Charles University

Faculty of Science

Study programme: Botany

Mgr. Veronika Konečná

Evolutionary drivers and consequences of parallel evolution in plants

Evoluční mechanismy a důsledky paralelní evoluce u rostlin

Doctoral thesis

Supervisor: RNDr. Filip Kolář, Ph.D.

Prague, 2022

Content

Declaration

I hereby declare that this thesis has been composed by myself using the mentioned references and that it has not been submitted elsewhere, in whole or in part, to obtain the same or other academic degree.

Prohlašuji, že jsem tuto práci zpracovala samostatně a že jsem uvedla všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze, 3. 2. 2022 Veronika Konečná

Author contribution statement

I confirm that I have substantially contributed to all papers included to the following extent:

Case study 1: **Konečná, V.**, Yant, L. & Kolář, F. (2020). The evolutionary genomics of serpentine adaptation. *Frontiers in Plant Science* 11: 574616. doi.org/10.3389/fpls.2020.574616
Literature synthesis, manuscript writing – 70 %

Case study 2: Knotek, A., **Konečná, V.**, Wos, G., Požárová, D., Šrámková, G., Bohutínská, M., Zeisek, V., Marhold, K. & Kolář, F. (2020). Parallel alpine differentiation in *Arabidopsis arenosa*. *Frontiers in Plant Science* 11: 561526. doi.org/10.3389/fpls.2020.561526
Population genomic data analyses, manuscript editing – 20 %

Case study 3: **Konečná, V.**, Nowak, M. D. & Kolář, F. (2019). Parallel colonization of subalpine habitats in the central European mountains by *Primula elatior*. *Scientific Reports* 9: 3294. doi.org/10.1038/s41598-019-39669-2
Study design, field sampling, lab work, data analyses, manuscript writing – 90 %

Case study 4: **Konečná, V.**, Bray, S., Vlček, J., Bohutínská, M., Požárová, D., Choudhury, R. R., Bollmann-Giolai, A., Parisod, C., Yant, L. & Kolář, F. (2021). Parallel adaptation in autopolyploid *Arabidopsis arenosa* is dominated by repeated recruitment of shared alleles. *Nature Communications* 12: 4979. doi.org/10.1038/s41467-021-25256-5
Study design, field sampling, lab work, experiments, data analyses, manuscript writing – 90 %

Case study 5: **Konečná, V.**, Šustr, M., Požárová, D., Čertner, M., Krejčová, A., Tylová, E. & Kolář, F. Genomic basis and phenotypic manifestation of (non-)parallel serpentine adaptation in *Arabidopsis arenosa.* Manuscript.
Study design, experiments, data analyses, manuscript writing – 80 %

Acknowledgements

I would like to thank my supervisor Filip Kolář for being such a great mentor and friend in one person, for his unlimited guidance during my studies, for teaching me how to ask interesting scientific questions, write manuscripts, and for his valuable advices regarding data analyses. I highly appreciate that he has always been happy to discuss not only scientific challenges no matter if he was in Prague, Innsbruck, Oslo or the African mountains. He showed me the world of science with its pros and cons and provided me space for scientific growth. Thanks to Filip we also established great collaborations and I fully leveraged the advantages of a big team and large collaboration network. I will always be grateful to him for giving me a chance to work on exciting projects and believing in my skills! I am also very grateful to Levi Yant and Christian Parisod for hours of inspiring scientific discussions and for providing me an option for research stays in their labs. Further, I would like to specifically thank Levi for giving me valuable lessons in scientific writing and for his unlimited contagious enthusiasm. I thank also to other co-authors of our publications.

I am grateful to my colleagues from Plant ecological genomics group for having fun with them during retreats and for the friendly inspiring atmosphere in the group. I also really appreciate their help with experiments, bioinformatics, wet lab and field works (especially driving and collecting tens of litres of serpentine soil!). I enjoyed all serpentine field works and specifically I thank Anita, Majda, Sonia, and Paolo for being great companions. I greatly thank Doubravka, without her enormous help I would not have finished the experiments, and Timothée for sharing passion for serpentines with me.

Besides the group, I also thank my other colleagues and schoolmates for the time, which we spent together not only during lunches and coffee breaks (the pink cytometric/chilling lab will always have a place in my heart). Importantly, thanks to my dear friends Lenka, Honza P., and Pajarito, I could take a short rest from my projects and enjoy an incredible adventurous month in Argentina.

My greatest thanks go to my fiancé Marek for his patience (spending hours and hours of listening to stories of my scientific falls and rises), kindness and understanding, sharing a passion for (data) science, and importantly ability to always cheer me up! Finally, I thank to my parents for their support during my studies.

Abstract

Parallel adaptation to similar environmental pressures provides ideal model systems to study the repeatability of evolution in nature. Such replicated natural experiments can also provide important insights into the genomic basis of adaptations. However, well-documented examples are rare, particularly in plants. Here, I brought new evidence of parallel evolution from three plant systems facing one of the most challenging selective environments – alpine stands and toxic serpentine soils. Further, I leveraged the cases of naturally replicated parallel adaptation in *Arabidopsis arenosa* to study the extent of phenotypic and genomic parallelism and to address the evolutionary sources of the parallel genetic underpinnings.

By combining population genetic and experimental approaches I documented the complex interplay of adaptive, historical, and ecological processes in parallel evolution. The island-like distribution of high-elevation and serpentine habitats promoted their independent colonizations by distinct genetic lineages of *A. arenosa* and *Primula elatior* separately in each geographic region. Further, I showed how challenging environments structure genetic diversity within a species. For instance, I found higher genetic differentiation among (sub)alpine populations than among foothill populations. This suggests that mountain ridges act as migration barriers reducing gene flow among (sub)alpine populations. Moreover, colonization of (sub)alpine or serpentine habitats did not result in loss of genetic diversity suggesting rather gradual colonization by large populations than a strong bottleneck.

Taking advantage of multiple natural replicates of alpine and serpentine populations of *A. arenosa*, I quantified the magnitude of phenotypic parallelism and investigated its neutral and adaptive determinants and showed considerable differences among the systems. While in the alpine system, only a subset of traits showed a parallel response, I showed pervasive parallelism in serpentine *A. arenosa* in functional traits including similarly modified ion uptake differentiating serpentine and non-serpentine populations.

Further, I studied the genetic basis of five-fold parallel serpentine adaptation in *A. arenosa*. I detected significant parallelism, both at the gene and functional level involving e.g. ion homeostasis, inorganic anion transport, calcium transmembrane transporter activity, and response to metal ions. Next, I inferred the evolutionary sources of the parallel adaptive variation. I found that shared variation is the predominant source of parallel adaptive variants, in line with the population genomic properties of the highly variable and recently diverged tetraploid populations of *A. arenosa*. However, I also discovered an exceptional parallel locus candidate, *TPC1*, with parallel de novo mutations in a single codon in two distinct serpentine populations. Such a finding demonstrates that the rapid selection of novel alleles is still feasible in autopolyploids, perhaps reflecting the maintenance of a large pool of pre-existing variation and increased rates of beneficial alleles in organisms with doubled genomes.

In summary, cases of parallel evolution provide important insights into evolutionary drivers of adaptation and the identification of novel models of parallel evolution is a fruitful approach. The next step will be to deconstruct complex adaptations and find the crucial link between the phenotypic effect of a locus and its adaptive value in challenging environments. With a deeper understanding of the genetic architecture of repeated adaptations, closing this gap is not far off.

Abstrakt

Paralelní adaptace na podobné tlaky prostředí poskytuje ideální modelový systém pro studium opakovatelnosti evoluce v přírodě. Takovéto přírodní experimenty nám mohou také poskytnout důležité poznatky o genomické podstatě adaptací. Dobře zdokumentované příklady paralelní evoluce jsou však zejména u rostlin vzácné. Má práce dokumentuje paralelní evoluci u tří rostlinných systémů, které se přizpůsobily na jedny z nejnáročnějších selektivních prostředí – vysokohorská stanoviště a toxické hadcové půdy. Dále jsem jsem studovala rozsah paralelismu na fenotypové a genotypové úrovni u *Arabidopsis arenosa* spolu s evolučními zdroji paralelních genetických variant.

Kombinací populačně genetických a experimentálních přístupů jsem ukázala na komplexní souhru adaptivních, historických a ekologických procesů během paralelní evoluce. Izolované ostrovy horských a hadcových stanovišť umožnily jejich nezávislou kolonizaci odlišnými genetickými liniemi *A. arenosa* a *Primula elatior* v jednotlivých regionech. Ukázala jsem také, jakým způsobem se v rámci druhu mění genetická diverzita po kolonizaci nových prostředí. Zjistila jsem například vyšší genetickou diferenciaci mezi (sub)alpskými populacemi než mezi populacemi podhorskými. To naznačuje, že horské hřebeny představují migrační bariéru snižující genový tok mezi (sub)alpskými populacemi. Navíc kolonizace (sub)alpínských nebo hadcových stanovišť nevedla ke ztrátě genetické diverzity, což naznačuje spíše postupnou kolonizaci než silný bottleneck.

S využitím mnoha adaptovaných alpských a hadcových populací *A. arenosa* jsem kvantifikovala rozsah fenotypového paralelismu a zkoumala jeho neutrální a adaptivní determinanty a ukázala jsem na značné rozdíly mezi oběma systémy. Zatímco v alpském systému vykazovala paralelismus pouze podskupina znaků, u hadcových populací *A. arenosa* byl paralelismus všudypřítomný, a to například ve funkčních vlastnostech včetně podobně modifikovaného příjmu iontů odlišujícího hadcové a nehadcové populace.

Dále jsem studovala genetickou podstatu pětinásobné paralelní hadcové adaptace u *A. arenosa*. Zjistila jsem signifikantní paralelismus, jak na genové, tak na funkční úrovni zahrnující např. iontovou homeostázu, transport anorganických aniontů, aktivitu transmembránového transportéru vápníku a reakci na ionty kovů. Dále jsem studovala evoluční zdroje paralelních adaptivních variant. Zjistila jsem, že převládajícím zdrojem paralelních adaptivních variant je jejich sdílení, což je v souladu s vysoce variabilními a nedávno divergovanými tetraploidními populacemi *A. arenosa*. Překvapivým zjištěným bylo objevení paralelních de novo mutací v konzervovaném místě genu *TPC1* u dvou nezávislých hadcových populací. Takové zjištění ukazuje, že rychlá selekce nových alel je stále možná i u autopolyploidů, což může odrážet jejich velkou genetickou variabilitu a zvýšenou míru výhodných alel v organismech se zdvojenými genomy.

Závěrem, doložené příklady spolu s novými systémy paralelní evoluce nám poskytují důležité poznatky o evolučních adaptačních mechanismech. Dalším krokem bude de-konstrukce procesu adaptace a nalezení klíčového spojení mezi fenotypovým efektem lokusu a jeho adaptivní hodnotou v náročných prostředích. Díky hlubšímu pochopením genetické architektury opakovaných adaptací k tomu nebudeme mít daleko.

Part A – General chapters

1 Concept of parallelism

The same environmental pressure can cause similar phenotypic responses in phylogenetically distant populations. In turn, if the same phenotypes independently emerge in similar habitats it is more probable that these features evolved under natural selection than only due to the stochastic forces (Lenormand et al., 2009). Depending on the phylogenetic relatedness, we traditionally recognize parallel or convergent evolution of closely or distantly related genetic lineages responding to similar selection pressure (Arendt and Reznick, 2008). In convergent evolution, typical for distantly related genetic lineages, similar phenotypes evolved from different initial phenotypes (Bolnick et al., 2018). In my thesis, I focus on the parallel evolution of intraspecific lineages evolving from similar initial conditions. These cases of repeated evolution in natural populations represent independent biological replicates, which provide us insights into the mechanisms that facilitate parallel adaptation in nature (Yeaman et al., 2018). To study these evolutionary mechanisms, we first need to characterize a suitable model and quantify the extent of genetic and phenotypic parallelism among its populations. Although repeated emergence of similar phenotypes has been documented in many species in nature (Levin, 2001), we still have only limited knowledge, especially in plants, about the extent and mechanisms governing parallelism at the genomic level. The study of parallel evolution is an emerging and rapidly developing field of evolutionary genomics because cases of parallel adaptation provide an ideal experimental set-up for testing if the same variants, genes or functions were repeatedly selected (Stern, 2013).

Independent colonization of challenging habitats is promoted by the island-like distribution, when populations from ecologically contrasting environments, which can eventually evolve into ecotypes, repeatedly occur in close proximity over multiple sites in the landscape (Rundle and Nosil, 2005). A first, although indirect, evidence of parallel evolution is when populations cluster according to geography but not ecology – i.e. a sister position of populations from contrasting environments within the same geographic region inferred from selectively neutral genome-wide markers (Fig. 1A; Quesada et al., 2007; Roda et al., 2013; Butlin et al., 2014; James, Arenas-Castro, et al., 2021). In the single origin scenario, we expect a sister position of populations from the same environments among geographic regions as both arose once from the ancestral population, which was followed by range expansion (Fig. 1B). These two scenarios can be hard to distinguish if high gene flow occurs between ecotypes within geographic regions (Coyne and Orr, 2004). Therefore, distinguishing between these scenarios in a formalized, testable way, as well as quantifying gene flow between ecotypes is a prerequisite for a better understanding of the dynamics of parallel evolution in nature. Relying on the fact that challenging habitats in each geographic region were colonized by different genetic groups, we can conclude that parallel adaptation arose independently via similar selection pressures.

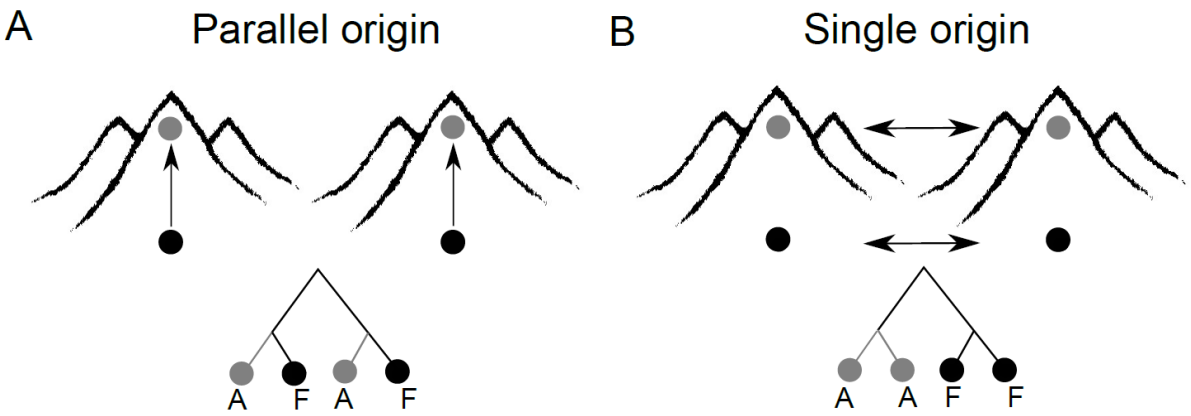


Fig. 1. Evolutionary scenarios of colonization of island-like habitats illustrated by an alpine environment. A) Parallel origin scenario of independent colonization of alpine habitats (A) from foothill populations (F) in each geographic region. B) Single origin of alpine populations followed by migration among alpine habitats across geographical regions. Note: grey dots = alpine populations; black dots = foothill populations; arrows = colonization events.

Repeated emergence of similar ecotypes due to ecological divergence was documented in some iconic examples from the animal kingdom (Fig. 2). A burst of genomic studies in the last decade showed parallel evolution as a more frequent phenomenon than was expected (Wood et al., 2005). Besides the well-known examples of parallel evolution from freshwater vs. marine ecotypes in threespine stickleback fish *Gasterosteus aculeatus* (Colosimo et al., 2005; Jones et al., 2012), crab vs. wave adapted ecotypes in marine snail *Littorina saxatilis* (Butlin et al., 2014), ecotypes adapted to different host plant species in stick insect *Timema cristinae* (Soria-Carrasco et al., 2014; Nosil et al., 2018), there are many more recent studies comprehensively showing parallel evolution in lake-stream cichlid fish *Astatotilapia burtoni* (Weber et al., 2021), saltmarsh beetle *Pogonus chalceus* adapted to different water regimes (Van Belleghem et al., 2018) or coastal vs. offshore bottlenose dolphin *Tursiops truncatus* (Louis et al., 2021). In contrast, well-documented examples of parallel evolution in plants are still rare. Parallel evolution has been comprehensively documented with genetic data for high- and low- elevation *Heliosperma pusillum* (Trucchi et al., 2017; Szukala et al., 2021), *Zea mays* (Fustier et al., 2017; Wang et al., 2021), and *Populus trichocarpa* (Holliday et al., 2016), annual vs. perennial ecotypes of *Oryza* (Cai et al., 2019), dune vs. non-dune *Helianthus petiolaris* (Andrew et al., 2012) and *Senecio lautus* (Roda et al., 2013; James, Wilkinson, et al., 2021), serpentine vs. non-serpentine *Solidago virgaurea* (Sakaguchi et al., 2017) and *Cerastium alpinum* (Berglund et al., 2004) or coastal vs. mine *Silene uniflora* (Papadopulos et al., 2021) (Fig. 2). Beyond that, we generally have only limited knowledge about the extent of phenotypic and genomic parallelism even in most of the above-mentioned examples and, importantly, about its mechanisms and evolutionary sources. In particular, there is a lack of such parallel examples in genetically well-tractable model species which allow more in-depth analyses and comprehensive interpretations.

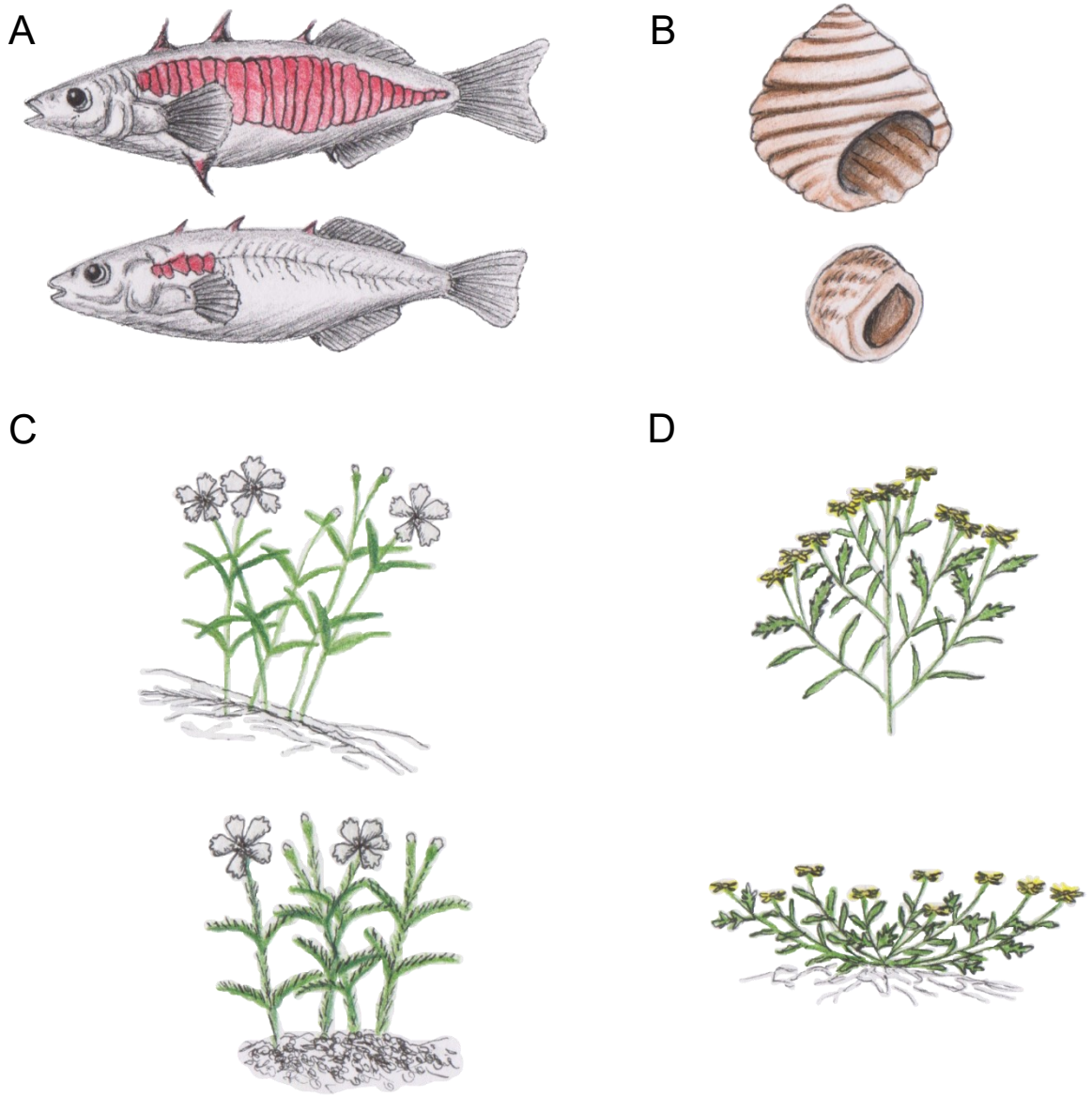


Fig. 2. Repeated emergence of similar ecotypes in selected model species for parallel evolution. A) marine vs. freshwater ecotypes in threespine stickleback fish *Gasterosteus aculeatus*. B) wave vs. crab ecotypes in marine snail *Littorina saxatilis*. C) alpine vs. montane ecotypes in *Heliosperma pusillum*. D) dune vs. headland ecotypes in *Senecio lautus*. Illustrations: V. Konečná

1.1 Varying extent of phenotypic and genomic parallelism

Although examples of particular parallel traits are ubiquitous in evolutionary literature, only few studies, mainly on animals quantified parallelism comprehensively across multiple morphological, life history or behavioural traits (but see James, Wilkinson et al., 2021). In general, individual phenotypes are more frequently measured in fully developed individuals, which prevents the exploration of potential phenotypic parallelism even in juvenile stages and further investigation of parallel trait development and dynamics over time. For instance, one of the most iconic parallel phenotypes evolved in threespine stickleback fish in freshwater populations with reduced defensive armour lateral plates (Colosimo et al., 2005) (Fig. 2). However, Stuart et al. (2017) showed that the direction and magnitude of phenotypic divergence in lake-stream threespine stickleback are not fully parallel among the population pairs.

Cases of repeatedly evolved phenotypes provide ideal systems for studying the genetic basis of parallel evolution (Elmer and Meyer, 2011; Stern, 2013). Selection on different biological levels from variants and genes up to the functions can lead to similar phenotypes (Stern, 2013). There are two major genetic mechanisms underlying parallel phenotypes. Firstly, the same locus can be under selection repeatedly in different populations, which may underlie parallel phenotypes. Repeated selection on the *Eda* locus in freshwater threespine stickleback fish (Colosimo et al., 2005) or on flowering time gene *FRI* in *Arabidopsis thaliana* (Korves et al., 2007) are examples of parallelism by genes. Secondly, repeated selection on different genes belonging to the same functional pathway in different populations – functional parallelism – may happen. This is a pervasive mechanism found generally across the majority of studies investigating the genomic basis of parallelism, e.g. in auxin hormone pathway leading to the selection of multiple genes related to shoot gravitropism in *Senecio lautus* (James, Allsopp, et al., 2021), in functions responsible for trichome development in *Heliosperma pusillum* (Szukala et al., 2021), or in response to heavy metal stress in *Silene uniflora* (Papadopulos et al., 2021).

1.2 Causes of non-parallel evolution

Independently evolved populations are rarely identical, representing rather a non-parallel continuum (Stuart et al., 2017; Bolnick et al., 2018). Deviations from parallel response may reflect a complex interplay of demographic history, population age, genetic background – genetic architecture of selected traits and availability of different pools of initial variation, genetic drift, local environmental heterogeneity, and of varying intensity of gene flow (Nosil and Crespi, 2004; Kolbe et al., 2012; Leinonen et al., 2013; Lucek et al., 2014; Rosenblum et al., 2014; Fraïsse and Welch, 2019; James, Arenas-Castro, et al., 2021). Therefore, both deterministic and stochastic processes can lead to non-parallel patterns identified frequently in systems of parallel evolution (Bolnick et al., 2018). Furthermore, phenotypic similarity across populations can also reflect short-term phenotypic plasticity, non-heritable differences in reaction to distinct environments, which may facilitate colonization of new habitats (Corl et al., 2018; Szukala et al., 2021). The combination of neutral and selective processes mentioned above can cause non-parallel deviations, which have been recently identified even in many iconic examples of parallel evolution animals such as threespine stickleback (Stuart et al., 2017; Magalhaes et al., 2021), salmonid fish *Salvelinus alpinus* (Jacobs et al., 2020), guppy *Poecilia reticulata* (Whiting et al., 2021), cichlid (Weber et al., 2021), marine snail (Morales et al., 2019), and songbird *Parus major* (Salmón et al., 2021), and also in plants – *Arabidopsis arenosa* or *Senecio lautus* (Knotek et al., 2020; Bohutínská, Vlček, et al., 2021; James, Wilkinson, et al., 2021).

Environmental heterogeneity poses multiple challenges, which can be enhanced or diminished by selection. Specifically, non-parallelism between similar environments can arise from other environmental factors – environmental constraints – across environmental gradients if they exist. We suppose that the extent of parallelism covaries with the levels of environmental heterogeneity as was shown e.g. in threespine stickleback (Stuart et al., 2017). Importantly, the choice of measured environmental factors describing the complex challenging environments has a large impact on the quantification of parallelism (Stuart et al., 2017; Langerhans, 2018).

Besides environmental constraints, also genetic constraints can lead to non-parallelism. For instance, the genetic architecture of traits, including the size of genetic effects underlying adaptive phenotypes, number of genomic regions, recombination rate, degree of genetic redundancy, or pleiotropy (when one gene is affecting multiple phenotypes (Hoban et al., 2016)), has a large impact on the extent of non-parallelism (Láruson et al., 2020). Genetic constraint, specifically genetic redundancy, i.e. mutations in different genes (possibly within

the same pathway), has been observed for example in colouration. The light/dark colour is achieved in the majority of animals including mice or lizards by changes at *Mc1r* gene (Manceau et al., 2010; Rosenblum et al., 2014). However, in some cases such as *Peromyscus polionotus* mice, similar light colouration evolved by multiple genetic solutions in Atlantic and Gulf beach lineages within the same species (Hoekstra et al., 2006; Steiner et al., 2009). Mutations in *Mc1r* associated with light colour are absent in Atlantic beach mice (Hoekstra et al., 2006), but instead, the light *Agouti* gene allele is fixed in both Atlantic and Gulf Coasts. This illustrates an example of non-parallelism in *Mc1r* gene due to genetic redundancy in light colouration.

Most of the adaptations have a polygenic genetic basis which promotes the evolution of multiple alternative routes to adaptation (Barghi et al., 2020). It can result in non-parallelism at variant- and gene-level as similar phenotypes evolved via changes at many different variants and genes within the same pathway or gene expression (Hermisson and Pennings, 2017; Höllinger et al., 2019; Barghi et al., 2020). Consequently, in systems with the polygenic architecture of genetic traits, there is a higher chance of functional parallelism (Yeaman, 2015). Prevailing non-parallelism at the level of variants and genes, but high functional parallelism has been experimentally shown e.g. in *Drosophila simulans* in temperature-mediated laboratory conditions (Barghi et al., 2019) and by sequencing studies of natural populations of *Heliosperma pusillum* (Szukala et al., 2021), *Senecio lautus* (James, Allsopp, et al., 2021), and *Silene uniflora* (Papadopulos et al., 2021). This is in congruence with functional redundancy, where at least two genes encode the same biochemical function (Láruson et al., 2020).

2 Evolutionary sources of genomic parallelism

Even if the same gene underlies parallelism, the adaptive allele(s) may originate from different evolutionary sources (Fig. 3). Firstly, repeated selection can act on standing genetic variation that is shared in the ancestral (non-adapted) populations including the potentially advantageous variants. When different populations are exposed to similar pressure, such advantageous variants will independently increase in their frequency (Barrett and Schluter, 2008). Secondly, gene flow among multiple adapted populations or adaptive introgression among species can promote sharing of adaptive alleles (Hufford et al., 2013). Both these scenarios are based on the presence of a pre-existing variation with shared origin as they operate on alleles of a single mutational origin, in contrast to the third, selection of independent de novo mutations targeting the same genomic regions (Stern, 2013). Different taxa or different populations within the same species can converge to similar solutions, although they can use various evolutionary sources. These are dependent mainly on the divergence among genetic lineages and on the intensity of gene flow (Bohutínská, Vlček, et al., 2021; Waters and McCulloch, 2021). The relative contribution of pre-existing versus de novo genomic variation to parallel adaptation is still poorly understood.

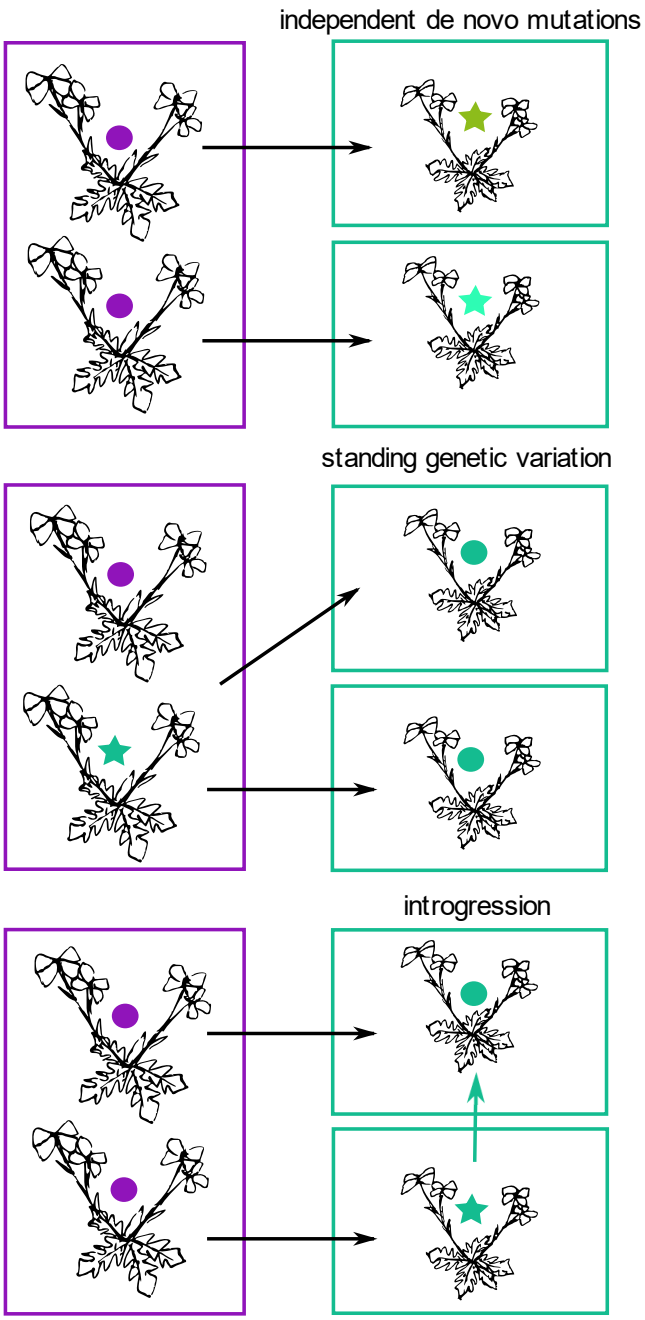


Fig. 3. Different sources of genetic variation in parallel evolution. Schematic illustrations show repeated adaptation to new challenging environments (in green frames) via independent de novo mutations, selection on standing genetic variation or adaptive introgression. The last two processes together operate on a shared pool of pre-existing variation. Note: dot = genetic variant; asterisk = origin of a new adaptive variant; arrow = colonization and adaptation event; green arrow = gene flow. Inspired by Waters and McCulloch (2021). Illustrations: V. Konečná

A recent synthesis of genomic studies by Waters and McCulloch (2021) showed that repeated sorting of standing genetic variation is a prevailing source for parallel evolution. When a large pool of pre-existing variation, inherited from the common ancestor, is available, selection on such standing variants may be particularly beneficial under intense selection during rapid adaptation as it saves the time otherwise needed for the emergence of novel adaptive mutations (Barrett and Schluter, 2008; Conte et al., 2012). In addition, standing genetic variation already exists in the genomic environment for some time before adaptation, thus negative pleiotropic effects of linked variants are likely to be minimized (Hermisson and Pennings, 2005; Przeworski et al., 2005; Van Etten et al., 2020). One of the classical examples comes from beach or deer mice, *Peromyscus* (Hoekstra et al., 2006; Steiner et al., 2009; Barrett et al., 2019). Specifically, the allele of the regulatory element in the *Agouti* gene associated with light pigmentation has been independently selected from standing genetic variation in both beach *Peromyscus polionotus* mice lineages from the Atlantic and Gulf Coast. This allele, which is fixed in both beach populations, is varying in frequency in mainland populations. Similarly, in the case of threespine stickleback fishes, the same *Eda* alleles were repeatedly selected in freshwater populations from ancestral marine populations where these alleles are present in low frequency (Colosimo et al., 2005; Jones et al., 2012; Magalhaes et al., 2021). Other examples come from copper mine populations of *Mimulus guttatus* (Lee and Coop, 2017), morning glory *Ipomoea purpurea* (Van Etten et al., 2020), black grass *Alopecurus myosuroides* (Kersten et al., 2021), saltmarsh beetle (Van Belleghem et al., 2018), vinous-throated parrotbill bird *Sinisuthora webbiana* (Lai et al., 2019), Asian honey bee *Apis cerana* (Ji et al., 2020), and fish *Coilia nasus* (Zong et al., 2020). In some species, the selection can act on relative ancient polymorphisms – ancestral genetic variants fluctuating in populations as

balanced polymorphisms for millions of years (Nelson and Cresko, 2018; Louis et al., 2021; Urban et al., 2021; Weber et al., 2021).

Another solution for adaptation to rapid environmental change specifically between hybridizing lineages is sharing the variants via gene flow followed by introgression. Introgression can be an important mechanism for spreading variants identical by descent (Meier et al., 2017; Svardal et al., 2020; Todesco et al., 2020). It has been shown in altitudinal adaptation and mimetic wing patterns in *Heliconius* butterfly (e.g. Consortium et al., 2012; Montejo-Kovacevich et al., 2021), highland adaptation in maize *Zea mays* (Wang et al., 2021), meiotic adaptation in plant *Arabidopsis arenosa* and *lyrata* (Marburger et al., 2019), winter colour coat variation in hare *Lepus timidus* (Giska et al., 2019), and anthropogenic adaptation to pesticides in mice *Mus musculus domesticus* (Song et al., 2011) and water pollution in fish *Fundulus grandis* (Oziolor et al., 2019).

Finally, in contrast to quickly growing literature on shared variants, empirical evidence for de novo mutations, which are not identical by descent, is rather rare especially among closely related genetic lineages. Such disproportion is logical when taking into account the time that novel mutations need to evolve and spread in comparison with standing genetic variation, which is immediately available and can be quickly reassembled into new combinations by recombination (Barrett and Schluter, 2008). Thus, de novo mutations are much more frequent in convergent evolution among distantly related lineages, for instance in the pigmentation gene *Mc1r* across animals (Hoekstra et al., 2006; Manceau et al., 2010; Stern, 2013; Rosenblum et al., 2014). However, emerging genomic studies revealed some exceptions with examples of de novo mutations even from recently diverged lineages (~10.000 years ago) and possibly also from the same amino acid (Waters and McCulloch, 2021). Such an example comes from rhodopsin gene evolution in fishes, this gene encodes receptors essential for vision in dim light which is necessary for the transition from marine to brackish or freshwater environments (Hill et al., 2019). Further, strong anthropogenic pressure can lead to novel solutions e.g. in cryptic colouration in stick insect *Timema* (Villoutreix et al., 2020), loss of song in Hawaiian crickets *Teleogryllus* (Zhang et al., 2021), or herbicide resistance in *Amaranthus tuberculatus* (Kreiner et al., 2019). The higher chance for the origin of de novo mutations is for example in genomic regions with high mutation rates, which was documented for highly-mutable enhancer region of *Pitx1* gene responsible for a pelvic fin loss in separate freshwater populations of stickleback fishes (Chan et al., 2010; Xie et al., 2019). However, rapid modifications via de novo mutations are possible even in highly conserved molecular pathways e.g. associated with essential mitochondrial processes mediating the tolerance to hydrogen sulfide in fish *Poecilia mexicana* (Greenway et al., 2020) or meiosis in *Arabidopsis arenosa* (Bohutínská, Vlček, et al., 2021). Another explanation can be related to the functional basis of adaptive traits, for instance loss of certain traits can be achieved by a simpler genetic change compared to gain of novel functional traits (Zhang et al., 2021).

Repeated evolution leading to similar phenotypes via distinct de novo mutations is the primary explanation for convergent evolution among distantly related lineages (reviewed in Waters and McCulloch, 2021). In contrast, sorting from pre-existing genetic variation is the main evolutionary driver of parallel evolution as closely related genetic lineages share ancestral genetic variation (Bohutínská, Vlček, et al., 2021). However, importantly, for some parallel multigenic traits, the combination of both evolutionary sources – de novo genomic variation and selection on pre-existing genetic variation – seems to be more frequent than previously thought even among close genetic lineages and small geographic scales (Waters and McCulloch, 2021). Such examples were found in plant *Arabidopsis arenosa* (Bohutínská, Handrick, et al., 2021; Bohutínská, Vlček, et al., 2021; Konečná et al., 2021), fish *Poecilia mexicana* (Greenway et al., 2020), and threespine stickleback (Jones et al., 2012; Xie et al., 2019; Magalhaes et al., 2021).

3 Selected study systems of parallel evolution

For establishing new systems and addressing the mechanisms and evolutionary sources of parallel evolution in plants I have chosen two extreme environments providing multiple challenges to plant life – challenging high-elevation alpine habitats and toxic serpentine substrates. The island-like distribution of alpine and serpentine habitats provides natural replicates for testing hypotheses of their independent colonizations and quantifying the extent of phenotypic and genomic parallelism in adapted populations. Such a heterogeneous landscape mosaic can trigger local adaptations, fitness advantage of population at its home site, especially in the presence of steep gradients (Jain and Bradshaw, 1966; Kawecki and Ebert, 2004).

3.1 Alpine environment and examples of parallel evolution

High elevations pose a spectrum of challenges such as freezing, fluctuating temperatures, strong winds, increased UV radiation, low partial pressure of carbon dioxide or lack of pollinators triggering various selection pressures (Körner, 2003). Such pressures can lead to distinct phenotypic changes e.g. contracted rosette plants, dense cushions, and large flowers. The island-like distribution of alpine habitats promotes parallel colonization from geographically proximal foothill populations (Levin, 2001). The further restriction of gene flow between lower and higher elevations as well as the reduction or absence of gene flow among mountain ridges can lead to the accumulation of reproductive barriers and/or hybrid incompatibilities (Funk et al., 2005).

Although the mountains provide an ideal model system for studying the evolution of recurrently adapted ecotypes, studies documenting the parallel colonization of alpine habitats within species level based on genetic data addressing the evolutionary mechanisms of adaptation with a combination of experimental data for quantification of phenotypic parallelism are still extremely rare. Parallel evolution of alpine ecotypes was documented in *Arabidopsis halleri* (Kubota et al., 2015; Šrámková-Fuxová et al., 2017; Bohutínská, Vlček, et al., 2021), *Arabidopsis lyrata* (Hämälä et al., 2018), maize (Takuno et al., 2015; Fustier et al., 2017; Wang et al., 2021), and *Populus trichocarpa* (Holliday et al., 2016). The opposite direction of colonization – from alpine to foothill (montane) habitats has been documented from *Heliosperma pusillum* (Trucchi et al., 2017). The alternative scenario of a single origin of alpine ecotype followed by dispersal has also been documented, for example in *Senecio halleri* single lineage colonized the Alps in a stepwise manner (Bettin et al., 2007).

Such naturally replicated cases of parallel evolution allow us also to test how elevation differentiation affects the structuring of the genetic diversity. The influence of habitat differentiation on the genetic diversity of populations has been previously studied mainly at lower elevations (Jacquemyn et al., 2004; Odat et al., 2004; Ortego et al., 2012). It is hypothesized that alpine populations would be genetically depauperate relative to lowland populations due to the effects of genetic bottleneck, genetic drift, reduced gene flow, and habitat fragmentation (Young et al., 1996). This has indeed been observed in alpine populations of *Arabidopsis thaliana* from the Italian Alps, which exhibit reduced genetic diversity relative to foothill populations (Günther et al., 2016), but in contrast, alpine populations of *A. thaliana* or *Papaver occidentale* from the Swiss Alps show no evidence of reduced genetic diversity (Luo et al., 2015; Pittet et al., 2020). In *Primula merrilliana* from eastern China, alpine populations exhibit even higher genetic diversity, because of large population size and inter-connection by gene flow (Shao et al., 2015). Therefore, we need more studies of parallel evolution to test the generality of how alpine conditions shape population genetic diversity.

3.2 Serpentine soils and examples of parallel evolution

Serpentine soils represent a powerful model for studying multi-challenge rapid adaptation (Rajakaruna, 2018), in contrast to alpine habitats, which possess a very complex set of challenges, they are rather of intermediate complexity. Their extreme and distinct chemical and physical properties act as strong, but tractable selective pressures for quantification (Konečná et al., 2020). Serpentine soils are derived from ultramafic rocks and are defined mainly by a highly skewed elemental content: (i) low content of nutrients such as calcium (Ca), nitrogen (N), potassium (K), and phosphorus (P), (ii) elevated levels of heavy metals e.g. chromium (Cr), cobalt (Co), and nickel (Ni) and (iii) extremely reduced Ca/Mg ratio (far below one on serpentines vs. well-above one in other soils). Additionally, serpentine soils are highly porous, prone to drought and erosions (Brady et al., 2005). As a result of these challenges, serpentines are characterised by poor plant productivity, a high rate of endemism, and different vegetation types (Whittaker, 1954; Brady et al., 2005; Harrison and Rajakaruna, 2011). Consequently, serpentines act as a strong filter on plant diversity, excluding most of the species occurring in the neighbouring areas and triggering a strong local adaptive response in the species that managed to pass this filter.

Due to extremely strong responses triggered by the serpentines and decades of intense experimental and evolutionary research, serpentines represent one of the best-described model systems for the study of local adaptation in plants. Since the classical textbook studies in the 50's (e.g. Kruckeberg, 1950, 1951, 1967) numerous experimental studies have demonstrated ubiquitous strong local adaptation towards the challenging substrate and revealed particular life-history and (eco-)physiological mechanisms standing behind these adaptations (Brady et al., 2005; Harrison and Rajakaruna, 2011; O'Dell and Rajakaruna, 2011; Rajakaruna, 2018). For instance, drought stress adaptations include slower growth rate, higher root/shoot biomass ratios, earlier flowering, modification of flowers and reduced leaf sizes (Brady et al., 2005; O'Dell and Rajakaruna, 2011; Von Wettberg et al., 2014). In terms of chemistry, adaptations can involve selective uptake of some micro/macronutrients such as Ca and exclusion/regulated accumulation/storage of heavy metals (Brady et al., 2005; Kazakou et al., 2008).

The island distribution of serpentines provides natural replicates for studying possible independent colonizations of serpentine barrens. The parallel evolution was documented e.g. in *Alyssum serpyllifolium* (Mengoni et al., 2003; Sobczyk et al., 2017), *Cerastium alpinum* (Berglund et al., 2004), *Lasthenia californica* complex (Rajakaruna et al., 2003), *Mimulus guttatus* (Selby and Willis, 2018), *Solidago virgaurea* (Sakaguchi et al., 2017), and *Streptanthus glandulosus* complex (Mayer and Soltis, 1994). In fact, single origins of serpentine populations are very rarely documented among species growing both on and off multiple serpentines (e.g. *Picris hieracioides*, Sakaguchi et al., 2018), and remains rather a property of genuine serpentine endemics (e.g., *Knautia serpentinicola*, Kolář et al., 2012; and *Halacsya sendtneri*, Cecchi and Selvi, 2009).

In contrast to decades of intense experimental research, the genetic basis of serpentine parallel adaptations is still poorly known. The exception is for example *Mimulus guttatus* in North America, where two populations shared the major quantitative trait loci (QTL) containing genes related to the transport of heavy metals and metal binding (Selby and Willis, 2018). Indirect evidence has been also provided by comparing allelic frequencies at particular loci, e.g. *TPC1,* suggesting genetic parallelism in European and North American serpentine populations of *Arabidopsis lyrata* (Turner et al., 2010). The genetic convergence was described between *A. arenosa* and *A. lyrata* by Arnold et al. (2016) in nine gene coding loci with serpentine relevant predicted functions, such as Ca, K, and Ni homeostasis. Specifically, Arnold et al. (2016) identified almost 200 differentiated loci between serpentine and non-serpentine populations with top candidates for selection including genes related to ion homeostasis traits

– heavy metal transporters, root macronutrient transporters, and dehydration tolerance genes suggesting a polygenic basis of serpentine adaptation. To get a complete view, we have to test in multiple populations if candidate genes for selection are associated with a fitness effect, and thus understand their true adaptive value.

An intriguing open question concerns the origin of selected variants in different serpentine populations. Lastly, with the growing number of studies on parallel serpentine adaptation in various species, we will be also able to identify convergent loci mediating serpentine adaptation across species (see a summary in Konečná et al., 2020).

4 Objectives

The main aim of the thesis was to establish novel systems for studying the evolutionary drivers and consequences of parallel evolution in plants and use their potential to uncover evolutionary mechanisms shaping genomic parallelism. I asked those general evolutionary questions:

Is plant evolution in alpine and serpentine environments repeatable?

If so, to what extent is parallelism manifested at phenotypic and genomic level?

What are the main evolutionary sources of adaptive genetic variation during rapid adaptation?

What is the extent of phenotypic non-parallelism in recently diverged parallel ecotypes?

I addressed three systems facing two strong selective environments – alpine stands and toxic serpentine soil. I brought new evidence for parallel colonization of (sub)alpine habitats by *Arabidopsis arenosa* and *Primula elatior* and serpentine soils by *A. arenosa*. Further, I leveraged the cases of naturally replicated parallel adaptation in *A. arenosa* to study the extent of phenotypic and genomic parallelism together with evolutionary sources of its genetic basis. Finally, I also elucidated the causes underlying non-parallel evolution of certain traits. I addressed the following specific objectives:

Establishing novel systems of parallel evolution in plants

a) What is known about the frequency and genomic architecture of parallel plant adaptation to serpentine soil (CS1)?
b) Were challenging alpine and serpentine environments colonized independently in each geographic region? (CS2, CS3, CS4)
c) How does elevational differentiation affect population genetic diversity and differentiation? (CS2, CS3)
d) Does the elevation difference act as a barrier to gene flow between foothill and subalpine populations? (CS3)

Quantifying the extent of phenotypic parallelism

e) Is parallel colonisation of challenging environments associated with parallel fitness response? (CS4, CS5)
f) Does parallel environmental differentiation cause the same heritable changes in phenotype over independent environmental transitions? (CS2, CS5)
g) To which extent is the direction and magnitude of phenotypic response different among independently adapted populations? (CS2, CS5)

Quantifying the extent of genomic parallelism and identifying its evolutionary sources

h) How frequently does evolution repeat itself at the genomic level? (CS1, CS4, CS5)
i) Is selection on shared variation a major source of the parallel adaptive genetic variation within a species? (CS4)
j) Is adaptation via convergent de novo alleles even feasible in recently diverged plant populations? (CS4)

Elucidating the causes of non-parallelism

k) To which extent do independent colonizers of similar environments exhibit non-parallel phenotypes? (CS2, CS5)
l) Does the extent of phenotypic parallelism correspond to gene-level parallelism, or does the genetic redundancy lead to limited gene parallelism? (CS5)

5 Methods

5.1 Model species *Arabidopsis arenosa*

Arabidopsis arenosa is an established model species for studying drivers and consequences of adaptive evolution in a genomic and molecular genetic context. This perennial outcrosser encompassing large genetically diverse diploid and autotetraploid populations (Kolář et al., 2016; Yant and Bomblies, 2017; Monnahan et al., 2019), which are widespread at low and mid elevations (up to ~1000 m a.s.l.), but scattered in serpentine sites in Central Europe and treeless alpine habitats (above ~1500 m a.s.l.) in four mountain ranges – Eastern Alps and Eastern, Southern, and Western Carpathians. Except for the Western Carpathians occupied by both diploids and tetraploids (diploids in Vysoké Tatry Mts. and tetraploids in Nízké and Západné Tatry Mts.), the alpine zone of the other three mountain ranges was colonized only by tetraploids. Serpentine habitats in Central Europe were colonized solely by tetraploids.

Tetraploid *A. arenosa* cytotype originated through natural whole-genome duplication from a single diploid lineage in Western Carpathians ~19-31k generations ago (Monnahan et al., 2019). *Arabidopsis arenosa* occurs on islands of environmentally suitable conditions represented by open rocky outcrops, screes, and steep slopes typical for heliophilous species; these spatially discrete areas are surrounded by forests or dense grasslands. There is a clear distribution gap between alpine and foothill ecotypes spanning at least 500 m of elevation corresponding with timberline and a sharp boundary between serpentine and siliceous/neutral soils with no intermediate habitats.

Arabidopsis arenosa is a genetically well tractable species because of its easy cultivation, rapid life-cycles, the availability of close well-annotated reference genome (*A. lyrata* (Hu et al., 2011)), and small genome size (~200 Mb) (Hollister et al., 2012; Yant and Bomblies, 2017). Moreover, tetraploids harbour increased genome-wide adaptive diversity (Monnahan et al., 2019) and currently occupy a broader ecological niche than their diploid sisters (Molina-Henao and Hopkins, 2019), including serpentine habitats (Arnold et al., 2016), railway lines (Baduel et al., 2016; Baduel, Hunter, et al., 2018), and contaminated mine tailings (Przedpełska and Wierzbicka, 2007; Preite et al., 2019). This makes *A. arenosa* a promising model for empirical enquiries of adaptation in natural conditions (Yant and Bomblies, 2017; Monnahan et al., 2019).

5.2 Model species *Primula elatior*

Primula elatior is a promising model for studying the genetic consequences of elevational differentiation. This perennial outcrossing species has broad ecological preferences including a large elevational range in several mountain ranges in Europe (Valentine and Kress, 1972). This considerable ecological breadth is linked to high morphological variation primarily in leaf and calyx shape (Valentine and Kress, 1972; Richards, 2003). In contrast to morphological investigations, the genetic structure of *P. elatior* remains unknown except for studies of lowland populations at fine-scale in Belgium (Jacquemyn et al., 2002, 2004, 2008; Van Rossum and Triest, 2006). In mountain ranges in Central Europe, *P. elatior* grows along an elevational gradient – from foothill meadows, riverbanks, and forest edges, up to subalpine meadows, snow beds, and rocky outcrops in glacial cirques. Specifically, in Krkonoše and Jeseníky mountains, the subalpine populations are restricted to specific microhabitats within several glacial cirques. Generally, the difference between foothill and subalpine populations across mountain ranges is defined by treeline; foothill populations occur below the treeline in contrast to subalpine populations. However, the distributional gap along the elevational gradient is highly variable (~300 m) across mountain ranges (V. Konečná, personal observations).

5.3 Approaches

I combined various population genetic and experimental approaches in my studies of parallel evolution. I collected widespread foothill and non-serpentine populations, which colonized scattered (sub)alpine and serpentine habitats. I further conducted experimental cultivations, inferred the evolutionary history of populations including reconstruction of genetic structure, identified candidate genes under directional selection followed by validation using gene ontology enrichment, overlapping replicates of adaptation, statistical modelling and experiments.

Field sampling included seeds for cultivations, soils for elemental analyses and cultivations, leaf samples for ploidy determination, genetic analysis, elemental analyses, and vegetation samples. In CS2, we conducted a common garden experiment to test the persistence of distinct vegetative and floral traits of alpine plants in foothill-like conditions and to determine parallel vs. non-parallel traits. Using a similar approach as Stuart et al. (2017), the large ecotypic effect of a particular trait was an indication of parallelism while the strong contribution of ecotype × geographic region (population pair) interaction implied non-parallel differentiation in the particular trait. In CS4 and CS5, I conducted reciprocal transplant experiments for serpentine and non-serpentine plants from three population pairs. I cultivated them in the soil from their original and foreign sites and measured life-history traits such as fitness proxies, leaf elemental concentrations, and root growth to test for local substrate adaptation. Further, focusing only on serpentine challenging soils I used the same approach mentioned above for the quantification of phenotypic parallelism.

To explore the genetic structure and diversity of populations, I used complementary approaches. In CS3, I inferred genetic structure by genotyping of nuclear microsatellite loci, in CS2 I used single nucleotide polymorphism (SNP) data from RAD sequencing approach

together with genome-wide resequencing data, which I also used in CS4 and CS5. In both cases, sequences were mapped to the annotated reference genome of *A. lyrata* following Monnahan et al. (2019). I explored evolutionary history among populations by allele frequency covariance graphs, Bayesian clustering, distance networks, and ordinations. I calculated genetic diversities of populations using various genetic indices and analysis of molecular variance. Further, I specifically tested for parallel origin of alpine and serpentine populations in CS2 and CS4 using coalescent simulations in fastsimcoal (Excoffier and Foll, 2011). Briefly, I compared the fit of the data with two competing topologies: (i) parallel origin – sister position of populations from contrasting habitats from the same geographic region and (ii) single origin – sister position of populations from distinct regions but belonging to the same ecotype. For each topology, I either assumed or not assumed between ecotype gene flow. Further, in CS4 I estimated divergence times for each pair of serpentine and non-serpentine populations from the same geographic region. In CS3, I identified the direction and extent of gene flow between foothill and subalpine populations in each mountain range using a coalescent framework in Migrate-n (Beerli and Palczewski, 2010).

I leveraged five-fold natural replicates of serpentine adaptation in CS4 to identify candidate genes that show repeated footprints of selection. I combined two approaches for detection of selection – divergence scans and environmental association analysis to refine the list of loci for parallel selection modelling. I identified genes exhibiting excessive differentiation between paired populations using 1 % outlier F_{ST} window-based scans. I also annotated these genes into biological processes, molecular pathways, and cellular components using gene ontology enrichment analysis (Alexa and Rahnenfuhrer, 2020) and explored predictions of protein-protein interactions (Szklarczyk et al., 2015). Further, I overlapped these candidate gene lists across population pairs and identified parallel differentiation candidates that represent divergence outliers in at least two population pairs. Next, I inferred candidates directly associated with individual soil elemental concentrations by performing environmental association analysis using latent factor mixed models (LFMM; Caye et al., 2019). Finally, I overlapped LFMM candidates with parallel differentiation candidates to produce a final refined list of serpentine adaptation candidates. Next, to infer the sources of adaptive variation in each serpentine adaptation candidate gene, I modelled allele frequency covariance around repeatedly selected sites and identified the most likely evolutionary scenario using a designated "Distinguishing among Modes of Convergence" (DMC) approach (Lee and Coop, 2017). As DMC does not provide fine-scale information about the distribution of sequence variation at particular alleles, I further investigated the candidate alleles of *TPC1* gene, the only candidate with de novo mutations in independent serpentine populations. I investigated natural sequence variation in this locus from additional Sanger sequencing and short-read data available for *A. arenosa* from its whole geographic range. To investigate potential functional impact we performed structural homology modelling of selected alleles using crystallographically determined structures as a template.

6 Key results

I studied the nature of parallel evolution, which is a central component of the study of evolutionary repeatability and mechanisms of adaptation. Leveraging replicated cases of environmental transition in nature I showed how challenging environments structure genetic diversity among populations within a species. By further experimental and population genomic inquiry in one such system, I asked to what extent it results in similar solutions when populations are facing the same environmental challenges. By combining population genetic and experimental approaches I documented the complex interplay of adaptive, historical and ecological processes in parallel evolution.

6.1 Establishing novel systems of parallel evolution in plants

I brought the evidence for parallel colonization of (sub)alpine and serpentine habitats by genetically distinct lineages of *A. arenosa* (CS2, CS4) and *P. elatior* (CS3). In CS2, we used SNP genotyping of 200 individuals from 29 alpine and adjacent 30 foothill populations to reveal repeated colonization of alpine habitats in geographically distinct mountain regions of the Eastern Alps, Eastern, Southern, and Western Carpathians. I confirmed independent colonization of alpine habitats within each mountain region by coalescent simulations and genome resequencing data. In CS3, by genotyping nuclear microsatellite loci in 202 individuals from eight foothills and adjacent eight subalpine populations, I revealed three genetic groups corresponding with the three studied geographic regions (Jeseníky, Krkonoše, and Western Carpathians). Thus, in both studies (CS2 and CS3), the occurrence of distinct populations in isolated high-elevation stands is a result of parallel colonization events of (sub)alpine habitats from multiple foothill gene pools. Furthermore, I found higher genetic differentiation among (sub)alpine populations than among foothill populations and overall high variation within geographic regions (CS2 and CS3). Importantly, genetic differentiation between foothill and (sub)alpine populations was non-significant within any of the mountain regions in both model species. This suggests that mountain ridges act as migration barriers reducing gene flow more strongly than elevational differences between foothill and (sub)alpine populations and/or recent divergence. Moreover, colonization of (sub)alpine habitats did not result in loss of genetic diversity, which is in congruence with high migration rates between foothill and subalpine populations in *Primula* (CS3) and/or presumably large colonizing populations in *Arabidopsis* (CS2).

I further focused on toxic serpentine substrates as a promising system for studying genomic parallelism in a literature review (CS1). The toxicity of serpentine soils imposes strong selection pressure and their island-like distribution provides convenient natural replicates to study repeated evolution. Although there are numerous experimental studies documenting adaptation to serpentines, only a handful of studies provide insights into the genomic basis of such adaptations (CS1). To address this gap, I reconstructed the evolutionary history of the *A. arenosa* system and then characterized parallel genome evolution and its drivers. Firstly, I compared parallel vs. single origin scenarios by coalescent simulations using genome-wide resequencing data of 78 individuals of *A. arenosa* from five serpentine and adjacent non-serpentine populations from Central Europe (CS4). Consistently across all pairwise iterations of serpentine – non-serpentine population pairs, the parallel origin scenario had the highest support. I also demonstrated low genetic differentiation between serpentine and non-serpentine populations, in congruence with likely recent postglacial colonizations (< 12.000 years), and lack of bottlenecks during substrate transition.

6.2 Quantifying the extent of phenotypic parallelism

Taking advantage of multiple natural replicates of alpine (CS2) and serpentine (CS5) ecotypes of *A. arenosa*, we quantified the magnitude of phenotypic parallelism and investigated its neutral and adaptive determinants. In CS2, we measured morphological traits of alpine and foothill *Arabidopsis* plants from natural populations. Although the vast majority of scored traits (14/16) significantly differentiated alpine and foothill ecotypes, the level of parallelism remarkably differed across traits (nine parallel vs. seven non-parallel traits). Parallel traits were mainly related to height and flowering: regardless of the geographic region, all alpine plants were shorter with larger flowers. This is in contrast to traits describing leaf size and shape, which were mostly non-parallel, i.e. geographic-region specific. Using a subset of alpine and foothill populations cultivated in homogeneous common garden conditions resembling foothill habitats, we revealed that phenotypic differentiation between ecotypes, except for some leaf traits demonstrating plasticity, remained stable and significant even after two generations of cultivation. In summary, reduced stem height and larger floral organs consistently showed parallel genetically determined components of phenotypic differentiation, even though its magnitude differed across geographic regions.

In CS5, I investigated this question in more detail and with a more diverse set of functional traits and asked to what extent phenotypic evolution is repeatable during rapid parallel edaphic adaptation to similar stressful environments across three population pairs of serpentine and adjacent non-serpentine populations. For this aim, I used a substrate-focused reciprocal transplant experiment as specific soil chemistry is the main defining factor of serpentine habitats (Brady et al., 2005; O'Dell and Rajakaruna, 2011). Firstly, I demonstrated rapid repeated adaptation to serpentine soils. I showed that populations of serpentine origin attain consistently higher cumulative fitness estimates when growing in their native soils than their originally non-serpentine counterparts, altogether providing strong evidence for substrate-driven local adaptation. I inspected the overall patterns of adaptation to show that they evolved in the same direction and with none or utmost limited trade-offs. However, the magnitude of adaptations differed in congruence with the level of genome-wide differentiation, i.e. population pair with the highest genetic differentiation had also the highest phenotypic difference. Secondly, to assess the extent of phenotypic parallelism, I analyzed the differences in individual life-history traits such as fitness proxies, plant ion uptake, and root growth. Almost all traits (11/13) significantly differentiated serpentine and non-serpentine ecotypes and phenotypic parallelism was pervasive (only three traits were non-parallel). Both direction and magnitude of phenotypic response were very similar regardless of population pair. Serpentine plants had larger root biomass and rosette area, higher number of additive rosettes, bigger seed mass, bolted earlier, had lower uptake of Ca and Mg and higher exclusion of heavy metals (Co, Cr, and Ni) from the shoots than non-serpentine plants when cultivated in serpentine soils. Overall, the response of serpentine populations to serpentine challenge led to similar phenotypic solutions.

6.3 Quantifying the extent of genomic parallelism and identifying its evolutionary sources

In the literature review (CS1) I identified serpentine soils as promising, yet underexplored natural models for studying the genomic basis of plant adaptation. I thus empirically addressed parallel genome adaptation in serpentine *A. arenosa* (CS4). Specifically, I leveraged five natural replicates of serpentine adaptation to seek the genomic basis and evolutionary source of the parallel candidate alleles. Combining divergence scans and environmental association analysis, I inferred a list of reliable selection candidate genes and functions. I detected significant

parallelism, at both gene and functional levels, over these five pairs and I found parallel candidate genes involved in ion homeostasis, inorganic anion transport, calcium transmembrane transporter activity and response to metal ions. For instance, the candidates included the *NRT2.1* and *NRT2.2* high-affinity nitrate transporters, *TPC1*, a central calcium channel, and potassium transporters *AKT5* and *KUP9*. Interestingly some of such genes were found also in the few previous studies (as summarized in CS1), and I identified candidates for convergent evolution of *A. arenosa* with *A. lyrata* (e.g. *KUP9* and *TPC1*), *Alyssum serpyllifolium* (*FPN2=IREG2*), and *Mimulus guttatus* (AT4G19670) representing potential evolutionary hotspots for serpentine adaptation. Interestingly, serpentine populations of these species come from different continents, which provides evidence for some similar selection pressures posed by temperate zone serpentine sites worldwide, leading to convergence at specific loci mediating serpentine adaptation.

