and introgression can occur even when strong barriers exist (37). Indeed, hybridization has been reported between *A. arenosa* and *A. lyrata* (15), and a substantial signal of this hybridization is clear in our coalescent models (Fig. 1*C*), which also provide evidence for a history of introgression specifically between the Gulsen *A. arenosa* population and *A. lyrata*. Importantly, however, even though *A. lyrata* introgression contributed alleles, the Gulsen *A. arenosa* population is not a true hybrid population (i.e., there is no evidence of rampant hybrid formation or widespread retention of *A. lyrata* polymorphisms outside these few selected loci).

Convergent Evolution Between Serpentine *A. arenosa* and *A. lyrata*. We compared our results to a genome scan of diploid serpentine populations of *A. lyrata* in Scotland and the United States (7) and observed evidence of convergent evolution that independently targeted the same loci. Using the same reference genome assembly as our study, the *A. lyrata* study detected 96 SNPs that exhibit allele frequency differences of greater than 80% between serpentine and nonserpentine populations. We tested whether any of our SNPs matching the identical criterion are situated near outliers in the *A. lyrata* study. We found that 9 of our 77 most differentiated SNPs lie very near (within 2 kb) 9 of the 96 top candidate SNPs reported in serpentine *A. lyrata* ($P < 6.1e-09$; hypergeometric test, nearby SNPs collapsed into single observations to ensure independence; *SI Appendix*, Section S5). These 9 SNPs overlap or are directly adjacent to six gene loci, among which are *KUP9* and *TPC1* (*SI Appendix*, Table S10). This underscores the importance of K^+ and Ca^{2+} in serpentine adaptation in both *A. arenosa* and *A. lyrata* (1–5). Both genes contain high-frequency derived changes specific to independent serpentine populations (Austria in this study, Scotland in ref. 7). The use of distinct derived alleles at the same genes suggests that the possible solutions to serpentine-associated challenges may be relatively constrained, despite the abundance of genes that could in principle affect K^+ and Ca^{2+}.

The vacuolar channel encoded by *TPC1* is regulated by changes in Ca^{2+} levels, and a point mutant in *TPC1* increases vacuolar Ca^{2+} storage (38). *TPC1* levels control Ca^{2+}-mediated root-to-shoot stress signaling (28). Given the severely Ca^{2+}-challenged environment of serpentine sites, including Gulsen (Fig. 2), we speculate that the high-frequency changes we see in *TPC1* and other Ca^{2+}-related genes may potentially act as a molecular rheostat, compensating for globally decreased Ca^{2+} availability. In addition to *TPC1* and *KUP9*, we see nine additional genes among our top sweep candidates that are also under the strongest selection in *A. lyrata* (*SI Appendix*, Table S11) ($P < 1.3e-06$; hypergeometric test as in *SI Appendix*, Section S5). Among these are *ferroportin 2* (*FPN2*), which encodes a Ni transport protein, orthologous to the iron efflux transporter ferroportin in animals, as well as a hydrolase implicated in calmodulin binding (ortholog of AT5G37710). Of particular relevance to the very high Ni found at Gulsen, mutants of *FPN2* exhibit increased Ni sensitivity and it has been proposed that FPN2 transports Ni, Co, and Fe into the vacuole (39, 40). Why these genes and others are under selection in two independent serpentine colonizations merits further study (41, 42).

Conclusions

We have shown that an autotetraploid *A. arenosa* population adapted to a highly challenging serpentine site and exhibits strong evidence of selection in genes that control specific ion homeostasis-related traits, as well as drought adaptation, providing strong candidates for control of these traits. Several of the alleles under selection were likely introgressed from *A. lyrata*. Furthermore, by comparing to a genome scan in diploid *A. lyrata*, we present evidence of convergent evolution, with distinct alleles of 11 genes having been independently targeted following serpentine colonization in these two species. The overlap between selected genes in serpentine-endemic *A. arenosa* and *A. lyrata* suggests that diploid and tetraploid adaptations to serpentine are not qualitatively different. This work advances our understanding of the polygenic basis of multitrait adaptation and its repeatability across species and gives an example of selective sweeps that occurred in the context of substantial levels of inter- and intraspecific gene flow.

Methods

Detailed descriptions of samples and methods are provided in *SI Appendix*. All sequence data are freely available in the National Center for Biotechnology Institute SRA database (BioProject PRJNA325082).

Plant Growth and Treatment. Plant materials and growth conditions for genomic analysis were as previously described (9). Plants for inductively coupled plasma-mass spectrometry (ICP-MS) analysis were grown in an exclusive growth room to avoid plant pathogens, which obviated the need for pesticide applications that could interfere with the trace metal analyses or otherwise add noise to the experiment. Seeds were sown in 20-row trays with each accession occupying two separated rows in Pro-Mix (Premier Horticulture), a soilless mix. Excess seeds were sown to try to ensure full rows of six plants each, and plants were thinned to six per row after germination. The trays were stratified at 4 °C for 3 d. The plants were then grown in the growth room of the Purdue Ionomics Center with 8 h light (90 $mmol \cdot m^2 \cdot s$) and 16 h dark (to prevent bolting), at temperatures ranging from 19 °C to 22 °C. On subsequent days, plants were bottom watered twice a week with modified to one-quarter strength Hoagland's solution. Several leaves were harvested from 5-wk-old plants for analysis, with care being taken to harvest equivalent leaves from each plant.

Elemental Analysis of Leaf Tissue. Tissue samples were dried at 92 °C for 20 h in Pyrex tubes. After cooling in a desiccator for 45 min, samples were digested at 110 °C for 4 h with 0.7 mL of concentrated nitric acid to which indium had been added as an internal standard and diluted to 6.0 mL. Analysis was performed on an ICP-MS (Elan DRCe; PerkinElmer). A liquid reference material, composed of pooled leaf samples, was run to correct for drift and between-run variation. All samples were normalized, as determined with an iterative algorithm using the best-measured elements and implemented in the ionomicshub.org database (www.ionomicshub.org/home/PiiMS), under the Education > How-To drop menus).

Elemental Extraction of Soils. Soil samples were dried and about 5 g of each was weighed into 50-mL Falcon tubes. Each was extracted with 25 mL of water by shaking for 1 h and centrifuged before sampling, adding nitric acid to 5% (vol/vol), and analyzing with an Elan DRCe ICP-MS.

Flowering Time Measurements. To measure flowering time, we germinated seeds collected from Gulsen ($n = 39$), Kasparstein ($n = 17$), and Hochlantsch ($n = 30$) on 1/2× MS plates. We recorded germination date by root emergence on agar plates and then transferred seedlings to soil (1/2 Sunshine Mix no. 1, 1/2 vermiculite). We grew plants in Conviron MTPC-144 chambers for 8 h dark at 12 °C, 4 h light (cool-white fluorescent bulbs) at 18 °C, 8 h light at 20 °C, 4 h light at 18 °C. We quantified flowering time as the first day that flower buds were visible in the center of the rosette. We tested whether distributions differed using a two-tailed *t* test for each comparison.

Library Preparation and Sequencing. Genomic DNA was extracted from leaf material as in ref. 10. DNA libraries were prepared using Illumina library preparation kits and sequenced on a HiSeq2500 (*SI Appendix*, Section S2).

Read Mapping and Genotyping. Data generated in this study were processed through the entire alignment, genotyping, and analysis pipeline in parallel with raw reads from individuals generated in refs. 9 and 10. Briefly, reads were mapped to the repeatmasked Lyrata107 genome (18) using Stampy (43). *A. arenosa* autotetraploids retain tetrasomic inheritance (9), so there are no homeologs, meaning all reads are appropriately mapped to the same set of loci represented in the diploid genome build. Resultant bam files were processed with Samtools (44) and Picard (picard.sourceforge.net/), and genotyped following GATK best practices (*SI Appendix*, Section S2). Filtering information is given in (*SI Appendix*, Section S3). Gene information was inferred with the *A. lyrata* version 2 annotation (45).

Genomic and Demographic Analysis. Only sites passing all filters were retained for analysis (*SI Appendix*, Sections S2 and S3). We reconstructed the demographic history of Gulsen and Hochlantsch using coalescent simulations and neutral sites (fourfold degenerate). After observing evidence of interspecific admixture between Gulsen and *A. lyrata*, we included a single Austrian *A. lyrata* genome sequence to represent an outgroup population to quantify this interspecific gene flow. We fit various migration models to the data via coalescent simulations (*SI Appendix*, Section S4) using the program fastsimcoal2 to obtain likelihoods for each model.

For the model with the highest Akaike weight, we constructed 90% nonparametric bootstrap confidence intervals (sampling fourfold degenerate SNP matrix with replacement). To scan the genome for signs of selective sweep between groups, we used four metrics across 193,881 25-SNP nonoverlapping genomic windows: D_{xy}, F_{ST}, DD residual, and 2dSFS-CLR test (*SI Appendix*, Section S5). All analyses were performed using Python3, Perl, and R scripts and are freely available. To quantify levels of introgression across the genome, we constructed a four-taxon ABBA/BABA test similar to f_d in Martin et al. (35) (*SI Appendix*, Section S6).

ACKNOWLEDGMENTS. We thank members of the L.Y. and K.B. laboratories for helpful discussions. This work was supported through the European Research Council Grant StG CA629F04E (to L.Y.); a Harvard University Milton Fund Award (to K.B.); Ruth L. Kirschstein National Research Service Award 1 F32 GM096699 from the NIH (to L.Y.); National Science Foundation Grant IOS-1146465 (to K.B.); NIH National Institute of General Medical Sciences Grant 2R01GM078536 (to D.E.S.); and Biotechnology and Biological Sciences Research Council Grant BB/L000113/1 (to D.E.S.).

