Genetic Basis and Adaptive Significance of Repeated Scent Loss in Selfing *Capsella* Species

Friederike Jantzen

Dissertation

zur Erlangung des akademischen Grades

„doctor rerum naturalium“ (Dr. rer. nat.)

in der Wissenschaftsdisziplin Biologie

eingereicht an der

Mathematisch-Naturwissenschaftlichen Fakultät

Institut für Biochemie und Biologie

der Universität Potsdam

Potsdam, April 2019

Hauptbetreuer: Prof. Dr. Michael Lenhard

Gutachter: Prof. Dr. Alexander Vainstein

Prof. Dr. Florian P. Schiestl

Contents

Summary 1

Zusammenfassung 3

List of figures 5

List of tables 6

Abbreviations 7

1. Introduction 9

1. 1 Evolution of the Selfing Syndrome 9

1. 2 The role of floral scent (loss) 16

1. 3 Evolution of flower size 20

1. 4 Evolutionary history of *Capsella* 22

1. 5 Aims of this work 26

2. Manuscript: Retracing the molecular basis and evolutionary history of the loss of benzaldehyde emission in the genus *Capsell*a 27

2. 1 Summary 29

2. 2 Introduction 30

2. 3 Material and Methods 33

2. 3. 1 Plant material and growth conditions 33

2. 3. 2 Scent phenotyping 33

2. 3. 3 Molecular cloning and plant transformation 34

2. 3. 4 *In vitro* CNL enzyme assays 34

2. 3. 5 Transient expression assay in petunia flowers 35

2. 3. 6 QTL mapping 36

2. 3. 7 Population-genetic analyses 36

2. 3. 8 Statistical analysis 36

2. 4 Results 37

2. 4. 1 Loss of BAld emission in *C. rubella* Cr1504 is due to a single amino-acid exchange in CNL1 37

2. 4. 2 Evolutionarily independent amino-acid exchanges in CNL1 underlie the loss of BAld emission in additional *C. rubella* accessions 40

2. 4. 3 Evolutionary history of the *CNL1* mutations in *C. rubella* 41

2. 4. 4 Loss of BAld emission in *C. orientalis* is not due to a mutation in *CNL1* 43

2. 4. 5 Loss of BAld emission in *C. orientalis* results from polygenic downregulation of the benzenoid pathway 45
2. 5 Discussion 47
2. 5. 1 *CNL1* as a target for parallel evolution of BAld loss 47
2. 5. 2 Evolutionary history of *CNL1* mutations in *C. rubella* 48
2. 5. 3 The molecular basis of the loss of BAld emission in *C. orientalis* 48
2. 6 Acknowledgements 50
2. 7 Author Contributions 50
2. 8 References 51
2. 9 Supporting information 53
2. 9. 1 Supplemental Figures 53
2. 9. 2 Supporting Tables 60
2. 10 Transcriptome analysis of genes involved in the benzenoid pathway 63
3. Manuscript: A high-throughput amplicon-based method for estimating outcrossing rates 66
3. 1 Abstract 68
3. 2 Background 69
3. 3 Methods 71
3. 3. 1 Reagents 71
3. 3. 2 Equipment 71
3. 3. 3 Plant material and growth conditions 72
3. 3. 4 DNA extraction from pooled seeds 72
3. 3. 5 Design of PCR primers 73
3. 3. 6 PCR amplification and library generation 75
3. 3. 7 Amplicon sequencing 79
3. 3. 8 Sequence analysis and estimation of outcrossing rates 80
3. 4 Results 81
3. 5 Discussion 88
3. 6 References 91
3. 7 Declarations 93
3. 8 Supplemental Information 94
4. Identifying a candidate gene underlying PAQTL1 96
4. 1 Introduction 96

4. 2 Material and Methods 98
4. 2. 1 Plant Material and growth conditions 98
4. 2. 2 Molecular cloning 98
4. 2. 3 Morphological measurements and statistical analysis 98
4. 2. 4 Oligonucleotides 99
4. 3 Results 100
4. 3. 1 Confirming the position of PAQTL1 100
4. 3. 2 Identifying a candidate gene 101
4. 4 Discussion and Outlook 105
5. Concluding remarks 107
6. Additional Material and Methods 112
6.1 Chemicals 112
6.2 Disposable equipment 112
6.3 Microorganisms 113
6.4 Methods 113
6.4.1 Plant-related methods 113
6.4.2 DNA-related methods 114
6.4.3 RNA-related methods 115
Appendix: Oligonucleotide and vector lists 116
Acknowledgements 118
References 119
Affidavit 124

Summary

Floral scent is an important way for plants to communicate with insects, but scent emission has been lost or strongly reduced during the transition from pollinator-mediated outbreeding to selfing. The shift from outcrossing to selfing is not only accompanied by scent loss, but also by a reduction in other pollinator-attracting traits like petal size and can be observed multiple times among angiosperms. These changes are summarized by the term selfing syndrome and represent one of the most prominent examples of convergent evolution within the plant kingdom. In this work the genus *Capsella* was used as a model to study convergent evolution in two closely related selfers with separate transitions to self-fertilization.

Compared to their outbreeding ancestor *C. grandiflora,* the emission of benzaldehyde as main compound of floral scent is lacking or strongly reduced in the selfing species *C. rubella* and *C. orientalis*. In *C. rubella* the loss of benzaldehyde was caused by mutations to cinnamate:CoA ligase CNL1, but the biochemical basis and evolutionary history of this loss remained unknown, together with the genetic basis of scent loss in *C. orientalis*. Here, a combination of plant transformations, *in vitro* enzyme assays, population genetics and quantitative genetics has been used to address these questions. The results indicate that CNL1 has been inactivated twice independently by point mutations in *C. rubella*, leading to a loss of benzaldehyde emission. Both inactivated haplotypes can be found around the Mediterranean Sea, indicating that they arose before the species´ geographical spread. This study confirmed CNL1 as a hotspot for mutations to eliminate benzaldehyde emission, as it has been suggested by previous studies. In contrast to these findings, CNL1 in *C. orientalis* remains active. To test whether similar mechanisms underlie the convergent evolution of scent loss in *C. orientalis* a QTL mapping approach was used and the results suggest that this closely related species followed a different evolutionary route to reduce floral scent, possibly reflecting that the convergent evolution of floral scent is driven by ecological rather than genetic factors.

In parallel with studying the genetic basis of repeated scent loss a method for testing the adaptive value of individual selfing syndrome traits was established. The established method allows estimating outcrossing rates with a high throughput of samples and detects successfully insect-mediated outcrossing events, providing major advantages regarding time and effort compared to other

approaches. It can be applied to correlate outcrossing rates with differences in individual traits by using quasi-isogenic lines as demonstrated here or with environmental or morphological parameters.

Convergent evolution can not only be observed for scent loss in *Capsella* but also for the morphological evolution of petal size. Previous studies detected several QTLs underlying the petal size reduction in *C. orientalis* and *C. rubella*, some of them shared among both species. One shared QTL is PAQTL1 which might map to *NUBBIN,* a growth factor. To better understand the morphological evolution and genetic basis of petal size reduction, this QTL was studied. Mapping this QTL to a gene might identify another example for a hotspot gene, in this case for the convergent evolution of petal size.

Zusammenfassung

Blütenduft stellt einen wichtigen Kommunikationsweg für Pflanzen mit ihrer Außenwelt dar, jedoch ging die Emission von Duftstoffen während des Übergangs von Bestäuber-vermittelter Auskreuzung zur Selbstung verloren oder wurde stark reduziert. Dieser Wandel des Fortpflanzungssystems wurde nicht nur vom Duftverlust begleitet, sondern auch von einer Reduktion anderer Merkmale, die Bestäuber anlocken, so zum Beispiel auch einer verringerten Blütenblattgröße. Diese veränderten Merkmale werden unter dem Begriff Selbstungssyndrom zusammengefasst und können mehrfach in der Evolution von Angiospermen beobachtet werden. Der Übergang von Auskreuzung zu Selbstung und die damit einhergehende Veränderung von Blüten stellt eines der bekanntesten Beispiele für konvergente Evolution im Pflanzenreich dar. In dieser Arbeit wurde die Gattung *Capsella* als Modell für die Untersuchung konvergenter Evolution genutzt, da sie zwei eng verwandte selbstbestäubende Arten enthält, die unabhängig voneinander den Übergang zur Selbstbefruchtung vollzogen.

Verglichen mit dem auskreuzendem Vorfahr *C. grandiflora*, ist die Emission von Benzaldehyd als Hauptbestandteil des Blütenduftes in den selbstbestäubenden Arten *C. rubella* und *C. orientalis* verloren gegangen oder stark reduziert. Der Verlust von Benzaldehyd wurde in *C. rubella* durch Mutationen in der Cinnamat:CoA Ligase CNL1 verursacht, die biochemische Grundlage und evolutionäre Geschichte dieses Verlustes waren jedoch unbekannt, ebenso wie die genetische Grundlage für den Duftverlust von *C. orientalis*. In der vorliegenden Arbeit wurde eine Kombination aus transformierten Pflanzen, *in vitro* Enzymassays, Populationsgenetik und quantitativer Genetik genutzt, um diese Fragen zu beantworten. Die Ergebnisse sprechen dafür, dass in *C. rubella* CNL1 zweimal unabhängig voneinander durch Punktmutationen inaktiviert wurde und diese zu einem Verlust der Benzaldehydemission führten. Beide inaktive Haplotypen können in Populationen rund um das Mittelmeer detektiert werden, was dafürspricht, dass die Mutationen auftraten, bevor sich die Art geografisch ausbreiten konnte. Diese Arbeit bestätigt, dass CNL1 ein sogenannter Hotspot für die Eliminierung der Benzaldehydemission ist, wie es schon von vorhergehenden Arbeiten angedeutet wurde. Im Gegensatz dazu verbleibt CNL1 in *C. orientalis* funktionsfähig. Um herauszufinden, ob bei dieser Art ähnliche Mechanismen der konvergenten Evolution zum Duftverlust führten, wurde ein QTL-Kartierungsansatz gewählt. Die Ergebnisse sprechen dafür, dass hier die Duftreduktion auf einem evolutionär anderen Weg erreicht

wurde als in *C. rubella*, eventuell spiegelt dies wider, dass die Evolution von Blütenduft eher durch ökologische als genetische Faktoren angetrieben wird.

Parallel zur Aufklärung der genetischen Basis von wiederholtem Duftverlust wurde eine Methode zum Testen des adaptiven Wertes einzelner Merkmale des Selbstungssyndromes etabliert. Diese Methode erlaubt es, Auskreuzungsraten mit einem hohen Probendurchsatz abzuschätzen, und es konnte gezeigt werden, dass sie erfolgreich Auskreuzungsereignisse detektiert, die von Insekten vermittelt wurden. Diese neue Methode bringt einige Vorteile gegenüber bisher angewandten Protokollen mit sich; sie ist wesentlich schneller und eine Vielzahl an Proben lässt sich mit deutlich geringerem Aufwand analysieren. Dies ermöglicht eine Korrelation der Auskreuzungsrate mit unterschiedlichen Merkmalen, indem, wie hier demonstriert, quasi-isogene Linien verwendet werden. Denkbar wäre auch eine Kombination mit anderen morphologischen oder ökologischen Parametern, um deren Einfluss auf die Auskreuzungsrate zu untersuchen.

Konvergente Evolution kann nicht nur beim Duftverlust in der Gattung *Capsella* beobachtet werden, sondern auch bei der morphologischen Evolution der Petalengröße. Vorhergehende Studien fanden mehrere QTL, die der Reduktion der Petalengröße in *C. rubella* und *C. orientalis* unterliegen, manche von ihnen sind sogar in beiden Spezies zu finden. Einer der gemeinsamen QTL ist PAQTL1, der vielleicht auf *NUBBIN* zurückgeführt werden kann, einem Wachstumsfaktor. Um die morphologische Evolution und die genetische Grundlage der Reduktion der Petalengröße besser zu verstehen, soll dieser QTL weiter charakterisiert werden, da ihm eventuell ein Hotspot-Gen für die konvergente Evolution der Petalengröße unterliegen könnte.

