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.

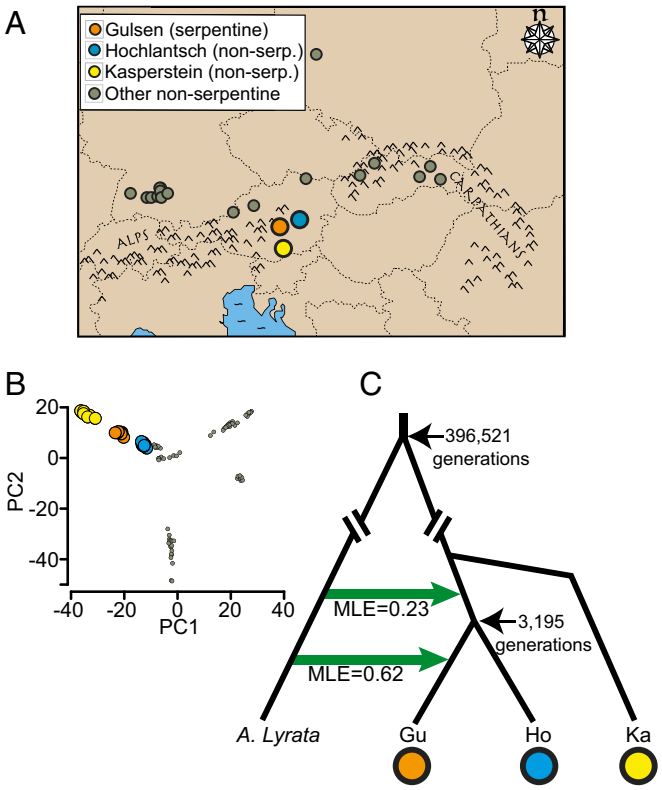

Fig. 1. *A. arenosa* populations sampled for this study. (*A*) Locations of the 29 *A. arenosa* populations sampled. Orange dot gives the location of the focal Gulsen (GU) serpentine population along with other highlighted populations at Hochlantsch (HO) and Kasparstein (KA). Note: one Swedish location 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

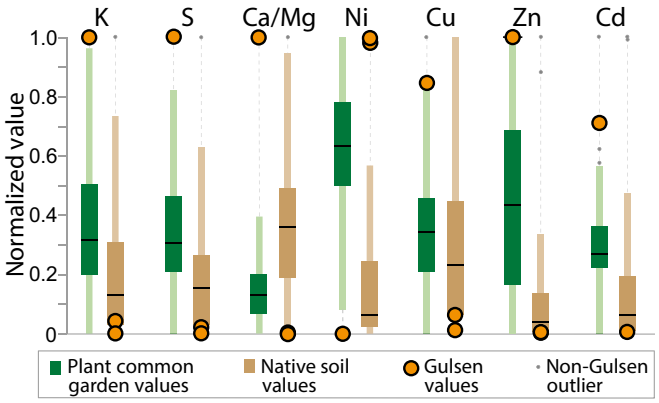

Fig. 2. Serpentine *A. arenosa* is an extreme outlier for the accumulation of many elements. Elemental profiling of 29 *A. arenosa* populations. Green distributions represent plant tissue data, brown distributions represent data from soil collected at plant sites. Orange dots indicate position in distribution 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 $x' = (x - x_{min})/(x_{max} - x_{min})$.

PLANT BIOLOGY

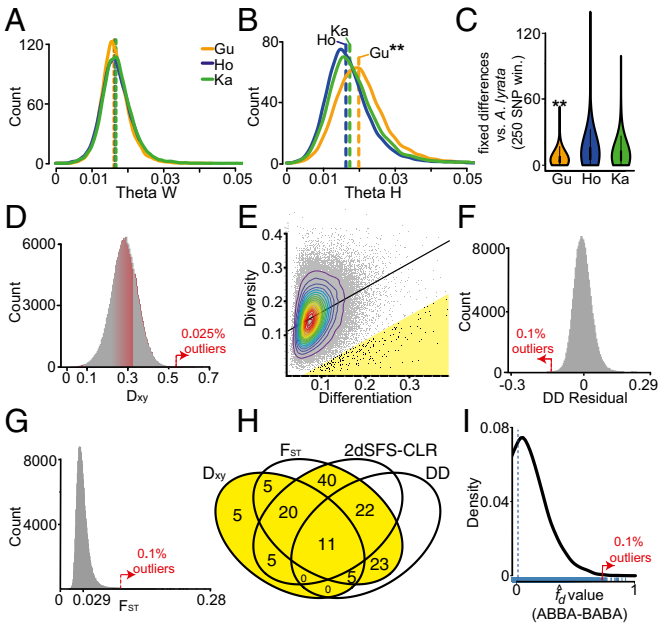

Fig. 3. Measures of differentiation. (*A*) Watterson estimator θ_W diversity in resequenced populations over genome windows. The vertical dashed line for each population gives the mean. (*B*) θ_H, a diversity metric sensitive to extreme frequency SNPs (double asterisk signifies that Gulsen distribution 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)

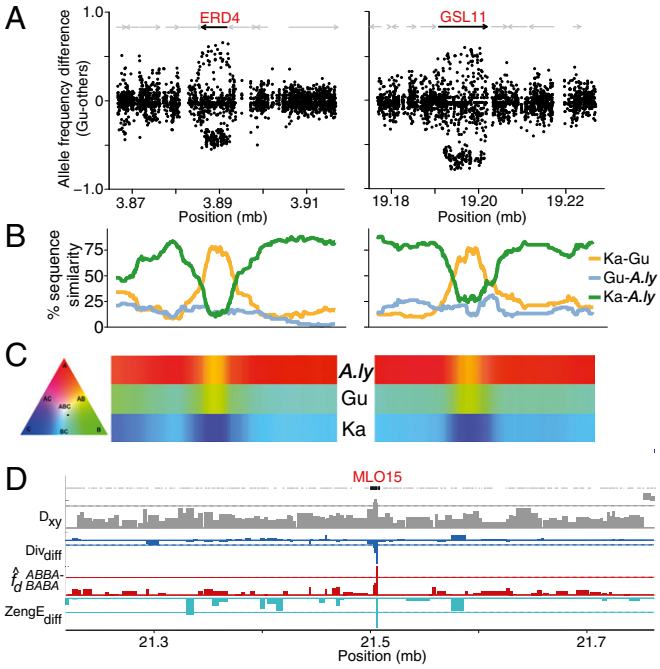

Fig. 4. Selective sweeps on *A. lyrata*-like alleles in serpentine *A. arenosa*. (*A*) Allele frequency differences in example differentiated regions. Dots represent polymorphic SNPs. The *x* axis gives chromosome location; *y* axis gives degree of differentiation calculated by plotting the difference 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·m^2·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.