Friederike Jantzen | Joseph H. Lynch | Christian Kappel | Jona Höfflin
Oded Skaliter | Natalia Wozniak | Adrien Sicard | Claudia Sas
Funmilayo Adebesin | Jasmin Ravid | Alexander Vainstein
Monika Hilker | Natalia Dudareva | Michael Lenhard

Retracing the molecular basis and evolutionary history of the loss of benzaldehyde emission in the genus Capsella

Suggested citation referring to the original publication:
New Phytologist 224 (2019) 3, pp. 1349–1360
DOI https://doi.org/10.1111/nph.16103
ISSN (print) 0028-646x
ISSN (online) 1469-8137

New Phytologist

Research

Retracing the molecular basis and evolutionary history of the loss of benzaldehyde emission in the genus *Capsella*

Friederike Jantzen[1], Joseph H. Lynch[2,3], Christian Kappel[1], Jona Höfflin[4], Oded Skaliter[5], Natalia Wozniak[1], Adrien Sicard[1,6], Claudia Sas[1], Funmilayo Adebesin[2,3], Jasmin Ravid[5], Alexander Vainstein[5], Monika Hilker[4], Natalia Dudareva[2,3] and Michael Lenhard[1]

[1]Institute for Biochemistry and Biology, University of Potsdam, Karl-Liebknecht-Straße 24–25, D-14476, Potsdam-Golm, Germany; [2]Department of Biochemistry, Purdue University, 175 South University St., West Lafayette, IN 47907-2063, USA; [3]Purdue Center for Plant Biology, Purdue University, West Lafayette, IN 47907, USA; [4]Institute of Biology, Dahlem Centre of Plant Sciences (DCPS), Freie Universität Berlin, Haderslebener Straße 9, 12163, Berlin, Germany; [5]Robert H. Smith Faculty of Agriculture, Food and Environment, The Hebrew University of Jerusalem, PO Box 12, 76100, Rehovot, Israel; [6]Department of Plant Biology, Uppsala BioCenter, Swedish University of Agricultural Sciences and Linnean Center for Plant Biology, Uppsala, Sweden

Author for correspondence:
Michael Lenhard
Tel: +49 331 9775580
Email: michael.lenhard@uni-potsdam.de

Received: *1 March 2019*
Accepted: *18 July 2019*

New Phytologist (2019) **224:** 1349–1360
doi: 10.1111/nph.16103

Key words: benzaldehyde, *Capsella*, cinnamate-CoA ligase, evolution, floral scent, selfing syndrome, shepherd's purse.

Summary

- The transition from pollinator-mediated outbreeding to selfing has occurred many times in angiosperms. This is generally accompanied by a reduction in traits attracting pollinators, including reduced emission of floral scent. In *Capsella*, emission of benzaldehyde as a main component of floral scent has been lost in selfing *C. rubella* by mutation of cinnamate-CoA ligase CNL1. However, the biochemical basis and evolutionary history of this loss remain unknown, as does the reason for the absence of benzaldehyde emission in the independently derived selfer *Capsella orientalis*.
- We used plant transformation, *in vitro* enzyme assays, population genetics and quantitative genetics to address these questions.
- CNL1 has been inactivated twice independently by point mutations in *C. rubella*, causing a loss of enzymatic activity. Both inactive haplotypes are found within and outside of Greece, the centre of origin of *C. rubella*, indicating that they arose before its geographical spread. By contrast, the loss of benzaldehyde emission in *C. orientalis* is not due to an inactivating mutation in *CNL1*.
- *CNL1* represents a hotspot for mutations that eliminate benzaldehyde emission, potentially reflecting the limited pleiotropy and large effect of its inactivation. Nevertheless, even closely related species have followed different evolutionary routes in reducing floral scent.

Introduction

Functionally identical evolutionary transitions have frequently occurred independently in multiple lineages. Such convergent evolution raises the question to what extent the genetic and molecular basis of a given trait change is shared between independent events, and thus whether phenotypic evolution is predictable (Arendt & Reznick, 2008; Stern & Orgogozo, 2008; Stern, 2013). Answering this question promises not only to illuminate how many genetic routes are accessible to natural evolution for a given trait change, but may also identify the most promising candidate genes for engineering a trait of interest (Lenser & Theissen, 2013). Theory predicts that convergent evolution will tend to target genes with minimal pleiotropic effects, whose modification nevertheless maximizes the phenotypic consequences (Stern, 2013). Several examples for such 'hotspot' genes targeted by mutations in independent lineages have been found in animals (Arendt & Reznick, 2008; Stern & Orgogozo, 2008; Stern, 2013). By contrast, less is known about the molecular basis of convergent evolution in plants, although some examples have been identified, such as the *ANTHOCYANIN2* (*AN2*) locus controlling floral pigmentation in petunia (Esfeld *et al.*, 2018), or the *FRIGIDA* locus controlling the vernalization requirement in *Arabidopsis* (Shindo *et al.*, 2005).

One example of convergent evolution in plants is the transition from animal-mediated outbreeding to autogamous selfing and the concomitant evolution of the selfing syndrome (Sicard & Lenhard, 2011); this has occurred hundreds of times independently in angiosperms. The selfing syndrome comprises both a shift in sexual allocation from male towards female function and the reduction of many previous adaptations for pollinator attraction, such as a reduction in flower size, reduced opening of the flowers, less nectar and less scent production. The genus *Capsella*, which diverged from the *Arabidopsis* lineage *c.* 10–14 million years ago (Ma), provides a promising model to study the genetics of the selfing syndrome, as the transition from outbreeding to selfing has occurred twice independently within this genus (Hurka *et al.*, 2012). Approximately 1–2 Ma the self-compatible

Capsella orientalis diverged from a self-incompatible *C. grandiflora*-like ancestor (Hurka *et al.*, 2012; Bachmann *et al.*, 2019), and *c.* 100 000–200 000 yr ago the self-compatible *C. rubella* was derived from *C. grandiflora* (Foxe *et al.*, 2009; Guo *et al.*, 2009; Koenig *et al.*, 2018). In contrast to the outbreeding *C. grandiflora*, both selfing species form much smaller flowers and lack the strong floral scent of *C. grandiflora*.

Floral scent compounds can be grouped into four different classes based on their biochemical origin: phenylpropanoids/benzenoids, terpenoids, fatty acid derivatives and amino acid derivatives (Dudareva *et al.*, 2013). Benzenoids are ultimately derived from phenylalanine, which is converted by phenylalanine-ammonia lyase to *trans*-cinnamic acid (Supporting Information Fig. S1) (Dudareva *et al.*, 2013; Widhalm & Dudareva, 2015). Cinnamic acid can be converted to C_6–C_1 benzenoids including benzoic acid or benzaldehyde (BAld) via β-oxidative or non-β-oxidative pathways. While the steps in the non-β-oxidative pathway remain largely unknown, the β-oxidative pathway for benzoic acid synthesis has been fully elucidated (Van Moerkercke *et al.*, 2009; Klempien *et al.*, 2012; Qualley *et al.*, 2012) and is localized in peroxisomes. It starts with cinnamate:CoA ligase (CNL) catalysing the formation of cinnamoyl-CoA, which then undergoes a shortening of the three-carbon side chain by two carbons, ultimately resulting in benzoyl-CoA. Benzoyl-CoA can be exported from peroxisomes to the cytoplasm directly to form benzylbenzoate and phenylethylbenzoate, or hydrolysed by peroxisomal thioesterase before export of the resulting benzoic acid (Adebesin *et al.*, 2018) and its conversion to methylbenzoate.

BAld is one of the phylogenetically most widespread, and thus likely most ancient floral volatiles (Schiestl, 2010), and elicits diverse responses in different animal species (Schiestl, 2010). However, the origins of BAld remain unresolved, but could be three-fold (Widhalm & Dudareva, 2015) (Fig. S1). First, it could be formed from benzoyl-CoA produced by the β-oxidative pathway, via the action of a carboxylic acid reductase. The second route is CNL- and CoA-dependent, but non-β-oxidative, and produces BAld directly from cinnamoyl-CoA via cinnamoyl-CoA hydratase/lyase. Finally, BAld synthesis could proceed via the incompletely characterized CoA-independent non-β-oxidative path from cinnamic acid (Fig. S1) (Boatright *et al.*, 2004; Sheehan *et al.*, 2012; Dudareva *et al.*, 2013).

Previous work has identified loss-of-function mutations in *CNL* genes as responsible for the loss of BAld emission from the bird-pollinated *Petunia exserta* and selfing *C. rubella* (Amrad *et al.*, 2016; Sas *et al.*, 2016). While in *P. exserta* a premature stop codon was responsible for inactivating *CNL*, in *C. rubella* a missense mutation leading to an amino acid exchange close to a catalytically important region in CNL1 has been suggested as causal for its loss of activity. An independently derived *CNL1* loss-of-function allele was found in a North African accession of *C. rubella*. Together, these findings suggest *CNL* as a hotspot for mutations underlying the evolutionary loss of BAld emission, possibly reflecting the enzyme's position as the first committed step in the CoA-dependent pathways of benzenoid synthesis. However, the biochemical basis for the loss of *CNL1* activity in *C. rubella* has not been proven, and the origin of the second inactive allele has not been elucidated. Also, the basis of scent loss in the independently derived *C. orientalis* is unknown.