1. Proctor J, Woodell SR (1975) The ecology of serpentine soils. *Adv Ecol Res* 9:255–366.
2. Brady KU, Kruckeberg AR, Bradshaw HD, Jr (2005) Evolutionary ecology of plant adaptation to serpentine soils. *Annu Rev Ecol Evol Syst* 36(1):243–266.
3. Harrison S, Rajakaruna N (2011) *Serpentine: The Evolution and Ecology of a Model System* (Univ of California Press, Berkeley).
4. Vlamis J, Jenny H (1948) Calcium deficiency in serpentine soils as revealed by adsorbent technique. *Science* 107(2786):549.
5. Walker RB, Walker HM, Ashworth PR (1955) Calcium-magnesium nutrition with special reference to serpentine soils. *Plant Physiol* 30(3):214–221.
6. Woodell SR, Mooney HA, Lewis H (1975) The adaptation to serpentine soils in California of the annual species *Linanthus androsaceus* (Polemoniaceae). *Bull Torrey Bot Club* 102(5):232–238.
7. Turner TL, Bourne EC, Von Wettberg EJ, Hu TT, Nuzhdin SV (2010) Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. *Nat Genet* 42(3):260–263.
8. Schmickl R, Paule J, Klein J, Marhold K, Koch MA (2012) The evolutionary history of the *Arabidopsis arenosa* complex: Diverse tetraploids mask the Western Carpathian center of species and genetic diversity. *PLoS One* 7(8):e42691.
9. Hollister JD, et al. (2012) Genetic adaptation associated with genome-doubling in autotetraploid Arabidopsis arenosa. *PLoS Genet* 8(12):e1003093.
10. Yant L, et al. (2013) Meiotic adaptation to genome duplication in Arabidopsis arenosa. *Curr Biol* 23(21):2151–2156.
11. Wright KM, et al. (2015) Selection on meiosis genes in diploid and tetraploid *Arabidopsis arenosa*. *Mol Biol Evol* 32(4):944–955.
12. Arnold B, Kim ST, Bomblies K (2015) Single geographic origin of a widespread autotetraploid *Arabidopsis arenosa* lineage followed by interploidy admixture. *Mol Biol Evol* 32(6):1382–1395.
13. Koch MA, Matschinger M (2007) Evolution and genetic differentiation among relatives of *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* 104(15):6272–6277.
14. Hoffmann MH (2005) Evolution of the realized climatic niche in the genus *Arabidopsis* (Brassicaceae). *Evolution* 59(7):1425–1436.
15. Schmickl R, Koch MA (2011) *Arabidopsis* hybrid speciation processes. *Proc Natl Acad Sci USA* 108(34):14192–14197.
16. Eggler J (1955) Ein Beitrag zur Serpentinvegetation in der Gulsen bei Kraubath in Obersteiermark. *Mitt Naturw Ver Steiermark* 85:27–72.
17. Arnold B, Corbett-Detig RB, Hartl D, Bomblies K (2013) RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. *Mol Ecol* 22(11):3179–3190.
18. Hu TT, et al. (2011) The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. *Nat Genet* 43(5):476–481.
19. Watterson GA (1975) On the number of segregating sites in genetical models without recombination. *Theor Popul Biol* 7(2):256–276.
20. Fu YX (1995) Statistical properties of segregating sites. *Theor Popul Biol* 48(2):172–197.
21. Arnold B, Bomblies K, Wakeley J (2012) Extending coalescent theory to autotetraploids. *Genetics* 192(1):195–204.
22. Johnson JB, Omland KS (2004) Model selection in ecology and evolution. *Trends Ecol Evol* 19(2):101–108.
23. Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M (2013) Robust demographic inference from genomic and SNP data. *PLoS Genet* 9(10):e1003905.
24. Kimura M, Ohta T (1969) The average number of generations until fixation of a mutant gene in a finite population. *Genetics* 61(3):763–771.
25. Smith J, Kronforst MR (2013) Do Heliconius butterfly species exchange mimicry alleles? *Biol Lett* 9(4):20130503.
26. Weir BS (1996) *Genetic Data Analysis II: Methods for Discrete Population Data* (Sinaur Assoc., Sunderland, MA).
27. Nielsen R, et al. (2009) Darwinian and demographic forces affecting human protein coding genes. *Genome Res* 19(5):838–849.
28. Pfister A, et al. (2014) A receptor-like kinase mutant with absent endodermal diffusion barrier displays selective nutrient homeostasis defects. *eLife* 3:e03115.
29. Hosmani PS, et al. (2013) Dirigent domain-containing protein is part of the machinery required for formation of the lignin-based Casparian strip in the root. *Proc Natl Acad Sci USA* 110(35):14498–14503.
30. Kamiya T, et al. (2015) The MYB36 transcription factor orchestrates Casparian strip formation. *Proc Natl Acad Sci USA* 112(33):10533–10538.
31. Choi W-G, Toyota M, Kim S-H, Hilleary R, Gilroy S (2014) Salt stress-induced Ca2+ waves are associated with rapid, long-distance root-to-shoot signaling in plants. *Proc Natl Acad Sci USA* 111(17):6497–6502.
32. Peiter E, et al. (2005) The vacuolar Ca2+-activated channel TPC1 regulates germination and stomatal movement. *Nature* 434(7031):404–408.
33. Kim B-H, von Arnim AG (2009) FIERY1 regulates light-mediated repression of cell elongation and flowering time via its 3′ (2′),5′-bisphosphate nucleotidase activity. *Plant J* 58(2):208–219.
34. Cai X, et al. (2006) Mutant identification and characterization of the laccase gene family in *Arabidopsis*. *J Exp Bot* 57(11):2563–2569.
35. Martin SH, Davey JW, Jiggins CD (2015) Evaluating the use of ABBA-BABA statistics to locate introgressed loci. *Mol Biol Evol* 32(1):244–257.
36. Mallet J, Besansky N, Hahn MW (2016) How reticulated are species? *BioEssays* 38(2):140–149.
37. Yatabe Y, Kane NC, Scotti-Saintagne C, Rieseberg LH (2007) Rampant gene exchange across a strong reproductive barrier between the annual sunflowers, *Helianthus annuus* and *H. petiolaris*. *Genetics* 175(4):1883–1893.
38. Beyhl D, et al. (2009) The fou2 mutation in the major vacuolar cation channel TPC1 confers tolerance to inhibitory luminal calcium. *Plant J* 58(5):715–723.
39. Schaaf G, et al. (2006) AtIREG2 encodes a tonoplast transport protein involved in iron-dependent nickel detoxification in *Arabidopsis thaliana* roots. *J Biol Chem* 281(35):25532–25540.
40. Morrissey J, et al. (2009) The ferroportin metal efflux proteins function in iron and cobalt homeostasis in *Arabidopsis*. *Plant Cell* 21(10):3326–3338.
41. Tiffin P, Ross-Ibarra J (2014) Advances and limits of using population genetics to understand local adaptation. *Trends Ecol Evol* 29(12):673–680.
42. Conte GL, Arnegard ME, Peichel CL, Schluter D (2012) The probability of genetic parallelism and convergence in natural populations. *Proc Biol Sci* 279(1749):5039–5047.
43. Lunter G, Goodson M (2011) Stampy: A statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Res* 21(6):936–939.
44. Li H, et al.; 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078–2079.
45. Rawat V, et al. (2015) Improving the annotation of *Arabidopsis lyrata* using RNA-Seq data. *PLoS One* 10(9):e0137391.

Supplementary Information for:

Borrowed alleles and convergence in serpentine adaptation

Brian J. Arnold[a, b], Brett Lahner[c], Jeffrey M. DaCosta[a], Caroline M. Weisman[a], Jesse D. Hollister[d], David E. Salt[c,e], Kirsten Bomblies[a,f], Levi Yant[a,f,1]

a. Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts, 02138, United States of America

b. Center for Communicable Disease Dynamics, Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts 02115, United States of America

c. Department of Horticulture and Landscape Architecture, Purdue University, West Lafayette, Indiana, 47907, United States of America

d. Department of Ecology and Evolution, Stony Brook University, Stony Brook, NY, 11794-5245, United States of America

e. Institute of Biological and Environmental Science, University of Aberdeen, Aberdeen, Scotland AB24 3UU, United Kingdom

f. Department of Cell and Developmental Biology, John Innes Centre, Norwich Research Park, Norwich, NR4 7UH, United Kingdom

1. Author for correspondence: levi.yant@jic.ac.uk

Author Contributions:

LY conceived and headed the project.

BJA, JDH, LY and KB collected plant and soil materials.

KB performed the flowering time analysis.

LY, JDC, and CW performed DNA extractions and constructed Illumina libraries.

JDC and LY performed read mapping data processing and variant calling.

BJA performed the coalescent analysis and ABBA-BABA and D_{xy} tests

BJA performed the demographic analyses on WGS and RAD data with input from LY.

BJA performed the genomic ABBA-BABA tests.

LY performed the analysis of selective sweep (F_{ST}, 2dSFS CLR, and DDresidual values and detected outlier regions).

JDH aided in scripting for SFS metric.

BL performed elemental analysis with input from DES.

LY, KB, and DES drafted the manuscript with input from all coauthors.

Supplemental Information:

Section S1. Phylogenetic analysis of Austrian *A. arenosa* population samples and calculation of Akaike weights.

We reconstructed phylogenies of four Austrian *A. arenosa* populations to confirm the PCA in Figure 1B that Hochlantsch is the closest relative to Gulsen. Using the RADseq dataset in (1), we constructed simple models of all possible relationships between four populations: Gulsen, Hochlantsch, Kasparstein, and Kößlbach, a geographically distant population sample from Northwestern Austria. We fit each model to SNP data using *fastsimcoal2* (2) and compared the model likelihoods using Akaike Information Criterion (AIC) to measure the relative fit of each demographic model. Following (3), we calculated the AIC value of model *i* with *d* parameters using

$$AIC_i = 2d - 2\ln(Likelihood_i)$$

and then used these values to obtain Akaike weights (*w*) for each model using

$$\Delta_i = AIC_i - \min(AIC)$$

$$w_i = \frac{e^{-0.5\Delta_i}}{\sum_r^R e^{-0.5\Delta_i}}$$

Akaike weights quantify the relative likelihood of each model given the candidate set of models and may be interpreted as the probability that model *i* is the best model among the set of candidates. For a simple model of four populations with no migration, there are *d* = 10 parameters: 7 population sizes (3 ancestral) and 3 divergence times. Table S2 shows the results for each phylogenetic model, and analysis of the Akaike weights

show the Newick phylogenetic tree (KO,(KA,(HO,GU))) has the highest relative likelihood by a large margin.

Section S2. Whole genome resequencing data generation and processing.

DNA extraction, sequencing and read trimming

Seeds collected in the wild were grown in common growth chamber conditions and genomic DNA was extracted from leaf material as in (4). Whole-genome Illumina libraries were constructed for each individual plant using Illumina's TruSeq PCR-free HT kit and, quantified by Qubit (ThermoFisher, Inc.) and characterized for fragment distributions by Tapestation (Agilent, Inc.). All samples were sequenced at Harvard University's Bauer sequencing core on an Illumina HiSeq 2500 with v4 chemistry. Adapter sequences in raw reads were identified and trimmed using cutadapt v1.8 (https://cutadapt.readthedocs.org) (5). All raw short read data for this study have been deposited in the NCBI Sequence Read Archive under BioProject number XXX-XXX.

Processing, alignment and variant calling

Trimmed reads were then mapped to the repeatmasked (hardmasked) Lyrata107 genome (6) using stampy v1.0.21 for single-end data (`stampy.py -g <reference> -h <reference_basename> --substitutionrate=0.001 -t10 --bamkeepgoodreads -M <input.fastq> > <output.sam>`), and a combination of bwa v0.7.4 and stampy for paired end data using default recommendations for BWA employment for PE reads in stampy (7). We then used Picard Tools v1.98

(http://broadinstitute.github.io/picard) to mark duplicate reads (MarkDuplicates.jar) and edit read group naming (AddOrReplaceReadGroups.jar). We next applied GATK v2.7-2 (8) for indel realignment (RealignerTargetCreator and IndelRealigner) and performed genotyping (UnifiedGenotyper) across all samples simultaneously (all scripts available online at https://github.com/jeffdacosta/MaleaeGenomics/tree/master/Genome-Scan). UnifiedGenotyper was used with the -ploidy 4 option: `GenomeAnalysisTK-2.7-2 -T UnifiedGenotyper -nt 8 -nct 3 --min_base_quality_score 25 -rf MappingQuality --min_mapping_quality_score 25 -rf DuplicateRead -rf BadMate -rf BadCigar -R <reference> -L <scaffold> -ploidy 4 -glm SNP -o <outfile.vcf> -stand_emit_conf 13.0 -stand_call_conf 25.0 --output_mode EMIT_ALL_SITES -dcov 200 -I <input_realigned.bam>`).