List of figures

Figure 1.1: Origin of self-fertilization in *Amsinckia* 11
Figure 1.2: Parallel loss of floral scent in *Capsella* and *Petunia* 19
Figure 1.3: Flowers of *C. grandiflora* and *C. rubella* 21
Figure 1.4: Evolutionary history of the genus *Capsella.* 23
Figure 1.5: Distribution map of *Capsella* species. 23
Figure 2.1: Molecular analysis of different *Capsella* CNL1 alleles. 37
Figure 2.2: Distribution of the three different *C. rubella* haplotypes in the Mediterranean. 42
Figure 2.3: Characterizing scent loss in *C. orientalis*. 44
Figure 2.4.: Analyzing scent production in *C. orientalis*. 46
Figure 2.5: Expression levels of putative *Capsella* homologues for phenylpropanoid/benzenoid pathway genes and additional genes involved in BAld emission. 65
Figure 2.S1: Overview over phenylpropanoid/benzenoid pathway in *Petunia*. 53
Figure 2.S2: Benzenoid emission after transient expression in *P. axillaris N.* 54
Figure 2.S3: HPLC assays of recombinant CNL enzymes. 54
Figure 2.S4: Multiple sequence alignment of CNL1 protein sequences from different organisms. 56
Figure 2.S5: Variant clustering of diploid CNL1 haplotypes in *C. grandiflora* and *C. rubella* accessions. 58
Figure 2.S6: Genetic basis of BAld loss in *C. orientalis.* 59
Figure 3.1: Plot set-up for common garden experiment. 82
Figure 3.2: Primer-test for individual primers and pooling strategy. 83
Figure 3.3: Examples for qPCR amplification traces. 83
Figure 3.4: Structure of the final amplicon. 83
Figure 3.5: Example of a final library as analyzed with TapeStation. 84
Figure 3.6: Non-parental haplotype frequencies across the amplicons. 86
Figure 3.7: Frequency of the two alternative parental haplotypes in reads for amplicon 6 (Carubv10023818m). 87
Figure 4.1: Genetic map based on the *C. grandiflora* x *C. rubella* RIL population. 97
Figure 4.2: Confirming the region for PAQTL1. 100
Figure 4.3: Narrowing down the position of PAQTL1. 101

List of tables

Table 2.1: Michaelis-Menten kinetics for different CNL1 proteins..... 39
Table 2.S1: Oligonucleotides used in this study..... 60
Table 2.S2: Origin of *C. rubella* accessions and assignment to *CNL1* haplotype clusters. 61
Table 3.1: Primers for Primary PCR. 73
Table 3.2: Primers for Indexing PCR. 75
Table 3.S1: Example indexing scheme. 94
Table 3.S2: Read statistics for the amplicons. 94
Table 4.1: Oligonucleotides for mapping PAQTL1. 99
Table 4.2: Oligonucleotides used for cloning. 99
Table 4.3: Annotated genes or transcripts from *C. rubella* in the 100 kb interval. 102
Table A1: Oligonucleotides used for qRT-PCR. 116
Table A2: Vectors generated and used for this work. 116

Abbreviations

AN2	ANTHOCYANIN2
BA	Benzoic acid
BAld	Benzaldehyde
BALDH	Benzaldehyde dehydrogenase
BPBT	Benzoyl-CoA:benzylalcohol/2-phenylethanol benzoyltransferase
BSMT	Benzoic acid/salicylic acid carboxyl methyltransferase
C4H	Cinnamate 4-hydroxylase
CNL1	Cinnamate:CoA ligase
CoA	Coenzyme A
Cg	*Capsella grandiflora*
Co	*Capsella orientalis*
Cr	*Capsella rubella*
DNA	Deoxyribonucleic acid
EGS	Eugenol synthase
EOBII	EMISSION OF BENZENOIDS II
GC-MS	Gas chromatography-mass spectrometry
GSI	Gametophytic self-incompatibility
HIF	Heterogeneous inbred family
HPLC	High performance liquid chromatography
IGS	Isoeugenol synthase
LMS	Limonene-myrcene synthase
MRCA	Most recent common ancestor
NIL	Nearly-isogenic line
ODO1	ODORANT1
OS	Ocimene synthase

PA	Petal area
PAAS	Phenylacetaldehyde synthase
PAL	Phenylalanine ammonia lyase
QTL	Quantitative trait locus
RIL	Recombinant inbred line
RNA	Ribonucleic acid
SAP	STERILE APETALA
SC	Self-compatibility
SCR	*S*-locus cysteine-rich protein
SI	Self-incompatibility
SLF	*S*-locus F-box protein
SLiCE	Seamless ligation cloning extract
SNP	Single nucleotide polymorphisms
SRK	*S*-locus receptor kinase
SSI	Sporophytic self-incompatibility
TPS02	Terpene synthase 2
VOCs	Volatile organic compounds

1. Introduction

1. 1 Evolution of the Selfing Syndrome

Flowering plants (Angiosperms) are the most successful plant group on earth. In contrast to gymnosperms, they maximized their reproductive success by evolving flowers. Their species richness is at least partly based on diversification driven by interaction with their pollinators (Stebbins, 1970). Being immobile, they rely on different mechanisms for pollination as pollen vectors enable mating amongst individuals. Depending on the pollinator, plants show different adaptations to reach high rates of fertilization by outcrossing. Pollination often occurs through abiotic factors such as water and wind or by animals as biotic factors. Plant-pollinator interaction is a heavily studied field and a lot of examples for co-evolution have been investigated, especially when it comes to plants and their adaptations to animal pollinators. For example the pollination mode seems to be correlated with floral scent, as demonstrated by Farré-Armengol et al. (2015). This study compared floral emissions of wind-pollinated (anemophilous) and insect-pollinated (entemophilus) species and showed that insect-pollinated species have significantly higher emissions of volatile organic compounds (VOCs), especially terpenes and benzenoids. Even within a relatively short time frame changes in floral traits through different pollinator communities can be detected as shown by a quite recent study. Real-time divergent evolution in plants driven by pollinators has been observed in *Brassica rapa*; plants pollinated by bumblebees evolved a taller stature and more fragrant flowers within few generations only (Gervasi and Schiestl, 2017).

Sets of specialized traits that attract the same pollinators across different angiosperm species are thought to be a result of convergent evolution and are summarized by the term pollination syndrome. This includes floral scent and size as mentioned already, but also flower color (visible and UV), morphology and nectar volume and composition. Studying the genetic basis of these traits is of particular interest for many researchers, as gaining more insights into the function of genes may elucidate their role in the evolution of pollination syndrome traits.

One of the most prominent models to study plant-insect interaction is the genus *Petunia*. The species in this genus exhibit different pollination syndromes and the genetic basis and adaptive value of floral traits such as color and scent have been studied in more detail. For example, experiments with *Manduca*

sexta moths showed a significant preference for scented *Petunia* lines over non-scented lines (Klahre et al., 2011). These lines only differed in one QTL for scent, *ODORANT1*.

Also in other species a high adaptation to pollinators is documented. Nunes et al. (2017) showed that there may be convergent evolution in two different orchid species, *Catasetum cernuum* and *Gongora bufonia*, to a unique pollination niche. Both species share the emission of two VOCs, (*E*)-β-ocimene and (*E*)-epoxyocimene, which probably play a major role in the specific association of these orchids and their exclusive, very specialized pollinator, while emitting different scent compounds otherwise. In orchids the adaption to solitary Hymenoptera and other relatively unspecialized insects is thought to be the primitive condition (Stebbins, 1970), but *C. cernuum* and *G. bufonia* are pollinated by a unique species of male fragrance-seeking orchid bees.

It is widely recognized that most features of flowers influence patterns of pollen dispersal, and outcrossing is seen as the ancestral state of mating systems. Through outcrossing, heterozygosity can be maintained and this protects plants from the expression of deleterious recessive alleles. It also increases the chances of overdominant allelic interactions that can contribute to heterosis (Charlesworth and Willis, 2009). Angiosperms are not only adapted to the promotion of outcrossing, but also evolved self-incompatibility (SI) systems to avoid self-fertilization. These systems can be found in about 40% of angiosperm species and in most cases, they are controlled by one polymorphic locus, the *S*-locus, containing the male and female specificity component (Fujii et al., 2016). Usually these mechanisms are based on a self- and nonself-recognition process between pollen and pistil, followed by inhibition of self-pollen growth. Two main types of SI have been identified: sporophytic SI (SSI) and gametophytic SI (GSI). SSI was identified first in Brassicaceae and there the pollen SI phenotype is determined by the diploid genome of the parent (sporophyte). If sporophyte and gametophyte carry the same *S*-locus, the pollen grains express a small ligand peptide (SCR – *S*-locus cysteine-rich protein) that is recognized by the *S*-locus receptor kinase (SRK) expressed in the pistil. The interaction of the peptide and the kinase activates a phosphorylation cascade that leads to self-pollen rejection. Non-self pollen grains carry different SCRs, which are not recognized by the sporophytic kinase and can fertilize the ovules (Fujii et al., 2016).

GSI has been characterized best in Solanaceae. Here, an S-RNase is expressed in the pistil and degrades RNAs in the pollen tube of self-pollen, thereby stopping pollen tube growth. If the pollen is non-self, the *S*-locus F-box proteins (SLF) expressed in the pollen tube can bind the S-RNAse and lead to degradation of this RNAse; as a result, the pollen tube can continue growing (Fujii et al., 2016).

SI mechanisms are still under investigation and more and more factors are being identified. Self-incompatibility is one of the most important systems to prevent selfing and different molecular mechanisms have evolved several times in the history of angiosperms to promote outcrossing (Takayama and Isogai, 2005). Still, one main disadvantage of obligate outcrossers is the reliance on mates and pollinators for reproductive success.

In situations when the number of mating partners or pollinators is limited, for example when seeds get dispersed widely and a colonization of new habitats is possible, a transition from outcrossing to selfing can be beneficial. This transition is actually one of the most frequent changes in the evolutionary history of angiosperms. It has been observed that selfing has evolved multiple times independently (Barrett, 2002), even within the same genus. For example in the genus *Amsinckia*, selfing evolved at least four times as illustrated in Figure 1.1 (Schoen et al., 1997). Without the necessity to attract pollinators, floral traits do not need to be as showy and a reduction in e. g. petal size and scent emission were observed, possibly due to a reallocation of the plants´ energy to other processes with fitness benefit (Brunet, 1992). Goodwillie et al. (2010) showed support for this by analyzing different datasets for outcrossing rates of higher plants and came to the following result: the outcrossing rate is positively associated with the product of flower size and number. This leads to the conclusion that flowers could be smaller, if a high outcrossing rate does not need to be maintained.

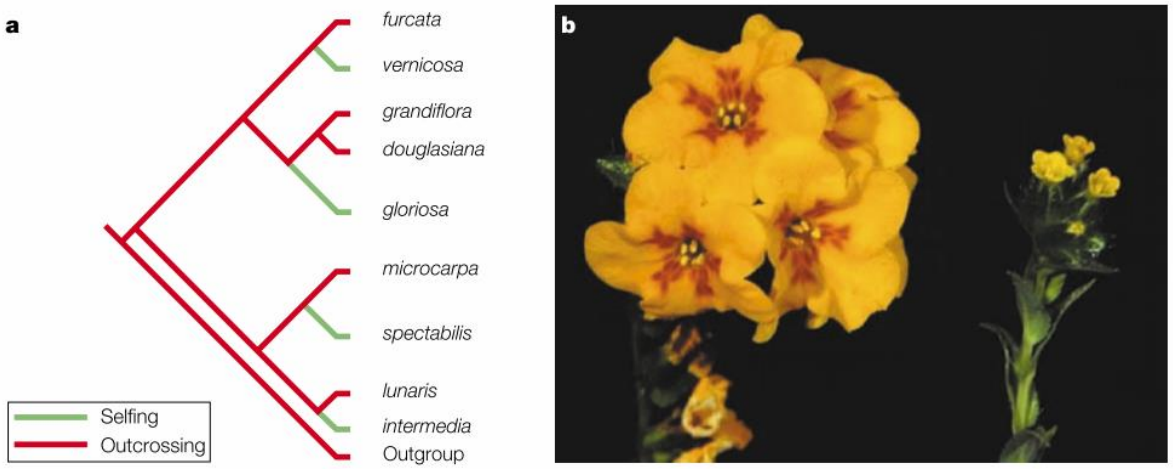

Figure 1.1: Origin of self-fertilization in *Amsinckia*, selfing arose four times independently as shown by phylogenetic reconstructions based on restriction-site variation in the chloroplast genome. Outcrossing species have larger flowers than selfers as shown by photo in b. Modified from Barrett (2002).

The change in floral size can be observed in multiple selfing species and is usually accompanied by other morphological and functional changes (Orndruff, 1969). These changes, not only limited to floral size and a reduction in scent emission, but also including a reduced opening angle of flowers (resulting in a reduced distance between anthers and stigma for more efficient self-pollination), a reduced pollen-to-ovule ratio and a reduction in nectar production, are termed the selfing syndrome, following the idea of pollination syndromes (Sicard and Lenhard, 2011). The selfing syndrome is a prominent example of convergent evolution in angiosperms and comparative studies between selfers and their outbreeding sister taxa can be used to test theoretical predictions about genetic evolution.

It is thought that the phenotypic convergence among selfing species suggests a close relationship between flower morphology and selfing efficiency, making it likely that conserved genetic limitations for flower development exist. The genetic networks controlling developmental traits are usually very complex and have an influence on evolutionary targets (Stern and Orogozo, 2008). It seems more feasible that genes with low pleiotropy but high phenotypic effect (like genes mediating between upstream patterning factors and downstream growth regulators) are altered by evolutionary changes than genes that have an influence on upstream processes. Evolutionary re-targeting of genes mediating floral development and traits may explain partially the phenotypic convergence observed among selfing species that arose independently (Stern, 2013).

For a transition from obligate outcrossing to predominant selfing a breakdown of self-incompatibility or a modification of other traits that prevent self-fertilization is thought to be necessary first (Schoen et al., 1997; Barrett et al., 2014). It has been hypothesized that the evolution of SC was often caused by a mutation in the male determinant of the SI system, as pollen grains then can fertilize ovules with the same haplotype, but also all other plants within the population, leading to a faster spread of mutated male alleles compared to mutations in female components of the system, which can only spread through seeds (Tsuchimatsu and Shimizu, 2013). This has been shown for *Arabidopsis thaliana* accessions across Europe by Tsuchimatsu et al. (2010). In 95% of the analyzed accessions the male determinant is mutated and complementation experiments suggest that all female components remain functional, confirming theoretical predictions.