In the present work, we use a combination of genetic and biochemical analyses with population-genetic studies to address the following three questions: (1) What is the molecular and biochemical basis for the loss of CNL1 function in different *C. rubella* accessions? (2) What is the evolutionary history and the geographical distribution of different inactive *CNL1* haplotypes in *C. rubella*? (3) How has BAld emission been lost in the independently derived selfing species *C. orientalis* at the genetic and molecular level?

Materials and Methods

Plant material and growth conditions

All *C. grandiflora* (Fauché & Chaub.) Boiss., *C. rubella* Reut. and *C. orientalis* Klokov accessions used for the experimental studies have been described (Sicard *et al.*, 2011; Hurka *et al.*, 2012; Koenig *et al.*, 2018); their properties are summarized in Table S1. Lines used for transformation via floral dip (HIFgg/rr) have been described (Sicard *et al.*, 2011; Sas *et al.*, 2016); these are lines that were derived from crossing the two closest recombinants from our previous fine-mapping and should only differ in *c.* 10 kb around the *CNL1* locus, but should otherwise be essentially isogenic. Plants for scent measurements were grown in the glasshouse with supplemental lighting under a 16 h : 8 h, light : dark cycle. Temperature during the day was 21°C and during the night 16°C.

Scent phenotyping

Dynamic headspace collection To determine BAld emission from plants of the different *Capsella* genotypes, emitted floral volatiles of *Capsella* inflorescences were collected as described (Sas *et al.*, 2016), using dynamic headspace sampling adapted to small organ size. Details are described in Methods S1.

Internal pools To determine whether and if so how much BAld and other benzenoids the flowers of the different *Capsella* genotypes contained, internal pools of these compounds were analysed. For this, three replicates of at least 100 mg open flowers with sepals removed were harvested between 16:00 and 18:00 h. Flowers were extracted overnight at 4°C in 5 ml dichloromethane with naphthalene added as an internal standard. Samples were concentrated under nitrogen to *c.* 200 µl and analysed by GC-MS on an Agilent 7890B gas chromatograph (GC) (Agilent Technologies, Santa Clara, CA, USA) equipped with an HP-5MS column (30 m, 0.25 mm, 0.25 µm; Agilent Technologies) and coupled to an Agilent 5977B high-efficiency electro impact mass spectrometer (Agilent Technologies). A 1 µl aliquot of sample was injected at 1 : 10 split using a Gerstel cooled injector system (CIS4, Gerstel, Germany) with an injection gradient of $12° s^{-1}$ from 60°C to 280°C. Column temperature was held at 40°C for 0.5 min, then heated to 220°C (held for 1 min) at $8°C min^{-1}$. Helium was used as a carrier gas at a flow rate of

1 ml min^{-1}. MS ionization energy was set at 70 eV, and the mass spectrum was scanned from 50 to 550 amu.

Molecular cloning and plant transformation

To directly test the causal roles of individual polymorphisms in *CNL1* in plants, a number of constructs were generated and transformed into different *Capsella* genotypes to be used for volatile measurements. Details of the oligonucleotides used can be found in Table S2 and molecular cloning using SLiCE fusion cloning (Zhang *et al.*, 2012) is described in Methods S1.

Plant transformation was performed with *Agrobacterium tumefaciens* strain GV3101 using floral dip (Clough & Bent, 1998). The transformation medium contained 10% sucrose and 0.016% Silwet L-77. Dipping was repeated three times at 2–3 d intervals. Transgenic *Capsella* plants were selected on plates containing ½MS salts and 50 μM glufosinate-ammonium (phosphinothricin; Sigma-Aldrich). For each experiment, at least nine independent T1 transformants were genotyped and phenotyped in scent measurements (exact numbers are in the respective figures).

In vitro CNL enzyme assays

To test the effect of individual amino acid exchanges in the CNL1 protein on its biochemical activity, different versions were expressed as recombinant proteins and assayed *in vitro* for enzymatic activity. Enzymes were assayed for activity as described previously (Klempien *et al.*, 2012), and the cinnamoyl-CoA product was analysed by HPLC (Qualley *et al.*, 2012). Further details can be found in Methods S1.

Transient expression assay in petunia flowers

To test the activity of different variants of CNL1 protein in a complementary *in planta* system to *Capsella*, transient expression was performed in petunia flowers. *Petunia axillaris N* flowers were infiltrated at anthesis with *A. tumefaciens* AGL-0 containing either pBAR-35SΩ:Cr1504CNL1, pBAR-35SΩ:Cr1504CNL1rec or pRCS2-35S:PhCNL1 as described previously (Spitzer-Rimon *et al.*, 2012). Details of the analysis can be found in Methods S1.

QTL mapping

To study the genetic basis of the reduction of BAld emission in *C. orientalis*, we performed quantitative trait locus (QTL) mapping using a population derived from a cross of *C. grandiflora* with *C. orientalis*. Generation and genotyping of the mapping population derived from crossing *C. grandiflora* and *C. orientalis* is described in detail elsewhere (Wozniak *et al.*, 2019). Briefly, accession *Co1983* was crossed with *Cr4.23*, and *Co1983* with *Cg926* followed by ovule rescue as previously described (Sicard *et al.*, 2015). *C. orientalis* served as the female parent in both crosses. F1 plants were allowed to self, and phenotyping analysis in the F2 population was conducted using the progenies of one F1 plant each. Note that accession *Cr4.23* (not sequenced, but without BAld emission – see Fig. 1d) was used instead of *Cr1504* in the genetic mapping experiment because genetic incompatibility between *Co1983* and *Cr1504* strongly impaired hybrid performance (Sicard *et al.*, 2015). Genotyping was performed on 462 F2 individuals from the *Co* × *Cg* population using double digest restriction-site associated DNA sequencing (ddRAD-seq) (Peterson *et al.*, 2012) with digestion by *Eco*RI/*Msp*I, resulting in 1420 usable markers. A linkage map was constructed using R/ASMap (Taylor & Butler, 2017) with the Kosambi function. For phenotyping, head space samples were taken from five to seven plants per F3 family as pools, derived from the genotyped F2 plants. Due to limited capacity, only 165 randomly chosen F3 families were phenotyped, representing 165 F2 plants. Measurements were repeated on two consecutive days. The QTL for benzaldehyde emission was mapped using R/QTL (Broman *et al.*, 2003). The single-QTL genome scan was run using a nonparametric model. QTLs were tested in 1 cM intervals and assessed by using LOD scores (the $\log_{10}$–likelihood ratio comparing the hypothesis that there is a QTL at the marker to the hypothesis that there is no QTL). The QTL significance threshold was determined by running the permutation test under the nonparametric model with 1000 permutation replicates and by using the 5% and 10% significance cutoff.

Population-genetic analyses

To determine the evolutionary history of the causal *CNL1* mutations in Cr80TR1 discovered in the present work and study the geographical distribution of the different *CNL1* haplotypes across the range of *C. rubella*, we performed population-genetic analyses. Genome sequencing data for the 182 individuals from a *C. grandiflora* population in the Zagory mountains in Greece (PRJNA275635; Josephs *et al.*, 2015) and from a species-wide sample of 13 *C. grandiflora* individuals (PRJNA254516; Agren *et al.*, 2014; Douglas *et al.*, 2015) were downloaded from NCBI, those for *C. rubella* accessions from EBI ENA (PRJEB6689; Koenig *et al.*, 2018). Details are provided in Methods S1.

Statistical analysis

Statistical analysis was performed using R. Differences between means were compared using Student's *t*-test.