Section S3. Genomic data filtering.

Discovered variants were filtered to remove ambiguous or low quality genotypes. We masked variable loci that failed any of the following filters:

1) >2 alleles (retaining biallelic sites only)

2) <4X coverage in any sample

3) Phred-scaled probability score of less than 25.0

4) all GATK “best practices” filtering expressions (`--filterExpression "QD < 2.0 || FS > 60.0 || MQ < 40.0 || HaplotypeScore > 13.0 || MappingQualityRankSumTest < -12.5 || ReadPosRankSum < -8.0"`),

5) Following (1), we also excluded loci exhibiting the highest levels of heterozygosity, which may arise from potentially paralogous loci in any given individual mapping to single locus in the reference. Briefly, using 8 *A. arenosa* diploid whole-genome sequences, we previously identified regions (no longer than 2kb) of high heterozygosity in *A. arenosa* in which all 8 diploids were heterozygous at 3 or more sites (1). Here we excluded loci in these regions from downstream analyses, reasoning that it is very unlikely for all 8 diploids from two separate populations to be heterozygous at 3 or more sites, according to Hardy-Weinberg equilibrium.

Section S4. Construction of demographic model and coalescent simulations.

We constructed a demographic model for three populations in which the Gulsen (GU) and Hochlantsch (HO) populations fuse, backwards-in-time, before the ancestor of these two populations fuses with *A. lyrata*. We allowed five different backwards-in-time migration parameters: HO to GU, GU to HO, GU to *A. lyrata*, HO to *A. lyrata*, and the ancestor of HO and GU to *A lyrata*. In order to ascertain which migration parameters were statistically supported by the data and thus potentially biologically meaningful, we permuted all possible combinations of these five migration rates to create 32 demographic models. We fit each model to four-fold degenerate SNP data via

coalescent simulations using *fastsimcoal2* (2) and compared the resulting model likelihoods with Akaike weights as in Section S1 above (Table S4).

To find model likelihoods and maximum likelihood estimates of parameters, *fastsimcoal2* starts with random initial parameter values taken from user-specified distributions (uniform or Log-uniform). Then, *fastsimcoal2* uses a conditional maximization (ECM) algorithm that maximizes each model parameter in turn. The lower range limit of these user-specified distributions is an absolute minimum for the parameter value, but no upper limit exists; the upper bound serves only as a limit for choosing the initial parameter value, but subsequent draws of parameter values may surpass this initial bound as the ECM algorithm proceeds. For each model we used 50 independent optimizations, each initialized with different starting values for parameters to avoid local maxima in the likelihood surface. To obtain confidence intervals (CIs) for parameter estimates, we sampled with replacement from the four-fold degenerate SNP matrix to create 100 replicate datasets and performed inference as above with 50 independent optimizations for each replicate. Lastly, we used the same mutation rate as (1) to calibrate coalescent simulations and obtain absolute values of population sizes and divergence times.

An example of the input files we constructed to specify both the model and the distributions for initial parameter values:

Example "*tpl*" *fastsimcoal2* input file used to specify demographic model in Figure 1C

```
//Parameters for the coalescence simulation program : simcoal.exe
3 samples to simulate :
//Population effective sizes (number of genes)
HOpopsize
GUpopsize
LYRpopsize
//Samples sizes and samples age
24
24
1
//Growth rates: negative growth implies population expansion
0
0
0
//Number of migration matrices : 0 implies no migration between demes
3
//Migration matrix 0
0       MigHG MigHL
MigGH       0 MigGL
0       0       0
//Migration matrix 1
0       0       MigAL
0       0       0
0       0       0
//Migration matrix 2
0       0       0
0       0       0
0       0       0
//historical event: time, source, sink, migrants, new deme size, new growth rate, migration matrix index
2 historical event
TDIV1 1 0 1 ResizeTIME1 0 1
TDIV2 2 0 1 ResizeTIME2 0 2
//Number of independent loci [chromosome]
1 0
//Per chromosome: Number of contiguous linkage Block: a block is a set of contiguous loci
1
//per Block:data type, number of loci, per generation recombination and mutation rates and optional parameters

FREQ 1 0 4.3e-8
```

Example "*est*" *fastsimcoal2* input file used to specify demographic model in Figure 1C

```
// Priors and rules file
// *********************
[PARAMETERS]
//#isInt? #name   #dist.#min  #max
//all N are in number of haploid individuals
1       HOpopsize  unif     100  1000000   output
1       GUpopsize  unif     100  1000000   output
1       LYRpopsize unif     100  1000000   output
1       ANCpopsize1        unif     100000  10000000   output
1       ANCpopsize2        unif     100000  10000000   output
1       TDIV1         unif    1       10000output
1       TIMEextra     unif    1       1000000      hide
0       MigHG         logunif         1e-7   1e-4   output
0       MigHL         logunif         1e-7   1e-4   output
0       MigGH         logunif         1e-7   1e-4   output
0       MigGL         logunif         1e-7   1e-4   output
0       MigAL         logunif         1e-7   1e-4   output

[RULES]

[COMPLEX PARAMETERS]
1       TDIV2 = TDIV1+TIMEextra ouput
0       ResizeTIME1       =       ANCpopsize1/HOpopsize        hide
0       ResizeTIME2       =       ANCpopsize2/LYRpopsize       hide
0       NmHG = MigHG*HOpopsize    output
0       NmHL = MigHL*HOpopsize    output
0       NmGH = MigGH*GUpopsize    output
0       NmGL = MigGL*GUpopsize    output
0       NmAL = MigAL*ANCpopsize1  output
```

With the above *tpl* and *est* files, we performed each independent optimization for the demographic model in Figure 1C using the following command-line options:

```
./fsc25 -t TemplateFile.tpl -n100000 -N100000 -d -e EstimationFile.est -M
0.001 –l 10 -L 40 -k 50000 --multiSFS
```

Section S5. Selective sweep analysis, gene coding locus overlap, comparative analysis, enrichment tests, and quantitative tests for introgression (four taxon ABBA-BABA tests).

In addition to calculating group-wise allele frequency differences at the 4.9 million SNPs discovered, we also used three window-based tests to scan for the most robust signatures of selective sweep. For all window-based metrics, we conscribed the genome into 193,937 twenty-five SNP windows (excluding 56 windows that were larger than 50kb, for a final number of 193,881 windows). Because detection of selective sweep in the context of high migration is challenging (e.g. (9)), we employed several methods to robustly detect sweeps:

1) D_{xy} metric: Following (10), we calculated absolute levels of divergence for windows of *n* SNPs as:

$$d_{xy} = \frac{1}{n}\sum_{i=1}^{n} p_{ix}(1-p_{iy}) + p_{iy}(1-p_{ix})$$

Here p_x and p_y are the derived allele frequencies in population *x* and *y*, respectively.

2) DD residual metric: To test for deviations from expected patterns of Gulsen diversity and divergence ("DD residual" test) between Gulsen and non-serpentine groups, we calculated nucleotide diversity in Gulsen and average allele frequency difference between Gulsen and the non-serpentine samples for each window. We then calculated the linear regression fit to these data. Residuals were calculated based on

vertical deviations from the regression line fit and were used to identify outliers with the most negative regression values, which in this analysis identified regions with excess divergence relative to diversity, compared to the genome-wide relationship. We previously devised the DD metric to identify genomic windows with excess differentiation for a given level of diversity in this same autotetraploid *A. arenosa* system (4). This DD metric is based on the logic of the HKA test; by capitalizing on the positive relationship between diversity and differentiation across the genome, the DD metric scans for outliers from this relationship by plotting diversity against differentiation for all genomic windows, obtaining the genome-wide best fit and calculating the residual values from this fit. **Negative outlier values identify regions of the genome that exhibit high differentiation for their level of diversity in the serpentine population, a selective sweep signature not unlike low pi/high F_{ST}, but collapsing both into a single, locus-normalized measure of lowered diversity** (yellow shading in Fig 3E).

3) **F_{ST} metric:** We calculated F_{ST} between groups following (11).

4) **2dSFS CLR metric:** We implemented a composite likelihood ratio test of the 2dSFS CLR test, following the framework of (12). Briefly, we calculated the likelihoods of observing *k* SNPs at frequency *i* in the non-serpentine sample and frequency *j* in the Gulsen sample, over all *i* and *j*, compared the ratio of composite likelihoods for window-specific versus genome-wide models as in (4). We note that confident haplotype determination from short read data in autopolyploids is currently not reliable, so we did not use haplotype-based tests.

Gene ID assignment: Windows in outlier tails of selection metric lists were intersected and candidate selected genes were identified by proximity (< 2kb) between *A. lyrata* version-2 (13) reference annotations and outlier windows. Suspected copy number variants were filtered as in (1). Genomic location relationships were determined using bedtools 2.25.0 (14).

Overlap of metrics: Outliers for absolute divergence, D_{xy}, overlapped highly with outliers of F_{ST}, DDresidual, and the 2dSFS CLR outliers, with 46 of the 51 D_{xy} 0.025% outliers already present on the other 0.1% outlier lists (Figure 3H). There was considerable overlap in our highest stringency 0.1% outlier lists for differentiation metrics F_{ST} and DD (Figure 3H), as 61/197 DD genes within 2kb or containing outlier windows were also overlapping or proximal to F_{ST} outlier windows. In addition, there was much overlap of the 2dSFS and F_{ST} outliers, and 33 gene coding loci were represented on all three 0.1% outliers lists (Supplementary Dataset S3-S6; Figure 3H). The window-based scan yielded a total of 136 gene-coding loci as selective sweep candidates. In addition to these window-based metric candidates, we retained the most differentiated single SNPs with allele frequency differences between Gulsen and nonserpentine populations >0.8 (following (15)); this gave 77 SNPs genome-wide, or the top 0.0016%. We identified 47 coding loci within 2kb of these SNPs (Supplementary Dataset S7), 21 of which were already among the 136 window-based loci. The addition of the remaining 26 outlier-SNP-proximal coding loci brought our list of Top Sweep Candidates to a final total of 162 coding loci (*SI Appendix*, Tables S6 and S7). These

most differentiated regions are scattered across the genome in small, dispersed footprints of selection, consistent with an emerging picture of multigenic selection and adaptation. These sweep regions consistently displayed sharp peaks of differentiation with immediate decay to background (e.g. Fig. 4), leaving little ambiguity in candidate identification. This, along with the low correlation of diversity between adjacent windows (*SI Appendix*, Figure S3), indicates that negligible linkage disequilibrium observed in other *A. arenosa* populations (1, 4, 16) also applies to the Gulsen population.

Gene function inference and locus visualization: Gene functions/identities were inferred by nearest homology with *A. thaliana* loci. Allele frequency difference outlier data from (15) were reprocessed from raw between-population SNP frequencies using the updated *A. lyrata* version-2 annotation (13) to more closely parallel the data generated in this study. We used HybridCheck for similarity investigation of candidate borrowed alleles and visualization of loci in Figures 4B and 4C (17).