If self-fertilization has evolved within a population, it will lead to an increase in homozygosity and limit the consequences of recombination. As a result, the expression of recessive mutations will increase as well as linked selection and the influence of genetic drift (Barrett et al., 2014). As a consequence, shifts toward high selfing rates are often accompanied by a strong reduction in genetic diversity and a

'population bottleneck' as shown by Slotte et al. 2013 for *Capsella rubella*, a derived selfer. The reduced effective population size is expected to increase the importance of drift versus selection, which can lead to the accumulation of deleterious mutations and could peak in the extinction of a species. Therefore, the evolution of selfing depends on a balance between reproductive assurance (and the advantage of being independent) and the intensity of inbreeding depression by accumulating recessive deleterious mutations (Woźniak and Sicard, 2018).

The comparison of outcrossing and selfing species has attracted a lot of attention; already Charles Darwin studied floral traits as adaptations to promote cross-pollination in plants (Darwin, 1876). Charlesworth et al. (1990) showed that inbreeding depression can be very large in moderately selfing populations, but in more inbred populations inbreeding depression is lower and selection favors alleles that increase the selfing rate. Furthermore, it has been hypothesized that deleterious recessive alleles leading to inbreeding depression can be purged by selection: through selfing, their homozygosity and expression increases within the population, and they thus become visible to selection, potentially leading to their elimination. After purging, inbreeding depression is thought to be too low to overcome the advantages of selfing and a reversion to outcrossing seems very unlikely (Lande and Schemske, 1984; Glémin and Ronfort, 2013). However, in Solanaceae these predictions could not be confirmed, because self-incompatible species diversify at a significantly higher rate than self-compatible plants (Goldberg et al., 2010). This suggests that the short-term advantage of potentially selfing individuals is offset by strong species selection that favors obligate outcrossing (Goldberg et al., 2010). However, a lower genetic diversity in small, isolated populations may limit the chances to get fertilized by a compatible allele and self-incompatibility increases the chance of extinction. It seems very likely that in such scenarios selfing could arise and provide an escape.

Probably different dynamics led to the independent evolution of self-compatible plant species in different angiosperm families. Concerning the contrasting effects of obligate outcrossing or predominant selfing on plant fitness it has been hypothesized that these two different mating systems were most likely stable end points of mating-system evolution in plants (Lande and Schemske, 1984; Barett, 2003).

The transition to selfing in hermaphrodite flowers is generally associated with the selfing syndrome, and the similarities in morphological evolution raised a lot of interest. One main question is whether the mutations inducing these changes were fixed, because they conferred a fitness benefit, or whether they were a result of the reduced efficiency of purging selection in selfers, maybe even including a fitness

disadvantage. Identifying the genetic basis for these morphological changes may answer this question, but will also give insights into demographic contexts and the selective mechanisms that were involved in the evolution of selfing lineages (Barrett et al., 2014). If reproductive assurance was the main driver for this shift in plant mating systems, gene flow was very likely limited at the origin of a new species and new mutations would be expected to underlie the evolution of the selfing syndrome, rather than being captured from standing genetic variation within the ancestral populations (Wozniak and Sicard, 2018). Once a high selfing rate is established, the contribution of exogenous pollen to genetic variation is unlikely, and therefore the genetic basis of traits that act as mating system modifiers improving the selfing rate in early steps of adaptation will be most informative. An important point to consider is the correlation of the evolution of these mating system modifying traits and their adaptive value. Adaptive mutations are thought to be the fuel of evolution and appearance of novel genotypes with reproductive advantages allow populations to persist when facing threatening changes (Hartfield et al., 2017).

Arguing for an adaptive value of the selfing syndrome are the morphological changes that can be observed between independent transitions to selfing. Remarkably, plants from different genera across the whole angiosperm kingdom show these, even if it does not seem to correlate with the time of selfing evolution (Wozniak and Sicard, 2018). Two different scenarios are thought to be responsible for the convergence of these changes: either the selfing syndrome traits were established through a relaxation of selective pressures on floral traits for pollinator attraction or they are based on the adaptive advantage of resource allocation. Maybe even an interaction of both resulted in the selfing syndrome traits that can be observed today (Wozniak and Sicard, 2018). To gain better insights and prove the predictions made by theoretical studies the genetic basis of the selfing syndrome evolution is studied.

To date, some quantitative genetic studies have been conducted and these partly support the view that large effect mutations underlie the breakdown of self-incompatibility and the distance between anther and stigma, while many smaller- to moderate-effect loci may be responsible for traits such as flower size and pollen-to-ovule ratio (Goodwillie et al., 2006; Sicard, et al., 2011; Slotte et al., 2012; Wozniak, 2019). It has been noted that the number of underlying mutations seems to be larger in some systems than others. So far, data on dominance is limited, but the dominance relationship of mutations seems to confirm a predicted higher dominance level of mutations having a strong influence on the selfing rate, such as mutations to the *S*-locus, than on traits which have a more indirect influence on selfing efficiency (Sicard and Lenhard, 2011; Wozniak and Sicard, 2018). To further prove theoretical predictions, it is necessary to not only have quantitative genetic studies, but also elucidate the

molecular function of different mutations underlying the selfing syndrome trait evolution and backtrace their evolutionary history to see if they arose from standing genetic variation within the ancestral population or arose *de novo*. One would expect that the potential for adaptation from standing genetic variation is reduced in selfers, as theory predicts a strong reduction of genetic diversity after the transition to selfing (Barett et al., 2014; Hartfield et al., 2017).

One model system to answer the raised questions and prove predictions made about the genetic basis of the selfing syndrome is the genus *Capsella*. Within this genus, outcrossing and selfing species can be found and selfing arose at least twice independently. *Capsella grandiflora* is an outcrossing species with genetic self-incompatibility to prevent selfing. *Capsella rubella* and *Capsella orientalis* are two derived selfers that show well-established selfing syndrome traits (Foxe et al., 2009; Hurka et al., 2012). Several quantitative genetic studies examined the evolution of selfing syndrome traits and their role in population genetics in this genus (Sicard et al., 2011; Slotte et al., 2012; Bachmann et al., 2018; Wozniak, 2019).

Furthermore, morphological evolution is being studied in *Capsella* and some genes that contribute to the relevant trait changes have been identified already. For example, two different genetic factors regulating petal size have been described. Sicard et al. (2016) showed that the small petals in *Capsella rubella* are partially caused by a specific reduction of the activity of a general growth factor in petals. Fujikura et al. (2018) demonstrated that a variation in splicing efficiency of a brassinosteroid-biosynthesis enzyme resulted in a higher level of brassinosteroids, and that these restrict petal growth by limiting cell proliferation. Not only for petal size, but also for scent reduction in *Capsella rubella* a genetic factor has been identified by Sas et al. (2016). They showed that the inactivation of one enzyme is responsible for the loss of the main compound of floral scent in *Capsella*.

In this work, the evolution of selfing syndrome traits in *Capsella* was studied for a better understanding of convergent evolution and its genetic basis.

1. 2 The role of floral scent (loss)

Floral scent and color are important attractants for insects, pollinators and herbivores, but our knowledge about volatile production is poor compared to what is known about flower pigment formation (Raguso, 2008; Sheehan et al., 2012). Nevertheless, different approaches and techniques have been developed to characterize scent compounds and also understand the biosynthetic pathways for scent production. As floral fragrances are also used in the perfume industry and many food crop species rely on insects for pollination (Klein et al., 2006), unraveling plants´ biosynthesis of volatile organic compounds (VOCs) generates broad interest.

Scent is a channel of communication between plants and their environment. A variation in compounds, the ratio of compounds and the time of emission can influence the perception of scent signals (Raguso 2008). To determine the chemical composition of floral fragrances the dynamic headspace collection method was developed. To collect the emitted volatiles from a flower still connected to the plant, the flowers are put in an almost fully closed container and air gets drawn out through a filter, where volatiles are trapped. After a set amount of time the volatiles get eluted from the filter and can be analyzed using gas chromatography, which separates the volatiles by size, and mass spectrometry, which is used to identify the different volatiles (gas chromatography-mass spectrometry, GC-MS). With the help of these investigations it has been shown that floral scents are often a complex mixture of small volatile molecules. The dominant groups that were detected are terpenoids, phenylpropanoid and benzoid compounds. Other chemicals like fatty acid derivates or molecules containing nitrogen or sulfur could also be detected (Knudsen et al., 1993; Dudareva and Pichersky, 2000). Floral scent compounds are very diverse and a good overview is provided by Knudsen et al. (1993, 2006). In the past years, more and more studies tried to elucidate the role of individual compounds in insect attraction. For example, methyl benzoate, eugenol and benzyl alcohol elicit positive behavioral responses in different bee-species (Dötterl and Vereecken, 2010).

It has been shown by Sas et al., (2016), that the main compound of floral scent in *Capsella grandiflora* is benzaldehyde, a product of the phenylpropanoid/benzenoid pathway. Also, benzaldehyde is the phylogenetically most widespread benzenoid and it is therefore likely to be one of the most ancient volatiles emitted by flowers (Schiestl, 2010). Due to the ubiquitous distribution of benzenoids in scent bouquets this pathway has been studied in more detail in different species. Some regulating factors and enzymes have been identified as shown in Figure 2.S1. The MYB transcription factor *ODORANT1* (ODO1)

is a main regulator of this pathway and the amino acid phenylalanine is the precursor for its products (Verdonk et al., 2005). Enzymes of downstream reactions have been partially identified but others remain unknown (dashed lines in Figure 2.S1). The cinnamate:CoA ligase CNL1 is of particular interest for this work as it catalyzes the first committed step for benzoic acid (BA) synthesis via the peroxisomal CoA-dependent β-oxidative pathway in *Petuni*a (Klempien et al., 2012). In *Capsella* it is thought to have a role in an alternative pathway, the CoA-dependent non-β-oxidative pathway leading to benzaldehyde synthesis (Sas et al., 2016; Abd El-Mawla et al., 2002). As illustrated by Figure 1.2, a mutation in CNL1 leads to a reduced emission of scent compounds in both species. This mutation of the same gene for the loss of benzenoid production is a remarkable example for parallel evolution (Amrad et al., 2016; Raguso, 2016).

Only few scent-related transcriptional regulators are identified to date. Four transcription factors have been shown to play important roles for scent emission in Petunia, which is widely used as model plant to study floral scent emission. As mentioned above, ODO1 is a main regulator of the phenylpropanoid/benzenoid pathway and is regulated by *EMISSION OF BENZENOIDS I* and *II* (EOBI and II), both are R2R3-MYB transcription factors (Spitzer-Rimon et al., 2010; Spitzer-Rimon et al., 2012). *EOBII* is a regulator of the biosynthesis of phenylpropanoid volatiles in Petunia and its expression is flower-specific and temporally and spatially associated with scent emission. Downstream of EOBII acts EOBI, another R2R3-MYB regulatory factor which directly regulates scent production by activation of *ODO1* and other scent-related genes in *Petunia.* It has been shown that *PH4*, not only plays a role in vacuolar acidification, but also in floral volatile emission (Cna`ani et al., 2015). The exact role of this factor of the R2R3-MYB family is not understood in detail yet.

To explain the variety of scent bouquets it has been hypothesized that the metabolic costs of volatile biosynthesis may not be as high as the ecological cost of herbivores feeding on plants and their reproductive structures. In *Petunia*, floral scent consists of attractive and repellent compounds at the same time and the reduction of particular components can decrease damage by florivores (Kessler et al., 2013). Additionally, western flower thrips *Frankliniella occidentalis* were attracted by monoterpenes and benzenoids, including benzaldehyde (Koschier et al., 2000). Thrips are phytophagous and a heavy infestation can destroy the plant´s reproductive structures. This suggests that a strong pollinator-mediated selection is needed to outweigh the negative effects of herbivore attraction to scented flowers. During a transition to selfing this selective force disappears and it is possible that scent emission not only gets lost as a secondary effect, but that herbivore pressure could actually drive the reduction in

scent emission. *CNL1* could be a target of selective processes in *Petunia* and *Capsella,* because it influences only few compounds in the pathway and others, for example coumaric acid, remain the same. Coumaric acid is a precursor for salicylic acid and important for plant defense (Widhalm and Dudareva, 2015). A reduction of a factor further upstream could have a pleiotropic influence on other important building blocks for various secondary metabolites in plants. This could explain why *CNL1* has been targeted more than once for the alteration of benzenoid emission and matches the theoretical prediction about genes with a low pleiotropy, but big phenotypic effect likely being targeted more than once during evolution as mentioned before. These genes are considered as 'speciation' or 'hotspot' genes. Several examples have been identified in animal model organisms, but many fewer examples are known in plants (Martin and Orgogozo, 2013).