Results

Loss of BAld emission in *C. rubella* Cr1504 is due to a single amino acid exchange in CNL1

We had previously demonstrated that a mutation in the transcribed region of *CNL1* underlies the major QTL for the loss of BAld emission in *C. rubella* accession Cr1504 and several other accessions with the same *CNL1* haplotype (Sas *et al.*, 2016). In addition, we had identified another *CNL1* haplotype carrying an independent loss-of-function mutation – a 4 bp deletion – in accession CrTAAL-1-TS3 from northern Africa (hereafter termed

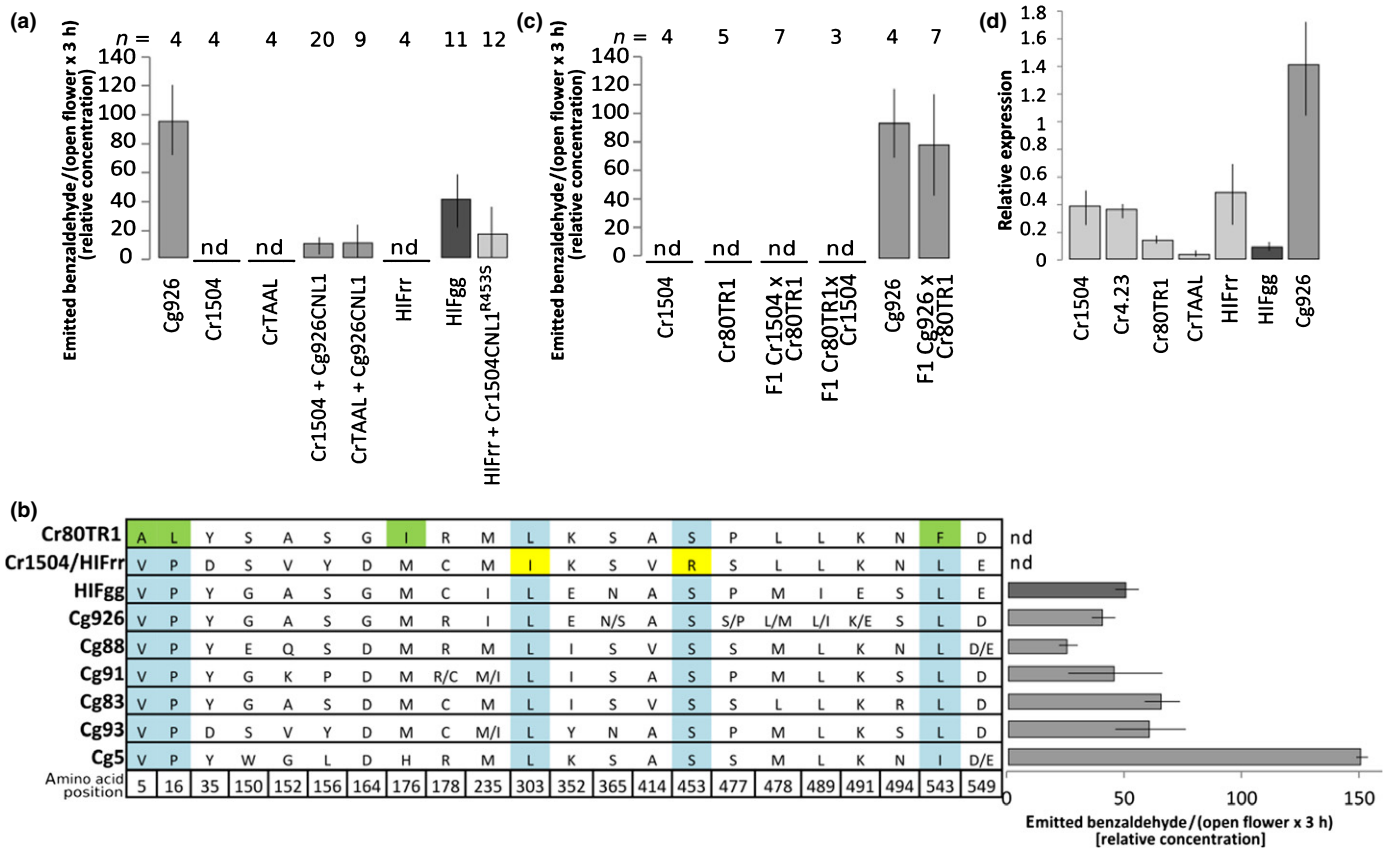

Fig. 1 Molecular analysis of different *Capsella* CNL1 alleles. (a) Quantification of emitted BAld in the headspace of the indicated accessions or transgenic plants. Values are mean ± SD from measurements of the indicated number of individuals or indicated number of independent transformants per construct, with three measurements on consecutive days from the same individuals. nd, Not detected. (b) Amino acid differences between Cr80TR1, Cr1504 and different *C. grandiflora* accessions. Green, mutations found exclusively in Cr80TR1; yellow, mutations found exclusively in Cr1504. Bars on the right show benzaldehyde (BAld) emission for the respective individuals. Values are mean ± SD from three technical replicates per individual. nd, Not detected. Relative concentration of BAld was normalized to 10 ng tridecane as an internal standard in the samples here and throughout the figures. (c) Quantification of BAld emission in a complementation test for Cr1504 and Cr80TR1. Cr80TR1 was crossed to Cg926 as a control. Maternal parents are named first for the F1 plants. Values are mean ± SD from measurements of the indicated number of individuals, with three measurements on consecutive days from the same individuals. (d) qRT-PCR-based quantification of *CNL1* expression in different *Capsella* accessions in flowers. Expression was normalized to the constitutively expressed *TUB* gene. Values are mean ± SD from three biological replicates.

CrTAAL). To confirm that the loss of *CNL1* function was the major cause for the loss of BAld emission from these accessions, we transformed both with a functional *CNL1* allele from *C. grandiflora* (Fig. 1a). This was sufficient to restore BAld emission in both backgrounds, albeit to a lesser extent than what is seen in *C. grandiflora*; thus, inactivation of *CNL1* is the main cause for the loss of BAld emission in these accessions, while additional minor mutations may underlie the reduced BAld levels obtained in the transformants relative to *C. grandiflora*.

Correlational evidence had suggested a single amino acid exchange, serine-to-arginine at position 453 close to a predicted AMP- and CoA-binding site, as the likely causal mutation in CNL1 from *C. rubella* accession Cr1504 (Fig. 1b). To test the role of this mutation directly, we performed complementary *in vivo* and *in vitro* experiments. For the former, we changed this codon in *Cr1504 CNL1* back to one encoding serine and expressed this version, termed *Cr1504 CNL1*^{R453S}, in a heterogeneous inbred family homozygous for the *Cr1504 CNL1* allele (HIFrr) (Sas *et al.*, 2016). These HIFs are lines that were derived from crossing the two closest recombinants from our previous fine-mapping and should only differ in *c.* 10 kb around the *CNL1* locus, but be essentially isogenic otherwise. In contrast to nontransgenic HIFrr plants, transformants expressing the *Cr1504 CNL1*^{R453S} allele emitted detectable levels of BAld, comparable to HIF plants homozygous for the *CNL1* allele from *C. grandiflora* Cg926 (HIFgg) (Fig. 1a). In addition, we transiently expressed *Cr1504 CNL1* and *Cr1504 CNL1*^{R453S} in *Petunia axillaris* flowers. In contrast to flowers expressing *Cr1504 CNL1* that only emitted low levels of methyl benzoate and benzyl benzoate (cf. Fig. S1), flowers that transiently expressed the reconstituted version emitted several-fold higher levels of these benzenoids, comparable to flowers expressing the *Petunia hybrida CNL* gene (Fig. S2). Thus, these results from two complementary *in vivo* settings demonstrate that reinstating the original serine at position 453 is sufficient to restore *CNL1* activity.

For the *in vitro* assays, we expressed the CNL1 proteins encoded by the following alleles in *Escherichia coli*: *Cr1504*

CNL1, *Cg926 CNL1*, *Cr1504 CNL1*^{R453S} and *Cg926 CNL1*^{S453R}; the last carries the Cr1504-like serine-to-arginine mutation at position 453 in an otherwise *Cg926 CNL1* background. The purified proteins were used for enzyme assays to test their cinnamate:CoA-ligase activity (Fig. S3) (Klempien *et al.*, 2012). While Cg926 CNL1 protein showed robust cinnamate:CoA-ligase activity, none was detected for the Cr1504 CNL1 protein (Table 1). Replacing serine 453 by arginine in the Cg926 CNL1^{S453R} protein abolished enzyme activity. Conversely, the Cr1504 CNL1^{R453S} protein, carrying the arginine 453 to serine mutation, again showed cinnamate:CoA-ligase activity, albeit at a lower level than the Cg926 CNL1 protein. These results provide direct biochemical evidence that the serine-to-arginine exchange at position 453 in CNL1 is sufficient to inactivate the enzyme and that this mutation explains the loss of *CNL1* function in the Cr1504 haplotype; nevertheless, the reduced activity of the Cr1504 CNL1^{R453S} compared to the Cg926 CNL1 protein suggests that additional mutations may also contribute to the altered enzymatic activity.

Evolutionarily independent amino acid exchanges in CNL1 underlie the loss of BAld emission in additional *C. rubella* accessions

While investigating the origin of the 4 bp pair deletion in *CNL1* of accession CrTAAL, we identified a further *C. rubella* accession, Cr80TR1-TS1 from Turkey (hereafter termed Cr80TR1), that carried the same *CNL1* haplotype as CrTAAL except for the 4 bp deletion. Thus, this allele could code for a full-length, potentially functional CNL1 protein. However, no BAld emission was detectable from Cr80TR1 inflorescences (Fig. 1b,c). We crossed Cr80TR1 to Cr1504 to test for complementation of the BAld-emission defect. F1 plants from this cross did not emit detectable levels of BAld, in contrast to F1 plants from a cross between Cg926 and Cr80TR1. This lack of complementation in the former F1 indicates that Cr80TR1 also carries an inactive *CNL1* allele.

Loss of *CNL1* activity could be due to *cis*-regulatory mutations causing a lack of expression or to missense mutations inactivating the protein. Although analysis of whole-genome sequencing data for Cr80TR1 identified a 114 bp deletion in its promotor at positions 7541 229–7541 342, upstream of the predicted transcriptional start site, *CNL1* expression levels in floral tissue of Cr80TR1 were comparable to those in Cr1504 or the HIFgg plants that robustly emit BAld (Fig. 1d). Thus, the promoter deletion is unlikely to explain the loss of *CNL1* activity in Cr80TR1; instead, alterations in protein primary structure are probably responsible.