Additional metrics: To scan for regions with localized low diversity specifically in the Gulsen population vs. other *A. arenosa* populations, we devised the Div_{diff} metric. This is calculated for each window simply as (θ_W diversity in Gulsen - θ_W diversity in nonserpentine *A. arenosa*). Therefore, more negative values indicate specifically lower diversity in Gulsen at that genomic window, a classic sweep signature. The $ZengE_{diff}$ metric is calculated in the same way using Zeng's E for each window and indicates a relatively negative Zeng's E value in Gulsen over a particular window, relative to other *A. arenosa* populations.

Enrichment tests: To conservatively ensure independence of observations in hypergeometric tests of enrichment, we collapsed nearby SNPs (or gene loci) into single observations if they were within the distance of linkage decay in *A. arenosa*, based the observed lack of correlation of diversity estimates at fourfold degenerate sites over this scale in the present data (*SI Appendix*, Figure S3) and *(1, 4, 16)*. Having performed this, no enriched overlap SNPs or gene coding loci were < 10MB apart in the contrasts of SNP proximity between the current study and (15) or the comparison or the four taxon test outliers and Top Sweep Candidates, ensuring no linkage. In this way the loci sampled in these tests are effectively independent observations.

Section S6. Quantitative tests for introgression. To test for introgression between *A. arenosa* and *A. lyrata* in localized windows of *n* SNPs, we calculated an ABBA-BABA statistic using two *A. arenosa* populations (P_1 and P_2), an *A. lyrata* sequence (*L*), and an *A. thaliana* sequence (*T*) as an outgroup:

$$D(P_1, P_2, L, T) = \frac{\sum_{i=1}^{n} X_{ABBA}(i) - X_{BABA}(i)}{\sum_{i=1}^{n} Y_{ABBA}(i)}$$

where

$$X_{ABBA}(i) = (1 - p_{i1})p_{i2}$$

$$X_{BABA}(i) = p_{i1}(1 - p_{i2})$$

This effectively weights each SNP by its fit to the ABBA pattern, in which P_2 has the same allele as *A. lyrata* with frequency p_{i2}, or the BABA pattern, in which P_2 has the opposite allele as *A. lyrata* with frequency 1-p_{i2}. Y_{ABBA} is calculated similarly but assuming complete introgression from *A. lyrata* to P_2, homogenizing allele frequencies (which is 1 in *A. lyrata* since only a single sequence was available). This method is similar to f_d in (18).

To assess the genome-wide significance of the ABBA-BABA statistic, we used both bootstrap and uneven m-delete jackknife (19) resampling techniques with a block size of 100,000 bp. Using Hochlantsch as P_1, we tested for admixture between *A. lyrata* (P_3) and Gulsen or Kasparstein (P_2). There were slightly more ABBA than BABA patterns with Gulsen as P_2 as the mean genome-wide f_d = 0.004. However, this mean was not significantly greater than zero since bootstrapped 95% CIs contained zero (-0.002 – 0.009) and jackknife estimates of the standard error suggest zero was not significantly different (Z-score = -1.41; p-value > 0.10). This trend was specific to Gusen and was abolished when Kasparstein is used as the recipient P_2 (mean f_d = -0.0005; bootstrapped 95% CIs = -0.007 – 0.005; jackknife Z-score = 0.16)

Supplemental Figures:

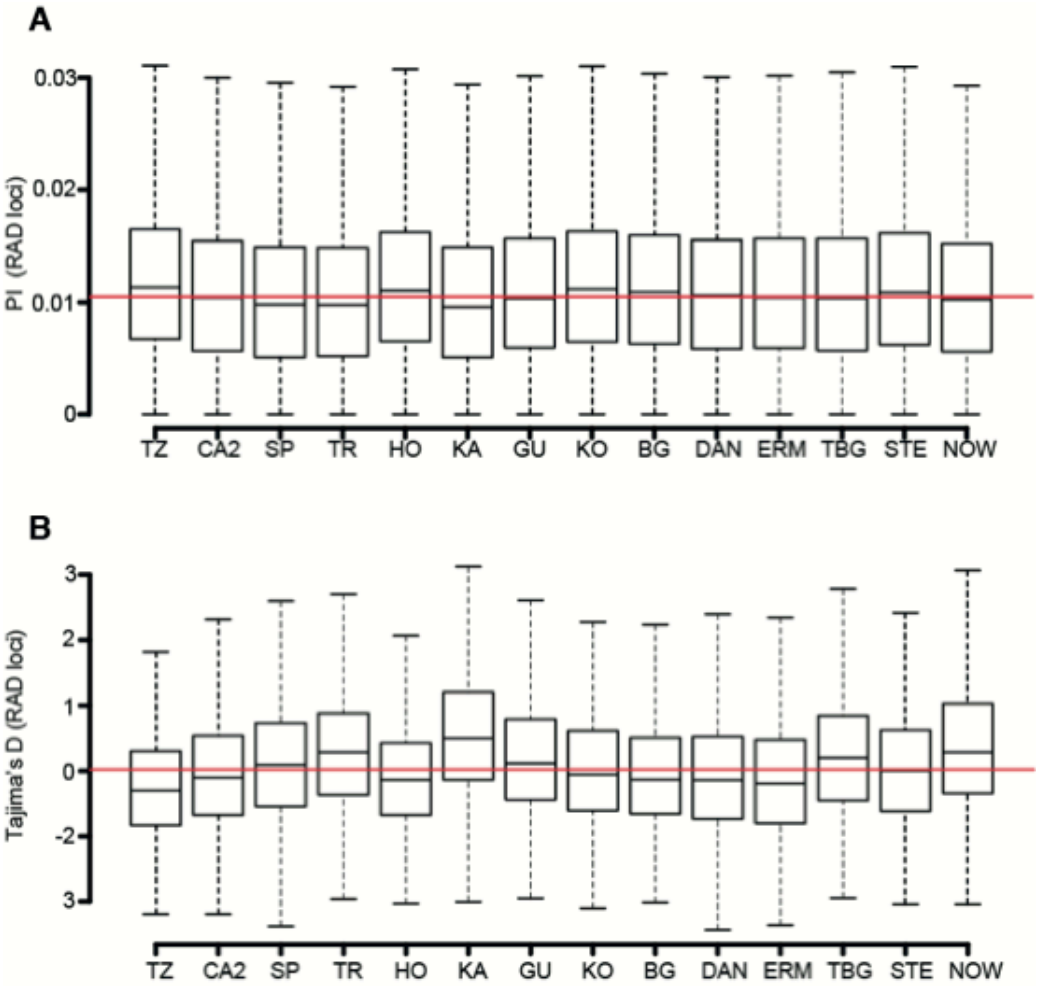

Figure S1. Summary statistics for *Arabidopsis arenosa* range-wide autotetraploid populations at RAD loci indicate lack of extreme bottleneck in Gulsen or bottleneck with substantial gene flow. (A) Diversity (π) and **(B)** Tajima's D. RAD data were reprocessed from (1) for range-wide population survey here shown.

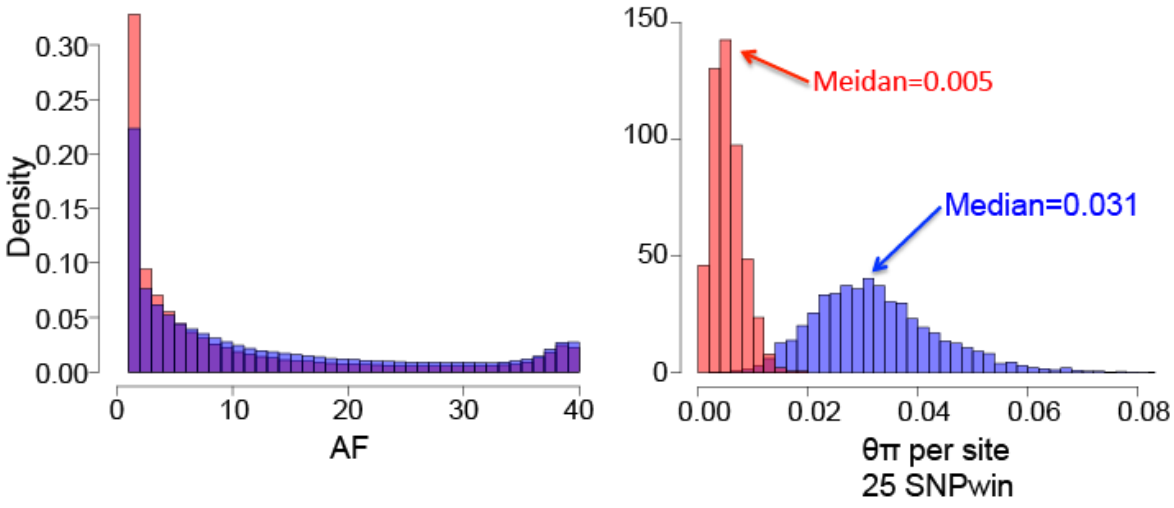

Figure S2. Allele frequency spectra and diversity in Gulsen population. Blue histograms represent fourfold degenerate sites and red represent all sites. Data here shown are from the genome resequencing of 10 Gulsen autotetraploid individuals (40 alleles per site).

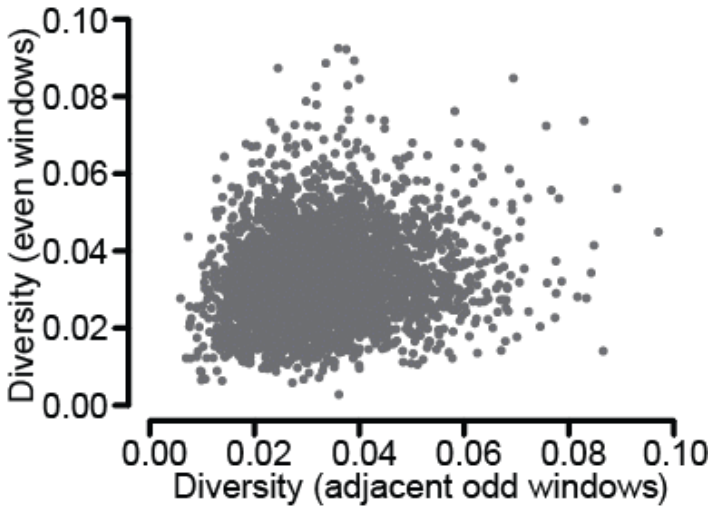

Figure S3. Minimal correlation (Pearson Correlation = 0.147) of diversities between adjacent 25 SNP windows genome-wide, indicates rapid linkage decay in *A. arenosa*. Diversities at fourfold degenerate sites of adjacent windows are plotted.