Next, I inferred the evolutionary sources of the parallel adaptive variation in the top 61 serpentine adaptation candidate genes using statistical modelling. In line with theoretical expectations, I found that shared variation is a vastly prevalent source of parallel adaptive variants (in 97 % cases) across the highly variable and recently diverged tetraploid populations of *A. arenosa*. This is in congruence with large effective population size and polysomic masking of allelic variation in polyploids (Van De Peer et al., 2017; Baduel, Bray, et al., 2018). However, I also discovered an exceptional parallel locus candidate, *TPC1*, with parallel de novo mutations in a single codon in two distinct serpentine populations. We further used protein structure modelling to demonstrate the potential functional impact of this mutation on the conductance of divalent cations Ca^{2+} and Mg^{2+}. Overall, the study brings one of the first empirical examples for parallel candidate adaptive de novo mutations within a plant species and the first such example for autopolyploids. Although the theory suggests the reduced efficacy of selection on novel variants in autopolyploids (Otto and Whitton, 2000; Gerstein and Otto, 2009), beneficial alleles are introduced at increased rates and additional variation may accumulate due to polysomic masking (Otto and Whitton, 2000). Such a finding demonstrates that the rapid selection of novel alleles is still feasible in autopolyploids, perhaps reflecting the maintenance of a large pool of pre-existing variation and increased rates of beneficial alleles in organisms with doubled genomes.

6.4 Elucidating the causes of non-parallelism

Even in systems exhibiting generally strong and genetically determined parallelism, both neutral and selective processes may cause significant non-parallel deviations in particular traits or populations. Such preferential focus on parallel traits and genes may lead to overlooking of non-parallels trends, which however also represent informative and relevant evolutionary trajectories (Bolnick et al., 2018). For example, in CS2, I observed a continuum of non-parallelism and alpine *A. arenosa* populations were generally more phenotypically homogeneous than foothill populations. For most of the traits exhibiting non-parallelism in the field, ecotypic differentiation was no longer present under cultivation in the common garden, demonstrating that phenotypic plasticity is the likely major driver of phenotypic non-parallelism in alpine *A. arenosa*. The emergence of non-parallel traits in response to local adaptation is less likely due to overall lower phenotypic variation among the alpine populations and very similar environmental conditions across alpine sites, speaking against strong selection triggered by locally specific conditions.

In serpentine *A. arenosa* (CS5), I found only three traits showing non-parallel variation: above-ground biomass, flowering time, and Ca/Mg ratio in the leaves. Specifically, the S3 population in contrast to the other two serpentine populations flowered earlier, which is in congruence with a previous study by Arnold et al. (2016), suggesting genetically determined non-parallelism. Regarding elemental uptake, the S1 population exhibited an inverse pattern in

the leaf Ca/Mg ratio than the other two population pairs. Complex genetic architecture of the physiological traits such as Ca uptake and Ca/Mg homeostasis in the plant cells can lead to the evolution of different mechanisms to cope with the low Ca/Mg ratio in the tissues.

In CS4, although I found significant parallelism by genes (2 to 4 % of shared candidates between any two pairs), there is a generally high proportion of non-parallel gene candidates and absence of candidate loci shared across all the population pairs. It can be related to the polygenic basis of serpentine adaptation represented by allele frequency shifts in many genes across population pairs (Yeaman, 2015; Wilkinson et al., 2021). Even though significant, gene-level parallelism is relatively rare, as compared to high similarity in functional pathways. Likely, these similarly modified phenotypes arose mainly via selection of different genes in similar functional pathways and developmental processes showing the role of genetic redundancy and stochasticity in rapid adaptation with a polygenic basis (Boyle et al., 2017; Barghi et al., 2020; Láruson et al., 2020).

7 Conclusions

Cases of parallel evolution provide important insights into evolutionary drivers of adaptation. I showed significant parallelism manifested at both the genomic (CS4) and phenotypic levels (CS2 and CS4) in two plant species in which the parallel origin of ecotypes is of recent origin. In general, a higher level of parallelism among populations with similar evolutionary histories is more likely than in systems with differences in the divergence times and variation in the environments (Stern et al., 2009; Thompson et al., 2019). I confirmed that in closely related genetic lineages with a large pool of shared variants repeated sorting of standing genetic variation is a prevailing source for adaptive variation (CS4). We can expect a lower extent of shared variation among distantly related lineages, but this can be overcome by gene flow followed by introgression at reasonable geographic scales (e.g. Lewis et al., 2019). A major mechanism underlying non-parallelism observed in alpine *A. arenosa* in the field is likely phenotypic plasticity (CS2). In serpentine populations of *A. arenosa,* genetic architecture of adaptive traits possibly largely influenced the repeatability of adaptation in this system (CS5). The genetic architecture of the selected adaptive trait has a large impact on the extent of parallelism. Theory suggests that large-effect loci showing more predictable evolutionary paths via constrained genetic pathways are usually involved in the selection of a single well-defined stress factor (Coyne and Orr, 2004; Yeaman et al., 2018; Kim et al., 2021). In contrast, many loci with small effects are often under selection when facing more complex environmental challenges. In such cases, evolution takes different paths as the same adaptive phenotypes can be reached via a combination of different alleles/genes due to genetic redundancy in polygenic systems (Barghi et al., 2020; Láruson et al., 2020; Montejo-Kovacevich et al., 2021). This is also the case of serpentine *A. arenosa* with a polygenic basis of adaptation (CS4).

Approaches used in presented studies still have limitations. Distinguishing between adaptive differentiated candidates and candidates identified solely based on the effect of linkage is usually beyond the limits of the approaches used, given the studied populations are autotetraploid. As a result, the extent of parallelism can be underestimated. The solution for future studies on how to unlink non-causal alleles is to use recombinant populations as was recently applied in *Senecio lautus* (James, Allsopp, et al., 2021). In recombinant populations putatively causative alleles are decoupled from surrounding non-causal variation. The advantage of *A. arenosa* is large effective population sizes, which makes the selection easy to detect, and allows randomization of linked polymorphism.

Although serpentines fascinated evolutionary biologists for decades, there is a lack of studies associating genes underlying ecologically important traits, with their fitness consequences. Thus, we miss the crucial link between the phenotypic effect of a locus and its adaptive value. To detect genomic regions associated with fitness, in future studies, we can bulk segregant analysis combined with shallow individual sequencing. If we want to address parallel genomic basis of serpentine adaptation, it is necessary to create segregating F2 population for each serpentine – non-serpentine population pair and identify genomic regions underlying fitness variation in such mapping populations. Finally, we will be able to compare such fitness candidates with selection candidates previously detected from genome-wide divergence scans of the original populations. Future studies should also involve functional genomics using genetic knock-outs or gene transformation, which will help us validate the promising candidates.

The new natural systems of parallel evolution, established in my thesis, will allow us to further examine the mechanisms of predictability of evolution. Indeed this has already been partly done both in the alpine system of *A. arenosa* (Bohutínská, Vlček, et al., 2021) as well as in serpentine *A. arenosa* (this thesis CS4 and CS5). Uncovering the extent of parallelism and causes of non-parallelism can help us to better understand the predictability of evolution i.e. to

which degree the extent of parallelism matches the theoretical expectations (Stern et al., 2009; Waters and McCulloch, 2021). Understanding repeatability is also essential for predicting how organisms will behave under environmental change. Further, it will help us to better assess the role of genetic redundancy in polygenic systems and functional constraints with impact on the extent of parallelism. Rapid adaptations in geographically separated populations can further help us to understand if de novo mutations in closely related lineages, such as we documented in *TPC1* gene*,* are more common than we previously thought. The next step is to deconstruct complex adaptations and link the genotype – candidate genes or directly alleles – with their fitness effects in challenging environments. With an increasing number of promising model systems of parallel evolution and specifically with a deeper understanding of the genetic architecture of repeated adaptations, closing this gap is not far off.

8 References

Alexa, A., and J. Rahnenfuhrer. 2020. topGO: Enrichment Analysis for Gene Ontology. R package version 2.42.0.

Andrew, R. L., K. L. Ostevik, D. P. Ebert, and L. H. Rieseberg. 2012. Adaptation with gene flow across the landscape in a dune sunflower. *Molecular Ecology* 21: 2078–2091.

Arendt, J., and D. Reznick. 2008. Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? *Trends in Ecology and Evolution* 23: 26–32.

Arnold, B. J., B. Lahner, J. M. DaCosta, C. M. Weisman, J. D. Hollister, D. E. Salt, K. Bomblies, and L. Yant. 2016. Borrowed alleles and convergence in serpentine adaptation. *Proceedings of the National Academy of Sciences* 113: 8320–8325.

Baduel, P., B. Arnold, C. M. Weisman, B. Hunter, and K. Bomblies. 2016. Habitat-Associated life history and stress-tolerance variation in Arabidopsis arenosa. *Plant Physiology* 171: 437–451.

Baduel, P., S. Bray, M. Vallejo-Marin, F. Kolář, and L. Yant. 2018. The 'Polyploid Hop': Shifting challenges and opportunities over the evolutionary lifespan of genome duplications. *Frontiers in Ecology and Evolution* 6: 1–19.

Baduel, P., B. Hunter, S. Yeola, and K. Bomblies. 2018. Genetic basis and evolution of rapid cycling in railway populations of tetraploid Arabidopsis arenosa. *PLoS Genetics* 14: 1–26.

Barghi, N., J. Hermisson, and C. Schlötterer. 2020. Polygenic adaptation: a unifying framework to understand positive selection. *Nature Reviews Genetics* 21: 769–781.

Barghi, N., R. Tobler, V. Nolte, A. M. Jaksic, F. Mallard, K. A. Otte, M. Dolezal, et al. 2019. Polygenic adaptation fuels genetic redundancy in Drosophila. *PLoS Biology* 17: 1–31.

Barrett, R. D. H., S. Laurent, R. Mallarino, S. P. Pfeifer, C. C. Y. Xu, M. Foll, K. Wakamatsu, et al. 2019. Linking a mutation to survival in wild mice. *Science* 363: 499–504.

Barrett, R. D. H., and D. Schluter. 2008. Adaptation from standing genetic variation. *Trends in Ecology and Evolution* 23: 38–44.

Beerli, P., and M. Palczewski. 2010. Unified framework to evaluate panmixia and migration direction among multiple sampling locations. *Genetics* 185: 313–326.

Van Belleghem, S. M., C. Vangestel, K. De Wolf, Z. De Corte, M. Möst, P. Rastas, L. De Meester, and F. Hendrickx. 2018. Evolution at two time frames: Polymorphisms from an ancient singular divergence event fuel contemporary parallel evolution. *PLoS Genetics* 14: 1–26.

Berglund, A. N., S. Dahlgren, and A. Westerbergh. 2004. Evidence for parallel evolution and site-specific selection of serpentine tolerance in Cerastium alpinum during the colonization of Scandinavia. *New Phytologist* 161: 199–209.

Bettin, O., C. Cornejo, P. J. Edwards, and R. Holderegger. 2007. Phylogeography of the high alpine plant Senecio halleri (Asteraceae) in the European Alps: In situ glacial survival with postglacial stepwise dispersal into peripheral areas. *Molecular Ecology* 16: 2517–2524.

Bohutínská, M., V. Handrick, L. Yant, R. Schmickl, F. Kolář, K. Bomblies, and P. Paajanen. 2021. De Novo Mutation and Rapid Protein (Co-)evolution during Meiotic Adaptation in Arabidopsis arenosa . *Molecular Biology and Evolution*: 1–15.

Bohutínská, M., J. Vlček, S. Yair, B. Leanen, V. Konečná, M. Fracassetti, T. Slotte, and F. Kolář. 2021. Genomic basis of parallel adaptation varies with divergence in Arabidopsis and its relatives. *Proceedings of the National Academy of Sciences* 118: e2022713118.

Bolnick, D. I., R. D. H. Barrett, K. B. Oke, D. J. Rennison, and Y. E. Stuart. 2018. (Non) Parallel Evolution. *Annual Review of Ecology, Evolution, and Systematics* 49: 303–330.

Boyle, E. A., Y. I. Li, and J. K. Pritchard. 2017. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* 169: 1177–1186.

Bradshaw, H. D. 2005. Mutations in CAX1 produce phenotypes characteristic of plants tolerant to serpentine soils. *New Phytologist* 167: 81–88.

Brady, K. U., A. R. Kruckeberg, and H. D. Bradshaw Jr. 2005. Evolutionary Ecology of Plant Adaptation to Serpentine Soils. *Annual Review of Ecology, Evolution, and Systematics* 36: 243–266.

Butlin, R. K., M. Saura, B. Jackson, C. Andr, A. Caballero, J. A. Coyne, J. Galindo, et al. 2014. Parallel evolution of local adaptation and reproductive isolation in the face of gene flow. *Evolution* 68: 935–949.

Cai, Z., L. Zhou, N. N. Ren, X. Xu, R. Liu, L. Huang, X. M. Zheng, et al. 2019. Parallel Speciation of Wild Rice Associated with Habitat Shifts. *Molecular Biology and Evolution* 36: 875–889.

Caye, K., B. Jumentier, J. Lepeule, and O. François. 2019. LFMM 2: Fast and accurate inference of gene-environment associations in genome-wide studies. *Molecular Biology and Evolution* 36: 852–860.

Cecchi, L., and F. Selvi. 2009. Phylogenetic relationships of the monotypic genera Halacsya and Paramoltkia and the origins of serpentine adaptation in circummediterranean Lithospermeae (Boraginaceae): Insights from ITS and matK DNA sequences. *Taxon* 58: 700–714.

Chan, Y. F., M. E. Marks, F. C. Jones, G. Villarreal, M. D. Shapiro, S. D. Brady, A. M. Southwick, et al. 2010. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a pitxl enhancer. *Science* 327: 302–305.

Colosimo, P. F., K. E. Hosemann, S. Balabhadra, G. V. Jr, M. Dickson, J. Grimwood, J. Schmutz, et al. 2005. Widespread Parallel Evolution in Sticklebacks by Repeated Fixation of Ectodysplasin Alleles. *Science* 307: 1928–1933.

Consortium, T. H. G., K. K. Dasmahapatra, J. R. Walters, A. D. Briscoe, J. W. Davey, A. Whibley, N. J. Nadeau, et al. 2012. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* 487: 94–98.

Conte, G. L., M. E. Arnegard, C. L. Peichel, and D. Schluter. 2012. The probability of genetic parallelism and convergence in natural populations. *Proceedings of the Royal Society B: Biological Sciences* 279: 5039–5047.

Corl, A., K. Bi, C. Luke, A. S. Challa, A. J. Stern, B. Sinervo, and R. Nielsen. 2018. The Genetic Basis of Adaptation following Plastic Changes in Coloration in a Novel Environment. *Current Biology* 28: 2970-2977.

Coyne, J. A., and H. A. Orr. 2004. Speciation. MA: Sinauer Associates, Sunderland.

Elmer, K. R., and A. Meyer. 2011. Adaptation in the age of ecological genomics: Insights from parallelism and convergence. *Trends in Ecology and Evolution* 26: 298–306.

Van Etten, M., K. M. Lee, S. M. Chang, and R. S. Baucom. 2020. Parallel and nonparallel genomic responses contribute to herbicide resistance in Ipomoea purpurea, a common agricultural weed. *PLoS genetics* 16: e1008593.

Excoffier, L., and M. Foll. 2011. fastsimcoal: A continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics* 27: 1332–1334.

Fang, B., P. Kemppainen, P. Momigliano, and J. Merilä. 2021. Population Structure Limits Parallel Evolution in Sticklebacks. *Molecular Biology and Evolution* 38: 4205–4221.

Fraïsse, C., and J. J. Welch. 2019. The distribution of epistasis on simple fitness landscapes. *Biology Letters* 15: 20180881.

Funk, W. C., M. S. Blouin, P. S. Corn, B. A. Maxell, and D. S. Pilliod. 2005. Population structure of Columbia spotted frogs (Rana lueteiventris) is strongly affected by the landscape. *Molecular Ecology* 14: 1–14.

Fustier, M., J. Brandenburg, S. Boitard, J. Lapeyronnie, and L. E. Eguiarte. 2017. Signatures of local adaptation in lowland and highland teosintes from whole-genome sequencing of pooled samples. *Molecular Ecology* 26: 2738–2756.

Gerstein, A. C., and S. P. Otto. 2009. Ploidy and the causes of genomic evolution. *Journal of Heredity* 100: 571–581.

Geyer, C. J., S. Wagenius, and R. G. Shaw. 2007. Aster models for life history analysis. *Biometrika* 94: 415–426.

Giska, I., L. Farelo, J. Pimenta, F. A. Seixas, M. S. Ferreira, J. P. Marques, I. Miranda, et al. 2019. Introgression drives repeated evolution of winter coat color polymorphism in hares. *Proceedings of the National Academy of Sciences of the United States of America* 116: 24150–24156.

Greenway, R., N. Barts, C. Henpita, A. P. Brown, L. A. Rodriguez, C. M. Rodríguez Peña, S. Arndt, et al. 2020. Convergent evolution of conserved mitochondrial pathways underlies repeated adaptation to extreme environments. *Proceedings of the National Academy of Sciences of the United States of America* 117: 21822.

Günther, T., C. Lampei, I. Barilar, and K. J. Schmid. 2016. Genomic and phenotypic differentiation of Arabidopsis thaliana along altitudinal gradients in the North Italian Alps. *Molecular Ecology* 25: 3574–3592.

Hämälä, T., A. J. Gorton, D. A. Moeller, and P. Tiffin. 2020. Pleiotropy facilitates local adaptation to distant optima in common ragweed (Ambrosia artemisiifolia). *PLoS Genetics* 16: 1–23.

Hämälä, T., T. M. Mattila, and O. Savolainen. 2018. Local adaptation and ecological differentiation under selection, migration, and drift in Arabidopsis lyrata. *Evolution* 72: 1373–1386.

Hämälä, T., and O. Savolainen. 2019. Genomic patterns of local adaptation under gene flow in Arabidopsis lyrata. *Molecular Biology and Evolution* 36: 2557–2571.

Harrison, S., and N. Rajakaruna. 2011. Serpentine: the evolution and ecology of a model system. University of California Press.

Hereford, J. 2009. A quantitative survey of local adaptation and fitness trade-offs. *American Naturalist* 173: 579–588.

Hermisson, J., and P. S. Pennings. 2005. Soft sweeps: Molecular population genetics of adaptation from standing genetic variation. *Genetics* 169: 2335–2352.

Hermisson, J., and P. S. Pennings. 2017. Soft sweeps and beyond: understanding the patterns and probabilities of selection footprints under rapid adaptation. *Methods in Ecology and Evolution* 8: 700–716.

Hill, J., E. D. Enbody, M. E. Pettersson, C. G. Sprehn, D. Bekkevold, A. Folkvord, L. Laikre, et al. 2019. Recurrent convergent evolution at amino acid residue 261 in fish rhodopsin. *Proceedings of the National Academy of Sciences of the United States of America* 116: 18473–18478.

Hoban, S., J. L. Kelley, K. E. Lotterhos, M. F. Antolin, G. Bradburd, D. B. Lowry, M. L. Poss, et al. 2016. Finding the Genomic Basis of Local Adaptation: Pitfalls, Practical Solutions, and Future Directions. *The American naturalist* 188: 379–397.

Hoekstra, H. E., R. J. Hirschmann, R. A. Bundey, P. A. Insel, and J. P. Crossland. 2006. A Single Amino Acid Mutation Contributes to Adaptive Beach Mouse Color Pattern. *Science* 313: 101–104.

Holliday, J. A., L. Zhou, R. Bawa, M. Zhang, and R. W. Oubida. 2016. Evidence for extensive parallelism but divergent genomic architecture of adaptation along altitudinal and latitudinal gradients in Populus trichocarpa. *New Phytologist* 209: 1240–1251.

Höllinger, I., P. S. Pennings, and J. Hermisson. 2019. Polygenic adaptation: From sweeps to subtle frequency shifts. *PLoS Genetics* 15: 1–26.

Hollister, J. D., B. J. Arnold, E. Svedin, K. S. Xue, B. P. Dilkes, and K. Bomblies. 2012. Genetic Adaptation Associated with Genome-Doubling in Autotetraploid Arabidopsis arenosa. *PLoS Genetics* 8: e1003093.

Hu, T. T., P. Pattyn, E. G. Bakker, J. Cao, J. Cheng, R. M. Clark, N. Fahlgren, et al. 2011. The Arabidopsis lyrata genome sequence and the basis of rapid genome size change. *Nature genetics* 43: 476–481.

Hufford, M. B., P. Lubinksy, T. Pyhäjärvi, M. T. Devengenzo, N. C. Ellstrand, and J. Ross-Ibarra. 2013. The Genomic Signature of Crop-Wild Introgression in Maize. *PLoS Genetics* 9.

Jacobs, A., M. Carruthers, A. Yurchenko, N. V. Gordeeva, S. S. Alekseyev, O. Hooker, J. S. Leong, et al. 2020. Parallelism in eco-morphology and gene expression despite variable evolutionary and genomic backgrounds in a Holarctic fish. *PLoS Genetics* 16: e1008658.

Jacquemyn, H., R. Brys, and M. Hermy. 2002. Patch occupancy, population size and reproductive success of a forest herb (Primula elatior) in a fragmented landscape. *Oecologia* 130: 617–625.

Jacquemyn, H., O. Honnay, P. Galsbusera, and I. Roldán-Ruiz. 2004. Genetic structure of the forest herb Primula elatior in a changing landscape. *Molecular Ecology* 13: 211–219.

Jacquemyn, H., K. Vandepitte, I. Roldán-Ruiz, and O. Honnay. 2008. Rapid loss of genetic variation in a founding population of Primula elatior (Primulaceae) after colonization. *Annals of Botany* 103: 777–783.

Jain, S. K., and A. D. Bradshaw. 1966. Evolutionary divergence among adjacent plant populations I. The evidence and its theoretical analysis. *Heredity* 21: 407–441.

James, M. E., R. N. Allsopp, J. S. Groh, K. Avneet, M. J. Wilkinson, and D. Ortiz-Barrientos. 2021. Uncovering the genetic architecture of replicated adaptation. *SSRN preprint. 3981902.*

James, M. E., H. Arenas-Castro, J. S. Groh, S. L. Allen, J. Engelstädter, and D. Ortiz-Barrientos. 2021. Highly Replicated Evolution of Parapatric Ecotypes. *Molecular Biology and Evolution* 38: 4805–4821.

James, M. E., M. J. Wilkinson, D. M. Bernal, H. Liu, H. L. North, J. Engelstädter, and D. Ortiz-Barrientos. 2021. Phenotypic and genotypic parallel evolution in parapatric ecotypes of Senecio. *Evolution* 75: 3115–3131.

Ji, Y., X. Li, T. Ji, J. Tang, L. Qiu, J. Hu, J. Dong, et al. 2020. Gene reuse facilitates rapid radiation and independent adaptation to diverse habitats in the Asian honeybee. *Science Advances* 6: eabd3590.

Jones, F. C., M. G. Grabherr, Y. F. Chan, P. Russell, E. Mauceli, J. Johnson, R. Swofford, et al. 2012. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* 484*:* 55–61.

Kawecki, T. J., and D. Ebert. 2004. Conceptual issues in local adaptation. *Ecology Letters* 7: 1225–1241.

Kazakou, E., P. G. Dimitrakopoulos, A. J. M. Baker, R. D. Reeves, and A. Y. Troumbis. 2008. Hypotheses , mechanisms and trade-offs of tolerance and adaptation to serpentine soils: from species to ecosystem level. 495–508.

Kersten, S., J. Chang, C. D. Huber, Y. Voichek, C. Lanz, P. Lang, U. Lutz, et al. 2021. Standing genetic variation fuels rapid evolution of herbicide resistance in blackgrass. *bioRxiv*: 2021.12.14.472587.

Kim, K.-W., R. De-Kayne, I. J. Gordon, K. S. Omufwoko, D. J. Martins, Ff.-C. Richard, and S. H. Martin. 2021. Stepwise evolution of a butterfly supergene via duplication and inversion. *bioRxiv*: 2021.12.06.471392.

Knotek, A., V. Konečná, G. Wos, D. Požárová, G. Šrámková, M. Bohutínská, V. Zeisek, et al. 2020. Parallel Alpine Differentiation in Arabidopsis arenosa. *Frontiers in Plant Science* 11: 561526.

Kolář, F., T. Fér, M. Štech, P. Trávníček, E. Dušková, P. Schönswetter, and J. Suda. 2012. Bringing together evolution on serpentine and polyploidy: Spatiotemporal history of the diploid-tetraploid complex of Knautia arvensis (Dipsacaceae). *PLoS ONE* 7: e39988.

Kolář, F., G. Fuxová, Z. Eliška, A. J. Nagano, L. Hyklová, M. Lučanová, H. Kudoh, and K. Marhold. 2016. Northern glacial refugia and altitudinal niche divergence shape genome-wide differentiation in the emerging plant model Arabidopsis arenosa. *Molecular Ecology* 25: 3929–3949.

Kolbe, J. J., M. Leal, T. W. Schoener, D. A. Spiller, and J. B. Losos. 2012. Founder effects persist despite adaptive differentiation: A field experiment with lizards. *Science* 335: 1086–1089.

Konečná, V., S. Bray, J. Vlček, M. Bohutínská, D. Požárová, R. R. Choudhury, A. Bollmann-Giolai, et al. 2021. Parallel adaptation in autopolyploid Arabidopsis arenosa is dominated by repeated recruitment of shared

alleles. *Nature Communications* 12: 4979.
Konečná, V., L. Yant, and F. Kolář. 2020. The Evolutionary Genomics of Serpentine Adaptation. *Frontiers in Plant Science* 11: 574616.
Körner, C. 2003. Alpine plant life: functional plant ecology of high mountain ecosystems. 2nd ed. Springer Science & Business Media.
Korves, T. M., K. J. Schmid, A. L. Caicedo, C. Mays, J. R. Stinchcombe, M. D. Purugganan, and J. Schmitt. 2007. Fitness effects associated with the major flowering time gene FRIGIDA in Arabidopsis thaliana in the field. *The American naturalist* 169: E141-E157.
Kreiner, J. M., D. A. Giacomini, F. Bemm, B. Waithaka, J. Regalado, C. Lanz, J. Hildebrandt, et al. 2019. Multiple modes of convergent adaptation in the spread of glyphosate-resistant Amaranthus tuberculatus. *Proceedings of the National Academy of Sciences* 116: 21076–21084.
Kruckeberg, A. R. 1950. An Experimental Inquiry into the Nature of Endemism on Serpentine Soils. University of California, Berkeley.
Kruckeberg, A. R. 1967. Ecotypic Response to Ultramafic Soils by Some Plant Species of Northwestern United States. *Brittonia* 19: 133–151.
Kruckeberg, A. R. 1951. Intraspecific Variability in the Response of Certain Native Plant Species to Serpentine Soil. *American Journal of Botany* 38: 408–419.
Kubota, S., T. Iwasaki, K. Hanada, A. J. Nagano, A. Fujiyama, A. Toyoda, S. Sugano, et al. 2015. A Genome Scan for Genes Underlying Microgeographic-Scale Local Adaptation in a Wild Arabidopsis Species. *PLoS Genetics* 11: 1–26.
Lai, Y. T., C. K. L. Yeung, K. E. Omland, E. L. Pang, Y. Hao, B. Y. Liao, H. F. Cao, et al. 2019. Standing genetic variation as the predominant source for adaptation of a songbird. *Proceedings of the National Academy of Sciences of the United States of America* 116: 2152–2157.
Langerhans, R. B. 2018. Predictability and Parallelism of Multitrait Adaptation. *Journal of Heredity* 109: 59–70.
Láruson, Á. J., S. Yeaman, and K. E. Lotterhos. 2020. The Importance of Genetic Redundancy in Evolution. *Trends in Ecology and Evolution* 35: 809–822.
Lee, K. M., and G. Coop. 2017. Distinguishing among modes of convergent adaptation using population genomic data. *Genetics* 207: 1591–1619.
Leinonen, P. H., D. L. Remington, J. Leppälä, and O. Savolainen. 2013. Genetic basis of local adaptation and flowering time variation in Arabidopsis lyrata. *Molecular Ecology* 22: 709–723.
Lenormand, T., D. Roze, and F. Rousset. 2009. Stochasticity in evolution. *Trends in Ecology and Evolution* 24: 157–165.
Levin, D. A. 2001. The recurrent origin of plant races and species. *Systematic Botany* 26: 197–204.
Lewis, J. J., R. C. Geltman, P. C. Pollak, K. E. Rondem, S. M. van Belleghem, M. J. Hubisz, P. R. Munn, et al. 2019. Parallel evolution of ancient, pleiotropic enhancers underlies butterfly wing pattern mimicry. *Proceedings of the National Academy of Sciences of the United States of America* 116: 24174–24183.
Louis, M., M. Galimberti, F. Archer, S. Berrow, A. Brownlow, R. Fallon, M. Nykänen, et al. 2021. Selection on ancestral genetic variation fuels repeated ecotype formation in bottlenose dolphins. *Science Advances* 7: 1–14.
Lucek, K., A. Sivasundar, and O. Seehausen. 2014. Disentangling the role of phenotypic plasticity and genetic divergence in contemporary ecotype formation during a biological invasion. *Evolution* 68: 2619–2632.
Luo, Y., A. Widmer, and S. Karrenberg. 2015. The roles of genetic drift and natural selection in quantitative trait divergence along an altitudinal gradient in Arabidopsis thaliana. *Heredity* 114: 220–228.
Magalhaes, I. S., J. R. Whiting, D. D'Agostino, P. A. Hohenlohe, M. Mahmud, M. A. Bell, S. Skúlason, and A. D. C. MacColl. 2021. Intercontinental genomic parallelism in multiple three-spined stickleback adaptive radiations. *Nature Ecology and Evolution* 5: 251–261.
Manceau, M., V. S. Domingues, C. R. Linnen, E. B. Rosenblum, and H. E. Hoekstra. 2010. Convergence in pigmentation at multiple levels: Mutations, genes and function. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365: 2439–2450.
Marburger, S., P. Monnahan, P. J. Seear, S. H. Martin, J. Koch, P. Paajanen, M. Bohutínská, et al. 2019. Interspecific introgression mediates adaptation to whole genome duplication. *Nature Communications* 10: 1–11.
Mayer, M., and P. S. Soltis. 1994. The Evolution of Serpentine Endemics : A Chloroplast DNA Phylogeny of the Streptanthus glandulosus Complex (Cruciferae). *Systematic Botany* 19: 557–574.
Meier, J. I., V. C. Sousa, D. A. Marques, O. M. Selz, C. E. Wagner, L. Excoffier, and O. Seehausen. 2017. Demographic modelling with whole-genome data reveals parallel origin of similar Pundamilia cichlid species after hybridization. *Molecular Ecology* 26: 123–141.
Mengoni, A., A. J. M. Baker, M. Bazzicalupo, R. D. Reeves, N. Adigüzel, E. Chianni, F. Galardi, et al. 2003. Evolutionary dynamics of nickel hyperaccumulation in Alyssum revealed by ITS nrDNA analysis. *New Phytologist* 159: 691–699.

Molina-Henao, Y. F., and R. Hopkins. 2019. Autopolyploid lineage shows climatic niche expansion but not divergence in Arabidopsis arenosa. *American Journal of Botany* 106: 61–70.
Monnahan, P., F. Kolář, P. Baduel, C. Sailer, J. Koch, R. Horvath, B. Laenen, et al. 2019. Pervasive population genomic consequences of genome duplication in Arabidopsis arenosa. *Nature ecology & evolution* 3: 457.
Montejo-Kovacevich, G., J. I. Meier, C. N. Bacquet, I. A. Warren, Y. F. Chan, M. Kucka, C. Salazar, et al. 2021. Repeated genetic adaptation to high altitude in two tropical butterflies. *bioRxiv*: 2021.11.30.470630.
Morales, H. E., R. Faria, K. Johannesson, T. Larsson, M. Panova, A. M. Westram, and R. K. Butlin. 2019. Genomic architecture of parallel ecological divergence: Beyond a single environmental contrast. *Science Advances* 5: eaav9963.
Nelson, T. C., and W. A. Cresko. 2018. Ancient genomic variation underlies repeated ecological adaptation in young stickleback populations. *Evolution Letters* 2: 9–21.
Nosil, P., and B. J. Crespi. 2004. Does gene flow constrain adaptive divergence or vice versa? A test using ecomorphology and sexual isolation in Timema cristinae walking-sticks. *Evolution* 58: 102–112.
Nosil, P., R. Villoutreix, C. F. De Carvalho, T. E. Farkas, V. Soria-Carrasco, J. L. Feder, B. J. Crespi, and Z. Gompert. 2018. Natural selection and the predictability of evolution in timema stick insects. *Science* 359: 765–770.
O'Dell, R. E., and N. Rajakaruna. 2011. Intraspecific variation, adaptation, and evolution. *In* S. Harrison, and N. Rajakaruna [eds.], Serpentine: Evolution and ecology in a model system, 97–137. University of California Press.
Odat, N., G. Jetschke, and F. H. Hellwig. 2004. Genetic diversity of Ranunculus acris L. (Ranunculaceae) populations in relation to species diversity and habitat type in grassland communities. *Molecular Ecology* 13: 1251–1257.
Ortego, J., E. C. Riordan, P. F. Gugger, and V. L. Sork. 2012. Influence of environmental heterogeneity on genetic diversity and structure in an endemic southern Californian oak. *Molecular Ecology* 21: 3210–3223.
Otto, S. P., and J. Whitton. 2000. Polyploid incidence and evolution. *Annual Review of Genetics* 34: 401–437.
Oziolor, E. M., N. M. Reid, S. Yair, K. M. Lee, S. Guberman VerPloeg, P. C. Bruns, J. R. Shaw, et al. 2019. Adaptive introgression enables evolutionary rescue from extreme environmental pollution. *Science* 364: 455–457.
Papadopulos, A. S. T., A. J. Helmstetter, O. G. Osborne, A. A. Comeault, D. P. Wood, E. A. Straw, L. Mason, et al. 2021. Rapid Parallel Adaptation to Anthropogenic Heavy Metal Pollution. *Molecular Biology and Evolution* 38: 3724–3736.
Van De Peer, Y., E. Mizrachi, and K. Marchal. 2017. The evolutionary significance of polyploidy. *Nature Reviews Genetics* 18: 411–424.
Pittet, L., Y. Fragnière, S. Grünig, S. Bétrisey, B. Clément, E. Gerber, M. Ronikier, et al. 2020. Genetic structure of the endemic Papaver occidentale indicates survival and immigration in the Western Prealps. *Alpine Botany* 130: 129–140.
Preite, V., C. Sailer, L. Syllwasschy, S. Bray, U. Kraemer, and L. Yant. 2019. Convergent evolution in Arabidopsis halleri and Arabidopsis arenosa on calamine metalliferous soils. *Philosophical Transactions of the Royal Society B* 374: 20180243.
Przedpełska, E., and M. Wierzbicka. 2007. Arabidopsis arenosa (Brassicaceae) from a lead-zinc waste heap in southern Poland - A plant with high tolerance to heavy metals. *Plant and Soil* 299: 43–53.
Przeworski, M., G. Coop, and J. D. Wall. 2005. The Signature of Positive Selection on Standing Genetic Variation. *Evolution* 59: 2312.
Quesada, H., D. Posada, A. Caballero, P. Morán, and E. Rolán-Alvarez. 2007. Phylogenetic evidence for multiple sympatric ecological diversification in a marine snail. *Evolution* 61: 1600–1612.
Rajakaruna, N. 2018. Lessons on Evolution from the Study of Edaphic Specialization. *The Botanical Review* 84: 39–78.
Rajakaruna, N., B. G. Baldwin, R. Chan, A. M. Desrochers, B. A. Bohm, and J. Whitton. 2003. Edaphic races and phylogenetic taxa in the Lasthenia californica complex (Asteraceae: Heliantheae): An hypothesis of parallel evolution. *Molecular Ecology* 12: 1675–1679.
Richards, J. 2003. Primula. 2nd ed. Timber Press, Portland, OR, USA.
Roda, F., L. Ambrose, G. M. Walter, H. L. Liu, A. Schaul, A. Lowe, P. B. Pelser, et al. 2013. Genomic evidence for the parallel evolution of coastal forms in the Senecio lautus complex. *Molecular Ecology* 22: 2941–2952.
Rosenblum, E. B., C. E. Parent, and E. E. Brandt. 2014. The Molecular Basis of Phenotypic Convergence. *Annual Review of Ecology, Evolution, and Systematics* 45: 203–226.
Van Rossum, F., and L. Triest. 2006. Fine-scale genetic structure of the common Primula elatior (Primulaceae) at an early stage of population fragmentation. *American Journal of Botany* 93: 1281–1288.
Rundle, H. D., and P. Nosil. 2005. Ecological speciation. *Ecology Letters* 8: 336–352.
Sakaguchi, S., K. Horie, N. Ishikawa, A. J. Nagano, M. Yasugi, H. Kudoh, and M. Ito. 2017. Simultaneous

evaluation of the effects of geographic, environmental and temporal isolation in ecotypic populations of Solidago virgaurea. *New Phytologist* 216: 1268–1280.
Sakaguchi, S., K. Horie, T. Kimura, A. J. Nagano, Y. Isagi, and M. Ito. 2018. Phylogeographic testing of alternative histories of single-origin versus parallel evolution of early flowering serpentine populations of Picris hieracioides L. (Asteraceae) in Japan. *Ecological Research* 33: 537–547.
Salmón, P., A. Jacobs, D. Ahrén, C. Biard, N. J. Dingemanse, D. M. Dominoni, B. Helm, et al. 2021. Continent-wide genomic signatures of adaptation to urbanisation in a songbird across Europe. *Nature Communications* 12: 2983.
Schmickl, R., J. Paule, J. Klein, K. Marhold, and M. A. Koch. 2012. The evolutionary history of the Arabidopsis arenosa complex: Diverse tetraploids mask the Western Carpathian center of species and genetic diversity. *PLoS ONE* 7: e42691.
Selby, J. P., and J. H. Willis. 2018. Major QTL controls adaptation to serpentine soils in Mimulus guttatus. *Molecular Ecology* 27: 5073–5087.
Shao, J., J. Wang, Y. Xu, Q. Pan, Y. A. Shi, S. Kelso, and G. Lv. 2015. Genetic diversity and gene flow within and between two different habitats of Primula merrilliana (Primulaceae), an endangered distylous forest herb in eastern China. *Botanical Journal of the Linnean Society* 179: 172–189.
Shaw, R. G., C. J. Geyer, S. Wagenius, H. H. Hangelbroek, and J. R. Etterson. 2008. Unifying life-history analyses for inference of fitness and population growth. *American Naturalist* 172: E35–E47.
Sobczyk, M. K., J. A. C. Smith, A. J. Pollard, and D. A. Filatov. 2017. Evolution of nickel hyperaccumulation and serpentine adaptation in the Alyssum serpyllifolium species complex. *Heredity* 118: 31–41.
Song, Y., S. Endepols, N. Klemann, D. Richter, F. R. Matuschka, C. H. Shih, M. W. Nachman, and M. H. Kohn. 2011. Adaptive introgression of anticoagulant rodent poison resistance by hybridization between old world mice. *Current Biology* 21: 1296–1301.
Soria-Carrasco, V., Z. Gompert, A. A. Comeault, T. E. Farkas, T. L. Parchman, J. S. Johnston, C. A. Buerkle, et al. 2014. Stick insect genomes reveal natural selection's role in parallel speciation. *Science* 344: 738–742.
Šrámková-Fuxová, G., E. Záveská, F. Kolář, M. Lučanová, S. Španiel, and K. Marhold. 2017. Range-wide genetic structure of Arabidopsis halleri (Brassicaceae): glacial persistence in multiple refugia and origin of the Northern Hemisphere disjunction. *Botanical Journal of the Linnean Society* 185: 321–342.
Steiner, C. C., H. Römpler, L. M. Boettger, T. Schöneberg, and H. E. Hoekstra. 2009. The genetic basis of phenotypic convergence in beach mice: Similar pigment patterns but different genes. *Molecular Biology and Evolution* 26: 35–45.
Stern, D. L. 2013. The genetic causes of convergent evolution. *Nature Reviews Genetics* 14: 751–764.
Stern, D. L., V. Orgogozo, U. Pierre, A. Bâtiment, and Q. Saint. 2009. Is Genetic Evolution Predictable ? Nonrandom Distribution of Evolutionarily Relevant Mutations. *Science* 323: 746–751.
Stuart, Y. E., T. Veen, J. N. Weber, D. Hanson, M. Ravinet, B. K. Lohman, C. J. Thompson, et al. 2017. Contrasting effects of environment and genetics generate a continuum of parallel evolution. *Nature Ecology and Evolution* 1: 1–7.
Svardal, H., F. X. Quah, M. Malinsky, B. P. Ngatunga, E. A. Miska, W. Salzburger, M. J. Genner, et al. 2020. Ancestral hybridization facilitated species diversification in the lake malawi cichlid fish adaptive radiation. *Molecular Biology and Evolution* 37: 1100–1113.
Szklarczyk, D., A. Franceschini, S. Wyder, K. Forslund, D. Heller, J. Huerta-Cepas, M. Simonovic, et al. 2015. STRING v10: Protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Research* 43: D447–D452.
Szukala, A., J. Lovegrove-Walsh, H. Luqman, S. Fior, T. Wolfe, B. Frajman, P. Schönswetter, and O. Paun. 2021. Polygenic routes lead to parallel altitudinal adaptation in Heliosperma pusillum (Caryophyllaceae). *bioRxiv:* 2021.07.05.451094.
Takuno, S., P. Ralph, K. Swarts, R. J. Elshire, J. C. Glaubitz, E. S. Buckler, M. B. Hufford, and J. Ross-Ibarra. 2015. Independent Molecular Basis of Convergent Highland Adaptation in Maize. *Genetics* 200: 1297–1312.
Thompson, K. A., M. M. Osmond, and D. Schluter. 2019. Parallel genetic evolution and speciation from standing variation. *Evolution Letters* 3: 129–141.
Todesco, M., G. L. Owens, N. Bercovich, J. S. Légaré, S. Soudi, D. O. Burge, K. Huang, et al. 2020. Massive haplotypes underlie ecotypic differentiation in sunflowers. *Nature* 584: 602–607.
Trucchi, E., B. Frajman, T. H. A. Haverkamp, P. Schönswetter, and O. Paun. 2017. Genomic analyses suggest parallel ecological divergence in Heliosperma pusillum (Caryophyllaceae). *New Phytologist* 216: 267–278.
Turner, T. L., E. C. Bourne, E. J. Von Wettberg, T. T. Hu, and S. V Nuzhdin. 2010. Population resequencing reveals local adaptation of Arabidopsis lyrata to serpentine soils. *Nature Genetics* 42: 260–263.
Urban, S., A. Nater, A. Meyer, and C. F. Kratochwil. 2021. Different Sources of Allelic Variation Drove Repeated Color Pattern Divergence in Cichlid Fishes. *Molecular Biology and Evolution* 38: 465–477.
Valentine, D. H., and A. Kress. 1972. Primula L. *In* T. Tutin [ed.], Flora Europaea 3, Cambridge University

Press, Cambridge.
Villoutreix, R., C. F. De Carvalho, V. Soria-Carrasco, D. Lindtke, M. De-La-Mora, M. Muschick, J. L. Feder, et al. 2020. Large-scale mutation in the evolution of a gene complex for cryptic coloration. *Science* 369: 460–466.
Wang, L., E. B. Josephs, K. M. Lee, L. M. Roberts, R. Rellán-Álvarez, J. Ross-Ibarra, and M. B. Hufford. 2021. Molecular Parallelism Underlies Convergent Highland Adaptation of Maize Landraces. *Molecular Biology and Evolution* 38: 3567–3580.
Wang, M., Y. Zhao, and B. Zhang. 2015. Efficient Test and Visualization of Multi-Set Intersections. *Scientific Reports* 5: 1–12.
Waters, J. M., and G. A. McCulloch. 2021. Reinventing the wheel? Reassessing the roles of gene flow, sorting and convergence in repeated evolution. *Molecular Ecology* 30: 4162–4172.
Weber, A. A. T., J. Rajkov, K. Smailus, B. Egger, and W. Salzburger. 2021. Speciation dynamics and extent of parallel evolution along a lake-stream environmental contrast in African cichlid fishes. *Science Advances* 7: 1–21.
Von Wettberg, E. J. B., J. Ray-Mukherjee, N. D'Adesky, D. Nesbeth, and S. Sistla. 2014. The Evoltionary Ecology and Genetics of Stress Resistance Syndrome (SRS) Traits: Revisiting Chapin, Autumn and Pugnaire (1993). *In* N. Rajakaruna, R. S. Boyd, and T. B. Harris [eds.], Plant ecology and evolution in harsh environments, 201–226. Nova Science Publishers, New York.
Whiting, J. R., J. R. Paris, M. J. van der Zee, P. J. Parsons, D. Weigel, and B. A. Fraser. 2021. Drainage-structuring of ancestral variation and a common functional pathway shape limited genomic convergence in natural high- And low-predation guppies. *PLoS Genetics* 17: 1–29.
Whittaker, R. H. 1954. The ecology of serpentine soils. *Ecology* 35: 258–288.
Wilkinson, M. J., F. Roda, G. M. Walter, M. E. James, R. Nipper, J. Walsh, S. L. Allen, et al. 2021. Adaptive divergence in shoot gravitropism creates hybrid sterility in an Australian wildflower. *Proceedings of the National Academy of Sciences of the United States of America* 118: 1–11.
Wood, T. E., J. M. Burke, and L. H. Rieseberg. 2005. Parallel genotypic adaptation: When evolution repeats itself. *Genetica* 123: 157–170.
Xie, K. T., G. Wang, A. C. Thompson, J. I. Wucherpfennig, T. E. Reimchen, A. D. C. Maccoll, D. Schluter, et al. 2019. DNA fragility in the parallel evolution of pelvic reduction in stickleback fish. *Science* 84: 81–84.
Yant, L., and K. Bomblies. 2017. Genomic studies of adaptive evolution in outcrossing Arabidopsis species. *Current Opinion in Plant Biology* 36: 9–14.
Yeaman, S. 2015. Local adaptation by alleles of small effect. *American Naturalist* 186: S74–S89.
Yeaman, S., A. C. Gerstein, K. A. Hodgins, and M. C. Whitlock. 2018. Quantifying how constraints limit the diversity of viable routes to adaptation. *PLoS Genetics* 14: 1–25.
Young, A., T. Boyle, and T. Brown. 1996. The population genetic consequences of habitat fragmentation for plants. *Trends in ecology & evolution* 11: 413–418.
de Zelicourt, A., J. Colcombet, and H. Hirt. 2016. The Role of MAPK Modules and ABA during Abiotic Stress Signaling. *Trends in Plant Science* 21: 677–685.
Zhang, X., J. G. Rayner, M. Blaxter, and N. W. Bailey. 2021. Rapid parallel adaptation despite gene flow in silent crickets. *Nature Communications* 12: 1–15.
Zong, S.-B., Y.-L. Li, and J.-X. Liu. 2020. Genomic Architecture of Rapid Parallel Adaptation to Fresh Water in a Wild Fish. *Molecular Biology and Evolution*: 1–13.

Part B – Case studies

Case study 1

The evolutionary genomics of serpentine adaptation

frontiers in Plant Science

REVIEW
published: 16 December 2020
doi: 10.3389/fpls.2020.574616

The Evolutionary Genomics of Serpentine Adaptation

Veronika Konečná[1,2], Levi Yant[3] and Filip Kolář[1,2,4]**

[1]Department of Botany, Faculty of Science, Charles University, Prague, Czechia, [2]Institute of Botany, The Czech Academy of Sciences, Průhonice, Czechia, [3]Future Food Beacon and School of Life Sciences, University of Nottingham, Nottingham, United Kingdom, [4]Natural History Museum, University of Oslo, Oslo, Norway

Serpentine barrens are among the most challenging settings for plant life. Representing a perfect storm of hazards, serpentines consist of broadly skewed elemental profiles, including abundant toxic metals and low nutrient contents on drought-prone, patchily distributed substrates. Accordingly, plants that can tolerate the challenges of serpentine have fascinated biologists for decades, yielding important insights into adaptation to novel ecologies through physiological change. Here we highlight recent progress from studies which demonstrate the power of serpentine as a model for the genomics of adaptation. Given the moderate – but still tractable – complexity presented by the mix of hazards on serpentine, these venues are well-suited for the experimental inquiry of adaptation both in natural and manipulated conditions. Moreover, the island-like distribution of serpentines across landscapes provides abundant natural replicates, offering power to evolutionary genomic inference. Exciting recent insights into the genomic basis of serpentine adaptation point to a partly shared basis that involves sampling from common allele pools available from retained ancestral polymorphism or *via* gene flow. However, a lack of integrated studies deconstructing complex adaptations and linking candidate alleles with fitness consequences leaves room for much deeper exploration. Thus, we still seek the crucial direct link between the phenotypic effect of candidate alleles and their measured adaptive value – a prize that is exceedingly rare to achieve in any study of adaptation. We expect that closing this gap is not far off using the promising model systems described here.

Keywords: serpentine, adaptation, population genomics, edaphic extremes, ionomics

OPEN ACCESS

Edited by:
Thomas L. P. Couvreur, IRD UMR232 Diversité, adaptation, développement des plantes (DIADE), France

Reviewed by:
Nishanta Rajakaruna, California Polytechnic State University, United States
Simon Cornelis Groen, New York University, United States

***Correspondence:**
Filip Kolář
filip.kolar@natur.cuni.cz;
filip.kolar@gmail.com
Levi Yant
levi.yant@nottingham.ac.uk

Specialty section:
This article was submitted to Plant Systematics and Evolution, a section of the journal Frontiers in Plant Science

Received: *20 June 2020*
Accepted: *23 November 2020*
Published: *16 December 2020*

Citation:
Konečná V, Yant L and Kolář F (2020) The Evolutionary Genomics of Serpentine Adaptation. Front. Plant Sci. 11:574616. doi: 10.3389/fpls.2020.574616

INTRODUCTION

Local adaptation optimizes fitness to the environment, often at the scale of meters. The resultant spatially varying selection leads to between-population genomic divergence that, depending on the intensity of gene flow, may maintain intraspecific adaptive diversity or lead to ecological speciation (Rundle and Nosil, 2005; Savolainen et al., 2013). In sessile plants, heterogeneous landscape mosaics, such as mountains or patchy soils, can trigger dramatic cases of local adaptation, especially in the presence of a steep gradient in the selective agent (Jain and Bradshaw, 1966). However, despite recent progress in the genomics of adaptation, there are still a few empirical inquiries into spatially varying selection (e.g., Hämälä et al., 2018; Hämälä and Savolainen, 2019) that provide empirical verification of the theory concerning adaptation under migration scenarios *via* finding correlation between

fitness and environmental factors underlying local selection in natural populations (Yeaman and Whitlock, 2011). Only few studies assess the fine-scale genomic architecture of complex adaptive traits adequately (Holliday et al., 2016).

Serpentine barrens (**Figure 1**) represent powerful models to understand genome modification to local conditions because their extreme chemical and physical properties act as strong, quantifiable selective pressures. Derived from ultramafic rocks, serpentine soils are highly skewed in their content of many elements, being typically: (i) low in macronutrients such as Ca, K, N, and P, (ii) high in metals Co, Cr, and Ni, and (iii) greatly reduced in Ca relative to Mg. Worldwide, it is this highly skewed Ca/Mg ratio that defines serpentines (O'Dell and Rajakaruna, 2011), despite considerable diversity in other qualities. To add insult to injury, serpentine soils are typically very porous, with low water holding capacity and, due to their dark color, are frequently prone to substrate over-heating (Proctor and Woodell, 1975; Brady et al., 2005). These multifarious chemical and physical characteristics together have been termed "the serpentine syndrome" (Jenny, 1980), a state that results in very low ecosystem productivity with low competition and frequent endemism (Kruckeberg, 1954; Whittaker, 1954; Brady et al., 2005; Harrison and Rajakaruna, 2011). Moreover, the island-like distribution of serpentines provides abrupt edaphic contrasts that are replicated frequently across landscapes, triggering parallel adaptation (Roberts and Proctor, 1992). Such natural replicates can be leveraged to discern consistent trends in mechanisms and genetic bases of adaptation, as well as ecological speciation (Rundle and Nosil, 2005; Losos, 2011).

The distinctive floristic composition of serpentines attracted botanists as early as the beginning of the twentieth century. What started as a general fascination with floristic peculiarities (Pančić, 1859; Lämmermayr, 1927; Novák, 1928; reviewed by Whittaker, 1954; Eggler, 1955) continued with experiments testing local adaptation (Kruckeberg, 1951, 1954) and targeted genetic investigations (Bradshaw, 2005; Bratteler et al., 2006). Contributions have emerged from a broad diversity of species, including *Achillea, Cerastium, Collinsia, Gilia, Helianthus, Knautia, Mimulus, Silene,* and *Streptanthus* (reviewed by Brady et al., 2005; O'Dell and Rajakaruna, 2011; Rajakaruna, 2018), significantly contributing to our understanding of the importance of local adaptation and ecotypic differentiation in plants.

FIGURE 1 | Extreme environment of serpentine barrens, illustrating sharp boundaries to adjacent land and low productivity (Pindos Mountains, Greece). Photos: F. Kolář.

As historical context is well summarized elsewhere (e.g., Brady et al., 2005; Anacker, 2014), we here focus on recent advances providing context for genomic studies. We first highlight advances in our understanding of serpentine adaptation at the phenotypic, physiological, and genomic levels. We discuss: (i) advances in knowledge of the selective factors imposed by serpentines, (ii) progress in experimental verification of local adaptation to serpentine soils, and (iii) plant responses to serpentines: the "phenotype" of serpentine adaptation. Finally, (iv) we summarize the first genomic studies that have very recently been built on previous insights and outline ways forward to integrate the study of serpentine adaptation.

DRIVERS OF SELECTION AT SERPENTINE SOILS

Given the heterogeneity between various serpentine sites, in order to understand the mechanistic basis of serpentine adaptation it is first necessary to define the exact selective agents in play for any given case. While serpentine syndrome represents a complex set of selection pressures that vary from site to site, since the 1950's there has been broad evidence that Ca availability plays a leading role (Vlamis and Jenny, 1948; Vlamis, 1949; Kruckeberg, 1954; Walker et al., 1955). More recently, the advent of high-throughput inductively coupled plasma mass spectrometry (ICP-MS; "ionomics" – Salt et al., 2008; Huang and Salt, 2016) has accelerated the characterization of that 'hidden half' of the plant environment: the underground soil matrix. Ionomics has allowed the rapid characterization of elemental accumulation in plant tissues in common garden experiments in diverse soil types from serpentines, to toxic mines, to saline soils (e.g., Arnold et al., 2016; Stein et al., 2017; Busoms et al., 2018; Preite et al., 2019). This and other recent advances in soil profiling have supported the salient role of distorted Ca/Mg ratio on serpentines (Palm and Van Volkenburgh, 2014), but has also identified other players, such as elevated heavy metal concentrations (Co, Cr, Ni, and/or Zn), for example in *Arabidopsis lyrata* (Veatch-Blohm et al., 2017) or *Knautia serpentinicola* (Čertner et al., 2018). These elevated metal levels are sometimes accompanied by lower concentrations of nutrients K, P, and S, as seen in *Helianthus exilis* (Sambatti and Rice, 2006), *Cerastium alpinum* (Berglund et al., 2004), and *Arabidopsis arenosa* (Arnold et al., 2016). Additional factors, such as drought, can interact with major chemical factors contributing to local adaptation to specific stresses, especially in drought-prone areas (e.g., Salehi Eskandari et al., 2017). The roles of biotic interactions, such as with bacteria, archaea, or mycorrhiza, are largely unknown with the little available evidence suggesting highly diverse effects (Mengoni et al., 2001; Pal and Paul, 2004; Davoodian et al., 2012; Doubková et al., 2012; reviewed by Schechter and Branco, 2014 and Southworth et al., 2014). For instance, higher diversity of arbuscular mycorrhizal fungi (AMF) was observed in serpentine populations of *Collinsia sparsiflora* compared to non-serpentine ones (Schechter and Bruns, 2008).

Further, AMF more efficiently promoted growth and P uptake in serpentine *K. serpentinicola* (Doubková et al., 2012). On the other hand, the bacterial communities from serpentine and non-serpentine soils in Northern California were not different from each other (Oline, 2006).

To add to this complexity, there commonly exists fine-scale, site-specific differences in the serpentine-defining factors themselves. Indeed, Berglund et al. (2004) leveraged such variability among different serpentine sites occupied by *C. alpinum* to demonstrate that strength of tolerance to Mg and Ni was related particularly to effective concentrations of these elements in soil at each site. Similarly, the relative roles of other components depend on particular species – or even the site – studied. For example, variation in B, Ca, Fe, Na, and Zn is observed between serpentine barrens harboring *Mimulus guttatus* (Selby and Willis, 2018). In line with this, differences in physical properties also seem to be regionally specific rather than a universal property of all serpentines. For example, while drought and erosion characterize serpentines in drought-prone regions such as California or the Middle East (Kruckeberg, 1984; Salehi-Eskandari et al., 2018), they do not distinguish serpentine and non-serpentine sites of otherwise similar geomorphology in Central and Northern Europe (Novák, 1928; Rune, 1953), (Teptina et al., 2018), and tropical regions in South and Southeast Asia (Galey et al., 2017).

In summary, the study of serpentine adaptation requires an initial decision: one must choose whether to address serpentine adaptation holistically (including physical properties, biotic interactions, and site-specific soil chemistry) or to instead focus on a universally dominating parameter (such as altered Ca/Mg ratio or elevated Ni content).

EXPERIMENTAL EVIDENCE FOR ADAPTATION TO SERPENTINE SOILS

A long tradition of reciprocal transplant experiments since the 1950's (e.g., Kruckeberg, 1950, 1951, 1954, 1967) provided broad evidence of local adaptation to serpentine soils. These approaches have recently been expanded to diverse species, e.g., *Helianthus exilis* (Sambatti and Rice, 2006), *Collinsia sparsiflora* (Wright et al., 2006), *Achillea millefolium* (O'Dell and Claassen, 2006), *Mimulus guttatus* (reviewed by Selby et al., 2014; Selby and Willis, 2018), and *Arabidopsis arenosa* (**Figure 2**). The observed adaptive differences were compromised of a wide range of fitness proxies, from the extent of juvenile mortality in *Mimulus*, to higher biomass production in *Achillea*, and seed production in *Helianthus*.

Taking our understanding of adaptive differences a step further, specific deconstruction of serpentine tolerance to individual elements has been performed in several contexts. Because serpentine chemical stress can be simply modeled, researchers have modulated the cardinal factors: Ca/Mg ratios and Ni levels in hydroponics and custom growth media (Proctor, 1971; Gabbrielli and Pandolfini, 1984; O'Dell and Rajakaruna, 2011). In *Knautia arvensis* and *Cerastium alpinum*, higher tolerance of serpentine populations to elevated Mg and Ni is evidenced by greater root growth (Berglund et al., 2004; Kolář et al., 2014). A specific effect of Ca/Mg ratio on both total biomass and photosynthetic rates was shown in *M. guttatus* (Palm et al., 2012). Furthermore, the effect of specific elements (Cr, Ni, and Ca/Mg ratio) on seed germination has been examined in *Arabidopsis lyrata* (Veatch-Blohm et al., 2013, 2017). There, while Ni caused a slower seedling growth, especially in non-serpentine accessions, there was no differential response to elevated Cr and Mg that could be related to fitness. Therefore, we can conclude that mechanisms reducing the Ni and/or Mg toxicity evolved in serpentine populations of many species; their phenotypes, such as length of the root, however, differ.

The origins of these phenotypes have been probed with genetic investigations of population history, which have documented striking manifold parallel colonizations of sites for a majority of the sufficiently sampled species or species groups (e.g., *Alyssum serpyllifolium*, Mengoni et al., 2003; Sobczyk et al., 2017; *C. alpinum*, Berglund et al., 2004; *Lasthenia californica* complex, Rajakaruna et al., 2003a; *M. guttatus*, Selby and Willis, 2018; *Minuartia verna* complex, Nunvářová Kabátová et al., 2019; *Solidago virgaurea*, Sakaguchi et al., 2017; *Streptanthus glandulosus* complex, Mayer and Soltis, 1994). In fact, single origins of serpentine populations are very rarely documented among species growing both on and off multiple serpentines (e.g., *Picris hieracioides*, Sakaguchi et al., 2018), and remains rather a property of genuine serpentine endemics (e.g., *K. serpentinicola*, Kolář et al., 2012, *Halacsya sendtneri*, Cecchi and Selvi, 2009). Unfortunately, only in a few cases have such genetic investigations been coupled with reciprocal transplants (Sakaguchi et al., 2017, 2019; Selby and Willis, 2018) or hydroponic experiments (Rajakaruna et al., 2003b,c; Berglund

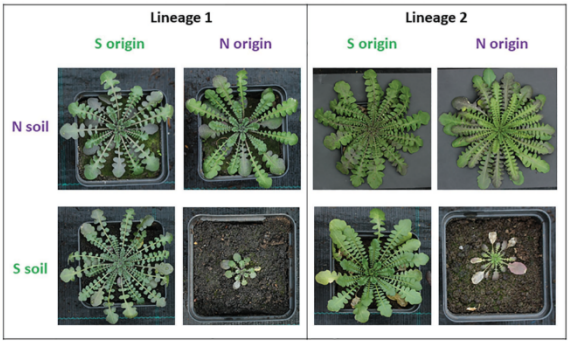


FIGURE 2 | Transplant experiment demonstrating parallel fitness response of two lineages of *Arabidopsis arenosa* to serpentine soil, depending on the substrate of origin. Two pairs of originally serpentine (S) and non-serpentine (N) populations representing different genetic lineages were cultivated in their native and foreign (native to the other member of the pair) soil, illustrative photo of representatives from all populations and treatments are depicted. Note the considerably smaller plants of non-serpentine origin when cultivated in serpentine soil but lack of such response for serpentine population in the non-serpentine soil suggesting the absence of the substrate-related trade-offs for serpentine plants. Photos by V. Konečná.

et al., 2004), ultimately demonstrating parallel substrate adaptation across multiple serpentine populations. Such rare cases provide particularly valuable naturally replicated model systems for further, finer-scale investigations of serpentine adaptation.

In other cases, serpentine adaptation is a constitutive trait present in all populations of a species regardless of their native soil chemistry. Such constitutive tolerance towards Ni has been reported for instance in *Silene dioica* (Westerbergh, 1994), *Noccaea goesingense* (Reeves and Baker, 1984), and *Noccaea montana* (Peer et al., 2006). Constitutive tolerance to both high Ni and low Ca/Mg ratios, at least as indicated by root growth, was documented for *Galium valdepilosum*, a likely "pre-adapted" species that colonized nearly all scattered serpentine outcrops throughout its overall species range (Kolář et al., 2014). Moreover, even plants that do not occur on serpentines but grow on dry and nutrient-poor habitats, such as granite outcrops, can tolerate extremely low Ca/Mg ratios, such as *Phacelia dubia* (Taylor and Levy, 2002). Overall, such cases of constitutive tolerance are only rarely documented (reviewed by O'Dell and Rajakaruna, 2011), potentially due to a publication bias towards "positive" results (i.e., where clear within-species local adaptation is evident).