Comparing the sequence of the predicted CNL1 protein from Cr80TR1 with those from other *C. rubella* and *C. grandiflora* accessions identified four amino acid exchanges that were unique to *C. rubella* Cr80TR1 (Fig. 1b). Of these, the exchange of a very highly conserved proline at position 16 for a leucine and a leucine-to-phenylalanine exchange at position 543 represent plausible candidates for the causal mutations (Fig. S4). The leucine-to-phenylalanine mutation is located close to a predicted CoA-binding region around amino acid 520 and replaces the aliphatic leucine with the bulky aromatic phenylalanine. We therefore tested the causal role of the two mutations using *in vitro* assays as above (Table 1). In particular, we expressed the following proteins: Cr80TR1 CNL1; Cg926 CNL1^{P16L} (leucine at position 16 in an Cg926 CNL1 background); Cg926 CNL1^{L543F} (phenylalanine at position 543); Cg926 CNL1^{P16L,L543F}; Cr80TR1 CNL1^{L16P} (proline at position 16 in an otherwise Cr80TR1 CNL1 background); Cr80TR1 CNL1^{F543L} (leucine at position 543); Cr80TR1 CNL1^{L16P,F543L}. As expected, the Cr80TR1 CNL1 protein did not show any detectable *in vitro* enzyme activity, and nor did the Cg926 CNL1^{P16L} and Cg926 CNL1^{P16L,L543F} proteins. The Cg926 CNL1^{L543F} protein retained some weak activity. Conversely, the reconstituted Cr80TR1 CNL1^{L16P} and Cr80TR1 CNL1^{L16P,F543L} proteins regained some and very strong CNL activity, respectively, while the Cr80TR1 CNL1^{F543L} protein did not show activity.

Thus, the proline-to-leucine mutation at position 16 is sufficient to inactivate the enzyme; however, to reconstitute high enzyme activity starting from the Cr80TR1 CNL1 protein requires changing both amino acids back to the *C. grandiflora* sequence. In which order these two mutations arose, and whether the L543F mutation represents secondary degeneration or the first mutational step towards abolishing CNL1 activity remains unknown.

Table 1 Michaelis–Menten kinetics for different CNL1 proteins

Protein	v_{max} (pkat mg^{-1})	K_M (μM)
Cg926 CNL1	27 337 ± 914	81 ± 11
Cr1504 CNL1	nd	nd
Cr1504 CNL1^{R453S}	3472 ± 216	105 ± 24
Cg926 CNL1^{S453R}	nd	nd
Cr80TR1 CNL1	nd	nd
Cg926 CNL1^{P16L}	nd	nd
Cg926 CNL1^{L543F}	498 ± 20	173 ± 23
Cg926 CNL1^{P16L,L543F}	nd	nd
Cr80TR1 CNL1^{L16P}	1693 ± 94	156 ± 29
Cr80TR1 CNL1^{F543L}	nd	nd
Cr80TR1 CNL1^{L16P,F543L}	16 015 ± 684	105 ± 14
Co1983CNL1	114 243 ± 5963	213 ± 33

Cinnamoyl-CoA ligase activities were determined using 25–1200 μM cinnamic acid, and using CoA-SH as the cosubstrate at a fixed concentration of 2 mM. Assays were performed in triplicate, and the results fit to the Michaelis–Menten equation. Results are shown as mean ± SEM.
nd, Not detectable; pkat, picokatal.

Evolutionary history of the *CNL1* mutations in *C. rubella*

Our previous analysis had indicated that the serine-to-arginine mutation in *Cr1504 CNL1* had probably arisen *de novo* in the *C. rubella* lineage, rather than having been captured from standing variation in *C. grandiflora* (Sas *et al.*, 2016). Having identified the causal mutations in the *CNL1* locus of Cr80TR1, we also sought to understand its origin. Also, a large number of additional *C. rubella* accessions have by now been sequenced (Koenig

et al., 2018), allowing us to study the geographical distribution of the different *CNL1* haplotypes in more detail.

To understand the origin of the inactivating mutations in Cr80TR1, we identified all polymorphisms between the *CNL1* haplotype from Cr80TR1 and that from *C. grandiflora Cg926* segregating in our original recombinant inbred line (RIL) population as a definitively active reference haplotype. We then determined the allelic state of all polymorphisms identified in this way across the resequenced *C. grandiflora* individuals (182 individuals from one population in Greece and 13 individuals from a species-wide sample) and *C. rubella* accessions and clustered the accessions (Fig. S5). This identified three clusters containing the *C. rubella* haplotypes, one corresponding to the *Cr1504 CNL1* haplotype, one to the Cr80TR1 haplotype and a small cluster of only four accessions with a third haplotype, whose functionality is currently unknown (Fig. S5; Table S3). The L543F mutation (nucleotide polymorphism at position 7 539 156; Fig. S5) was shared by all the *C. rubella* accessions in the Cr80TR1 cluster and the P16L mutation (nucleotide polymorphism at position 7540 987) was found in all but one of these accessions. However, neither mutation was found in any of the other *C. rubella* accessions, nor in any of the *C. grandiflora* accessions. The *C. grandiflora* accessions contained a polymorphism at position 7539 156; however, this only resulted in a conservative leucine-to-isoleucine exchange. Therefore, the absence of the inactivating mutations from all available *C. grandiflora* samples suggests strongly that both the P16L and the L543F mutations only arose as *de novo* mutations in the *C. rubella* lineage, rather than having been captured from standing variation in *C. grandiflora*. Incidentally, the 4 bp deletion was only seen in the single North African CrTAAL accession whose haplotype clustered with that of Cr80TR1.

Plotting the geographical distribution of the three *CNL1* clusters showed that all three are present in Greece, the presumed geographical origin and centre of diversity of *C. rubella*. Outside of Greece, the Cr1504-like haplotype was found throughout Western Europe, while the Cr80TR1-like haplotype was found in south-eastern Greece, Turkey and northern Africa (Fig. 2). The accession from Sicily carries a haplotype that appears to have resulted from recombination between the Cr1504 and the Cr80TR1 haplotypes (bottom-most in Fig. S5). This pattern suggests that the westward geographical spread of *C. rubella* that occurred *c.* 13 500 yr ago (Koenig *et al.*, 2018) carried with it the Cr1504-like haplotype, while the Cr80TR1-like haplotype spread mainly towards the south and east. The third cluster was restricted to Greece and Cyprus.

Loss of BAld emission in *C. orientalis* is not due to a mutation in *CNL1*

The genus *Capsella* contains another diploid selfing species, *C. orientalis*, that diverged from the *C. grandiflora*/*C. rubella* lineage more than 1 Ma. We investigated whether *C. orientalis* flowers still emit BAld and determined floral scent in five *C. orientalis* accessions from across its range in Central Asia (Table S4). No BAld emission was detected in two *C. orientalis* accessions tested, while the other three tested accessions still emitted low to very low levels, indicating that this scent compound has also been lost or strongly reduced in this species (Fig. 3a). We used several approaches to test whether the loss of BAld emission in *C. orientalis* was also due to inactivating mutations in the *CNL1* locus. First, we genotyped F2 plants from a cross between *C. grandiflora* and *C. orientalis* at the *CNL1* locus and measured BAld emission from their inflorescences. There was no difference in the average level of BAld emission between plants homozygous for the *C. orientalis CNL1* allele and plants carrying at least one copy of the *C. grandiflora CNL1* allele (Fig. 3b). This suggests that the *C. orientalis CNL1* allele remains functional.

Second, if this is the case, crossing non-BAld-emitting *C. rubella* with *C. orientalis* should result in complemented offspring that do emit BAld. To test this, we used F3 families derived from a cross of *C. rubella* Cr4.23 and *C. orientalis* Co1983, both of which do not emit BAld. The use of F3 families rather than F1 plants was necessary, as F1 plants can only be obtained via embryo rescue in small numbers, making their analysis impractical. We genotyped F2 plants from this cross for the *CNL1* locus, selected plants that were homozygous for either of the two parental alleles and collected emitted floral scent from five to seven F3 plants per F2 individual as a pool. As expected, when the F3 plants were homozygous for the *C. orientalis CNL1* allele, BAld emission was readily detectable (in 11 out of 12 tested F3 families), while no BAld was emitted from F3 plants homozygous for the nonfunctional *C. rubella CNL1* allele (Fig. 3c). This indicates both that *C. orientalis CNL1* is still functional and that the mutation(s) underlying the loss of BAld emission in *C. orientalis* affects one or more genes that remain functional in *C. rubella*, allowing for complementation in crosses between the two species. Furthermore, expression of *Co1983 CNL1* in HIFrr plants resulted in transgenic plants with partially restored BAld emission (Fig. S6a), consistent with a functional CNL1 allele.