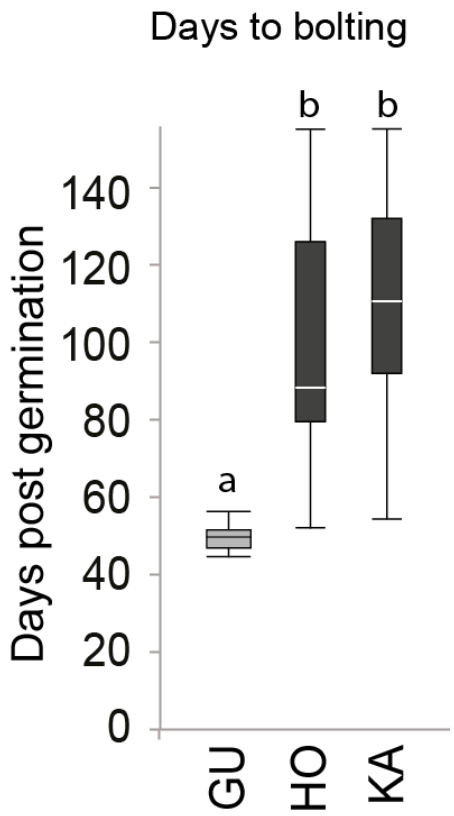

Figure S4. The serpentine population, GU, flowers much earlier than nearest relatives, nonserpentine HO and KA. Box plots showing flowering time (as days from germination to first visible buds). Letters indicate significantly different distributions: (a) GU flowers significantly earlier than either KA (Two-tailed T-test, p-value = 9.7 x 10-7) or HO (Two-tailed T-test, p-value = 3.4 x 10-9). (b) KA and HO are not significantly different (Two-tailed T-test, p-value = 0.59). GU=Gulsen population; HO=Hochlantsch population; KA= Kasparstein population.

Table S1: *Arabidopsis arenosa* population locations sampled for elemental accumulation and population genomic resequencing.

ID	N	E	Location	Country
BGS	47 37 41	13 00 06	Berchtesgaden	Germany
BK	48 04 49	9 05 16	Donau/Neumühle	Germany
BU	48 00 58	8 57 41	Donau/Durchbruch-Beuron	Germany
CA1	48 52 39	20 31 39	Hnilčík	Slovakia
CA2	48 51 06	21 05 24	Sivec	Slovakia
DFS	62 55 20	15 57 30	Dead Falls, Lake Ragunda	Sweden
EZ	48 28 54	9 22 23	Eppenzil, Bad Urach	Germany
FT	48 4 33	8 5 27	Finstertal, Donau	Germany
GS	48 26 52	9 25 20	Grindelsteige, Bad Urach	Germany
GU	47 17 24	14 55 54	Gulsen Mountain	Austria
HA	48 51 06	15 51 30	Dyje River (Hardegg)	Austria
HB	48 05 01	9 09 22	Schmeie/Donau (Höhnberg)	Germany
HF	48 29 19	9 23 21	Bad Urach	Germany
HO	47 22 12	15 23 12	Hochlantsch	Austria
KA	46 41 18	14 52 18	Kasparstein	Austria
KO	47 44 49	13 41 23	Kößlbach	Austria
MT	48 26 33	9 27 54	Mühltal/Seeburg	Germany
NM	48 05 05	9 03 57	Neumühle/Donau	Germany
OF	48 30 45	9 19 38	Rossfeld/Glems	Germany
RB	48 1 19	8 57 54	Roggenbusch/Donau	Germany
RF	48 06 04	9 03 02	Reiftal/Donau	Germany
SN	49 10 27	18 51 42	Strečno	Slovakia
SP	48 59 20	20 46 30	Dreveník	Slovakia
ST	52 16 49	16 42 34	Stęszew	Poland
TBG	48 00 33	8 56 50	Triberg	Germany
TR	48 53 39	18 2 41	Trenčín	Slovakia
US	48 28 31	9 23 44	Upfinger Steig, Bad Urach	Germany
WF	48 30 13	9 19 23	Wiesfels, Rossfeld/Glems	Germany
WT	48 00 33	8 56 50	Wolfental/Donau	Germany

Table S2. Likelihood analysis of all possible phylogenies of Austrian *Arabidopsis arenosa* population samples

Population Phylogeny	***Max*** $\log_{10}(Lhood_i)$	AIC_i	Δ_i	w_i
((HO,KA),(GU,KO))*	-281314.979	1295523.354	3032.237	~0
((HO,KA),(GU,KO))**	-281307.008	1295486.646	2995.530	~0
((HO,GU),(KA,KO))**	-281188.224	1294939.626	2448.509	~0
((HO,KO),(KA,GU))*	-281163.259	1294824.658	2333.541	~0
((HO,KO),(KA,GU)) **	-281028.528	1294204.199	1713.082	~0
((HO,GU),(KA,KO))*	-281018.913	1294159.92	1668.803	~0
(HO,(KA,(GU,KO)))	-280972.098	1293944.329	1453.212	~0
(GU,(KA,(HO,KO)))	-280875.083	1293497.558	1006.442	$2.84*10^{-219}$
(KA,(HO,(GU,KO)))	-280851.153	1293387.356	896.240	$2.42*10^{-195}$
(KA,(GU,(HO,KO)))	-280830.3	1293291.325	800.208	$1.72*10^{-174}$
(GU,(KO,(HO,KA)))	-280827.266	1293277.353	786.236	$1.86*10^{-171}$
(GU,(HO,(KA,KO)))	-280795.04	1293128.947	637.830	$3.14*10^{-139}$
(HO,(KO,(KA,GU)))	-280786.023	1293087.422	596.305	$3.26*10^{-130}$
(HO,(GU,(KA,KO)))	-280749.142	1292917.578	426.462	$2.48*10^{-93}$
(KA,(KO,(HO,GU)))	-280736.548	1292859.581	368.464	$9.74*10^{-81}$
(KO,(GU,(HO,KA)))	-280714.23	1292756.803	265.686	$2.02*10^{-58}$
(KO,(HO,(KA,GU)))	-280667.19	1292540.176	49.059	$2.22*10^{-11}$
(KO,(KA,(HO,GU)))	-280656.537	1292491.117	0	1

Note: Population phylogenies are in Newick format. Likelihoods (Lhood) were estimated for each phylogenetic model using *fastsimcoal2*. See Section S1 of *SI appendix* for calculation of AIC values and Akaike weights (*w*). HO=Hochlantsch population; GU=Gulsen population; KA=Kasparstein population; KO=Kößlbach outgroup population.
*left clade fuses first
**right clade fuses first

Table S3. Aligned genome resequencing coverage per individual sample from populations GU, HO, and KA.

Individual	Fold Coverage
GU01	22
GU02	26
GU03	12
GU04	15
GU05	37
GU06	24
GU07	24
GU08	20
GU09	18
GU10	23
HO04	24
HO10	18
HO15	17
HO17	16
HO20	19
HO21	16
KA02	36
KA06	19
KA16	17
KA18	18
KA19	22
KA27	12
KA29	17
KA30	25
Mean	21

Note: mean successfully aligned read coverage per individual in all gene coding regions is given.

Table S4. Model selection results for all permutations of five migration rates.

Backwards-in-time Migration Parameter					Analysis of Relative Likelihoods				
Ancestral *A .arenosa* to *A. lyrata*	GU to *A. lyrata*	HO to *A. lyrata*	HO to GU	GU to HO	*Max* $\log_{10}(Lhood_i)$	*Number of parameters*	AIC_i	Δ_i	w_i
-	-	-	✔	✔	-994044.201	9	4577760.718	27788.9389	~0
-	-	-	✔	-	-993920.703	8	4577189.989	27218.20959	~0
-	-	-	-	-	-993828.003	7	4576761.089	26789.31032	~0
-	-	✔	✔	-	-993654.062	9	4575964.062	25992.28241	~0
-	-	-	-	✔	-993591.496	8	4575673.934	25702.15533	~0
-	-	✔	✔	✔	-993573.413	10	4575594.659	25622.88004	~0
-	-	✔	-	✔	-993548.72	9	4575478.944	25507.16457	~0
-	-	✔	-	-	-993366.787	8	4574639.111	24667.33214	~0
✔	-	✔	✔	✔	-990622.018	11	4562004.983	12033.20378	~0
✔	-	✔	-	-	-990600.93	9	4561903.869	11932.08995	~0
✔	-	-	-	✔	-990591.638	9	4561861.078	11889.29871	~0
✔	-	-	✔	-	-990590.394	9	4561855.349	11883.56988	~0
✔	-	-	-	-	-990550.27	8	4561668.571	11696.79203	~0
✔	-	-	✔	✔	-990508.595	10	4561480.651	11508.87156	~0
✔	-	✔		✔	-990432.303	10	4561129.313	11157.53392	~0
✔	-	✔	✔	-	-989951.002	10	4558912.84	8941.0609	~0
-	✔	-	-	✔	-989877.823	9	4558573.838	8602.059151	~0
-	✔	-	-	-	-989559.649	8	4557106.593	7134.813732	~0
-	✔	✔	✔	-	-989187.687	10	4555397.645	5425.865419	~0
-	✔	-	✔	-	-989093.484	9	4554961.824	4990.044572	~0
-	✔	-	✔	✔	-988975.598	10	4554420.939	4449.159479	~0
-	✔	✔	-	✔	-988933.039	10	4554224.947	4253.168042	~0
-	✔	✔	✔	✔	-988739.76	11	4553336.864	3365.085353	~0
✔	✔	-	✔	-	-988733.579	10	4553306.4	3334.620796	~0
-	✔	✔	-	-	-988611.865	9	4552743.886	2772.107112	~0
✔	✔	-	-	✔	-988606.905	10	4552723.045	2751.265468	~0
✔	✔	✔	✔	-	-988432.574	11	4551922.221	1950.441544	~0
✔	✔	✔	-	-	-988352.181	10	4551549.997	1578.218098	~0
✔	✔	✔	✔	✔	-988330.867	12	4551455.843	1484.0635	~0
✔	✔	-	✔	✔	-988287.02	11	4551251.92	1280.140603	$1.05*10^{-278}$
✔	✔	-	-	-	-988222.163	9	4550949.242	977.4630804	$5.58*10^{-213}$
✔	✔	✔	-	✔	-988009.041	11	4549971.779	0	1

Note: HO=Hochlantsch population; GU=Gulsen population; “Ancestral *A. arenosa*”=ancestor of HO and GU. AIC analyses and Akaike weights calculated as in Section S1 of *SI Appendix*. Each model was fit to four-fold degenerate SNPs using 50 runs of *fastsimcoal2*.

Table S5. Maximum Likelihood Estimates (MLEs) for parameters of the migration model with the highest Akaike weight along with 90% and 95% confidence intervals (CIs).