Benzaldehyde can elicit different behavioral responses in different animal species (Schiestl, 2010), and terpenoids play an important role in insect behavior, too. In *Mimulus*, the three floral monoterpenes D-limonene, β-myrcene and (*E*)-β-ocimene contribute to differential pollinator attraction. The bumblebee-pollinated *M. lewisii* produces a scent bouquet consisting of different volatiles, but dominated by the three monoterpenes. *M. cardinalis* is hummingbird pollinated and emits only D-limonene in much smaller amounts than *M. lewisii* (Byers et al., 2014a). Pollinator-choice assays showed that *Bombus vosnesenskii* bumblebees, the natural pollinators of *M. lewisii*, are attracted by mixtures of the three monoterpenes and that floral scent alone is sufficient to elicit differential visitation by this bumblebee species (Byers et al, 2014a). An additional study investigated the genetic basis of the different volatile blends and there seem to be two loci responsible, *LIMONENE-MYRCENE SYNTHASE* (*LMS*) and *OCIMENE SYNTHASE* (*OS*). The results suggested that allelic variation in the *OS* locus contributes to differential pollinator visitation and is considered to be a 'speciation gene' (Byers et al., 2014b). A follow-up study looked at other *Mimulus* species and showed that *M. lewisii* is the only species specialized to bumblebee-pollination and also the only one emitting (*E*)-β-ocimene. Different mutations in the *OS* gene lead to the loss of (*E*)-β-ocimene emission in *M. cardinalis* and *M. verbenaceus,* both arose independently. These findings underline the speciation-gene concept, stating that genes with low pleiotropic effect but high phenotypic output are re-targeted during evolution.

A parallel study to this work analyzed additional scent compounds in the *Capsella* scent bouquet (Wozniak, 2019). The emission of (*E*)-β-ocimene can be robustly detected in *Capsella grandiflora*, the outcrosser, yet it cannot be detected in *Capsella rubella* and is strongly reduced in *Capsella orientalis.* The loss of (*E*)-β-ocimene emission in the selfer *C. rubella* is very likely to be due to mutations resulting

in amino acid changes in the gene *TERPENE SYNTHASE 02* (*TPS02).* The reduction in *C. orientalis* is probably due to polygenic effects (Wozniak, 2019).

The reduction in scent emission or the complete loss of floral scent during the transition to selfing has not only been shown in *Capsella* but also in other species, for example *Abronia umbellata* (Doubleday et al., 2013) and *Oenothera flava* (Raguso et al., 2007). The molecular basis of these changes is still not known.

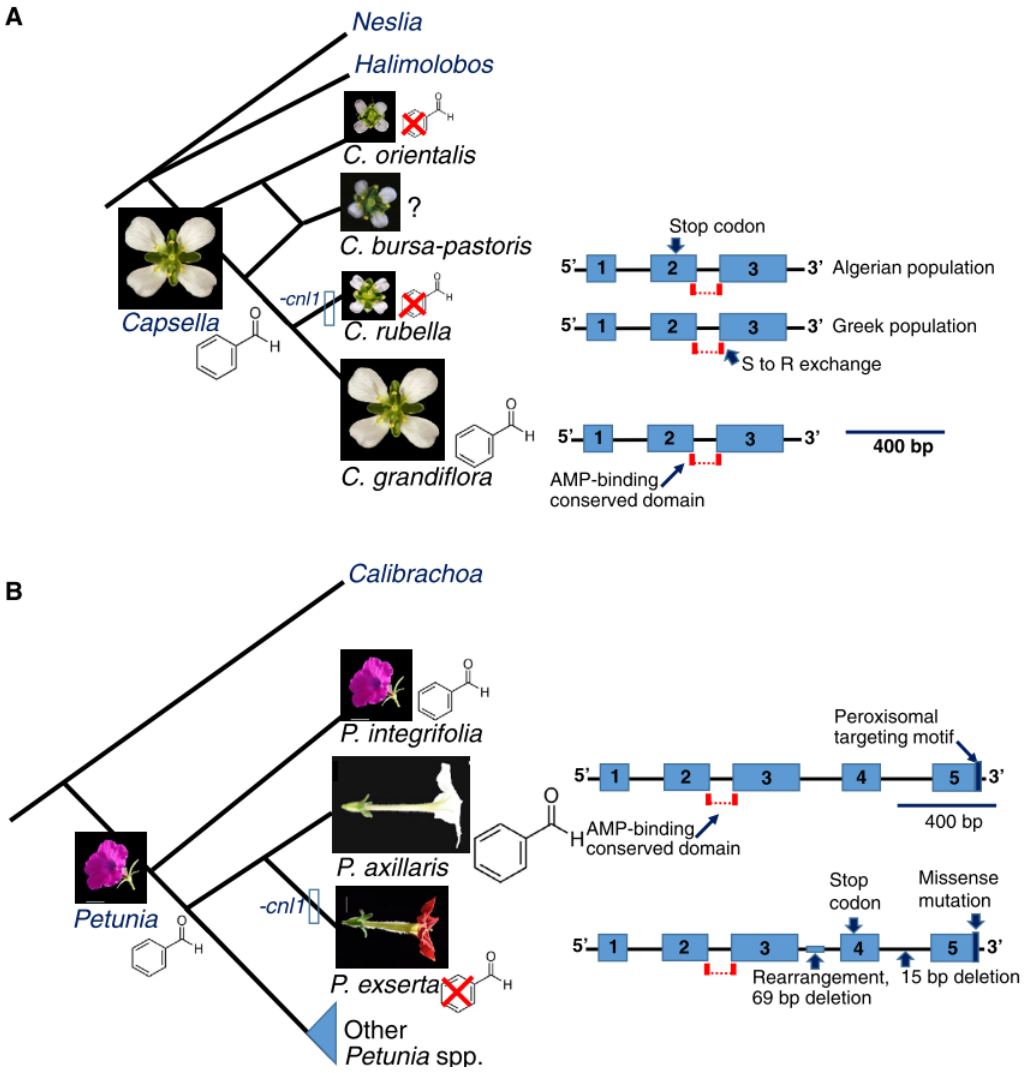

Figure 1.2: Parallel loss of floral scent in *Capsella* (A) and *Petunia* (B). Modified after Robert A. Raguso, 2016. The putative ancestral condition is indicated by the benzaldehyde molecule at the ancestral node and an empty bar across a branch indicates a loss-of-function mutation. The coding sequence diagrams on the right show the putative mutations to the *CNL1* gene. Exons are indicated by blue blocks, introns by black lines and sites of mutations by arrows.

1. 3 Evolution of flower size

Flowers show a wide variation in size. The world´s largest flowers grow up to almost a meter in diameter and are produced within the genus *Rafflesia*. Natural selection on flower size is associated with pollination efficiency. When the energetic costs of building larger flowers are lower than the gain obtained by the increase in pollination efficiency, large flowers tend to be favored (Davis et al., 2008). This leads to the suggestion that the selective pressure should relax when flower size does not play a role for pollination efficiency anymore. Confirmation for this hypothesis is provided by the morphological consequences of the transition to selfing, where a strong reduction in petal size can be observed (Orndruff, 1969; Sicard and Lenhard, 2011). Once self-fertilization is established, plants do not rely on animals for their reproduction anymore and traits associated with pollinator attraction will underlie a weaker selective pressure.

The transition to selfing has independently evolved multiple times and was often followed by a strong reduction in flower size. Therefore, the repeated organ size reduction argues for the pollinator-driven selection of large flowers (Wozniak and Sicard, 2017). The context of selfing syndrome evolution has been used to understand the genetic basis of natural variation in flower size, and it was concluded that most of the variation in flower size seems to have a complex genetic basis with different individual mutations having weak to moderate effects (Sicard and Lenhard, 2011).

An example for this is provided by a comparative study between *Capsella grandiflora* and *C. rubella* (see Figure 1.3 for difference in flower size). Six QTLs for differences in petal area were detected, explaining about 55% of the variation between species (Sicard et al., 2011). In following studies, two of these QTLs have been fine-mapped and identified. Polymorphisms within the intron of an organ-specific enhancer, a *cis*-regulatory element, regulating the level of STERILA APETALA (SAP) protein in developing petals are responsible for the size differences caused by PAQTL6. Population-genetic analysis showed that none of the detected polymorphisms at the causal region in the selfing species are *de novo* mutations, suggesting that the small-petal allele was captured from standing genetic variation in the ancestral outcrossing population (Sicard et al., 2016). PAQTL5 was identified a bit later, here a different splicing efficiency in the brassinosteroid-biosynthesis enzyme CYP724A1 contributed to the reduced petal size in *C. rubella*. Surprisingly the underlying mutations of this phenotypic change arose probably by *de novo* mutations and were not captured from standing genetic variation, as population-genetic analysis showed (Fujikura et al., 2018). These two examples highlight the role of both *de novo* mutations and

standing variation in the evolution of the selfing syndrome. Interestingly both mutations reduce the petal size by affecting similar mechanisms. The mutations in the petal-specific enhancer of SAP lead to a decrease in petal cell number due to a shorter proliferation period and the more efficient splicing of CYP724A1 results in higher brassinosteroid levels which limit cell proliferation as well (Sicard et al., 2016; Fujikura et al., 2018). In both cases the cell size remains the same, but the cell number is reduced in the selfer. The molecular basis of four other QTLs explaining the floral size variation between *C. grandiflora* and *C. rubella* remains to be investigated.

The genetic basis of selfing syndrome traits has also been studied in other species such as *Mimulus,* where at least 11 QTLs underlie the difference in corolla width between the selfer *M. guttatus* and the outcrosser *M. nasutus* (Fishman et al., 2002). Morphological traits often show a polygenic basis, enabling the observed floral diversity among angiosperms.

Figure 1.3: Flowers of *C. grandiflora* (left) and *C. rubella* (right). Lateral views of mature flowers, bar = 1 mm. Photo taken by Adrien Sicard.

1. 4 Evolutionary history of *Capsella*

The genus *Capsella* is used widely to study its population genetics and established selfing-syndrome traits. As mentioned before, selfing arose at least twice independently in this genus. The three species studied in this context are *Capsella grandiflora*, *C. rubella* and *C. orientalis*. As Figure 1.4 shows, *C. grandiflora* and *C. rubella* are closely related and both derived from a *grandiflora*-like ancestor which very likely was self-incompatible and outcrossing. *C. orientalis* probably diverged about 1 to 2 million years ago from that lineage as discussed by Hurka et al. (2012). This hypothesis is supported by the species´ distribution, *C. grandiflora* and *rubella* have overlapping ranges, but *C. orientalis´* distribution is not connected. Also the geographic range size of *C. grandiflora* is quite small, restricted to Greece and the Adriatic region, whereas *C. rubella* can be found around the whole Mediterranean Sea and beyond. *C. orientalis´* distribution range is also quite wide as it can be found in Southern Russia and Kazakhstan (Fig. 1.5). This supports the hypothesis that selfing species have a wider distribution because they can colonize new habitats more easily than outcrossing species, also referred to as "Baker´s law" (Baker, 1967). It states that that a single, selfing individual is sufficient to start a new population and is more likely to do so than two self-incompatible individuals, who need to be close together spatially and temporally to reproduce.

According to Grossenbacher et al. (2015) the plant mating system can even predict the geographic size range. They compared species from 20 different genera, and plants reproducing via self-fertilization have consistently larger geographic ranges than their outcrossing relatives.

The genus *Capsella* proved to be a great resource to study and understand the patterns of genetic diversity between populations. This provides a better understanding of evolutionary dynamics (Barrett, 2003). Furthermore, the spatial movement of single genes and their transmission in time under the influence of different mating systems are studied in this genus.

Information about natural pollinators of *Capsella* is scarce. According to personal observations and communication with other researchers, different hymenoptera (e.g. bees) and diptera (e.g. hover flies) species might be the main pollinators, as they were observed visiting the flowers in Greece and Germany (M. Lenhard and N. Wozniak, personal communication).

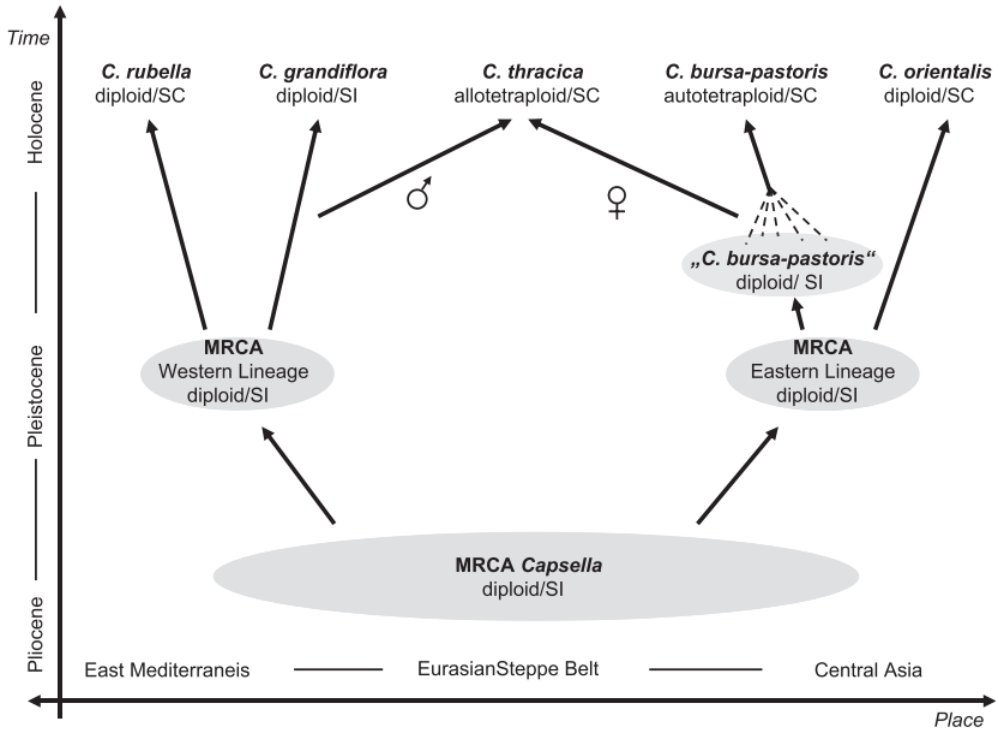

Figure 1.4: Evolutionary history of the genus *Capsella.* Modified from Hurka et al., 2012. MRCA = most recent common ancestor.