Third, to test directly whether *C. orientalis* CNL1 protein retains enzymatic activity, we purified the protein from *E. coli* and used it for the above *in vitro* assay. Indeed, *C. orientalis* CNL1 protein showed high catalytic activity *in vitro* (Table 1), confirming our conclusions from the genetic experiments described above.

Thus, in summary, *C. orientalis* has lost BAld emission via a genetic route that is independent of the one seen in *C. rubella*, that is by one or more mutations to genes in the benzenoid pathway other than *CNL1*.

Loss of BAld emission in *C. orientalis* does not result from a single major-effect mutation in the benzenoid pathway

To identify the mutation(s) underlying the loss of BAld emission in *C. orientalis*, we used a genetic approach, exploiting the availability of 165 F3 families derived from genotyped F2 plants from a *C. grandiflora* × *C. orientalis* population. Floral volatiles were collected from five to seven plants of each of the 165 available F3 families as pools, with two replicate samples collected per family on consecutive days. The F3 families showed a continuous

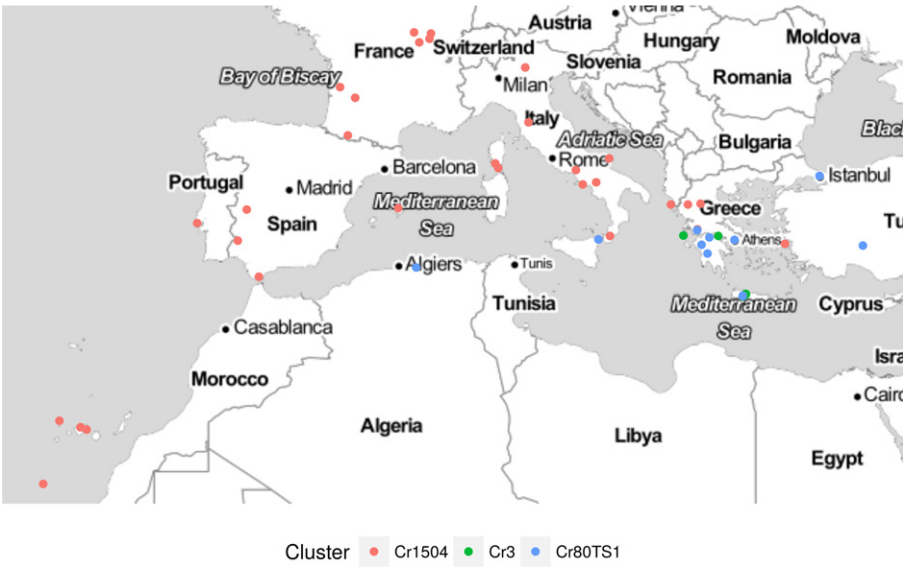

Fig. 2 Distribution of the three different *Capsella rubella* haplotypes around the Mediterranean. The Cr1504-like cluster (red dots) spread westward and the CR80TR1-like cluster (blue dots) east- and southward. The third, Cr3-like haplotype (green dots) is restricted to Greece, including Crete. All three haplotypes are present in Greece, the centre of diversity for this species.

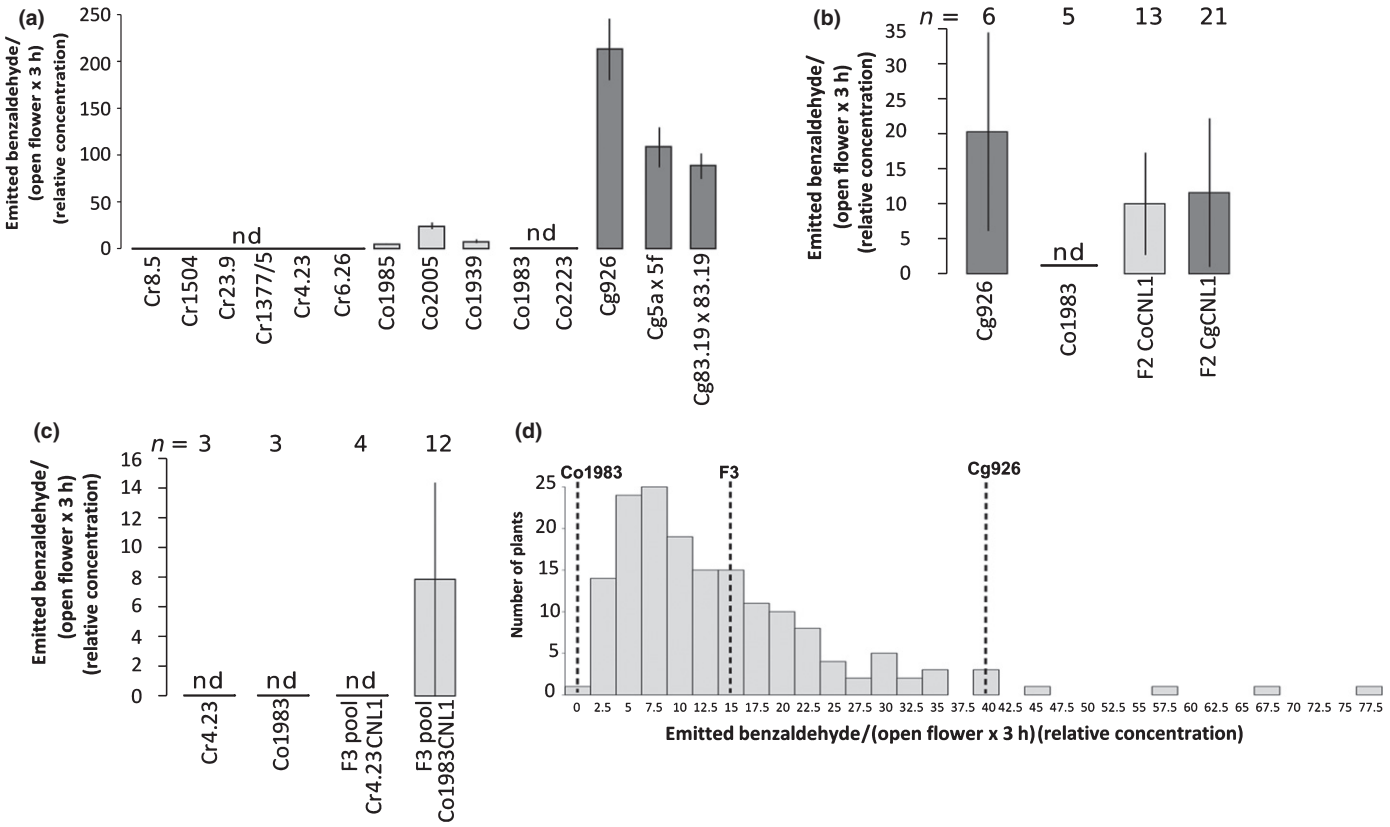

Fig. 3 Characterizing scent loss in *Capsella orientalis*. (a) Quantification of emitted BAld in the headspace of the indicated accessions. Values are mean ± SD from measurements of five individuals with three measurements on consecutive days from the same individuals. (b) BAld emission of parents and F2 plants of a Co1983 and Cg926 cross. F2 plants were genotyped for their *CNL1* allele. Values are mean ± SD from measurements of the indicated number of individuals, with two measurements on consecutive days from the same individuals. (c) BAld emission of parents and F3 plants of a Cr4.23 and Co1983 cross. F2 plants were genotyped for their *CNL1* allele and those homozygous for the parental alleles were selected for further investigations. Scent was collected for five to seven F3 plants per F2 individual as pools. Values are mean ± SD from measurements of the indicated number of pools, with two measurements on consecutive days from the same individuals selected for the pooled measurements. (d) Distribution of BAld emission in F3 pools of genotyped F2 plants from a Co1983 and Cg926 cross. Floral scent was collected from five to seven plants per F3 family as pools, with two measurements on consecutive days. Average values for parents and the whole F3 population are indicated. nd, Not detected.

distribution of the amounts of BAld emitted (Fig. 3d), with only one family showing no detectable BAld. This pattern is in stark contrast to that seen in a *C. grandiflora* × *C. rubella* RIL population, where a large number of families do not emit BAld. However, interpretation of the above results is complicated by the observation that the 165 F2 plants showed substantial segregation distortion (Fig. S7), with < 10% *C. orientalis* homozygotes for the most severely affected genomic regions. Still, if there was a major inactivating mutation as in *C. rubella* in a gene in such a strongly distorted region, one would predict *c.* 10 or more families without BAld emission amongst the phenotyped ones. As this was not seen, our results are instead consistent with mutations in at least two genes distinct from *CNL1* underlying the loss of BAld emission in the used *C. orientalis* accession. We used the phenotype values of the F3 pools in combination with the F2 genotypes for QTL mapping. However, this did not identify any statistically significant QTL (Fig. S6b). While the power of our analysis to detect QTL is limited by the number of F3 families phenotyped, this result is consistent with the continuous phenotype distribution across the F3 families in arguing for a polygenic basis for the loss of BAld emission in *C. orientalis* without major-effect mutations.