Parameter	MLEs	Lower / Upper 90% CI	Lower / Upper 95% CI
HOpop	64858	31402 / 109954	27993 / 111825
GUpop	36928	14245 / 55275	12027 / 57530
LYRpop	790849	447355 / 925985	417807 / 987603
ANCpop1	314143	285367 / 382699	279573 / 388059
ANCpop2	843556	541894 / 1141009	482374 / 1213314
DivTime1	3195	1398 / 4555	1215 / 4766
DivTime2	396521	272227 / 426678	264813 / 451432
MigHL	$5.86*10^{-6}$	$2.96*10^{-7}$ / $9.23*10^{-6}$	$2.193*10^{-7}$ / $1.16*10^{-5}$
MigGH	$3.85*10^{-6}$	$4.08*10^{-7}$ / $3.95*10^{-5}$	$2.352*10^{-7}$ / $5.01*10^{-5}$
MigGL	$1.69*10^{-5}$	$9.99*10^{-6}$ / $4.95*10^{-5}$	$8.094*10^{-6}$ / $5.79*10^{-5}$
MigAL	$7.20*10^{-7}$	$3.06*10^{-7}$ / $1.09*10^{-6}$	$2.394*10^{-7}$ / $1.24*10^{-6}$
N*MigHL	0.380	0.028 / 0.466	0.0151 / 0.502
N*MigGH	0.142	0.010 / 1.573	0.006 / 1.707
N*MigGL	0.624	0.434 / 0.853	0.388 / 0.899
N*MigAL	0.226	0.116 / 0.352	0.088 / 0.383

Note: Shown are **population sizes** of Hochlantsch (HOpop), Gulsen (GUpop), *A. lyrata* (LYRpop), the ancestral *A. arenosa* population of HO and GU (ANCpop1), and the ancestral population of *A. arenosa* and *A. lyrata* (ANCpop2). These populations sizes are $4N_e$ or $2N_e$ haploid number of for tetraploids and diploids, respectively. Also shown are **divergence times** between HO and GU (DivTime1) and between the ancestral *A. arenosa* population and *A. lyrata* (DivTime2). Last are the backwards **migration probabilities** from HO to GU (MigHG), from GU to HO (MigGH), from GU to *A. lyrata* (MigGL), and from the ancestral *A. arenosa* population and *A. lyrata* (MigAL). Migration is also presented as these migration probabilities multiplied by population size (N), which may be interpreted as the average number of individuals in the sink population that migrated from the source population in the previous generation. Confidence intervals were constructed by bootstrapping the four-fold degenerate SNP matrix 100 times and performing inference on each replicate with 50 runs of *fastsimcoal2*.

Table S6: Top Sweep Candidate list selection summary.

Candidate List	Loci	Gene IDs within 2kb	Inclusion Criteria
0.025% D_{xy} windows	48	51	0.025% outliers of empirical distribution of 25SNP windows All are included in final Top Sweep Candidate list
0.1%Fst windows	194	209	0.1% outliers of empirical distribution of 25SNP windows
0.1%DD windows	194	197	0.1% ouliters of empirical distribution of 25SNP windows
0.1%2dSFS windows	194	239	0.1% outliers of empirical distribution of 25SNP windows
0.1%_Triple_positive	-	33	overlap of Fst, DD, 2DSFS distribution 0.1% outliers gene IDs
Single_Outlier_SNPs	77	47	Gu vs HoKa AFD > 0.80 (77/4,889,615 SNPs = top 0.0016% outliers)
Top Sweep Candidates	-	162	0.1% Fst overlap and either 0.1%DD or 0.1%2dSFS overlap unions; 0.025% D_{xy} windows; Single_Outlier_SNP list overlap loci

Table S7. Gene IDs in windows of highest absolute net divergence D_{xy}. Extreme (top 0.025%) outliers are given.

A.lyrata ID	*A.thaliana* ID	A. *thaliana* name and description	Serpentine relevance
AL1G15400	AT1G05550		
AL1G15410	(none)		
AL1G15420	(none)		
AL1G20210	AT1G09660	RNA-binding KH domain-containing protein	Drought (20)
AL1G20220	AT1G09665	TIR domain protein	
AL1G20230	AT1G09680	PPR superfamily protein	
AL1G20240	AT1G09590	SH3-like translation protein	
AL1G29750	AT1G17460	TRFL3: MYB family transcription factor	
AL1G29760	AT1G17470	DRG1: DRG (developmentally regulated G-protein) protein.	Drought (20)
AL1G29770	AT1G17480	IQ-DOMAIN 7, Contains IQ calmodulin-binding region	Calcium, Drought (20)
AL1G50400	(none)		
AL1G51900	(none)		
AL2G23100	AT1G65810	P-loop containing nucleoside triphosphate hydrolase	
AL3G20210	AT3G09130	unknown protein	
AL3G20220	AT3G09140	unknown protein	
AL3G20230	AT3G09150	ELONGATED HYPOCOTYL 2 (HY2);GENOMES UNCOUPLED 3 (GUN3), a ferredoxin-dependent biliverdin reductase.	Iron, drought (20)
AL3G26610	AT3G14290	20S PROTEASOME ALPHA SUBUNIT E2 (PAE2)	
AL3G38380	AT3G23640	HETEROGLYCAN GLUCOSIDASE 1 (HGL1)	Drought (20)
AL3G53920	AT2G20810	LGT4	Drought (20)
AL3G53930	AT2G20815	QWRF3	Drought (20)
AL4G20110	AT2G26890	GRAVITROPISM DEFECTIVE 2: gravitropic response in hypocotyls and shoots. The mutants are defective in amyloplast sedimentation.	root function, drought (20)
AL4G34790	AT2G37840	unknown protein; has Calcium/calmodulin-dependent protein kinase-like domain	Calcium
AL4G34800	AT2G37860	LOWER CELL DENSITY 1 (LCD1)	Calcium (21), drought (20)
AL5G28090	AT3G48670	RDM12: Double-stranded RNA-binding protein involved in de novo methylation and siRNA-mediated maintenance methylation	
AL5G28100	AT3G48675		
AL5G28110	AT3G48680	GCAL2: A mitochondrial gamma carbonic anhydrase-like protein. Component of the NADH dehydrogenase complex	
AL5G30250	(none)		
AL5G30260	(none)		
AL5G37210			
AL5G37220	AT3G55940	Phosphoinositide-specific phospholipase C family protein; EF-hand-like, C2 Ca2+-dependent membrane targeting	Calcium (22), DREB2A target(23), Calcium-dependent, Drought (20)
AL5G40990	AT3G59100	GSL11: Similar to callose synthase	
AL6G13040	AT5G03555	NCS1: PLUTO (plastidic nucleobase transporter), a member of the Nucleobase:Cation-Symporter1 protein family	Solute transport, cation symporters (24, 25), cation stress response (21)
AL6G13050	AT5G03560	TPR-like; FUNCTIONS IN: nucleobase:cation symporter activity	Solute transport, cation symporter (24, 25), drought (20, 21)
AL6G13060	AT5G03560	TPR-like; FUNCTIONS IN: nucleobase:cation symporter activity	Solute transport, cation symporter (23, 24), drought (20)
AL6G13070	AT5G03570	IREG2: Nickel transport protein. Ortholog of the iron efflux transporter ferroportin (FPN) identified in animals	Also found in (15); Iron
AL6G25010	AT5G14310	CXE16	
AL6G25020	AT5G14320	EMB3137: Ribosomal protein S13/S18 family	Drought (20)
AL6G25030	AT5G14330	Unknown protein	
AL6G29070	AT5G17970	TIR-NBS-LRR	
AL6G40930	AT5G28900	Calcium-binding EF-hand family protein; FUNCTIONS IN: calcium ion binding	Calcium
AL6G41390			
AL6G42430	AT4G08620	SULPHATE TRANSPORTER 1;1: Encodes a sulfate transporter. Contains STAS domain. Expressed in roots and guard cells. Up-regulated by sulfur deficiency.	Sulfate, transporter, root function; Calcium (21)
AL7G21910	AT4G30240	Syntaxin/t-SNARE family protein; INVOLVED IN: Golgi vesicle transport, vesicle-mediated transport	
AL7G21920	AT4G30230	unknown protein	Calcium (27), drought (20)
AL7G21930	AT4G30220	SMALL NUCLEAR RIBONUCLEOPROTEIN F (RUXF)	Drought (20)
AL7G30250	AT4G23220	CRK14: A cysteine-rich receptor-like protein kinase	
AL7G40510	(none)		

AL7G40520	AT3G22110	PAC1: Alpha-3 subunit of 20s proteasome	Drought (20)
AL7G42040	AT4G14270	Containing PAM2 motif which mediates interaction with the PABC domain of polyadenyl binding proteins	Drought (20)
AL7G42050	AT4G14260		
AL7G42060	(none)		

Table S8. Top Sweep Candidates based on window-based D_{xy}, triple metric (F_{ST}, DD, 2dSFS) and extreme differentiated SNPs.

A.lyrata ID	*A.thaliana* ID	*A. thaliana* name and description	Serpentine relevance
AL1G14770	AT1G05020	ENTH/ANTH/VHS superfamily; FUNCTIONS IN: phospholipid binding	
AL1G14780	AT1G05030	Major facilitator superfamily protein; FUNCTIONS IN: transmembrane transporter activity, sugar:hydrogen symporter activity general substrate transporter	
AL1G14890	(none)		
AL1G14900	AT1G05140	Peptidase M50 family protein; FUNCTIONS IN: metalloendopeptidase activity; CONTAINS putative membrane-associated Zn metallopeptidase	Zinc; Drought (20)
AL1G14910	AT1G05150	Ca-binding tetratricopeptide family protein; FUNCTIONS IN: Zinc ion binding, Ca^{2+}ion binding	Calcium, Zinc; Drought (20)
AL1G15390	AT1G05540		
AL1G15400	AT1G05550		
AL1G15410	(none)		
AL1G15420	(none)		
AL1G15910	AT1G05960	ARM repeat superfamily protein	Drought (20)
AL1G18370	AT1G08035	Unknown protein	
AL1G18380	AT1G08040		Drought (20)
AL1G19510	AT1G09070	SRC2: Involved in protein storage vacuole targeting	Vacuole; Potassium (26)
AL1G20210	AT1G09660	RNA-binding KH domain-containing protein	Drought (20)
AL1G20220	AT1G09665	TIR domain protein	
AL1G20230	AT1G09680	PPR superfamily protein	
AL1G20240	AT1G09590	SH3-like translation protein	
AL1G20770	AT1G10070	BCAT2: Branched-chain amino acid aminotransferase	Calcium (27); Drought (20)
AL1G20780	AT1G10090	ERD4: Early-responsive to dehydration stress protein	Dehydration (28); vacuolar (28)
AL1G20790	AT1G10095	Protein prenylyltransferase superfamily protein	
AL1G29750	AT1G17460	TRFL3: MYB family transcription factor	
AL1G29760	AT1G17470	DRG1: DRG (developmentally regulated G-protein) protein. Has GTPase activity	Drought (20)
AL1G29770	AT1G17480	IQ-DOMAIN 7, Contains IQ calmodulin-binding region	Calcium, Drought (20)
AL1G32190	AT1G19440	KCS4: A member of the 3-ketoacyl-CoA synthase family	Drought (20)
AL1G32200	AT1G19450	Major facilitator superfamily; FUNCTIONS IN: transmembrane transporter activity, sugar:hydrogen symporter; LOCATED IN: plasma membrane, vacuole	Calcium (21); Vacuole (29); Drought (20)
AL1G40950	AT1G27520	MNS5: Glycosyl hydrolase family 47 protein; Ca^{2+} ion binding; LOCATED IN: endomembrane system, membrane	Calcium; Drought (20)
AL1G40960	AT1G27530		
AL1G48060	AT1G33612	LRR family protein; LOCATED IN: endomembrane system	
AL1G50400	(none)		
AL1G51900	(none)		
AL1G54340	(none)		
AL1G54350	AT1G47900		
AL1G59090	AT1G51260	LPAT3: Acyl-CoA: 1-acylglycerol-3-phosphate acyltransferase	
AL1G59100			Also found in (15)
AL1G59110	AT1G51310	Transferases;tRNA (5-methylaminomethyl-2-thiouridylate)-methyltransferases	Also found in (15); Drought (20)
AL1G65400	AT1G54920	Unknown protein	Drought (20)
AL2G11840	AT1G63310	Unknown protein	Drought (20)
AL2G12140	(none)		
AL2G13380	AT1G62500	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily	
AL2G13390	(none)		
AL2G13400	AT1G62480	Vacuolar Ca^{2+}-binding protein-related; INVOLVED IN: response to Cd ion, response to salt stress	Calcium; Cadmium; Drought (20)
AL2G23100	AT1G65810	P-loop containing nucleoside triphosphate hydrolase	
AL3G20210	AT3G09130	unknown protein	
AL3G20220	AT3G09140	unknown protein	
AL3G20230	AT3G09150	ELONGATED HYPOCOTYL 2 (HY2);GENOMES UNCOUPLED 3 (GUN3), a ferredoxin-dependent biliverdin reductase.	Iron, drought (20)
AL3G21440	AT3G10116	COBRA-like extracellular glycosyl-phosphatidyl inositol-anchored protein	
AL3G21450	AT3G10130	Heme-binding protein; Has SOUL haem-binding domain	Iron
AL3G21460	AT3G10120	Unknown protein	