MECHANISMS OF SERPENTINE ADAPTATION

Experimental studies have revealed a range of life-history and physiological mechanisms potentially underlying serpentine adaptation (Brady et al., 2005; Harrison and Rajakaruna, 2011; O'Dell and Rajakaruna, 2011). Drought stress adaptations include slower growth rate, reduced height, higher root/shoot biomass ratios, early flowering and specific modification of flowers (O'Dell and Rajakaruna, 2011; von Wettberg et al., 2014), and reduced leaf size and sclerophylly observed in some serpentine plants (Brady et al., 2005). In terms of chemistry, selective uptake of some micronutrients and macronutrients such as Ca and exclusion (or regulated accumulation and storage) of different metals are major adaptive mechanisms (Brady et al., 2005; Kazakou et al., 2008). Modulating the status of particular nutrients linked with monitoring uptake thus allows for valuable insight. For example, Kolář et al. (2014) compared serpentine populations of *Knautia serpentinicola* with non-serpentine populations of closely related *Knautia arvensis* in hydroponics with different concentrations of Ni and Mg. Interestingly, serpentine-origin plants accumulated less Ni when cultivated in a high Mg solution. On the contrary, non-serpentine plants accumulated approximately the same concentration of Ni regardless of the Mg concentration in media. Concordantly, in *A. lyrata* under Ni treatment, serpentine-origin plants had lower shoot and root Ni levels compared to non-serpentine plants (Veatch-Blohm et al., 2017).

Given that serpentine is a multi-hazard environment, it is only natural that serpentine adaptation results in physiological changes touching on a range of chemical challenges. A high-throughput approach to assessing this mixture is represented by performing common garden experiments incorporating a broad natural variation in a given species and to measure the relative accumulation of a panel of mineral nutrients. In this way, Arnold et al. (2016) assayed 20 elements in plants from 29 *A. arenosa* populations, after growing them in common conditions. This matrix was contrasted with data from source soil samples, providing a direct comparison of specific natural genetic variation across the ionome. Plants of serpentine origin accumulated the highest levels of K and S, excluded Ni, and exhibited the highest Ca/Mg ratios. These changes indicate a suite of specific, refined physiological adaptations in serpentine *A. arenosa*, which has a genetic basis. In this way, ionomics brings a key tool for understanding the actual "adaptive phenotype" distinguishing serpentine-adapted plants. Such approach, combined with common garden or transplant experiments, should become a standard practice in the characterization of the genetic basis of edaphic adaptation.

Adding to the complexity of multi-hazard adaptation, the rich experimental literature documents multiple solutions to identical environmental triggers (reviewed by Palm and Van Volkenburgh, 2014). For example, plants react to a skewed Ca/Mg ratio with a stunning diversity of mechanisms: either by selective translocation of Ca from roots to shoots (e.g., *A. millefolium*, O'Dell and Claassen, 2006), restriction of Mg uptake (e.g., *Gilia capitata*, Kruckeberg, 1951), or tolerance to higher concentrations of Mg in shoots (e.g., *Streptanthus polygaloides*, Boyd et al., 2009). Analogously, serpentine-adapted plants respond to elevated levels of heavy metals such as Ni in soil either by restricted uptake to shoots (Doubková et al., 2012; Kolář et al., 2014; Salehi Eskandari et al., 2017) or tolerance to high Ni levels in tissues (particularly in metal hyperaccumulators, e.g., Assunção et al., 2003; Galardi et al., 2007; Leigh Broadhurst et al., 2009).

While we have learned much about different adaptation mechanisms at a species level, considerably less is known about the variation in these traits at a population level within species. Experiments leveraging multiple cases of repeated within-species colonization of serpentine patches have demonstrated a broad variation in the strength of responses in a handful of studies (e.g., Berglund et al., 2004; Galardi et al., 2007; Veatch-Blohm et al., 2017). Together with variation in the soil chemistry across serpentine barrens, this suggests that there may be independent solutions even within a species, each fine-tuned to particular conditions at each site.

In summary, because serpentines produce clear challenges that can be experimentally dissected into specific factors (and moreover, which are replicated within species), decades of research have prepared a solid foundation for studies of their genomic basis.

FIRST INSIGHTS INTO THE GENOMIC BASIS OF SERPENTINE ADAPTATION

Given the multi-challenge nature of serpentines and the diverse adaptive phenotypes generated in response, we may expect a highly complex, polygenic basis. Further, the variation in elemental soil composition between serpentine barrens suggests

no single "basis" of serpentine adaptation. On the other hand, as compared to other multi-hazard "extreme" environments (coastal, alpine, or high-arctic sites), serpentines are fairly well-defined, dominated by the effect of few major elements. This, together with strong selective pressures, makes discovery of major-effect candidates seem likely, similar to what has been found for other soils such as metal contaminated and saline sites (e.g., in *Arabidopsis halleri*, Courbot et al., 2007; Willems et al., 2007; in *Noccaea caerulescens*, Deniau et al., 2006; and in *Mimulus guttatus*, Wright et al., 2013). In addition, theory suggests that specific aspects of serpentine colonization – such as abrupt fitness differences and patchily-distributed habitats – favor the emergence of large-effect alleles (Dittmar et al., 2016; Gilbert and Whitlock, 2017). Theory also suggests that adaptation in the face of gene flow, such as that from nearby non-serpentine sites, may promote fixation of smaller numbers of large-effect loci (Yeaman and Whitlock, 2011).

In line with this, the few quantitative trait locus (QTL) studies applied to serpentine ecotypes so far have provided evidence for a simple genetic architecture of single traits (e.g., Ni tolerance in *Silene vulgaris*, Bratteler et al., 2006 and *Caulanthus amplexicaulis*, Burrell et al., 2012). Additionally, studies focused on particular genes known to have an ion homeostasis effect show strong natural differentiation in its sequence variation and/or in associated phenotypic responses (e.g., *A. serpyllifolium*, Sobczyk et al., 2017; *Arabidopsis thaliana*, Bradshaw, 2005; Agrawal et al., 2012). While such hypothesis-driven studies bring valuable insights into the basis of a particular gene or trait, they capture neither genetic architecture nor the complexity of the full serpentine syndrome. Accordingly, neither classical QTL studies nor candidate gene-targeted inquiries can alone provide a picture of the broad genomic remodeling which we speculate is required for robust establishment in such a multi-hazard environment.

In contrast to QTL studies, high-density genomic divergence scans detect signatures of directional selection in a purely natural system from a holistic perspective, both in terms of genes (the entire genome) and parameters screened (the entire serpentine syndrome in nature: both known and unknown factors). In other words, genome scans can be both genetically and phenotypically agnostic. Such scans have been performed in two wild, outcrossing *Arabidopsis* species – *A. arenosa* (Arnold et al., 2016) and *A. lyrata* (Turner et al., 2008, 2010). Both indicate a highly polygenic basis of serpentine adaptation. The study by Turner et al. (2010), notably one of the first truly genome-scale scans for selection used sequencing of pooled samples to reveal outlier differentiated SNPs at loci involved in ion transmembrane transport, metal tolerance, and calcium ion binding. Clearer detection of candidate loci was achieved by individual-level genome resequencing by Arnold et al. (2016), discriminating approximately 160 genes exhibiting multiple signatures of selection in a serpentine-adapted population of *A. arenosa*. These included genes related to Ca signaling, ion homeostasis, metal transport, root macronutrient transport, and dehydration tolerance. These works provided a broad view on the genetic basis of serpentine adaptation and have served as hypothesis generators that can now guide functional assessment of particular alleles in natural conditions.

A complementary study in *M. guttatus* innovatively combined field and genomic approaches by assessing survival differences of F2 mapping populations grown on serpentines by bulk segregant analysis (Selby and Willis, 2018). This enabled the identification of a major QTL (containing several 100 candidate genes) that contributed to the survival of plants by 33%. However, because of the limited size of the survivor pool, the study was underpowered to detect smaller effect loci. Indeed, mapping approaches commonly fail to identify small effect loci and tend to overestimate the influence of large effect loci due to the linkage (Rockman, 2012). Nevertheless, this study was an important step towards linking genomic variation and fitness consequences and demonstrates that the *Mimulus* system has a strong potential to reveal more refined results in the future (Selby et al., 2014).

GENOMICS OF ECOLOGICAL SPECIATION ON SERPENTINE

Ecological speciation can occur when strong divergent selection is associated with the rise of reproductive incompatibilities between ecologically distinct populations (Nosil et al., 2009). Such reproductive incompatibilities may emerge in association with serpentine either as a direct consequence of selection against maladaptive gene flow (e.g., reinforcement to avoid hybrids that are unfit in either environment) or as a by-product of local adaptation, e.g., through physical linkage (as observed in copper-tolerant *Mimulus*, Rajakaruna and Whitton, 2004; Wright et al., 2013) or shifts in flowering time (found in serpentine *Solidago*, Sakaguchi et al., 2018, 2019). Indeed, serpentine barrens provide intriguing candidate cases of incipient – and even parallel – ecological speciation (e.g., Rajakaruna and Whitton, 2004; Moyle et al., 2012; reviewed by Ostevik et al., 2012). The genomic basis of these incompatibilities remains unknown. Aeschbacher et al. (2017) used coalescent modeling on pooled sequence data to detect genome-wide signals of selection against maladaptive gene flow from non-serpentine *M. guttatus* to serpentine populations. Whether such selection promotes accumulation of reproductive barriers through reinforcement or is directly linked with between ecotype-specific reproductive incompatibilities is unknown. However, given experimental evidence for the accumulation of reproductive incompatibilities associated with colonization of serpentines across plant taxa (reviewed by Rajakaruna and Whitton, 2004; Rajakaruna, 2018) the genomics of ecological speciation in serpentine plants promises to be a fruitful area for future research.

CONVERGENCE AND PARALLELISM IN SERPENTINE ADAPTATION

The patchy, island-like distribution of serpentines provides convenient natural replicates to study repeated evolution both within species (parallelism) and between them (convergence). Given the many independent evolutionary paths that species (and indeed populations within species) have taken to serpentine-adaptive phenotypes, it is reasonable to expect

reuse of common pathways, or even proximal molecular actors to the common serpentine challenges.

While phenotypic parallelism and convergence are obvious from a rich literature, only recently has genetic parallelism and convergence been tested. Genetic parallelism has been indicated in *M. guttatus*, where two serpentine populations share a major QTL important for serpentine adaptation (Selby and Willis, 2018). Some indication of genetic parallelism has also been provided by Turner et al. (2010) in European and North American serpentine populations of *A. lyrata*, which shared the same non-synonymous mutation that codes for a change in the *TPC1* locus encoding Ca-ion channel. This is almost fixed in three serpentine populations in two continents. Genetic convergence was described in *A. arenosa* and *A. lyrata* by Arnold et al. (2016) in nine gene coding loci with serpentine-relevant predicted functions, such as Ca, K, and Ni homeostasis (**Table 1**). Importantly, studies of evolutionary history and fitness responses of several other plant systems, such as *C. alpinum* (Scandinavia – Berglund et al., 2004), *L. californica* complex (California – Rajakaruna et al., 2003a,b,c), and *Solidago virgaurea* (Japan – Sakaguchi et al., 2017, 2019), suggest parallel evolution of serpentine ecotypes, which may be a frequent phenomenon (see also **Figure 2**). Further leveraging such truly non-model systems for genomic studies should bring novel vital insights into the basis of local adaptation.

An intriguing open question concerns the origin of the variants that repeatedly exhibit signatures of selection in different serpentine populations. Such repeated signatures may arise in three major ways (Stern, 2013; Lee and Coop, 2019): (i) the causal variants might have been repeatedly sampled from the standing variation present in the ancestral population; (ii) they may have been transferred between adapted populations by gene flow; or (iii) they may have arisen by independent *de novo* mutations. That adaptive gene flow may be involved in serpentine adaptation has been indicated in *A. arenosa* and *A. lyrata* despite overall weak genome-wide signal for introgression between these two species (Arnold et al., 2016). Yet, the relative contribution of these pathways to repeated adaptation remains unknown, particularly in closely related populations and thus represents a fruitful area for future research.

CONCLUSION

Given the perfect storm of challenges that serpentine presents, a comprehensive understanding of serpentine adaptation is a challenging task that requires multidisciplinary investigation (Wright and Von Wettberg, 2009). However, recent progress in quantitative genetic, ionomic, and population genomic study has shown that understanding the basis for the interacting phenotypes which constitute serpentine adaptation is well worth the effort. Specifically, recent inquiries in the *Arabidopsis* and *Mimulus* genera have provided proof-of-concept that the genetic basis of serpentine adaptation is accessible, and the time is now ripe for synthetic studies (**Figure 3**). Here, we sketched gaps in our understanding of serpentine adaptation and considered integrative approaches to link candidate adaptive alleles with fitness effects as a promising avenue to make progress.

We suggest that ongoing ecological and population genomic studies of serpentine adaptation hold strong potential to contribute to our understanding of fundamental evolutionary principles, similar to the contributions of early eco-evolutionary studies achieved in the 1950s and 1960s (Kruckeberg, 1950, 1951, 1954, 1967). Aside from the obvious benefit of improved knowledge about soil-stress adaptations for rational crop breeding, serpentine models can efficiently address our understanding of environmentally driven adaptation and ecological speciation. Specifically, serpentine soils represent a strong selective pressure leading to well-tractable adaptive phenotypes. Yet, in contrast to similarly poised mine sites, serpentines provide fully natural setups with longer time frames on which selection has acted.

TABLE 1 | Candidate convergent loci mediating serpentine adaptation in *Alyssum serpyllifolium, Arabidopsis arenosa, Arabidopsis lyrata, and Mimulus guttatus*.

Species	*Arabidopsis thaliana* homolog	Potential function[1]
Arabidopsis arenosa[2] and *Arabidopsis lyrata*[3]	*AT1G51310*	A tRNA-methyltransferase.
	AT3G10985	A senescence-associated gene. Expression is induced in response to treatment with Nep1, a fungal protein that causes necrosis. The mRNA is cell-to-cell mobile.
	AT5G04320	Encodes a protein that protects meiotic centromere cohesion.
	AT5G04330	Cytochrome P450 superfamily protein.
	AT4G03560	*TPC1*, a vaculolar Ca^{2+} channel.
	AT4G19960	*KUP9*, a K+ ion transmembrane transporter that can also mediate Cs uptake if expressed in *Escherichia coli*.
	AT5G37710	Alpha/beta-Hydrolases superfamily protein.
	AT5G37720	*ALY4*, a member of a protein family involved in RNA export from the nucleus and transcriptional coactivation.
A. arenosa[2], *A. lyrata*[3], and *Alyssum serpyllifolium*[4]	*AT5G03570*	*FPN2*, a tonoplast-localized Ni transporter.
A. lyrata[3] and *Mimulus guttatus*[5]	*AT4G19670*	RING/U-box superfamily protein.

[1]*The Arabidopsis Information Resource (TAIR), www.arabidopsis.org, July, 2019.*
[2]*(Arnold et al., 2016).*
[3]*(Turner et al., 2010).*
[4]*(Sobczyk et al., 2017).*
[5]*(Selby, 2014; Selby and Willis, 2018).*

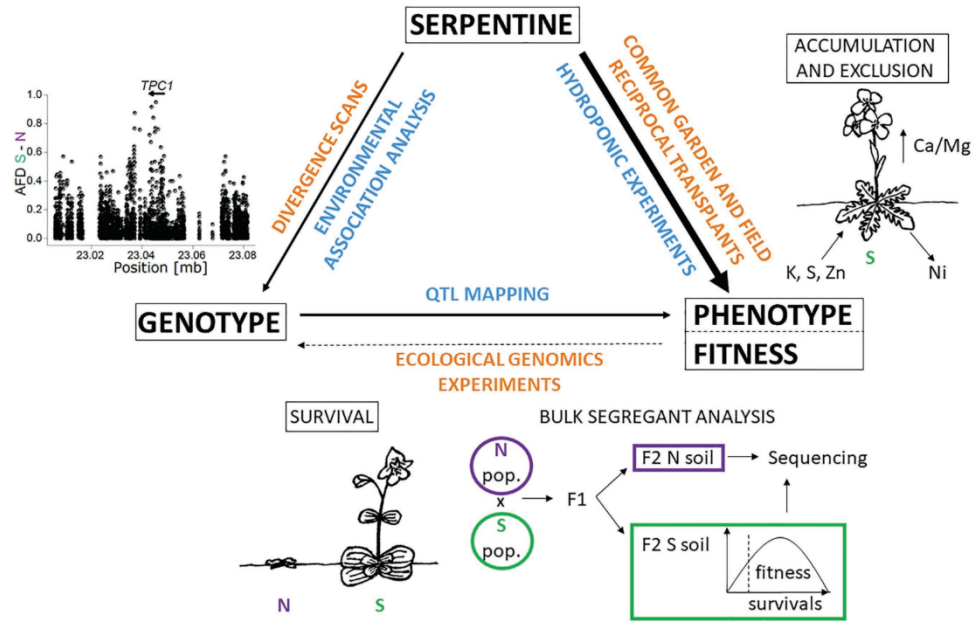


FIGURE 3 | Approaches to study the basis of serpentine adaptation: Integration of holistic (orange) and one factor/hypothesis driven approaches (blue) to investigate links between genotype, phenotype, fitness, and serpentine environments. The right side of the triangle depicts ionomic analysis of *Arabidopsis arenosa* (Arnold et al., 2016), the bottom side is accompanied by the scheme of bulk segregant analysis applied in the recent study of *Mimulus guttatus* (Selby and Willis, 2018), and the left side is supplemented with allele frequency difference graph (AFD) of S and N population for *TPC1* locus Arnold et al. (2016). Line thickness scales with the quantity of empirical studies available. S – serpentine adapted plant, N – non-serpentine plant. The scheme is inspired by Sork et al. (2013) and Rellstab et al. (2015). Illustrations by V. Konečná.

Frequent reports of serpentine endemism and the reproductive isolation experienced by edaphic ecotypes also suggest serpentine plants as attractive, yet still too rarely leveraged models for ecological speciation. Beyond this, serpentine is a highly defined environment characterized primarily by a limited number of soil chemistry components (also with potentially linked biotic interactions). Such "intermediate complexity" makes serpentine an attractive model of adaptation towards a suite of naturally relevant factors, which can still be experimentally manipulated using, e.g., reciprocal transplants. Finally, the repeated origins of serpentine adaptation allow leveraging natural replicates empowering genomic and experimental inference. Genome-wide studies of serpentine populations combined with ecological genomic experiments and/or forward genetic validation may thus bring significant contribution to our general knowledge on the adaptation mechanisms towards environmental challenges.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work and approved it for publication.

FUNDING

The study was supported by the Charles University, project GA UK No. 410120 to VK and the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme (grant number ERC-StG 679056 HOTSPOT), *via* a grant to LY. Additional support was provided by Charles University (Primus/SCI/35 to FK) and by long-term research development project No. RVO 67985939 of the Czech Academy of Sciences.

REFERENCES

Aeschbacher, S., Selby, J. P., Willis, J. H., and Coop, G. (2017). Population-genomic inference of the strength and timing of selection against gene flow. *Proc. Natl. Acad. Sci. U. S. A.* 114, 7061–7066. doi: 10.1073/pnas.1616755114

Agrawal, B., Lakshmanan, V., Kaushik, S., and Bais, H. P. (2012). Natural variation among *Arabidopsis* accessions reveals malic acid as a key mediator of nickel (Ni) tolerance. *Planta* 236, 477–489. doi: 10.1007/s00425-012-1621-2

Anacker, B. L. (2014). The nature of serpentine endemism. *Am. J. Bot.* 101, 219–224. doi: 10.3732/ajb.1300349

Arnold, B. J., Lahner, B., DaCosta, J. M., Weisman, C. M., Hollister, J. D., Salt, D. E., et al. (2016). Borrowed alleles and convergence in serpentine adaptation. *Proc. Natl. Acad. Sci. U. S. A.* 113, 8320–8325. doi: 10.1073/pnas.1600405113

Assunção, A. G. L., Bookum, W. M., Nelissen, H. J. M., Vooijs, R., Schat, H., and Ernst, W. H. O. (2003). Differential metal-specific tolerance and accumulation patterns among *Thlaspi caerulescens* populations originating from different soil types. *New Phytol.* 159, 411–419. doi: 10.1046/j.1469-8137.2003.00819.x

Berglund, A. N., Dahlgren, S., and Westerbergh, A. (2004). Evidence for parallel evolution and site-specific selection of serpentine tolerance in *Cerastium alpinum* during the colonization of Scandinavia. *New Phytol.* 161, 199–209. doi: 10.1046/j.1469-8137.2003.00934.x

Boyd, R. S., Wall, M. A., Santos, S. R., and Davis, M. A. (2009). Variation of morphology and elemental concentrations in the California nickel hyperaccumulator *Streptanthus polygaloides* (Brassicaceae). *Northeast. Nat.* 16, 21–38. doi: 10.1656/045.016.0503

Bradshaw, H. D. (2005). Mutations in CAX1 produce phenotypes characteristic of plants tolerant to serpentine soils. *New Phytol.* 167, 81–88. doi: 10.1111/j.1469-8137.2005.01408.x

Brady, K. U., Kruckeberg, A. R., and Bradshaw Jr., H. D. (2005). Evolutionary ecology of plant adaptation to serpentine soils. *Annu. Rev. Ecol. Evol. Syst.* 36, 243–266. doi: 10.1146/annurev.ecolsys.35.021103.105730

Bratteler, M., Baltisberger, M., and Widmer, A. (2006). QTL analysis of intraspecific differences between two *Silene vulgaris* ecotypes. *Ann. Bot.* 98, 411–419. doi: 10.1093/aob/mcl113

Burrell, M. A., Hawkins, A. K., and Pepper, A. E. (2012). Genetic analyses of nickel tolerance in a north American serpentine endemic plant, *Caulanthus amplexicaulis var. barbarae* (Brassicaceae). *Am. J. Bot.* 99, 1875–1883. doi: 10.3732/ajb.1200382

Busoms, S., Paajanen, P., Marburger, S., Bray, S., Huang, X. Y., Poschenrieder, C., et al. (2018). Fluctuating selection on migrant adaptive sodium transporter alleles in coastal *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U. S. A.* 115, E12443–E12452. doi: 10.1073/pnas.1816964115

Cecchi, L., and Selvi, F. (2009). Phylogenetic relationships of the monotypic genera Halacsya and Paramoltkia and the origins of serpentine adaptation in circummediterranean Lithospermeae (Boraginaceae): insights from ITS and matK DNA sequences. *Taxon* 58, 700–714. doi: 10.1002/tax.583002

Čertner, M., Sudová, R., Weiser, M., Suda, J., and Kolář, F. (2018). Ploidy-altered phenotype interacts with local environment and may enhance polyploid establishment in *Knautia serpentinicola* (Caprifoliaceae). *New Phytol.* 221, 1117–1127. doi: 10.1111/nph.15426

Courbot, M., Willems, G., Motte, P., Arvidsson, S., Roosens, N., Saumitou-laprade, P., et al. (2007). A major quantitative trait locus for cadmium tolerance in *Arabidopsis halleri* colocalizes with HMA4, a gene encoding a heavy metal ATPase 1 [OA]. *Plant Physiol.* 144, 1052–1065. doi: 10.1104/pp.106.095133

Davoodian, N., Bosworth, J., and Rajakaruna, N. (2012). Mycorrhizal colonization of *Hypericum perforatum* L. (Hypericaceae) from serpentine and granite outcrops on the deer isles, Maine. *Northeast. Nat.* 19, 517–526. doi: 10.1656/045.019.0312

Deniau, A. X., Pieper, B., Ten Bookum, W. M., Lindhout, P., Aarts, M. G. M., and Schat, H. (2006). QTL analysis of cadmium and zinc accumulation in the heavy metal hyperaccumulator Thlaspi caerulescens. *Theor. Appl. Genet.* 113, 907–920. doi: 10.1007/s00122-006-0350-y

Dittmar, E. L., Oakley, C. G., Conner, J. K., Gould, B. A., and Schemske, D. W. (2016). Factors influencing the effect size distribution of adaptive substitutions. *Proc. R. Soc. B Biol. Sci.* 283:20153065. doi: 10.1098/rspb.2015.3065

Doubková, P., Suda, J., and Sudová, R. (2012). The symbiosis with arbuscular mycorrhizal fungi contributes to plant tolerance to serpentine edaphic stress. *Soil Biol. Biochem.* 44, 56–64. doi: 10.1016/j.soilbio.2011.09.011

Eggler, J. (1955). Ein Beitrag zur Serpentinvegetation in der Gulsen bei Kraubath in Obersteiermark. Mitt Naturw Ver Steiermark 85, 27–72.

Gabbrielli, R., and Pandolfini, T. (1984). Effect of Mg2+ and Ca2+ on the response to nickel toxicity in a serpentine endemic and nickel-accumulating species. *Physiol. Plant.* 62, 540–544. doi: 10.1111/j.1399-3054.1984.tb02796.x

Galardi, F., Corrales, I., Mengoni, A., Pucci, S., Barletti, L., Barzanti, R., et al. (2007). Intra-specific differences in nickel tolerance and accumulation in the Ni-hyperaccumulator *Alyssum bertolonii*. *Environ. Exp. Bot.* 60, 377–384. doi: 10.1016/j.envexpbot.2006.12.011

Galey, M. L., van der Ent, A., Iqbal, M. C. M., and Rajakaruna, N. (2017). Ultramafic geoecology of south and Southeast Asia. *Bot. Stud.* 58:18. doi: 10.1186/s40529-017-0167-9

Gilbert, K. J., and Whitlock, M. C. (2017). The genetics of adaptation to discrete heterogeneous environments: frequent mutation or large-effect alleles can allow range expansion. *J. Evol. Biol.* 30, 591–602. doi: 10.1111/jeb.13029

Hämälä, T., Mattila, T. M., and Savolainen, O. (2018). Local adaptation and ecological differentiation under selection, migration, and drift in *Arabidopsis lyrata*. *Evolution* 72, 1373–1386. doi: 10.1111/evo.13502

Hämälä, T., and Savolainen, O. (2019). Genomic patterns of local adaptation under gene flow in *Arabidopsis lyrata*. *Mol. Biol. Evol.* 36, 2557–2571. doi: 10.1093/molbev/msz149

Harrison, S., and Rajakaruna, N. (2011). *Serpentine: The evolution and ecology of a model system*. University of California Press.

Holliday, J. A., Zhou, L., Bawa, R., Zhang, M., and Oubida, R. W. (2016). Evidence for extensive parallelism but divergent genomic architecture of adaptation along altitudinal and latitudinal gradients in *Populus trichocarpa*. *New Phytol.* 209, 1240–1251. doi: 10.1111/nph.13643

Huang, X. Y., and Salt, D. E. E. (2016). Plant ionomics: from elemental profiling to environmental adaptation. *Mol. Plant* 9, 787–797. doi: 10.1016/j.molp.2016.05.003

Jain, S. K., and Bradshaw, A. D. (1966). Evolutionary divergence among adjacent plant populations I. the evidence and its theoretical analysis. *Heredity* 21, 407–441. doi: 10.1038/hdy.1966.42

Jenny, H. (1980). *The soil resource. Origin of behaviour*. New York: Springer.

Kazakou, E., Dimitrakopoulos, P. G., Baker, A. J. M., Reeves, R. D., and Troumbis, A. Y. (2008). Hypotheses, mechanisms and trade-offs of tolerance and adaptation to serpentine soils: from species to ecosystem level. *Biol. Rev. Camb. Philos. Soc.* 83, 495–508. doi: 10.1111/j.1469-185X.2008.00051.x

Kolář, F., Dortová, M., Lepš, J., Pouzar, M., and Krejčová, A. (2014). Serpentine ecotypic differentiation in a polyploid plant complex: shared tolerance to Mg and Ni stress among di- and tetraploid serpentine populations of. *Plant Soil* 374, 435–447. doi: 10.1007/s11104-013-1813-y

Kolář, F., Fér, T., Štech, M., Trávníček, P., Dušková, E., Schönswetter, P., et al. (2012). Bringing together evolution on serpentine and polyploidy: spatiotemporal history of the diploid-tetraploid complex of *Knautia arvensis* (Dipsacaceae). *PLoS One* 7:e39988. doi: 10.1371/journal.pone.0039988

Kruckeberg, A. R. (1950). *An experimental inquiry into the nature of endemism on serpentine soils*. Berkeley: University of California.

Kruckeberg, A. R. (1951). Intraspecific variability in the response of certain native plant species to serpentine soil. *Am. J. Bot.* 38, 408–419. doi: 10.1002/j.1537-2197.1951.tb14842.x

Kruckeberg, A. R. (1954). The ecology of serpentine soils: a symposium. III. Plant species in relation to serpentine soils. *Ecology* 35, 267–274.

Kruckeberg, A. R. (1967). Ecotypic response to ultramafic soils by some plant species of northwestern United States. *Brittonia* 19, 133–151. doi: 10.2307/2805271

Kruckeberg, A. R. (1984). *California serpentines: Flora, vegetation, geology, soils and management problems*. University of California Press.

Lämmermayr, L. (1927). Materialien zur Systematik und Ökologie der Serpentinflora. II. Das Problem der, Serpentinpflanzen "–Eine kritische ökologische Studie, Sitzungsber. Akad. Wiss. Wien, math.-naturw. Kl., Abt., 25–69.

Lee, K. M., and Coop, G. (2019). Population genomics perspectives on convergent adaptation. *Philos. Trans. R Soc. Lond. B Biol. Sci.* 374, 1–19. doi: 10.1098/rstb.2018.0236

Leigh Broadhurst, C., Tappero, R. V., Maugel, T. K., Erbe, E. F., Sparks, D. L., and Chaney, R. L. (2009). Interaction of nickel and manganese in accumulation and localization in leaves of the Ni hyperaccumulators Alyssum murale and *Alyssum corsicum*. *Plant Soil* 314, 35–48. doi: 10.1007/s11104-008-9703-4

Losos, J. B. (2011). Convergence, adaptation, and constraint. *Evolution* 65, 1827–1840. doi: 10.1111/j.1558-5646.2011.01289.x

Mayer, M., and Soltis, P. S. (1994). The evolution of serpentine endemics : a chloroplast DNA phylogeny of the *Streptanthus glandulosus* complex (Cruciferae). *Syst. Bot.* 19, 557–574. doi: 10.2307/2419777

Mengoni, A., Baker, A. J. M., Bazzicalupo, M., Reeves, R. D., Adigüzel, N., Chianni, E., et al. (2003). Evolutionary dynamics of nickel hyperaccumulation in alyssum revealed by ITS nrDNA analysis. *New Phytol.* 159, 691–699. doi: 10.1046/j.1469-8137.2003.00837.x

Mengoni, A., Barzanti, R., Gonnelli, C., Gabbrielli, R., and Bazzicalupo, M. (2001). Characterization of nickel-resistant bacteria isolated from serpentine soil. *Environ. Microbiol.* 3, 691–698. doi: 10.1046/j.1462-2920.2001.00243.x

Moyle, L. C., Levine, M., Stanton, M. L., and Wright, J. W. (2012). Hybrid sterility over tens of meters between ecotypes adapted to serpentine and non-serpentine soils. *Evol. Biol.* 39, 207–218. doi: 10.1007/s11692-012-9180-9

Nosil, P., Funk, D. J., and Ortiz-Barrientos, D. (2009). Divergent selection and heterogeneous genomic divergence. *Mol. Ecol.* 18, 375–402. doi: 10.1111/j.1365-294X.2008.03946.x

Novák, F. A. (1928). Ekologické úvahy o hadcových rasách a hadcové vegetaci. Věda přírodní 9.

Nunvářová Kabátová, K., Kolář, F., Jarolímová, V., Krak, K., and Chrtek, J. (2019). Does geography, evolutionary history or ecology drive ploidy and genome size variation in the *Minuartia verna* group (Caryophyllaceae) across Europe? *Plant Syst. Evol.* 305, 1019–1040. doi: 10.1007/s00606-019-01621-2

O'Dell, R. E., and Claassen, V. P. (2006). Serpentine and nonserpentine *Achillea millefolium* accessions differ in serpentine substrate tolerance and response to organic and inorganic amendments. *Plant Soil* 279, 253–269. doi: 10.1007/s11104-005-2360-y

O'Dell, R. E., and Rajakaruna, N. (2011). "Intraspecific variation, adaptation, and evolution" in *Serpentine: Evolution and ecology in a model system*. eds. S. Harrison and N. Rajakaruna (University of California Press), 97–137.

Oline, D. K. (2006). Phylogenetic comparisons of bacterial communities from serpentine and nonserpentine soils. *Appl. Environ. Microbiol.* 72, 6965–6971. doi: 10.1128/AEM.00690-06

Ostevik, K. L., Moyers, B. T., Owens, G. L., and Rieseberg, L. H. (2012). Parallel ecological speciation in plants? *Int. J. Ecol.* 2012, 1–17. doi: 10.1155/2012/939862

Pal, A., and Paul, A. K. (2004). Aerobic chromate reduction by chromium-resistant bacteria isolated from serpentine soil. *Microbiol. Res.* 159, 347–354. doi: 10.1016/j.micres.2004.08.001

Palm, E., Brady, K., and Van Volkenburgh, E. (2012). Serpentine tolerance in *Mimulus guttatus* does not rely on exclusion of magnesium. *Funct. Plant Biol.* 39, 679–688. doi: 10.1071/FP12059

Palm, E. R., and Van Volkenburgh, E. (2014). "Physiological adaptation of plants to serpentine soil" in *Plant ecology and evolution in harsh environments*. eds. N. Rajakaruna, R. S. Boyd and T. B. Harris (New York: Nova Science Publishers), 129–148.

Pančić, J. (1859). Die Flora der Serpentinberge in Mittel-Serbien. Verhandl. Zool.-Bot. Gesell. Wien 9, 139–150.

Peer, W. A., Mahmoudian, M., Freeman, J. L., Lahner, B., Richards, E. L., Reeves, R. D., et al. (2006). Assessment of plants from the Brassicaceae family as genetic models for the study of nickel and zinc hyperaccumulation. *New Phytol.* 172, 248–260. doi: 10.1111/j.1469-8137.2006.01820.x

Preite, V., Sailer, C., Syllwasschy, L., Bray, S., Kraemer, U., and Yant, L. (2019). Convergent evolution in *Arabidopsis halleri* and *Arabidopsis arenosa* on calamine metalliferous soils. *Philos. Trans. R. Soc. B* 374:20180243. doi: 10.1098/rstb.2018.0243

Proctor, J. (1971). The plant ecology of serpentine: II. *J. Ecol.* 59, 397–410. doi: 10.2307/2258320

Proctor, J., and Woodell, S. R. (1975). The ecology of serpentine soils. *Adv. Ecol. Res.* 9, 255–366.

Rajakaruna, N. (2018). Lessons on evolution from the study of edaphic specialization. *Bot. Rev.* 84, 39–78. doi: 10.1007/s12229-017-9193-2

Rajakaruna, N., Baldwin, B. G., Chan, R., Desrochers, A. M., Bohm, B. A., and Whitton, J. (2003a). Edaphic races and phylogenetic taxa in the *Lasthenia californica* complex (Asteraceae: Heliantheae): an hypothesis of parallel evolution. *Mol. Ecol.* 12, 1675–1679. doi: 10.1046/j.1365-294x.2003.01843.x

Rajakaruna, N., Bradfield, G. E., Bohm, B. A., and Whitton, J. (2003b). Adaptive differentiation in response to water stress by edaphic races of *Lasthenia californica* (Asteraceae). *Int. J. Plant Sci.* 164, 371–376. doi: 10.1086/368395

Rajakaruna, N., Siddiqi, M. Y., Whitton, J., Bohm, B. A., and Glass, A. D. M. (2003c). Differential responses to Na+/K+ and Ca2+/Mg2+ in two edaphic races of the *Lasthenia californica* (Asteraceae) complex: a case for parallel evolution of physiological traits. *New Phytol.* 157, 93–103. doi: 10.1046/j.1469-8137.2003.00648.x

Rajakaruna, N., and Whitton, J. (2004). "Trends in the evolution of edaphic specialists with an example of parallel evolution in the *Lasthenia californica* complex" in *Plant adaptation: Molecular genetics and ecology*. NRC Research Press, 103–110.

Reeves, R. D., and Baker, A. J. M. (1984). Studies on metal uptake by plants from serpentine and non-serpentine populations of Thlaspi goesingense Hálácsy (Crycuferae). *New Phytol.* 98, 191–204. doi: 10.1111/j.1469-8137.1984.tb06108.x

Rellstab, C., Gugerli, F., Eckert, A. J., Hancock, A. M., and Holderegger, R. (2015). A practical guide to environmental association analysis in landscape genomics. *Mol. Ecol.* 24, 4348–4370. doi: 10.1111/mec.13322

Roberts, B. A., and Proctor, J. (1992). *The ecology of areas with Serpentinized rocks. A World View*. Dordrecht: Kluwer Academic Press.

Rockman, M. V. (2012). The QTN program and the alleles that matter for evolution: all that's gold does not glitter. *Evolution* 66, 1–17. doi: 10.1111/j.1558-5646.2011.01486.x

Rundle, H. D., and Nosil, P. (2005). Ecological speciation. *Ecol. Lett.* 8, 336–352. doi: 10.1111/j.1461-0248.2004.00715.x

Rune, O. (1953). Plant life on serpentines and related rocks in the north of Sweden. Sv. växtgeografiska sällsk.

Sakaguchi, S., Horie, K., Ishikawa, N., Nagano, A. J., Yasugi, M., Kudoh, H., et al. (2017). Simultaneous evaluation of the effects of geographic, environmental and temporal isolation in ecotypic populations of *Solidago virgaurea*. *New Phytol.* 216, 1268–1280. doi: 10.1111/nph.14744

Sakaguchi, S., Horie, K., Ishikawa, N., Nishio, S., Worth, J. R. P., Fukushima, K., et al. (2019). Maintenance of soil ecotypes of *Solidago virgaurea* in close parapatry via divergent flowering time and selection against immigrants. *J. Ecol.* 107, 418–435. doi: 10.1111/1365-2745.13034

Sakaguchi, S., Horie, K., Kimura, T., Nagano, A. J., Isagi, Y., and Ito, M. (2018). Phylogeographic testing of alternative histories of single-origin versus parallel evolution of early flowering serpentine populations of *Picris hieracioides* L. (Asteraceae) in Japan. *Ecol. Res.* 33, 537–547. doi: 10.1007/s11284-017-1536-2

Salehi Eskandari, B., Ghaderian, S. M., and Schat, H. (2017). The role of nickel (Ni) and drought in serpentine adaptation: contrasting effects of Ni on osmoprotectants and oxidative stress markers in the serpentine endemic, *Cleome heratensis*, and the related non-serpentinophyte, *Cleome foliolosa*. *Plant Soil* 417, 183–195. doi: 10.1007/s11104-017-3250-9

Salehi-Eskandari, B., Ghaderian, S. M., and Schat, H. (2018). Differential interactive effects of the Ca/Mg quotient and PEG-simulated drought in Alyssum inflatum and Fortuynia garcinii. *Plant Soil* 428, 213–222. doi: 10.1007/s11104-018-3649-y

Salt, D. E., Baxter, I., and Lahner, B. (2008). Ionomics and the study of the plant Ionome. *Annu. Rev. Plant Biol.* 59, 709–733. doi: 10.1146/annurev.arplant.59.032607.092942

Sambatti, J. B. M., and Rice, K. J. (2006). Local adaptation, patterns of selection, and gene flow in the Californian serpentine sunflower (*Helianthus exilis*). *Evolution* 60, 696–710. doi: 10.1111/j.0014-3820.2006.tb01149.x

Savolainen, O., Lascoux, M., and Merilä, J. (2013). Ecological genomics: genes in ecology and ecology in genes. *Nat. Rev. Genet.* 14, 807–820. doi: 10.1038/nrg3522

Schechter, S., and Branco, S. (2014). "The ecology and evolution of mycorrhizal fungi in extreme soils" in *Plant ecology and evolution in harsh environments*. eds. N. Rajakaruna, R. S. Boyd and T. B. Harris (New York: Nova Science Publishers), 33–52.

Schechter, S. P., and Bruns, T. D. (2008). Serpentine and non-serpentine ecotypes of *Collinsia sparsiflora* associate with distinct arbuscular mycorrhizal fungal assemblages. *Mol. Ecol.* 17, 3198–3210. doi: 10.1111/j.1365-294X.2008.03828.x

Selby, J. P. (2014). *The genetic basis of local adaptation to serpentine soils in Mimulus guttatus*. Dissertation, Duke University.

Selby, J. P., Jeong, A. L., Toll, K., Wright, K. M., and Lowry, D. B. (2014). "Methods and discoveries in the pursuit of understanding the genetic basis of adaptation to harsh environments in Mimulus" in *Plant ecology and evolution in harsh environments*. eds. N. Rajakaruna, R. S. Boyd and T. B. Harris (New York: Nova Science Publishers), 243–266.

Selby, J. P., and Willis, J. H. (2018). Major QTL controls adaptation to serpentine soils in *Mimulus guttatus*. *Mol. Ecol.* 27, 5073–5087. doi: 10.1111/mec.14922

Sobczyk, M. K., Smith, J. A. C., Pollard, A. J., and Filatov, D. A. (2017). Evolution of nickel hyperaccumulation and serpentine adaptation in the *Alyssum serpyllifolium* species complex. *Heredity* 118, 31–41. doi: 10.1038/hdy.2016.93

Sork, V. L., Aitken, S. N., Dyer, R. J., Eckert, A. J., Legendre, P., and Neale, D. B. (2013). Putting the landscape into the genomics of trees: approaches for understanding local adaptation and population responses to changing climate. *Tree Genet. Genomes* 9, 901–911. doi: 10.1007/s11295-013-0596-x

Southworth, D., Tackaberry, L. E., and Massicotte, H. B. (2014). Mycorrhizal ecology on serpentine soils. *Plant Ecol. Diversity* 7, 445–455. doi: 10.1080/17550874.2013.848950

Stein, R. J., Höreth, S., de Melo, J. R. F., Syllwasschy, L., Lee, G., Garbin, M. L., et al. (2017). Relationships between soil and leaf mineral composition are element-specific, environment-dependent and geographically structured in the emerging model *Arabidopsis halleri*. *New Phytol.* 213, 1274–1286. doi: 10.1111/nph.14219

Stern, D. L. (2013). The genetic causes of convergent evolution. *Nat. Rev. Genet.* 14, 751–764. doi: 10.1038/nrg3483

Taylor, S. I., and Levy, F. (2002). Responses to soils and a test for preadaptation to serpentine in *Phacelia dubia* (Hydrophyllaceae). *New Phytol.* 155, 437–447. doi: 10.1046/j.1469-8137.2002.00478.x

Teptina, A., Paukov, A., and Rajakaruna, N. (2018). Ultramafic vegetation and soils in the circumboreal region of the northern hemisphere. *Ecol. Res.* 33, 609–628. doi: 10.1007/s11284-018-1577-1

Turner, T. L., Bourne, E. C., Von Wettberg, E. J., Hu, T. T., and Nuzhdin, S. V. (2010). Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. *Nat. Genet.* 42, 260–263. doi: 10.1038/ng.515

Turner, T. L., von Wettberg, E. J., and Nuzhdin, S. V. (2008). Genomic analysis of differentiation between soil types reveals candidate genes for local adaptation in *Arabidopsis lyrata*. *PLoS One* 3:e3183. doi: 10.1371/journal.pone.0003183

Veatch-Blohm, M. E., Roche, B. M., and Campbell, M. J. (2013). Evidence for cross-tolerance to nutrient deficiency in three disjunct populations of *Arabidopsis lyrata ssp. lyrata* in response to substrate calcium to magnesium ratio. *PLoS One* 8:e63117. doi: 10.1371/journal.pone.0063117

Veatch-Blohm, M. E., Roche, B. M., and Dahl, E. E. (2017). Serpentine populations of *Arabidopsis lyrata ssp. lyrata* show evidence for local adaptation in response to nickel exposure at germination and during juvenile growth. *Environ. Exp. Bot.* 138, 1–9. doi: 10.1016/j.envexpbot.2017.02.017

Vlamis, J. (1949). Growth of lettuce and barley as influenced by degree of calcium-saturation of soil. *Soil Sci.* 67:453–466. doi: 10.1097/00010694-194906000-00005

Vlamis, J., and Jenny, H. (1948). Calcium deficiency in serpentine soils as revealed by adsorbent technique. *Science* 107:549. doi: 10.1126/science.107.2786.549

von Wettberg, E. J. B., Ray-Mukherjee, J., D'Adesky, N., Nesbeth, D., and Sistla, S. (2014). "The Evoltionary ecology and genetics of stress resistance syndrome (SRS) traits: revisiting chapin, autumn and pugnaire (1993)" in *Plant ecology and evolution in harsh environments*. eds. N. Rajakaruna, R. S. Boyd and T. B. Harris (New York: Nova Science Publishers), 201–226.

Walker, R. B., Walker, H. M., and Ashworth, P. (1955). Calcium-magnesium nutrition with special reference to serpentine soils. *Plant Physiol.* 30, 214–221. doi: 10.1104/pp.30.3.214

Westerbergh, A. (1994). Serpentine and non-serpentine *Silene dioica* plants do not differ in nickel tolerance. *Plant Soil* 167, 297–303. doi: 10.1007/BF00007956

Whittaker, R. H. (1954). The ecology of serpentine soils. *Ecology* 35, 258–288. doi: 10.2307/1931126

Willems, G., Dräger, D. B., Courbot, M., and Godé, C. (2007). The genetic basis of zinc tolerance in the metallophyte *Arabidopsis halleri* ssp. halleri (Brassicaceae): an analysis of quantitative trait loci. *Genetics* 176, 659–674. doi: 10.1534/genetics.106.064485

Wright, K. M., Lloyd, D., Lowry, D. B., Macnair, M. R., and Willis, J. H. (2013). Indirect evolution of hybrid lethality due to linkage with selected locus in *Mimulus guttatus*. *PLoS Biol.* 11:e1001497. doi: 10.1371/journal.pbio.1001497

Wright, J. W., Stanton, M. L., and Scherson, R. (2006). Local adaptations to serpentine and non-serpentine soils in *Collinsia sparsiflora*. *Evol. Ecol. Res.* 8, 1–21.

Wright, J. W., and Von Wettberg, E. (2009). 'Serpentinomics' - an emerging new field of study. *Northeast. Nat.* 16, 285–296. doi: 10.1656/045.016.0521

Yeaman, S., and Whitlock, M. C. (2011). The genetic architecture of adaptation under migration-selection balance. *Evolution* 65, 1897–1911. doi: 10.1111/j.1558-5646.2011.01269.x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Case study 2

Parallel alpine differentiation in *Arabidopsis arenosa*

frontiers in Plant Science

ORIGINAL RESEARCH
published: 08 December 2020
doi: 10.3389/fpls.2020.561526

Parallel Alpine Differentiation in *Arabidopsis arenosa*

Adam Knotek[1,2], Veronika Konečná[1,2], Guillaume Wos[1], Doubravka Požárová[1], Gabriela Šrámková[1], Magdalena Bohutínská[1,2], Vojtěch Zeisek[1,2], Karol Marhold[1,3] and Filip Kolář[1,2,4]*

[1] Department of Botany, Charles University, Prague, Czechia, [2] Institute of Botany, The Czech Academy of Sciences, Průhonice, Czechia, [3] Institute of Botany, Slovak Academy of Sciences, Bratislava, Slovakia, [4] Department of Botany, University of Innsbruck, Innsbruck, Austria

Parallel evolution provides powerful natural experiments for studying repeatability of evolution and genomic basis of adaptation. Well-documented examples from plants are, however, still rare, as are inquiries of mechanisms driving convergence in some traits while divergence in others. *Arabidopsis arenosa*, a predominantly foothill species with scattered morphologically distinct alpine occurrences is a promising candidate. Yet, the hypothesis of parallelism remained untested. We sampled foothill and alpine populations in all regions known to harbor the alpine ecotype and used SNP genotyping to test for repeated alpine colonization. Then, we combined field surveys and a common garden experiment to quantify phenotypic parallelism. Genetic clustering by region but not elevation and coalescent simulations demonstrated parallel origin of alpine ecotype in four mountain regions. Alpine populations exhibited parallelism in height and floral traits which persisted after two generations in cultivation. In contrast, leaf traits were distinctive only in certain region(s), reflecting a mixture of plasticity and genetically determined non-parallelism. We demonstrate varying degrees and causes of parallelism and non-parallelism across populations and traits within a plant species. Parallel divergence along a sharp elevation gradient makes *A. arenosa* a promising candidate for studying genomic basis of adaptation.

Keywords: adaptation, alpine environments, *Arabidopsis*, convergence, parallel evolution, phenotypic parallelism

OPEN ACCESS

Edited by:
Michael Eric Schranz, Wageningen University and Research, Netherlands

Reviewed by:
Tuomas Hämälä, University of Minnesota Twin Cities, United States
Karl Josef Schmid, University of Hohenheim, Germany

***Correspondence:**
Filip Kolář
filip.kolar@natur.cuni.cz;
filip.kolar@gmail.com

Specialty section:
This article was submitted to Plant Systematics and Evolution, a section of the journal Frontiers in Plant Science

Received: *12 May 2020*
Accepted: *16 November 2020*
Published: *08 December 2020*

Citation:
Knotek A, Konečná V, Wos G, Požárová D, Šrámková G, Bohutínská M, Zeisek V, Marhold K and Kolář F (2020) Parallel Alpine Differentiation in Arabidopsis arenosa. Front. Plant Sci. 11:561526. doi: 10.3389/fpls.2020.561526

INTRODUCTION

When and why does evolution result in a predictable outcome remain challenging questions in evolutionary biology. Parallel emergence of identical phenotypes in similar environments brings one of the most intuitive examples of the action of natural selection and provides independent biological replicates allowing for identifying the genetic basis and phenotypic outcomes of adaptation (Elmer and Meyer, 2011; Losos, 2011). In particular, parallel evolution of genetically determined phenotypically distinct entities within a species, so-called ecotypes, provide valuable insights into processes of local adaptation and incipient ecological speciation in nature (Bolnick et al., 2018; Thompson et al., 2019). In contrast to well-studied examples of ecotypic parallelism from animals such as fishes (Rundle et al., 2000; Jones et al., 2012; Reid et al., 2016; Stuart et al., 2017), snails (Butlin et al., 2014; Ravinet et al., 2016), and stick insects (Nosil et al., 2002; Soria-Carrasco et al., 2014), mechanisms driving parallel evolution in plants are largely unknown (but see Roda et al., 2013b; Bertel et al., 2018).

While directional selection driven by similar environment shall lead to convergence, alternative neutral and selective forces may counteract it, leading to larger genetic and phenotypic divergence among the derived ecotypes (reviewed in Bolnick et al., 2018). Specifically, both stochastic processes such as genetic drift, varying intensity of gene flow and different pools of initial variation available for selection as well as deterministic processes such as selection in response to locally distinct micro-environments and genetic architecture of the traits can lead to non-parallel patterns in phenotypic variation (Elmer et al., 2014; Langerhans, 2018). Consequently, there is no clear-cut border between parallel and non-parallel phenotypes, and the extent of parallelism may vary over traits or particular sets of compared populations (Stuart et al., 2017). Finally, apparent similarity across populations can reflect short-term environmentally induced traits (phenotypic plasticity), however, the relative contribution of the heritable vs plastic forces in the manifestation of phenotypic parallelism remains unknown (Pfennig et al., 2010).

Multiple ($N > 2$) independent environmental transitions within a species provide powerful naturally replicated systems to infer the mechanisms driving parallel phenotypic evolution as well as the extent of variation in parallel vs non-parallel response. Such systems allow multiple pairwise comparisons, in contrast to systems comparing only two parallel transitions. Indeed, independently evolved animal ecotypes documented that a combination of both stochastic and deterministic processes led to a varying extent of parallelism (Bolnick et al., 2018). In contrast, such detailed inquiries in plants are scarce due to paucity of well-established systems that involve multiple replicates of environmental transitions within a species. The limited evidence from the so far investigated systems, *Heliosperma pusillum* (Trucchi et al., 2017; *Caryophyllaceae*; five pairs of alpine-foothill ecotypes, Bertel et al., 2018) and *Senecio lautus* (Roda et al., 2013a, *Asteraceae*; seven pairs of dune-heathland ecotypes, Roda et al., 2013b), suggest considerable morphological divergence among the independently formed ecotypes despite similar environmental triggers, perhaps as a result of genetic drift (Trucchi et al., 2017) or gene flow (Roda et al., 2013a). However, to infer mechanisms driving parallel evolution in plants in general and its genetic basis in particular, additional genetically well-tractable plant systems are needed.

Alpine populations of *Arabidopsis arenosa* represent a promising system for addressing drivers and consequences of parallel evolution in a genomic and molecular genetic context of the well-researched *Arabidopsis* genus. The species thrives mostly in low to mid-elevations (up to ∼1,000 m a.s.l.) of Central and Eastern Europe, but scattered occurrences in treeless alpine habitats (∼1,500–2,500 m a.s.l.) have been described from four mountain regions in the floristic literature (Melzer, 1960; Měsíček and Goliašová, 2002; Bartok et al., 2016). Alpine environments are generally well-poised for inquiry of plant parallel adaptation. On one hand, high elevations pose a spectrum of challenges to plant life, potentially triggering directional selection, such as freezing and fluctuating temperatures, strong winds, increased UV radiation and low partial pressure of carbon dioxide. Such pressures are believed to jointly trigger emergence of distinct alpine phenotypic "syndromes" (e.g., contracted rosette plants, dense cushions, and giant rosettes) that have been recurrently formed in distinct mountain ranges throughout the world (Hedberg and Hedberg, 1979; Körner, 2003; Halbritter et al., 2018; Konečná et al., 2019). Indeed, alpine *A. arenosa* constitutively exhibits a distinct morphotype characterized by lower stature, less-lobed and thicker leaves, larger flowers and wider siliques (Měsíček and Goliašová, 2002). On the other hand, the island-like distribution of alpine habitats promotes parallel colonization of individual alpine "islands" by the spatially closest foothill populations (Levin, 2001). In line with this, a recent genomic study of *A. arenosa* (Monnahan et al., 2019) demonstrated geographical structuring of its range-wide genetic diversity, suggesting that the alpine ecotype might be of a polytopic origin. This hypothesis, as well as the extent of phenotypic parallelism and its plastic vs. genetic basis, however, remains untested.

In this study, we sampled multiple alpine and adjacent foothill *A. arenosa* populations covering all mountain regions known to harbor the alpine ecotype. Using genome-wide SNP genotyping we tested our main hypothesis that alpine environment within each mountain region has been colonized independently (parallel origin scenario) as opposed to clustering of the alpine populations together (single origin scenario). Indeed, a combination of genetic structure analyses and modeling revealed multi-parallel origin of the alpine populations. Thus, we combined analyses of field-sampled phenotypic data and a common garden experiment to assess whether the independent alpine transitions are also associated with phenotypic parallelism. Specifically, we ask: (1) Had independent alpine colonization triggered similar phenotypic transitions in distinct mountain regions? (2) Does the alpine phenotype remain divergent from the "typical" foothill form in standardized conditions and, if so, which characters contribute to the genetically determined alpine syndrome? (3) In contrast, are there traits exhibiting a rather opposite, non-parallel response?

MATERIALS AND METHODS

Field Sampling

Arabidopsis arenosa is a perennial outcrosser encompassing diploid and autotetraploid populations (Kolář et al., 2016b; Monnahan et al., 2019) that is widespread in foothill (colline to sub-montane) elevations across Central and Eastern Europe. Scattered occurrences of morphologically distinct populations have been recorded from four distinct mountain regions in Europe, sometimes treated as a separate species "*A. neglecta*": Eastern Alps (Melzer, 1960), Eastern (Pachschwöll and Pachschwöll, 2019), Southern (Bartok et al., 2016), and Western Carpathians (Měsíček and Goliašová, 2002). While only tetraploid populations colonized the alpine stands in the former three regions, both diploid and tetraploid populations reached the alpine belt in the Western Carpathians. As the two cytotypes still occupy distinct alpine sub-regions there (diploids in Vysoké Tatry Mts. and tetraploids in Západné Tatry Mts.; Wos et al., 2019), we kept the ploidies as separate units for

the sake of clarity. In sum, we hereafter refer to five regions: Niedere Tauern and surrounding foothills of the Eastern Alps (NT region, tetraploid), Rodna Mts. and adjacent regions of Eastern Carpathians (RD, tetraploid), Făgăraş Mts. in Southern Carpathians (FG, tetraploid), Vysoké Tatry Mts. and adjacent foothill diploid populations in Western Carpathians (VT, diploid), Západné Tatry Mts. and adjacent foothill tetraploid populations in Western Carpathians (ZT, tetraploid).

In each mountain region we sampled multiple populations from foothill habitats (semi-shaded rocky outcrops, screes and steep slopes; "foothill ecotype") and from alpine sites (screes and rocky outcrops above the timberline; "alpine ecotype"). As this heliophilous species occurs on "islands" of environmentally suitable conditions (rocky outcrops in lower elevations, rocks and screes in isolated glacial cirques in the alpine zone), the population was defined as a set of individuals occurring in a spatially discrete area in a homogeneous vegetation type surrounded by habitats with unsuitable conditions for the species (e.g., forests, dense grasslands, and arable land). Both ecotypes are separated by a clear distribution gap spanning at least 500 m of elevation which also corresponds with the timberline. We avoided sampling plants in riverbeds immediately below the alpine populations that could represent recent colonizers germinated from washed seeds of the originally alpine plants. We sampled adult individuals from a total of 58 populations – 30 from foothill and 28 from alpine habitats. Within each population, we collected on average 17 individuals at the full-flowering stage: a small part of fresh tissue for ploidy determination, leaf tissue desiccated in silica gel for genotyping and vouchers for morphometrics. We selected largest rosette leaf, second stem leaf from the base and one random flower from the terminal inflorescence and fixed them onto paper with transparent tape for detailed measurements of organ sizes; the entire individual was then press-dried. For each population we also sampled the following local environmental parameters at a microsite with abundant occurrence of *A. arenosa*: (i) vegetation samples (phytosociological relevés, each covering an area of 3×3 m) recording percentage of the area covered by herb layer and listing all vascular plant species and (ii) mixed rhizosphere soil samples from five microsites within the vegetation sample.

Ploidy level of each sampled individual was determined using flow cytometry as described by Kolář et al. (2016b).

Experimental Cultivation

In addition to field sampling, we established a common garden experiment to test whether the plants of alpine origin keep their distinct appearance when cultivated in the foothill-like conditions. We used seeds of *A. arenosa* collected from 16 natural populations overlapping with those sampled for field phenotyping and genotyping (except for pop. AA254 that was used in the experiment as a replacement of spatially close, <10 km, pop. AA145 which was included in the field dataset). The populations represent four regions (NT, VT, ZT, and FG) and two distinct elevations (two foothill and two alpine populations per region; see **Supplementary Table S1** for locality details). To minimize maternal effect of the original localities, we firstly raised one generation in growth chambers under constant conditions. To simulate outcrossing in natural populations, each plant was hand-pollinated by a mixture of pollen from the same population (∼14 individuals) for a period of one week; the offspring thus represented a genetically variable mixture of full- and half-sibs. The plants used for phenotyping were raised in the Botanical Garden of the University of Innsbruck (Austria) situated in the Alpine valley, i.e., in conditions resembling the foothill habitat. Phenotypic traits were collected on all plants in the full-flowering stage in an identical way as for the field sampling. For details on cultivation conditions, see **Supplementary Methods S1**.

Inference of Parallel Origins

We genotyped 156 individuals from 46 populations (2–8 individuals per population, 3.5 on average) using the double-digest RADseq protocol of Arnold et al. (2015). For an additional 44 individuals from 11 populations genome-wide SNP data were already available from a genome resequencing study (Monnahan et al., 2019). Raw reads processing and filtration generally followed our earlier study (31); for details see the **Supplementary Methods S1**.

We inferred population grouping using several complementary approaches. Firstly, we ran *K*-means clustering, a non-parametric method with no assumption on ploidy, in adegenet v1.4-2 using 1000 random starts for each *K* between 1 and 20 and selected the partition with the lowest Bayesian Information Criterion (BIC) value (Jombart, 2008). Secondly, we used model-based method FastStructure (Raj et al., 2014). We randomly sampled two alleles per tetraploid individual (using a custom script) – this approach has been demonstrated not to lead to biased clustering in autotetraploid samples in general (Stift et al., 2019) and *Arabidopsis* in particular (Monnahan et al., 2019). We ran the FastStructure with 10 replicates under $K = 5$ (the same number as for *K*-means clustering, representing the number of the regions) and additionally only for tetraploid individuals under $K = 4$. Finally, we displayed genetic distances among individuals using principal component analysis (PCA) as implemented in adegenet v.2.1.1. For clustering analyses, we used random thinning over 150 kb windows (length of our RAD-locus) to reduce the effect of linkage and removed singletons resulting in a dataset of 4,341 SNPs. PCA, AMOVA and genetic distances (see below) were calculated using the full set of 103,928 filtered SNPs.

Additionally, we tested for parallel origin of alpine populations using coalescent simulations in *fastsimcoal v.26* (Excoffier et al., 2013) – an approach that is also suitable for autotetraploids and mixed-ploidy systems (Arnold et al., 2012). We constructed unfolded (polarization following Monnahan et al., 2019) four-dimensional joint allele frequency spectra from genome-wide SNPs from a subset of sufficiently sampled populations, one alpine and one foothill per region (∼ 176–417 k SNPs per population, see **Supplementary Table S2** and **Supplementary Methods S3** for details; one-dimensional AFS of the populations are published in Monnahan et al., 2019). We compared all regions occupied by tetraploid populations (NT, RD, FG, and ZT) in a pairwise manner. Taking into account single origin of the widespread *A. arenosa* tetraploid cytotype (Arnold et al., 2015; Monnahan et al., 2019), we had not compared all

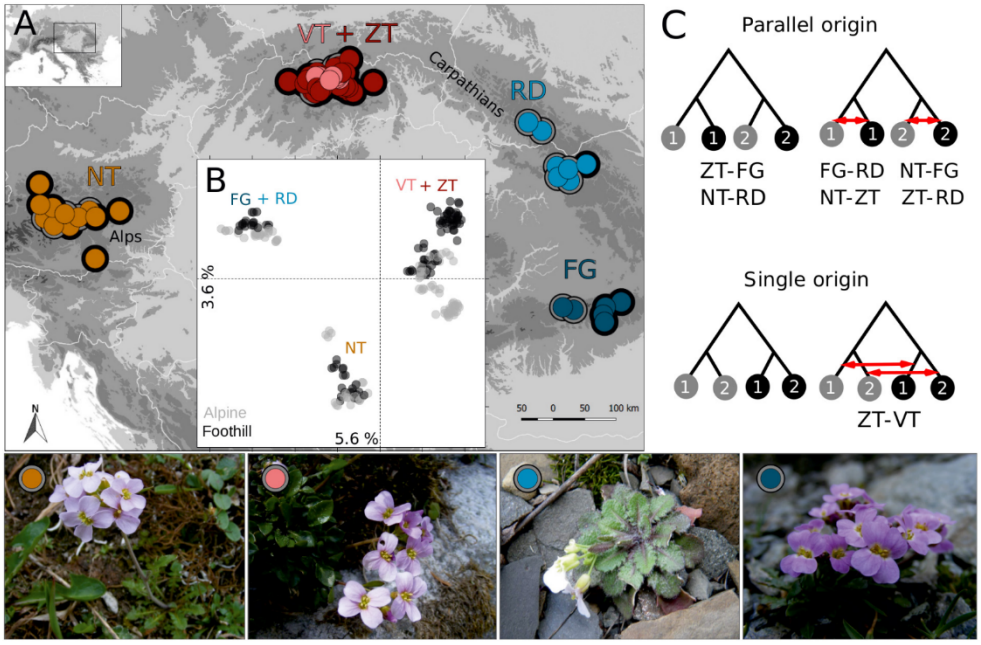


FIGURE 1 | Parallel alpine differentiation of *Arabidopsis arenosa* in Central Europe inferred from genome-wide SNP variation. **(A)** Geographic distribution of sampled populations colored by *K*-means clusters (circles with different colors) and classified according to the ecotype (black rim = foothill, gray rim = alpine). **(B)** Principal component analysis of the individuals colored according to ecotype and labeled according to region (color corresponding to panel **A**). **(C)** Four scenarios addressed by coalescent simulations; gray dots = alpine populations, black dots = foothill populations, numbers = regions, red arrows = migration events. The scenario with best fit to the observed allele frequency spectra for each pair of regions tested is indicated by its respective code below the scheme. Illustrative photos of convergent alpine *A. arenosa* ecotypes from four regions are presented at the bottom. Note that the VT region is occupied by diploids while all other regions are represented by autotetraploid populations.

diploid-tetraploid pairs but only the spatially closest diploid and tetraploid populations from Western Carpathians (VT and ZT) for which a complex reticulation relationships was suggested previously (Wos et al., 2019). Briefly, for each pair of regions, we compared the fit of our data with two competing topologies: (i) sister position of alpine and foothill populations from the same region (i.e., parallel origin) and (ii) sister position of populations from distinct regions but belonging to the same ecotype (i.e., single origin of each ecotype). For each topology we either assumed or not assumed secondary contact (between-ecotype gene flow) within each region, what resulted in a total of four scenarios per regional pair (**Figure 1C**). As the main aim of the study was testing for single vs parallel origin of the ecotypes (i.e., model comparison) we had not followed with additional analyses quantifying population divergence (i.e., parameter estimates) within each region.