To investigate the reason for the loss of BAld emission in more detail, we investigated whether BAld was still synthesized in *C. orientalis*. Thus, we measured the internal pools of various benzenoids in open flowers (excluding sepals) of *C. grandiflora* and *C. orientalis*, and for comparison also in *C. rubella* (Fig. 4a). This analysis revealed that BAld was produced in *C. orientalis*, and that its internal pool was even larger than that in *C. grandiflora*. Even *C. rubella* still contained BAld internally, albeit at lower levels compared to *C. grandiflora*. (Note that in our previous study, we had not detected BAld in the internal pool of *C. rubella* (Sas *et al.*, 2016), yet this had been measured from entire inflorescences, rather than only mature, partly dissected flowers as done here, and the present study used a different, more sensitive analytical set-up.) In contrast to BAld, several other benzenoids that are not CNL-dependent were severely reduced in *C. orientalis*, including hydrocinnamic acid, benzoic acid and 2-phenylethanol, consistent with a generally lower biosynthetic potential for benzenoid-related compounds. The same three compounds were undetectable in the *C. rubella* internal pools. Thus, consistent with the finding that *CNL1* remains active in *C. orientalis*, flowers of this species still contain substantial levels of BAld in their internal pools, but fail to emit them at detectable levels, and have reduced levels of related compounds.

To test whether *C. orientalis* flowers are still able to emit BAld in principle, we aimed to boost BAld synthesis by feeding inflorescences with exogenous phenylalanine as the precursor to benzenoids. As expected, no BAld emission was detectable from *C. rubella* inflorescences even after phenylalanine feeding, consistent with the inactivation of CNL1 (Fig. 4b). By contrast, phenylalanine feeding caused robust BAld emission from *C. orientalis* inflorescences to comparable levels as seen in untreated *C. grandiflora*. In the latter, BAld emission was also strongly boosted by exogenous phenylalanine.

Thus, the capacity to synthesize and emit BAld is maintained in *C. orientalis*, although BAld emission is not detectable. It is possible that, for yet unknown reason(s), substrate is limiting in *C. orientalis* and therefore the emission phenotype can be restored through provision of excess substrate. Reduced transport capacity out of the cells, and/or compartmentalization of the produced BAld that renders it unavailable for emission may explain why the BAld that is synthesized is not emitted.

Discussion

Here we have addressed the molecular and biochemical basis and the evolutionary history of the loss of BAld emission in the two selfing species in *Capsella*. Our results indicate that the *CNL1* locus has been inactivated twice independently by *de novo* missense mutations that abolish enzyme activity in the *C. rubella* lineage, and that these two inactive haplotypes have spread along different geographical routes when *C. rubella* expanded its range out of Greece. By contrast, the loss of BAld emission in *C. orientalis* evolved without inactivation of *CNL1* and appears to be due to more than one smaller-effect mutations.

CNL1 as a target for parallel evolution of BAld loss

Which factors might explain the repeated loss of BAld emission via parallel mutations to *CNL1* in *C. rubella* and also in petunia, and can this repeated targeting of *CNL1* be rationalized based on the theory of convergent evolution? From a population-genetic perspective, three factors influence the probability of parallel evolution (Stern, 2013): the size of the mutational target; the net effect on fitness, which in turn depends on the extent of pleiotropy of mutations in a gene; and the magnitude of the phenotypic change. It is plausible that all three factors have contributed in the case of *CNL1*.

First, as loss of BAld emission can result from simple loss-of-function mutations in *CNL1*, and many missense or nonsense mutations cause such loss of function, *CNL1* presents a large mutational target for the loss of BAld emission. However, this is probably also true for other genes involved in BAld synthesis and emission. Second, mutations in CNL1 should only affect the C6–C1 branch of the phenylpropanoid path, thus limiting pleiotropic effects on other aspects of phenylpropanoid metabolism. Such limited pleiotropy of CNL1 mutations will increase their chance of fixation, irrespective of whether the loss of BAld is positively selected for or only occurs by genetic drift. Third, the phenotypic effect of mutations in CNL1 on BAld emission is large, given that it represents the first committed enzyme for the synthesis of BAld and other C6–C1 phenylpropanoids in the CoA-dependent pathways (Fig. S1). The same three aspects – a large mutational target for loss-of-function mutations, limited pleiotropy and a large phenotypic effect – hold true for several other targets of parallel evolution in plants, such as *AN2* and *FLAVONOID 3′5′ HYDROXYLASE* affecting flower pigmentation, or *FLOWERING LOCUS C* and *FRIGIDA* controlling flowering time (reviewed by Martin & Orgogozo, 2013). A similar case to *CNL1* is provided by the repeated

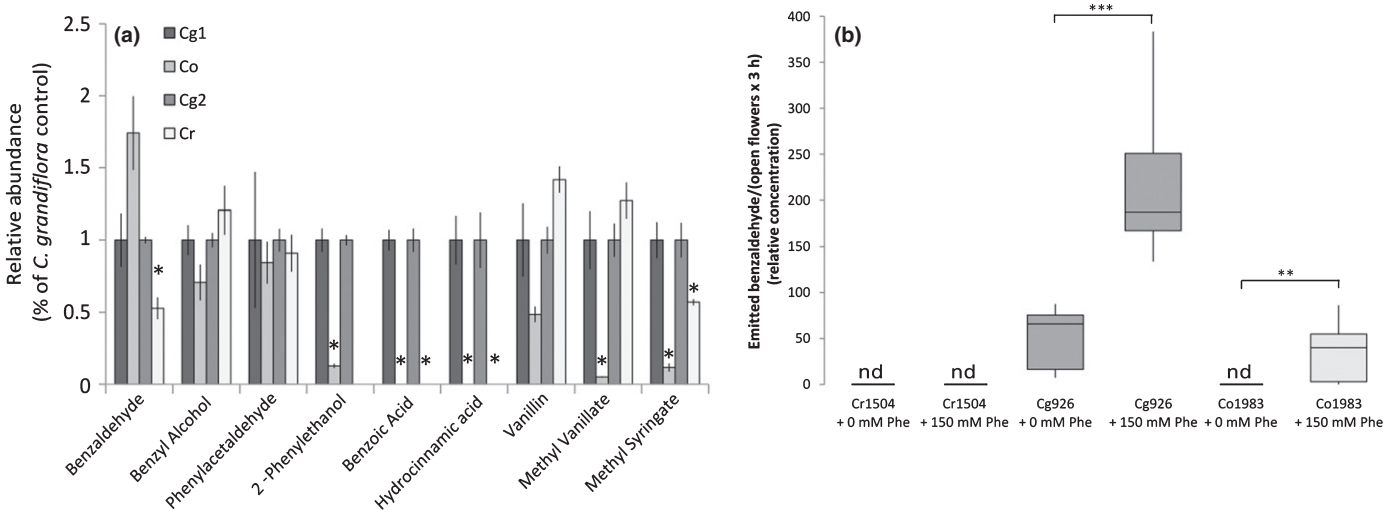

Fig. 4 Analysing scent production in *Capsella orientalis*. (a) Internal pool measurements for benzenoids in *Capsella*. *C. orientalis* 1983 (Co) and *C. rubella* 1504 (Cr) samples were each normalized to the respective *C. grandiflora* 926 samples analysed in parallel (Cg1 and Cg2, respectively). Three biological replicates each were analysed. Except for methyl syringate and hydrocinnamic acid, which are predictions based on mass spectra, the identities of all other compounds were confirmed by comparison to authentic standards. Values shown are mean ± SD. (b) Average BAld emission in phenylalanine-feeding experiment with three standard accessions. Feeding was performed by placing cut inflorescence stems in a solution of 150 mM phenylalanine and measuring BAld emission after 2 h of feeding. Box plots show the median value (black line), the 25 and 75 percentiles (bottom and top bounds of the box), and the top and bottom whiskers show the location of maximum and minimum. Scent was collected from five different samples per treatment and accession, and the experiment was repeated twice. One sample consisted of two to three inflorescences of the respective accession. Asterisks indicate statistically significant differences as determined by a Student's *t* test at: *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$. nd, Not detected.

inactivation of *OCIMENE SYNTHASE* in bird-pollinated and selfing *Mimulus* species, that shows the same hallmarks of loss-of-function mutations in key pathway genes with little pleitropic effects (Byers *et al.*, 2014; Peng *et al.*, 2017).

Thus, it is plausible that all three of these aspects also contributed to the repeated loss of BAld emission via mutations to *CNL1*, yet their relative contribution will require further study. Beyond this, it will be important to answer the question of whether loss of BAld emission is adaptive or only results from genetic drift after purifying selection on the gene has been removed in the selfing lineages.

Evolutionary history of *CNL1* mutations in *C. rubella*

Our population-genetic analyses presented here and previously (Sas *et al.*, 2016) indicate that the mutations inactivating *CNL1* in the two *C. rubella* haplotypes probably arose *de novo* in the *C. rubella* lineage, rather than having been captured from standing genetic variation in the ancestral outbreeding population. This conclusion is based on the observation that the causal mutations in the *Cr80TR1 CNL1* haplotype were not found in any of almost 200 re-sequenced *C. grandiflora* accessions, while the causal mutation in the *Cr1504 CNL1* haplotype was only found on long *C. rubella*-like haplotypes in a few *C. grandiflora* accessions, which probably resulted from more recent introgression, rather than representing ancestral variation (Sas *et al.*, 2016). For *Cr80TR1 CNL1* it is not clear in which sequence the two inactivating mutations arose, and thus whether the L543F mutation was the first step towards complete inactivation by the P16L mutation, or whether it merely represents secondary degeneration.