AL3G22530	(none)		Also found in (15)
AL3G22540	AT3G10985	SAG20: A senescence-associated gene whose expression is induced in response to treatment with Nep1, a fungal protein that causes necrosis	Also found in (15); Calcium (27); Drought (20)
AL3G26610	AT3G14290	20S PROTEASOME ALPHA SUBUNIT E2 (PAE2)	
AL3G38380	AT3G23640	HETEROGLYCAN GLUCOSIDASE 1 (HGL1)	Drought (20)
AL3G43150	AT2G05710	ACO3	
AL3G43160	(none)		
AL3G49430	(none)		
AL3G49440	(none)		
AL3G53120	AT2G18800	XTH21: Xyloglucan endotransglucosylase/hydrolase 21 (XTH21); INVOLVED IN: primary root development, cell wall modification; LOCATED IN: endomembrane	Calcium (27); Root functions
AL3G53130	AT2G18790	PHYB : Red/far-red photoreceptor	Drought (20)
AL3G53920	AT2G20810	LGT4	Drought (20)
AL3G53930	AT2G20815	QWRF3	Drought (20)
AL4G20110	AT2G26890	GRAVITROPISM DEFECTIVE 2: gravitropic response in hypocotyls and shoots. The mutants are defective in amyloplast sedimentation.	root function, drought (1720
AL4G23500	AT2G28950	ATEXPA6: Expansin. Involved in the syncytia formation in *A. thaliana* roots.	Root function
AL4G32180	AT2G36090	F-box family protein	Drought (20)
AL4G32190	AT2G36100	CASP1: A membrane bound protein involved in formation of the casparian strip. Required for the localization of ESB1.	Casparian strip (30)
AL4G32200	(none)		
AL4G34490	(none)		
AL4G34500	(none)		
AL4G34510	(none)		
AL4G34520	(none)		
AL4G34530	(none)		
AL4G34790	AT2G37840	unknown protein; has Calcium/calmodulin-dependent protein kinase-like domain	Calcium
AL4G34800	AT2G37860	LOWER CELL DENSITY 1 (LCD1)	Calcium (21), drought (20)
AL4G35360	(none)		
AL4G35380	AT2G38290	AMT2;1: High-affinity ammonium transporter, expressed in root. Expression in root and shoot is under nitrogen and carbon dioxide regulation, respectively.	Nitrongen, Root function, transporter; Drought (20)
AL4G35390	AT2G38300	MYB-like HTH transcriptional regulator family protein	
AL4G42980	(none)		
AL4G42990	AT2G44110	MLO15: Seven-transmembrane domain protein expressed in root	Root expression
AL5G19840	AT2G10440	Unknown protein	
AL5G28090	AT3G48670	RDM12: Double-stranded RNA-binding protein involved in de novo methylation and siRNA-mediated maintenance methylation	
AL5G28100	AT3G48675		
AL5G28110	AT3G48680	GCAL2: A mitochondrial gamma carbonic anhydrase-like protein. Component of the NADH dehydrogenase complex	
AL5G30250	(none)		
AL5G30260	(none)		
AL5G37200	AT3G55920	Cyclophilin-like peptidyl-prolyl cis-trans isomerase family protein	
AL5G37210			
AL5G37220	AT3G55940	Phosphoinositide-specific phospholipase C family protein; EF-hand-like, C2 Ca^{2+}-dependent membrane targeting	Calcium (31), DREB2A target (23), Calcium-dependent, Drought (20)
AL5G38300	AT3G56900	Transducin/WD40 repeat-like superfamily protein	Drought (20)
AL5G38310	AT3G56910	PSRP5	
AL5G40820	(none)		
AL5G40830	AT3G59000	F-box/RNI-like protein	
AL5G40840	(none)		
AL5G40900	AT3G59020	ARM repeat superfamily protein; FUNCTIONS IN: transporter activity.	
AL5G40990	AT3G59100	GSL11: Similar to callose synthase	
AL6G11190	AT5G01040	LAC8: Putative laccase, knockout mutant showed early flowering	Flowering time
AL6G11200	AT5G01030		
AL6G13040	AT5G03555	NCS1: PLUTO (plastidic nucleobase transporter), a member of the Nucleobase:Cation-Symporter1 protein family	Solute transport, cation symporters (24, 25), cation stress response (21)
AL6G13050	AT5G03560	TPR-like; FUNCTIONS IN: nucleobase:cation symporter activity	Solute transport, cation symporter (24, 25), drought (20, 21)
AL6G13060	AT5G03560	TPR-like; FUNCTIONS IN: nucleobase:cation symporter activity	Solute transport, cation symporter (24, 25), drought (20)

AL6G13070	AT5G03570	IREG2: Nickel transport protein. Ortholog of the iron efflux transporter ferroportin (FPN) identified in animals	Also found in (15); Iron
AL6G13940	AT5G04320	SGO2: Protects meiotic centromere cohesion	Also found in (15); Drought (20)
AL6G13950	AT5G04330	CYP84A4: Cytochrome P450 protein; FUNCTIONS IN: electron carrier activity, monooxygenase activity, iron ion binding, oxygen binding, heme binding	Also found in (15); Iron
AL6G15770	AT5G05860	UGT76C2: Cytokinin N-glucosyltransferase. Induced by ABA and drought stress. Functions in response to osmotic and drought stress.	Drought (20)
AL6G15780	AT5G05870	UGT76C1	Drought (20)
AL6G16670	AT5G06580	Glycolate dehydrogenase activity	Drought (20)
AL6G17670	AT5G07370	IPK2a	Drought (20)
AL6G17680	AT5G07380	Unknown protein	
AL6G17690	AT5G07390	RBOHA: Ferredoxin reductase domain; EF-Hand 1, Ca^{2+}-binding site	Calcium; Iron
AL6G20720	AT5G10230	ANN7: Ca^{2+}-binding	Calcium; Drought (32)
AL6G20740	AT5G10240	ASN3: Asparagine synthetase	Drought (20)
AL6G25010	AT5G14310	CXE16	
AL6G25020	AT5G14320	EMB3137: Ribosomal protein S13/S18 family	Drought (20)
AL6G25030	AT5G14330	Unknown protein	
AL6G29070	AT5G17970	TIR-NBS-LRR	
AL6G34110	AT5G22980	SCPL47	
AL6G40930	AT5G28900	Calcium-binding EF-hand family protein; FUNCTIONS IN: calcium ion binding	Calcium
AL6G41020	AT5G30490		Drought (20)
AL6G41390			
AL6G42430	AT4G08620	SULPHATE TRANSPORTER 1;1: Encodes a sulfate transporter. Contains STAS domain. Expressed in roots and guard cells. Up-regulated by sulfur deficiency.	Sulfate, transporter, root function; Calcium (21)
AL6G43330	(none)		
AL6G49300	(none)		
AL6G49310	AT4G03560	TPC1: Ca^{2+} channel that mediates a voltage-activated Ca^{2+}influx. Mutants lack detectable SV channel activity.	Calcium; Also found in (15); Vacuole (29); Drought (20)
AL7G11870	AT4G39675	Unknown protein	Drought (20)
AL7G11880	AT4G39680	SAP domain-containing protein	Drought (20)
AL7G12060	AT4G39830	Cupredoxin superfamily protein; FUNCTIONS IN: oxidoreductase activity, copper ion binding	Copper; Cesium (26),
AL7G12820	AT4G37860	SPT2 chromatin protein	
AL7G12830	AT4G37840	HKL3: A putative hexokinase	
AL7G12840	AT4G37830	Cytochrome c oxidase-related	
AL7G13170	AT4G37590	MEL1: Involved in auxin-mediated organogenesis	Drought (20), Salt (33)
AL7G13850	(none)		
AL7G19030	AT4G32710	PERK14: Proline-rich extensin-like receptor kinase	
AL7G21910	AT4G30240	Syntaxin/t-SNARE family protein; INVOLVED IN: Golgi vesicle transport, vesicle-mediated transport	
AL7G21920	AT4G30230	unknown protein	Calcium (27), drought (20)
AL7G21930	AT4G30220	SMALL NUCLEAR RIBONUCLEOPROTEIN F (RUXF)	Drought (20)
AL7G27590	AT4G25500	RS40: An arginine/serine-rich splicing factor. Root expressed	Drought (20)
AL7G27600	AT4G25490	CBF1: Transcriptional activator that binds to the DRE/CRT regulatory element and induces COR (cold-regulated) gene expression	
AL7G30250	AT4G23220	CRK14: A cysteine-rich receptor-like protein kinase	
AL7G33950	AT4G19960	KUP9: A potassium ion transmembrane transporter. Also mediates cesium uptake when expressed in E. coli	Potassium; Also found in (15); Drought (20)
AL7G34990	AT4G19100	PAM68	Drought (20)
AL7G35000	AT4G19090		
AL7G40500	AT4G15180	SDG2	
AL7G40510	(none)		
AL7G40520	AT3G22110	PAC1: Alpha-3 subunit of 20s proteasome	Drought (20)
AL7G40530	(none)		
AL7G40540	AT4G15160	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily	Drought (20)
AL7G42040	AT4G14270	Containing PAM2 motif which mediates interaction with the PABC domain of polyadenyl binding proteins	Drought (20)
AL7G42050	AT4G14260		
AL7G42060	(none)		
AL7G46090	(none)		
AL7G46550	AT5G37710	Alpha/beta-Hydrolases; FUNCTIONS IN: calmodulin binding	Also found in (15); Calcium; Drought (20)
AL7G46560	AT5G37720	ALY4: FUNCTIONS IN: nucleotide binding, nucleic acid binding	Also found in (15)
AL7G46700	AT5G37800	RSL1: *Best A.thaliana* protein match is: ROOT HAIR DEFECTIVE6	Root function