We used hierarchical analysis of molecular variance (AMOVA) implemented in the R package pegas to test for genetic differentiation (i) among regions and among populations within regions and (ii) among ecotypes and among populations within each ecotype (both in the total dataset and in each region separately). We calculated nucleotide diversity (π) and Tajima's D for each population with at least four individuals (N = 26) and pairwise differentiation (F_{ST}) between these populations using custom python3 scripts (available at https://github.com/mbohutinska/ScanTools_ProtEvol).

Ecological Data

To assess environmental differences among the populations, we combined broad-scale climatic parameters acquired from a database (SolarGIS, version 1.9, operated by GeoModel Solar, Bratislava, Slovakia) with local conditions sampled at the original site of each population. First, we estimated the average values of three climatic variables: Precipitation, Temperature and Photosynthetic Active Radiation (PAR), over April, May, and June, which correspond to the main growth period of *A. arenosa*. Second, we measured soil pH of each site by thermo-corrected electrode (WTW Multilab 540, Ionalyzer, pH/mV meter) at the Analytical Laboratory of the Institute of Botany (Průhonice, Czech Republic). Finally, we inferred local environmental conditions from the vegetation samples: (i) total vegetation cover of herb layer and (ii) Ellenberg Indicators Values (EIVs) calculated from the species list using JUICE (Tichý, 2002). EIVs provide estimates of environmental characteristics of the sites inferred from species composition based on expert knowledge (Ellenberg, 1992). In total, we recorded 671 plant species of which 76% have EIVs. We used only EIVs for Light, Nutrients and Moisture as the remaining EIVs (for Temperature, Continentality, and Soil Reaction) overlapped with

the above-described climatic and soil variables. In sum, our dataset of environmental parameters comprised the following eight variables: Precipitation, Temperature, PAR, soil_pH, Vegetation_cover, EIV_Light, EIV_Nutrients and EIV_Moisture. No pair of the parameters exhibited >0.8 Pearson correlation (**Supplementary Table S3**).

Phenotypic Traits

To test for parallel phenotypic response to alpine environment, we described the morphology of each plant sampled using 16 traits (999/223 vouchers in total scored for the field and common garden datasets, respectively). On each plant we measured 12 phenotypic characters describing the overall shape and size of both vegetative and reproductive organs (except for siliques which were not fully developed in the full-flowering stage which our sampling aimed at). To assess variation in shape of the organs independent of absolute size, we further derived four ratios (all characters are described and listed in **Supplementary Table S4**). Missing values in the field dataset (1.5% in total) were replaced by population means. For statistical analyses all characters except four (PL, PW, SL and SW), were log-transformed to approach normal distribution. No pair of traits was very strongly correlated (>0.8, **Supplementary Table S3**).

Statistical Analyses of Ecological and Morphological Data

We used PCA calculated by base R function *prcomp* to visualize ecological differences among populations and morphological differences among individuals. Separate PCAs have been calculated on standardized (zero mean and unit variance) sets of (i) the eight climatic and local environmental variables and the 16 morphological traits recorded on the individuals collected (ii) in field and (iii) in a common garden experiment. Overall differentiation in environmental conditions and morphology of the ecotypes were tested by permutation multivariate analyses of variance (permanova). We first calculated Euclidean distances among population (environmental data) or individual (field and common garden morphological data) values using *dist* function and then ran a permanova test with *adonis2* function (number of permutations = 30,000) in R package vegan 2.5–4. In addition, we quantified the range of morphological variation of each ecotype by calculating disparity as the median distance between each individual and centroid of their corresponding ecotype in the ordination space using the R package dispRity.

To quantify morphological differentiation between pre-defined groups (ecotypes and regions) across all traits we ran classificatory discriminant analysis with cross-validation as implemented in Morphotools 1.1 (Koutecký, 2014). To assess relative contribution of individual morphological characters to the between foothill-alpine differentiation individuals we calculated a constrained ordination (linear discriminant analysis, LDA) in Morphotools. We calculated the discriminant analyses and permanova tests (i) for complete datasets and then (ii) separately for each region with ecotype as a factor to assess major drivers of foothill-alpine phenotypic differentiation within each region and (iii) separately for each ecotype with region as a factor to quantify variation in between-region phenotypic differentiation within each ecotype.

Finally, we ran generalized linear mixed models (GLM) for each of the 16 phenotypic traits to test for the effects of ecotype, region and their interaction (individual-based data with population as a random factor, *lme* function) in nlme package v3.1-137. Parallelism in foothill-alpine differentiation was considered for traits with significant effect of ecotype but non-significant ecotype × region interaction (lack of regionally specific differences between ecotypes) that was revealed consistently in both field and common garden datasets. In turn, a trait with significant effect of ecotype × region interaction was indicative of non-parallelism (evidence for regional-specific differences). Finally, traits exhibiting significant ecotypic effect in the field dataset but not in common garden were considered as plastic with respect to alpine differentiation. To assess the degree of parallelism and non-parallelism in a more quantitative way, effect sizes of each factor and their interaction were estimated from a linear model using a partial eta-squared method (Etasq function) in R package heplots. Similarly as above, large ecotypic effect only was indicative of parallelism while strong contribution of the ecotype × region interaction implied non-parallel foothill-alpine differentiation in that particular trait (Stuart et al., 2017). All the above analyses were run separately for the field and common garden datasets.

RESULTS

Genetic Structure and Diversity

We confirmed the presence of tetraploid populations in four mountain regions (NT, ZT, RD, and FG) while only diploids represented the Vysoké Tatry (VT) region in our sample. For the 200 RAD-sequenced individuals we gathered 103,928 SNPs of average depth 30× (1.37% missing data). Non-hierarchical K-means clustering revealed that the populations cluster under K = 5 (partition supported by the lowest Bayesian information criterion, **Supplementary Figure S1**) according to the geographic regions disregarding their alpine or foothill origin (i.e., the ecotype; **Figure 1**). Populations from the spatially close VT and ZT regions, which differed by ploidy, were the single exception: here, the majority of the alpine diploid populations (from VT region) and some alpine tetraploid (ZT) populations formed a separate cluster distinct from the remaining Western Carpathian samples, regardless of ploidy. FastStructure run under $K = 5$ supported this clustering and revealed clear separation of all groups but the VT and ZT groups, which were remarkably admixed (**Supplementary Figure S2**). Principal component analysis confirmed spatial, not ecotypic, clustering and revealed three main clusters: populations from the RD and FG regions were separated from the VT and ZT cluster along the first axis, while NT populations differentiated from the rest along the second axis (**Figure 1B**).

Coalescent simulations demonstrated that parallel origin scenarios are more likely than scenarios assuming single

origin of alpine ecotype and this result was consistent across all combinations of tetraploid populations (Akaike weights supporting parallel origin ranged 0.99–1 across the pairwise comparisons of regions; **Supplementary Table S5**, see also **Supplementary Figure S3** for the distribution of AIC values). In contrast, for the spatially close diploid (VT) and tetraploid (ZT) populations the scenario involving sister position of alpine populations of both ploidies was preferred (Akaike weight for the scenario assuming single origin followed by migration was 1; **Supplementary Table S5**). In summary, genetic analysis coherently demonstrated polytopic origin of the alpine ecotype in four distinct mountain ranges (Alps – NT, Eastern Carpathians – RD, Southern Carpathians – FG, and Western Carpathians – VT and ZT; **Figure 1C**). In order to account for the ploidy difference among the Western Carpathian populations, we kept the VT and ZT populations separate in the following analyses. Additional analyses when the VT and ZT populations were merged into a single unit (**Supplementary Table S6**), demonstrated that such alternative grouping does not lead to qualitatively different results with respect to the patterns in alpine-foothill differentiation.

Ecotypic differentiation explained a negligible proportion of genetic variance (3%) while the five regions accounted for 20% of variation (AMOVA analysis, see **Supplementary Table S7**). Consequently, genetic differentiation between foothill and alpine populations was non-significant within any of the mountain regions in separate AMOVA analyses (**Table 1**, see also pairwise F_{ST} among populations, **Supplementary Table S8**). These results suggest there is overall low inter-population divergence within each region, and such observation is consistent across regions. The alpine colonization was not accompanied by a reduction in genetic diversity, as populations belonging to both ecotypes exhibited similar nucleotide diversity overall (Wilcoxon test, $W = 77$, $p = 0.93$) as well as within each mountain region separately and also Tajima's D values were close to neutrality (i.e., zero) in both ecotypes (**Table 1**).

Habitat Differentiation

Environmental conditions of foothill and alpine sites significantly differed (permutational multivariate analysis of variance of populations, permanova, $F = 16.92$, $p < 0.001$; **Figures 2A,B**). In contrast, populations from different mountain regions did not differ based on the environmental parameters recorded ($F = 1.21$, $p = 0.30$).

Morphological Differentiation in Natural Populations

Field sampled alpine individuals were overall morphologically differentiated from their foothill counterparts as revealed by their separation in PCA (**Figures 2C,D** and **Supplementary Figure S4**), high (87%) classification success in the classificatory discriminant analysis and significant effect of ecotype in permanova ($F = 49.25$, $p < 0.001$). Morphological differentiation by ecotype was also significant for each region separately, with high classification success (89–100% across regions, **Table 1**). Overall, morphological differentiation measured by disparity was significantly higher among the foothill than among the alpine individuals ($F_{1,997} = 113.4$, $p < 0.001$) suggesting that alpine individuals were morphologically more similar to each other than were their foothill counterparts.

Fourteen of the 16 scored characters significantly differentiated between the ecotypes, however, the level of parallelism remarkably differed across traits (**Figure 3**). Stem height and traits reflecting flower size consistently varied between ecotypes across regions as demonstrated by significant effect of ecotype but non-significant ecotype × region interaction (GLM, **Figure 3**, **Supplementary Table S9**, and **Supplementary Figure S5**) and strong contribution to the ordination constrained by ecotype (highest loadings to the discriminant axis in LDA, **Supplementary Table S10**). Overall, alpine plants, regardless of their region of origin, were shorter with larger calyces and petals (**Supplementary Table S4**). In contrast, traits describing leaf size and shape were those with the strongest ecotype × region interaction indicating region-specific (i.e., non-parallel) morphological differentiation between ecotypes (**Figure 3**). LDAs run separately for each region revealed such non-parallel response reflected distinct foothill-alpine differentiation in the FG region which was, apart from the height, most strongly driven by traits on leaves (LDA, **Supplementary Table S10**). Such regionally specific effect was primarily driven by the foothill FG morphotypes which is apparent from their very distinct position in the ordination space (**Figure 2C**) and higher distinctness of the foothill-FG than the alpine-FG populations when contrasted to populations from other regions belonging to the same ecotype (classification success as an FG group was 86% vs 77% for the foothill vs. alpine populations, respectively).

Phenotypic Variation in Common Garden

Overall phenotypic differentiation between plants originating from alpine and foothill environment remained highly significant even under two generations of cultivation in a common garden (a subset of 16 populations from four regions; **Table 1**, **Figures 2E,F**). Differences between originally foothill and alpine populations were also significant within each region studied (permanova, **Table 1**), however, with markedly varying strength among the four regions examined, being weakest for NT plants (63% classification success) and strongest for plants from FG region (95%, **Table 1**). This implies that while alpine populations from some regions (FG and ZT in particular) keep their high morphological distinctness regardless of growth conditions, foothill and alpine populations from the NT region became more similar to each other when grown in a common garden. Morphological disparity of the originally foothill and alpine individuals was still significantly higher than that of their alpine counterparts ($F_{1,221} = 30.96$, $p < 0.001$).

Similarly to the field data, traits describing plant height and floral size contributed most strongly to the ecotypic differentiation across all regions (strong effect of ecotype but no ecotype × region interaction; **Figure 3** and **Supplementary Table S9**). On the other hand, ecotypic differentiation

disappeared under common garden cultivation for nearly all leaf traits, with the exception of leaf lobe characters. However, leaf lobe traits also exhibited significant region × ecotype interaction due to their strong discriminative power between ecotypes in the FG region but not elsewhere (LDA, **Supplementary Table S10**).

In summary, reduced stem height and larger floral organs showed the strongest parallelism in the genetic component of morphological differentiation across regions (significant effect of ecotype but non-significant ecotype × region interaction both in field and in the common garden; **Figure 3**). On the other hand, leaf traits generally showed utmost regionally specific discriminative power in the field (no effect of ecotype and/or non-significant ecotype × region interaction) and such regionally specific discriminative effect mostly disappeared in a common garden, demonstrating plasticity in alpine-foothill differentiation for the majority of leaf traits. The only exception were leaf lobe traits that strongly discriminated foothill and alpine ecotype in one particular region (FG) constantly under both field and common garden conditions, demonstrating genetically determined non-parallelism.

DISCUSSION

Here, we combined a survey of genetic, ecological and morphological variation in natural populations with a common garden experiment to demonstrate multi-parallel ecotypic differentiation in a wild *Arabidopsis*. Replicated emergence of similar genetically based traits associated with alpine colonization suggests selection triggered by the challenging alpine environment shaped morphological variation of the ancestrally foothill *Arabidopsis* species.

Parallel Origin of the Alpine Ecotype

A mosaic distribution of ecotypes within a species' phylogeny, with genetic diversity rather structured by geographical proximity than by environment (the ecotypes), serves as evidence of repeated ecological divergence (Nosil et al., 2008; Johannesson et al., 2010). The clustering, distance-based and coalescent analyses of genome-wide SNPs congruently demonstrated that the sampled *Arabidopsis arenosa* populations exhibit distinct regional clustering regardless of their alpine vs foothill origin in four regions, i.e., supporting the parallel origin scenario. In line with biogeography (Pawłowski, 1970; Mráz and Ronikier, 2016), the major genetic groups corresponded to the four spatially well-defined and floristically distinct mountain regions: Eastern Alps (NT group) and Southern (FG), Eastern (RD) and Western Carpathians (VT + ZT groups). Genetic similarity of Western Carpathian diploids (VT) and tetraploids (ZT) has been detected previously and probably reflects recent origin of the tetraploid cytotype in the area and/or subsequent interploidy gene flow (Arnold et al., 2015; Monnahan et al., 2019; Wos et al., 2019). However, due to distinct ploidy and spatial arrangement reducing the chance for extant gene flow among the VT-alpine and ZT-alpine populations, we considered the two regions as separate units in the following discussion for the sake of clarity.

Parallel origin of the alpine *A. arenosa* ecotype is consistent with previous range-wide studies of the species' genetic structure (Arnold et al., 2015; Monnahan et al., 2019). Taking into account spatial arrangement of populations and overall phylogeny of the species, colonization of alpine stands from foothill populations likely underlies the observed foothill-alpine differentiation. Firstly, the foothill morphotype represents an ancestral state for the entire species; all three early diverged

TABLE 1 | Genetic and morphological diversity and differentiation of foothill and alpine populations of *A. arenosa* from the five mountain regions sampled in Central Europe.

	Genetic diversity and differentiation					Field morphology			Common garden morphology		
	N[1]	AMOVA (%)[2]	Genetic diversity[3]	Pairwise Fst[4]	Tajima's D[3]	N[1]	CDA[5] (%)	Differentiation[6]	N[1]	CDA[5] (%)	Differentiation[6]
Grouping by region											
All populations	200	20	0.048	0.132	−0.096	999	43	6.43***	223	47	4.00*
Foothill	109	23	0.049	0.120	−0.148	559	65	14.1***	117	62	6.58***
Alpine	91	30	0.048	0.139	−0.025	440	48	10.75***	106	75	19.94***
Grouping by ecotype											
All regions	200	3	0.049/0.048		−0.148/ −0.025	999	89	49.25***	223	83	108.5***
NT (Niedere Tauern)	41	n.s.	0.047/0.045	0.105/0.089	−0.022/0.069	232	100	8.83**	60	63	6.52**
VT (Vysoké Tatry)	73	n.s.	0.047/0.047	0.066/0.068	0.128/ −0.075	380	89	17.15***	48	83	60.91***
ZT (Západné Tatry)	42	n.s.	0.046/0.051	0.049/0.045	−0.251/ −0.229	204	90	16.36***	59	93	17.00***
RD (Rodna)	22	n.s.	0.058/0.060	–/–	−0.263/ −0.104	85	93	3.42*	–	–	–
FG (Făgăraş)	22	n.s.	0.055/0.042	0.062/0.119	−0.173/0.225	98	100	50.96***	56	95	113.61***

[1] *N RAD-sequenced/phenotyped individuals.*
[2] *Among-group genetic variation component (in %) as explained by hierarchical AMOVA.*
[3] *Pairwise nucleotide diversity (π) and Tajima's D averaged over populations with ≥ 4 individuals (foothill/alpine ecotypes, respectively).*
[4] *Pairwise F_{st} averaged over populations with ≥ 4 individuals (foothill/alpine ecotypes, respectively).*
[5] *% of correct classification into ecotype/regional group as inferred by classificatory discriminant analysis of the 16 morphological characters.*
[6] *F-values and significance (*P < 0.05, ***P < 0.01, and ***P < 0.001) of permanova analysis of the 16 morphological characters. Parallel origin of alpine ecotype.*

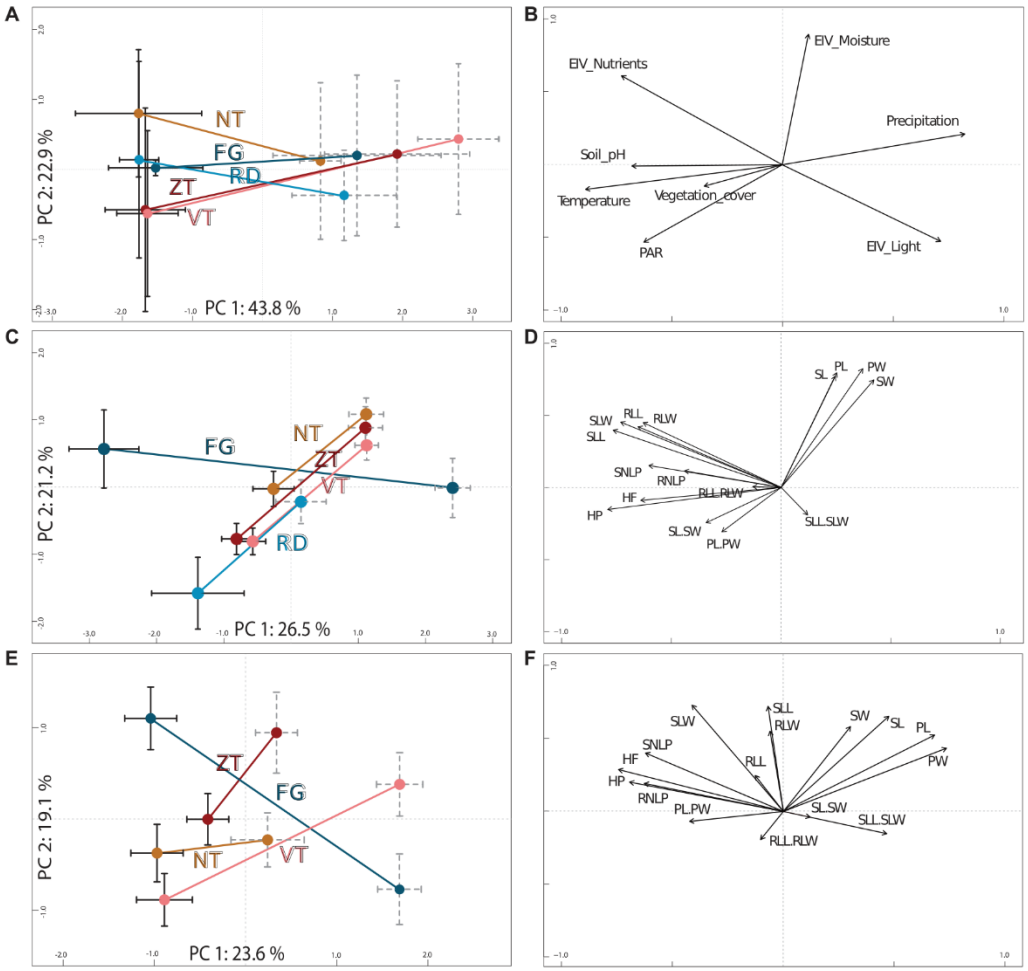


FIGURE 2 | Parallel ecological **(A,B)** and field- **(C,D)** and common garden-based **(E,F)** morphological divergence of foothill and alpine populations of *A. arenosa* from the five mountain regions. The diagrams display results of principal component analyses (PCA) based on **(A)** eight environmental parameters of the original sampling sites of all populations with **(C)** 16 morphological traits collected in 58 wild populations, and **(E)** the same 16 traits sampled in a subset of 16 populations cultivated in a common garden. Centroids (dots) and standard error bars (black solid = foothill, gray dashed = alpine) are depicted for each region and ecotype; the ecotypes from the same region are connected by a line and colored as in **Figure 1**. The corresponding loadings of each environmental parameter **(B)** or morphological trait **(D,F)** are depicted. See Table S3 for explanation of individual codes of the traits.

diploid lineages of *A. arenosa* comprise only foothill populations (Kolář et al., 2016a). Secondly, foothill populations are spread over the entire area of Central and Eastern Europe wherever suitable rocky habitats are available (Kolář et al., 2016b) while the alpine populations are generally rare and confined to isolated spots within the five mountain regions investigated. Presence of such morphologically distinct populations in other areas is highly unlikely given the floristic knowledge of European mountains (Měsíček and Goliašová, 2002; Fischer, 2008; Bartok et al., 2016).

In summary, spatial distribution of genetic variation demonstrated parallel colonization of the alpine habitats by prevalently foothill *Arabidopsis* species in four biogeographically distinct mountain ranges of Central and Eastern Europe.

Parallel Ecotypic Differentiation

Independent colonization of alpine stands followed by environmentally driven selection should lead to similar phenotypic changes across the alpine populations. Alternatively, drift, divergent selection to locally specific conditions and/or limited variation in the source populations would lead to regionally specific patterns of foothill-alpine differentiation (Bolnick et al., 2018). Independently of the region of origin, alpine *A. arenosa* populations exhibited consistently shorter stems and larger flowers demonstrating parallelism in typical traits associated with "alpine syndrome" that are considered adaptive in alpine environments (Galen, 1989; Halbritter et al., 2018). Elevation gradients belong to the most frequently studied environmental gradients in plant evolutionary ecology since

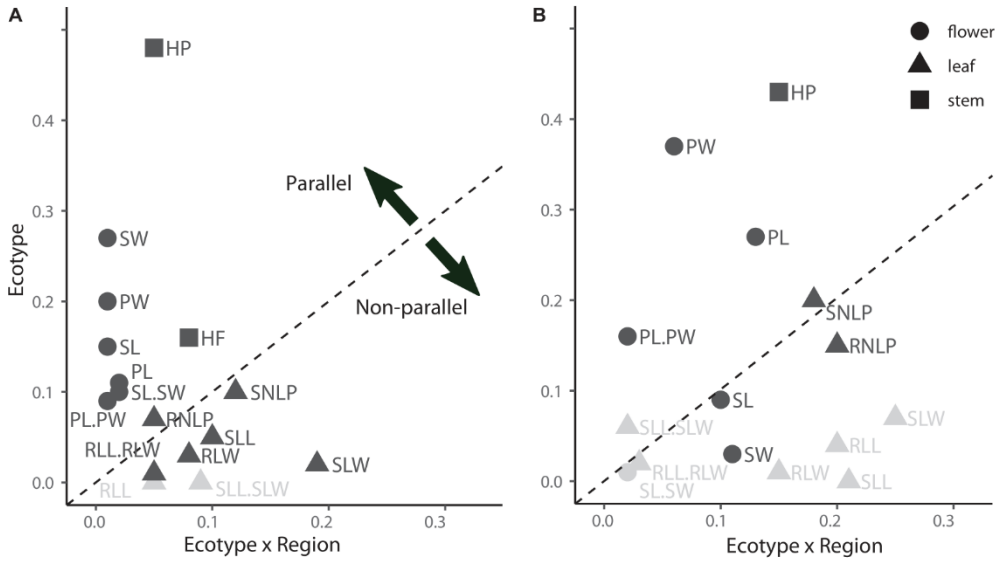


FIGURE 3 | Variation in the extent of parallelism among individual phenotypic traits sampled in the field **(A)** and common garden **(B)**. The extent of parallelism was estimated as effect sizes (Eta^2) in linear model addressing the effect of Ecotype (foothill-alpine morphological divergence), Region (divergence between mountain regions) and their interaction, calculated separately for each trait across all populations sampled. The effect size of Ecotype (*y*-axis) shows the extent to which a given trait diverges predictably between ecotypes, i.e., in parallel, while the Ecotype × Region effect size (*x*-axis) quantifies the extent to which foothill-alpine phenotypic divergence varies across regions (i.e., deviates from parallel). Points falling above the dashed line have a larger Ecotype effect (i.e., parallel) than Ecotype × Region interaction effect (i.e., non-parallel) while points falling below the dashed line have stronger Ecotype × Region interaction effect than the Ecotype effect. Correspondingly, Ecotype is non-significant for multiple traits (marked in gray). The traits are categorized according to plant organ to which they belong (different symbols). See **Supplementary Table S4** for explanation of individual codes of the traits.

the rise of this field (Turesson, 1930; Clausen et al., 1940), however, cases of phenotypic parallelism demonstrated by a combination of genetic and experimental data are still very rare in plant literature (Fustier et al., 2017). Interestingly, elevational differentiation has been revealed also in other *Arabidopsis* species (Fischer et al., 2013; Günther et al., 2016), often replicated across the studied populations (Kubota et al., 2015; Šrámková-Fuxová et al., 2017; Hämälä et al., 2018), making this genus also an attractive model for the study of alpine adaptation. However, except for European *A. halleri* (Šrámková-Fuxová et al., 2017) we lack a systematic quantitative assessment of morphological variation, leaving mostly unclear to which extent has parallel colonization been accompanied by parallel phenotypic shifts in other *Arabidopsis* species.

Alternatively, morphological similarities among the alpine populations may also reflect plastic response toward similar environmental conditions. To separate both effects, we grew plants of alpine and foothill origin from four regions in a common garden. Originally alpine plants kept their overall distinct appearance even in standard conditions, thus corresponding to the original definition of an ecotype *sensu* Turesson (1922). Importantly, the traits exhibiting the strongest parallelism in foothill-alpine divergence in the field were also highly distinctive in the common garden (stem height and flower size), supporting the hypothesis of parallel selection over phenotypic plasticity. It shall be noted, however, that alternative genetic yet non-adaptive mechanisms such as developmental constraints or mere chance may still stand behind such parallelism (Losos, 2011; Bolnick et al., 2018). In our case, however, we consider the adaptive scenario most likely given the power of four independent transitions ruling out pure chance, very similar direction of the environmental differentiation across regions, and correspondence of our traits with typical "alpine syndrome" observed across floras worldwide (Körner, 2003; Halbritter et al., 2018). In contrast, the regionally specific ecotypic differentiation detected in the field samples disappeared for the majority of such traits in cultivation, suggesting they rather manifested a plastic response to regionally specific conditions. Such plastic response may still represent a way how to cope with the local nuances of the challenging alpine environments (e.g., Anderson and Gezon, 2015), such a hypothesis would, however, require follow-up experimental tests.

The Degree and Sources of Non-parallelism

Even in systems exhibiting generally strong and genetically determined parallelism, both neutral and selective processes may cause significant non-parallel deviations in particular traits or populations (Stuart et al., 2017; Thompson et al., 2019). Although we observed significant non-parallelism within our set of traits, alpine *A. arenosa* populations showed overall phenotypic homogeneity that was relatively higher than that of their foothill ancestors (as indicated by relatively lower disparity

of the alpine ecotype). This contrasts to remarkably divergent phenotypic outcomes of independent shifts along the gradient of elevation in the other multi-parallel plant system, *Heliosperma* (Bertel et al., 2018).

For most of the traits exhibiting non-parallelism in the field data, ecotypic differentiation was no longer present under cultivation in the common garden, demonstrating phenotypic plasticity is the likely major driver of non-parallelism in our data. Only in one case (FG region) an additional trait (leaf lobes) discriminated foothill and alpine populations both in the field and in the common garden suggesting genetic determination of this component of non-parallelism. We consider a neutral scenario that initially divergent variation of the source (foothill) populations is likely responsible for this difference. Notably, the FG-foothill populations were more strongly phenotypically divergent from the other foothill regions than were their FG-alpine counterparts from the other alpine populations. Emergence of non-parallel traits in response to local adaptation is less likely due to overall lower phenotypic variation among the alpine populations and very similar environmental conditions across alpine sites, speaking against strong selection triggered by locally specific conditions. As parallelism informs about the role of selection, its detection is usually the prime aim of most studies (Sackton and Clark, 2019), although the regionally specific deviations provide valuable information on additional evolutionary processes affecting the predictability of evolution (Stuart et al., 2017).

CONCLUSION

Our study demonstrated that spatially isolated alpine environments harbor populations exhibiting remarkable phenotypic parallelism that is manifested by traits that are associated with alpine adaptation in general. Replicated phenotypic shifts in response to similar environmental pressures that are stable over two generations in cultivation suggest the role of selection triggered by the stressful alpine environment. Our findings open new avenues for studying convergent genetic underpinnings of phenotypic parallelism currently being addressed by a follow-up study (Bohutínská et al., 2020). Indeed, alpine *A. arenosa* is a particularly well-suited model for following such questions due to negligible neutral divergence between ecotypes within each region and lack of strong bottleneck, both of which mitigate potentially confounding signals of past demographic events. This, complemented by the genomic and genetic tractability of the *Arabidopsis* genus in general and this species in particular (highly variable outcrosser) make *Arabidopsis arenosa* a highly promising model for studying the genomic basis of parallel adaptation in plants.

DATA AVAILABILITY STATEMENT

The datasets presented in this study are provided as **Supplementary Datasets** (environmental parameters and morphological traits) and in online repositories (DNA sequences): NCBI BioProjects, accession nos: PRJNA484107 and PRJNA633005.

AUTHOR CONTRIBUTIONS

FK and KM designed the study. AK, MB, GŠ, DP, FK, and KM collected the data. AK, VK, GW, and VZ analyzed the data. FK and AK drafted the manuscript with contribution of all authors. All authors contributed to the article and approved the submitted version.

FUNDING

The project was funded by Czech Science Foundation (project 19-06632S to KM and 17-20357Y to FK), Charles University (GAUK project 708216 to AK), and Research Council of Norway (FRIPRO Mobility fellowship 262033 to FK). Additional support was provided by Ministry of Education Youth and Sports of the Czech Republic (7AMB18AT022 to GW). This work was also supported by the long-term research development project No. RVO 67985939 of the Czech Academy of Sciences.

ACKNOWLEDGMENTS

We thank Eliška Záveská, Magdalena Lučanová, Jana Mořkovská, Jana Smatanová, Peter Schönswetter, Karl Hülber, Stanislav Španiel, Jakub Hojka, Daniel Bohutínský, Jindřich Chrtek, Klára Kabátová, Frederick Rooks, and Martin Kolník for help with fieldwork. Erwann Arc, Dominik Kaplenig, and Ilse Kranner kindly provided plants cultivated in common garden. Computational resources were provided by the CESNET LM2015042 and the CERIT Scientific Cloud LM2015085. This manuscript has been released as a pre-print at BioRxiv, https://doi.org/10.1101/2020.02.13.948158 (Knotek et al., 2020).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.561526/full#supplementary-material

REFERENCES

Anderson, J. T., and Gezon, Z. J. (2015). Plasticity in functional traits in the context of climate change: a case study of the subalpine forb *Boechera stricta* (Brassicaceae). *Glob. Chang. Biol.* 21, 1689–1703. doi: 10.1111/gcb.12770

Arnold, B., Bomblies, K., and Wakeley, J. (2012). Extending coalescent theory to autotetraploids. *Genetics* 192, 195–204. doi: 10.1534/genetics.112.140582

Arnold, B., Kim, S.-T., and Bomblies, K. (2015). Single geographic origin of a widespread autotetraploid *Arabidopsis arenosa* lineage followed by interploidy admixture. *Mol. Biol. Evol.* 32, 1382–1395. doi: 10.1093/molbev/msv089

Bartok, A., Hurdu, B., Szatmari, P.-M., Ronikier, M., Puşcaş, M., Novikov, A., et al. (2016). New records for the high-mountain flora of the Făgăraş Mts. (southern Carpathians) with discussion on ecological preferences and distribution of studied taxa in the Carpathians. *Contr. Bot.* 51, 77–153.

Bertel, C., Rešetnik, I., Frajman, B., Erschbamer, B., Hülber, K., and Schönswetter, P. (2018). Natural selection drives parallel divergence in the mountain plant *Heliosperma pusillum* s.l. *Oikos* 127, 1355–1367. doi: 10.1111/oik.05364

Bohutínská, M., Vlcek, J., Yair, S., Laenen, B., Konecná, V., Fracassetti, M., et al. (2020). Genomic basis of parallel adaptation varies with divergence in *Arabidopsis* and its relatives. *bioRxiv [Preprint]*. doi: 10.1101/2020.03.24.005397

Bolnick, D. I., Barrett, R. D. H., Oke, K. B., Rennison, D. J., and Stuart, Y. E. (2018). (Non) parallel evolution. *Annu. Rev. Ecol. Evol. Syst.* 49, 303–330. doi: 10.1146/annurev-ecolsys-110617-062240

Butlin, R. K., Saura, M., Charrier, G., Jackson, B., André, C., Caballero, A., et al. (2014). Parallel evolution of local adaptation and reproductive isolation in the face of gene flow. *Evolution* 68, 935–949. doi: 10.1111/evo.12329

Clausen, J., Keck, D. D., and Hiesey, W. M. (1940). *Experimental Studies on the Nature of Species. I. Effect of Varied Environments on Western North American Plants*. Washington, DC: Carnegie Institution for Science.

Ellenberg, H. (1992). *Zeigerwerte der Pflanzen in Mitteleuropa. 3., Erweit. Aufl.* Göttingen: E. Goltze.

Elmer, K. R., Fan, S., Kusche, H., Spreitzer, M. L., Kautt, A. F., Franchini, P., et al. (2014). Parallel evolution of Nicaraguan crater lake cichlid fishes via non-parallel routes. *Nat. Commun.* 5:5168. doi: 10.1038/ncomms6168

Elmer, K. R., and Meyer, A. (2011). Adaptation in the age of ecological genomics: insights from parallelism and convergence. *Trends Ecol. Evol.* 26, 298–306. doi: 10.1016/j.tree.2011.02.008

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., and Foll, M. (2013). Robust demographic inference from genomic and SNP data. *PLoS Genet.* 9:e1003905. doi: 10.1371/journal.pgen.1003905

Fischer, M. A. (2008). *Exkursionsflora für Österreich, Liechtenstein und Südtirol*. Land Oberösterreich: OÖ. Landesmuseen.

Fischer, M. C., Rellstab, C., Tedder, A., Zoller, S., Gugerli, F., Shimizu, K. K., et al. (2013). Population genomic footprints of selection and associations with climate in natural populations of *Arabidopsis halleri* from the Alps. *Mol. Ecol.* 22, 5594–5607. doi: 10.1111/mec.12521

Fustier, M.-A., Brandenburg, J.-T., Boitard, S., Lapeyronnie, J., Eguiarte, L. E., Vigouroux, Y., et al. (2017). Signatures of local adaptation in lowland and highland teosintes from whole-genome sequencing of pooled samples. *Mol. Ecol.* 26, 2738–2756. doi: 10.1111/mec.14082

Galen, C. (1989). Measuring pollinator-mediated selection on morphometric floral traits: bumblebees and the alpine sky pilot, *Polemonium Viscosum*. *Evolution* 43, 882–890. doi: 10.1111/j.1558-5646.1989.tb05185.x

Günther, T., Lampei, C., Barilar, I., and Schmid, K. J. (2016). Genomic and phenotypic differentiation of *Arabidopsis thaliana* along altitudinal gradients in the North Italian Alps. *Mol. Ecol.* 25, 3574–3592. doi: 10.1111/mec.13705

Halbritter, A. H., Fior, S., Keller, I., Billeter, R., Edwards, P. J., Holderegger, R., et al. (2018). Trait differentiation and adaptation of plants along elevation gradients. *J. Evol. Biol.* 31, 784–800. doi: 10.1111/jeb.13262

Hämälä, T., Mattila, T. M., and Savolainen, O. (2018). Local adaptation and ecological differentiation under selection, migration, and drift in *Arabidopsis lyrata*. *Evolution* 72, 1373–1386. doi: 10.1111/evo.13502

Hedberg, I., and Hedberg, O. (1979). Tropical-alpine life-forms of vascular plants. *Oikos* 33, 297–307. doi: 10.2307/3544006

Johannesson, K., Panova, M., Kemppainen, P., André, C., Rolán-Alvarez, E., and Butlin, R. K. (2010). Repeated evolution of reproductive isolation in a marine snail: unveiling mechanisms of speciation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 365, 1735–1747. doi: 10.1098/rstb.2009.0256

Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24, 1403–1405. doi: 10.1093/bioinformatics/btn129

Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J., et al. (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* 484, 55–61. doi: 10.1038/nature10944

Knotek, A., Wos, G., Požárová, D., Konečná, V., Šrámková, G., Bohutínská, M., et al. (2020). Parallel alpine differentiation in *Arabidopsis arenosa*. *bioRxiv* [Preprint]. doi: 10.1101/2020.02.13.948158

Kolář, F., Fuxová, G., Záveská, E., Nagano, A. J., Hyklová, L., Lučanová, M., et al. (2016a). Northern glacial refugia and altitudinal niche divergence shape genome-wide differentiation in the emerging plant model *Arabidopsis arenosa*. *Mol. Ecol.* 25, 3929–3949. doi: 10.1111/mec.13721

Kolář, F., Lučanová, M., Záveská, E., Fuxová, G., Mandáková, T., Španiel, S., et al. (2016b). Ecological segregation does not drive the intricate parapatric distribution of diploid and tetraploid cytotypes of the *Arabidopsis arenosa* group (Brassicaceae). *Biol. J. Linn. Soc. Lond.* 119, 673–688. doi: 10.1111/bij.12479

Konečná, V., Nowak, M. D., and Kolář, F. (2019). Parallel colonization of subalpine habitats in the central European mountains by *Primula elatior*. *Sci. Rep.* 9:3294. doi: 10.1038/s41598-019-39669-2

Körner, C. (2003). *Alpine Plant Life: Functional Plant Ecology of High Mountain Ecosystems*. Heidelberg: Springer.

Koutecký, P. (2014). MorphoTools: a set of R functions for morphometric analysis. *Plant Syst. Evol.* 301, 1115–1121. doi: 10.1007/s00606-014-1153-2

Kubota, S., Iwasaki, T., Hanada, K., Nagano, A. J., Fujiyama, A., Toyoda, A., et al. (2015). A genome scan for genes underlying microgeographic-scale local adaptation in a wild *Arabidopsis* species. *PLoS Genet.* 11:e1005361. doi: 10.1371/journal.pgen.1005361

Langerhans, R. B. (2018). Predictability and parallelism of multitrait adaptation. *J. Hered.* 109, 59–70. doi: 10.1093/jhered/esx043

Levin, D. A. (2001). The recurrent origin of plant races and species. *Syst. Bot.* 26, 197–204. doi: 10.1043/0363-6445-26.2.197

Losos, J. B. (2011). Convergence, adaptation, and constraint. *Evolution* 65, 1827–1840. doi: 10.1111/j.1558-5646.2011.01289.x

Melzer, H. (1960). Neues und kritisches zur flora der Steiermark und des angrenzenden Burgenlandes. *Mitt. Naturwiss. Ver. Steiermark* 90, 85–102. doi: 10.1007/bf01633859

Měsíček, J., and Goliašová, K. (2002). "*Cardaminopsis* (C. A. Mey.) Hayek," in *Flóra Slovenska*, eds K. Goliašová and H. Šípošová (Bratislava: Veda), 388–415.

Monnahan, P., Kolář, F., Baduel, P., Sailer, C., Koch, J., Horvath, R., et al. (2019). Pervasive population genomic consequences of genome duplication in *Arabidopsis arenosa*. *Nat. Ecol. Evol.* 3, 457–468. doi: 10.1038/s41559-019-0807-4

Mráz, P., and Ronikier, M. (2016). Biogeography of the Carpathians: evolutionary and spatial facets of biodiversity. *Biol. J. Linn. Soc.* 119, 528–559. doi: 10.1111/bij.12918

Nosil, P., Crespi, B. J., and Sandoval, C. P. (2002). Host-plant adaptation drives the parallel evolution of reproductive isolation. *Nature* 417, 440–443. doi: 10.1038/417440a

Nosil, P., Egan, S. P., and Funk, D. J. (2008). Heterogeneous genomic differentiation between walking-stick ecotypes: "isolation by adaptation" and multiple roles for divergent selection. *Evolution* 62, 316–336. doi: 10.1111/j.1558-5646.2007.00299.x

Pachschwöll, C., and Pachschwöll, T. (2019). A new find of *Arabidopsis neglecta* (Brassicaceae) in the Svydovets Massif (Ukrainian Carpathians). *Ukr. Bot. J.* 76, 60–66. doi: 10.15407/ukrbotj76.01.060

Pawłowski, B. (1970). Remarques sur l'endémisme dans la flore des Alpes et des Carpates. *Vegetatio* 21, 181–243. doi: 10.1007/bf02269663

Pfennig, D. W., Wund, M. A., Snell-Rood, E. C., Cruickshank, T., Schlichting, C. D., and Moczek, A. P. (2010). Phenotypic plasticity's impacts on diversification and speciation. *Trends Ecol. Evol.* 25, 459–467. doi: 10.1016/j.tree.2010.05.006

Raj, A., Stephens, M., and Pritchard, J. K. (2014). fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics* 197, 573–589. doi: 10.1534/genetics.114.164350

Ravinet, M., Westram, A., Johannesson, K., Butlin, R., André, C., and Panova, M. (2016). Shared and nonshared genomic divergence in parallel ecotypes of *Littorina saxatilis* at a local scale. *Mol. Ecol.* 25, 287–305. doi: 10.1111/mec.13332

Reid, N. M., Proestou, D. A., Clark, B. W., Warren, W. C., Colbourne, J. K., Shaw, J. R., et al. (2016). The genomic landscape of rapid repeated evolutionary adaptation to toxic pollution in wild fish. *Science* 354, 1305–1308. doi: 10.1126/science.aah4993

Roda, F., Ambrose, L., Walter, G. M., Liu, H. L., Schaul, A., Lowe, A., et al. (2013a). Genomic evidence for the parallel evolution of coastal forms in the *Senecio lautus* complex. *Mol. Ecol.* 22, 2941–2952. doi: 10.1111/mec.12311

Roda, F., Liu, H., Wilkinson, M. J., Walter, G. M., James, M. E., Bernal, D. M., et al. (2013b). Convergence and divergence during the adaptation to similar environments by an Australian groundsel. *Evolution* 67, 2515–2529. doi: 10.1111/evo.12136

Rundle, H. D., Nagel, L., Boughman, J. W., and Schluter, D. (2000). Natural selection and parallel speciation in sympatric sticklebacks. *Science* 287, 306–308. doi: 10.1126/science.287.5451.306

Sackton, T. B., and Clark, N. (2019). Convergent evolution in the genomics era: new insights and directions. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 374:20190102. doi: 10.1098/rstb.2019.0102

Soria-Carrasco, V., Gompert, Z., Comeault, A. A., Farkas, T. E., Parchman, T. L., Johnston, J. S., et al. (2014). Stick insect genomes reveal natural selection's role in parallel speciation. *Science* 344, 738–742. doi: 10.1126/science.1252136

Šrámková-Fuxová, G., Záveská, E., Kolář, F., Lučanová, M., Španiel, S., and Marhold, K. (2017). Range-wide genetic structure of *Arabidopsis halleri* (Brassicaceae): glacial persistence in multiple refugia and origin of the Northern Hemisphere disjunction. *Bot. J. Linn. Soc.* 185, 321–342. doi: 10.1093/botlinnean/box064

Stift, M., Kolář, F., and Meirmans, P. G. (2019). Structure is more robust than other clustering methods in simulated mixed-ploidy populations. *Heredity* 123, 429–441. doi: 10.1038/s41437-019-0247-6

Stuart, Y. E., Veen, T., Weber, J. N., Hanson, D., Ravinet, M., Lohman, B. K., et al. (2017). Contrasting effects of environment and genetics generate a continuum of parallel evolution. *Nat. Ecol. Evol.* 1:0158. doi: 10.1038/s41559-017-0158

Thompson, K. A., Osmond, M. M., and Schluter, D. (2019). Parallel genetic evolution and speciation from standing variation. *Evol. Lett.* 3, 129–141. doi: 10.1002/evl3.106

Tichý, L. (2002). JUICE, software for vegetation classification. *J. Veg. Sci.* 13, 451–453. doi: 10.1111/j.1654-1103.2002.tb02069.x

Trucchi, E., Frajman, B., Haverkamp, T. H. A., Schönswetter, P., and Paun, O. (2017). Genomic analyses suggest parallel ecological divergence in *Heliosperma pusillum* (Caryophyllaceae). *New Phytol.* 216, 267–278. doi: 10.1111/nph.14722

Turesson, G. (1922). The species and the variety as ecological units. *Hereditas* 3, 100–113. doi: 10.1111/j.1601-5223.1922.tb02727.x

Turesson, G. (1930). The selective effect of climate upon the plant species. *Hereditas* 14, 99–152. doi: 10.1111/j.1601-5223.1930.tb02531.x

Wos, G., Mořkovská, J., Bohutínská, M., Šrámková, G., Knotek, A., Lučanová, M., et al. (2019). Role of ploidy in colonization of alpine habitats in natural populations of *Arabidopsis arenosa*. *Ann. Bot.* 124, 255–268. doi: 10.1093/aob/mcz070

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Case study 3

Parallel colonization of subalpine habitats in the central European mountains by *Primula elatior*

SCIENTIFIC REPORTS

OPEN

Parallel colonization of subalpine habitats in the central European mountains by *Primula elatior*

Received: 20 July 2018
Accepted: 30 January 2019
Published online: 01 March 2019

Veronika Konečná[1,2], Michael D. Nowak[3] & Filip Kolář[1,2,4]

The island-like distribution of subalpine habitats across mountain ranges can trigger the parallel evolution of locally adapted ecotypes. Such naturally replicated scenarios allow testing hypotheses on how elevational differentiation structures genetic diversity within species. Nevertheless, the parallel colonization of subalpine habitats across different mountain ranges has only rarely been documented with molecular data. We chose *Primula elatior* (Primulaceae), naturally spanning entire elevation range in multiple mountain regions of central Europe, to test for the origin of its scattered subalpine populations. Nuclear microsatellite variation revealed three genetic groups corresponding with the distinct study regions. We found that genetic differentiation between foothill and subalpine populations within each region was relatively low, suggesting that the colonization of subalpine habitats occurred independently within each mountain range. Furthermore, the strongest differentiation was usually found between the subalpine populations suggesting that mountain ridges may act as migration barriers that can reduce gene flow more strongly than elevational differences between foothill and subalpine populations. Finally, we found that subalpine colonization did not result in a loss of genetic diversity relative to foothill populations in agreement with the high migration rates that we document here between the subalpine and the foothill populations. In summary, our study shows subalpine *Primula elatior* populations are genetically diverse and distinct results of parallel colonization events from multiple foothill gene pools.

Alpine and subalpine habitats represent challenging and often unpredictable environments for plants. Plants have to face a complex set of stresses in such environments including freezing, fluctuating temperatures, increased UV radiation, and terrain disturbances implying various selection pressures[1]. Local adaptation to these environmental stresses often results in the evolution of specific morphological and/or physiological traits that confer increased fitness in these challenging environments[2,3]. Populations characterized by these adaptations are often called "ecotypes" and their level of overall genetic differentiation from ancestral populations may be still very low[2,4]. In cases where reproductive barriers arise between ecotypes and their ancestral populations, either as a consequence of natural selection or simply due to spatial isolation, the ecotype may represent the first step towards the founding of a new species[5,6]. In such cases, we can observe speciation as a continuum of divergence along an elevational gradient leading to textbook examples of ecotypic differentiation[7,8]. The restriction of gene flow between lower and higher elevations as well as the reduction or absence of gene flow among mountain ridges can lead to the accumulation of reproductive barriers and/or hybrid incompatibilities[9]. This process is but one of several potential explanations for relatively high species diversity in alpine/subalpine habitats throughout the world[10].

The island-like distribution of alpine/subalpine habitats can trigger allopatric differentiation and parallel colonization from lower elevations independently in each mountain range[11]. It can also lead to phenotypic differentiation potentially resulting in the convergent evolution of traits that confer local adaptation (i.e. ecotypes)[2]. Although the mountains provide an ideal model system for studying the evolution of recurrently adapted ecotypes, the parallel colonization of alpine/subalpine habitats has only been rarely documented with molecular data. Rare exceptions to this are represented by studies in *Arabidopsis halleri*[12], *Zea mays*[13], and *Populus*

[1]Department of Botany, Faculty of Science, Charles University, Benátská 2, CZ-128 00, Prague, Czech Republic. [2]Institute of Botany, The Czech Academy of Sciences, Zámek 1, CZ-252 43, Průhonice, Czech Republic. [3]Natural History Museum, University of Oslo, Sars' gate 1, NO-0562, Oslo, Norway. [4]Department of Botany, University of Innsbruck, Sternwartestraße 15, AT-6020, Innsbruck, Austria. Correspondence and requests for materials should be addressed to V.K. (email: konecnv@natur.cuni.cz)

trichocarpa[14]. The alternative scenario of a single origin of alpine/subalpine ecotype followed by dispersal has also been documented, for example in *Senecio halleri* whose alpine populations represent a single lineage that colonized the Alps in a stepwise manner[15].

The influence of habitat differentiation on the genetic diversity of populations has been previously studied mainly at lower elevations[16–18]. It is poorly known how the colonization of alpine/subalpine h abitats has shaped the genetic diversity of populations, but it is hypothesized that alpine/subalpine populations would be genetically depauperate relative to lowland populations due to the effects of genetic bottleneck, genetic drift, reduced gene flow, and habitat fragmentation[19], all of which can promote the genetic isolation of alpine/subalpine populations from their foothill relatives[20]. This has indeed been observed in alpine populations of *Arabidopsis thaliana* from the Italian Alps, which exhibit reduced genetic diversity relative to foothill populations[21], but in contrast, alpine *A. thaliana* populations from the Swiss Alps show no evidence of reduced genetic diversity[22]. In *Primula merrilliana* from eastern China, alpine populations exhibit even higher genetic diversity, moreover, they are larger and more inter-connected with gene flow than foothill populations[23]. This pattern implies that the foothill populations may have been colonized by alpine populations. In summary, studies of recurrently originated ecotypes of a single species may provide valuable replicates to test the generality of how alpine conditions shape population genetic diversity.

In this study, we address the genetic consequences of elevational differentiation in *Primula elatior* (Primulaceae), a species with broad ecological preferences including a large elevational range in several mountain ranges in Europe[24]. The wide ecological breadth is linked to high morphological variation primarily in leaf and calyx shape[24,25]. Individuals originally from subalpine populations tended to have urceolate calyxes (the narrowest in the top part of calyx), compared to individuals from foothill populations mainly with tubular calyxes (the same wide along the calyx)[26]. The morphological traits are plastic, except the shape of calyx, which remained stable after cultivation of subalpine population under uniform foothill conditions in a common garden experiment in one of the mountain ranges (the Krkonoše mountain range, V. Konečná unpubl.). In contrast to morphological investigations, the genetic structure of *P. elatior* remains unknown except studies at fine-scale in Belgium[16,27–29]. We focus on three mountain ranges in central Europe, where *P. elatior* grows along an elevational gradient from foothill meadows, river banks, and forest edges, up to subalpine meadows, snowbeds, and rocky outcrops in the glacial cirques. In the Krkonoše and the Jeseníky mountains, subalpine populations are restricted to glacial cirques. In the Tatry mountains, subalpine populations grow in valley meadows and snowbeds. Generally, the difference between foothill and subalpine populations across mountain ranges is defined by treeline, foothill populations occur below the treeline in contrast to subalpine populations, which occur above the treeline. We assume colonization of subalpine habitats from foothill habitats during warmer periods of the postglacial Holocene. Our assumption of upslope colonization is likely because temperate species are highly unlikely to have survived past glaciations that affected these subalpine habitats[30].

Using *P. elatior* as a suitable system, we examined how elevation shapes genetic structure within a species, and by comparing these results across three distinct mountain ranges (the Jeseníky, the Krkonoše, and the Tatry), we evaluated how generally applicable these patterns of differentiation might be in *P. elatior*. First, we tested whether the subalpine populations in three mountain ranges represent parallel colonization events of subalpine habitats occurring independently in each mountain range or if the subalpine ecotype evolved once and later spread across the different mountain ranges. Second, we tested if the elevation acts as a barrier to gene flow between foothill and subalpine populations, and whether gene flow is asymmetric; e.g. are migration events from high to low elevations following the downslope transport of seeds and pollen more common than upslope migration events. Finally, we tested for a reduction of genetic diversity associated with the colonization of subalpine habitats relative to lowland habitats.

Results

Genetic structure and evolutionary relationships among populations. By genotyping 12 nuclear microsatellite loci in 202 individuals from 16 populations we detected a total of 120 alleles with maximum of 24 and minimum of three alleles per locus.

We explored the genetic structure of *P. elatior* populations across the three target mountain ranges, where it occupies both foothill and subalpine habitats, using Bayesian clustering (STRUCTURE), distance networks (Neighbor-joining networks), and ordinations (principal component analysis, PCA). The results of the STRUCTURE analyses showed that populations from each mountain range (the Jeseníky, the Krkonoše, and the Tatry) formed a separate cluster, regardless of their foothill-subalpine differentiation, under the corresponding partition of $K = 3$ (Fig. 1A). The model of $K = 3$ also exhibited the highest similarity among the replicated runs and at this partition the rise of likelihood values started to flatten, suggesting that the inclusion of additional parameters do not significantly improve the fit of the model beyond $K = 3$ (Supplementary Fig. S1). We observed three clusters according to the three mountain ranges in PCA (Fig. 1C) as well as in Neighbor-joining network based on F_{ST} distances (Fig. 1B). The separation of populations into three clusters that largely corresponded to geographical range and the absence of any genetic structure associated with elevation strongly suggested the parallel evolution of the subalpine populations in each mountain range.

In subsequent analyses, we performed separate STRUCTURE analyses for each region to investigate finer structure within each mountain range (Fig. 2). All the analyses tended to separate the subalpine and foothill populations within each region: this was apparent already under $K = 2$ in the Jeseníky and under $K = 3$ in the other two mountain ranges. In the Krkonoše and the Tatry, one subalpine population (S4 and S5, respectively) separated from the remaining populations under $K = 2$, suggesting that not only one group of subalpine populations exists in these mountain ranges.

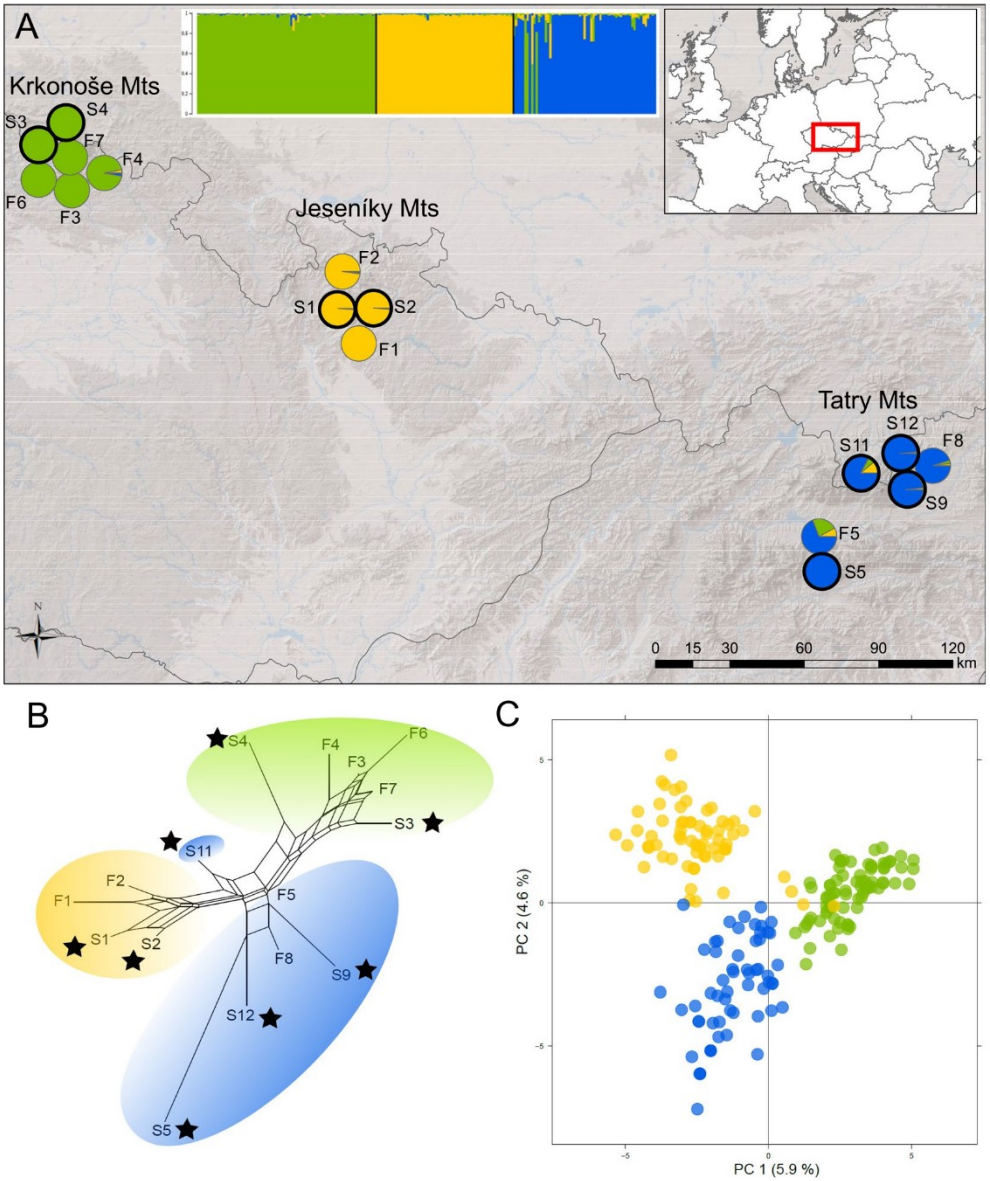


Figure 1. Parallel origin of subalpine populations of *Primula elatior* in the three mountain ranges in central Europe. (**A**) Geographical distribution of the complete dataset of 16 foothill and subalpine (marked by bold circles) populations of *P. elatior*, pie charts and barplots show the proportional assignment of individuals from each population to the three clusters inferred by STRUCTURE, (**B**) neighbor-joining network of the same populations based on among-population F_{ST} distances (subalpine populations marked by asterisks), (**C**) principal component analysis (PCA) of all 202 individuals (coloured according to STRUCTURE results). Map was processed in ArcMap version 10.0 (http://desktop.arcgis.com/en/arcmap/) by V. Konečná, map layer was modified (http://desktop.arcgis.com/en/arcmap/10.3/manage-data/raster-and-images/hillshade-function.htm) by V. Konečná.