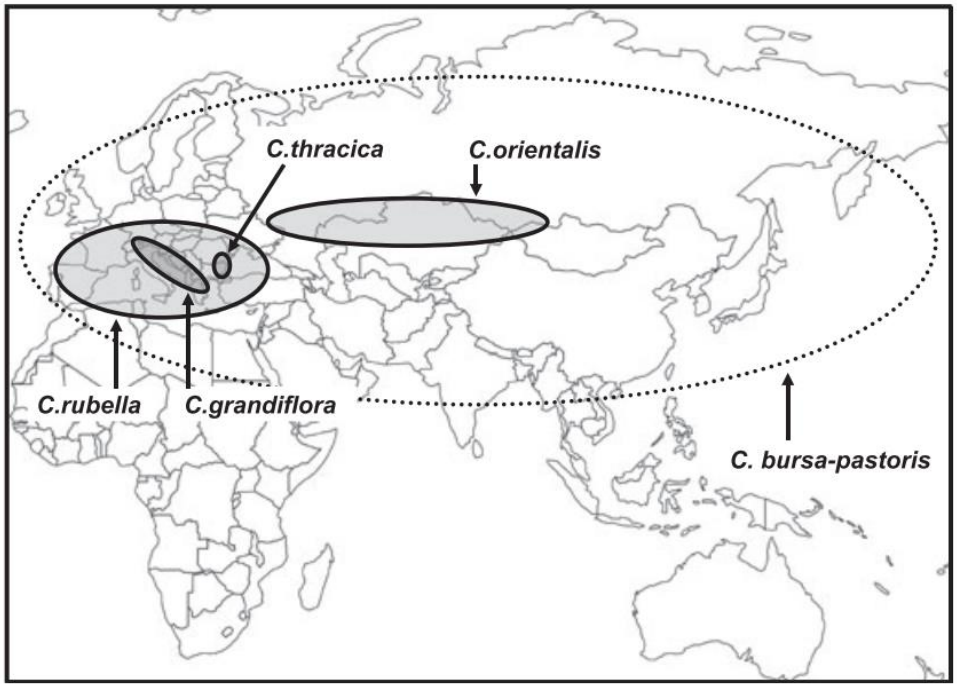

Figure 1.5: Distribution map of *Capsella* species. Modified from Hurka et al., 2012. Not indicated: Worldwide distribution of *C. bursa-pastoris* and colonized regions from *C. rubella* in Australia, North and South America.

Capsella rubella and *grandiflora* separated quite recently, about 100,000 – 200,000 years ago (Guo et al., 2009; Slotte et al., 2013). Analysis of comparative sequence information showed that *C. rubella* has only one or two alleles at most loci, suggesting that this species derived after an extreme population bottleneck and probably a small number of individuals or maybe even a single, selfing individual from Greece is the progenitor of this species (Guo et al., 2009). A more recent study identified more haplotypes that might have founded the selfing population and diversity probably got lost after the transition to self-fertilization (Brandvain et al., 2013). The offspring of the newly emerged selfers probably colonized the whole Mediterranean via Italy and Spain in one direction or Turkey and further in the other direction. These routes are referred to as northern and southern route (Guo et al., 2009) or eastern and western route (Koenig et al., 2019). It has been shown recently that there might be ongoing gene flow between *C. grandiflora* and the eastern *C. rubella* populations leading to introgression of *grandiflora*-alleles in *rubella* populations and vice versa (Koenig et al., 2019). This underlines how demography can influence population genetics, making genetic scans for adaptive traits more difficult. Demography and selection are strongly intertwined in self-fertilizing species (Hartfield et al., 2017) and as more information about the genus *Capsella* becomes available, more general theoretical predictions made about the evolution of selfing can be proven.

As described by Hurka et al. (2012) there are also two different lineages in the population history of the genus *Capsella* (Fig. 1.4). *C. orientalis* derived from an ancestral eastern lineage and has different phylogenetic roots than *C. grandiflora* and *rubella*. The split between these different lineages was estimated to have happened at the end of the Pliocene (Hurka et al., 2012). The two selfing species *C. rubella* and *orientalis* arose independently. Despite their different phylogenetic history, they share very well established selfing-syndrome traits. Identifying the genetic basis of these traits and searching for evidence for parallel evolution in these species is a key interest of researchers in the field.

A breakdown of self-incompatibility in ancestors of the selfers was necessary to enable the speciation of the different *Capsella* species shown in Figure 1.4. The loss of SI in *C. orientalis* occurred within the past 2.6 million years. A frameshift deletion in the male specificity gene SCR is fixed in *C. orientalis* and is responsible for the loss of male SI specificity, confirming predictions that during the transition to selfing the selection for mutations disrupting male SI specificity should be stronger (Bachmann et al., 2018; Tsuchimatsu and Shimizu, 2013). Furthermore a small *S*-locus-linked RNA causes dominance of self-compatibility, suggesting that the degeneration of pollen SI specificity in dominant *S*-alleles is important for mating system shifts in Brassicaceae (Bachmann et al., 2018).

C. orientalis shows strongly reduced genome-wide polymorphism levels in comparison to *C. rubella.* This fits to *C. orientalis´* estimated transition to selfing being more than 2 million years earlier than *C. rubella*, as selfing is expected to result in reduced genome-wide polymorphisms (Wright et al, 2013; Douglas et al., 2015). Both selfers carry different *S*-locus haplotypes, confirming the independent shift to selfing (Bachmann et al., 2018). In *C. rubella* the male determinant SCR is a pseudogene and probably responsible for breakdown of SI, whereas functional versions of SRK, the female determinant, have persisted (Guo et al., 2009).

Some genes playing a role in selfing syndrome traits have been identified in *Capsella rubella* lately (Sas et al., 2016; Sicard et al., 2016; Fujikura et al., 2018). One of these genes is *CNL1*, being responsible for the production of benzaldehyde (BAld) as major compound of floral scent in *Capsella grandiflora*. Sequence analysis of *CNL1* alleles from different populations led to the hypothesis that two different alleles of *CNL1* can be detected in *Capsella rubella* populations and these led to two independent inactivations of CNL1 resulting in a loss of BAld emission. One of them is an amino acid change which was detected in most of the *C. rubella* accessions studied at the time. The other one is a 4 bp deletion resulting in a premature stop codon (Sas et al., 2016; see also 1. 2). To prove if benzaldehyde emission has been lost twice independently, more evidence is necessary.

During previous work it was possible to design quasi-isogenic *Capsella* lines differing in only 12 kb around the locus for loss of BAld emission (Sas et al., 2016). This allows directly testing for the effects of a single selfing-syndrome trait on the outbreeding rate. This could potentially determine if scent loss evolved as an adaptive response, as a product of the relaxation of selective pressure or as a consequence of the increased effect of genetic drift in selfing lineages. So far, most hypotheses are supported by empirical data based on correlations in highly genetically heterogeneous populations, and it would be a definite advance to measure the adaptive value of a selfing-syndrome trait in an otherwise homogeneous background (Wozniak and Sicard, 2018).

Asking if the same or different factors are involved in selfing syndrome evolution, a comparative QTL study by Wozniak (2019) was done for *C. grandiflora*, *rubella* and *orientalis*. Interspecific crosses of all species were genotyped and phenotyped for flower size and pollen-to-ovule ratio to see if the same or different loci are involved in selfing-syndrome evolution. Some shared regions could be observed, indicating parallel evolution using the same genetic pathways. If this is also applicable to scent emission remains to be investigated.

1. 5 Aims of this work

Following up the work of Sas et al. (2016), one aim of this work is to understand the molecular basis and adaptive value of scent loss in *Capsella rubella* by:

I) Investigating if the S-to-R mutation in *CNL1* that can be detected in many accessions is causal for the loss of benzaldehyde emission;
II) Studying the second independent inactivation of *CNL1* which was detected in one Algerian accession in more detail and analyzing its distribution within different *C. rubella* populations;
III) Establishing a high-throughput method to compare two different *CNL1* alleles in a common garden experiment to mimic natural conditions and to test if one trait can be enough to elicit behavioral differences in pollinators resulting in different outcrossing rates.

As resources for a comparative QTL study have been established in the lab by N. J. Wozniak (2019), they were used to study the loss of benzaldehyde emission in *C. orientalis* aiming for a better comprehension of scent loss in this species. Claudia Sas already showed that *CNL1* is not responsible for BAld loss in *C. orientalis*, suggesting that convergent evolution of scent loss in the two selfing species was not achieved by mutations to the same gene but probably due to a different genetic pathway. To further investigate the loss of benzaldehyde emission in *C. orientalis,* expression data was analyzed and a precursor-feeding experiment conducted.

For additional elucidation of the genetic basis of petal size difference in the genus *Capsella*, the position of PAQTL1 explaining 9% of petal size variation between *C. grandiflora* and *C. rubella* should be confirmed and characterized.

2. Manuscript: Retracing the molecular basis and evolutionary history of the loss of benzaldehyde emission in the genus *Capsell*a

This manuscript was written by Michael Lenhard and me. Except for manuscript Figures 2.2, 2.3b, 2.4a, 2.S2, 2.S3 and 2.S5 all the experiments were performed, analyzed and prepared for publication by me. Christian Kappel did the population genetics illustrated by Figures 2.2 and 2.S5. Figure 2.3b is based on data from Claudia Sas, I prepared them for publication. Joseph H. Lynch performed and analyzed experiments for Figure 2.4 and Table 2.1. Oded Skaliter performed experiments for Figure 2.S2 and I prepared the figure for publication.

This manuscript was submitted to New Phytologist on March 1st 2019.

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

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

[1] University of Potsdam, Institute for Biochemistry and Biology, 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, U.S.A. and Purdue Center for Plant Biology, Purdue University, West Lafayette, IN 47907, U.S.A.

[3] Institute of Biology, Dahlem Centre of Plant Sciences (DCPS), Freie Universität Berlin, Haderslebener Straße 9, 12163 Berlin, Germany.

[4] Robert H. Smith Faculty of Agriculture, Food and Environment, The Hebrew University of Jerusalem, POB 12, 76100 Rehovot, Israel

* present address: Department of Plant Biology, Uppsala BioCenter, Swedish University of Agricultural Sciences and Linnean Center for Plant Biology, Uppsala, Sweden

[#]present address: Bayer Corporation, 407 Davis Dr., Morrisville, NC 27560, U.S.A.

§ author for correspondence:

email: michael.lenhard@uni-potsdam.de

telephone: +49-331-9775580

2. 1 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 a 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 *C. 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 outside Greece, the center 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.

Keywords: benzaldehyde, *Capsella*, cinnamate:CoA ligase, evolution, floral scent, selfing syndrome, Shepherd's Purse

2. 2 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 about 10-14 million years ago, 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). About 1 to 2 million years ago the self-compatible *C. orientalis* diverged from a self-incompatible *C. grandiflora*-like ancestor (Hurka *et al.*, 2012; Bachmann *et al.*, 2018), and around 100,000 to 200,000 years 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). Phenylpropanoids/benzenoids, the second largest class of floral volatiles, are ultimately derived from phenylalanine, which is converted by phenylalanine-ammonia lyase to *trans*-cinnamic acid for the production of C_6-C_3 and C_6-C_1 compounds (Figure 2.S1). Alternatively, C_6-C_2 compounds, such as phenylacetaldehyde and phenylethanol, are formed from phenylalanine via a cinnamic acid-independent pathway initiated by phenylacetaldehyde synthase (Dudareva *et al.*, 2013; Widhalm & Dudareva, 2015). Cinnamic acid can be metabolized to volatile C_6-C_3 phenylpropenes via coumarate; it can also be converted to C_6-C_1 benzenoids including benzoic acid via β-oxidative or non-β-oxidative pathways. While the steps in the non-β-oxidative pathway remain incompletely understood, the β-oxidative pathway for benzoic acid synthesis has been fully elucidated (Qualley *et al.*, 2012) and shown to be localized in peroxisomes. It starts with cinnamate:CoA ligase (CNL) catalyzing the formation of cinnamoyl-CoA, which then undergoes a shortening of the three-carbon side chain by two carbons via steps analogous to fatty-acid β-oxidation, ultimately resulting in benzoyl-CoA. Benzoyl-CoA can be exported from peroxisomes to the cytoplasm directly to form benzylbenzoate and phenylethylbenzoate, or hydrolyzed 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 wide-spread, and thus likely most ancient floral volatiles (Schiestl, 2010), and elicits diverse responses in different animal species (Schiestl, 2010). However, the origins of BAld are still unresolved, but could be three-fold (Widhalm & Dudareva, 2015). 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 (Figure 2.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 from selfing *C. rubella* (Amrad *et al.*, 2016; Sas *et al.*, 2016). Additionally to the mutation in *P. exserta CNL1*, a reduced expression of the transcription factor *ODO1* may also contribute to the loss of several scent compounds in this species. 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* (Amrad *et al.*, 2016; Sas *et al.*, 2016). 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, neither has the biochemical basis for the loss of *CNL1* activity in *C. rubella* been proven, nor has the origin of the second inactive allele been elucidated. Also, the basis of scent loss in the independently derived *C. orientalis* is unknown.