All three *C. rubella CNL1* haplotypes are found today in Greece, the centre of diversity and probable geographical origin of the species, and the region where its range continues to overlap with that of *C. grandiflora* (Koenig *et al.*, 2018). A recent population-genomic study of *C. rubella* defined two subpopulations based on genome-wide polymorphism data, an Eastern one found in Greece, Cyprus and Turkey, and a Western one found around the Mediterranean west of Greece, on the Cape Verde islands and beyond (Koenig *et al.*, 2018). Demographic modelling suggested that the geographical spread to the west occurred around 13 500 yr ago. These subpopulations are also reflected in the *CNL1* haplotypes, with the Western subpopulation containing the *Cr1504* haplotype, and the Eastern one the *Cr80TR1*-like and the rarer third haplotypes. The only discrepancy between the genome-wide assignment and our *CNL1* haplotype clusters is found in a North African accession, which was assigned to the Western subpopulation, but carries the *Cr80TR1* haplotype; this suggests that both haplotypes were initially carried along with the subpopulation spreading westward. The *C. rubella* accession from Sicily with its recombinant haplotype indicates residual outcrossing between *C. rubella* accessions before or during their spread.

The molecular basis of the loss of BAld emission in *C. orientalis*

CNL1 was targeted by repeated inactivating mutations in *C. rubella* and petunia. *C. rubella* and *C. orientalis* were both derived from ancestral outbreeding populations, albeit at very different times (*c.* 200 000 yr ago vs 1–2 Ma), and probably evolved

selfing in a similar manner as a mode of reproductive assurance. Therefore, it appeared very likely that the loss or reduction of BAld emission in *C. orientalis* would also be due to inactivation of CNL1. However, this is clearly not the case. The *C. orientalis CNL1* locus remains fully functional as demonstrated by our genetic and *in vitro* studies. Consistent with this notion, comparatively high levels of BAld were found in the internal pool of *C. orientalis*, and its emission could be induced by feeding *C. orientalis* with phenylalanine, probably boosting flux through the phenylpropanoid pathway. The results from a segregating population derived from a *C. grandiflora* × *C. orientalis* cross did not support the hypothesis that a single major-effect mutation abolished BAld emission in *C. orientalis*. Rather, the continuous phenotype distribution in this segregating population, the failure to detect significant QTLs for BAld emission in it, and the high frequency at which BAld emission was regained in families derived from a *C. rubella* × *C. orientalis* cross (as long as these did not carry the nonfunctional *C. rubella CNL1* allele) are all consistent with the notion that the lack or strong reduction of BAld emission in *C. orientalis* is due to more than one mutation with smaller individual effects. At least in some *C. orientalis* accessions, their combination appears to reduce the rate of emission and potentially flux through the pathway to levels that make BAld emission undetectable by our method. As emission is a biologically mediated process (Adebesin *et al.*, 2017), the lack of correlation between BAld emission and internal pools in *C. orientalis* vs *C. grandiflora* might indicate inactivating mutations in genes responsible for export of BAld out of the cell in addition to its biosynthesis. In petunia, growing plants at increased ambient temperatures reduces BAld synthesis and emission by down-regulation of many genes in the phenylpropanoid and shikimate pathways (Cna'ani *et al.*, 2015). Thus, future work should test whether growing *C. orientalis* at different conditions that may more closely resemble the environment in its native range can induce BAld emission. The mutations underlying the loss of BAld emission in *C. orientalis* also await further study, as does the question of why the genetic routes to the same phenotypic outcome differed between *C. rubella* and *C. orientalis*.

Recent experimental–evolution studies have shown that scent profiles of plants can respond quickly to divergent selection by different pollinators (Gervasi & Schiestl, 2017; Schiestl *et al.*, 2018). However, it remains a major unresolved question whether the loss of floral scent compounds seen in selfing species (Sicard & Lenhard, 2011; Peng *et al.*, 2017) merely represents degradation after purifying selection on the associated genes has been removed, or has an adaptive benefit, for example by making the plants less conspicuous to antagonists. A number of cases have been studied in which floral scent also attracts antagonists, such as florivores, herbivores or seed feeders (Theis, 2006; Andrews *et al.*, 2007; Theis & Adler, 2012; Kessler *et al.*, 2013; Theis *et al.*, 2014). In particular, BAld attracts female Western flower thrips that oviposit in and damage the flowers (Koschier *et al.*, 2000). Self-incompatible plants that are dependent on attracting pollinators for reproduction have evolved different solutions to this problem, for example the emission of complex scent mixtures, some components of which act to deter florivores (Kessler *et al.*, 2013). A simpler path would seem to be open to self-compatible species that no longer need to attract pollinators for reproduction and thus could simply lose the emission of floral scent. This would provide a scenario for adaptive loss of BAld emission in *Capsella* as emission is an energetically expensive process. Our work provides the genetic resources to begin answering these questions using field experiments.

Acknowledgements

We thank Doreen Mäker and Christiane Schmidt for plant care, Cris Kuhlemeier for seeds of *P. axillaris* N, and Orit Edelbaum for help with petunia experiments. We thank members of the Lenhard lab and Isabel Bäurle for discussion. This work was funded by grant no. G-1310-203.13/2015 from the German-Israeli Foundation for Scientific Research and Development (GIF) to ML, MH and AV, by grant SI1967/1-1 from the Deutsche Forschungsgemeinschaft to AS, and by grant IOS-1655438 from National Science Foundation to ND.

Author contributions

AV, MH, ND and ML designed the research. FJ, JHL, JH, CK, OS, NW, AS, CS, FA and JR performed the research. FJ, JHL, CK, OS, FA, ND and ML analysed and interpreted data. FJ and ML wrote the manuscript with input from all authors.

ORCID

Funmilayo Adebesin https://orcid.org/0000-0002-1288-0944
Natalia Dudareva https://orcid.org/0000-0003-0777-7763
Monika Hilker https://orcid.org/0000-0001-7519-7395
Christian Kappel https://orcid.org/0000-0002-1450-1864
Michael Lenhard https://orcid.org/0000-0001-8661-6911
Joseph H. Lynch https://orcid.org/0000-0003-0685-1562
Adrien Sicard https://orcid.org/0000-0002-4104-6844
Alexander Vainstein https://orcid.org/0000-0002-2667-3081
Natalia Wozniak https://orcid.org/0000-0002-9131-0666

References

Adebesin F, Widhalm JR, Boachon B, Lefevre F, Pierman B, Lynch JH, Alam I, Junqueira B, Benke R, Ray S *et al.* 2017. Emission of volatile organic compounds from petunia flowers is facilitated by an ABC transporter. *Science* **356**: 1386–1388.

Adebesin F, Widhalm JR, Lynch JH, McCoy RM, Dudareva N. 2018. A peroxisomal thioesterase plays auxiliary roles in plant beta-oxidative benzoic acid metabolism. *The Plant Journal* **93**: 905–916.

Agren JA, Wang W, Koenig D, Neuffer B, Weigel D, Wright SI. 2014. Mating system shifts and transposable element evolution in the plant genus *Capsella*. *BMC Genomics* **15**: 602.

Amrad A, Moser M, Mandel T, de Vries M, Schuurink RC, Freitas L, Kuhlemeier C. 2016. Gain and loss of floral scent production through changes in structural genes during pollinator-mediated speciation. *Current Biology* **26**: 3303–3312.

Andrews ES, Theis N, Adler LS. 2007. Pollinator and herbivore attraction to cucurbita floral volatiles. *Journal of Chemical Ecology* **33**: 1682–1691.

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

Bachmann JA, Tedder A, Laenen B, Fracassetti M, Désamoré A, Lafon-Placette C, Steige KA, Callot C, Marande W, Neuffer B *et al.* 2019. Genetic basis and timing of a major mating system shift in *Capsella*. *New Phytologist* **224**: 505–517.

Boatright J, Negre F, Chen X, Kish CM, Wood B, Peel G, Orlova I, Gang D, Rhodes D, Dudareva N. 2004. Understanding *in vivo* benzenoid metabolism in petunia petal tissue. *Plant Physiology* **135**: 1993–2011.

Broman KW, Wu H, Sen S, Churchill GA. 2003. R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **19**: 889–890.

Byers KJRP, Vela JP, Peng F, Riffell JA, Bradshaw HD. 2014. Floral volatile alleles can contribute to pollinator-mediated reproductive isolation in monkeyflowers (*Mimulus*). *The Plant Journal* **80**: 1031–1042.

Clough SJ, Bent AF. 1998. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *The Plant Journal* **16**: 735–743.