Genetic diversity of populations and among population variation. The populations varied considerably in terms of within population variation. Allelic richness ranged from 2.29 to 3.72 with the highest values (3.33–3.72) being present in foothill populations in the Tatry (Table 1, Supplementary Fig. S2). Similarly, expected heterozygosity (H_E), ranged from 0.4 to 0.6 in the whole dataset and foothill populations from the Tatry exhibited the highest values (0.52–0.6). Populations also varied in gene diversity (H_S) ranging from 0.43 to 0.65 with the highest values (0.65–0.61) in one foothill and two subalpine populations from the Tatry. The proportion of rare alleles (DW index)[31] exhibited a similar trend as well, with the range from 82.44 in the Jeseníky subalpine population to 411.79 in the Tatry foothill population. Despite this variation, we did not detect consistent differences between the group of foothill and subalpine populations across three mountain ranges in diversity index

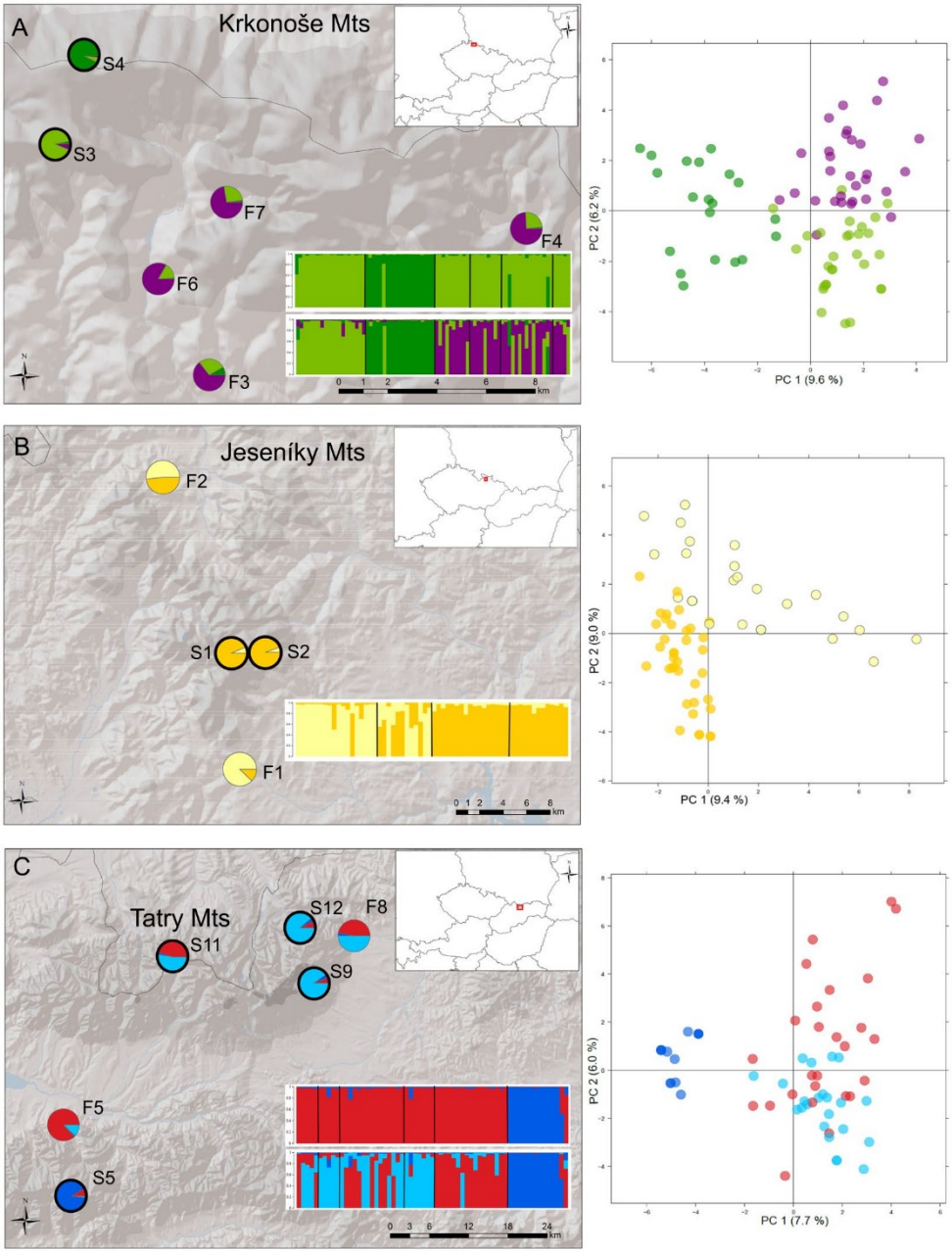


Figure 2. Genetic sub-structuring of *P. elatior* populations in each target mountain range. Pie charts and barplots show the proportional assignment of individuals from each population to clusters inferred by a separate STRUCTURE analysis of (**A**) the Krkonoše (79 individuals), (**B**) the Jeseníky (60 individuals), and (**C**) the Tatry (63 individuals); accompanied by principal component analysis (PCA) of individuals. Subalpine populations are denoted by bold circles. Maps were processed in ArcMap version 10.0 (http://desktop.arcgis.com/en/arcmap/) by V. Konečná, map layers were modified (http://desktop.arcgis.com/en/arcmap/10.3/manage-data/raster-and-images/hillshade-function.htm) by V. Konečná.

(non-significant ANOVA with region as a random factor) (Table 2). Instead, each mountain range tended to show a different pattern (Supplementary Fig. S2).

Among-population differentiation (pairwise F_{ST}) was low, but varied considerably from 0.04 to 0.28, with an average of 0.13 (Supplementary Table S3). In the Krkonoše and the Tatry, we observed lower F_{ST} values between foothill populations (average of 0.06 for six comparisons in the Krkonoše, and 0.06 for one comparison in the Tatry) compared to those between subalpine populations (on average 0.14 for six comparisons in the Tatry and 0.13 for one comparison in the Krkonoše). In contrast, two subalpine populations in the Jeseníky, that occupied

Pop. ID	Locality	Range	Group	Elevation	No. Ind.	No. Alleles	Allelic richness	H_O	H_E	H_S	DW index	No. private alleles	HWE test[a]	MIGRATE-N
F1	Žďárský potok u Rýmařova	Jeseníky	foothill	735	18	41	2.6	0.43 ± 0.28	0.44 ± 0.26	0.43	240.76	2		✓
F2	V Mlýnkách	Jeseníky	foothill	618	13	44	2.71	0.35 ± 0.29	0.46 ± 0.28	0.45	121.95	1	PRIV 4**, PACA 78**, PV 4767*, PV 23741*	✓
S1	Velká Kotlina	Jeseníky	subalpine	1397	13	36	2.61	0.49 ± 0.36	0.46 ± 0.30	0.54	82.44	0	PACA 78*, PV 23741*	✓
S2	Velká Kotlina edge	Jeseníky	subalpine	1301	16	39	2.6	0.46 ± 0.30	0.48 ± 0.27	0.51	84.78	0	PRIV 4*, PACA 78*	✓
F3	Strážné	Krkonoše	foothill	776	15	45	2.75	0.41 ± 0.37	0.46 ± 0.33	0.49	175.15	2	PRIV 4***, PACA 38**	✓
F4	Spálený mlýn	Krkonoše	foothill	788	10	39	2.68	0.41 ± 0.35	0.46 ± 0.30	0.49	84.45	0	PACA 38*	✓
F6	Michlův mlýn	Krkonoše	foothill	658	9	35	2.52	0.39 ± 0.29	0.43 ± 0.26	0.45	150.71	1	PACA 38*	
F7	Svatý Petr	Krkonoše	foothill	816	5	34	2.82	0.43	0.5	0.57	138.8	0		
S3	Velká Kotelní jáma	Krkonoše	subalpine	1280	20	45	2.62	0.40 ± 0.32	0.45 ± 0.31	0.53	145.06	4	PRIV 4***, PACA 78**	✓
S4	Malá Sněžná jáma	Krkonoše	subalpine	1368	20	41	2.52	0.34 ± 0.32	0.44 ± 0.32	0.52	227.05	1	PRIV 4***, PACA 78**, PV 4767*, PACA 38*	✓
F5	Poludnica	Tatry	foothill	1055	17	68	3.72	0.57 ± 0.33	0.62 ± 0.29	0.61	411.79	9	PRIV 4***, PV 1973**	✓
F8	Ždiar, Tatranská Kotlina	Tatry	foothill	769	15	63	3.33	0.48 ± 0.34	0.54 ± 0.33	0.58	380.05	5	PACA 78***	
S5	Štefánikova chata	Tatry	subalpine	1709	14	33	2.29	0.39 ± 0.43	0.41 ± 0.29	0.49	167.63	1	PV 279**, PACA 78***, PRIV 7**, PV 4767*, PV 23741*	✓
S9	Zamkovského chata	Tatry	subalpine	1470	7	34	2.66	0.54 ± 0.34	0.49 ± 0.28	0.57	314.54	1		
S11	Kondraczka	Tatry	subalpine	1950	5	43	3.53	0.5	0.51	0.65	227.41	1		
S12	Tristar	Tatry	subalpine	1450	5	34	2.82	0.43	0.44	0.62	195.04	0		

Table 1. Genetic diversity of 16 populations of *Primula elatior* from foothill and subalpine habitats in the three mountain ranges investigated. [a]Only loci with significant deviation from HWE are listed; P-values were estimated by 1000 permutations ($^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$).

different parts of one glacial cirque, were less differentiated from each other (0.04) than were their two foothill counterparts (0.06). Between the foothill and subalpine populations, we observed at most a moderate differentiation, with the highest mean pairwise F_{ST} values in the Krkonoše (0.10) followed by the Tatry (0.09), and the lowest values in the Jeseníky (0.08).

The results of an AMOVA (analysis of molecular variance) confirmed high intrapopulation variability by assigning the highest proportion of variation always among individuals within populations (from 97.24% to 73.27%). In congruence with pairwise F_{ST} values, two subalpine populations in the Krkonoše and four in the Tatry were more differentiated from each other than their foothill counterparts: 21.42% and 24.18% of among-subalpine population variation compared to 5.47% and 3.03% of among-foothill population variation in the Krkonoše (four populations) and the Tatry (two populations), respectively. The lowest variation (2.76%), although still significant, was found between subalpine populations from the Jeseníky. Finally, hierarchical AMOVAs within each mountain range showed that the foothill-subalpine differentiation was non-significant and accounted for little variation (from 2.32% to 6.11%, Table 2). To conclude, even though the STRUCTURE results revealed some differentiation between foothill and subalpine populations, this differentiation is still markedly lower than differentiation observed between subalpine populations in two of the three mountain ranges.

Gene flow, migration rates and model selections. We further addressed the role of migration in foothill-subalpine population differentiation by testing for the presence of gene flow and estimating the strength of gene flow between populations in a coalescent framework using MIGRATE-N. Within each region, we analysed pairs of representatively sampled (≥10 individuals per population) subalpine-subalpine, foothill-subalpine, and foothill-foothill populations; for each pair, we modelled past evolutionary history under Bayesian inference of five migration models differing in presence and directionality of migration. The models with unidirectional migration gained the best support in all population pairs analysed (Table 3, Fig. 3, Supplementary Table S4). In the case of the foothill-subalpine pairs, models of unidirectional migration from subalpine to foothill populations were consistently the best models fit across different population pairs from different mountain ranges. The estimated number of immigrants (*Nm*) was overall high (>1) (ranging from 5 to 551 with an average of 160), which suggests that migration via ongoing gene flow has probably a greater effect on the extent of differentiation between

Group	No. of ind./ pop.	No. of alleles	No. of private alleles	Allelic richness	H_E	H_S	DW index	F_{ST}	Among populations variation (%)	Pairwise F_{ST} among populations (min-max values)	Among groups variation (%)
Jeseníky Mts											
Subalpine populations	29/2	44	0	2.61 ± 0.005	0.47 ± 0.01	0.53 ± 0.015	83.61 ± 1.17	0.03	2.76*	0.035	
Foothill populations	31/2	53	3	2.66 ± 0.055	0.45 ± 0.01	0.44 ± 0.01	181.36 ± 59.41	0.08	7.82***	0.057	
Subalpine × foothills											6.11 (p = 0.33)
Krkonoše Mts											
Subalpine populations	40/2	56	5	2.57 ± 0.05	0.45 ± 0.005	0.53 ± 0.005	186.06 ± 41	0.21	21.42***	0.132	
Foothill populations	39/4	60	3	2.69 ± 0.093	0.46 ± 0.019	0.50 ± 0.035	137.28 ± 26.41	0.05	5.47***	(0.035) 0.059 (0.08)	
Subalpine × foothills											4.26 (p = 0.14)
Tatry Mts											
Subalpine populations	31/4	68	3	2.83 ± 0.352	0.46 ± 0.038	0.58 ± 0.053	226.16 ± 44.82	0.24	24.18***	(0.065) 0.125 (0.174)	
Foothill populations	32/2	87	14	3.53 ± 0.195	0.58 ± 0.04	0.6 ± 0.015	395.92 ± 15.87	0.03	3.03 (p = 0.1)	0.06	
Subalpine × foothills											2.32 (p = 0.25)
Significance of differences between groups of foothill and subalpine populations				$F_{1,12} = 0.09$	$F_{1,12} = 0.09$	$F_{1,12} = 0.3$	$F_{1,12} = 0.08$				

Table 2. Genetic diversity and differentiation of foothill and subalpine populations within the target mountain ranges. P-values were estimated by 1000 permutations (*P < 0.05, **P < 0.01, ***P < 0.001). Additionally, we calculated for subalpine pop. S5 from the Nízké Tatry Mts and foothill pop. F5 and F8 among population variation (subalpine × foothill pop.) = 23.82% (p = 0.33), $F_{ST} = 0.27$, and pairwise F_{ST} among populations = (0.06) 0.11 (0.137).

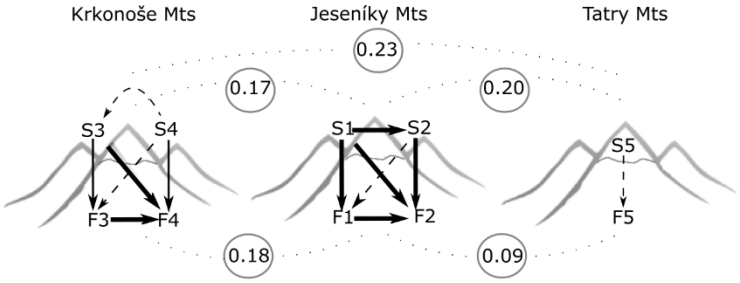


Figure 3. Migration and differentiation of foothill (F) and subalpine (S) populations of *P. elatior* in the three mountain ranges of central Europe. The values between mountain ranges are overall mean pairwise F_{ST} between foothill and subalpine populations. The arrows indicate preferred migration scenario, their width reflects estimated number of immigrants per generation (*Nm*): dashed arrow (5–21), standard arrow (68–73), and thick arrow (161–551). Created by V. Konečná.

populations than genetic drift (*Nm* values >1)[32]. However, besides migration, the large estimates for the number of immigrants could also at least partly reflect the recent shared ancestry of populations[33].

We observed the lowest estimated migration rates between subalpine populations in the Krkonoše (*Nm* = 5), where subalpine populations are separated by a mountain ridge. In contrast, in the Jeseníky, a relatively weak barrier to gene flow appears to exist between the two populations residing within one glacial cirque based on the high migration rate estimated (*Nm* = 266). The highest number of immigrants in the entire dataset was also found between two foothill populations (Krkonoše, F3 and F4; *Nm* = 551) (Table 3, Fig. 3).

Discussion

Our population genetic investigation of central European *Primula elatior* populations provided strong evidence for parallel colonization of subalpine habitats within each of the three mountain ranges that are the focus of our study (Fig. 1). In each case, subalpine populations are more closely related to the geographically closest foothill populations from the same mountain range than to their subalpine counterparts from other mountain ranges. Although island-like distributed subalpine environments are likely to trigger the evolution of recurrently adapted populations, the parallel colonization of subalpine habitats across different mountain ranges in one species has only rarely been documented with genetic data[12–14]. Future studies on parallel ecotype evolution should also

Range	Model	*Nm*	Pairwise F_{ST}
Jeseníky	F1 → F2	206	0.06
Krkonoše	F3 → F4	551	0.05
Jeseníky	S1 → F1	261	0.08
Jeseníky	S2 → F1	21	0.09
Jeseníky	S1 → F2	161	0.07
Jeseníky	S2 → F2	173	0.07
Krkonoše	S3 → F3	68	0.07
Krkonoše	S4 → F3	17	0.11
Krkonoše	S3 → F4	257	0.09
Krkonoše	S4 → F4	73	0.13
Nízké Tatry	S5 → F5	14	0.13
Jeseníky	S1 → S2	266	0.04
Krkonoše	S4 → S3	5	0.13

Table 3. Summary of the best fitting migration models (based on Bayes factor, see Supplementary Table S4) for foothill and subalpine populations of *P. elatior*, their differentiation, and numbers of immigrants per generation (*Nm*) between populations estimated by MIGRATE-N. F = foothill populations, S = subalpine populations.

focus on mechanisms behind the genetic parallelism. For instance, whether repeated selection from the standing genetic variation has important role in formation of parallel ecotypes[14,34], whether the selection of independent mutations takes place in parallel evolution[13], or if adaptive introgression via borrowed alleles from adapted species can facilitate this process[35].

The example of the parallel origin of ecotypes along an elevational gradient, although with the opposite direction of colonization than in our study, was recently documented for *Heliosperma pusillum*, where distinct isolated ecotypes occupying cave entrances in foothills originated from a widespread alpine lineage[36]. The impact of elevational difference on divergence between foothill (montane) and alpine populations has been documented in *Dianthus callizonus*[37], *Primula merrilliana*[23], and *Solidago virgaurea*[38]. The parallel origin of ecotypes has been also documented previously with genetic data in different environments such as sand dune vs. non-dune in *Helianthus petiolaris*[39], sand dune vs. rock in *Senecio lautus*[40], serpentine vs. non-serpentine soils in *Solidago virgaurea*[41], wave vs. crab predation in the mollusc *Littorina saxatilis*[42], and freshwater vs. saltwater in the fish *Gasterosteus aculeatus*[34]. The absence of similar studies in subalpine plants might simply reflect less attention to this phenomenon of differentiation in subalpine environment.

In contrast to differentiation among the three distinct mountain ranges (mean pairwise $F_{ST} = 0.13$), foothill and subalpine populations within each region were less differentiated (pairwise F_{ST} ranging from 0.08 to 0.10) and accounted for a negligible and non-significant proportion of variation in AMOVA tests (from 2.32% to 6.11% variation). The variation in these values across mountain ranges likely reflects different spatial and elevational distances between subalpine and foothill populations in each region as well as heterogeneity of habitats. In the Jeseníky and northern part of the Tatry (S9, S11; Vysoké/Belanské Tatry) we observed a nearly continuous occurrence of *P. elatior* along an elevational gradient (corresponding with mean pairwise F_{ST} between foothill and subalpine populations of 0.08 and 0.07, respectively)[43,44]. In contrast, in the Krkonoše and the Nízké Tatry, there was a gap in the occurrence of several hundred meters, which can restrict gene flow (reflected by higher mean pairwise F_{ST} of 0.10 and 0.13, respectively), as was also revealed in *Knautia* in the Krkonoše[45]. Available population genetic studies have documented that the levels of differentiation among foothill and alpine/subalpine populations may vary considerably among species, from high (mean pairwise $F_{ST} = 0.19$ in *Primula merrilliana*[23]) to low (mean pairwise $F_{ST} = 0.02$ in *Arabidopsis arenosa*[46] also from the Tatry).

Our study shows that genetic differentiation along an elevational gradient may vary within a species, from one mountain range to another, although this pattern appears to be less extreme in *Primula* relative to *Heliosperma*, where the individual alpine-foothill pairs varied considerably in differentiation (mean F_{ST} 0.17–0.43)[36]. The extent of differentiation likely indicates if populations have an older or younger origin/divergence. Besides elevational differences, lower or comparable differentiation to our study was also observed in two *Primula* species between grassland and forest habitats in lowland habitats at a fine-scale of several kilometres (mean $F_{ST} = 0.02$ in *P. elatior*[16] and 0.08 in *P. veris*)[47]. Our observation in *P. elatior* thus suggests that elevation is generally a weak isolating barrier whose strength may further vary across independent colonization events in different mountains.

The relatively low foothill-subalpine differentiation that we observed in *P. elatior* is likely due to high levels of retained ancestral polymorphism and/or persisting gene flow along the elevational gradient. This was suggested by coalescent modelling in which we found consistent support for models with unidirectional migration from subalpine to foothill populations. The unidirectional migration likely reflected gene flow after the colonization of subalpine habitats in postglacial times. This was the best supported model in all three mountain ranges, however, with varying estimated strength across the regions. Such unidirectional downslope migration has been observed also in *P. merrilliana*[23], in which the seed mobility is generally restricted in a similar way to *P. elatior* due to the lack of any long-distance seed dispersal adaptation[48]. In the mountains, seed flow may be further enhanced by washing downslope by mountain streams, but pollen flow may be restricted along an elevational gradient due to differences in flowering phenology between populations at different elevations[4]. In the closely related species *P. veris*, differences in the flowering phenology affect gene flow even among distinct habitats in the same (lowland) elevation[47].

In contrast to the low foothill-subalpine differentiation that we observed, individual subalpine populations were the most highly differentiated entities within two of the three studied regions (21.42% and 24.18% variation in the Krkonoše and the Tatry, respectively, split into distinct genetic clusters, Fig. 2). For generalization of the differentiation of subalpine populations, however, more populations shall be compared. High differentiation between subalpine populations in the Krkonoše has been documented also in another species growing in the glacial cirques – *Gentiana pannonica* (14.15%)[49]. The Jeseníky was the only exception, in which both subalpine populations were restricted to a single glacial cirque with likely high opportunities for gene flow and this is likely responsible for the low among-population differentiation and F_{ST} values. These results suggest that mountain ridges may act as relatively strong migration barriers, which can affect gene flow more strongly than elevational differences between subalpine and foothill populations. Although higher differentiation among subalpine populations may also reflect past bottleneck events during their origin from their foothill counterparts, we do not consider this likely. We have not observed significantly reduced levels of diversity in these populations (Table 2) – a clear sign of past bottleneck in populations[50].

In contrast to the differentiation observed among subalpine populations, differentiation among foothill populations was low (3.03–7.82% in different regions). This relatively low differentiation is may be the product of a higher density and large population size of foothill populations. Pollination is more efficient in large populations and the larger population size may provide more opportunities for gene flow among populations[16,51].

Our results show that subalpine colonization did not appear to lead to a loss of genetic diversity relative to foothill populations. The differences in genetic diversity of subalpine vs. foothill populations were non-significant over the replicated mountain ranges (Table 2). Moreover, neither index of genetic diversity shows any consistent trend across the regions with respect to the elevational groups (Supplementary Fig. S2). On the other hand, foothill populations from the Tatry were together the most genetically diverse of all populations included in our study. This could be a consequence of the foothill habitat[16,17,52], which may have experienced the long-term isolation of populations in this area serving as a glacial refugium for temperate species[46,53]. Generally, subalpine populations of *P. elatior* appear to be able to maintain genetic diversity equal to foothill populations despite smaller population sizes and the spatial isolation of the glacial cirques in which they occur. This could be due to a large size of the initial colonizing population (ruling out founder effect) and the relative stability of the habitat in postglacial time providing good conditions for the persistence of sufficiently large populations. An additional non-exclusive explanation might be a gradual and relatively slow pace of the shifts of the treeline during the Holocene, which might have maintained sufficient population sizes without bottlenecks, and thus preserve genetic diversity during the colonization process[23].

In summary, we describe a case of parallel colonization of subalpine habitats from multiple foothill gene pools. Our results imply that there is a distinct mountain diversity in the subalpine habitats – the subalpine populations are genetically differentiated from their foothill counterparts and from each other. In addition, subalpine populations regularly preserve genetic diversity at similar levels relative to their foothill counterparts. However, ongoing gene flow, in particular from subalpine to foothill habitats together with low levels of differentiation, likely linked to a recent (postglacial) origin of the subalpine populations, seems to prevent any stronger differentiation along the elevational gradient that may lead to speciation.

Methods

Study species and sampling. *Primula elatior* is an outcrossing perennial plant with a basal rosette of leaves. Flowers are produced in an umbel, characterized by distyly and self-incompatibility[25]. The main pollinators are Hymenoptera (mostly bumblebees) and Diptera[54,55]. Seed mobility in this species is generally restricted due to lack of any adaptations for long-distance dispersal, for instance, compared to closely related species *P. vulgaris*[48]. Therefore, seeds are dispersed autonomously for short distances, and occasionally they can be washed downslope from streamside habitats.

Plant material was collected in 2015–2016 in the three mountain ranges in central Europe: the Jeseníky, the Krkonoše, and the Tatry. In the Jeseníky and the Krkonoše, we sampled all known populations in subalpine glacial cirques and representative set from distinct valleys in the foothills. In the Tatry, we included populations from two distinct subalpine parts: the southern mountain range of the Nízké Tatry, the northern-eastern mountain range of the Vysoké/Belanské Tatry, and representative foothill populations in the basins between these two mountain ranges (called overall the Tatry). Our dataset comprises in total 202 individuals from 16 populations (Table 1 and Supplementary Table S1). In each mountain range, we sampled multiple foothill (below treeline) and subalpine (above treeline) populations. Individuals in the populations were sampled randomly, but with a minimum distance of approximately 1 m between individuals to avoid collecting clones. Leaf tissue from five to twenty individuals per population was immediately dried in silica gel for subsequent DNA extraction and microsatellite genotyping.

Microsatellite genotyping. Genomic DNA was extracted from dry tissue using a modified NaCl/CTAB protocol[56]. We employed 12 microsatellite loci developed by Van Geert *et al.*[57], Bickler *et al.*[58], and Seino *et al.*[59] (Supplementary Table S2). From those, seven loci, originally developed for *P. veris*[58], were successfully cross-amplified in *P. elatior*: PV 23741, PV 21795, PV 27775, PV 4767, PV 23424, PV 19773, and PV 279. The others (Paca 11, Paca 38, Paca 78, Priv 4, and Priv 7) have been already cross-amplified in *P. elatior* by Seino *et al.*[59] and Van Geert *et al.*[57]. Our cross-amplification of the 12 loci was successful, meaning that we consistently amplified variable loci across the sample set, contrary to three additional loci Paca 404[59], PV 8880[58], and PV 26720[58], which were not consistently amplified, and therefore not employed in the study. The fluorescently labelled primers (dyes: PET, NED, 6-FAM, and VIC; Applied Biosystems) were designed into two multiplexes based on the results from a complementary threshold analysis in Multiplex manager 1.2. Microsatellite loci were amplified using the QIAGEN Multiplex PCR Kit, with a total reaction volume of 5 µl of QIAGEN mix. The mix contained 0.25 µl of forward primer and 0.25 µl of reverse

primer, 1 μl of ddH_2O, 2.5 μl of Master Mix, and we added 1 μl (10 ng) of DNA. The PCR amplification was conducted in a thermocycler (Eppendorf Mastercycler Pro) under the following conditions for both multiplexes: 5 min of denaturation at 95 °C, followed by 35 cycles of 95 °C at 30 s, 57 °C for 90 s, 72 °C for 40 s, and a final extension of 68 °C for 30 min. Amplification products were separated using 3130xl Genetic Analyser (DNA laboratory of Faculty of Science, Charles University, Prague) with GeneScan 500 LIZ (Applied Biosystems) as an internal standard.

Genetic structure and diversity. Allele sizes were determined in GENEMARKER 2.6 (SoftGenetics). We checked possible presence of null alleles, stuttering, large allele dropout by the program MICRO-CHECKER 2.2.3[60]. None of the loci showed presence of null alleles in more than half of the populations (maximum seven for Priv 4 and six for Paca 78), and we thus retained all 12 microsatellite loci in analyses. We tested if loci significantly deviated from HWE in each population in R, package pegas[61]. None of the populations had significant deviation from HWE in more than half of the loci (maximum seven in S5) (Table 1).

First, we explored population structure in the entire dataset as well as separately for each region using Bayesian clustering in STRUCTURE 2.3.3[62] employing Abel HPC cluster of the University of Oslo. We used independent allele frequencies model with admixture, which allows for mixed ancestry of individuals. The number of clusters was set from K = 1 to K = 10 for entire dataset and from K = 1 to K = max, in which "max" equalled number of sampled populations in particular regions (K = 6 for each of the Krkonoše and the Tatry datasets, K = 4 for the Jeseníky dataset). Analysis for each K was performed with 20 replicates, the initial length of burn-in period 100,000 and 1,000,000 of Markov chain Monte Carlo (MCMC) replicates after burn-in. Similarity coefficients among runs of the same K[63] were calculated using Structure-sum-2009 script[31] in R 3.3.2[64]. For an optimal number of clusters (K), we considered the partition, where the rising likelihood of K values started to flatten and which also exhibited high similarity among replicated runs for that particular K (Supplementary Fig. S1). Some analyses allowed several possibilities for the optimal number of K, due to the hierarchical genetic structure of populations[65], in that case, we presented several partitions. Outputs from STRUCTURE analyses were graphically visualized in Structure Plot $V_{2.0}$[66].

Further, we visualized relationships among individuals and populations using distance-based approaches. Firstly, genetic relationships among individuals were plotted in centred principal component analysis (PCA) calculated in R, package adegenet[67]. Secondly, networks of pairwise F_{ST} distance among populations were created based on neighbor-net algorithm in SplitsTree 4.13.1[68]. Nei's pairwise F_{ST}s[69], in which heterozygosities are weighted by group sizes, and therefore comparison between populations with different sizes of individuals is possible, were calculated in R, packages adegenet[67] and hierfstat[70].

In order to test differences in population genetic properties, we calculated descriptive statistics of all populations with respect to small number of individuals and imbalance sampling. Due to varying number of individuals samples per populations (from five to 20 individuals) we employed following subsampling strategy to rule out the effect of varying sample size: five individuals per population were randomly selected, we calculated the corresponding statistics and repeated the process 100 times and we further presented mean value with standard deviation from 100 replicates (https://github.com/MarekLipan/Population-subsampling/blob/master/Genind_subsampling_func.R). We calculated observed heterozygosity (H_O) and expected heterozygosity (H_E) by subsampling in R, packages adegenet[67] and hierfstat[70]. Further, we calculated the numbers of alleles, allelic richness (with reference population size of five individuals per population, 1,000 permutations), and Nei's unbiased estimator for gene diversity (H_S)[71], which is corrected for small sample size, in Microsatellite analyser (MSA) 4.05[72]. Furthermore, we quantified the proportion of rare alleles using the frequency of down-weighted marker index (DW index), calculated as a ratio of means from the presence-absence matrix of alleles, which makes the measure less sensitive to different number of individuals per population[31]. DW index was calculated using R-script AFLPdat[31]. Finally, a number of private alleles was counted in R, package PopGenKit[73]. Differences in allelic richness, H_E, H_S, and in DW index between foothill and subalpine group were tested by hierarchical ANOVA with region as a random effect factor in R, package stats. Further, hierarchical structuring of genetic variation among populations within foothill vs. subalpine group in each region was revealed by analysis of molecular variance (AMOVA) in Arlequin 3.1[74]. AMOVAs were calculated by the method of the number of different alleles (F_{ST}-like) with 1,000 permutations. Finally, the structuring of variation among foothill and subalpine populations in each region was also explored by hierarchical AMOVAs.

Estimation of gene flow direction and migration rates. To identify the direction and intensity of migration among the foothill and subalpine populations, we searched for optimal models of migrations between populations within each target mountain range in a coalescent framework using MIGRATE-N version 3.6.11[75]. We have chosen two pairs of foothill and subalpine populations from the Jeseníky, the Krkonoše, and one pair from sub-range of the Tatry (Nízké Tatry), focusing on populations with the maximum number of genotyped individuals ($n \geq 10$). To keep the simulation scenarios feasible, we worked with two-population models that were iterated among all possible population combinations within each mountain range: we analysed pairs of subalpine-subalpine, foothill-subalpine, and foothill-foothill populations. For each pair, we modelled past evolutionary history under Bayesian inference of five migration models differing in presence and directionality of migration. We allowed bidirectional migration between two populations (model 1), unidirectional migration (models 2 and 3), panmixia (model 4 assuming that two populations belong to one panmictic population), and zero migration between two separate populations (model 5). In case of foothill-subalpine comparisons, the model 2 allowed migration from foothill to subalpine while the model 3 assumed only migration from subalpine to foothill populations.

We used microsatellite data type with Brownian motion microsatellite model. The mutation rate was set constant over all loci. MIGRATE-N estimated two parameters Theta - θ (mutation scaled population size) and M (mutation scaled immigration rate). The starting values of θ and M were calculated from Wright's F_{ST} using prior values (for θ: minimum = 0.004, delta ranged from 0.9 to 2.0, maximum ranged from 9.0 to 20, bins = 500; for M: minimum = 0, delta ranged from 80 to 100, maximum ranged from 800 to 1000, bins = 500). Prior distributions

of population sizes and migration rates were set based on the personal knowledge of the populations from the field with a broader range of both parameters with the aim to achieve the best searching of the space. To reach the stable states we ran each model for each pair multiple times (at least five). We have also checked the effective sample size, which was well over 500 (approximately 3000–5000). After burn-in of 20,000, we sampled 500,000 states from a single Markov chain, one every 5,000 steps. Four chains were run in parallel with heating static scheme (temperatures: 1.0, 1.5, 3.0, 10,000). According to Hodel *et al.*[32], based on θ and M parameters we calculated a number of immigrants per generation following the formula: $Nm = [(\theta_x * M_{y-x})/4]$ (for nuclear loci). For calculating Nm values, we used median values for both θ and M parameters.

To select the most likely model among the five models for each population pair, we used Bayes factors comparison. Bayes factor allows comparing nested and non-nested models, without assuming normality, or large samples[76,77]. We calculated natural log Bayes factors following the formula: $LBF = \ln [mL(\text{model}_1)] - \ln [mL(\text{model}_2)]$, in which model_1 is a model with the highest marginal likelihood and model_2 is each of the other models. We used "Bezier" approximated marginal likelihood calculated using the thermodynamic integration with the heating scheme described above). Marginal likelihood is the integral of the likelihood function over the complete parameter range. Afterwards, we calculated the probability of each model following the formula:

$$Prob_{model_i} = \frac{mLmodel_i}{\sum_j^n mLmodel_j},$$

in which $mLmodel_i$ is the marginal likelihood of $model_i$ and $\sum_j^n mLmodel_j$ is the sum of the marginal likelihoods of all other models.

Data Availability

All data generated or analysed during this study are included in this article (and its Supplementary Information Files).

References

1. Körner, C. *Alpine plant life: functional plant ecology of high mountain ecosystems*. (Springer Science & Business Media, 2003).
2. Flatscher, R., Frajman, B., Schönswetter, P. & Paun, O. Environmental heterogeneity and phenotypic divergence: Can heritable epigenetic variation aid speciation? *Genet. Res. Int.* 1–9, https://doi.org/10.1155/2012/698421 (2012).
3. Pfennig, D. W. *et al.* Phenotypic plasticity's impacts on diversification and speciation. *Trends Ecol. Evol.* **25**, 459–467 (2010).
4. Bertel, C., Hülber, K., Frajman, B. & Schönswetter, P. No evidence of intrinsic reproductive isolation between two reciprocally non-monophyletic, ecologically differentiated mountain plants at an early stage of speciation. *Evol. Ecol.* **30**, 1031–1042 (2016).
5. Clausen, J., Keck, D. D. & Hiesey, W. M. The Concept of Species Based on Experiment. *Am. J. Bot.* **26**, 103–106 (1939).
6. Lowry, D. B. Ecotypes and the controversy over stages in the formation of new species. *Biol. J. Linn. Soc.* **106**, 241–257 (2012).
7. Clausen, J., Keck, D. D. & Hiesey, W. M. Regional Differentiation in Plant Species. *Am. Nat.* **75**, 231–250 (1941).
8. Hiesey, W. M., Clausen, J. & Keck, D. D. Relations between Climate and Intraspecific Variation in Plants. *Am. Nat.* **76**, 5–22 (1942).
9. Funk, W. C., Blouin, M. S., Corn, P. S., Maxell, B. A. & Pilliod, D. S. Population structure of Columbia spotted frogs (*Rana lueteiventris*) is strongly affected by the landscape. *Mol. Ecol.* **14**, 1–14 (2005).
10. Ohsawa, T. & Ide, Y. Global patterns of genetic variation in plant species along vertical and horizontal gradients on mountains. *Glob. Ecol. Biogeogr.* **17**, 152–163 (2008).
11. Levin, D. The recurrent origin of plant races and species. *Syst. Bot.* **26**, 197–204 (2001).
12. Šrámková-Fuxová, G. *et al.* Range-wide genetic structure of *Arabidopsis halleri* (Brassicaceae): glacial persistence in multiple refugia and origin of the Northern Hemisphere disjunction. *Bot. J. Linn. Soc.* **185**, 321–342 (2017).
13. Fustier, M., Brandenburg, J., Boitard, S., Lapeyronnie, J. & Eguiarte, L. E. Signatures of local adaptation in lowland and highland teosintes from whole-genome sequencing of pooled samples. *Mol. Ecol.* **26**, 2738–2756 (2017).
14. Holliday, J. A., Zhou, L., Bawa, R., Zhang, M. & Oubida, R. W. Evidence for extensive parallelism but divergent genomic architecture of adaptation along altitudinal and latitudinal gradients in *Populus trichocarpa*. *New Phytol.* **209**, 1240–1251 (2016).
15. Bettin, O., Cornejo, C., Edwards, P. J. & Holderegger, R. Phylogeography of the high alpine plant *Senecio halleri* (Asteraceae) in the European Alps: *In situ* glacial survival with postglacial stepwise dispersal into peripheral areas. *Mol. Ecol.* **16**, 2517–2524 (2007).
16. Jacquemyn, H., Honnay, O., Galsbusera, P. & Roldán-Ruiz, I. Genetic structure of the forest herb *Primula elatior* in a changing landscape. *Mol. Ecol.* **13**, 211–219 (2004).
17. Odat, N., Jetschke, G. & Hellwig, F. H. Genetic diversity of *Ranunculus acris* L. (Ranunculaceae) populations in relation to species diversity and habitat type in grassland communities. *Mol. Ecol.* **13**, 1251–1257 (2004).
18. Ortego, J., Riordan, E. C., Gugger, P. F. & Sork, V. L. Influence of environmental heterogeneity on genetic diversity and structure in an endemic southern Californian oak. *Mol. Ecol.* **21**, 3210–3223 (2012).
19. Young, A., Boyle, T. & Brown, T. The population genetic consequences of habitat fragmentation for plants. *Trends Ecol. Evol.* **11**, 413–418 (1996).
20. Ellstrand, N. C. & Elam, D. R. Population genetic consequences of small population size: Implications for plant conservation. *Annu. Rev. Ecol. Syst.* **24**, 217–242 (1993).
21. Günther, T., Lampei, C., Barilar, I. & Schmid, K. J. Genomic and phenotypic differentiation of *Arabidopsis thaliana* along altitudinal gradients in the North Italian Alps. *Mol. Ecol.* **25**, 3574–3592 (2016).
22. Luo, Y., Widmer, A. & Karrenberg, S. The roles of genetic drift and natural selection in quantitative trait divergence along an altitudinal gradient in *Arabidopsis thaliana*. *Heredity (Edinb).* **114**, 220–228 (2015).
23. Shao, J. *et al.* Genetic diversity and gene flow within and between two different habitats of *Primula merrilliana* (Primulaceae), an endangered distylous forest herb in eastern China. *Bot. J. Linn. Soc.* **179**, 172–189 (2015).
24. Valentine, D. H. & Kress, A. In *Flora Europaea* 3 (ed. Tutin, T.) (Cambridge University Press, 1972).
25. Richards, J. *Primula*. (Timber Press, 2003).
26. Konečná, V. Phenotypic variability and evolutionary relationships among populations of *Primula elatior* along an altitudinal gradient (Master's thesis). (Charles University, Prague, Czech Republic, 2017).
27. Jacquemyn, H., Brys, R. & Hermy, M. Patch occupancy, population size and reproductive success of a forest herb (*Primula elatior*) in a fragmented landscape. *Oecologia* **130**, 617–625 (2002).
28. Jacquemyn, H., Vandepitte, K., Roldán-Ruiz, I. & Honnay, O. Rapid loss of genetic variation in a founding population of *Primula elatior* (Primulaceae) after colonization. *Ann. Bot.* **103**, 777–783 (2008).

29. Van Rossum, F. & Triest, L. Fine-scale genetic structure of the common *Primula elatior* (Primulaceae) at an early stage of population fragmentation. *Am. J. Bot.* **93**, 1281–1288 (2006).
30. Krahulec, F. Species of vascular plants endemic to the Krkonoše Mts (Western Sudetes). *Preslia* **78**, 503–516 (2006).
31. Ehrich, D. AFLPDAT: A collection of R functions for convenient handling of AFLP data. *Mol. Ecol. Notes* **6**, 603–604 (2006).
32. Hodel, R. G. J., D S Cortez, M. B., Soltis, P. S. & Soltis, D. E. Comparative phylogeography of black mangroves (*Avicennia germinans*) and red mangroves (*Rhizophora mangle*) in Florida: Testing the maritime discontinuity in coastal plants. *Am. J. Bot.* **103**, 730–739 (2016).
33. Beerli, P. & Felsenstein, J. Maximum-Likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics* **152**, 763–773 (1999).
34. Jones, F. C. *et al.* The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484**, 55–61 (2012).
35. Hufford, M. B. *et al.* The Genomic Signature of Crop-Wild Introgression in Maize. *PLoS Genet.* **9**, (2013).
36. Trucchi, E., Frajman, B., Haverkamp, T. H. A., Schönswetter, P. & Paun, O. Genomic analyses suggest parallel ecological divergence in *Heliosperma pusillum* (Caryophyllaceae). *New Phytol.* **216**, 267–278 (2017).
37. Gabel, A. & Sattler, J. & Reisch Christoph. Genetic variation and performance of the alpine plant species *Dianthus callizonus* differ in two elevational zones of the Carpathians. *Alp. Bot.* **127**, 65–74 (2017).
38. Sakaguchi, S. *et al.* Phylogeographic analysis of the East Asian goldenrod (*Solidago virgaurea* complex, Asteraceae) reveals hidden ecological diversification with recurrent formation of ecotypes. *Ann. Bot.* **121**, 489–500 (2018).
39. Andrew, R. L., Ostevik, K. L., Ebert, D. P. & Rieseberg, L. H. Adaptation with gene flow across the landscape in a dune sunflower. *Mol. Ecol.* **21**, 2078–2091 (2012).
40. Roda, F. *et al.* Genomic evidence for the parallel evolution of coastal forms in the *Senecio lautus* complex. **22**, 2941–2952 (2013).
41. Sakaguchi, S. *et al.* Simultaneous evaluation of the effects of geographic, environmental and temporal isolation in ecotypic populations of *Solidago virgaurea*. *New Phytol.* **216**, 1268–1280 (2017).
42. Butlin, R. K. *et al.* Parallel evolution of local adaptation and reproductive isolation in the face of gene flow. *Evolution (N. Y).* **68**, 935–949 (2014).
43. Kochjarová, J. In *Flóra Slovenska VI/4* (VEDA, 2017).
44. Bureš, L. *Chráněné a ohrožené rostliny CHKO Jeseníky*. (Agentura Rubico, s.r.o., 2013).
45. Kolář, F., Kaplan, Z., Suda, J. & Štech, M. Populations of *Knautia* in ecologically distinct refugia on the Hercynian massif belong to two endemic species. *Preslia* **87**, 363–386 (2015).
46. Kolář, F. *et al.* Northern glacial refugia and altitudinal niche divergence shape genome-wide differentiation in the emerging plant model *Arabidopsis arenosa*. *Mol. Ecol.* **25**, 3929–3949 (2016).
47. Deschepper, P., Brys, R., Fortuna, M. A. & Jacquemyn, H. Analysis of spatial genetic variation reveals genetic divergence among populations of *Primula veris* associated to contrasting habitats. *Sci. Rep.* 1–12, https://doi.org/10.1038/s41598-017-09154-9 (2017).
48. Taylor, K. & Woodell, S. R. J. Biological Flora of the British Isles: *Primula elatior* (L.) Hill. *J. Ecol.* **96**, 1098–1116 (2008).
49. Ekrtová, E., Štech, M. & Fér, T. Pattern of genetic differentiation in *Gentiana pannonica* Scop.: did subalpine plants survive glacial events at low altitudes in Central Europe? *Plant Syst. Evol.* **298**, 1383–1397 (2012).
50. England, P. R. *et al.* Effects of intense versus diffuse population bottlenecks on microsatellite genetic diversity and evolutionary potential. *Conserv. Genet.* **4**, 595–604 (2003).
51. Shao, J., Zhang, X., Zhang, Z. & Zhu, G. Identification of effective pollinators of *Primula merrilliana* and effects of flower density and population size on pollination efficiency. *J. Syst. Evol.* **46**, 537–544 (2008).
52. Lowry, D. B., Rockwood, R. C. & Willis, J. H. Ecological reproductive isolation of coast an inland races. *Evolution (N. Y).* **62**, 2196–2214 (2008).
53. Juřičková, L., Horáčková, J. & Ložek, V. Direct evidence of central European forest refugia during the last glacial period based on mollusc fossils. *Quat. Res.* **82**, 222–228 (2014).
54. Schou, O. The distyly in *Primula elatior* (L.) Hill (Primulaceae), with a study of flowering phenology and pollen flow. *Bot. J. Linn. Soc.* **86**, 261–274 (1983).
55. Proctor, M., Yeo, P. & Lack, A. *The natural history of pollination*. (HarperCollins Publishers, 1996).
56. Štorchová, H. *et al.* An improved method of DNA isolation from plants collected in the field and conserved in saturated NaCl/CTAB solution. *Taxon* **49**, 79–84 (2000).
57. Van Geert, A. *et al.* Isolation and characterization of microsatellite loci in primrose (*Primula vulgaris*). *Belgian J. Bot.* **139**, 261–264 (2006).
58. Bickler, C., A'Hara, S., Cottrell, J., Rogers, L. & Bridle, J. Characterisation of thirteen polymorphic microsatellite markers for cowslip (*Primula veris* L.) developed using a 454 sequencing approach. *Conserv. Genet. Resour.* **5**, 1185–1187 (2013).
59. Seino, M. M., de Vega, C., Bazaga, P., Jacquemyn, H. & Herrera, C. M. Development and characterization of microsatellite loci for the primrose *Primula vulgaris* and successful cross-amplification in the congeneric *P. elatior* and *P. veris*. *Conserv. Genet. Resour.* 653–655, https://doi.org/10.1007/s12686-014-0171-2 (2014).
60. Van Oosterhout, C., Hutchinson, W. F., Wills, D. P. M. & Shipley, P. Micro-checker: software for identifying and correcting genotyping errors in microsatellite data. *Mol. Ecol. Notes* **4**, 535–538 (2004).
61. Paradis, E. pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics* **26**, 419–420 (2010).
62. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of Population Structure Using Multilocus Genotype Data. *Genetics* **155**, 945–959 (2000).
63. Nordborg, M. *et al.* The pattern of polymorphism in *Arabidopsis thaliana*. *Plos Biol.* **3**, 1289–1299 (2005).
64. R Core Team. R: a language and environment for statistical computing, Vienna, Austria, https://www.r-project.org/ (2016).
65. Ehrich, D. *et al.* Genetic consequences of Pleistocene range shifts: contrast between the Arctic, the Alps and the East African mountains. *Mol. Ecol.* **16**, 2542–2559 (2007).
66. Ramasamy, R. K., Ramasamy, S., Bindroo, B. B. & Naik, V. G. Structure plot: a program for drawing elegant STRUCTURE bar plots in user friendly interface. *Springerplus* **3**, 431 (2014).
67. Jombart, T. Adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403–1405 (2008).
68. Huson, D. H. & Bryant, D. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **23**, 254–267 (2006).
69. Nei, M. Analysis of gene diversity in subdivided populations. *Proc. Natl. Acad. Sci.* **70**, 3321–3323 (1973).
70. Goudet, J. Hierfstat, a package for R to compute and test hierarchical F-statistics. *Mol. Ecol. Notes* **2**, 184–186 (2005).
71. Nei, M. *Molecular evolutionary genetics*. (Columbia university press, 1987).
72. Dieringer, D. & Schlötterer, C. Microsatellite analyser (MSA): a platform independent analysis tool for large microsatellite data sets. *Mol. Ecol. Notes* **3**, 167–169 (2003).
73. Rioux Paquette, S. PopGenKit: useful functions for (batch) file conversion and data resampling in microsatellite datasets. R package version 1.0. (2012).
74. Excoffier, L., Laval, G. & Schneider, S. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol. Bioinformatic* **1**, 47–50 (2005).
75. Beerli, P. In *Population Genetics for Animal Conservation* (eds Bertorelle, G., Bruford, M. W., Hauffe, H. C., Rizzoli, A. & Vernesi, C.) (Cambridge University Press, 2009).
76. Beerli, P. & Palczewski, M. Unified framework to evaluate panmixia and migration direction among multiple sampling locations. *Genetics* **185**, 313–326 (2010).
77. Kass, R. E. & Raftery, A. E. Bayes factors. *J. Am. Stat. Assoc.* **90**, 773–795 (1995).

Acknowledgements

The computations were performed using resources provided by UNINETT Sigma2 - the National Infrastructure for High Performance Computing and Data Storage in Norway (http://www.hpc.uio.no/). Access to computing and storage facilities provided under the programme "Projects of Large Research, Development, and Innovations Infrastructures" (CESNET LM2015042), was greatly appreciated. We are grateful to Kristýna Šemberová, who helped us with the maps. This study was supported by Grant No. 925916 from the Charles University Grant Agency. Additional support was provided by Charles University, Primus-SCI-35.

Author Contributions

V.K., M.D.N. and F.K. designed research; V.K. performed research; V.K. and F.K. analysed data; V.K. and F.K. wrote the paper and all authors revised the paper.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-019-39669-2.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Case study 4

Parallel adaptation in autopolyploid *Arabidopsis arenosa* is dominated by repeated recruitment of shared alleles

nature COMMUNICATIONS

ARTICLE

https://doi.org/10.1038/s41467-021-25256-5 OPEN

Parallel adaptation in autopolyploid *Arabidopsis arenosa* is dominated by repeated recruitment of shared alleles

Veronika Konečná[1,2], Sian Bray[3], Jakub Vlček[1,4,5], Magdalena Bohutínská[1,2], Doubravka Požárová[1], Rimjhim Roy Choudhury[6,7], Anita Bollmann-Giolai[8], Paulina Flis[3], David E. Salt[3], Christian Parisod[6], Levi Yant[9,10]✉ & Filip Kolář[1,2,10]✉

Relative contributions of pre-existing vs de novo genomic variation to adaptation are poorly understood, especially in polyploid organisms. We assess this in high resolution using autotetraploid *Arabidopsis arenosa*, which repeatedly adapted to toxic serpentine soils that exhibit skewed elemental profiles. Leveraging a fivefold replicated serpentine invasion, we assess selection on SNPs and structural variants (TEs) in 78 resequenced individuals and discover significant parallelism in candidate genes involved in ion homeostasis. We further model parallel selection and infer repeated sweeps on a shared pool of variants in nearly all these loci, supporting theoretical expectations. A single striking exception is represented by *TWO PORE CHANNEL 1*, which exhibits convergent evolution from independent de novo mutations at an identical, otherwise conserved site at the calcium channel selectivity gate. Taken together, this suggests that polyploid populations can rapidly adapt to environmental extremes, calling on both pre-existing variation and novel polymorphisms.

[1] Department of Botany, Faculty of Science, Charles University, Prague, Czech Republic. [2] The Czech Academy of Sciences, Institute of Botany, Průhonice, Czech Republic. [3] Future Food Beacon and School of Biosciences, University of Nottingham, Nottingham, UK. [4] Biology Centre, Czech Academy of Sciences, České Budějovice, Czech Republic. [5] Department of Zoology, Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic. [6] Institute of Plant Sciences, University of Berne, Bern, Switzerland. [7] Department of Systematic and Evolutionary Botany, University of Zurich, Zurich, Switzerland. [8] John Innes Centre (JIC), Norwich Research Park, Norwich, UK. [9] Future Food Beacon and School of Life Sciences, University of Nottingham, Nottingham, UK. [10] These authors jointly supervised this work: Levi Yant, Filip Kolář. ✉email: levi.yant@gmail.com; filip.kolar@natur.cuni.cz

Rapid adaptation to novel environments is thought to be enhanced by the availability of genetic variation; however, the relative contribution of standing variation versus the role of novel mutation is a matter of debate[1,2], especially in higher ploidy organisms. Whole-genome duplication (WGD; leading to polyploidisation) is a major force underlying diversification across eukaryotic kingdoms, seen most clearly in plants[3–5] with various effects on genetic variation[6–8]. While WGD is clearly associated with environmental change or stress[5,9], the precise impact of WGD on adaptability is largely unknown in multicellular organisms, and there is virtually no work assessing the evolutionary sources of adaptive genetic variation in young polyploids. Work in autopolyploids, which clearly isolate effects of WGD from hybridisation (which is confounded in allopolyploids), indicates that subtle genomic changes may follow WGD alone[8,10,11], which raises the question of when their adaptive value may originate.

Autopolyploidy is expected to alter selective and adaptative process in many ways, but a dearth of empirical data prevents synthetic evaluation. Besides immediate phenotypic[12–14] and genomic[10,11] changes following WGD, theory is unsettled regarding how adaptation proceeds as the autopolyploid lineage diversifies and adapts to novel challenges. On the one hand, autopolyploids can mask deleterious alleles and accumulate cryptic allelic diversity[7]. In addition, the number of mutational targets is multiplied in autopolyploids, meaning that new alleles are introduced more quickly[6,15,16]. This could promote adaptation[6,17]. On the other hand, reduced rates of allele frequency changes may retard adaptation[6,18], particularly for de novo mutations, which emerge in a population at initially low frequencies[19]. Recent advances in theory and simulations suggest potential solutions to this controversy. Polyploidy may promote adaptation under scenarios of rapid environmental change (e.g. colonisation of challenging habitats) when selection is strong and originally neutral or mildly deleterious alleles standing in polyploid populations may become beneficial[5,20]. However, empirical evidence supporting this scenario is fragmentary. There is broad correlative evidence that polyploids are good colonisers of areas experiencing environmental flux (e.g. the Arctic[21,22], stressful habitats[8,23,24], and heterogeneous environments[25]). However, the genomic basis of such polyploid adaptability—and whether their primary source of adaptive alleles is high diversity (standing variation) or large mutational target size (de novo mutations)—remains unknown.

We focus on natural autotetraploid *Arabidopsis arenosa* populations repeatedly facing one of the greatest environmental challenges for plant life—naturally toxic serpentine soils. Serpentines occur as islands in the landscape with no intermediate habitats and are defined by peculiar elemental contents (highly skewed Ca/Mg ratio and elevated heavy metals such as Cr, Co, and Ni), that may be further combined with low nutrient availability and propensity for drought[26]. *Arabidopsis arenosa* is a well-characterised, natural diploid-autotetraploid species with large and genetically diverse outcrossing populations[27]. The widespread autotetraploids, which originated from a single diploid lineage ~19–31k generations ago[8], harbour increased adaptive diversity genome wide[8] and currently occupy a broader ecological niche than their diploid sisters[28], including serpentine outcrops[29], railway lines[30,31], and contaminated mine tailings[32,33]. This makes *A. arenosa* a promising model for empirical inquiries of adaptation in autopolyploids[8,27]. As a proof of concept, selective ion uptake phenotypes and a polygenic basis for serpentine adaptation have been suggested from a single *A. arenosa* serpentine population[29]. However, limited sampling left unknown whether the same genes are generally (re)used and what is the evolutionary source of the selected alleles, i.e. leaving unresolved the evolutionary dynamics and mechanism underlying these striking adaptations. We ask specifically: (1) Does gene-level parallelism in autotetraploid *A. arenosa* dominantly reflect repeated sampling from the large pool of shared variation that is expected to be maintained in autopolyploids? and (2) Is repeated adaptation from novel mutations feasible in autotetraploid populations?

In this work, we deconstruct the sources of parallel adaptive variation in *A. arenosa*. First, we sample five serpentine/non-serpentine population pairs of autotetraploid *A. arenosa* and demonstrate rapid parallel adaptation by combining demographic analysis and reciprocal transplant experiments. Taking advantage of the power of this fivefold replicated natural selection experiment, we identify candidate adaptive loci from population resequencing data and find significant parallelism underlying serpentine adaptation. We then model parallel selection using a designated framework and statistically infer the evolutionary sources of parallel adaptive variation for all candidate loci. In line with theory, we find that shared variation is the vastly prevalent source of parallel adaptive variants in serpentine *A. arenosa*. However, we also discover an exceptional locus exhibiting footprints of selection on alleles originating from two distinct de novo mutations. In line with the latter hypothesis, this demonstrates that the rapid selection of novel alleles is still feasible in autopolyploids, indicating broad evolutionary flexibility of lineages with doubled genomes.

Results

Parallel serpentine adaptation. First, we inferred independent colonisation of each serpentine site by different local *A. arenosa* populations. To do this, we resequenced five pairs of geographically proximate serpentine (S) and non-serpentine (N) populations covering all known serpentine sites occupied by the species to date (8 individuals per population on average, mean sequencing depth 21×; Fig. 1a, Supplementary Fig. 1, Supplementary Data 1, and Supplementary Tables 1–3). Phylogenetic, ordination, and Bayesian analyses based on nearly neutral fourfold-degenerate (4dg) sites demonstrated overall grouping of populations by spatial proximity, not by substrate. In all but one case, the adjacent S and N populations occupied sister position in the population tree and belonged to the same Bayesian cluster; only the population S3 occupied somewhat isolated position yet still within the lineage of Eastern Alpine populations (Fig. 1b and Supplementary Fig. 1). We thus further tested the independent colonisation of each serpentine site by coalescent simulations. Consistently over all possible pairwise iterations of S–N population pairs ($n = 10$), the scenario of independent colonisation of each serpentine site was more likely than any scenario assuming sister position of two S populations (Fig. 1c). Note that subsequent gene flow between substrate types within each S–N population pair was unlikely as the assumption of migration within each population pair had not significantly improved the model fit (Supplementary Fig. 2 and Supplementary Data 2). Reflecting the independent origin of the five S populations, we analysed each serpentine colonisation event separately in the following analyses to take into account neutral population structure in the data, using the spatially closest N population as a contrast where needed. The very low differentiation between S and proximate N populations and consistently low population split times (Table 1) indicate very recent, postglacial serpentine invasions. There is no evidence of bottleneck associated with colonisation, as S and N populations exhibited similar nucleotide diversity and Tajima's *D* values (Table 1).

To assess whether the colonisation of serpentines was accompanied by substrate adaptation, we combined ionomics

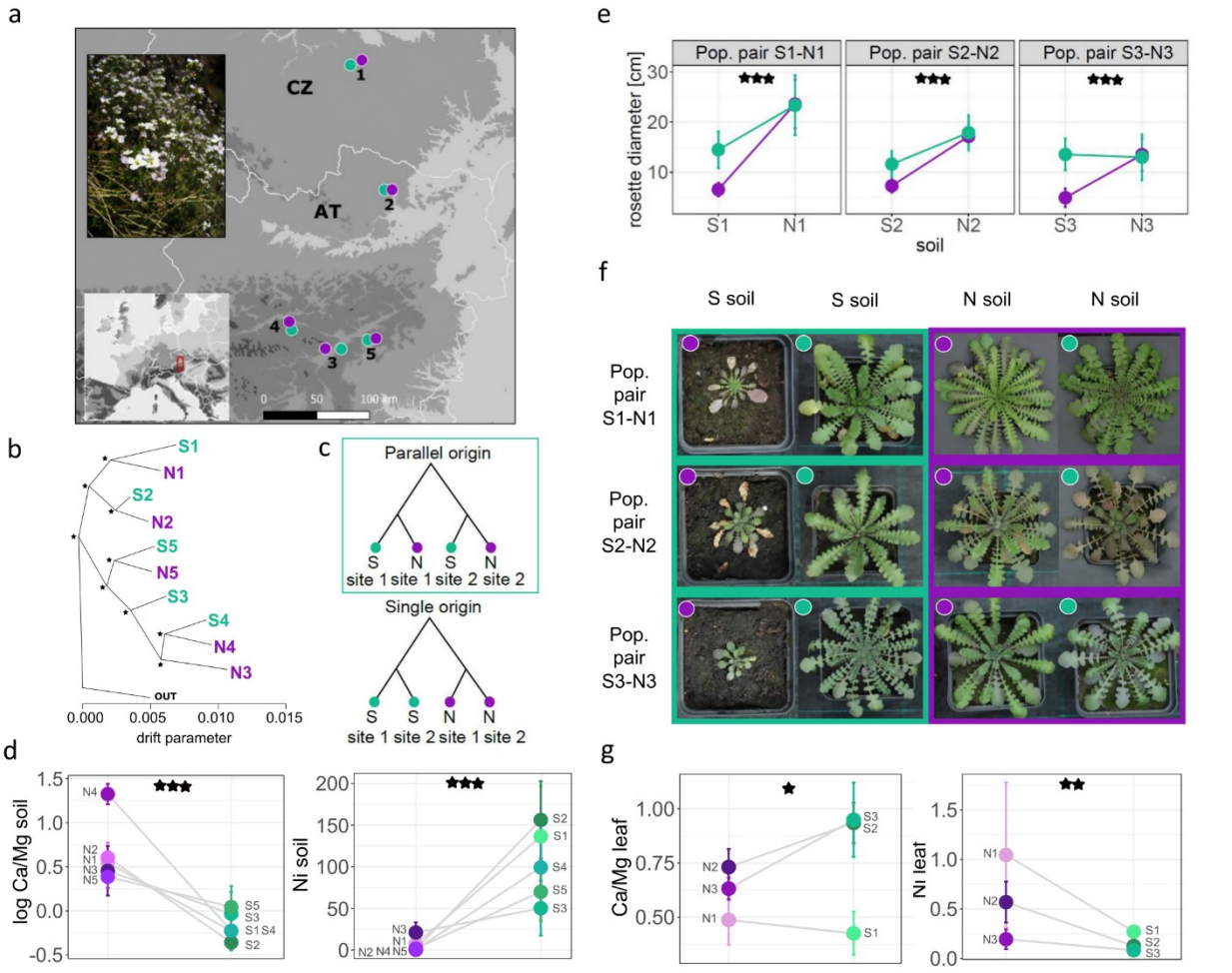


Fig. 1 Parallel adaptation of *Arabidopsis arenosa* to challenging serpentine soils. a Locations of the investigated serpentine (S, green) and non-serpentine (N, violet) populations sampled as spatially proximate pairs (numbers) in Central Europe with an illustrative photo of an S population (photo was taken by F. Kolář). **b** Allele frequency covariance graph of populations based on ~870,000 fourfold-degenerate SNPs; asterisks show the 100 bootstrap branch support. The outgroup (OUT) is represented by a tetraploid population from Western Carpathians, the ancestral area of tetraploid *A. arenosa*[93]. **c** Two contrasting evolutionary scenarios of serpentine colonisation compared in coalescent simulations; the topology assuming independent serpentine colonisations (framed in green) received the highest support consistently across all 10 pairwise combinations of S–N population pairs. **d** Differences in Ca/Mg ratio and in Ni concentrations [µg/g] in S and N soils from the original sampling sites ($n = 78$ individual samples, one-way ANOVAs: $F_{1,77} = 26.5$, $p = 1.94e{-}06$ and $F_{1,77} = 117.4$, $p = 2.01e{-}16$ for Ca/Mg and Ni, respectively). **e** Differences in maximum rosette size of three population pairs attained after 3 months of cultivation in local serpentine and non-serpentine substrates (significance of the soil treatment × soil origin interaction in a two-way ANOVA is indicated: $F_{1,90} = 21.6$, $p = 1.17e{-}05$, $F_{1,96} = 12.3$, $p = 6.88e{-}04$, and $F_{1,85} = 42$, $p = 5.68e{-}09$ for population pairs 1, 2 and 3, respectively). **f** Example photos illustrating parallel growth response in the three population pairs to serpentine soils (green frame) depending on the soil of origin (dot colour) (photo was taken by V. Konečná). **g** Differences in ion uptake between originally S and N individuals when cultivated in serpentine soils; Ni concentrations were standardised by corresponding soil Ni values ($n = 28$ individual samples, one-way ANOVAs: $F_{1,27} = 6.2$, $p = 0.019$ and $F_{1,27} = 13.5$, $p = 0.001$ for Ca/Mg ratio and Ni, respectively). Points denote mean, error bars depict standard error of mean in charts **e**, **d**, **g**. Source data underlying Fig. 1d, g are provided as a Source data file.

with a reciprocal transplant experiment. First, using ionomic profiling of native soil associated with each sequenced individual, we characterised major chemical parameters differentiating on both substrates (Fig. 1d and Supplementary Figs. 3 and 4). Among the 20 elements investigated (Supplementary Fig. 3a), only the bioavailable concentration of Mg, Ni, Co, and Ca/Mg ratio consistently differentiated both soil types (Bonferroni-corrected one-way analysis of variance (ANOVA) taking population pair as a random variable). Serpentine sites were not macronutrient poor (Supplementary Table 4) and were not differentiated from non-serpentines by bioclimatic parameters (annual temperature, precipitation, and elevation; Supplementary Fig. 3), indicating that skewed Ca/Mg ratios and elevated heavy metal content are likely the primary selective agents on the sampled serpentine sites[34–36].