In the present work, we employ 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 are the evolutionary history and 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?

2. 3 Material and Methods

2. 3. 1 Plant material and growth conditions

All *C. grandiflora*, *C. rubella* and *C. orientalis* accessions used for the experimental studies have been described (Sicard *et al.*, 2011; Hurka *et al.*, 2012; Koenig *et al.*, 2018). Lines used for transformation via floral dip (HIFgg/rr) have been described (Sicard *et al.*, 2011; Sas *et al.*, 2016). Plants for scent measurements were grown in the greenhouse with supplemental lighting under a 16 hours light/ 8 hours dark cycle. Temperature during the day was 21°C and during the night 16°C.

2. 3. 2 Scent phenotyping

Dynamic headspace collection

Scent collections of *Capsella* inflorescences were performed as described (Sas *et al.*, 2016), using dynamic headspace sampling with the following modifications. Headspace samples were analyzed by coupled gas-chromatography-mass spectrometry (Agilent 7890A, Agilent 5975C, autosampler Agilent G4513A, DB-5 column 30m Ø 0.25mm). Floral volatiles were quantified as described (Sas *et al.*, 2016), yet here using the software OpenChrom.

Internal pools

For internal pool analysis, three replicates of at least 100 mg open flowers with sepals removed were harvested between 4 pm and 6 pm. Flowers were extracted overnight at 4°C in 5 mL dichloromethane with naphthalene added as internal standard. Samples were concentrated under nitrogen to approximately 200 µL and analyzed by GC-MS on an Agilent 7890B gas chromatograph (GC) (Agilent Technologies, Santa Clara, CA) equipped with a 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, Santa Clara, CA). 1 µL of sample was injected at 1:10 split using a Gerstel cooled injector system (CIS4, Gerstel, Germany) with an injection gradient of 12° per sec 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.

2. 3. 3 Molecular cloning and plant transformation

The following constructs were used to complement *Capsella* plants: Cg926CNL1 (described by (Sas *et al.*, 2016)), Cr1504CNL1^{R453S} and Co1983CNL1. Cr1504CNL1^{R453S} was generated by introducing a point mutation into a construct containing Cr1504CNL1 generated before (Sas *et al.*, 2016), using primers OFJ1, OFJ2, OFJ3 and OFJ4 (all oligonucleotide sequences are given in Supporting Table 2.S1). Co1983CNL1 was generated by amplification from genomic DNA extracted from *C. orientalis* 1983 plants. The Co1983CNL1 promoter sequence was amplified using OFJ20 and OFJ21 (1. 7 kb; spanning the whole intergenic region between CNL1 and the gene before), the coding region was amplified using OFJ19 and OFJ22 (2 kb). The PCR-amplified fragments were joined by subcloning into a modified version of the pBluescript II KS (StrataGen) using SLiCE fusion cloning (Zhang *et al.*, 2012). The fragments were cloned into the PacI site of the plant transformation vector pBarMAP, a derivate of pGPTVBAR (Becker *et al.*, 1992).

To generate constructs for transient expression in petunia, the Ω leader was amplified using OFJ9 and OFJ10; CNL1 fragments were amplified using OFJ11 and OFJ12 or OFJ26 and OFJ27 (see primer list) and cloned into the plant transformation vector pBarM35SXFB, a derivate of pGPTVBAR using the SLiCE fusion cloning.

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 in 2-3 days intervals.

Constructs for enzyme assays were generated by amplification from cDNA from the respective accessions and cloned using the CloneJET PCR cloning kit (Thermo Fisher Scientific). Primers introduced a NcoI restriction site before the start and removed the stop codon and introduced a NotI site at the end for cloning into the expression vector.

2. 3. 4 *In vitro* CNL enzyme assays

CNL1 open reading frames were subcloned into the pET28 expression vector using NcoI and NotI restriction sites. *Escherichia coli* Rosetta cells carrying the expression constructs were cultured in LB medium with 50 mg/mL kanamycin at 37°C. When the culture density reached A_{600} = 0.6, expression was induced by addition of isopropyl 1-thio-b-D-galactopyranoside (IPTG) to a final concentration of 0.4 mM and the temperature was decreased to 18°C. After a 16-h incubation on a rotary shaker (220 rpm) at 18°C, the *E. coli* cells were harvested by centrifugation and resuspended in buffer containing 50 mM

potassium phosphate, pH 8.0, 150 mM KCl, 20% (v/v) glycerol, and 1 mM phenylmethanesulfonyl fluoride. Cells were lysed by the addition of lysozyme (0.5 mg/mL) and DNase I (10 mg/mL in 4 mM MgCl2, final), incubated on ice for 30 min and sonicated for 3 min. Extracts were centrifuged for 30 min (13,000 x g, 4°C), and the recombinant enzymes were purified using Ni-NTA agarose (Qiagen) following the manufacturer's protocol, using a wash buffer composed of 50 mM potassium phosphate, pH8.0, 500 mM NaCl, and 20 mM imidazole and an elution buffer composed of 50 mM potassium phosphate, pH8.0, 500 mM NaCl, and 500 mM imidazole. Eluted protein was desalted on EconoPac 10 DG columns (Bio-Rad) into the buffer containing 20 mM potassium phosphate, pH 8.0, and 20% (v/v) glycerol. Desalted enzymes were aliquoted, frozen in liquid N_2, and stored at -80°C until use. The protein concentration was determined by the Bradford method (Bradford, 1976). Enzymes were assayed for activity as described previously (Klempien *et al.*, 2012), and the cinnamoyl-CoA product analyzed by HPLC (Qualley *et al.*, 2012).

2. 3. 5 Transient expression assay in petunia flowers

Petunia axillaris N flowers were infiltrated at anthesis with *A. tumefaciens* AGL-0 containing either pBAR-35SΩ:Cr1504CNL1, pBAR-35SΩ:Cr1504CNL1^{R453S} or pRCS2-35S:PhCNL1 as described previously (Spitzer-Rimon *et al.*, 2012). Three days following infiltration, dynamic headspace analysis was performed on detached flowers for 24h. Emitted volatiles were collected using an adsorbent trap consisting of a glass tube containing 200 mg Porapak Type Q polymer (80/100 mesh; Alltech) held in place with steel mesh plugs. Trapped volatiles were eluted using 1.5 ml hexane, 0.5 ml acetone and 2 µg isobutylbenzene was added to each sample as an internal standard. GC–MS analysis (1 µl sample) was performed using a device composed of a Pal autosampler (CTC Analytic), a TRACE 1300 Mainframe MS equipped with an Rtx-5SIL mass spectrometer fused-silica capillary column (Restek; i.d. 0.25µm, 30m × 0.25mm) and a TRACE ISQ LT EI single quadruple mass spectrometer (ThermoFinnigan). Helium was used as the carrier gas at a flow rate of 1 ml/min. The injector temperature was set to 220°C (splitless mode) and the interface to 240°C, and the ion source was adjusted to 200°C. The analysis was performed under the following temperature program: 2 min of isothermal heating at 40°C followed by a 7°C/min oven temperature ramp to 250°C then 2 min of isothermal heating. The system was equilibrated for 1 min at 70°C before injection of the next sample. Mass spectra were recorded at 3.15 scan/s with a scanning range of 40–450 mass-to-charge ratio and electron energy of 70 eV. Compounds were tentatively identified (>95% match) based on NIST 14 Mass Spectral Library data version (software version 2.2) using the XCALIBUR v4.1 2017 (ThermoFinnigan). Further identification of major compounds was based

on comparisons of mass spectra and retention times with those of authentic standards (Sigma-Aldrich) analyzed under similar conditions.

2. 3. 6 QTL mapping

Generation and genotyping of the mapping population derived from crossing *C. grandiflora* and *C. orientalis* will be described elsewhere (Wozniak and Sicard, in preparation). For phenotyping, head space samples were taken from five to seven plants per F3 family as pools, derived from the genotyped 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 non-parametric model. QTLs were tested in 1cM intervals and assessed by using LOD scores (the log10 likelihood ratio comparing the hypothesis that there is a QTL at the marker to the hypothesis that there is no QTL anywhere in the genome). The QTL significance threshold was determined by running the permutation test under the non-parametric model with 1000 permutation replicates and by using the 5% and 10% significance cutoff.

2. 3. 7 Population-genetic analyses

Genome sequencing data for the *C. grandiflora* population were downloaded from NCBI (PRJNA275635; (Josephs *et al.*, 2015)), those for *C. rubella* accessions from EBI ENA (PRJEB6689; (Koenig *et al.*, 2018)). Reads were mapped to the *C. rubella* reference genome (Slotte *et al.*, 2013) using bwa mem (Li, 2013). Reads mapping to the region of interest were extracted using samtools (Li *et al.*, 2009). Local variant calling was done using samtools. Further data processing was done using R (www.r-project.org). Hierarchical clustering was done using Euclidean distances of allele calls. Plotting was done using the R/lattice package. Identified *C. rubella* samples with cluster associations were plotted on a map using the R/ggmap package, sample coordinates were given by (Koenig *et al.*, 2018).

2. 3. 8 Statistical analysis

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

2. 4 Results

2. 4. 1 Loss of BAld emission in *C. rubella* Cr1504 is due to a single amino-acid exchange in CNL1

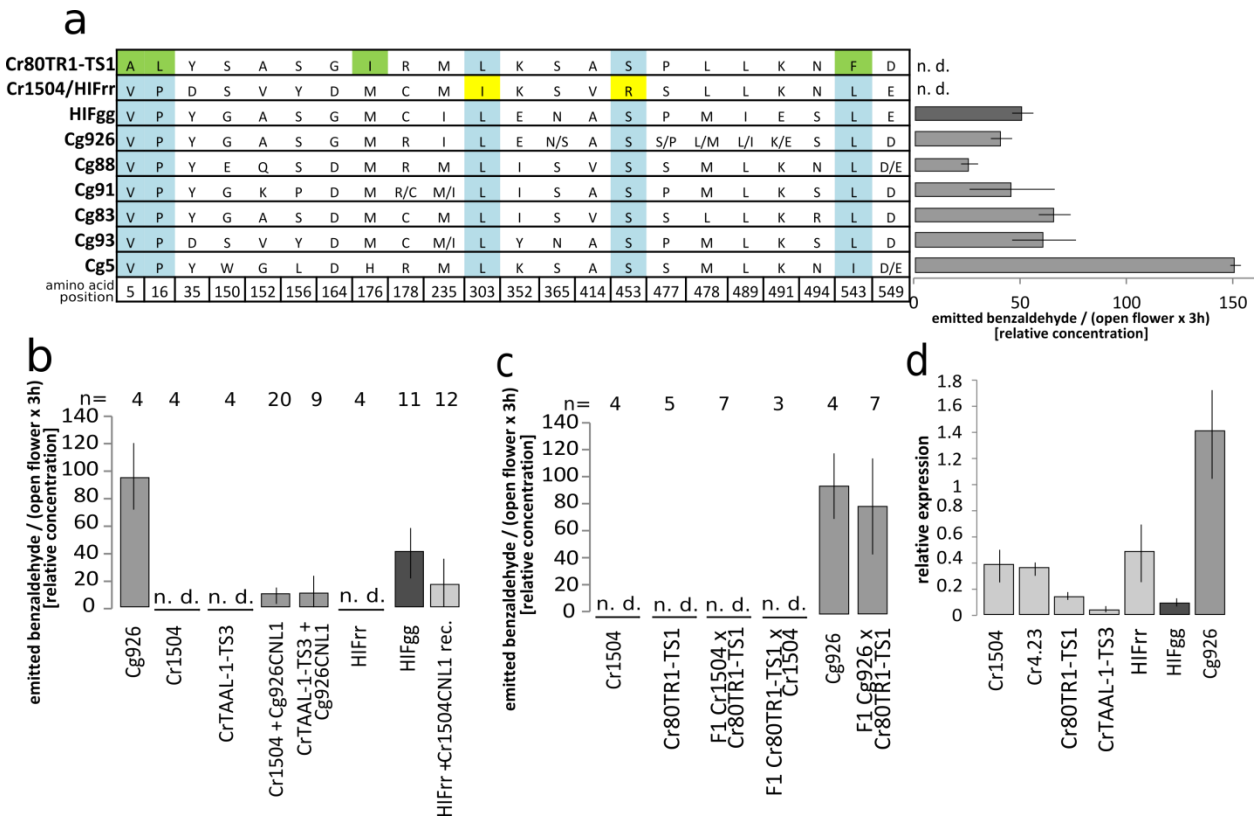

Figure 2.1: Molecular analysis of different *Capsella* CNL1 alleles.

a) 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. n.d., not detected. Relative concentration of BAld was normalized to 10 ng tridecane as internal standard in the samples here and throughout all figures.

b) 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. n.d., not detected.

c) Quantification of BAld emission in complementation test for Cr1504 and Cr80TR1. Cr80TR1 was crossed to Cg926 as 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.