Cna'ani A, Muhlemann JK, Ravid J, Masci T, Klempien A, Nguyen TT, Dudareva N, Pichersky E, Vainstein A. 2015. *Petunia* × *hybrida* floral scent production is negatively affected by high-temperature growth conditions. *Plant, Cell & Environment* **38**: 1333–1346.

Douglas GM, Gos G, Steige KA, Salcedo A, Holm K, Josephs EB, Arunkumar R, Agren JA, Hazzouri KM, Wang W *et al.* 2015. Hybrid origins and the earliest stages of diploidization in the highly successful recent polyploid *Capsella bursa-pastoris*. *Proceedings of the National Academy of Sciences, USA* **112**: 2806–2811.

Dudareva N, Klempien A, Muhlemann JK, Kaplan I. 2013. Biosynthesis, function and metabolic engineering of plant volatile organic compounds. *New Phytologist* **198**: 16–32.

Esfeld K, Berardi AE, Moser M, Bossolini E, Freitas L, Kuhlemeier C. 2018. Pseudogenization and resurrection of a speciation gene. *Current Biology* **28**: 3776–3786 e3777.

Foxe JP, Slotte T, Stahl EA, Neuffer B, Hurka H, Wright SI. 2009. Recent speciation associated with the evolution of selfing in *Capsella*. *Proceedings of the National Academy of Sciences, USA* **106**: 5241–5245.

Gervasi DDL, Schiestl FP. 2017. Real-time divergent evolution in plants driven by pollinators. *Nature Communications* **8**: 14691.

Guo YL, Bechsgaard JS, Slotte T, Neuffer B, Lascoux M, Weigel D, Schierup MH. 2009. Recent speciation of *Capsella rubella* from *Capsella grandiflora*, associated with loss of self-incompatibility and an extreme bottleneck. *Proceedings of the National Academy of Sciences, USA* **106**: 5246–5251.

Hurka H, Friesen N, German DA, Franzke A, Neuffer B. 2012. 'Missing link' species *Capsella orientalis* and *Capsella thracica* elucidate evolution of model plant genus *Capsella* (Brassicaceae). *Molecular Ecology* **21**: 1223–1238.

Josephs EB, Lee YW, Stinchcombe JR, Wright SI. 2015. Association mapping reveals the role of purifying selection in the maintenance of genomic variation in gene expression. *Proceedings of the National Academy of Sciences, USA* **112**: 15390–15395.

Kessler D, Diezel C, Clark DG, Colquhoun TA, Baldwin IT. 2013. Petunia flowers solve the defence/apparency dilemma of pollinator attraction by deploying complex floral blends. *Ecology Letters* **16**: 299–306.

Klempien A, Kaminaga Y, Qualley A, Nagegowda DA, Widhalm JR, Orlova I, Shasany AK, Taguchi G, Kish CM, Cooper BR *et al.* 2012. Contribution of CoA ligases to benzenoid biosynthesis in petunia flowers. *Plant Cell* **24**: 2015–2030.

Koenig D, Hagmann J, Li R, Bemm F, Slotte T, Nueffer B, Wright S, Weigel D. 2018. Long-term balancing selection drives evolution of immunity genes in *Capsella*. *eLife* **8**: e43606.

Koschier EH, De Kogel WJ, Visser JH. 2000. Assessing the attractiveness of volatile plant compounds to western flower thrips *Frankliniella occidentalis*. *Journal of Chemical Ecology* **26**: 2643–2655.

Lenser T, Theissen G. 2013. Molecular mechanisms involved in convergent crop domestication. *Trends in Plant Science* **18**: 704–714.

Martin A, Orgogozo V. 2013. The loci of repeated evolution: a catalog of genetic hotspots of phenotypic variation. *Evolution* **67**: 1235–1250.

Peng F, Byers KJRP, Bradshaw HD. 2017. Less is more: independent loss-of-function *OCIMENE SYNTHASE* alleles parallel pollination syndrome diversification in monkeyflowers (*Mimulus*). *American Journal of Botany* **104**: 1055–1059.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. 2012. Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PLoS ONE* **7**: e37135.

Qualley AV, Widhalm JR, Adebesin F, Kish CM, Dudareva N. 2012. Completion of the core beta-oxidative pathway of benzoic acid biosynthesis in plants. *Proceedings of the National Academy of Sciences, USA* **109**: 16383–16388.

Sas C, Muller F, Kappel C, Kent TV, Wright SI, Hilker M, Lenhard M. 2016. Repeated inactivation of the first committed enzyme underlies the loss of benzaldehyde emission after the selfing transition in *Capsella*. *Current Biology* **26**: 3313–3319.

Schiestl FP. 2010. The evolution of floral scent and insect chemical communication. *Ecology Letters* **13**: 643–656.

Schiestl FP, Balmer A, Gervasi DD. 2018. Real-time evolution supports a unique trajectory for generalized pollination. *Evolution* **72**: 2653–2668.

Sheehan H, Hermann K, Kuhlemeier C. 2012. Color and scent: how single genes influence pollinator attraction. *Cold Spring Harbor Symposia on Quantitative Biology* **77**: 117–133.

Shindo C, Aranzana MJ, Lister C, Baxter C, Nicholls C, Nordborg M, Dean C. 2005. Role of FRIGIDA and FLOWERING LOCUS C in determining variation in flowering time of *Arabidopsis*. *Plant Physiology* **138**: 1163–1173.

Sicard A, Kappel C, Josephs EB, Lee Y-W, Marona C, Stinchcombe JR, Wright SI, Lenhard M. 2015. Divergent sorting of a balanced ancestral polymorphism underlies the establishment of gene-flow barriers in *Capsella*. *Nature Communications* **6**: 7960.

Sicard A, Lenhard M. 2011. The selfing syndrome: a model for studying the genetic and evolutionary basis of morphological adaptation in plants. *Annals of Botany* **107**: 1433–1443.

Sicard A, Stacey N, Hermann K, Dessoly J, Neuffer B, Baurle I, Lenhard M. 2011. Genetics, evolution, and adaptive significance of the selfing syndrome in the genus *Capsella*. *Plant Cell* **23**: 3156–3171.

Spitzer-Rimon B, Farhi M, Albo B, Cna'ani A, Ben Zvi MM, Masci T, Edelbaum O, Yu YX, Shklarman E, Ovadis M *et al.* 2012. The R2R3-MYB-like regulatory factor EOBI, acting downstream of EOBII, regulates scent production by activating ODO1 and structural scent-related genes in *Petunia*. *Plant Cell* **24**: 5089–5105.

Stern DL. 2013. The genetic causes of convergent evolution. *Nature Reviews Genetics* **14**: 751–764.

Stern DL, Orgogozo V. 2008. The loci of evolution: how predictable is genetic evolution? *Evolution* **62**: 2155–2177.

Taylor J, Butler D. 2017. R package ASMap: efficient genetic linkage map construction and diagnosis. *Journal of Statistical Software* **79**: 1–29.

Theis N. 2006. Fragrance of Canada thistle (*Cirsium arvense*) attracts both floral herbivores and pollinators. *Journal of Chemical Ecology* **32**: 917–927.

Theis N, Adler LS. 2012. Advertising to the enemy: enhanced floral fragrance increases beetle attraction and reduces plant reproduction. *Ecology* **93**: 430–435.

Theis N, Barber NA, Gillespie SD, Hazzard RV, Adler LS. 2014. Attracting mutualists and antagonists: plant trait variation explains the distribution of specialist floral herbivores and pollinators on crops and wild gourds. *American Journal of Botany* **101**: 1314–1322.

Van Moerkercke A, Schauvinhold I, Pichersky E, Haring MA, Schuurink RC. 2009. A plant thiolase involved in benzoic acid biosynthesis and volatile benzenoid production. *The Plant Journal* **60**: 292–302.

Widhalm JR, Dudareva N. 2015. A familiar ring to it: biosynthesis of plant benzoic acids. *Molecular Plant* **8**: 83–97.

Wozniak NJ, Kappel C, Marona C, Altschmied L, Neuffer B, Sicard A. 2019. A common molecular basis to the convergent evolution of the selfing syndrome in *Capsella*. *bioRxiv*: 653139.

Zhang Y, Werling U, Edelmann W. 2012. SLiCE: a novel bacterial cell extract-based DNA cloning method. *Nucleic Acids Research* **40**: e55.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 Overview of the phenylpropanoid/benzenoid pathway in *Petunia*.

Fig. S2 Benzenoid emission after transient expression in *P. axillaris N*.

Fig. S3 HPLC assays of recombinant CNL enzymes.

Fig. S4 Multiple sequence alignment of CNL1 protein sequences from different organisms.

Fig. S5 Variant clustering of diploid CNL1 haplotypes in *C. grandiflora* and *C. rubella* accessions.

Fig. S6 Genetic basis of BAld loss in *C. orientalis*.

Fig. S7 Segregation distortion among F2 families used for F3 phenotyping.

Methods S1 Details on experimental procedures.

Table S1 Accessions and genotypes used in the different experiments.

Table S2 Primers used in this study.

Table S3 Origin of *C. rubella* accessions and assignment to *CNL1* haplotype clusters.

Table S4 Origin of *C. orientalis* accessions.

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.