We then tested for differential fitness response towards serpentine soil between populations of S versus N origin using reciprocal transplant experiments. We cultivated plants from three population pairs (S1–N1, S2–N2, and S3–N3) on both native soil types within each pair for three months (until attaining maximum rosette size), observing significantly better germination and growth of the S plants in their native serpentine substrate as compared to their closest N relatives. First, we found a significant interaction between soil type and soil of origin at germination

Table 1 Between-population divergence and within-population diversity of the five investigated serpentine/non-serpentine population pairs inferred from genome-wide fourfold-degenerate single nucleotide polymorphisms.

Population pair	Divergence (generations)[a]	Pairwise F_{ST}	Nucleotide diversity[b]	Tajima's D[b]
S1–N1	(774) 4317 (6284)	0.069	0.0292/0.0276	0.266/0.485
S2–N2	(500) 2690 (3826)	0.029	0.0306/0.0285	0.131/0.372
S3–N3	(794) 3912 (6316)	0.085	0.0307/0.0287	−0.046/0.350
S4–N4	(812) 2918 (3542)	0.057	0.0304/0.0297	0.132/0.177
S5–N5	(546) 3539 (4869)	0.047	0.0292/0.0296	0.267/0.089

[a]Divergence between proximal S–N populations. Mean and 95% confidence intervals inferred by bootstrapping, estimated by coalescent simulations. Assuming 2-year generation time[93], all estimates indicate recent postglacial divergence.
[b]Genome-wide nucleotide diversity (π) and Tajima's D of each S–N population.

(generalised linear model (GLM) with binomial errors taking population pair as a random variable, $\chi^2 = 22.436$, $p < 0.001$), although the fitness disadvantage of N plants in serpentine soil varied across population pairs (Supplementary Fig. 5). During subsequent cultivation, we recorded zero mortality but found a significant interaction effect between soil treatment and soil of origin on growth, as approximated by maximum rosette sizes (two-way ANOVA taking population pair as a random variable, $F_{1,277} = 55.5$, $p < 0.001$, Fig. 1e; see Supplementary Fig. 6 for rosette size temporal development). Once again, the S plants consistently produced significantly larger rosettes (by 47% on average) than their N counterparts when grown in serpentine soil, indicating consistent substrate adaptation (Fig. 1e, f). Finally, we evaluated differences in Ni and Ca/Mg accumulation in leaves harvested on plants cultivated in serpentine soils. Consistent with adaptive responses to soil chemistry, we found higher Ca/Mg ratio and reduced uptake of Ni (lower leaf/soil ratio) in tissue of serpentine plants relative to their non-serpentine counterparts (Fig. 1g). Taken together, our demographic analysis complemented by transplant experiments support recent parallel serpentine adaptation of autotetraploid *A. arenosa* at five distinct sites, exhaustively covering all known serpentine populations of the species.

Parallel genomic footprints of selection on serpentine at single-nucleotide polymorphisms (SNPs) and transposable elements (TE)s. Using these five natural replicates of serpentine adaptation, we sought the genomic basis and evolutionary source of the parallel adaptations. To do this, we combined divergence scans and environmental association analysis to refine the list of loci for parallel selection modelling only to the candidates that repeatedly differentiated across multiple population pairs and were significantly associated with the selective soil environment. First, we identified initial inclusive lists of gene-coding loci exhibiting excessive differentiation between paired populations using 1% outlier F_{ST} window-based scans (490–525 candidate genes per pair; details in 'Methods'; Supplementary Data 3). These most inclusive lists must be interpreted with caution, as they are based on a simple assumption that the most differentiated regions are under directional selection[37]. However, in support of their relevance a gene ontology (GO) analysis of the 2245 candidates from all five population pairs shows significant enrichment (Fisher's exact test; $p < 0.05$) of 'biological processes', 'molecular functions', and 'cellular components' considered relevant to serpentine adaptation[26,34], such as inorganic anion transport, ion homoeostasis, post-embryonic development, and calcium transmembrane transporter activity (Fig. 2b and Supplementary Data 4).

To refine this broad list and pinpoint parallel evolution candidate genes, we overlapped these candidate gene lists across population pairs, identifying 207 'parallel differentiation candidates' that represent divergence outliers in at least two S–N population pairs. The level of parallelism was greater than expected by chance for all pairs of S–N contrasts (Fisher's exact test; Fig. 2a and Supplementary Data 4) and we hereafter refer to this as 'significant parallelism'. Such a fraction of parallel gene candidates (0.02–0.04 out of all candidates from that particular population pair) is in line with other naturally adapting systems of comparable divergence[33,38–40]. The parallel differentiation candidates were significantly enriched ($p < 0.05$) for GO terms, such as regulation of ion transmembrane transport, voltage-gated calcium and potassium channel activity or plasma membrane (Supplementary Data 4). The absence of common candidates across all five population pairs may reflect a complex genetic basis of the traits allowing for the modulation of the same pathway by different genes in some populations. This is supported by significant functional parallelism, i.e. higher than random number of overlapping GO terms that were repeatedly identified by separate enrichment analyses of outlier gene list from each population pair (Supplementary Fig. 7 and Supplementary Data 4). Additionally, adaptation via partial (soft) sweeps, which are likely to occur in autotetraploids[19], might have further limited the power of our divergence scans in some loci and populations.

As a complementary approach, we inferred candidates directly associated with the distinctive chemical characteristics of serpentine soil by performing environmental association analysis using latent factor mixed models (LFMM)[41]. This analysis quantitatively determines the association between each soil elemental concentration and SNPs across the genome in both S and N populations at the level of individual plants (in total 78). We identified 2,809 genes (LFMM candidates) harbouring ≥1 SNP significantly associated with at least one distinctive serpentine soil parameter previously identified by ionomic analysis (Ca/Mg ratio, high Mg, Ni, and Co; Supplementary Data 5). Finally, we overlapped the LFMM candidates with the parallel differentiation candidates to produce a final refined list of 61 'serpentine adaptation candidates' (Fig. 2c, Supplementary Fig. 8, and Supplementary Data 6 and 7). This conservative approach aims to identify the strongest candidates underlying serpentine adaptation for further model-based inference of the sources of variation in the next section. We note that this approach discards population-specific (private) candidates and cases of distinct genetic architecture of a trait (e.g. distinct genes affecting the same pathway) and thus cannot quantify the overall genome proportion that evolves in parallel. Importantly, however, it also minimises false positives from population-specific selection and genetic drift.

These 61 serpentine adaptation candidates were significantly enriched ($p < 0.05$) for categories related, for example, to regulation of ion transmembrane transport, and specifically, voltage-gated calcium and potassium channel activity (Supplementary Data 7). Candidates included the *NRT2.1* and *NRT2.2* high-affinity nitrate transporters, which act as repressors of lateral root initiation[42,43]; *RHF1A*, which is involved in gametogenesis and transferase activities[44,45]; *TPC1*, a central calcium channel

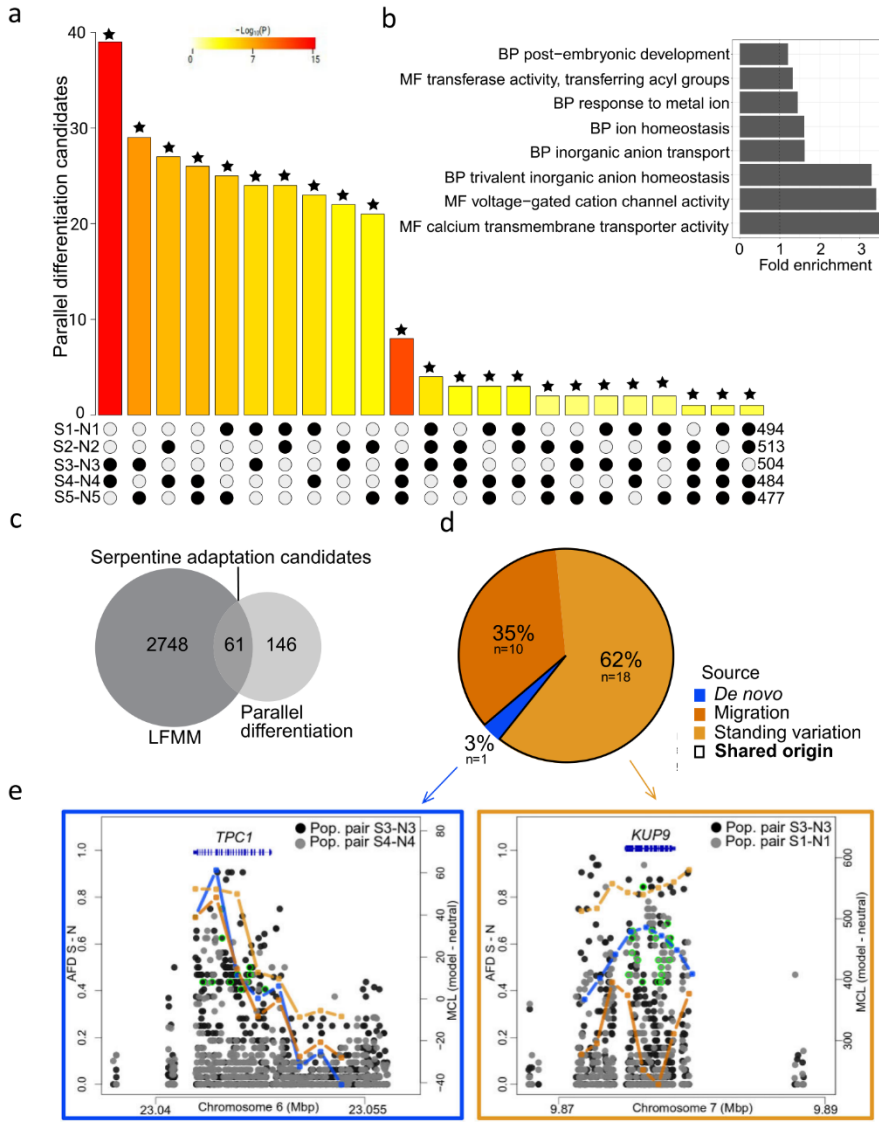


Fig. 2 Parallel serpentine adaptation candidates and the sources of parallel variants in *A. arenosa*. **a** Intersection of candidates from each population pair (S1–N1 to S5–N5) demonstrating more genes repeatedly found as candidates across two, three, and four population pairs than expected by chance alone (all intersections were significant at $p < 0.01$ (highlighted by asterisks), one-sided Fisher's exact test); note: the colour intensity of the bars represents the p value significance of the intersections. **b** Gene ontology (GO) enrichment of the candidates (across all population pairs); GO categories: biological process (BP) and molecular function (MF); for complete list of GO terms, see Supplementary Data 4. **c** Overlap between parallel differentiation candidates and latent factor mixed model (LFMM) candidates resulting in 61 serpentine adaptation candidate genes. **d** Proportions of serpentine adaptation candidates originating from de novo mutations or being of shared origin out of the total of 29 cases of non-neutral parallelism as inferred by the Distinguishing among Modes of Convergence approach (DMC; see text for details). **e** Two examples of parallel candidate loci, illustrating nucleotide divergence and maximum composite log-likelihood (MCL) estimation of the source of the selected alleles in these particular loci inferred in DMC. Allele frequency difference (AFD) for locus with independent de novo mutations (left) and with parallel recruitment of shared ancestral standing variation (right). Left y-axis: AFD between S and N populations. Dots: AFD values of individual SNPs; bright green circles: non-synonymous SNPs with AFD ≥ 0.4; lines (right y-axis): MCL difference between neutral versus parallel selection scenario following colour scheme in **d**; gene models are in blue.

that mediates plant-wide stress signalling and tolerance[46]; and potassium transporters *AKT5* and *KUP9*. Furthermore, when we compared our serpentine adaptation candidates ($n = 61$) to candidate loci for parallel serpentine adaptation in *Arabidopsis lyrata* ($n = 62$) from a previous study[47], we found two loci in common (significant overlap; $p < 0.007$), *KUP9* and *TPC1*, further supporting important roles of these two ion transporters in repeated adaptation to serpentine soil. In addition, when overlapping the candidate genes detected at least in one of our five population pairs with serpentine *A. lyrata* study we revealed additional convergent loci involved in ion homoeostasis, calcium, nickel, and potassium transmembrane transport (Supplementary Table 5), suggesting existence of 'hotspot' regions in *Arabidopsis* genome in response to serpentine stress. An additional candidate gene (*FPN2* = *IREG2*) investigated in *Alyssum* (Brassicaceae; Sobczyk et al.[48]) has been found to be shared between three population pairs. Finally, when comparing to the only genomically investigated serpentine system outside Brassicaceae

(*Mimulus*, Phrymaceae; Selby[49]), there was only limited overlap in two loci with one of our population pair. On the other hand, similar functions were enriched altogether suggesting parallel adaptation through similar pathways in very divergent (~140 myr) species.

SNP data only present part of the picture and, despite linkage, do not capture structural variation. Specific TE families can be activated by abiotic stresses and possibly contribute to adaptation to challenging environments[11,50–52]. Thus, we also investigated divergence at TEs in population pairs 1 to 4 (relatively lower coverage of the N5 population did not permit this analysis) based on 21,690 TE variants called using the TEPID approach that is specifically designed for population TE variation studies[53]. Assuming linkage between each TE variant and surrounding SNPs (in the proximity of ±100 bp), we applied a similar differentiation outlier window-based workflow as specified above and identified 92–115 TE-associated candidate genes per S–N contrasts (Supplementary Data 8). In comparison with the list of candidates based on SNPs (for the same four population pairs, $n = 1{,}853$), we observed the overlap of 46 genes. The GO enrichment of TE-associated candidates from all four population pairs resulted in significant enrichment ($p < 0.05$) of functions such as transmembrane transport, water channel activity and symporter activity (Supplementary Data 9). By overlapping the lists of TE-associated candidates across S–N pairs, we identified 13 parallel TE-associated differentiation candidates (Supplementary Fig. 9 and Supplementary Data 10; significant overlap, $p < 0.05$). These loci included the plasma membrane protein *PIP2*, the putative apoplastic peroxidase *PRX37*, and *RALF-LIKE 28*, which is involved in calcium signalling. This suggests a potential impact of TEs on serpentine adaptation and gives discrete candidates for future study.

Sources of adaptive variation. Next, we tested whether variants in each serpentine adaptation candidate have arisen by parallel de novo mutations or instead came from pre-existing variation shared across populations. To do so, we modelled allele frequency covariance around repeatedly selected sites for each locus and identified the most likely of the four possible evolutionary scenarios using a designated 'Distinguishing among Modes of Convergence' (DMC) approach[54]: (i) a null-model assuming no selection (neutral model), (ii) independent de novo mutations at the same locus, (iii) repeated sampling of shared ancestral variation, and (iv) sharing of adaptive variants via migration between adapted populations. For simplicity, we considered scenarios (iii) and (iv) jointly as 'shared origin' because both processes operate on alleles of a single mutational origin, in contrast to scenario (ii). To choose the best fitting scenario for each of the 61 candidate genes, we compared the maximum composite log-likelihoods (MCLs) between the four scenarios (see 'Methods'; Supplementary Table 6). This analysis indicated that parallel selection exceeded the neutral model for 62 out of the total 84 candidate cases of parallelism (i.e. cases when two population pairs shared one of the 61 serpentine adaptation candidates). To focus only on well-justified candidates of adaptation *within* the serpentine populations, we excluded an additional 33 cases where the scenario of parallel selection with the highest MCL estimate in serpentine populations was not considerably higher (>10%) than this estimate in non-serpentine populations, which resulted in 29 candidate cases of serpentine adaptation parallelism. Shared origins dominated these results, representing 97% of the cases (28/29; Fig. 2d and Supplementary Data 7). The alternate non-neutral scenario, parallel de novo origin, was supported only for a single locus, *TWO PORE CHANNEL 1* (*TPC1*) in one case (S3–N3 and S4–N4; Fig. 2e). Using a more permissive threshold for identifying differentiation candidates (3% outliers, leading to a fivefold increase in parallel candidates) resulted in a similar DMC estimate of the proportion of the shared variation scenario (103/114 cases, i.e. 90%; Supplementary Fig. 10 and Supplementary Data 11), indicating that our inference of the dominant role of shared variation in genic parallelism is not dependent on a particular stringent outlier threshold. Finally, we applied a similar approach to parallel TE-associated differentiation candidates ($n = 13$) assuming selection on TE variants left a footprint in surrounding SNP–allele frequency covariance. We found a single non-neutral candidate, *ATPUX7*, for which parallel selection on standing variation was inferred (Supplementary Data 10). In summary, by a combination of genome-wide scanning with a designated modelling approach, we find that a non-random fraction of loci is likely reused by selection on serpentine, sourcing almost exclusively from a pool of alleles shared across the variable autotetraploid populations. Note that our conservative approach, focussed on identifying regions of repeated excessive differentiation and significant soil-related allele frequency differences, is not designed to cover the entire range of adaptive loci. Further research is thus needed to comprehensively cover the complete landscape of adaptation in autotetraploid *A. arenosa*.

Rapid recruitment of convergent de novo mutations at the calcium channel *TPC1*. One advantage of the DMC approach is an objective model selection procedure. However, it does not give fine scale information about the distribution of sequence variation at particular alleles. Therefore, we further investigated candidate alleles of the *TPC1* gene, for which DMC results suggested the sweep of different de novo mutations in independent serpentine populations. Upon closer inspection of all short-read sequences complemented by Sanger sequencing of additional 40 individuals from the three serpentine populations, a remarkably specific selection signal emerged. We found two absolutely serpentine-specific, high-frequency, non-synonymous mutations only at residue 630, overlapping the region of the highest MCL estimate for the de novo scenario in DMC, and directly adjacent to the selectivity gate of the protein in structural homology models (Fig. 3 and Supplementary Table 7). Of the two, the polymorphism Val630Leu is nearly fixed in the S3 population (25 homozygous Leu630 individuals and four heterozygous Leu630/Val630 individuals out of 29 individuals) and is at a high frequency in the S5 population (one homozygous Leu630 individual and 17 heterozygous out of 20 individuals); the second convergent Val630Tyr mutation is at high frequency in the S4 population (four homozygous Tyr630 individuals and 18 heterozygous Tyr630/Val630 out of 25 individuals; Fig. 3a, b and Supplementary Fig. 11). Strikingly, Val630Tyr requires a three-nucleotide mutation covering the entire codon (GTA to TAT). Neither of these variants were found in any other *A. arenosa* population in a range-wide catalogue[8] encompassing 1724 *TPC1* alleles (including 368 alleles from the focal area of Eastern Alps; Fig. 3a) nor in the available short-read data of the other two *Arabidopsis* outcrossing species (224 *A. lyrata* and 178 *Arabidopsis halleri* alleles, respectively, Supplementary Fig. 12), indicating that both are private to serpentine populations. Altogether, the absolute lack of either *A. arenosa* serpentine-specific variant in non-serpentine sampling across the genus strongly supports the conclusions of the DMC modelling of their independent de novo mutation origin.

To investigate the potential functional impact of these high-frequency, convergent amino acid changes, we first generated an alignment for *TPC1* homologues across the plant and animal kingdoms. Residue 630 (634/633 in *Arabidopsis thaliana*/*A. lyrata*) is conserved as either a Val or Ile across kingdoms,

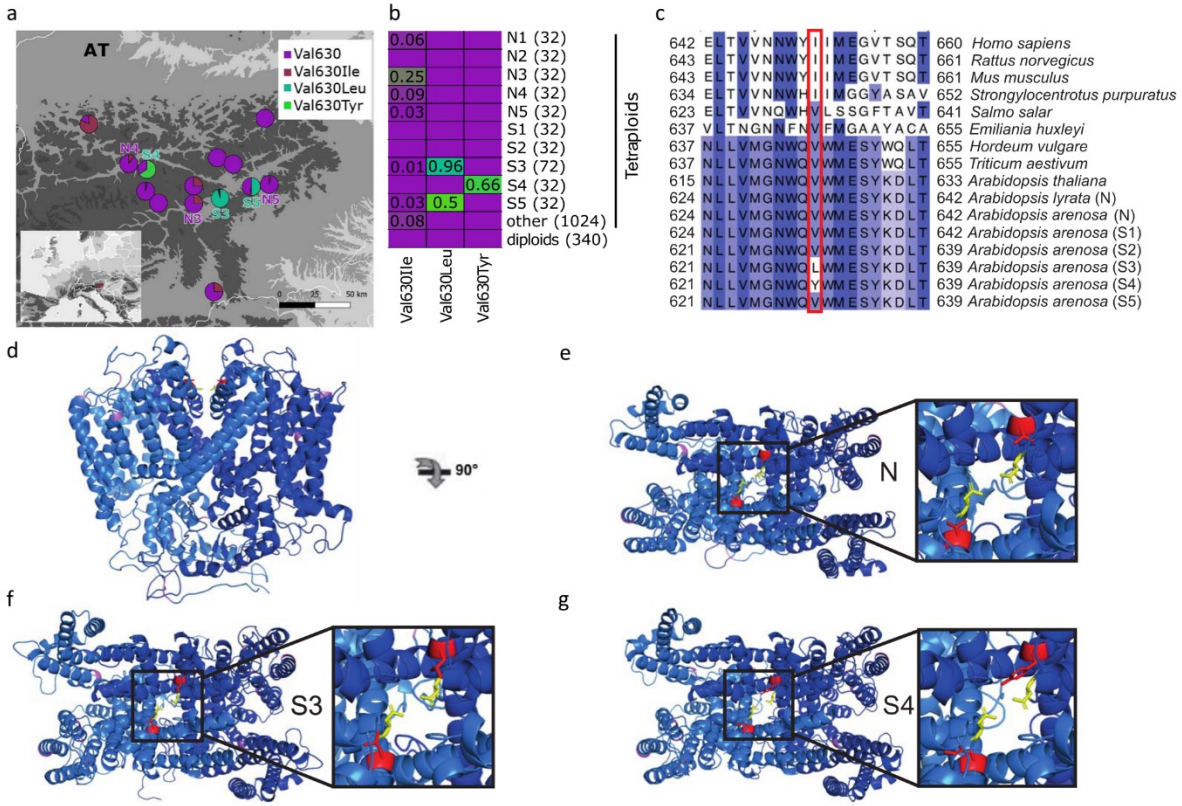


Fig. 3 Serpentine-private, convergent de novo high-impact protein changes in the *TWO PORE CHANNEL 1* (*TPC1*) locus. a All populations with serpentine-specific variants and all other resequenced *A. arenosa* populations in the focal area of Eastern Alps, showing frequencies of amino acid substitutions at residue 630 as pie charts (map drawn by V. Konečná). **b** Population frequencies of substitutions in the residue 630 among 1,724 alleles from range-wide *A. arenosa* resequenced samples. Colours denote frequencies from ancestral non-serpentine (violet) to serpentine-specific alleles (green) and number in brackets denotes total *N* of alleles screened. **c** Cross-kingdom conservation of the site shown by multiple sequence alignment of surrounding exon, including consensus sequences (AF > 0.5) from all serpentine *A. arenosa* populations (in S5 population, the frequency of Val630Leu is 0.5). Residues are coloured according to the percentage that matches the consensus sequence from 100% (dark blue) to 0% (white), the position of the serpentine-specific high-frequency non-synonymous polymorphism is highlighted in red. **d**–**g** Structural homology models of *A. arenosa TPC1* alleles. Dimeric subunits are coloured blue or marine. Non-synonymous variation that is not linked to serpentine soil is coloured deep purple. Residue 630 is coloured red and drawn as sticks. The adjacent residue, 627 (631 in *A. thaliana*), which has an experimentally demonstrated key role in selectivity control, is yellow and drawn as sticks. **d** Side view of the non-serpentine allele. **e** Top view of the non-serpentine allele with the detail of the pore opening depicted in the inset. **f** Top view of the Leu630 allele private for S3 and S5 populations. **g** Top view of the Tyr630 allele private for S4 population. Source data underlying Fig. 3a, b are provided as a Source data file.

except for the serpentine *Arabidopsis* populations (Fig. 3b, c and Supplementary Fig. 13). Val/Ile and Tyr are disparate amino acids in both size and chemical properties, so this substitution has high potential for a functional effect. Although the difference between Val/Ile and Leu is not as radical, we predicted that the physical difference between the side chains of Val/Ile versus Leu (the second terminal methyl group on the amino acid chain) may also have a functional effect by making new contacts in the tertiary structure. To test this, we performed structural homology modelling of all alleles found in *A. arenosa* (Fig. 3d–g), using two crystallographically determined structures as a template (PDB codes 5DQQ and 5E1J[55,56]). In the tertiary structure, residue 630 sits adjacent to the Asn residue (Asn627 in *A. arenosa*), which forms the pore's constriction point and has been shown to control ion selectivity in *A. thaliana*[55–58]. In *A. thaliana*, this Asn627 residue, when substituted by site-directed mutagenesis to the human homologue state, can cause Na^+ non-selective *A. thaliana TPC1* to adopt the Na^+ selectivity of human *TPC1*[56]. Depending on the rotameric conformation adopted, the Leu630 allele forms contacts with the selectivity-determining Asn627 residue, while the non-serpentine Val630 does not (Fig. 3e).

Our modelling suggests that the Tyr630 residue is even more disruptive: the Tyr side chain can adopt one of the two broad conformations, either sticking into the channel, where it occludes the opening, or sticking away from the channel and directly into surrounding residues, which have been shown to be important for stabilising Asn627 in *A. thaliana*. Both of the conformations seen, shown in heterodimer form (Fig. 3g), are highly likely to disrupt the stability of Asn627, thereby modifying the selectivity of the channel. Finally, to determine likely rotameric conformations, we generated 100 models of the S4 homodimer (two Tyr630 alleles). In this case, the two residues were significantly less likely to both point into the channel (Supplementary Table 8), suggesting that this is difficult for the structure to accommodate. Because the S4 Tyr630 allele is predominantly in a heterozygous state with the Val630 allele in nature, we also generated 100 models of the S4 heterodimer (one Tyr630 and one Val630 allele). In the heterodimer, there was no significant difference between the occurrence of the two side-chain conformations (Supplementary

Table 8), suggesting that either disruptive conformation, sticking into the pore or affecting nearby functional residues, is possible. Taken together, these results suggest that even the presence of a single Tyr630 variant within a dimer will have a substantial impact on *TPC1* function and that the Tyr630 allele is likely to be dominant or partially dominant. This prediction is consistent with its predominant occurrence in a heterozygous state and a ~50% allele frequency in S4 (Supplementary Fig. 11). We also speculate that if homodimers of the Tyr allele are too disruptive to *TPC1* function this may result in a heterozygote advantage maintained by balancing selection. In conclusion, our results suggest that the serpentine-linked allelic variation at residue 630 impacts the selectivity of *TPC1*, with functional implications, and that independent convergent de novo mutations have been repeatedly selected upon during adaptation to serpentine soils. Dedicated, electrophysiological single vacuole conductance experiments are required to explore functional changes in detail at *TPC1*. However, the exceptionally suggestive convergent changes we discovered and modelled to structures at the pore selectivity gate force the speculation that they mediate change in the relative conductance of the divalent cations Ca^{2+} and Mg^{2+}, the highly skewed ratios of which stand as hallmarks of serpentine soils[26].

Discussion

Here we inferred the evolutionary sources of adaptive variation in autotetraploid populations by taking advantage of naturally replicated adaptation to toxic serpentine soil in wild *A. arenosa*. Using a designated statistical approach leveraging parallelism, we inferred that nearly all parallel serpentine adaptation candidates were sourced from a common pool of alleles that was shared across populations. However, for one exceptional candidate—a central calcium channel shown to mediate stress signalling[46]—we identified independent de novo mutations at the same otherwise highly conserved site with likely functional consequences. Our approach informs on natural sources of parallel adaptive variation in autotetraploid populations, which has not yet been investigated genome-wide in a natural polyploid system. Yet we refrain from direct comparison with diploid ancestors because diploid serpentine populations are not known in *A. arenosa*, leaving a space for further study investigating other species encompassing multiple ploidies facing the same environmental challenge.

The potential of polyploidy to enhance adaptation is a matter of ongoing debate that is mainly fuelled by theory-based controversies revolving around the efficiency of selection[6,19,59,60] and observations of the frequencies of WGD events in space and time[5,22]. In contrast, empirical population-level investigations that unravel evolutionary mechanisms operating in natural autopolyploids are very scarce. It has been shown that autopolyploids may rapidly react to new challenges by landscape genetic[23] and experimental studies[24,61]. Our transplant experiments coupled with demographic investigations support this and further demonstrate that such rapid adaptation may be repeated many times within a species, partially drawing on the same variants. Furthermore, a previous study in *A. arenosa* showed that the genome-wide proportion of non-synonymous polymorphisms fixed by directional selection was higher in tetraploids than in diploids, suggesting increased adaptive variation in natural autopolyploids[8]. However, it remains unclear whether such variation reflects increased input of novel mutations (as observed, e.g. in experimental yeast populations[17]) or sampling from increased standing variation (as predicted by theory[5,6]). Here we find nearly exclusive repeated sampling from a shared pool of variants. Such a dominant role of pre-existing variation is in line with the studies of parallel adaptation in diploid systems such as *Littorina* snails[62,63], stickleback fishes[64,65], *Helliconius* butterflies[66], *Sinosuthora webbiana* vinous-throated parrotbill birds[39], *Ipomoea purpurea* morning glories[67], *Apis cerana* Asian honebees[68], or *Coilia nasus* fishes[69]. Yet examples are lacking from autopolyploid systems, where a large pool of standing variation is expected by theory due to larger effective population size and polysomic masking of allelic variation[5,70]. Sharing alleles that have persisted in a specific genomic environment already for some time may be particularly beneficial under intense selection when rapid adaptive responses are needed[67,71–73]. In addition, standing genetic variation likely minimises negative pleiotropic effects of linked variants[67,74,75]. It should be noted, however, that our estimates may be biased upward for shared alleles by focussing only on cases of parallelism, which provided a testable framework for our inference of the sources of variants. Larger fractions of novel mutations may be represented among the non-parallel adaptive variation, which is, however, harder to identify.

In contrast to shared variants, empirical evidence for parallel de novo mutations within species is rare even in diploids[73,76–78] and we lack any example from polyploids. Theory suggests that such a scenario is unlikely for autopolyploids, as reduced efficacy of selection on a novel, initially low-frequency variants is predicted for most dominance states in autopolyploids[6,18,19]. On the other hand, beneficial alleles are introduced at increased rates in doubled genomes[6,17] and additional variation may accumulate due to polysomic masking[6]. Here we provide an example of parallel recruitment of two distinct de novo mutations with likely phenotypic effect in separate polyploid populations within one species, demonstrating that adaptive sourcing from novel polymorphisms is in fact feasible even in autopolyploids. Interestingly, high frequencies of homozygous individuals in one population demonstrates that such novel variants may approach fixation, in stark contrast to theory, which predicts incomplete sweeps of dominant mutations to be prevalent in autotetraploids[6,19]. On the other hand, the prevalence of heterozygotes in the other serpentine population together with results of structural modelling suggest (at least partial) dominance of the serpentine allele.[6]

Overall, our study demonstrates that rapid environmental adaptation may repeatedly occur in established autopolyploid populations. Footprints of selection at similar genomic positions mostly occur because of the repeated recruitment from a large pool of pre-existing variation, yet exceptionally also from recurrent de novo mutations. Thus, these results support the emerging view of autopolyploids as diverse evolutionary amalgamates, capable of flexible adaptation in response to environmental challenge.

Methods

Field sampling. Serpentines occur in Central Europe as scattered edaphic 'islands' surrounded by open rocky habitats on other substrates in which autotetraploid *A. arenosa* frequently occur. In contrast, *A. arenosa* colonised only some serpentine sites in this area[79,80] indirectly suggesting that colonisation of serpentine sites by surrounding non-serpentine populations happened in parallel and was probably linked with local substrate adaptation. To test this hypothesis, we sampled all five serpentine (S) populations of *A. arenosa* known to date and complemented each by a proximal (<19 km distant) non-serpentine (N) population. All N populations grew in similar vegetation (rocky outcrops in open forests or grasslands) and soil type (siliceous to neutral rocks; Supplementary Table 1). Although we observed a considerable variation in the overall soil chemistry in our samples (Supplementary Fig. 14), the principal soil factors differentiating between S and N populations were always the same—higher Mg, Ni, and Co and lower Ca/Mg in S populations. Diploid serpentine *A. arenosa* is not known, even though serpentine barrens are frequent in some diploid-dominated areas, such as the Balkan peninsula. The sampled populations covered considerable elevational gradient (414–1750 m a.s.l.), but the differences in elevation within the pairs were small except for one pair where no nearby subalpine non-serpentine population exists (population pair S4–N4, difference 740 m).

We sampled eight individuals per every population for genomic analysis and confirmed their tetraploid level by flow cytometry. For each individual, we also sampled soil from very close proximity to the roots (~10–20 cm below ground),

except for N5 population for which genotyped data were already taken from the previous study[8]. There we collected an additional eight soil samples and use their average in the following environmental association analysis. For the transplant experiment, we also collected seeds (~20–30 maternal plants/population) from three population pairs (S1–N1, S2–N2, S3–N3) and bulks of soil ~80 l (sieved afterwards) from the natural sites occupied by these six populations.

DNA extraction, library preparation, sequencing, raw data processing, and filtration. We stored all samples for this study in RNAlater (R0901-500ML, SIGMA-ALDRICH CO LTD) to avoid genomic DNA degradation and we further prepared the leaf material as described in refs. [81–83]. We extracted DNA as described in Supplementary Method 1. Genomic libraries for sequencing were prepared using the Illumina TRUSeq PCR-free library. Libraries were sequenced as 150 bp paired-end reads on a HiSeq 4000 (3 lanes in total) by Norwegian Sequencing Centre, University of Oslo.

We used trimmomatic-0.36[84] to remove adaptor sequences and low-quality base pairs (<15 PHRED quality score). Trimmed reads >100 bp were mapped to reference genome of North American *A. lyrata*[85] by bwa-0.7.15[86] (https://rcc.uchicago.edu/docs/software/modules/bwa/midway2/0.7.15.html) with default setting. Duplicated reads were identified by picard-2.8 (https://github.com/broadinstitute/picard) and discarded together with reads that showed low mapping quality (<25). Afterwards we used GATK v.3.7 to call and filter reliable variants and invariant sites according to best practices[87] (complete variant calling pipeline available at https://github.com/vlkofly/Fastq-to-vcf). Namely, we used the HaplotypeCaller module to call variants per individual using the ploidy = 4 option, which enables calling full tetraploid genotypes. Then we aggregated variants across all individuals by module GenotypeGVCFs. We selected only biallelic SNPs and removed those that matched the following criteria: Quality by Depth (QD) < 2.0, FisherStrand (FS) > 60.0, RMSMappingQuality (MQ) < 40.0, MappingQualityRankSumTest (MQRS) < −12.5, ReadPosRankSum < −8.0, StrandOddsRatio (SOR) > 3.0. We called invariant sites also with the GATK pipeline similarly to variants, and we removed sites where QUAL was <15. Both variants and invariants were masked for sites with average read depth (RD) >2 times standard deviation as these sites were most likely located in duplicated regions and we also masked regions with excessive heterozygosity, representing likely paralogous mis-assembled regions, following Monnahan et al.[8]. One individual per each S2 and N5 populations was excluded due to exceptionally bad data quality (low percentage of mapped reads and low RD, <10 on average), leaving us with a final data set of 78 individuals that were used in genomic analyses. This pre-filtered data set contained 110,358,565 sites (of which 11,744,200 were SNPs) with average depth of coverage 21× (Supplementary Data 1). This way of genotyping leads to allele frequency estimates that are well comparable with previous estimates[88] including a broad range-wide *A. arenosa* sampling[8] (the site frequency spectra (SFS) are presented in Supplementary Fig. 15). Our site frequency estimates, which were constructed by program est-sfs[89], are likely not biased by the number of individuals as was demonstrated by consistent site frequency estimates when subsampling one more deeply sampled population to the most common number of 32 chromosomes (S3; when including additional nine individuals from Arnold et al.[29]; Supplementary Fig. 16).

Reconstruction of population genetic structure. We inferred the population genetic structure, diversity, and relationships among individuals from putatively neutral 4dg SNPs filtered for DP >8 per individual and maximum fraction of filtered genotypes (MFFG) of 0.2, i.e. allowing max. 20% missing calls per site (1,042,793 SNPs with a total of 0.49% missing data; see Supplementary Tables 1–3 and Supplementary Data 1 for description of data sets and filtration criteria). We used several complementary approaches. First, we ran principal component analysis (PCA) on individual genotypes using glPCA function in adegenet v.2.1.1 replacing the missing values by average allele frequency for that locus. Second, we applied model-based clustering with accelerated variational inference in fastStructure v.1.0[90]. To remove the effect of linkage, we randomly selected one SNP per a 1 kb window, keeping 10 kb distance between the windows and, additionally, filtered for minimum minor allele frequency (MAF) = 0.05 resulting in a data set of 9,923 SNPs. As fastStructure does not handle the polyploid genotypes directly, we randomly subsampled two alleles per each tetraploid site using a custom script. This approach has been demonstrated to provide unbiased clustering in autotetraploid samples in general[91] and *Arabidopsis* in particular[8]. We ran fastStructure with 10 replicates under $K = 5$ (corresponding to the number of population pairs) with default settings. Third, we inferred relationships among populations using allele frequency covariance graphs implemented in TreeMix v.1.13[92]. We used custom python3 scripts (available at https://github.com/mbohutinska/TreeMix_input) to create the input files. We ran TreeMix analysis rooted with an outgroup population (tetraploid *A. arenosa* population 'Hranovnica' from the area of origin of the autotetraploid cytotype in Western Carpathians[93]). We repeated the analysis over the range of 0–6 migration edges to investigate the change in the explanatory power of the model when assuming migration event(s) and found that adding migration to the model did not lead to large improvement (Supplementary Fig. 17). We bootstrapped the scenario without migration (the topology did not change with adding the migrations) choosing bootstrap block size 1 kb (the same window size also for the divergence scan, see below) and 100 replicates and summarised the results using *SumTrees.py* function in DendroPy[78]. Finally, we calculated nucleotide diversity (π) and Tajima's D for each population and pairwise differentiation (F_{ST}) (Supplementary Table 9) for each population pair using custom python3 scripts (available at https://github.com/mbohutinska/ScanTools_ProtEvol; see Supplementary Table 3 for the number of sites per each population). For nucleotide diversity calculation, we down-sampled each population to six individuals on a per-site basis to keep equal sample per each population while also keeping the maximum number of sites with zero missingness.

Demographic inference. We performed demographic analyses in fastsimcoal v.2.6[94] to specifically test for parallel origin of serpentine populations and to estimate divergence time between serpentine and proximal non-serpentine populations. We constructed unfolded multidimensional SFS from the variant and invariant 4dg sites (filtered in the same ways as above, Supplementary Table 2) using custom python scripts published in our earlier study (FSC2input.py at https://github.com/pmonnahan/ScanTools/)[8]. We repolarized a subset of sites using genotyped individuals across closely related diploid *Arabidopsis* species to avoid erroneous inference of ancestral state based on a single reference *A. lyrata* individual following Monnahan[8].

First, we tested for parallel origin of serpentine populations using population quartets (two pairs of geographically proximal serpentine and non-serpentine populations) and iterated such pairs across all combinations of regions (10 pairwise combinations among the five regions in total). For each quartet, we created four-dimensional SFS and compared following the four evolutionary scenarios (Supplementary Fig. 2): (i) parallel origin of serpentine ecotype—sister position of serpentine and non-serpentine populations within the same region, (ii) parallel origin with migration—the same topology with additional gene flow between serpentine and the proximal non-serpentine population, (iii) single origin of serpentine ecotype—sister position of serpentine populations and of non-serpentine populations, respectively, and (iv) single origin with migration—the same topology with additional gene flow between serpentine and the proximal non-serpentine populations. For each scenario and population quartet, 50 fastsimcoal runs were performed. For each run, we allowed for 40 ECM optimisation cycles to estimate the parameters and 100,000 simulations in each step to estimate the expected SFS. We used wide range of initial parameters (effective population size, divergence times, migration rates; see the example *.est and *.tpl files provided for each model tested in the Supplementary Data 12) and assumed mutation rate of 4.3×10^{-8} inferred for *A. arenosa* previously[32]. Further, we extracted the best likelihood partition for each fastsimcoal run, calculated Akaike information criterion (AIC), and summarised the AIC values across the 50 fastsimcoal runs. The scenario with lowest median AIC values within each particular population quartet was preferred (Supplementary Fig. 2 and Supplementary Data 2).

Second, we estimated divergence time between S and N populations from each population pair (i.e. S1–N1, S2–N2, S3–N3, S4–N4, and S5–N5) based on two-dimensional SFS using the same fastsimcoal settings as above. We simulated according to models of two-population split, not assuming migration because the model with migration did not significantly increase the model fit across the quartets of populations (Supplementary Fig. 2 and Supplementary Data 2). To calculate 95% confidence intervals for parameter estimates (Table 1), we sampled with replacement the original SNP matrices to create 100 bootstrap replicates of the two-dimensional SFS per each of the five population pairs.

Window-based scans for directional selection. We leveraged the fivefold-replicated natural set-up to identify candidate genes that show repeated footprints of selection across multiple events of serpentine colonisation. First, we identified genes of excessive divergence for each pair of proximal serpentine (S)–non-serpentine (N) populations (five pairs in total). Reflecting hierarchical structure in the data, we avoided merging multiple populations into larger units for the estimation of F_{ST} and strictly worked in a pairwise design. We admit that population S3 occupies a somewhat separate position and its ancestral non-serpentine population might have thus remained unsampled (or got extinct)—we therefore used the spatially closest population N3 as the most representative paired population available in our sampling. We calculated pairwise F_{ST}[95] for non-overlapping 1 kbp windows along the genome with the minimum of 10 SNPs per window to exclude potential biases in F_{ST} estimation caused by low-informative windows[96]. We used the custom script (https://github.com/mbohutinska/ScanTools_ProtEvol) based on ScanTools pipeline that has been successfully applied in our previous analyses of autotetraploid *A. arenosa*[8]. The window size of 1 kbp was selected to properly account for the average genome-wide linkage disequilibrium (LD) decay of genotypic correlations (150–800 bp) previously estimated in autotetraploid *A. arenosa*[40]. We used windows of fixed length and thus with homogeneous position on genome across all population pairs (in contrast to windows defined by number of SNPs and thus varying in exact position and length) to facilitate comparisons of selection candidates across distinct population pairs. Our F_{ST} estimates are unlikely to be strongly affected by varying numbers of SNPs per window, as the correlation between F_{ST} and number of SNPs per window was very weak (Spearman's rank correlation coefficient varied from 0.02 to 0.04 across population pairs; Supplementary Fig. 18). We identified the upper 99% quantile of all 1 kbp windows in the empirical distribution of F_{ST} metric per each population pair. Then we identified

initial lists of genes with excessive differentiation for each S–N population pair as genes overlapping with the 1% outlier windows using *A. lyrata* gene annotation[97], where the gene includes 5' untranslated regions (UTRs), start codons, exons, introns, stop codons, and 3' UTRs. Then we refined this inclusive list and identified parallel differentiation candidates as genes identified as candidates in at least two population pairs. We tested whether such overlap is higher than a random number of overlapping items given the sample size using Fisher's exact test in SuperExactTest R package[98]. Our F_{ST}-based detection of outlier windows was not largely biased towards regions with low recombination rate (based on the available *A. lyrata* recombination map[99]; Supplementary Fig. 19).

Environmental association analysis. To further refine the candidate list to genes associated with the discriminative serpentine soil parameters, we performed environmental association analysis using LFMMs—LFMM 2 (https://bcm-uga.github.io/lfmm/)[41]. We tested the association of allele frequencies at each SNP for each individual with associated soil concentration of the key elements differentiating serpentine and non-serpentine soils: Ca/Mg ratio and bioavailable soil concentrations of Co, Mg, and Ni. Only those elements were significant in one-way ANOVAs (Bonferroni corrected) testing differences in elemental soil concentration between S and N population, taking population pair as a random variable. We retained 1,783,055 SNPs without missing data and MAF > 0.05 as an input for the LFMM analysis. LFMM accounts for a discrete number of ancestral population groups as latent factors. We used five latent factors reflecting the number of population pairs. Due to hierarchical structure in the data (PCA based on ~1 M 4dg SNPs indicated five main components, yet the first three axes alone also explained considerable variation; Supplementary Fig. 20), we also performed the additional analysis assuming three latent factors. As such analysis had only a minor effect on the total number of serpentine adaptation candidates (reducing their number by only two), we further used a candidate list based on five latent factors that corresponds to the total number of population pairs, thus it is also better comparable with the parallel differentiation candidates. To identify SNPs significantly associated with soil variables, we transformed p values to false discovery rate (<0.05) based on q values using the qvalue R package v.2.20[100]. Finally, we annotated the candidate SNPs to genes, termed 'LFMM candidates' (at least one significantly associated SNP per candidate gene).

We made a final shortlist of serpentine adaptation candidates by overlapping the LFMM candidates, reflecting significant association with important soil elements, with the previously identified parallel differentiation candidates, mirroring regions of excessive differentiation repeatedly found across parallel population pairs. For visualisation purposes (Fig. 2e), we annotated SNPs in the serpentine adaptation candidates using SnpEff v. 4.3[101] following *A. lyrata* version 2 genome annotation[97].

TE variant calling and analysis. TE variants (insertions or deletions) among sequenced individuals were identified and genotyped in population pairs 1–4 using TEPID v.0.8[53] following the approach described in Rogivue et al.[50] (relatively lower coverage of the N5 population did not permit this analysis in the last pair). We annotated TEs based on available *A. lyrata* TE reference[102]. TEPID is based on split and discordant read mapping information and employs read mapping quality, sequencing breakpoints, and local variation in sequencing coverage to call the absence of reference TEs as well as the presence of non-reference TE copies. This method is specifically suited for studies at the population level as it takes intra-population polymorphism into account to refine TE calls in focal samples by supporting reliable call of non-reference alleles under lower thresholds when found in other individuals of the population. We filtered the data set by excluding variants with MFFG > 0.2 and DP < 8, which resulted in 21,690 TE variants (13,542 deletions and 8,148 insertions as compared to the reference).

Assuming linkage between TE variants and nearby SNPs[53], we calculated pairwise F_{ST}[95] using SNP frequencies (in the same way as specified above) for non-overlapping 1 kbp windows containing TE variant(s) for each population pair. The candidate windows for directional selection were identified as the upper 99% quantile of all windows containing a TE variant in the empirical distribution of F_{ST} metric per each population pair. Further, we identified candidate genes (TE-associated candidates) as those present up to +/−2 kbp upstream and downstream from the candidate TE variant (assuming functional impact of TE variant until such distance, following Hollister et al.[103]). Finally, we identified parallel TE-associated differentiation candidates as those loci that appeared as candidates in at least two population pairs.

GO enrichment analysis. We inferred potential functional consequences of the candidate gene lists using GO enrichment tests within biological processes, molecular functions, and cellular components domains. We applied Fisher's exact test ($p < 0.05$) with the 'elim' algorithm implemented in topGO v.2.42 R package[104]. We worked with *A. thaliana* orthologues of *A. lyrata* genes obtained using biomaRt[105] and *A. thaliana* was also used as the background gene universe in all gene set enrichment analyses. The used 'elim' algorithm traverses the GO hierarchy from the bottom to the top, discarding genes that have already been mapped to significant child terms while accounting for the total number of genes annotated in the GO term[104,106].

Modelling the sources of adaptive variation. For each serpentine adaptation candidate ($n = 61$ and $n = 13$ identified using SNPs and TE-associated variants, respectively), we modelled whether it exhibits patterns of parallel selection that is beyond neutrality and if so whether the parallel selection operated on de novo mutations or rather called on pre-existing variation shared across populations. We used model-based likelihood approach that is specifically designed to identify loci involved in parallel evolution and to distinguish among their evolutionary sources (DMC[54]). Convergent is analogous to parallel in this case as the entire approach is designed for closely related populations.

We considered the following four evolutionary scenarios assuming distinct variation sources: (i) no selection (neutral model), (ii) independent de novo mutations at the same locus, (iii) repeated sampling of ancestral variation that was standing in the non-adapted populations prior the onset of selection, and (iv) transfer of adaptive variants via gene flow (migration) from another adapted population. We interpreted the last two scenarios together as a variation that is shared across populations as both operate on the same allele(s) that do not reflect independent mutations.

We estimated composite log-likelihoods for each gene and under each scenario using a broad range of realistic parameters taking into account demographic history of our populations inferred previously and following recommendations in Lee and Coop[54] (positions of selected sites, selection coefficients, migration rates, times for which allele was standing in the populations prior to onset of selection, and initial allele frequencies prior to selection; see Supplementary Table 6 for the summary of all parameters and their ranges). We chose to place selected sites at eight locations at equal distance (default value recommended by authors https://github.com/kristinmlee/dmc) from each other along the particular gene. Such a density (one site per ~500 bp on average, as the mean length of serpentine adaptation candidate is ~4,000 bp) is in fact well within the range of the LD decay of 150–800 bp, which was estimated in *A. arenosa*[40]. For the calculation of co-ancestry decay, we also considered a 25 kbp upstream and downstream region from each gene. To choose the best fitting scenario for each candidate, we first estimated the MCL over the parameters for each of the three parallel selection scenarios and a neutral scenario. We selected among the parallel selection models by choosing the model with the highest MCL, following the approach of ref. [49]. Further, we considered the case significantly non-neutral only if the MCL difference between the selected parallel model and the corresponding neutral model was higher than the maximum of the distribution of the differences from the simulated neutral data in *A. arenosa* inferred in Bohutínská et al.[40] (i.e. MCL difference >21, a conservative estimate). Further, to focus only on divergence caused by selection in serpentine populations (i.e. eliminating divergence signals caused by selection in N populations), only cases of selection in the serpentine populations in which the scenario of parallel selection had a considerably higher MCL estimate (>10%) than selection in non-serpentine populations were taken into account.

To ensure that our inference on the relative importance of shared versus de novo variation is not biased by arbitrary outlier threshold selection, we re-analysed the SNP data set using a 3% outlier F_{ST} threshold for identifying differentiation candidates. We overlapped the resulting 1,179 parallel differentiation candidates with the LFMM candidates and subjected the resulting 246 serpentine adaptation candidates to DMC modelling in the same way as described above (Supplementary Data 11). For each of the 246 genes and parallel quartets of populations (in total 420 cases of parallelism) we again compared the four scenarios as described above (Supplementary Fig. 10 and Supplementary Data 11).

Reciprocal transplant experiment. To test for local substrate adaptation in three serpentine populations, we compared plant fitness in the native versus foreign soil in a reciprocal transplant experiment. We reciprocally transplanted plants of serpentine and non-serpentine origin from three population pairs (S1–N1, S2–N2, S3–N3) that served as representatives of independent colonisation in each broader geographic region (Bohemian Massif, lower Austria, and Eastern Alps, respectively). As the most stressful factor for *A. arenosa* populations growing on serpentine sites is the substrate (Arnold et al.[29] and Supplementary Fig. 3a, b), we isolated the soil effect by cultivating plants in similar climatic conditions in the greenhouse.

For each pair, we cultivated the plants in serpentine and non-serpentine soil originating from their original sites (i.e. S1 plant cultivated in S1 and N1 soil and vice versa) and tested for the interaction between the soil treatment and soil of origin in selected fitness indicators (germination and rosette diameter sizes). We germinated seeds from 12 maternal plants (each representing a seed family of a mixture of full- and half-sibs) from each population in Ppetri dishes filled by either type of soil (15 seeds/family/treatment). Seeds germinated in the growth chamber (Conviron) under conditions approximating spring season at the original sites: 12 h dark at 10 °C and 12 h light at 20 °C. We recorded the germination date as the appearance of cotyledon leaves for the period of 20 days after which there were no new seedlings emerging. We tested for the effect of substrate of origin (serpentine versus non-serpentine), soil treatment (serpentine versus non-serpentine), and their interaction on germination proportion using GLM with binomial errors. To account for lineage-specific differences between population pairs, which are uninformative for the overall assessment of the fitness response towards serpentine, we treated population pairs as a random variable.

Due to zero germination of N1 seeds in S1 soil, we measured differential growth response on plants that were germinated in the non-serpentine soils and were subjected to the differential soil treatment later, in a seedling stage. We chose 44–50 seedlings equally representing progeny of 11 maternal plants per each population (in total 284 seedlings), transferred each plant to a separate pot filled either with ~1 L of the original or the alternative paired soil (i.e. S1 soil for N1 population and vice versa). We randomly swapped the position of each pot twice a week and watered them with tap water when needed. We measured the rosette diameter and counted the number of leaves (which correlated with the rosette diameter, $R = 0.85$, $p < 0.001$) twice a week for five weeks until rosette growth reached a plateau (Supplementary Fig. 6). By that time, we observed zero mortality and only negligible flowering (1%, four of the 284 plants). We tested whether soil treatment (serpentine versus non-serpentine) with the interaction of soil of origin (serpentine versus non-serpentine) had a significant effect on rosette diameter sizes (the maximum rosette diameter sizes from the last tenth measurement as a dependent variable), using two-way ANOVA taking population pair (1–3) as a random factor.

Elemental analysis of soil and leaf samples. We quantified the soil elemental composition by inductively coupled plasma mass spectrometry (ICP-MS; PerkinElmer NexION 2000, University of Nottingham). We monitored 21 elements (Na, Mg, P, S, K, Ca, Ti, Cr, Mn, Fe, Co, Ni, Cu, Zn, As, Se, Rb, Sr, Mo, Cd, and Pb) in the soil extract samples. Individual soil samples of genotyped individuals (80 samples in total) were dried at 60 °C. Soil samples were sieved afterwards. Samples were prepared according to a protocol summarised in Supplementary Method 2. We quantified the elemental soil and leaf composition of the elements in samples from reciprocal transplant experiment by ICP OES spectrometer INTEGRA 6000 (GBC, Dandenong Australia). We monitored three elements that were identified as key elements differentiating S and N soils of the natural populations (Ca, Mg, and Ni) and decomposed samples prior the analysis. For details, see Supplementary Method 3.

Screening natural variation in the *TPC1* locus. To screen a broader set of *TPC1* genotypes in the relevant serpentine populations S3–S5, we Sanger-sequenced additional individuals sampled at the original sites of the focal S3, S4, and S5 populations (11, 17, and 12 individuals, respectively) exhibiting non-synonymous variation in the *TPC1* locus. For the amplification of the exon around the candidate site, we used specifically designed primers (Supplementary Table 10). The mix for PCR contained 0.3 µL of forward and reverse primer each, 14.2 µL of ddH2O, 0.2 µL of MyTag DNA polymerase, and 4 µL of reaction buffer MyTag, and we added 1 µL (10 ng) of DNA. The PCR amplification was conducted in a thermocycler (Eppendorf Mastercycler Pro) under the following conditions: 1 min of denaturation at 95 °C, followed by 35 cycles: 20 s at 95 °C, 25 s at 60 °C, 45 s at 72 °C, and a final extension for 5 min at 72 °C. Amplification products of high purity were sequenced at 3130xl Genetic Analyser (DNA laboratory of Faculty of Science, Charles University, Prague).

Then we checked whether the candidate alleles, inferred as serpentine specific in our data set, are also absent in a published broad non-serpentine sampling among outcrossing *Arabidopsis* species[8,33,107–112]. We downloaded all the available short-read genomic sequences published with the referred studies, called variants using the same approach as described above, and checked the genotypes at the candidate site (residue 630 in *A. arenosa* and 633 in *A. lyrata* and *A. halleri*). In total, we screened 1724 alleles of *A. arenosa*, 178 alleles of *A. halleri*, and 224 alleles of *A. lyrata*.

To visually compare variation at the entire *TPC1* locus, we generated consensus sequences for the group of all five non-serpentine *A. arenosa* populations and for each separate serpentine population using the Variant Call Format (VCF) with all variants in the region of scaffold_6:23,042,733–23,048,601 using bcftools (Supplementary Fig. 21). Sites were included in the consensus sequence if they had AF > 50%. As the original VCF contained only biallelic sites, an additional multiallelic VCF was also created using GATK and any variants with AF > 50% were manually added to the biallelic consensus sequence. The *A. lyrata* and *A. thaliana* non-serpentine sequences were assumed to match the corresponding reference for each species.

Finally, we screened the variation in the *TPC1* locus at deep phylogenetic scales. We generated multiple sequence alignments using Clustal-Omega[113] from the available mRNA sequences from GenBank (*Emiliania huxleyi*, *Hordeum vulgare*, *A. thaliana*, *Triticum aestivum*, *Salmo salar*, *Strongylocentrotus purpuratus*, *Homo sapiens*, *Rattus norvegicus*, and *Mus musculus*) that were complemented by the consensus *Arabidopsis* sequences described above. Alignments were manually refined and visualised in JalView[114].

Structural homology models. Structural homology models of dimeric TPC1 were generated with Modeller v. 9.24[115] using two *A. thaliana* crystal structures (5E1J and 5DQQ[55,57]) as templates. The final model was determined by its discrete optimised protein energy score.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Data supporting the findings of this work are available within the paper and its Supplementary Information files. A Reporting Summary for this Article is available as a Supplementary Information file. Sequence data generated in this study have been deposited in the GenBank SRA database as a BioProject PRJNA667586 (populations S1–S5, N1–N4). Additional sequence data used in this study are deposited in the GenBank SRA database within a BioProject PRJNA325082 (population N5). Source data are provided with this paper.

Code availability

Newly developed scripts are available at GitHub [https://github.com/vlkofly/Fastq-to-vcf] and [https://github.com/mbohutinska/ScanTools_ProtEvol].

Received: 25 January 2021; Accepted: 21 July 2021;
Published online: 17 August 2021

References

1. Stern, D. L. The genetic causes of convergent evolution. *Nat. Rev. Genet.* **14**, 751–764 (2013).
2. Barrett, R. D. H. & Schluter, D. Adaptation from standing genetic variation. *Trends Ecol. Evol.* **23**, 38–44 (2008).
3. Wood, T. E. et al. The frequency of polyploid speciation in vascular plants. *Proc. Natl Acad. Sci. USA* **106**, 13875–13879 (2009).
4. Soltis, D. E., Visger, C. J. & Soltis, P. S. The polyploidy revolution then…and now: Stebbins revisited. *Am. J. Bot.* **101**, 1057–1078 (2014).
5. Van De Peer, Y., Mizrachi, E. & Marchal, K. The evolutionary significance of polyploidy. *Nat. Rev. Genet.* **18**, 411–424 (2017).
6. Otto, S. P. & Whitton, J. Polyploid incidence and evolution. *Annu. Rev. Genet.* **34**, 401–437 (2000).
7. Otto, S. P. The evolutionary consequences of polyploidy. *Cell* **131**, 452–462 (2007).
8. Monnahan, P. et al. Pervasive population genomic consequences of genome duplication in *Arabidopsis arenosa*. *Nat. Ecol. Evol.* **3**, 457 (2019).
9. Van de Peer, Y., Ashman, T. L., Soltis, P. S. & Soltis, D. E. Polyploidy: an evolutionary and ecological force in stressful times. *Plant Cell* **33**, 11–26 (2020).
10. Bardil, A., Tayalé, A. & Parisod, C. Evolutionary dynamics of retrotransposons following autopolyploidy in the Buckler Mustard species complex. *Plant J.* **82**, 621–631 (2015).
11. Baduel, P., Quadrana, L., Hunter, B., Bomblies, K. & Colot, V. Relaxed purifying selection in autopolyploids drives transposable element over-accumulation which provides variants for local adaptation. *Nat. Commun.* **10**, 5818 (2019).
12. Ramsey, J. Polyploidy and ecological adaptation in wild yarrow. *Proc. Natl Acad. Sci. USA* **108**, 7096–7101 (2011).
13. Chao, D. et al. Polyploids exhibit higher potassium uptake and salinity tolerance in Arabidopsis. *Science* **341**, 658–659 (2013).
14. Bomblies, K. When everything changes at once: finding a new normal after genome duplication. *Proc. R. Soc. B Biol. Sci.* **287**, 20202154 (2020).
15. Soltis, P. S. & Soltis, D. E. The role of genetic and genomic attributes in the success of polyploids. *Proc. Natl Acad. Sci. USA* **97**, 7051–7057 (2000).
16. Haldane, J. B. S. *The Causes of Evolution* (Princeton University Press, 1932).
17. Selmecki, A. M. et al. Polyploidy can drive rapid adaptation in yeast. *Nature* **519**, 349–351 (2015).
18. Gerstein, A. C. & Otto, S. P. Ploidy and the causes of genomic evolution. *J. Hered.* **100**, 571–581 (2009).
19. Monnahan, P. & Brandvain, Y. The effect of autopolyploidy on population genetic signals of hard sweeps. *Biol. Lett.* **16**, 20190796 (2020).
20. Yao, Y., Carretero-Paulet, L. & Van de Peer, Y. Using digital organisms to study the evolutionary consequences of whole genome duplication and polyploidy. *PLoS ONE* **14**, e0220257 (2019).
21. Brochmann, C. et al. Polyploidy in arctic plants. *Biol. J. Linn. Soc.* **82**, 521–536 (2004).
22. Rice, A. et al. The global biogeography of polyploid plants. *Nat. Ecol. Evol.* **3**, 265–273 (2019).
23. Parisod, C. & Besnard, G. Glacial in situ survival in the Western Alps and polytopic autopolyploidy in *Biscutella laevigata* L. (Brassicaceae). *Mol. Ecol.* **16**, 2755–2767 (2007).
24. Martin, S. L. & Husband, B. C. Adaptation of diploid and tetraploid *Chamerion angustifolium* to elevation but not local environment. *Evolution* **67**, 1780–1791 (2013).