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-base pair deletion - in accession CrTAAL-1-TS3 from Northern Africa (termed CrTAAL from now on). 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. 2.1b). 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. 2.1a). 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). In contrast to non-transgenic 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. 2.1b). 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. 2.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. 2.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 *E. coli*: *Cr1504 CNL1*, *Cg926 CNL1*, *Cr1504 CNL1*^{R453S}, *Cg926 CNL1*^{S453R}; the latter 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. 2.S3) (Klempien *et al.*, 2012). While Cg926 CNL1 protein showed robust cinnamate:CoA-ligase activity, none was detected for the Cr1504 CNL1 protein. Replacing serine 453 by arginine in the Cg926 CNL1^{S453R} protein abolished enzyme activity (Table 2.1). 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; that said 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.

Table 2.1: Michaelis-Menten kinetics for different CNL1 proteins. 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.

Protein	v_{max} (pkat/mg)	K_M (µM)
Cg926 CNL1	27,337 ± 914	81 ± 11
Cr1504 CNL1	n. d.	n. d.
Cr1504 CNL1^{R453S}	3,472 ± 216	105 ± 24
Cg926 CNL1^{S453R}	n. d.	n. d.
Cr80TR1 CNL1	n. d.	n. d.
Cg926 CNL1^{P16L}	n. d.	n. d.
Cg926 CNL1^{L543F}	498 ± 20	173 ± 23
Cg926 CNL1$^{P16L, L543F}$	n. d.	n. d.
Cr80TR1 CNL1^{L16P}	1,693 ± 94	156 ± 29
Cr80TR1 CNL1^{F543L}	n. d.	n. d.
Cr80TR1 CNL1$^{L16P, F543L}$	16,015 ± 684	105 ± 14
Co1983CNL1	114,243 ± 5963	213 ± 33

2. 4. 2 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-base pair deletion in *CNL1* of accession CrTAAL, we identified a further *C. rubella* accession, Cr80TR1-TS1 from Turkey (termed Cr80TR1 from now on), that carried the same *CNL1* haplotype as CrTAAL except for the 4-base pair deletion. Thus, this allele could encode for a full-length, potentially functional CNL1 protein. However, no BAld emission was detectable from Cr80TR1 inflorescences (Fig. 2.1a, 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 promoter at positions 7,541,229-7,541,342, 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. 2.1d). Thus, the promoter deletion is unlikely to explain the loss of *CNL1* activity in Cr80TR1; instead, alterations in protein primary structure are likely 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. 2.1a). 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. 2.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 2.1). 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 neither 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.

2. 4. 3 Evolutionary history of the *CNL1* mutations in *C. rubella*

Our previous analysis had indicated that the serine-to-arginine mutation in *Cr1504 CNL1* had most likely 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 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* and *C. rubella* accessions and clustered the accessions (Fig. 2.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. The L543F mutation (nucleotide polymorphism at position 7,539,156; Fig. 2.S5) was shared by all accessions in the Cr80TR1 cluster and the P16L mutation (nucleotide polymorphism at position 7,540,987) 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 7,539,156; however, this only resulted in a conservative leucine-to-isoleucine exchange. Therefore, this analysis strongly suggests 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 geographic 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. 2.2). 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. 2.S5). This pattern suggests that the West-ward geographical spread of *C. rubella* that occurred about 13,500 years 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.

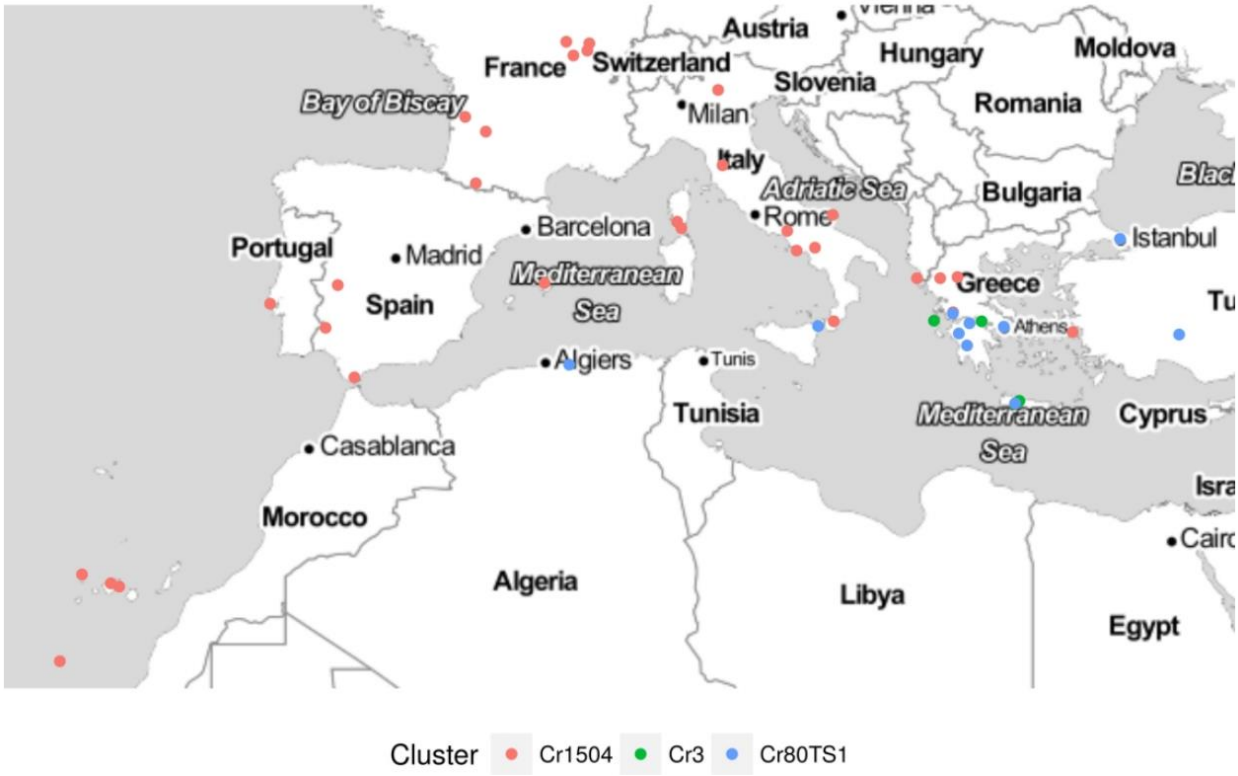

Figure 2.2: Distribution of the three different *C. rubella* haplotypes in the Mediterranean. The Cr1504-like cluster (red dots) spread West-ward and the CR80TR1-like cluster (blue dots) East- and South-ward. The third, Cr3-like haplotype (green dots) is restricted to Greece and Cyprus. All three haplotypes are present in Greece, the presumed center of diversity for this species.

2. 4. 4 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 one million years ago. We asked whether *C. orientalis* flowers still emit BAld. 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. 2.3a). 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. 2.3b). 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 non-functional *C. rubella CNL1* allele (Fig. 2.3c). 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 (Figure 2.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* (114,243 ± 5963 pkat/mg; Table 2.1), 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*, i.e. by one or more mutations to genes in the benzenoid pathway other than *CNL1*.

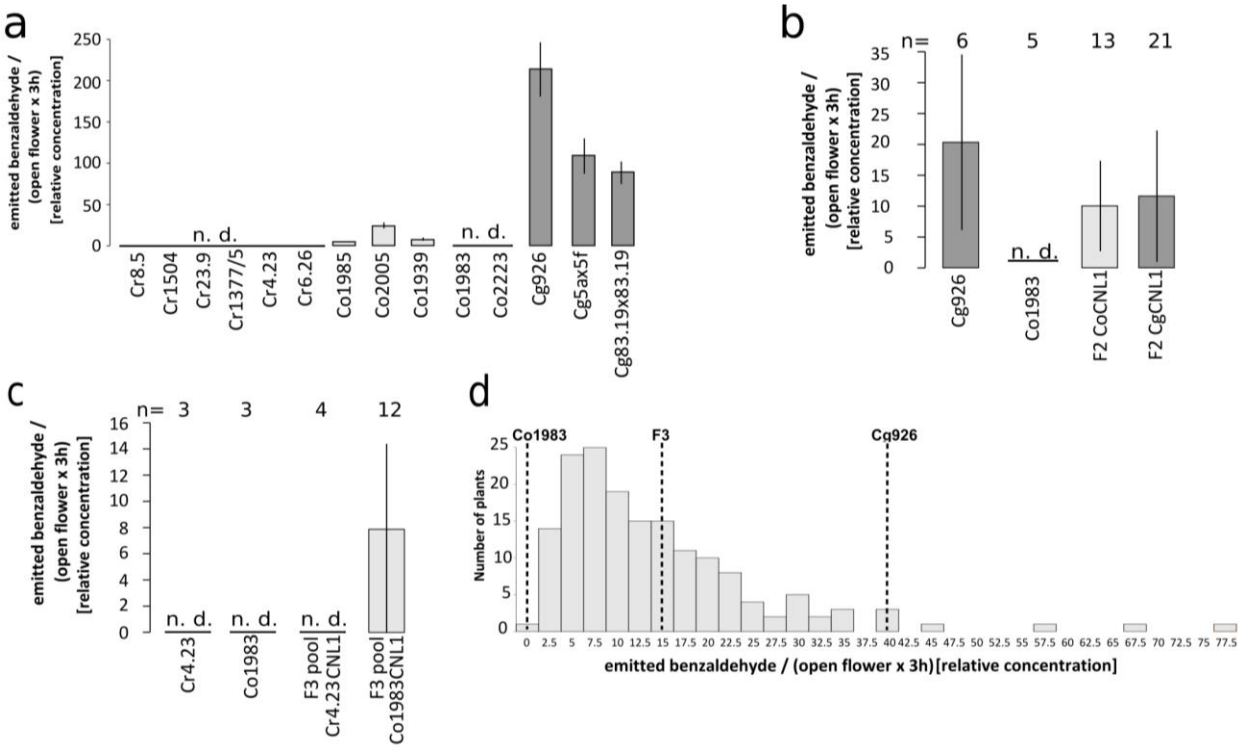

Figure 2.3: Characterizing scent loss in *C. orientalis*.

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 F2 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 pool. 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 whole F3 population are indicated.

2. 4. 5 Loss of BAld emission in *C. orientalis* results from polygenic downregulation of 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* x *C. orientalis* population. Floral volatiles were collected from five to seven plants per F3 family as pools, with two replicate samples collected per family on consecutive days. The F3 families showed a continuous distribution of the amounts of BAld emitted (Fig. 2.3d), with only one family showing no detectable BAld. This pattern is in stark contrast to that seen in a *C. grandiflora* x *C. rubella* RIL population, where a large number of families do not emit BAld, and argues against a single major mutation abolishing BAld emission in *C. orientalis*, because in this case we would expect around 25% of F3 families to emit no detectable BAld. 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. 2.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 asked whether BAld was still synthesized in *C. orientalis*. To this end, 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. 2.4a). This indicated that BAld was still 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, a number of 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. 2.4b). 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 strongly boosted by exogenous phenylalanine.

Thus, the capacity to synthesize and emit BAld is maintained in *C. orientalis*, yet BAld emission is not detectable. It is possible that, for unknown reason, substrate availability 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.

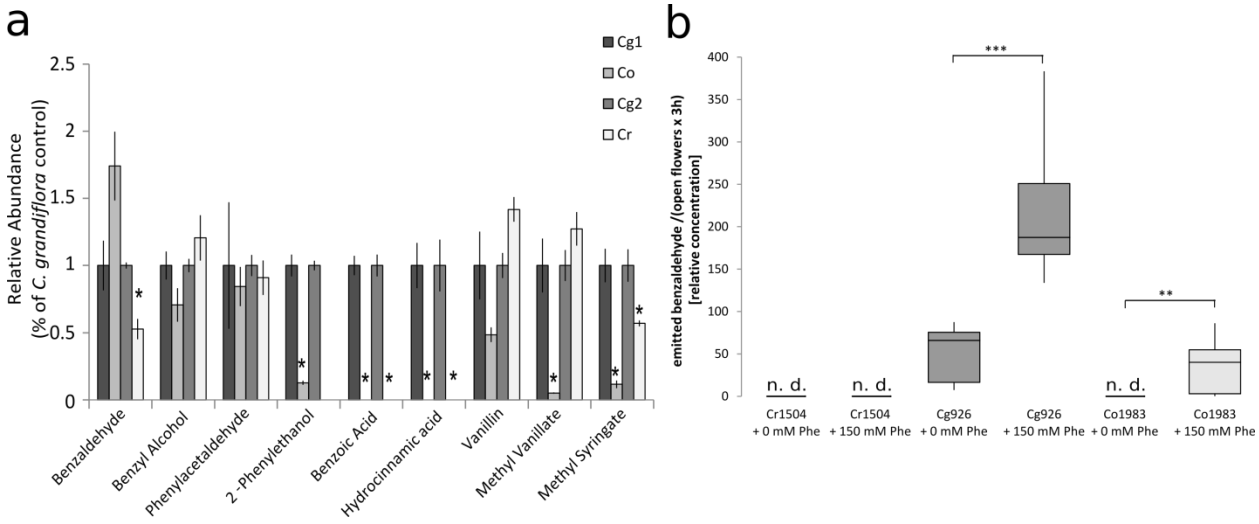

Figure 2.4.: Analyzing scent production in *C. orientalis*.