AL7G50000	(none)		
AL7G50010	(none)		
AL7G50020	(none)		
AL7G50850	AT5G39460	F-box family protein	
AL7G51660			
AL7G51670	AT5G38770	GDU7: Family involved in amino acid export	Drought (20)
AL8G19840	AT5G47850	CCR4: FUNCTIONS IN: kinase activity; INVOLVED IN: amino acid phosphorylation; LOCATED IN: endomembrane system	Calcium (27)
AL8G19850	AT5G47860		
AL8G31810	AT5G56180	ARP8	
AL8G41240	AT5G63980	HOS2: Rescues sulfur assimilation mutants in yeast. Involved in the response to cold, drought and ABA. Mutants exhibit induction of stress genes in response to cold, ABA, salt and dehydration Regulates flowering time.	Sulfur, drought, flowering time (34), Lithium tolerance (35)
AL8G41250	AT5G63990	Inositol monophosphatase protein; FUNCTIONS IN: inositol or phosphatidylinositol phosphatase; INVOLVED IN: sulfur metabolic process	Sulfur; Drought (20)

Table S9. Quantitative genomic assessment of population-specific introgression. Empirical genome-wide distribution ABBA/BABA four taxon test f_d 0.1% and 0.5% positive outliers present among Top Sweep Candidate loci.

f_d outlier percentile	*A.lyrata* ID	*A.thaliana* ID	*A.thaliana* name and description
0.1%	AL1G20780	AT1G10090	Early-responsive to dehydration stress (ERD4)
0.1%	AL4G42990	AT2G44110	MLO15: Seven-transmembrane domain protein expressed in root
0.1%	AL7G19030	AT4G32710	PERK14: Proline-rich extensin-like receptor kinase
0.1%	AL7G42040	AT4G14270	Containing PAM2 motif
0.1%	AL7G42050	AT4G14260	
0.1%	AL7G42060	(none)	
0.5%	AL5G40990	AT3G59100	GSL11: Similar to callose synthase
0.5%	AL1G40960	AT1G27530	
0.5%	AL2G23100	AT1G65810	P-loop containing nucleoside triphosphate hydrolase
0.5%	AL3G38380	AT3G23640	HETEROGLYCAN GLUCOSIDASE 1 (HGL1)
0.5%	AL6G29070	AT5G17970	TIR-NBS-LRR
0.5%	AL6G49310	AT4G03560	TPC1: Ca^{2+} channel that mediates a voltage-activated Ca^{2+}influx. Mutants lack detectable SV channel activity.
0.5%	AL7G27590	AT4G25500	RS40: An arginine/serine-rich splicing factor. Root expressed
0.5%	AL7G27600	AT4G25490	CBF1: Transcriptional activator that binds to the DRE/CRT regulatory element and induces COR (cold-regulated) gene expression
0.5%	AL7G50020	(none)	

Table S10. Loci overlapping or within 2kb of most differentiated single SNPs in this study and (15).

A.lyrata ID	*A.thaliana* ID	*A.thaliana* name and description
AL1G59100		
AL1G59110	AT1G51310	Transferases;tRNA (5-methylaminomethyl-2-thiouridylate)-methyltransferases; Drought (20)
AL3G22530		
AL3G22540	AT3G10985	SAG20: A senescence-associated gene whose expression is induced in response to treatment with Nep1, a fungal protein that causes necrosis; Calcium (27); Drought (20)
AL6G49310	AT4G03560	TPC1: Ca2+ channel that mediates a voltage-activated Ca2+influx. Mutants lack detectable SV channel activity Calcium; Vacuole (29); Drought (20)
AL7G33950	AT4G19960	KUP9: A potassium ion transmembrane transporter. Also mediates cesium uptake when expressed in E. coli; Potassium; Drought (20)

Table S11. Loci overlapping or within 2kb of our Top Sweep Candidates (window based lists and single differentiated SNPS) in this study and (15).

A.lyrata ID	*A.thaliana* ID	*A.thaliana* name and description
AL1G59100		
AL1G59110	AT1G51310	Transferases;tRNA (5-methylaminomethyl-2-thiouridylate)-methyltransferases; Drought (20)
AL3G22530		
AL3G22540	AT3G10985	SAG20: A senescence-associated gene whose expression is induced in response to treatment with Nep1, a fungal protein that causes necrosis; Calcium (27); Drought (20)
AL6G13070	AT5G03570	IRON REGULATED 2 (IREG2): encodes a tonoplast localized nickel transport protein.
AL6G13940	AT5G04320	SHUGOSHIN 2 (SGO2): Encodes a protein that protects meiotic centromere cohesion.
AL6G13950	AT5G04330	Cytochrome P450 superfamily protein; FUNCTIONS IN: electron carrier activity, monooxygenase activity, iron ion binding, oxygen binding, heme binding
AL6G49310	AT4G03560	TPC1: Ca2+ channel that mediates a voltage-activated Ca2+influx. Mutants lack detectable SV channel activity; Vacuole (29); Drought (20)
AL7G33950	AT4G19960	KUP9: A potassium ion transmembrane transporter. Also mediates cesium uptake when expressed in E. coli; Potassium; Drought (23)
AL7G46550	AT5G37710	alpha/beta-Hydrolases superfamily protein; FUNCTIONS IN: triglyceride lipase activity, calmodulin binding.
AL7G46560	AT5G37720	ALY4: FUNCTIONS IN: nucleotide binding, nucleic acid binding

Supplemental References

1. Arnold B, Kim ST, Bomblies K (2015) Single Geographic Origin of a Widespread Autotetraploid Arabidopsis arenosa Lineage Followed by Interploidy Admixture. *Molecular Biology and Evolution* 32(6):1382–1395.

2. Excoffier L, et al. (2013) Robust demographic inference from genomic and SNP data. *PLoS Genetics* 9(10): e1003905. doi:10.1371/journal.pgen.1003905

3. Johnson J, Omland K (2004) Model selection in ecology and evolution. *Trends in ecology & evolution* 19(2): 101-108

4. Yant L, et al. (2013) Meiotic Adaptation to Genome Duplication in Arabidopsis arenosa. *Current Biology* 23(21):2151–2156.

5. Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetjournal* 17(1):pp. 10–12.

6. Hu TT, et al. (2011) The Arabidopsis lyrata genome sequence and the basis of rapid genome size change. *Nat Genet* 43(5):476–481.

7. Lunter G, Goodson M (2011) Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Res* 21(6):936–939.

8. DePristo MA, et al. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43(5):491–498.

9. Vatsiou AI, Bazin É, Gaggiotti OE (2015) Detection of selective sweeps in structured populations: a comparison of recent methods. *Mol Ecol*. 25(1):89-103.

10. Smith J, Kronforst MR (2013) Do Heliconius butterfly species exchange mimicry alleles? *Biology Letters* 9(4):20130503–20130503.

11. Ross-Ibarra J, et al. (2008) Patterns of Polymorphism and Demographic History in Natural Populations of Arabidopsis lyrata. *PLoS ONE* 3(6):e2411.

12. Nielsen R, et al. (2009) Darwinian and demographic forces affecting human protein coding genes. *Genome Res* 19(5):838–849.

13. Rawat V, et al. (2015) Improving the Annotation of Arabidopsis lyrata Using RNA-Seq Data. *PLoS ONE* 10(9):e0137391–12.

14. Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26(6):841–842.

15. Turner TL, Bourne EC, Wettberg Von EJ, Hu TT, Nuzhdin SV (2010) Population resequencing reveals local adaptation of Arabidopsis lyrata to serpentine soils. *Nature Genetics* 42(3):260–263.

16. Hollister JD, et al. (2012) Genetic Adaptation Associated with Genome-Doubling in Autotetraploid Arabidopsis arenosa. *PLoS Genet* 8(12):e1003093.

17. Ward BJ, van Oosterhout C (2015) hybridcheck: software for the rapid detection, visualization and dating of recombinant regions in genome sequence data. *Mol Ecol Resour* doi: 10.1111/1755-0998.12469.

18. Martin SH, Davey JW, Jiggins CD (2014) Evaluating the Use of ABBA-BABA Statistics to Locate Introgressed Loci. *Molecular Biology and Evolution* 32(1):244–257.

19. Busing F, et al. (1999) Delete-m jackknife for unequal m. Statistics and Computing 9: 3-8.

20. Des Marais DL, et al. (2012) Physiological genomics of response to soil drying in diverse Arabidopsis accessions. *The Plant Cell* 24(3):893–914.

21. Maathuis FJM, et al. (2003) Transcriptome analysis of root transporters reveals participation of multiple gene families in the response to cation stress. *The Plant Journal* 35(6):675–692.

22. Curran A, et al. (2011) Calcium-dependent protein kinases from Arabidopsis show substrate specificity differences in an analysis of 103 substrates. *Frontiers in plant science* doi:10.3389/fpls.2011.00036.

23. Qin F, et al. (2007) Regulation and functional analysis of ZmDREB2A in response to drought and heat stresses in Zea mays L. *The Plant Journal* 50(1):54–69.

24. Mourad GS, et al. (2012) Genetic and molecular characterization reveals a unique nucleobase cation symporter 1 in Arabidopsis. *FEBS Letters* 586(9):1370–1378.

25. Schein JR, Hunt KA, Minton JA, Schultes NP, Mourad GS (2013) The nucleobase cation symporter 1 of Chlamydomonas reinhardtii and that of the evolutionarily distant Arabidopsis thaliana display parallel function and establish a plant-specific solute transport profile. *Plant Physiol Biochem* 70:52–60.

26. Hampton CR, et al. (2004) Cesium toxicity in Arabidopsis. *Plant Physiology* 136(3):3824–3837.

27. Wang J, et al. (2014) Arabidopsis transcriptional response to extracellular Ca 2+depletion involves a transient rise in cytosolic Ca 2. *J Integr Plant Biol* 57(2):138-50.

28. Seki M, et al. (2001) Monitoring the expression pattern of 1300 Arabidopsis genes under drought and cold stresses by using a full-length cDNA microarray. *The Plant Cell* 13(1):61–72.

29. Carter C, et al. (2004) The vegetative vacuole proteome of Arabidopsis thaliana reveals predicted and unexpected proteins. *The Plant Cell* 16(12):3285–3303.

30. Hosmani PS, et al. (2013) Dirigent domain-containing protein is part of the machinery required for formation of the lignin-based Casparian strip in the root. *Proc Natl Acad Sc* 110(35):14498–14503.

31. Curran A, Chang F, Chang CL, Garg S (2011) Calcium-dependent protein kinases from Arabidopsis show substrate specificity differences in an analysis of 103 substrates. *Frontiers in plant science* doi:10.3389/fpls.2011.00036.

32. del Pozo JC, Ramirez-Parra E (2014) Deciphering the molecular bases for drought tolerance in Arabidopsis autotetraploids. *Plant, Cell & Environment*. doi:10.1111/pce.12344.

33. Taji T (2004) Comparative Genomics in Salt Tolerance between Arabidopsis and Arabidopsis-Related Halophyte Salt Cress Using Arabidopsis Microarray. *Plant Physiology* 135(3):1697–1709.

34. Gašparič MB, et al. (2013) Insertion of a Specific Fungal 3′-phosphoadenosine-5′-phosphatase Motif into a Plant Homologue Improves Halotolerance and Drought Tolerance of Plants. *PLoS ONE* 8(12):e81872–12.

35. Xiong L, Lee H, Huang R, Zhu J-K (2004) A single amino acid substitution in the Arabidopsis FIERY1/HOS2 protein confers cold signaling specificity and lithium tolerance. *The Plant Journal* 40(4):536–545.