25. Wei, N., Cronn, R., Liston, A. & Ashman, T. L. Functional trait divergence and trait plasticity confer polyploid advantage in heterogeneous environments. *N. Phytol.* **221**, 2286–2297 (2019).
26. O'Dell, R. E. & Rajakaruna, N. in *Serpentine: Evolution and Ecology in a Model System* (eds Harrison, S. & Rajakaruna, N.) 97–137 (University of California Press, 2011).
27. Yant, L. & Bomblies, K. Genomic studies of adaptive evolution in outcrossing Arabidopsis species. *Curr. Opin. Plant Biol.* **36**, 9–14 (2017).
28. Molina-Henao, Y. F. & Hopkins, R. Autopolyploid lineage shows climatic niche expansion but not divergence in *Arabidopsis arenosa*. *Am. J. Bot.* **106**, 61–70 (2019).
29. Arnold, B. J. et al. Borrowed alleles and convergence in serpentine adaptation. *Proc. Natl Acad. Sci. USA* **113**, 8320–8325 (2016).
30. Baduel, P., Hunter, B., Yeola, S. & Bomblies, K. Genetic basis and evolution of rapid cycling in railway populations of tetraploid *Arabidopsis arenosa*. *PLoS Genet.* **14**, 1–26 (2018).
31. Baduel, P., Arnold, B., Weisman, C. M., Hunter, B. & Bomblies, K. Habitat-associated life history and stress-tolerance variation in *Arabidopsis arenosa*. *Plant Physiol.* **171**, 437–451 (2016).
32. Przedpełska, E. & Wierzbicka, M. *Arabidopsis arenosa* (Brassicaceae) from a lead-zinc waste heap in southern Poland - a plant with high tolerance to heavy metals. *Plant Soil* **299**, 43–53 (2007).
33. Preite, V. et al. Convergent evolution in *Arabidopsis halleri* and *Arabidopsis arenosa* on calamine metalliferous soils. *Philos. Trans. R. Soc. B* **374**, 20180243 (2019).
34. Brady, K. U., Kruckeberg, A. R. & Bradshaw, H. D. Jr Evolutionary ecology of plant adaptation to serpentine soils. *Annu. Rev. Ecol. Evol. Syst.* **36**, 243–266 (2005).
35. Kazakou, E., Dimitrakopoulos, P. G., Baker, A. J. M., Reeves, R. D. & Troumbis, A. Y. Hypotheses, mechanisms and trade-offs of tolerance and adaptation to serpentine soils: from species to ecosystem level. *Biol. Rev.* **83**, 495–508 (2008).
36. Konečná, V., Yant, L. & Kolář, F. The evolutionary genomics of serpentine adaptation. *Front. Plant Sci.* **11**, 574616 (2020).
37. Holsinger, K. E. & Weir, B. S. Genetics in geographically structured populations: defining, estimating and interpreting FST. *Nat. Rev. Genet.* **10**, 639–650 (2009).
38. Takuno, S. et al. Independent molecular basis of convergent highland adaptation in maize. *Genetics* **200**, 1297–1312 (2015).
39. Lai, Y. T. et al. Standing genetic variation as the predominant source for adaptation of a songbird. *Proc. Natl Acad. Sci. USA* **116**, 2152–2157 (2019).
40. Bohutínská, M. et al. Genomic basis of parallel adaptation varies with divergence in Arabidopsis and its relatives. *Proc. Natl Acad. Sci. USA* **118**, e2022713118 (2021).
41. Caye, K., Jumentier, B., Lepeule, J. & François, O. LFMM 2: fast and accurate inference of gene-environment associations in genome-wide studies. *Mol. Biol. Evol.* **36**, 852–860 (2019).
42. Remans, T. et al. A central role for the nitrate transporter NRT2.1 in the integrated morphological and physiological responses of the root system to nitrogen limitation in Arabidopsis. *Plant Physiol.* **140**, 909–921 (2006).
43. Little, D. Y. et al. The putative high-affinity nitrate transporter NRT2.1 represses lateral root initiation in response to nutritional cues. *Proc. Natl Acad. Sci. USA* **102**, 13693–13698 (2005).
44. Liu, J. et al. Targeted degradation of the cyclin-dependent kinase inhibitor ICK4/KRP6 by RING-type E3 ligases is essential for mitotic cell cycle progression during Arabidopsis gametogenesis. *Plant Cell* **20**, 1538–1554 (2008).
45. Stone, S. L. et al. Functional analysis of the RING-type ubiquitin ligase family of Arabidopsis. *Plant Physiol.* **137**, 13–30 (2005).
46. Choi, W. G., Toyota, M., Kim, S. H., Hilleary, R. & Gilroy, S. Salt stress-induced Ca^{2+} waves are associated with rapid, long-distance root-to-shoot signaling in plants. *Proc. Natl Acad. Sci. USA* **111**, 6497–6502 (2014).
47. Turner, T. L., Bourne, E. C., Von Wettberg, E. J., Hu, T. T. & Nuzhdin, S. V. Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. *Nat. Genet.* **42**, 260–263 (2010).
48. Sobczyk, M. K., Smith, J. A. C., Pollard, A. J. & Filatov, D. A. Evolution of nickel hyperaccumulation and serpentine adaptation in the *Alyssum serpyllifolium* species complex. *Heredity* **118**, 31–41 (2017).
49. Selby, J. P. *The Genetic Basis of Local Adaptation to Serpentine Soils in Mimulus guttatus*. Doctoral dissertation, Duke University (2014).
50. Rogivue, A. et al. Genome-wide variation in nucleotides and retrotransposons in alpine populations of *Arabis alpina* (Brassicaceae). *Mol. Ecol. Resour.* **19**, 773–787 (2019).
51. Wos, G., Choudhury, R. R., Kolář, F. & Parisod, C. Transcriptional activity of transposable elements along an elevational gradient in *Arabidopsis arenosa*. *Mob. DNA* **12**, 1–13 (2021).
52. Grandbastien, M.-A. et al. Stress activation and genomic impact of Tnt1 retrotransposons in Solanaceae. *Cytogenet. Genome Res.* **110**, 229–241 (2005).
53. Stuart, T. et al. Population scale mapping of transposable element diversity reveals links to gene regulation and epigenomic variation. *Elife* **5**, 1–27 (2016).
54. Lee, K. M. & Coop, G. Distinguishing among modes of convergent adaptation using population genomic data. *Genetics* **207**, 1591–1619 (2017).
55. Kintzer, A. F. & Stroud, R. M. Structure, inhibition and regulation of two-pore channel TPC1 from *Arabidopsis thaliana*. *Nature* **531**, 258–262 (2016).
56. Guo, J., Zeng, W. & Jiang, Y. Tuning the ion selectivity of two-pore channels. *Proc. Natl Acad. Sci. USA* **114**, 1009–1014 (2017).
57. Guo, J. et al. Structure of the voltage-gated two-pore channel TPC1 from *Arabidopsis thaliana*. *Nature* **531**, 196–201 (2016).
58. Kintzer, A. F. et al. Structural basis for activation of voltage sensor domains in an ion channel TPC1. *Proc. Natl Acad. Sci. USA* **115**, 9095–9104 (2018).
59. Griswold, C. K. & Williamson, M. W. A two-locus model of selection in autotetraploids: chromosomal gametic disequilibrium and selection for an adaptive epistatic gene combination. *Heredity* **119**, 314–327 (2017).
60. Mostafaee, N. & Griswold, C. K. Two-locus local adaptation by additive or epistatic gene combinations in autotetraploids versus diploids. *J. Hered.* **110**, 866–879 (2019).
61. Burgess, K. S., Etterson, J. R. & Galloway, L. F. Artificial selection shifts flowering phenology and other correlated traits in an autotetraploid herb. *Heredity* **99**, 641–648 (2007).
62. Morales, H. E. et al. Genomic architecture of parallel ecological divergence: beyond a single environmental contrast. *Sci. Adv.* **5**, eaav9963 (2019).
63. Ravinet, M. et al. Shared and nonshared genomic divergence in parallel ecotypes of *Littorina saxatilis* at a local scale. *Mol. Ecol.* **25**, 287–305 (2016).
64. Jones, F. C. et al. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484**, 55–61 (2012).
65. Colosimo, P. F. et al. Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science* **307**, 1928–1933 (2005).
66. Pardo-Diaz, C. et al. Adaptive introgression across species boundaries in *Heliconius* butterflies. *PLoS Genet.* **8**, e1002752 (2012).
67. Van Etten, M., Lee, K. M., Chang, S. M. & Baucom, R. S. Parallel and nonparallel genomic responses contribute to herbicide resistance in *Ipomoea purpurea*, a common agricultural weed. *PLoS Genet.* **16**, e1008593 (2020).
68. Ji, Y. et al. Gene reuse facilitates rapid radiation and independent adaptation to diverse habitats in the Asian honeybee. *Sci. Adv.* **6**, eabd3590 (2020).
69. Zong, S.-B., Li, Y.-L. & Liu, J.-X. Genomic architecture of rapid parallel adaptation to fresh water in a wild fish. *Mol. Biol. Evol.* **38**, 1317–1329 (2020).
70. Baduel, P., Bray, S., Vallejo-Marin, M., Kolář, F. & Yant, L. The 'Polyploid Hop': shifting challenges and opportunities over the evolutionary lifespan of genome duplications. *Front. Ecol. Evol.* **6**, 1–19 (2018).
71. Oziolor, E. M. et al. Adaptive introgression enables evolutionary rescue from extreme environmental pollution. *Science* **364**, 455–457 (2019).
72. Reid, N. M. et al. The genomic landscape of rapid repeated evolutionary adaptation to toxic pollution in wild fish. *Science* **354**, 1305–1308 (2016).
73. Kreiner, J. M. et al. Multiple modes of convergent adaptation in the spread of glyphosate-resistant *Amaranthus tuberculatus*. *Proc. Natl Acad. Sci. USA* **116**, 21076–21084 (2019).
74. Przeworski, M., Coop, G. & Wall, J. D. The signature of positive selection on standing genetic variation. *Evolution* **59**, 2312 (2005).
75. Hermisson, J. & Pennings, P. S. Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics* **169**, 2335–2352 (2005).
76. Chan, Y. F. et al. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a pitxl enhancer. *Science* **327**, 302–305 (2010).
77. Tishkoff, S. A. et al. Convergent adaptation of human lactase persistence in Africa and Europe. *Nat. Genet.* **39**, 31–40 (2007).
78. Xie, K. T. et al. DNA fragility in the parallel evolution of pelvic reduction in stickleback fish. *Science* **84**, 81–84 (2019).
79. Justin, C. Über bemerkenswerte vorkommen ausgewählter pflanzensippen auf serpentinstandorten Österreichs, Sloweniens sowie der Tschechischen Republik. *Linzer Biol. Beiträge* **25**, 1033–1091 (1993).
80. Punz, W., Aigner, B., Sieghardt, H., Justin, C. & Zechmeister, H. G. Serpentinophyten im Burgenland. *Verhandlungen Zool. Ges. Österreich* **147**, 83–92 (2010).
81. Bodenhausen, N., Horton, M. W. & Bergelson, J. Bacterial communities associated with the leaves and the roots of *Arabidopsis thaliana*. *PLoS ONE* **8**, e56329 (2013).
82. Horton, M. W. et al. Genome-wide association study of *Arabidopsis thaliana* leaf microbial community. *Nat. Commun.* **5**, 1–7 (2014).
83. Qvit-Raz, N., Jurkevitch, E. & Belkin, S. Drop-size soda lakes: transient microbial habitats on a salt-secreting desert tree. *Genetics* **178**, 1615–1622 (2008).
84. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
85. Hu, T. T. et al. The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. *Nat. Genet.* **43**, 476–481 (2011).
86. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

87. Mckenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
88. Hollister, J. D. et al. Genetic adaptation associated with genome-doubling in autotetraploid *Arabidopsis arenosa*. *PLoS Genet.* **8**, e1003093 (2012).
89. Keightley, P. D. & Jackson, B. C. Inferring the probability of the derived vs. the ancestral allelic state at a polymorphic site. *Genetics* **209**, 897–906 (2018).
90. Raj, A., Stephens, M. & Pritchard, J. K. FastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics* **197**, 573–589 (2014).
91. Stift, M., Kolář, F. & Meirmans, P. G. Structure is more robust than other clustering methods in simulated mixed-ploidy populations. *Heredity* **123**, 429–441 (2019).
92. Pickrell, J. K. & Pritchard, J. K. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* **8**, e100296 (2012).
93. Arnold, B., Kim, S. T. & Bomblies, K. Single geographic origin of a widespread autotetraploid *Arabidopsis arenosa* lineage followed by interploidy admixture. *Mol. Biol. Evol.* **32**, 1382–1395 (2015).
94. Excoffier, L. & Foll, M. fastsimcoal: A continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics* **27**, 1332–1334 (2011).
95. Weir, B. S. & Cockerham, C. C. Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358–1370 (1984).
96. Beissinger, T. M., Rosa, G. J., Kaeppler, S. M., Gianola, D. & De Leon, N. Defining window-boundaries for genomic analyses using smoothing spline techniques. *Genet. Sel. Evol.* **47**, 1–9 (2015).
97. Rawat, V. et al. Improving the annotation of *Arabidopsis lyrata* using RNA-Seq data. *PLoS ONE* **10**, 1–12 (2015).
98. Wang, M., Zhao, Y. & Zhang, B. Efficient test and visualization of multi-set intersections. *Sci. Rep.* **5**, 1–12 (2015).
99. Hämälä, T. & Savolainen, O. Genomic patterns of local adaptation under gene flow in *Arabidopsis lyrata*. *Mol. Biol. Evol.* **36**, 2557–2571 (2019).
100. Storey, J., Bass, A., Dabney, A. & Robinson, D. qvalue: Q-value estimation for false discovery rate control. R package version 2.20.0 (2020).
101. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* **6**, 80–92 (2012).
102. Legrand, S. et al. Differential retention of transposable element-derived sequences in outcrossing Arabidopsis genomes. *Mob. DNA* **10**, 1–17 (2019).
103. Hollister, J. D. et al. Transposable elements and small RNAs contribute to gene expression divergence between *Arabidopsis thaliana* and *Arabidopsis lyrata*. *Proc. Natl Acad. Sci. USA* **108**, 2322–2327 (2011).
104. Alexa, A. Rahnenfuhrer, J. topGO: Enrichment Analysis for Gene Ontology. R package version 2.44.0 (2021).
105. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* **4**, 1184 (2009).
106. Grossmann, S., Bauer, S., Robinson, P. N. & Vingron, M. Improved detection of overrepresentation of Gene-Ontology annotations with parent-child analysis. *Bioinformatics* **23**, 3024–3031 (2007).
107. Novikova, P. Y. et al. Sequencing of the genus Arabidopsis identifies a complex history of nonbifurcating speciation and abundant trans-specific polymorphism. *Nat. Genet.* **48**, 1077–1082 (2016).
108. Hämälä, T., Mattila, T. M., Leinonen, P. H., Kuittinen, H. & Savolainen, O. Role of seed germination in adaptation and reproductive isolation in *Arabidopsis lyrata*. *Mol. Ecol.* **26**, 3484–3496 (2017).
109. Mattila, T. M., Tyrmi, J., Pyhäjärvi, T. & Savolainen, O. Genome-wide analysis of colonization history and concomitant selection in *Arabidopsis lyrata*. *Mol. Biol. Evol.* **34**, 2665–2677 (2017).
110. Guggisberg, A. et al. The genomic basis of adaptation to calcareous and siliceous soils in *Arabidopsis lyrata*. *Mol. Ecol.* **27**, 5088–5103 (2018).
111. Hämälä, T., Mattila, T. M. & Savolainen, O. Local adaptation and ecological differentiation under selection, migration, and drift in *Arabidopsis lyrata*. *Evolution* **72**, 1373–1386 (2018).
112. Marburger, S. et al. Interspecific introgression mediates adaptation to whole genome duplication. *Nat. Commun.* **10**, 1–11 (2019).
113. Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
114. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2-a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).
115. Šali, A., Potterton, L., Yuan, F., van Vlijmen, H. & Karplus, M. Evaluation of comparative protein modeling by MODELLER. *Proteins Struct. Funct. Bioinformatics* **23**, 318–326 (1995).

Acknowledgements

This work was supported by the Czech Science Foundation (project 20-22783S to F.K.), Charles University (project Primus/SCI/35 to F.K. and Charles University Grant Agency grant No. 410120 to V.K.), and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme [grant number ERC-StG 679056 HOTSPOT to L.Y. and ERC-StG 850852 DOUBLE ADAPT to F.K.]. Additional support was provided by the long-term research development project No. RVO 67985939 of the Czech Academy of Sciences. Computational resources were provided by the CESNET LM2015042 and the CERIT Scientific Cloud LM2015085. The authors thank Lenka Flašková, Gabriela Šrámková, Anna Krejčová, and Mellieha Allen for help with laboratory work and result interpretation.

Author contributions

F.K., L.Y., and V.K. conceived the study. V.K., A.B.-G., L.Y., F.K., and M.B. performed field collections. A.B.-G., P.F., and V.K. did laboratory work. V.K., S.B., J.V., R.R.C., and M.B. performed analyses. V.K. and D.P. performed experiments. V.K., L.Y., and F.K. wrote the manuscript with input from all authors. All authors approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-021-25256-5.

Correspondence and requests for materials should be addressed to L.Y. or F.K.

Peer review information *Nature Communications* thanks John Willis and other anonymous reviewers for their contributions to the peer review of this work. Peer review reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Case study 5

Genomic basis and phenotypic manifestation of (non-)parallel serpentine adaptation in *Arabidopsis arenosa*

Genomic basis and phenotypic manifestation of (non-)parallel serpentine adaptation in *Arabidopsis arenosa*

Veronika Konečná, Marek Šustr, Doubravka Požárová, Martin Čertner, Anna Krejčová, Edita Tylová and Filip Kolář

ABSTRACT

Parallel evolution is common in nature and provides one of the most compelling examples of rapid environmental adaptation. In contrast to the recent burst of studies addressing genomic basis of parallel evolution, integrative studies linking genomic and phenotypic parallelism are scarce. Edaphic islands of toxic serpentine soils provide ideal systems for studying rapid parallel adaptation in plants. Serpentines are well-defined by a set of chemical selective factors and recent divergence between serpentine and geographically proximal non-serpentine populations allow mitigating confounding effects of drift. We leveraged threefold independent serpentine adaptation of *Arabidopsis arenosa* and used reciprocal transplant, ionomics, and available genome-wide polymorphisms to test if parallelism is manifested to a similar extent at both genomic and phenotypic levels. We found the varying magnitude of fitness differences, that was congruent with neutral genetic differentiation between populations, and limited costs of serpentine adaptation. Further, phenotypic parallelism in functional traits was pervasive. The genomic parallelism at the gene level was significant, although relatively minor. Therefore, the similarly modified phenotypes e.g. of ion uptake arose possibly by selection on different loci in similar functional pathways. In summary, we bring evidence for the important role of genetic redundancy in rapid adaptation involving traits with polygenic architecture.

INTRODUCTION

Adaptation is a key evolutionary mechanism for how organisms can cope with changing environments. Identifying adaptation, however, is challenging at both phenotypic and genetic levels as multiple confounding signals such as evolutionary constraints, demographic or stochastic genetic processes may produce patterns resembling adaptation. Multiple populations facing the same environmental challenge provide strong evidence for repeated adaptation within a species. If the same phenotypes independently emerge in similar habitats, it is more probable that these features evolved under natural selection than solely due to the stochastic forces (Lenormand et al., 2009). The same adaptive phenotypes can have a divergent genetic basis when e.g. different genes within a pathway are under selection in different independent populations (e.g. Fang et al., 2021) or a similar genetic basis when a source for adaptation is a shared pool of standing variation or pleiotropic constraint (Hämälä and Savolainen, 2019). Despite numerous examples of parallel adaptation, however, we have still a limited understanding of the genetic basis of parallel phenotypic changes and their interplay with the local environmental conditions.

The strength of correlation between environmental factors and particular phenotypic traits, which are under the selection in multiple adapted populations, provide the system for quantification of the extent of parallelism. However, independently evolved 'adaptive' phenotypes are rarely identical, representing rather a continuum of non-parallel evolution (Stuart et al., 2017; Bolnick et al., 2018), in response to the complex interplay of the genetic basis of a trait, population history, and of local environmental heterogeneity (Rosenblum et al., 2014; Fraïsse and Welch, 2019; James, Wilkinson, et al., 2021). Indeed, the varying extent of

phenotypic and genomic parallelism has been recently identified in many iconic examples of parallel evolution animals such as – threespine sticklebacks (Stuart et al., 2017; Magalhaes et al., 2021), salmonid fishes (Jacobs et al., 2020), guppies (Whiting et al., 2021), cichlids (Weber et al., 2021), marine snails (Morales et al., 2019), and songbirds (Salmón et al., 2021), but rarely investigated in plants (but see Knotek et al., 2020; James et al., 2021). For instance, Bohutínská et al. (2021) showed a divergence-dependent level of genomic parallelism in Brassicaceae i.e. with the increasing divergence between lineages the degree of gene reuse was decreasing. This is in congruence with findings in *Senecio lautus* (James, Arenas-Castro, et al., 2021; James, Wilkinson, et al., 2021), in which similar phenotypes with large divergence times within-population pairs have arisen via mutational changes in different genes, although many of these genes shared the same biological functions. However, even in the system of recent post-glacial origin such as alpine populations of *Heliosperma pusillum* (Trucchi et al., 2017), they also achieved similar phenotypes via selection of different genes in the same functional pathways. In such polygenic systems we expect genetic redundancy, i.e. mutations in different genes (possibly within the same pathway) can lead to similar phenotypes and thus to phenotypic parallelism (Hermisson and Pennings, 2017; Höllinger et al., 2019; Barghi et al., 2020). In systems with high heterogeneity among the compared 'parallel' habitat types such as alpine stands, which may strikingly differ by a range of local parameters such as substrate and exposition, the extent of genomic and phenotypic parallelism may be underestimated. We thus need empirical evidence from other systems encompassing recent polygenic adaptation to more narrowly defined environmental challenges to explore if genetic redundancy can lead to phenotypic parallelism.

Naturally toxic edaphic substrates, serpentines, represent one of the most challenging extreme environments for plants. Unlike alpine or coastal environments, selective conditions of serpentines can be easily manipulated in the experimental conditions as specific soil chemistry is the principal defining selective factor of serpentine habitats. Serpentines are defined by a highly skewed Ca/Mg ratio, high concentration of Mg, and elevated heavy metals such as Cr, Co, and Ni (Brady et al., 2005; O'Dell and Rajakaruna, 2011). Although occurring worldwide, serpentine outcrops are typically scattered as small edaphic 'islands', being surrounded by other less toxic substrates. This setup triggers parallel evolution via repeated colonization from the surrounding non-toxic substrates, which has been recently shown in the case of an autotetraploid *Arabidopsis arenosa* (Konečná et al., 2021). In its genetically highly diverse autotetraploid cytotype, low neutral genetic differentiation and recent split times between geographically proximal serpentine and non-serpentine populations indicate recent postglacial serpentine invasions that happened at least five times independently (Konečná et al., 2021). *Arabidopsis arenosa* thus represents an ideal plant model for studying the repeatability of adaptation towards a well-defined selective environment. Yet, while finding significant parallelism at the genomic level, the extent of phenotypic parallelism across relevant fitness and physiological traits and the question of whether both levels correlate remained unknown.

Here, we used reciprocal transplants and asked to what extent the phenotypic evolution is repeatable during rapid parallel edaphic adaptation across three serpentine – non-serpentine *A. arenosa* population pairs. We hypothesize that in this recently postglacially diversified system with pervasive sharing of adaptive polymorphism (Konečná et al., 2021), genetic and phenotypic parallelism will be largely congruent. We quantified (i) genomic parallelism at genes (shared selection candidate loci) and functional pathways (shared gene ontology categories) and (ii) parallel phenotypic changes by scoring a diverse set of traits: fitness proxies, morphological functional traits and ion uptake. We specifically asked: (i) Do serpentine populations exhibit fitness differences in a direction corresponding to local substrate adaptation? If so, does the direction and magnitude of fitness response differ among population pairs in congruence with the level of neutral genetic differentiation? (ii) Is there a cost of

serpentine adaptation (trade-off) and how its extent varies over the three parallel population pairs? (iii) Does the extent of phenotypic parallelism correspond to gene-level parallelism, or does the genetic redundancy lead to limited gene parallelism?

Local adaptation has been manifested in a congruent direction across all three parallel serpentine populations as the advantage of serpentine populations in their native soils and with utmost limited trade-offs. However, the magnitude of the adaptive response differed, following the level of genetic differentiation within population pairs. Further, we found pervasive phenotypic parallelism in functional traits with only a minority of traits showing non-parallel variation. Contrary to our initial hypothesis, gene-level parallelism was rare, especially when compared to high similarity in functional pathways. Likely, similar phenotypes arise mainly via selection of different genes in similar genetic pathways showing the role of genetic redundancy and stochasticity in rapid adaptation with polygenic basis.

MATERIALS AND METHODS

Study species and environment

Arabidopsis arenosa is an obligate outcrosser that is widespread on semi-open rocky habitats characterised by reduced competition on various substrates, predominantly on calcareous and siliceous non-serpentine soils across most of Europe (Schmickl et al., 2012). For this study we collected seeds from three autotetraploid serpentine (S1, S2, S3) and three geographically proximal (<19 km distant) sister non-serpentine siliceous populations in Central Europe (N1, N2, N3; STable 1; Konečná et al., 2021). The differences in elevation within the pairs were small for S1-N1 - 71 m and S2-N2 - 22 m and higher for S3-N3 - 322 m, and the variation in local vegetation structure (open mixed/coniferous forests with rocky outcrops) and climatic variables among individual sites was negligible (Konečná et al., 2021). In contrast, we observed strong differences in soil chemistry, mainly driven by high level of heavy metals (Co, Cr, and Ni), high concentration of Mg, and low Ca/Mg ratio consistently differentiating serpentine substrates occupied by the studied populations from their paired adjacent non-serpentine counterparts (see Konečná et al., 2021 for details) (also quantified in our set of populations, SFig. 1). Thus, we consider substrate adaptations as a primary source of selective divergence between replicated serpentine and non-serpentine populations and use different soils as experimental treatments in the analysis of serpentine adaptation.

Reciprocal transplant experiment/ Plant cultivation

To test for local substrate adaptation in the three serpentine – non-serpentine population pairs, we compared plant fitness in the native versus foreign soil in a reciprocal transplant experiment. We reciprocally transplanted plants of serpentine and non-serpentine origin from three population pairs (S1–N1, S2–N2, S3–N3) that served as representatives of independent serpentine colonisations (Konečná et al., 2021). For each population pair, we transplanted young seedlings into serpentine and non-serpentine soils from their original sites (i.e., S1 plant cultivated in S1 and N1 soil and vice versa) in a common garden (Faculty of Science Charles University, 200 m; methodological details are provided in Supplementary Methods) and tested for the interaction between the soil treatment and substrate of origin in selected fitness indicators (described further) as a proxy of local substrate adaptation. We observed initial differences already in germination and young rosette sizes (Konečná et al., 2021), and thus continued with the cultivation for the entire growing season until seed set.

Trait scoring

As the extent of the phenotypic parallelism is also dependent on the choice of measured traits, we scored a broad range of potentially relevant traits. Morphological traits directly linked to fitness (reproductive outputs) are the most obvious aims, in contrast, physiological or early developmental traits are only rarely scored due to technical limitations. To comprehensively explore the broad range of phenotypic variation we scored a diversity of functional traits. Specifically, we scored bolting and flowering time, rosette area, number of additive rosettes as morphological life-history traits and fruit production, total seed mass production, above ground and root biomass and survival as fitness proxies, ion uptake of five important elements characterizing serpentine substrate (Co, Cr, Ni, Ca, Mg), and functional traits of root growth and architecture in seedlings - total root growth, main root length, and density of lateral roots - in order to cover both early and later stages of plant development and both macroscopic and physiological properties. We included transitions to bolting, flowering, fruit production, and survival to cover all life stages.

We cultivated plants for 10 months (January – October 2019). We scored bolting time when the first flowering buds appeared on elongated stems (cca at least 1 cm tall) and flowering time as opening of the first flowers every three/four days. Further, we have measured the rosette area (which correlated with rosette diameter, see Konečná et al., 2021) in time of the maximum rosette diameter (SFig 6 in Konečná et al., 2021) in four months since germination. We have been collecting fruits regularly once/twice a week since mid-June (five months since germination), additionally, we collected fruits in two time points, when most plants produced matured fruits. We excluded fruits with less than five seeds, as those were possibly not well developed (mean number of seeds/fruit: 73). We provide the summary of flowering/fruit proportion of plants in STable 2. At the end of the experiment, we counted and weighted seeds per each harvested fruit (1-22 fruits/plant). We used this value of seed mass/fruit to calculate average seed mass/population/treatment, the total seed mass production was then calculated by multiplying the total number of fruits by average seed mass/population/treatment. To account for possible seasonal variability, we tested if seeds produced until the end of July varied in weight from those produced later in the summer and the autumn, but we did not find a significant difference. Next, we counted the total number of additive rosettes (plants produced additive sister rosettes) in early September, and we weighted the above-ground biomass (from all plants which survived) and root biomass (we randomly chose five plants/population/treatment due to challenging and time-consuming sample preparation) after the harvest at the end of October. Finally, we scored the survival of plants continuously and at the end of the experiment.

We quantified the concentration of the key serpentine-distinctive elements (Ca, Co, Cr, Mg, and Ni) in randomly selected 10 individuals from each serpentine and non-serpentine population cultivated in serpentine treatments (S1, S2, and S3). We collected leaves in mid-May (after 2.5 months of cultivation in serpentine soils) and dried and decomposed samples prior to the analysis (see Supplementary Methods for details). To identify potential contamination of leaf samples we calculated the interquartile range (difference between the 75th percentile and the 25th percentile) for all values of heavy metals (Co, Cr, and Ni) and removed all values (four samples, 6 % of all samples, in total) which had 1.5 times the interquartile range greater than the third quartile.

Root growth experiments in calcium/magnesium solutions

To analyse response to altered Ca/Mg ratio in the environments we conducted root growth experiments. We cultivated plants *in vitro* on agar-solidified media supplemented with various Ca/Mg ratios – 1.97 in control medium, 0.2 in medium simulating the mean Ca/Mg ratio in

natural serpentine populations involved in this study, and 0.04 in medium according to Bradshaw (2005). We measured the root traits (total root growth, main root length, and density of lateral roots) in plants 15 days since germination. Similarly to Berglund et al. (2004) we focused on root phenotypes, which directly mirror the stressful solutions, and they are visible even in the early stages of plant development. For cultivation details see Supplementary Methods.

Statistical analysis of trait variation

We assessed the variation in morphological life-history traits and fitness proxies by linear and generalized linear models (LM and GLM) and functional traits of root growth and architecture by linear mixed effect models (LMM). For each trait, we tested the effect of substrate of origin (S vs. N), population pair (1, 2, 3), treatment (S vs. N) and their interactions. Traits with deviation from Gaussian distribution were log or square root transformed to approach normality (indicated in Table 1, STable 6, STable 7). We used random effects of cultivation run and Petri dish for cultivation in LMM. No pair of measured traits was highly correlated ($R > 0.8$, SFig. 2). Tests of significance for individual fixed-effect factors, substrate of origin, treatment, population pair, as well as the substrate of origin*treatment interaction were conducted by Type III Wald $\chi 2$ tests with the R package car.

We used aster models (Geyer et al., 2007; Shaw et al., 2008) to analyze the hierarchical composite fitness differences among populations separately within each soil treatment. This approach is suitable for combining multiple fitness traits with different distributions into a single composite fitness variable. The hierarchical structure of the model, however, did not allow to incorporate final harvested above-ground biomass (plants, which survived, but did not reproduce). For each individual, we used the following hierarchy: 1 -> bolting (Bernoulli distribution, 0/1) -> flowering (Bernoulli distribution, 0/1) -> fruit production (Bernoulli distribution, 0/1) -> total seed mass production (Poisson distribution). By comparing the nested models with the effect of substrate of origin, population pair, and their interaction using likelihood ratio tests, we found the model with the highest likelihood (substrate of origin*population pair) (STable 3). Using likelihood ratio tests, we also tested the differences between populations within population pairs in order to assess if populations significantly differ in composite hierarchical fitness.

The aster model does not allow to incorporate possibly an important indicator of success in next season – above-ground biomass of plants, which did not produce fruits. Therefore, to further quantify fitness response using all five scored fitness proxies, we multiplied flowering proportion and early survival, fruit production, total seed mass production, late survival after fruit production, and above-ground biomass for both serpentine and non-serpentine plants in their native and foreign soil. The values of total seed mass production and above-ground biomass were standardized by the average of the serpentine – non-serpentine plants control values from non-serpentine treatment for each pair separately to account for population-pair variation. Besides the phenological and reproductive traits, we also included late survival and above-ground biomass, proxies for the plant's success in the following year. Then, we used cumulative fitness to quantify the overall fitness trade-offs (cost of adaptation; following Hereford, 2009). We calculated the difference in relative fitness between a pair of serpentine and non-serpentine populations at each population's native soil. These estimates were standardized by relative mean fitness at each soil. Further, we evaluated the relative magnitudes of local adaptations to serpentine soils by comparing serpentine and non-serpentine plants within each population pair cultivated in serpentine soil (following Hereford, 2009). The magnitudes of adaptations were expressed in terms of specific differences between the substrate of origins. We calculated these values based on overall cumulative fitness estimates. The

difference (native – foreign) represents the total increase in fitness due to local adaptation (if the values are positive).

We tested the differences in functional traits (morphological life-history traits and fitness proxies) and leaf elemental concentrations (ionomic phenotypes) between the group of serpentine and non-serpentine populations cultivated in the stress environment (serpentine soil) using LM with substrate of origin, population pair and their interaction as fixed effects. To assess the degree of parallelism and non-parallelism more quantitatively, effect sizes of each factor and their interaction were estimated from the output of LM models (trait ~ substrate of origin + population pair + substrate of origin*population pair) using a partial eta-squared method (anova_stats function) in R sjstats package. Larger effect of the substrate of origin than the effect of the substrate of origin*pop. pair interaction was an indication of parallelism while the opposite implied non-parallel serpentine/non-serpentine differentiation in that particular trait (Stuart et al., 2017).

Leaf and soil elemental concentration analyses

We used the soil elemental composition data from natural populations from Konečná et al. (2021) to identify major drivers of substrate adaptation in three serpentine populations. We further quantified the elemental composition of soils used in the reciprocal experiment for direct associations with leaf elemental concentrations using inductively coupled plasma mass spectrometry (ICP-OES; spectrometer INTEGRA 6000, GBC, Dandenong Australia). We monitored five elements (Ca, Co, Cr, Mg, and Ni) in both soil and leaf extract samples. Soil samples, which were collected from five pots/population/treatment at the same time as leaf samples in mid-May, were mixed, sieved and dried at 60 °C. Both soil and leaf samples were prepared according to a protocol summarised in Supplementary Methods.

Quantification of gene parallelism

We used already published individual resequencing data for serpentine (S1, S2, and S3) and non-serpentine (N1, N2, and N3) populations (7-8 individuals per population, 47 in total) from the study of Konečná et al. (2021). Raw data are available as a BioProject PRJNA667586 and PRJNA325082. To quantify the extent of gene parallelism we leveraged the threefold-replicated natural setup and already identified candidate genes that show repeated footprints of selection across three events of serpentine colonization in Konečná et al. (2021). Briefly, differentiation candidates were identified as gene-coding loci exhibiting excessive differentiation between serpentine – non-serpentine populations within each population pair using 1 % outlier F_{ST} window-based scans. Further, we identified parallel differentiation candidates as overlapping genes in differentiation candidate lists across at least two population pairs. We tested if such overlap is higher than a random number of overlapping items given the sample size using Fisher's exact test in SuperExactTest in R package (Wang et al., 2015). Further, we calculated the proportion of shared genes as number of shared differentiation candidate genes per each two population pairs divided by the total number of differentiation candidate genes for selected population pair combination.

Gene Ontology (GO) enrichment analysis and quantification of functional parallelism

To infer function-level parallelism, we annotated differentiation candidates into biological processes, molecular pathways, and cellular components using gene ontology (GO) enrichment analysis for each population pair separately. We applied Fisher's exact test with the classic algorithm implemented in the topGO R package (Alexa and Rahnenfuhrer, 2020). We worked with *A. thaliana* orthologs of *A. lyrata* genes obtained using biomaRt and *A. thaliana* was also

used as the background gene universe in all gene set enrichment analyses. With the "classic" algorithm each GO term is tested independently and it is not taking the GO hierarchy into account. Therefore, it is more suitable for comparisons across multiple lists of candidates to get overall quantification of parallelism. We accepted only significant GO terms ($p < 0.05$, Fisher's exact test). We calculated the proportion of shared functions as the number of shared GO terms per each two population pairs divided by the total number of GO terms for selected population pair combination. In order to test if functional parallelism in GO terms is not driven solely by parallel differentiation candidate genes, we ran the same analyses also without these candidates.

Functional protein association networks in STRING

To search for functional associations among identified differentiation candidates from three population pairs, we used STRING v11 database (Szklarczyk et al., 2015) of protein-protein association networks. We used 'multiple proteins' search in *Arabidopsis thaliana* and these active interaction sources: co-expression, co-occurrence, databases, experiments, gene fusion, neighborhood, and textmining. We required the minimum interaction score of medium confidence (0.4) and retained only 1st shell associations.

RESULTS

Varying magnitude and direction of parallel serpentine adaptation in Arabidopsis arenosa

In line with the hypothesis of parallel adaptation, originally serpentine populations grown in their native soils exhibited higher values in most of the fitness-related functional traits. Significant interaction of the substrate of origin with the soil treatment ($p < 0.05$, Type III Wald $\chi 2$ tests, Table 1, SFig. 3) was observed for the rosette area, number of additive rosettes, total seed mass production, above-ground and root biomass, and total root growth. To further quantify overall fitness differences, we further combined the joint contribution of four reproductive fitness traits in hierarchical aster models, which explicitly test the dependence of particular reproductive fitness traits on the others together with the transition among them in each treatment (Fig. 1A). We observed significantly higher composite fitness for originally serpentine populations S1 and S3 grown in their native serpentine substrate as compared to their paired non-serpentine populations (Fig. 1A; 45.63 and 56.42, respectively; STable 4). The difference in population pair 2 was negligible, reflecting the overall very low survival of both populations (62 %) in serpentine soil 2 (STable 2).

Table 1. Effects of substrate of origin (S vs. N)/ population pair (1, 2, 3)/ treatment (S vs. N soil)/ and their interactions on functional traits were tested by linear models (LM), binomial generalized linear models (GLM binomial), and linear mixed effect models (LMM). Tests of significance for individual fixed effect factors and interaction were conducted by Type III Wald $\chi 2$ tests. The random effects in LMM were cultivation batch and Petri dish for cultivation. The last three traits were assessed in a separate experiment using growth media (treatment here is represented by different Ca/Mg concentrations). *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; · $p < 0.1$

Response variable	**Transfor-mation**	**Model**	**Substrate of origin**			**Pop. pair**			**Treatment**			**Substrate of origin*treatment**		
			Df	$\chi 2$	*p*	Df	$\chi 2$	*p*	Df	$\chi 2$	*p*	Df	$\chi 2$	*p*
Bolting time [days]	log	LM	1	2.0524	n	2	26.3506	***	1	6.8092	**	1	2.4737	ns
Flowering time [days]	log	LM	1	0.0088	ns	2	3.5704	*	1	5.5588	*	1	0.0618	ns
Rosette area [mm^2]	sqrt	LM	1	74.0269	***	2	4.0045	*	1	382.8176	***	1	51.43337	***
N of additive rosettes	log	LM	1	34.4439	***	2	9.9119	***	1	91.4885	***	1	14.2435	***
Probability of fruit production (0/1)		GLM binomial	1	10.2295	**	2	3.576	ns	1	14.0429	***	1	2.8821	.
Total seed mass production [mg]	log	LM	1	27.2423	***	2	6.3113	**	1	26.9319	***	1	12.6945	***
Above-ground biomass [mg]	sqrt	LM	1	9.6231	**	2	9.6957	***	1	104.7775	***	1	13.8185	***
Root biomass [mg]	log	LM	1	15.6522	***	2	16.9575	***	1	61.3215	***	1	14.2662	***
Probability of late survival (after fruit production) (0/1)		GLM binomial	1	6.5438	*	2	4.7977	.	1	8.0743	**	1	2.1933	ns
Total root growth [cm]	log	LMM	1	123.7103	***	2	14.7846	***	2	60.0113	***	2	44.4136	***

Main root length [cm]		LMM	1	16.4607	***	2	8.9317	*	2	24.9806	***	2	6.2092	*
Density of lateral roots (number of lateral roots/main root length)	sqrt	LMM	1	12.9772	***	2	17.5417	***	2	20.2030	***	2	8.0637	*

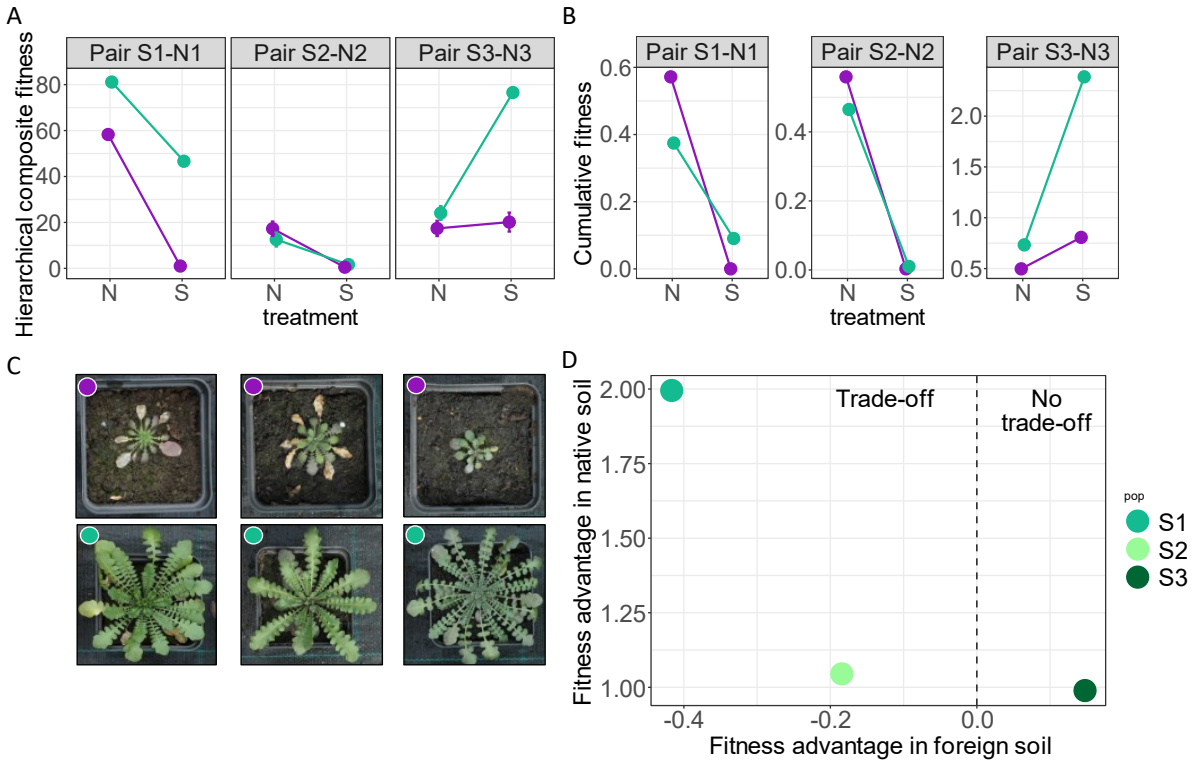


Fig. 1. Local adaptation to serpentine soils in three parallel population pairs of *A. arenosa*. A) Hierarchical composite fitness (+-SEs) for serpentine (green) and non-serpentine (violet) plants cultivated in local serpentine (S) and non-serpentine (N) soils. Estimates were taken from hierarchical aster models combining proportion of bolting plants, proportion of flowering plants, fruit production, and total seed mass. The significant ($p < 0.001$) interactions were found for the S1 - N1 and S3 - N3 population pairs (complete results in STable 4). B) Cumulative fitness calculated by combining five reproductive and vegetative traits (flower proportion and early survival, fruit production, total seed mass production, late survival after fruit production, and above-ground biomass). Note the y-axes have been adjusted to different ranges to aid visibility. C) Example photos illustrating parallel response to serpentine soils (dot colours denote the plant origin; from the left: population pair 1 – 2 – 3; photos taken by V. Konečná). D) Relative fitness advantage of the originally serpentine population in its native and foreign (i.e. non-serpentine) soils calculated based on cumulative fitness estimates. On the left side, both populations from the pair (as we are addressing serpentine adaptation only serpentine populations are visualized) had higher relative fitness in their native soils indicating fitness trade-off. On the right side, the serpentine population had relatively higher fitness than non-serpentine population in both soil treatments. Note: each fitness advantage estimate was standardized by the relative mean fitness of plants at each soil treatment.

Further, we calculated cumulative fitness (STable 5) based on vegetative and reproductive fitness proxies, to quantify differences in the magnitude of local adaptation among population pairs (Fig. 1B). In all cases, the magnitude of local adaptation was positive (Table 2), which indicates the advantage of serpentine populations in their native soil and selection against immigrants (Hereford, 2009). Yet, the magnitude of this response differs among population pairs, which is in line with varying relative fitness differences already observed in aster models. We found the lowest value for population pair 2 and highest for pair 3, which is fully in congruence with increasing genome-wide differentiation of population pairs 2 < 1 < 3 (Table 2). Overall, serpentine populations have higher fitness in serpentine soils than their originally non-serpentine counterparts, although its magnitude strongly varied, altogether providing strong evidence for substrate-driven local adaptation.

Table 2. Estimated magnitude of local adaptation in serpentine soils.

Population pair	**Genetic differentiation (F_{ST})[1]**	**Substrate environmental distance[2]**	**Magnitude of local adaptation — cumulative fitness[3]**	**Relative difference in hierarchical composite fitness[4]**
S1-N1	0.069	4.174	0.090	45.635
S2-N2	0.029	6.682	0.007	1.129
S3-N3	0.085	6.934	1.579	56.414

[1] inferred from genome-wide nearly-neutral fourfold degenerated single nucleotide polymorphisms (SNPs) by Konečná et al. (2021)
[2] environmental distances were calculated based on chemical analysis of 20 elements in in soils sampled at original locations
[3] cumulative fitness gathered by multiplication of flower production and early survival, fruit production, total seed mass production, late survival after fruit production, and above-ground biomass
[4] inferred by hierarchical aster model performed on proportion of bolting plants -> proportion of flowering plants -> fruit production -> total seed mass production

In line with the differences in the direction we also found variation in trade-offs. While population pairs 1 and 2 exhibited costs of serpentine adaptation in the non-serpentine soils (based on cumulative fitness values), a positive fitness advantage in the foreign non-serpentine soil was present in the S3 population (0.148), reflecting larger viability of S3 even in the non-serpentine soil (Fig. 1D). Yet, even in cases involving trade-offs, serpentine adaptation had a limited cost (1.52 average fitness advantage in native soil vs. -0.3 average fitness advantage in foreign soil, Fig. 1D). This is also in congruence with minor and nonsignificant differences in the survival of plants in non-serpentine treatment (89 % survival of originally serpentine plants vs. 87 % survival of originally non-serpentine plants).

Varying levels of phenotypic parallelism across functional traits

We further analyzed the magnitude of response to serpentine across distinct population pairs in the individual functional traits in order to identify traits exhibiting parallel responses. We identified traits exhibiting strong parallel response to the serpentine soil as those have a larger effect of the substrate of origin than the substrate of origin*population pair when grown in the selective serpentine soil (Fig. 2A, B). Regardless of the population pair, originally serpentine

plants bolted earlier, had larger rosette area, higher number of additive rosettes, higher seed mass, higher root biomass, and lower uptake of Ca, Co, Cr, Mg, and Ni than non-serpentine plants when cultivated in serpentine soil (STable 6, Fig. 2A, B, SFig. 4).

On the contrary, we found only a minority of traits that showed rather non-parallel variation i.e. a dominant effect of the substrate of origin*population pair interaction: above-ground biomass, flowering time, and Ca/Mg ratio in plants. The most remarkable difference was found for the Ca/Mg ratio, where S2 and S3 populations accumulated relatively higher Ca/Mg ratios than their paired N populations, while we observed a reverse trend in population pair 1 with lower (i.e. more skewed) values in S1 (Fig. 2C; as a consequence of the lower Ca uptake, Fig. S2). To test if such differences in Ca/Mg uptake also correspond to non-parallel response to altered Ca/Mg ratio in the environment, we cultivated plants of all populations *in vitro* on agar-solidified media with modified Ca/Mg ratio and scored root growth and architecture. We observed consistently better root growth of originally serpentine over non-serpentine plants in highly skewed Ca/Mg environments (Fig. 2D, E). In addition, lower Ca/Mg ratio affected root architecture: serpentine plants had longer main roots and a lower density of lateral roots. Interestingly, such results were consistent over the three population pairs (Fig. 2D), suggesting that the specific pattern of Ca and Mg uptake of S1 population in serpentine soil is not mirrored in distinct fitness response to skewed Ca/Mg concentrations *per se* and this population likely evolved a distinct mechanism how to cope with low Ca/Mg ratio in its above-ground tissues.

Altogether, in line with expectations, the recently diverged serpentine populations exhibited pervasive phenotypic parallelism in functional traits in early as well as in later developmental stages with only a minority of traits showing non-parallel variation.

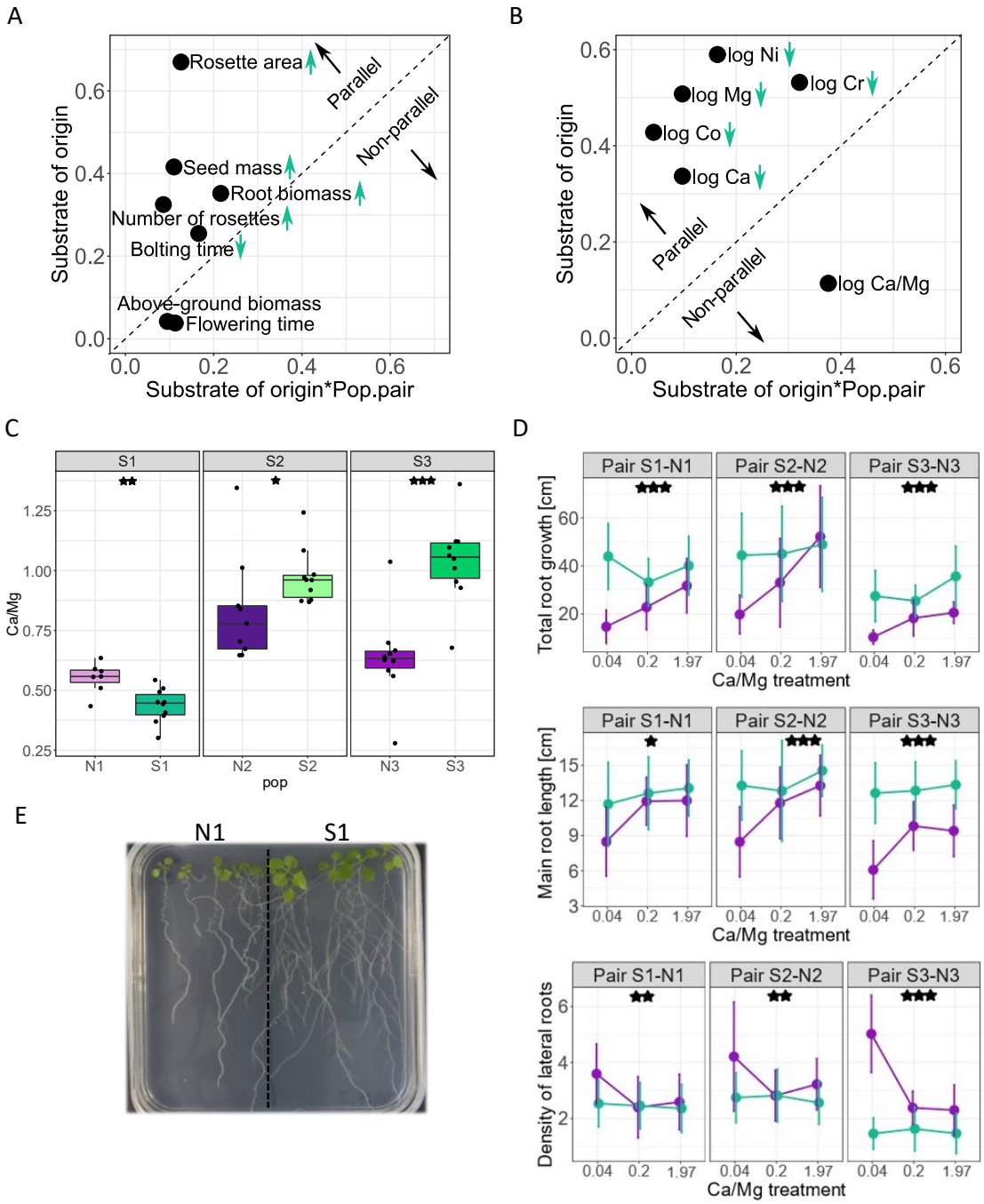


Fig. 2. Variation in the extent of parallelism among individual morphological life-history traits (A) and leaf elemental concentrations (B) scored on plants cultivated in the selective serpentine soils. The extent of parallelism was estimated as effect sizes (Eta2) in linear models addressing the effect of the substrate of origin (S vs. N), pop. pair (1, 2, 3), and their interaction, calculated separately for each trait (see STable 7 for input values). The effect size of the substrate of origin (y-axis) shows the extent to which a given trait diverges predictably between serpentine and non-serpentine plants, i.e., in parallel, while the substrate of origin*population pair effect size (x-axis) quantifies the extent to which serpentine/non-serpentine phenotypic divergence varies across population pairs (i.e., deviates from parallel). Points falling above the dashed line have a larger substrate of origin effect (i.e. parallel) than the substrate of origin*pop. pair interaction effect (i.e. non-parallel). Green arrows indicate the overall trend in trait value in populations of serpentine origin. C) Variation in the ratio of Ca/Mg content in the leaves of paired serpentine and non-serpentine populations cultivated in their corresponding serpentine soils. Note: asterisks denote the significance of the substrate of origin*treatment interaction (*** $p < 0.001$; ** $p < 0.01$; * $p <= 0.05$; STable6). D) Differences in root growth of the three population pairs after 15 days since germination; note: 1.97 is a control medium, green: plants of serpentine origin, violet: plants of non-serpentine origin. Asterisks denote the significance of the effect of substrate of origin*treatment interaction (*** $p < 0.001$; ** $p < 0.01$; * $p <= 0.05$). E) Illustrative photo from cultivation of N1 and S1 plant in 0.04 Ca/Mg treatment.

Genomic drivers of phenotypic parallelism

To uncover the genetic architecture underlying the observed parallel phenotypes, we also quantified and characterized parallelism at the level of genes and functions. First, we took differentiation candidate genes (1 % outliers in F_{ST} genetic differentiation between S and N populations; SData 1) identified by Konečná et al. (2021) for each pair and tested for gene-level parallelism. Although, we found a significant over-representation of shared genetic candidates across all combinations of population pairs (Fig. 3A; Fisher's exact test; $p < 0.05$; SData 1), the proportion of shared candidate genes was rather low, ranging from 2.16 % to 2.4 % (Table S8).

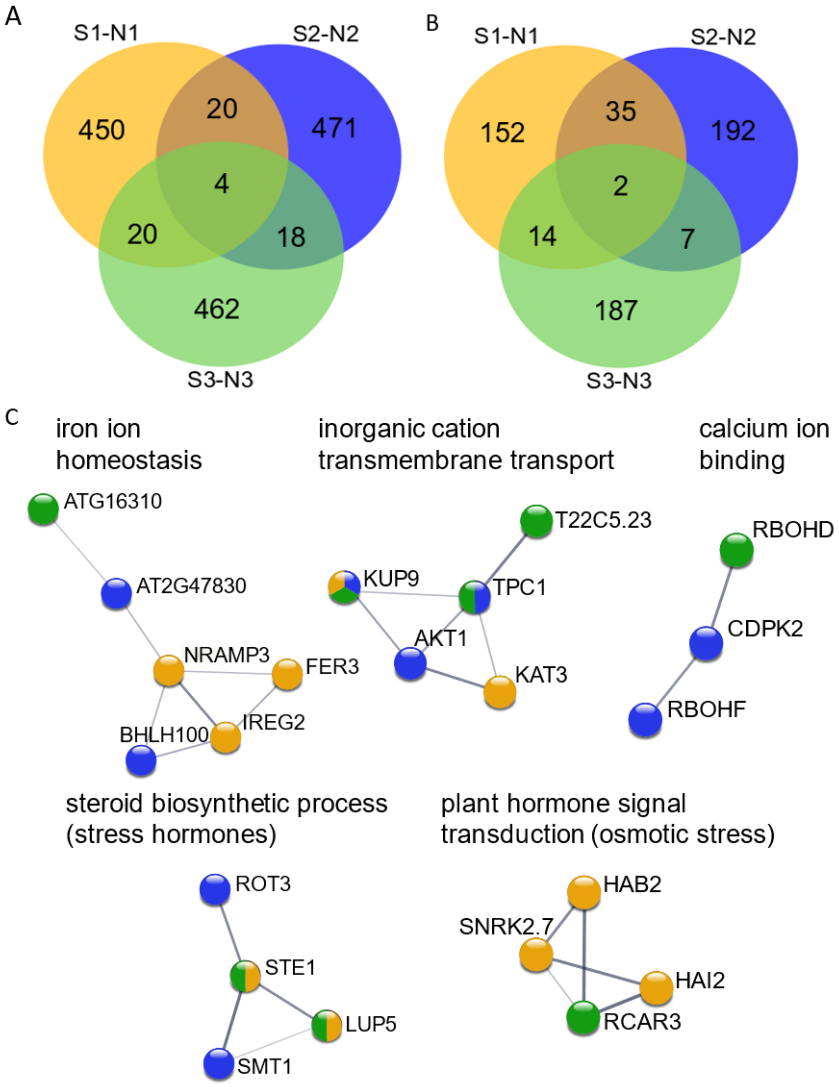


Fig. 3. Significant genomic parallelism among three investigated population pairs of *A. arenosa*. A) Intersection of differentiation candidate genes – parallelism by genes. B) Intersection of gene ontology terms in biological processes domain – parallelism by function. The number of the overlapping items was significantly higher than random across all intersections ($p < 0.05$; Fisher's exact test; SData 1 and SData 2). C) Parallelism by function – significantly enriched biological processes or KEGG pathways extracted by STRING database; we used medium and higher confidence associations – displayed by thickness of the lines (thick line is 0.4 and large line is 0.7 edge confidence). Note: colours are matching with part A) and B). Parallel differentiation candidates consist of multiple colours.

As the low gene-level parallelism might reflect genetic redundancy, we further associated all differentiation candidates with corresponding functional pathways (gene ontology – GO – enrichment analysis; SData 2) and quantified function-level parallelism (Fig. 3B) as the percentage of shared enriched GO terms between any two pairs of populations (STable 8). All overlaps were significant (Fisher's exact test; $p < 0.05$, SData 2) and the extent of pairwise overlap ranged from 2.02 % to 8.43 % for the most relevant category of biological process (BP; similar results were also achieved for molecular functions, and cellular components; STable 8). We also observed significant functional parallelism even after removing the parallel differentiation candidates for GO enrichment analysis (STable 8). Significantly enriched BP in all three population pairs were cellular process and metabolic

process. BP categories shared between two population pairs were more specific, reflecting multiple processes that are relevant for facing the toxic serpentine soils (Fig. 4): (i) cellular transport: transport (shared between S2-N2 and S3-N3), protein transport, protein localization, and organonitrogen compound metabolic process (all S1-N1 and S2-N2); (ii) abiotic stress: regulation of abscisic acid-activated signalling pathway, important in stress signalling (S1-N1 and S3-N3), response to endoplasmic reticulum stress (S1-N1 and S2-N2), and cellular response to stimulus (S1-N1 and S3-N3); and (iii) developmental processes: anatomical structure development (S1-N1 and S2-N2) and sexual sporulation (S1-N1 and S2-N2) (for complete lists see SData 2).

To further inspect function-level parallelism by a more specific pathway-oriented analyses, we inferred protein-protein functional networks from a combination of the differentiation candidates inferred for all three serpentine populations in STRING database. We found representative functional associations of distinct differentiation candidate genes inferred from different population pairs (Fig. 3C). For multiple functionally relevant parallel GO terms we showed that in each serpentine population, selection likely targeted different genes which belonged to the same pathway or developmental process. The STRING analysis also showed that candidate genes were highly inter-connected within such pathways (STable 9). Overall, we found significant parallelism by genes and functions associated with serpentine stress with evidence for selection on different genes from the same functional pathways possibly underlying the origin of functional parallel traits.

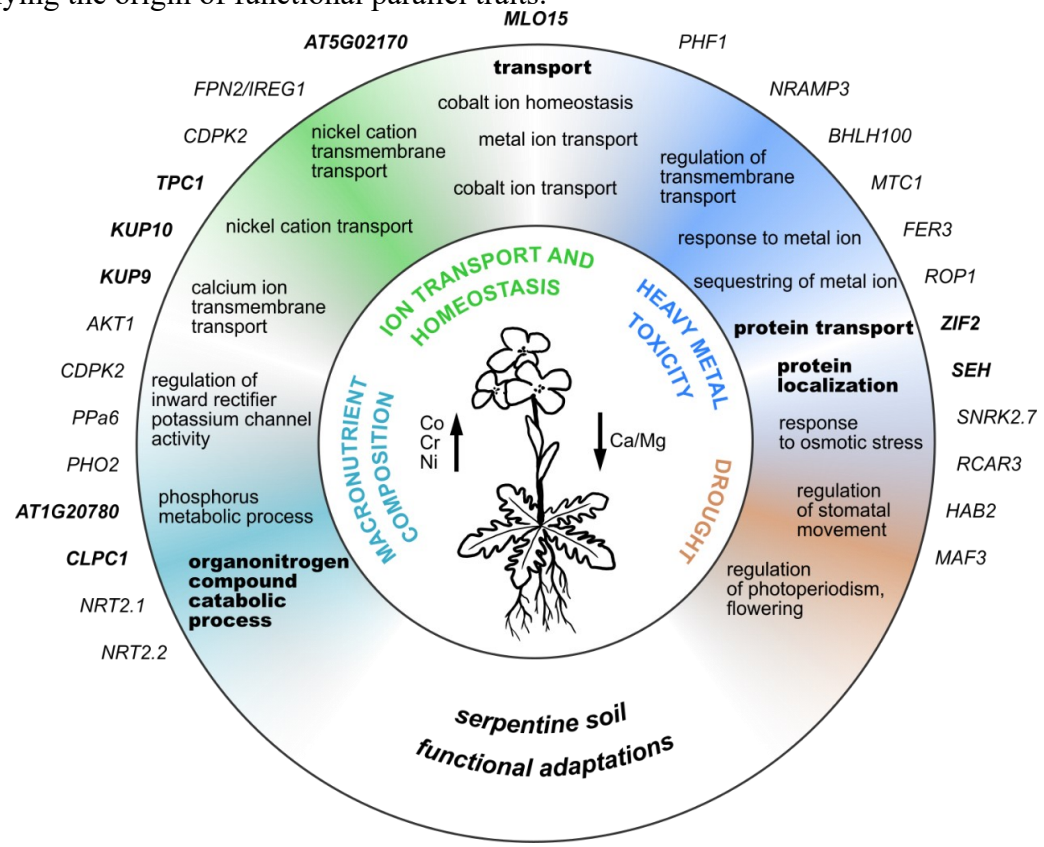


Fig. 4. Functional responses to the major environmental challenges of serpentine substrates detected by gene and functional pathway-level analysis of *A. arenosa*. The figure summarizes key serpentine-specific challenges following (Brady et al., 2005; O'Dell and Rajakaruna, 2011; Konečná et al., 2020), relevant significantly enriched biological process GO terms and differentiation candidate genes which were inferred for at least one (thin) and two/three (bold) serpentine populations. In the centre is an illustrative cartoon of the *A. arenosa* plant with phenotypic changes in line with serpentine adaptation detected in this study. Note: parallel genes and GO terms are in bold.

DISCUSSION

Varying magnitude of local adaptation and limited trade-offs

Despite generally significant serpentine adaptation, the magnitude of fitness differences between originally serpentine and non-serpentine populations strikingly varied among the three population pairs, raising a question on the reason for such variation. Firstly, the overall difference in selective environment may determine strength of the selection and thus also the extent of phenotypic divergence; positive correlation between magnitude of local adaptation and environmental distance, although with the small effect, was shown already by meta-analyses (Hereford, 2009). Although the S3 – N3 population pair exhibited the largest magnitude of local adaptation and was also the most diverged based on substrate environmental distance, this was not true for population pair 2 with comparable substrate environmental distance, yet by far the smallest magnitude of local adaptation (Table 2). Secondly, neutral genetic differentiation, mirroring the extent of genetic divergence and intensity of gene flow, may indicate that population genomic processes play an important role (Savolainen et al., 2013). Indeed, the magnitude of local adaptation was always in congruence with the genome-wide differentiation within population pairs. It suggests that higher genetic differentiation between geographically proximal serpentine and non-serpentine populations results in a higher difference in fitness of plants cultivated in the serpentine soils (Fig. 1A, B). Although there was broad evidence for local serpentine adaptation (reviewed e.g. in Brady et al., 2005; Konečná et al., 2020), experimental cultivations were not accompanied by genetic investigations of population history. Systematic studies on a larger number of S-N population contrasts are needed to understand the relationship between genetic differentiation and magnitude of local adaptation.

Using various combinations of fitness proxies, we detected utmost limited trade-offs of serpentine adaptation, i.e. typically no cost of serpentine adaptation in non-serpentine soil. Limited cost of adaptation seems to be in congruence with a major trend in other studies summarized by Hereford (2009). Generally, adaptation trade-offs can rather evolve in homogeneous environment (e.g. in copper tolerance in *Mimulus guttatus* (Wright et al., 2013) than in heterogenous environment (Bono et al., 2017). For example, there was no difference in survival in non-serpentine sites between serpentine and non-serpentine *M. guttatus* (Selby and Willis, 2018) and only a little cost of adaptation was identified in serpentine *Arabidopsis lyrata* (Veatch-Blohm et al., 2017). Even in the pioneering serpentine reciprocal transplant experiments, Kruckeberg (1951) demonstrated the absence of trade-offs over various serpentine plants. He also proposed that the causes of trade-offs are related to the presence/absence of competition. Although, there is not much evidence to support this hypothesis (Sianta and Kay, 2019), in *Helianthus exilis* local serpentine adaptation was observed only without competition (Sambatti and Rice, 2006). Due to the absence of competitors in our experiments, we cannot rule out this explanation for the limited cost of adaptation. On the other hand, similar habitat types, generally low-competitive open mixed/coniferous forests with rocky outcrops, of serpentine vs. non-serpentine *A. arenosa* populations rather suggest comparable level of competition in both habitats.

Pervasive phenotypic parallelism

We detected parallel response to the selective serpentine substrate over the three independent serpentine populations in nearly all scored phenotypic traits (Fig. 2A, B). According to our expectations, in this recently diverged system encompassing postglacially adapted populations (Konečná et al., 2021), parallelism is largely manifested at the phenotypic level. The parallel traits, with larger effect size of the substrate of origin than the effect size of substrate of

origin*population pair interaction, were prevailing – 10 out of 13 morphological life-history traits, fitness proxies, and traits related to ion uptake. Therefore, we showed that the phenotypic trait divergence between serpentine and non-serpentine populations seems to be largely predictable in *A. arenosa*. Experimental studies of serpentine adaptation revealed a range of life-history traits and physiological mechanisms showing similar responses among serpentine populations (e.g. Rajakaruna et al., 2003; Kolář et al., 2014), but also the variation in parallel traits depending on the similarity in the chemistry of the compared original serpentine soils (e.g. Berglund et al., 2004). The extent of parallelism was rarely studied and comprehensively quantified across multiple morphological and life-history traits, focusing mainly on animals (but see Knotek et al., 2020; James, Wilkinson, et al., 2021).

As specific soil chemistry is the principal selective factor of serpentines, regulation of micro-/macronutrients uptake and exclusion of heavy metals are expected to be the prime adaptive mechanisms (Brady et al., 2005; Kazakou et al., 2008). In line with this, serpentine adaptation in *A. arenosa* is likely dominantly driven by high levels of heavy metals (Co, Cr, and Ni), high concentration of Mg, and low Ca/Mg ratio, because concentration of these elements was consistently differentiating the original serpentine and non-serpentine soils (SFig. 1). Congruently, we found reduced uptake of all these elements in serpentine compared to non-serpentine plants when cultivated in serpentine soils, which was consistent across the three population pairs. *Arabidopsis arenosa* seems to reduce the uptake of heavy metals such as Ni, similarly as has been shown for closely related *A. lyrata* (Veatch-Blohm et al., 2017) and some other species from the same geographic region such as *Knautia arvensis* group (Kolář et al., 2014). Importantly, response to the altered Ca/Mg ratios, perhaps the most determining character of serpentines (Brady et al., 2005), has been observed already in the early life stages. We observed highly parallel trend of higher root growth of serpentine populations in media with highly skewed Ca/Mg ratios, well supporting similar studies on serpentine ecotypes of some other species (*Cerastium alpinum* and *K. arvensis;* Berglund et al., 2004; Kolář et al., 2014) yet not universally (e.g. *Galium valdepilosum;* Kolář et al., 2013 and *Streptanthus polygaloides;* Boyd et al., 2009). Similarly to *C. alpinum,* also non-serpentine plants of *A. arenosa* had a higher number of lateral roots than serpentine plants. Different root growth strategies possibly evolved in serpentine populations as a response to high chemical stress the majority of resources are invested in the main root growth and the formation of lateral roots is down-regulated (Berglund et al., 2004). Furthermore, the highest difference in this trait was in population pair 3, in which relevant differentiation candidate genes have been identified by a divergence scan (reduce lateral root formation (*RLF*), repressor of lateral root initiation (*NRT2.1*); SData 1) suggesting genetic basis of such trait.

Causes and consequences of non-parallel phenotypes

Despite the generally dominant phenotypic parallelism, we observed regionally specific patterns in three traits: above-ground biomass, flowering time and varying Ca/Mg ratio in leaves. Besides that, the non-parallelism was also observed in the germination rate in serpentine soils in a previous study by Konečná et al. (2021). While N2 and N3 populations had 93 % and 86 % germination rates, the N1 population did not germinate at all in serpentine soil (Konečná et al., 2021). Such locally-specific responses may reflect a complex interplay of the genetic basis of traits, genetic drift, and local environmental heterogeneity (Rosenblum et al., 2014; Fraïsse and Welch, 2019; James, Wilkinson, et al., 2021). In the case of leaf Ca/Mg ratio, S2 and S3 populations had relatively higher leaf Ca/Mg ratios than proximal non-serpentine populations, yet the opposite was observed for S1-N1 population pair. Highly skewed Ca/Mg ratio (<1) is the major factor defying serpentines worldwide (O'Dell and Rajakaruna, 2011) and the importance of Ca availability in the light of increased Mg has been highlighted for almost a century (Novák, 1928; Vlamis and Jenny, 1948; Vlamis, 1949; Kruckeberg, 1954; Walker et

al., 1955). For instance, adapted serpentine populations of *A. millefolium* had higher leaf Ca/Mg ratio than non-serpentine plants suggesting root to shoot transfer of calcium, selective uptake of Ca, or exclusion of Mg (O'Dell and Claassen, 2006). Variation in leaf Ca/Mg ratio among serpentine *A. lyrata* populations has been also shown by Veatch-Blohm et al. (2013). In *A. arenosa*, reduced uptake of Mg seems to be an important mechanism how to regulate the Ca/Mg ratio in serpentine plants (SFig. 4). Levels of Ca^{2+} cations in the cells are controlled by an array of channels and carriers pointing to the complex genetic basis of Ca homeostasis and signalling (Tang and Luan, 2017). Interestingly, one of the parallel differentiation candidate genes encoding central calcium channel and mediating stress signalling – *TPC1* (Choi et al., 2014) appeared as a selection candidate in population pairs 2 and 3, but not in pair 1 (Konečná et al., 2021) suggesting that observed non-parallelism may have a genetic basis.

Other non-parallel trait was flowering time, which was a subject of many studies on serpentine adaptation. The selection for flowering shifts with genetic basis has been shown in various taxa adapted to serpentine soils in California (Sianta and Kay, 2019). Sianta and Kay (2021) also showed that colonization of serpentine sites can cause maladaptive shifts to later flowering in the early stages of divergence. Indeed, this is in congruence with observations in *A. arenosa*, where we observed later flowering in the S2 population than in the N2 population, i.e. in the population pair exhibiting the most recent divergence time (Konečná et al., 2021). Although earlier flowering in serpentine plants was documented in many species, for instance in *Solidago virgaurea* (Sakaguchi et al., 2017), *Picris hieracioides* (Sakaguchi et al., 2018), and *Helianthus exilis* (Sambatti and Rice, 2006), Sianta and Kay (2021) showed shifts to later flowering in serpentine taxa. This is in high contrast to the general paradigm that serpentine adapted plants flower early to escape the drought in rocky serpentine sites in summer (Brady et al., 2005; Ferris and Willis, 2018). Such mixed evidence for plant earlier vs. later flowering in serpentine compared to non-serpentine populations is thus recapitulated also in our dataset and we hypothesize it might be due to variation among serpentine habitats. In particular, the S3 population with the earliest flowering time occupies the south-facing rocky outcrops that are most likely candidates for a selective extreme habitat affected by summer drought.

Genetic redundancy and polygenic architecture underlying observed parallelism

We observed significant gene parallelism but affecting a relatively low number of genes (ranging from 2.16 % to 2.4 %). It is rather surprising given the large pool of variation shared in this polyploid and recently diverged system (Konečná et al., 2021). However, it is consistent with an array of other studies showing similar levels of parallelism (Lai et al., 2019; Preite et al., 2019; Ji et al., 2020; Bohutínská et al., 2021; James, Wilkinson, et al., 2021; Papadopulos et al., 2021). Although, possibly high number of false-positive candidates resulting from F_{ST} scans can lead to a decrease in the percentage of parallels. This extent of parallelism corresponds with genetic redundancy and polygenic architecture of serpentine adaptation.

Relative low number of shared differentiation candidate genes can also reflect relatively simple genetic architecture of serpentine adaptation that is driven by only a few high-effect genes. Although this seems to be indeed the case for several serpentine plants (*Silene, Mimulus;* (Bratteler et al., 2006; Selby and Willis, 2018)), it is likely not the case of *Arabidopsis*. Polygenic basis of serpentine adaptation, i.e. allele frequency shifts in many genes across population pairs (Yeaman, 2015; Wilkinson et al., 2021), has been previously identified by high-density divergence scans for selection in *A. arenosa* and *A. lyrata* (Turner et al., 2010; Arnold et al., 2016; Konečná et al., 2021). These studies showed that many loci with small effects are often under selection when facing complex even well-defined environmental challenges. In contrast, evidence for simple genetic architecture is typically based on quantitative trait locus (QTL) studies, such as Ni tolerance in *Silene vulgaris* and *Caulanthus amplexicaulis* (Bratteler et al., 2006; Burrell et al., 2012) and survival of *Mimulus guttatus*

(Selby and Willis, 2018). Importantly, QTL studies are usually biased to detect major-effect loci thus providing limited insights into polygenic adaptations.

We also observed similarity in functional pathways, moreover, many of which were corresponding to observed parallel phenotypic differences in iron ion homeostasis, inorganic cation transmembrane transport, steroid biosynthetic process; molecular functions of calcium ion binding; and KEGG pathway (Kanehisa and Goto, 2000) of plant hormone signal transduction (Fig. 3C and Fig. 4 for other functional pathways). Importantly, significant overlaps in such GO terms remained even if parallel gene candidates were excluded, demonstrating selection on different genes in similar functional pathways and developmental processes. This shows the important role of genetic redundancy in rapid adaptations with polygenic basis (Boyle et al., 2017; Barghi et al., 2020; Láruson et al., 2020). That genetic redundancy as important factor underlying parallel adaptation towards complex environmental challenges has been also shown in other systems such as in alpine *Heliosperma pusillum* (Trucchi et al., 2017; Szukala et al., 2021) or dune *Senecio lauttus* (James, Allsopp, et al., 2021), but also from more narrowly defined environments – metal-polluted mines in *Silene uniflora* (Papadopulos et al., 2021) or temperature adaptations in *Drosophila simulans* (Barghi et al., 2019). In summary, given the expected polygenic basis of adaptation, it is unlikely that limited gene-level parallelism simply reflects genetic architecture but rather represent a combined effect of genetic redundancy and potential technical limitations of selection scans (e.g. increased rates of single-region false positives and limits in the detection of soft sweeps (Hoban et al., 2016)).

Conclusions

Here we document overall parallel response to serpentine substrate across a diversity of functional traits with utmost very limited trade-offs. Repeatedly identified enrichment of multiple relevant functional pathways, despite rather modest gene-level parallelism, demonstrates the complex interplay of allele sharing, likely polygenic architecture of the adaptation, and genetic redundancy underlies the observed parallel phenotypic manifestation of serpentine adaptation. Despite this general trend, there are also certain population- and trait-specific non-parallel deviations likely reflecting rather differences in genetic processes than variation in selective effect of the local environments. Further investigations of the genotype-phenotype links in well-defined strongly selective environments, such as serpentines, will help us to better assess the role of genetic redundancy in polygenic systems and its impact on the extent of parallelism in adaptation and thus evolutionary predictability.

Conflict of interest

The authors declare no conflict of interest.

Author contributions

V.K. and F.K. conceived the study. V.K., F.K., M.C., E.T., and M.S. designed experiments. V.K. and M.C. did field collections. A.K. performed the laboratory analyses. V.K., D.P. and M.S. performed the experiments. V.K. and D.P. performed data analyses. V.K. and F.K. wrote the manuscript with input from all authors.

Acknowledgements

This work was supported by Charles University (project Primus/SCI/35 to F.K. and the Grant Agency of Charles University project No. 410120 to V.K.) and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme [ERC-StG 850852 DOUBLE ADAPT to F.K.]. Additional support was provided by the long-term

research development project No. RVO 67985939 of the Czech Academy of Sciences. Computational resources were provided by the CESNET LM2015042 and the CERIT Scientific Cloud LM2015085. The authors thank Petr Knotek, Tereza Holcová, Veronika Vlčková, Eliška Konečná, Barbora Lepková, Timothée Lamotte, and Karolína Havlíková for help with field work, experiments, data collection and laboratory work.

REFERENCES

Alexa, A., and J. Rahnenfuhrer. 2020. topGO: Enrichment Analysis for Gene Ontology. R package version 2.42.0.

Arnold, B. J., B. Lahner, J. M. DaCosta, C. M. Weisman, J. D. Hollister, D. E. Salt, K. Bomblies, and L. Yant. 2016. Borrowed alleles and convergence in serpentine adaptation. *Proceedings of the National Academy of Sciences* 113: 8320–8325.

Barghi, N., J. Hermisson, and C. Schlötterer. 2020. Polygenic adaptation: a unifying framework to understand positive selection. *Nature Reviews Genetics* 21: 769–781.

Barghi, N., R. Tobler, V. Nolte, A. M. Jaksic, F. Mallard, K. A. Otte, M. Dolezal, et al. 2019. Polygenic adaptation fuels genetic redundancy in Drosophila. *PLoS Biology* 17: 1–31.

Berglund, A. N., S. Dahlgren, and A. Westerbergh. 2004. Evidence for parallel evolution and site-specific selection of serpentine tolerance in Cerastium alpinum during the colonization of Scandinavia. *New Phytologist* 161: 199–209.

Bohutínská, M., J. Vlček, S. Yair, B. Leanen, V. Konečná, M. Fracassetti, T. Slotte, and F. Kolář. 2021. Genomic basis of parallel adaptation varies with divergence in Arabidopsis and its relatives. *Proceedings of the National Academy of Sciences* 118: e2022713118.

Bolnick, D. I., R. D. H. Barrett, K. B. Oke, D. J. Rennison, and Y. E. Stuart. 2018. (Non) Parallel Evolution. *Annual Review of Ecology, Evolution, and Systematics* 49: 303–330.

Bono, L. M., L. B. Smith, D. W. Pfennig, and C. L. Burch. 2017. The emergence of performance trade-offs during local adaptation: insights from experimental evolution. *Molecular Ecology* 26: 1720–1733.

Boyd, R. S., M. A. Wall, S. R. Santos, and M. A. Davis. 2009. Variation of Morphology and Elemental Concentrations in the California Nickel Hyperaccumulator Streptanthus polygaloides (Brassicaceae) . *Northeastern Naturalist* 16: 21–38.

Boyle, E. A., Y. I. Li, and J. K. Pritchard. 2017. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* 169: 1177–1186.

Bradshaw, H. D. 2005. Mutations in CAX1 produce phenotypes characteristic of plants tolerant to serpentine soils. *New Phytologist* 167: 81–88.

Brady, K. U., A. R. Kruckeberg, and H. D. Bradshaw Jr. 2005. Evolutionary Ecology of Plant Adaptation to Serpentine Soils. *Annual Review of Ecology, Evolution, and Systematics* 36: 243–266.

Bratteler, M., M. Baltisberger, and A. Widmer. 2006. QTL analysis of intraspecific differences between two Silene vulgaris ecotypes. *Annals of Botany* 98: 411–419.

Burrell, M. A., A. K. Hawkins, and A. E. Pepper. 2012. Genetic analyses of nickel tolerance in a North American serpentine endemic plant, Caulanthus amplexicaulis var. barbarae (Brassicaceae). *American Journal of Botany* 99: 1875–1883.

Choi, W. G., M. Toyota, S. H. Kim, R. Hilleary, and S. Gilroy. 2014. Salt stress-induced Ca2+ waves are associated with rapid, long-distance root-to-shoot signaling in plants. *Proceedings of the National Academy of Sciences of the United States of America* 111: 6497–6502.

Fang, B., P. Kemppainen, P. Momigliano, and J. Merilä. 2021. Population Structure Limits Parallel Evolution in Sticklebacks. *Molecular Biology and Evolution* 38: 4205–4221.

Ferris, K. G., and J. H. Willis. 2018. Differential adaptation to a harsh granite outcrop habitat between sympatric Mimulus species. *Evolution* 72: 1225–1241.

Fraïsse, C., and J. J. Welch. 2019. The distribution of epistasis on simple fitness landscapes. *Biology Letters* 15: 20180881.

Geyer, C. J., S. Wagenius, and R. G. Shaw. 2007. Aster models for life history analysis. *Biometrika* 94: 415–426.

Hämälä, T., and O. Savolainen. 2019. Genomic patterns of local adaptation under gene flow in Arabidopsis lyrata. *Molecular Biology and Evolution* 36: 2557–2571.

Hereford, J. 2009. A quantitative survey of local adaptation and fitness trade-offs. *American Naturalist* 173: 579–588.

Hermisson, J., and P. S. Pennings. 2017. Soft sweeps and beyond: understanding the patterns and probabilities of selection footprints under rapid adaptation. *Methods in Ecology and Evolution* 8: 700–716.

Hoban, S., J. L. Kelley, K. E. Lotterhos, M. F. Antolin, G. Bradburd, D. B. Lowry, M. L. Poss, et al. 2016.

Finding the Genomic Basis of Local Adaptation: Pitfalls, Practical Solutions, and Future Directions. *The American naturalist* 188: 379–397.
Höllinger, I., P. S. Pennings, and J. Hermisson. 2019. Polygenic adaptation: From sweeps to subtle frequency shifts. *PLoS Genetics* 15: 1–26.
Jacobs, A., M. Carruthers, A. Yurchenko, N. V. Gordeeva, S. S. Alekseyev, O. Hooker, J. S. Leong, et al. 2020. Parallelism in eco-morphology and gene expression despite variable evolutionary and genomic backgrounds in a Holarctic fish. *PLoS Genetics* 16: e1008658.
James, M. E., R. N. Allsopp, J. S. Groh, K. Avneet, M. J. Wilkinson, and D. Ortiz-Barrientos. 2021. Uncovering the genetic architecture of replicated adaptation. *SSRN preprint. 3981902.*
James, M. E., H. Arenas-Castro, J. S. Groh, S. L. Allen, J. Engelstädter, and D. Ortiz-Barrientos. 2021. Highly Replicated Evolution of Parapatric Ecotypes. *Molecular Biology and Evolution* 38: 4805–4821.
James, M. E., M. J. Wilkinson, D. M. Bernal, H. Liu, H. L. North, J. Engelstädter, and D. Ortiz-Barrientos. 2021. Phenotypic and genotypic parallel evolution in parapatric ecotypes of Senecio. *Evolution* 75: 3115–3131.
Ji, Y., X. Li, T. Ji, J. Tang, L. Qiu, J. Hu, J. Dong, et al. 2020. Gene reuse facilitates rapid radiation and independent adaptation to diverse habitats in the Asian honeybee. *Science Advances* 6: eabd3590.
Kanehisa, M., and S. Goto. 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research* 28: 27–30.
Kazakou, E., P. G. Dimitrakopoulos, A. J. M. Baker, R. D. Reeves, and A. Y. Troumbis. 2008. Hypotheses , mechanisms and trade-offs of tolerance and adaptation to serpentine soils: from species to ecosystem level. 495–508.
Knotek, A., V. Konečná, G. Wos, D. Požárová, G. Šrámková, M. Bohutínská, V. Zeisek, et al. 2020. Parallel Alpine Differentiation in Arabidopsis arenosa. *Frontiers in Plant Science* 11: 561526.
Kolář, F., M. Dortová, J. Lepš, M. Pouzar, and A. Krejčová. 2014. Serpentine ecotypic differentiation in a polyploid plant complex: shared tolerance to Mg and Ni stress among di- and tetraploid serpentine populations of. *Plant and Soil* 374: 435–447.
Kolář, F., M. Lučanová, P. Vít, T. Urfus, J. Chrtek, T. Fér, F. Ehrendorfer, and J. Suda. 2013. Diversity and endemism in deglaciated areas: Ploidy, relative genome size and niche differentiation in the Galium pusillum complex (Rubiaceae) in Northern and Central Europe. *Annals of Botany* 111: 1095–1108.
Konečná, V., S. Bray, J. Vlček, M. Bohutínská, D. Požárová, R. R. Choudhury, A. Bollmann-Giolai, et al. 2021. Parallel adaptation in autopolyploid Arabidopsis arenosa is dominated by repeated recruitment of shared alleles. *Nature Communications* 12: 4979.
Konečná, V., L. Yant, and F. Kolář. 2020. The Evolutionary Genomics of Serpentine Adaptation. *Frontiers in Plant Science* 11: 574616.
Kruckeberg, A. R. 1951. Intraspecific Variability in the Response of Certain Native Plant Species to Serpentine Soil. *American Journal of Botany* 38: 408–419.
Kruckeberg, A. R. 1954. The ecology of serpentine soils: A symposium. III. Plant species in relation to serpentine soils. *Ecology* 35: 267–74.
Lai, Y. T., C. K. L. Yeung, K. E. Omland, E. L. Pang, Y. Hao, B. Y. Liao, H. F. Cao, et al. 2019. Standing genetic variation as the predominant source for adaptation of a songbird. *Proceedings of the National Academy of Sciences of the United States of America* 116: 2152–2157.
Láruson, Á. J., S. Yeaman, and K. E. Lotterhos. 2020. The Importance of Genetic Redundancy in Evolution. *Trends in Ecology and Evolution* 35: 809–822.
Lenormand, T., D. Roze, and F. Rousset. 2009. Stochasticity in evolution. *Trends in Ecology and Evolution* 24: 157–165.
Magalhaes, I. S., J. R. Whiting, D. D'Agostino, P. A. Hohenlohe, M. Mahmud, M. A. Bell, S. Skúlason, and A. D. C. MacColl. 2021. Intercontinental genomic parallelism in multiple three-spined stickleback adaptive radiations. *Nature Ecology and Evolution* 5: 251–261.
Morales, H. E., R. Faria, K. Johannesson, T. Larsson, M. Panova, A. M. Westram, and R. K. Butlin. 2019. Genomic architecture of parallel ecological divergence: Beyond a single environmental contrast. *Science Advances* 5: eaav9963.
Novák, F. A. 1928. Quelques remarques relatives au probleme de la vegetation sur les terrains serpentinques. *Preslia* 6: 42–71.
O'Dell, R. E., and V. P. Claassen. 2006. Serpentine and nonserpentine Achillea millefolium accessions differ in serpentine substrate tolerance and response to organic and inorganic amendments. *Plant and Soil* 279: 253–269.
O'Dell, R. E., and N. Rajakaruna. 2011. Intraspecific variation, adaptation, and evolution. *In* S. Harrison, and N. Rajakaruna [eds.], Serpentine: Evolution and ecology in a model system, 97–137. University of California Press.
Papadopulos, A. S. T., A. J. Helmstetter, O. G. Osborne, A. A. Comeault, D. P. Wood, E. A. Straw, L. Mason, et al. 2021. Rapid Parallel Adaptation to Anthropogenic Heavy Metal Pollution. *Molecular Biology and*

Evolution 38: 3724–3736.
Preite, V., C. Sailer, L. Syllwasschy, S. Bray, U. Kraemer, and L. Yant. 2019. Convergent evolution in Arabidopsis halleri and Arabidopsis arenosa on calamine metalliferous soils. *Philosophical Transactions of the Royal Society B* 374: 20180243.
Rajakaruna, N., M. Y. Siddiqi, J. Whitton, B. A. Bohm, and A. D. M. Glass. 2003. Differential responses to Na+/K+ and Ca2+/Mg2+ in two edaphic races of the Lasthenia californica (Asteraceae) complex: A case for parallel evolution of physiological traits. *New Phytologist* 157: 93–103.
Rosenblum, E. B., C. E. Parent, and E. E. Brandt. 2014. The Molecular Basis of Phenotypic Convergence. *Annual Review of Ecology, Evolution, and Systematics* 45: 203–226.
Sakaguchi, S., K. Horie, N. Ishikawa, A. J. Nagano, M. Yasugi, H. Kudoh, and M. Ito. 2017. Simultaneous evaluation of the effects of geographic, environmental and temporal isolation in ecotypic populations of Solidago virgaurea. *New Phytologist* 216: 1268–1280.
Sakaguchi, S., K. Horie, T. Kimura, A. J. Nagano, Y. Isagi, and M. Ito. 2018. Phylogeographic testing of alternative histories of single-origin versus parallel evolution of early flowering serpentine populations of Picris hieracioides L. (Asteraceae) in Japan. *Ecological Research* 33: 537–547.
Salmón, P., A. Jacobs, D. Ahrén, C. Biard, N. J. Dingemanse, D. M. Dominoni, B. Helm, et al. 2021. Continent-wide genomic signatures of adaptation to urbanisation in a songbird across Europe. *Nature Communications* 12: 2983.
Sambatti, J. B. M., and K. J. Rice. 2006. Local adaptation, patterns of selection, and gene flow in the Californian serpentine sunflower (Helianthus exilis). *Evolution* 60: 696–710.
Savolainen, O., M. Lascoux, and J. Merilä. 2013. Ecological Genomics: genes in ecology and ecology in genes. *Nature Reviews Genetics* 14: 807–820.
Schmickl, R., J. Paule, J. Klein, K. Marhold, and M. A. Koch. 2012. The evolutionary history of the Arabidopsis arenosa complex: Diverse tetraploids mask the Western Carpathian center of species and genetic diversity. *PLoS ONE* 7: e42691.
Selby, J. P., and J. H. Willis. 2018. Major QTL controls adaptation to serpentine soils in Mimulus guttatus. *Molecular Ecology* 27: 5073–5087.
Shaw, R. G., C. J. Geyer, S. Wagenius, H. H. Hangelbroek, and J. R. Etterson. 2008. Unifying life-history analyses for inference of fitness and population growth. *American Naturalist* 172: E35–E47.
Sianta, S. A., and K. M. Kay. 2019. Adaptation and divergence in edaphic specialists and generalists: serpentine soil endemics in the California flora occur in barer serpentine habitats with lower soil calcium levels than serpentine tolerators. *American Journal of Botany* 106: 690–703.
Sianta, S. A., and K. M. Kay. 2021. Parallel evolution of phenological isolation across the speciation continuum in serpentine-adapted annual wildflowers. *Proceedings of the Royal Society B: Biological Sciences* 288: 20203076.
Stuart, Y. E., T. Veen, J. N. Weber, D. Hanson, M. Ravinet, B. K. Lohman, C. J. Thompson, et al. 2017. Contrasting effects of environment and genetics generate a continuum of parallel evolution. *Nature Ecology and Evolution* 1: 1–7.
Szklarczyk, D., A. Franceschini, S. Wyder, K. Forslund, D. Heller, J. Huerta-Cepas, M. Simonovic, et al. 2015. STRING v10: Protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Research* 43: D447–D452.
Szukala, A., J. Lovegrove-Walsh, H. Luqman, S. Fior, T. Wolfe, B. Frajman, P. Schönswetter, and O. Paun. 2021. Polygenic routes lead to parallel altitudinal adaptation in. *bioRxiv:* 2021.07.05.451094.
Tang, R., and S. Luan. 2017. Regulation of calcium and magnesium homeostasis in plants: from transporters to signaling network. *Current Opinion in Plant Biology* 39: 97–105.
Trucchi, E., B. Frajman, T. H. A. Haverkamp, P. Schönswetter, and O. Paun. 2017. Genomic analyses suggest parallel ecological divergence in Heliosperma pusillum (Caryophyllaceae). *New Phytologist* 216: 267–278.
Turner, T. L., E. C. Bourne, E. J. Von Wettberg, T. T. Hu, and S. V Nuzhdin. 2010. Population resequencing reveals local adaptation of Arabidopsis lyrata to serpentine soils. *Nature Genetics* 42: 260–263.
Veatch-Blohm, M. E., B. M. Roche, and M. J. Campbell. 2013. Evidence for Cross-Tolerance to Nutrient Deficiency in Three Disjunct Populations of Arabidopsis lyrata ssp. lyrata in Response to Substrate Calcium to Magnesium Ratio. *PLoS ONE* 8: e63117.
Veatch-Blohm, M. E., B. M. Roche, and E. E. Dahl. 2017. Serpentine populations of Arabidopsis lyrata ssp. lyrata show evidence for local adaptation in response to nickel exposure at germination and during juvenile growth. *Environmental and Experimental Botany* 138: 1–9.
Vlamis, J. 1949. Growth of Lettuce and Barley As Influenced By Degree of Calcium-Saturation of Soil. *Soil Science* 67: 453–466.
Vlamis, J., and H. Jenny. 1948. Calcium deficiency in serpentine soils as revealed by adsorbent technique. *Science* 107: 549.
Walker, R. B., M. Walker, Helene, and P. Ashworth. 1955. Calcium-magnesium nutrition with special reference

to serpentine soils. *Plant Physiology* 30: 214.
Wang, M., Y. Zhao, and B. Zhang. 2015. Efficient Test and Visualization of Multi-Set Intersections. *Scientific Reports* 5: 1–12.
Weber, A. A. T., J. Rajkov, K. Smailus, B. Egger, and W. Salzburger. 2021. Speciation dynamics and extent of parallel evolution along a lake-stream environmental contrast in African cichlid fishes. *Science Advances* 7: 1–21.
Whiting, J. R., J. R. Paris, M. J. van der Zee, P. J. Parsons, D. Weigel, and B. A. Fraser. 2021. Drainage-structuring of ancestral variation and a common functional pathway shape limited genomic convergence in natural high- And low-predation guppies. *PLoS Genetics* 17: 1–29.
Wilkinson, M. J., F. Roda, G. M. Walter, M. E. James, R. Nipper, J. Walsh, S. L. Allen, et al. 2021. Adaptive divergence in shoot gravitropism creates hybrid sterility in an Australian wildflower. *Proceedings of the National Academy of Sciences of the United States of America* 118: 1–11.
Wright, K. M., D. Lloyd, D. B. Lowry, M. R. Macnair, and J. H. Willis. 2013. Indirect Evolution of Hybrid Lethality Due to Linkage with Selected Locus in Mimulus guttatus. *PLoS Biology* 11: e1001497.
Yeaman, S. 2015. Local adaptation by alleles of small effect. *American Naturalist* 186: S74–S89.
Zbíral, J., E. Čižmárová, E. Obdržálková, M. Rychlý, V. Vilamová, J. Srnková, and A. Žalmanová. 2016. Jednotné pracovní postupy: Analýza půd I. Ústřední kontrolní a zkušební ústav zemědělský.

SUPPLEMENTARY METHODS

Reciprocal transplant experiment

We germinated seeds from 12 mother plants (each representing a seed family of a mixture of full- and half-sibs) from each population in Petri dishes filled by either type of soil (15 seeds/family/treatment). Seeds germinated in the growth chamber (Conviron) under conditions approximating spring season at the original sites: 12 h dark at 10 °C and 12 h light at 20 °C. Due to zero germination of N1 seeds in S1 soil, we measured differential growth response on plants that were germinated in the non-serpentine soils and were subjected to the differential soil treatment later, in a seedling stage. We chose 44–50 seedlings equally representing progeny of 11 maternal plants per each population (in total 284 seedlings), transferred each plant to a separate pot filled either with ~1 L of the original or the alternative paired soil (i.e. S1 soil for N1 population and vice versa). We randomly swapped the position of each pot twice a week and watered them with tap water when needed. The germination rates were published in Konečná et al. (2021).

Elemental analysis of soil and leaf samples

Determination of Ca, Co, Cr, Mg, and Ni in plant tissues of *A. arenosa* and soil samples was carried out using inductively coupled plasma optical emission spectrometry (ICP OES).

Due to very small amounts of leaf samples in units to tens of mg, the samples were decomposed prior to the analysis using the microwave oven Speedwave ®Xpert (Berghof, Germany, maximal applied power 2000 W) with a multi-tube system. The plant tissue (2 replicates, 8 – 50 mg according to the available sample amount) was inserted into digestion tubes and treated with 2 ml of sub boilingly distilled (Berghof, Germany) nitric acid (Lachner, the Czech Republic) under the following conditions: 10 min hold on 170 °C, 30 % of maximal power, 10 min on 200 °C, 30 % of power, 30 min on 30 °C, 0 % of power. The mineralised samples were filled up to the final volume of 10 ml with deionised water (conductivity 0.055 μS/cm, Evoqua Water Technologies, Germany).

The elemental analysis of Ca, Co, Cr, Mg, and Ni was carried out using the sequential, radially viewed ICP OES spectrometer INTEGRA 6000 (GBC, Dandenong Australia) equipped with the ultrasonic nebulizer U5000AT+ (Teledyne Cetac Technologies, the USA), concentric nebulizer (2 $ml.min^{-1}$) and a glass cyclonic spray chamber (both Glass Expansion, Australia). The analytical lines used were Mg 285.2213 nm, Ca 422.673 nm, Ni 221.647 nm, Co 238.892 nm, and Cr 267.716 nm. The operation conditions of the ICP OES analysis were as follows: sample flow rate 1.5 $mL.min^{-1}$, plasma power 1000 W, plasma, auxiliary and nebulizer gas flow rates 10, 0.4, and 0.52 $L.min^{-1}$, respectively, photomultiplier voltage 600 V for Co, Cr, and Ni and 350 V for Ca and Mg, view height 6.5mm, three replicated reading on-peak 1 s, fixed point background correction. The multielement standards containing 10 – 5 – 1 – 0.5 – 0.1 $mg.L^{-1}$ of Mg and Ca and 0.1 – 0.05 – 0.01 – 0.005 – 0.001 $mg.L^{-1}$ of Co, Cr, and Ni were used for instrument calibration. The external calibration standards were prepared using standard solutions of Ca, Co, Cr, Mg, and Ni all containing 1 $g.L^{-1}$ (SCP, Canada). The limits of detection (concentration equal to three times the standard deviation at the point of the background correction) were 0.0005 $\mu g.L^{-1}$ for Co, Cr, and Ni and 2 $\mu g.L^{-1}$ for Ca and Mg. Certified reference material (Bush twigs and leaves GBW 07602 from the China National Analysis Centre for Iron and Steel, Beijing) was used to validate the method and for the quality control.

For the elemental soil analysis, we used extraction method of Mehlich III (Zbíral et al., 2016). The extraction buffer contained: 0.2 mol/L CH_3COOH, 0.015 mol/L NH_4F, 0.013 mol/L

HNO_3, 0.25 mol/L NH_4NO_3, 0.001 mol/L EDTA. We mixed 100 ±0,5 mL of extraction buffer with 10 g of fine-grained soil (2 mm) and spiked them for 10 min.

Root growth experiments in calcium/magnesium solutions

The control medium (Ca/Mg ratio 1.97) was based on one-fifth strength Murashige-Skoog medium and contained: 3.76 mM KNO_3, 0.25 mM KH_2PO_4, 0.3 mM $MgSO_4$, 0.59 mM $CaCl_2$, 20 µM H_3BO_3, 0.02 µM $CoCl_2$, 0.02 µM $CuSO_4$, 20 µM $FeSO_4$, 22.4 µM $MnSO_4$, 0.21 µM Na_2MoO_4, 1 µM KI, and 5.98 µM $ZnSO_4$. Medium with Ca/Mg ratio 0.04 contained 0.2 mM $CaCl_2$ and 4.5 mM of $MgCl_2$ according to Bradshaw (2005). Medium with Ca/Mg ratio 0.2 contained 2.15 mM $CaCl_2$ and 6.97 mM $MgCl_2$. Other salts were added in the same concentrations as in the control medium. All media were supplemented with 1 % w/v sucrose and solidified with 1 % w/v agar (Plant agar; Duchefa, Netherlands). pH of the media was adjusted to 6.0 by NaOH.

Seeds were surface sterilized with 20 % solution of commercial bleach with 0.1 % triton for 15 min and washed with distilled water three times. Sterilized seeds were sown on a control medium in the 120x120 mm sterile plates and vernalised at 4 °C and in the dark for 7 days. After vernalisation, seeds were germinated in a cultivation room with constant growth conditions (22/18 °C day/night temperature, 16/8 hours light/dark cycle). Six days after germination seedlings were transferred onto media with Ca/Mg ratios 0.2 and 0.04 and cultivated for the next nine days in the same cultivation conditions. We fixed plants on the 15th day after germination in 4 % formaldehyde overnight, degassed under vacuum, and gradually saturated with 15 % and 30 % glycerol. Roots of glycerol-saturated plants were scanned at high resolution (1200dpi, 24-bit) and root system traits were measured by Root analyser plug-in in NIS Elements AR 3.22.05 software (Laboratory Imaging).

SUPPLEMENTARY TABLES AND FIGURES

STable 1. Details on sampled populations.

Pop	Ploidy	Population name	Natural populations	Experiment	Bedrock type	Altitude	Latitude	Longitude	Country
			N ind genome/soil ionomics	N ind cultivated in total/ionomics S treatment					
S1	4x	Borovsko	7/8	47/10	serpentine	416	49.68381	15.13326	CZ
S2	4x	Steinegg	7/8	50/10	serpentine	414	48.62993	15.54256	AT
S3	4x	Gulsen	8/8	45/10	serpentine	628	47.28167	14.92764	AT
N1	4x	Vlastejovice	8/8	48/7	siliceous	345	49.73496	15.17484	CZ
N2	4x	Fuglau	8/8	50/9	siliceous	436	48.63149	15.55723	AT
N3	4x	Ingeringgraben	8/8	44/10	siliceous	950	47.28405	14.68154	AT

STable 2. Overview of survival and transition to reproduction of individuals in reciprocal transplant experiment.

Pop	Treatment	Total N of plants	Buds	Flowers	Fruits
N1	S	24	21	13	6
N1	N	24	24	22	20
N2	S	25	24	8	3
N2	N	25	24	19	14
N3	S	22	18	13	8
N3	N	22	19	17	15
S1	S	23	23	21	17
S1	N	24	23	22	21
S2	S	25	23	9	4
S2	N	25	25	22	15
S3	S	23	22	22	22
S3	N	22	22	19	17

STable 3. The effects of substrate of origin, population pair and their interaction on total fitness inferred by an aster hierarchical model. The aster models consisted of four fitness components, which were aligned in the following directional graph: proportion to bolting → proportion to flowering → fruit production → total seed mass production. All factors were tested by likelihood ratio tests using nested null models (model separately with substrate of origin or treatment effects were compared to model including both factors, further the model with both substrate of origin and population pair effects were compared to mode with substrate of origin*population pair interaction).

Treatment	Tested factor	Null df	Alternative df	Null deviance	Alternative deviance	Test df	Test deviance	*p* value
serpentine	Substrate of origin	5	7	20002	20346	2	-344	< 0.0001
	Population pair	6	7	20031	20346	1	-315	< 0.0001
	Substrate of origin + Population pair	7	9	20346	20384	2	-38	< 0.0001
non-serpentine	Substrate of origin	5	7	28718	29566	2	-848	< 0.0001
	Population pair	6	7	29554	29566	1	-12	0.0005
	Substrate of origin + Population pair	7	9	29566	29647	2	-81	< 0.0001

STable 4. Differences in total fitness inferred by hierarchical aster models separately for each population pair in serpentine and non-serpentine treatment. The differences were tested using likelihood ratio tests.

Population pair	Serpentine treatment		Non-serpentine treatment	
	$\chi 2$	*p* value	$\chi 2$	*p* value
S1-N1	19.52	< 0.0001	0.0006	0.9798
S2-N2	0.45	0.5002	1.14	0.2851
S3-N3	366.48	< 0.0001	2.35	0.1251

STable 5 Contribution of phenotypic traits to cumulative fitness of *A. arenosa* serpentine and non-serpentine plants.

Pop	Treatment	N indiv total/total seed mass production/above-ground biomass	Flower production and early survival	Fruit production	Late survival	Total seed mass production	Above-ground biomass	Cumulative fitness
N1	S	24/6/10	0.541667	0.461538	0.833333	0.02911	0.018178	0.00011
N2	S	25/3/16	0.32	0.375	1	0.103093	0.259713	0.003213
N3	S	22/8/18	0.590909	0.615385	1	1.989026	1.115717	0.806978
S1	S	23/17/17	0.913043	0.809524	0.823529	0.777088	0.191078	0.090382
S2	S	25/4/13	0.36	0.444444	0.5	0.3585	0.35725	0.010246
S3	S	23/22/22	0.956522	1	0.954545	2.818786	0.927159	2.386207
N1	N	24/20/16	0.916667	0.909091	0.7	0.857097	1.143467	0.571703
N2	N	25/14/12	0.76	0.736842	0.857143	1.21923	0.955973	0.559464
N3	N	22/15/14	0.772727	0.882353	0.8	0.893986	1.02013	0.497445
S1	N	24/21/11	0.916667	0.954545	0.47619	1.136098	0.791321	0.374591
S2	N	25/15/21	0.88	0.681818	0.933333	0.795385	1.044027	0.465026
S3	N	22/17/17	0.863636	0.894737	0.882353	1.093541	0.983422	0.733236

STable 6. Summary of linear models (LM) testing the effect of substrate of origin, the effect of population pair and their interactions on functional traits (ion uptake, morphological life-history traits, and fitness proxies) of plants cultivated in serpentine soils. Tests of significance for individual fixed effect factors and interaction were conducted by Type III Wald $\chi 2$ tests.

Response variable	**Transformation**	**Substrate of origin**			**Pop. pair**			**Substrate of origin*pop.pair**		
		Df	$\chi 2$	*p*	Df	$\chi 2$	*p*	Df	$\chi 2$	*p*
Ca [ppm]	log	1	18.1277	***	2	9.3029	***	2	2.6932	.
Mg [ppm]	log	1	5.4663	*	2	7.3841	**	2	2.6947	.
Ca/Mg	log	1	5.569	*	2	7.8672	**	2	15.0728	***
Co [ppm]	log	1	10.5126	**	2	10.7253	***	2	1.1056	ns
Cr [ppm]	log	1	22.911	***	2	41.466	***	2	11.793	***
Ni [ppm]	log	1	14.3601	***	2	62.4837	***	2	4.8947	*
Bolting time [days]	log	1	2.0948	ns	2	26.8956	***	2	12.4542	***
Flowering time [days]	log	1	0.007	ns	2	2.8453	.	2	5.116	**
Rosette area [mm^2]	sqrt	1	122.226	***	2	6.6118	**	2	9.7281	***
N of additive rosettes	log	1	38.815	***	2	11.1697	***	2	4.7695	*
Total seed mass production [mg]	log	1	31.0958	***	2	7.204	**	2	3.3422	*
Above-ground biomass [mg]	sqrt	1	12.2657	***	2	12.3582	***	2	4.7489	*
Root biomass [mg]	log	1	13.5719	**	2	14.7037	***	2	3.3024	.

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; · $p < 0.1$

STable 7. Effects of substrate of origin (S vs. N), treatment (S vs. N), and their interactions on functional traits tested by linear models (LM), binomial generalized linear models (GLM binomial), and linear mixed effect models (LMM) for each population pair separately. Tests of significance for individual fixed effect factors and interaction were conducted by Type III Wald $\chi 2$ tests. The random effects in LMM were cultivation run and Petri dish for cultivation.

Response variable	**Transformation**	**Model**	**Pop. pair**	**Substrate of origin**			**Treatment**			**Ecotype*treatment**		
				df	$\chi 2$	*p*	df	$\chi 2$	*p*	df	$\chi 2$	*p*
Bolting time [days]	log	LM	1	1	1.6173	ns	1	5.3656	*	1	1.9493	ns
Bolting time [days]	log	LM	2	1	5.4496	*	1	9.5134	**	1	0.1474	ns
Bolting time [days]	log	LM	3	1	60.53	***	1	10.794	**	1	12.1929	***
Flowering time [days]	log	LM	1	1	0.0067	ns	1	4.1831	*	1	0.0465	ns
Flowering time [days]	log	LM	2	1	2.1932	ns	1	2.1896	ns	1	6.3390	*
Flowering time [days]	log	LM	3	1	17.732	***	1	5.15737	**	1	5.1469	*
Rosette area [mm^2]	sqrt	LM	1	1	48.676	***	1	251.722	***	1	33.82	***
Rosette area [mm^2]	sqrt	LM	2	1	42.6912	***	1	144.2352	***	1	7.9263	**
Rosette area [mm^2]	sqrt	LM	3	1	85.345	***	1	53.6161	***	1	35.877	***
N of additive rosettes	log	LM	1	1	25.1521	***	1	66.8079	***	1	10.4011	**
N of additive rosettes	log	LM	2	1	13.2123	***	1	94.4882	***	1	2.2259	ns
N of additive rosettes	log	LM	3	1	6.4998	*	1	3.4469	*	1	4.1851	*
Probability of fruit production (0/1)		GLM binomial	1	1	10.2295	**	1	14.043	***	1	2.8821	.

Probability of fruit production (0/1)		GLM binomial	2	1	0.1651	ns	1	9.2197	**	1	0.0287	ns
Probability of fruit production (0/1)		GLM binomial	3	1	10.7342	**	1	4.3036	*	1	5.9476	*
Total seed mass production [mg]	log	LM	1	1	27.7629	***	1	27.4466	***	1	12.937	***
Total seed mass production [mg]	log	LM	2	1	1.5909	ns	1	7.0795	*	1	4.2835	*
Total seed mass production [mg]	log	LM	3	1	6.4173	*	1	0.0225	ns	1	3.0056	.
Above-ground biomass [mg]	sqrt	LM	1	1	6.7994	*	1	74.0321	***	1	9.7637	**
Above-ground biomass [mg]	sqrt	LM	2	1	1.8036	.	1	39.1427	***	1	0.3055	ns
Above-ground biomass [mg]	sqrt	LM	3	1	0.3318	ns	1	0.0068	ns	1	0.0252	ns
Root biomass [mg]	log	LM	1	1	19.738	***	1	77.331	***	1	17.991	***
Root biomass [mg]	log	LM	2	1	5.6086	*	1	25.3902	***	1	4.8871	*
Root biomass [mg]	log	LM	3	1	0.0142	ns	1	0.5516	ns	1	5.0671	*
Probability of late survival (after fruit production) (0/1)		GLM binomial	1	1	6.5438	*	1	8.0743	**	1	2.1933	ns

Probability of late survival (after fruit production) (0/1)		GLM binomial	2	1	0.0848	ns	1	4.8608	*	1	0.0621	ns
Probability of late survival (after fruit production) (0/1)		GLM binomial	3	1	0.0001	.	1	0.1392	ns	1	0.0001	.
Total root growth [cm]	log	LMM	1	1	126.12	***	2	61.19	***	2	45.36	***
Total root growth [cm]	log	LMM	2	1	72.507	***	2	102.225	***	2	44.424	***
Total root growth [cm]	log	LMM	3	1	119.729	***	2	79.918	***	2	26.266	***
Main root length [cm]		LMM	1	1	15.2207	***	2	26.2520	***	2	6.9497	*
Main root length [cm]		LMM	2	1	39.866	***	2	43.007	***	2	19.634	***
Main root length [cm]		LMM	3	1	92.897	***	2	54.702	***	2	23.9	***
Density of lateral roots (number of lateral roots/main root length)	sqrt	LMM	1	1	16.617	***	2	26.164	***	2	10.292	**
Density of lateral roots (number of lateral roots/main root length)	sqrt	LMM	2	1	22.098	***	2	20.394	***	2	10.882	**
Density of lateral roots (number of lateral roots/main root length)	sqrt	LMM	3	1	156.261	***	2	108.780	***	2	54.395	***

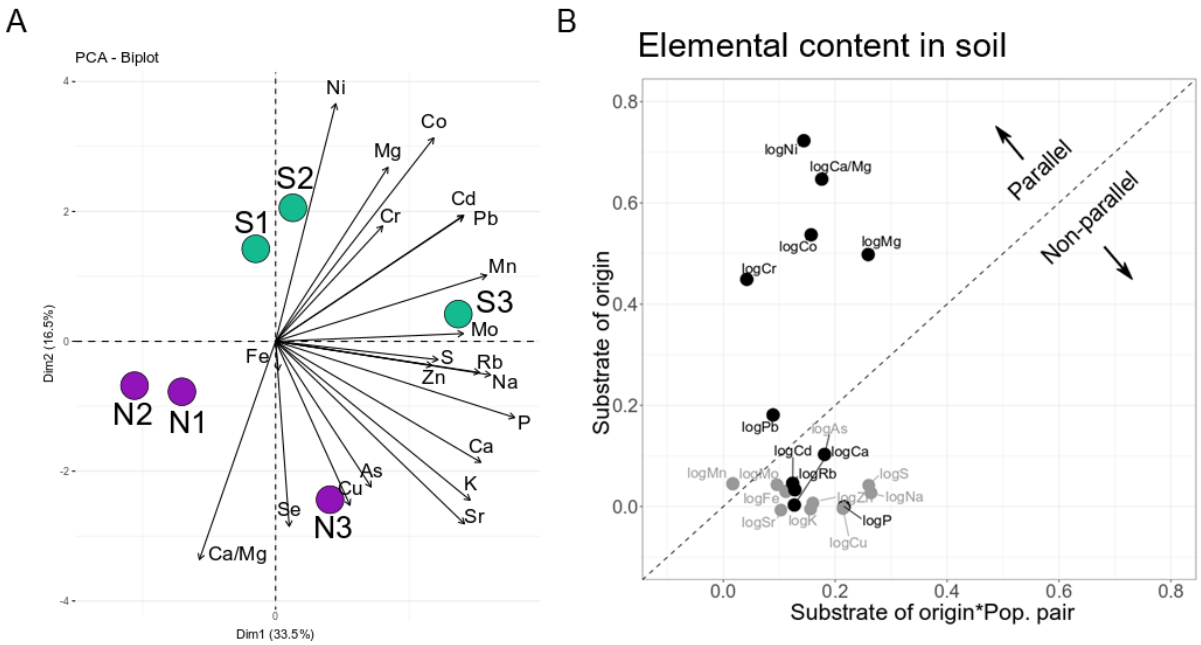
*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; · $p < 0.1$

STable 8. Quantification of parallelism by genes and functions. Percentages of shared differentiation candidate genes and significantly enriched gene ontology terms ($p < 0.05$) in biological processes domain - BP, molecular functions - MF, and cellular components - CC between any two population pairs when applying the classic algorithm (Alexa and Rahnenfuhrer, 2020). We calculated proportion of shared genes or functions as number of shared differentiation candidate genes or gene ontology (GO) terms per each two population pairs divided by the total number of differentiation candidate genes or GO terms for selected population pair combination. Percentages of shared enriched gene ontology terms, which were calculated based on GO enrichment without parallel differentiation candidates, are given in brackets.

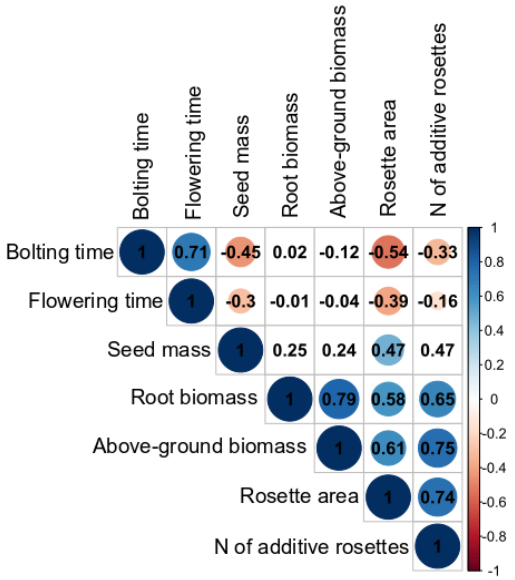
Level	**S1-N1 – S2-N2**	**S1-N1 – S3-N3**	**S2-N2 – S3-N3**
genes	2.38 %	2.40 %	2.16 %
enriched gene ontology terms BP	8.43 % (5.06 %)	3.87 % (4.43 %)	2.02 % (1.17 %)
enriched gene ontology terms MF	6.52 % (1.54 %)	5.43 % (1.6 %)	8.06 % (1.7 %)
enriched gene ontology terms CC	8.41 % (7.02 %)	5.56 % (5.77 %)	5.49 % (2.17 %)

STable 9. Number of protein interactions inferred by STRING for candidate differentiation genes for each population pair.

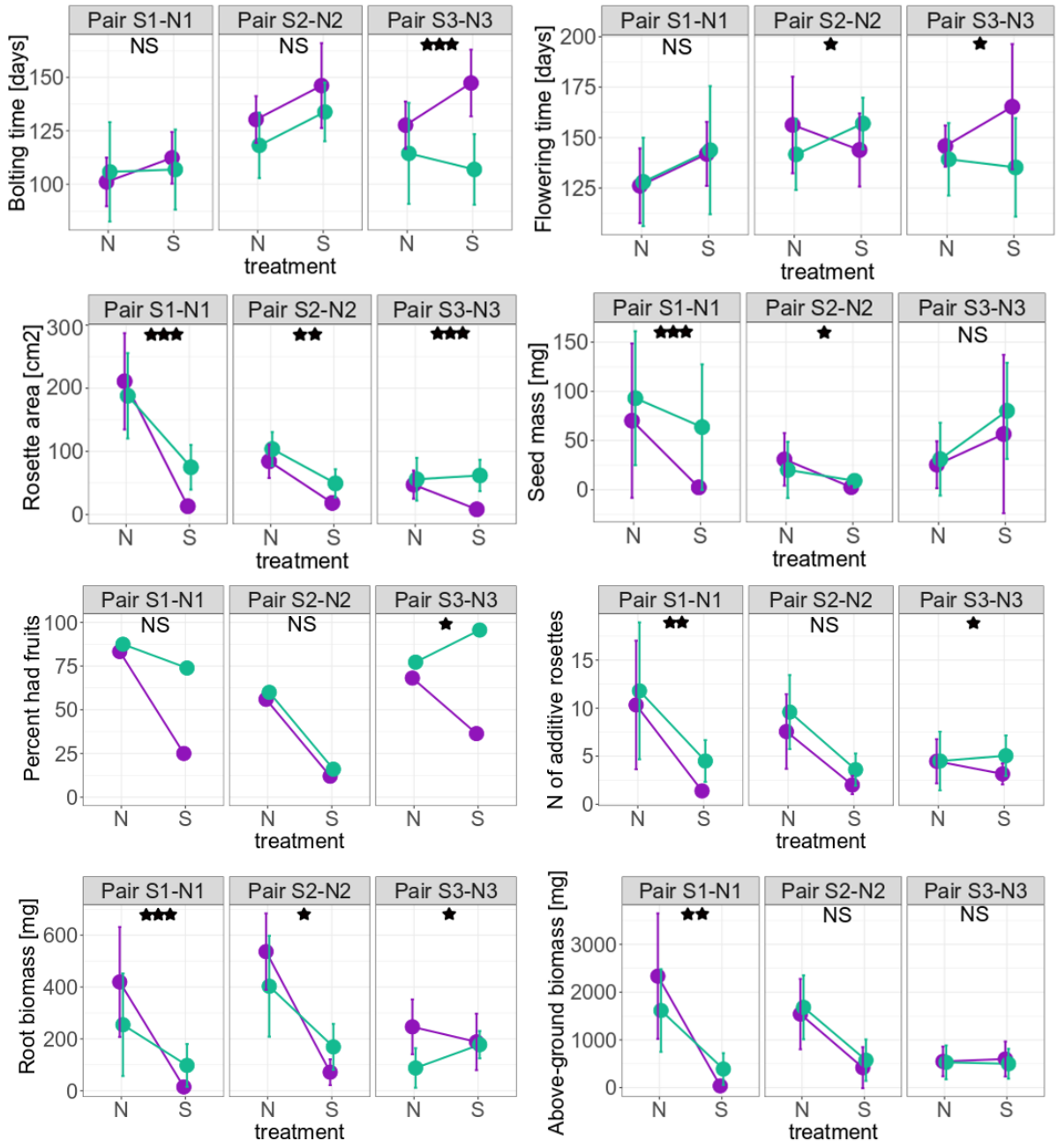
Population pair	**N differentiation candidate genes**	**N differentiation candidate genes interacted in total[1]**	**N differentiation candidate genes interacted with genes from at least one other population pair[2]**	**Proportion[1]**	**Proportion[2]**
S1-N1	494	358	340	0.72	0.69
S2-N2	513	357	332	0.7	0.65
S3-N3	504	363	333	0.72	0.66



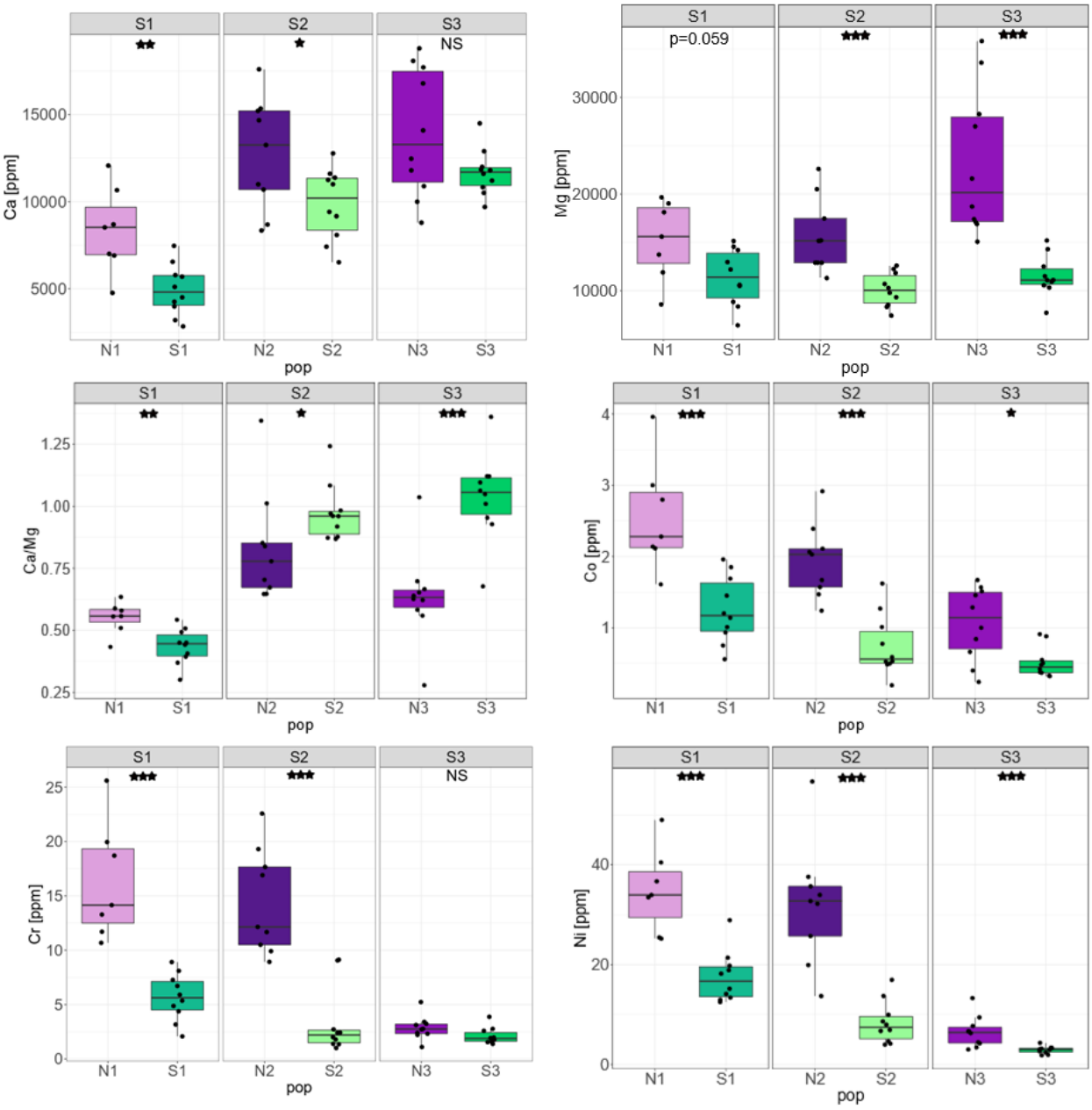
SFig. 1. A) PCA based on all soil variables (population means of 7/8 individual observations - STable 1) measured at original sites showing the differentiation between S and N soils. Serpentine (S, green dots) and non-serpentine (N, violet dots). B) Variation in the extent of parallelism among individual soil elemental concentrations from natural populations. The extent of parallelism was estimated as effect sizes (Eta2) in linear models addressing the effect of substrate of origin (S vs. N), pop. pair (1, 2, 3) and their interaction, calculated separately for each trait (particular elemental soil concentration and Ca/Mg ratio). The effect size of substrate of origin (y-axis) shows the extent to which a given trait diverges predictably between substrate of origins, i.e., in parallel, while the substrate of origin*pop. pair effect size (x axis) quantifies the extent to which serpentine/non-serpentine soil divergence varies across population pairs (i.e., deviates from parallel). Points falling above the dashed line have a larger substrate of origin effect (i.e. parallel) than substrate of origin*pop. pair interaction effect (i.e. non-parallel). Note: green arrows indicate the trend in all serpentine populations; grey points indicate the non-significant ($p < 0.05$) effect of substrate of origin.



SFig. 2. Pairwise Spearman's correlations among fitness traits. Note: circle size denotes significance (larger circle=lower *p* value), displayed are only circle sizes with $p < 0.05$).



SFig. 3. Differences in vegetative and reproductive fitness traits of three population pairs cultivated in local serpentine (S) and non-serpentine (N) soils. Note: asterisks denote the significance of substrate of origin*treatment interaction within each population pair (*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; see Table S7); the percentage of plants that had fruits was counted from the total number of cultivated plants.



SFig. 4. Variation in the uptake of Ca, Mg, and exclusion of Ni, Co, and Cr in serpentine (S) and non-serpentine (N) plants cultivated in serpentine soils (S1, S2, and S3). Note: asterisks denote the significance of the effect of substrate of origin (*** $p < 0.001$; ** $p < 0.01$; * $p <= 0.05$); elemental concentrations were assessed from plants after 6 months of cultivation; dots denote individuals.