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 analyzed in parallel (Cg1 and Cg2, respectively). Three biological replicates each were analyzed. 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 two hours of feeding. Box plots show the median value (black line), the 25 and 75 percentiles (bottom and top bounds of the box), the top and bottom whiskers show the location of maximum and minimum. Asterisks indicate statistically significant differences as determined by a Student´s t test at $P < 0.01$ (**), $P < 0.001$ (***). 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.

2. 5 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 geographic 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 rather appears to be based on a number of smaller-effect mutations.

2. 5. 1 *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? From a population-genetic perspective, three factors influence the probability of parallel evolution (Stern, 2013): (i) the size of the mutational target; (ii) the net effect on fitness, which in turn depends on the extent of pleiotropy of mutations in a gene; and (iii) 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 likely also true for other genes involved in BAld synthesis and emission. Second, mutations in CNL1 should only affect the C_6-C_1 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 C_6-C_1 phenylpropanoids in the CoA-dependent pathways (Fig. 2.S1).

Thus, it is plausible that all three of these aspects contributed to the repeated loss of BAld emission via mutations to *CNL1*, yet their relative contribution will require further study. In particular, 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.

2. 5. 2 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 most likely 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 few *C. grandiflora* accessions, which most likely 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 center of diversity and likely geographic origin of the species, and the region where its range continues to overlap with that of *C. grandiflora*. 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 years 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 haplotype. 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 may suggest that both haplotypes were initially carried along with the subpopulation spreading West-ward. The *C. rubella* accession from Sicily with its recombinant haplotype indicates residual outcrossing between *C. rubella* accessions before or during their spread.

2. 5. 3 The molecular basis of the loss of BAld emission in *C. orientalis*

CNL1 was targeted by repeated inactivating mutations in *C. rubella* and petunia. Also, *C. rubella* and *C. orientalis* are derived from what were likely very similar ancestral populations 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, likely boosting flux through the phenylpropanoid pathway. The results from a segregating population derived from a *C. grandiflora* x *C. orientalis* cross did not find evidence for a single major-effect mutation abolishing BAld emission in *C. orientalis* (Sas *et al.*, 2016). Rather, the continuous phenotype distribution in this segregating population; the failure to detect significant QTL for BAld emission in it; and the high frequency at which BAld emission was regained in families derived from a *C. rubella* x *C. orientalis* cross - as long as these did not carry the non-functional *C. rubella CNL1* allele - all argue that the lack or strong reduction of BAld emission in *C. orientalis* is due to a number of small-effect mutations. 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. 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 environmental conditions in its native range can induce BAld emission. Thus, the mutations underlying the loss of BAld emission in *C. orientalis* await further study, as does the question of why the genetic routes to the same phenotypic outcome differed so much between *C. rubella* and *C. orientalis*, despite their being derived from likely very similar ancestral populations.

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 herbivores. For example, BAld attracts female Western flower thrips that oviposit in and damage the flowers (Koschier *et al.*, 2000), providing a possible scenario for adaptive loss of BAld emission. While our work reported here does not address this question, it provides the genetic resources to begin answering it using field experiments.

2. 6 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 are grateful to members of the Lenhard lab and Isabel Bäurle for discussion and helpful suggestions. 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.

2. 7 Author Contributions

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

2. 8 References

Adebesin F, Widhalm JR, Lynch JH, McCoy RM, Dudareva N. 2018. A peroxisomal thioesterase plays auxiliary roles in plant beta-oxidative benzoic acid metabolism. *Plant J* **93**(5): 905-916.

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. *Curr Biol* **26**(24): 3303-3312.

Arendt J, Reznick D. 2008. Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? *Trends Ecol Evol* **23**(1): 26-32.

Bachmann JA, Tedder A, Laenen B, Fracassetti M, Désamoré A, Lafon-Placette C, Steige KA, Callot C, Marande W, Neuffer B, Bergès H, Köhler C, Castric V, Slotte T. 2018. Genetic basis and timing of a major mating system shift in Capsella. *biorxiv*: 425389.

Becker D, Kemper E, Schell J, Masterson R. 1992. New plant binary vectors with selectable markers located proximal to the left T-DNA border. *Plant Mol Biol* **20**(6): 1195-1197.

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 Physiol* **135**(4): 1993-2011.

Broman KW, Wu H, Sen S, Churchill GA. 2003. R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **19**(7): 889-890.

Clough SJ, Bent AF. 1998. Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant J* **16**(6): 735-743.

Cna'ani A, Muhlemann JK, Ravid J, Masci T, Klempien A, Nguyen TT, Dudareva N, Pichersky E, Vainstein A. 2015. Petunia x hybrida floral scent production is negatively affected by high-temperature growth conditions. *Plant Cell Environ* **38**(7): 1333-1346.

Dudareva N, Klempien A, Muhlemann JK, Kaplan I. 2013. Biosynthesis, function and metabolic engineering of plant volatile organic compounds. *New Phytol* **198**(1): 16-32.

Esfeld K, Berardi AE, Moser M, Bossolini E, Freitas L, Kuhlemeier C. 2018. Pseudogenization and Resurrection of a Speciation Gene. *Curr Biol* **28**(23): 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*. *Proc Natl Acad Sci U S A* **106**: 5241-5245.

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

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. *Proc Natl Acad Sci U S A* **106**(13): 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). *Mol Ecol* **21**(5): 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 of the United States of America* **112**(50): 15390-15395.

Klempien A, Kaminaga Y, Qualley A, Nagegowda DA, Widhalm JR, Orlova I, Shasany AK, Taguchi G, Kish CM, Cooper BR, D'Auria JC, Rhodes D, Pichersky E, Dudareva N. 2012. Contribution of CoA ligases to benzenoid biosynthesis in petunia flowers. *Plant Cell* **24**(5): 2015-2030.

Koenig D, Hagmann J, Li R, Bemm F, Slotte T, Nueffer B, Wright SI, Weigel D. **2019**. Long-term balancing selection drives evolution of immunity genes in *Capsella*. *eLife* **8**.

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**(12): 2643-2655.

Lenser T, Theissen G. 2013. Molecular mechanisms involved in convergent crop domestication. *Trends Plant Sci* **18**(12): 704-714.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv:1303.3997*.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Proc GPD. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**(16): 2078-2079.

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**(7): 1055-1059.

Qualley AV, Widhalm JR, Adebesin F, Kish CM, Dudareva N. 2012. Completion of the core beta-oxidative pathway of benzoic acid biosynthesis in plants. *Proc Natl Acad Sci U S A* **109**(40): 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. *Curr Biol* **26**(24): 3313-3319.

Schiestl FP. 2010. The evolution of floral scent and insect chemical communication. *Ecol Lett* **13**(5): 643-656.

Schiestl FP, Balmer A, Gervasi DD. 2018. Real-time evolution supports a unique trajectory for generalized pollination*. *Evolution* **72**(12): 2653-2668.

Sheehan H, Hermann K, Kuhlemeier C. 2012. Color and scent: how single genes influence pollinator attraction. *Cold Spring Harb Symp Quant Biol* **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 Physiol* **138**(2): 1163-1173.

Sicard A, Lenhard M. 2011. The selfing syndrome: a model for studying the genetic and evolutionary basis of morphological adaptation in plants. *Ann Bot* **107**(9): 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**(9): 3156-3171.

Slotte T, Hazzouri KM, Agren JA, Koenig D, Maumus F, Guo YL, Steige K, Platts AE, Escobar JS, Newman LK, Wang W, Mandakova T, Vello E, Smith LM, Henz SR, Steffen J, Takuno S, Brandvain Y, Coop G, Andolfatto P, Hu TT, Blanchette M, Clark RM, Quesneville H, Nordborg M, Gaut BS, Lysak MA, Jenkins J, Grimwood J, Chapman J, Prochnik S, Shu S, Rokhsar D, Schmutz J, Weigel D, Wright SI. 2013. The Capsella rubella genome and the genomic consequences of rapid mating system evolution. *Nat Genet* **45**(7): 831-835.

Spitzer-Rimon B, Farhi M, Albo B, Cna'ani A, Ben Zvi MM, Masci T, Edelbaum O, Yu YX, Shklarman E, Ovadis M, Vainstein A. 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**(12): 5089-5105.

Stern DL. 2013. The genetic causes of convergent evolution. *Nat Rev Genet* **14**(11): 751-764.

Stern DL, Orgogozo V. 2008. The loci of evolution: how predictable is genetic evolution? *Evolution* **62**(9): 2155-2177.

Widhalm JR, Dudareva N. 2015. A familiar ring to it: biosynthesis of plant benzoic acids. *Mol Plant* **8**(1): 83-97.

Zhang Y, Werling U, Edelmann W. 2012. SLiCE: a novel bacterial cell extract-based DNA cloning method. *Nucleic Acids Res* **40**(8): e55.

2. 9 Supporting information

2. 9. 1 Supplemental Figures

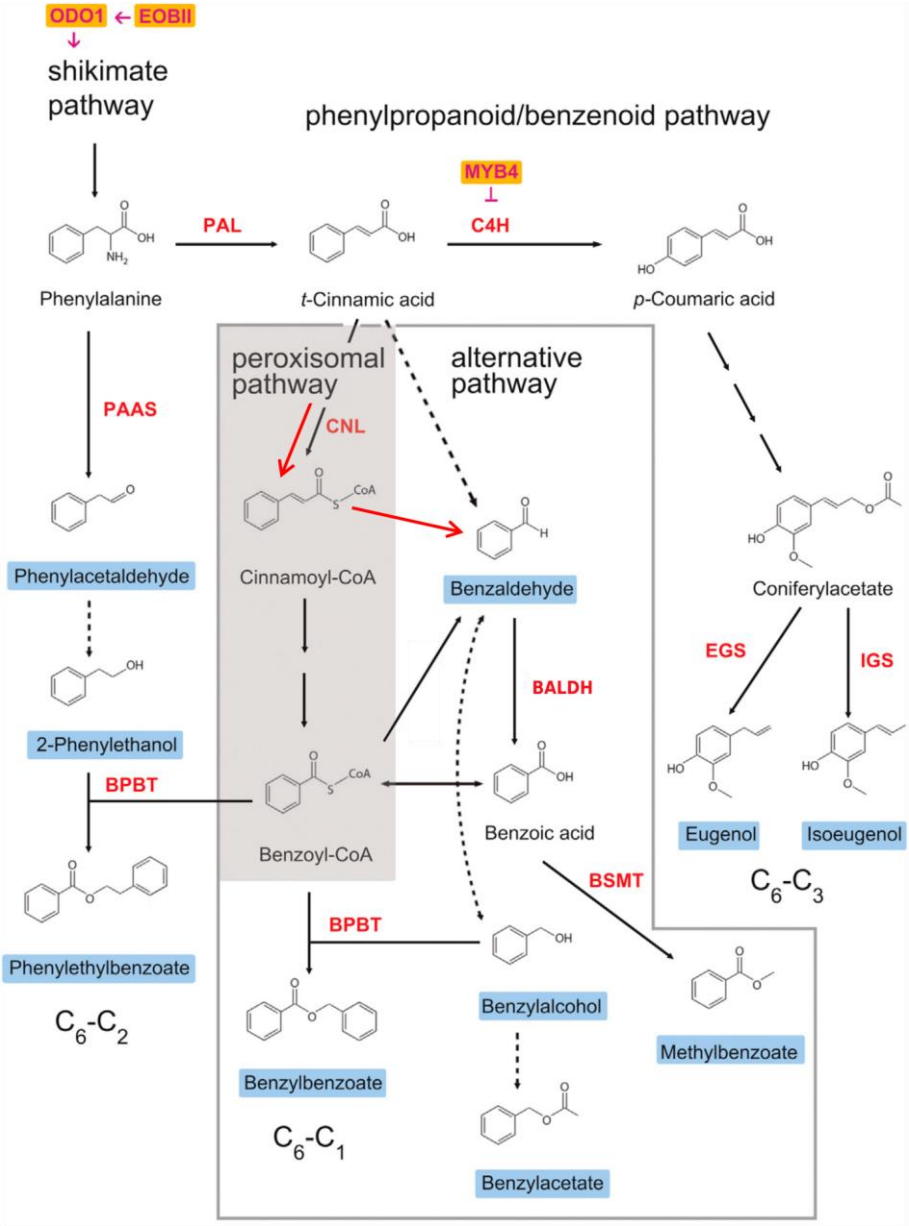

Figure 2.S1: Overview over phenylpropanoid/benzenoid pathway in *Petunia*. The pathway diagram was modified from (Amrad *et al.*, 2016). Volatile compounds emitted by fragrant petunia are highlighted in blue, key enzymes are shown in red, transcriptional regulators are highlighted in yellow. Dashed arrows indicate steps for which no enzymes have been identified and multiple arrows simplify enzymatic steps. BSMT: benzoic acid/salicylic acid carboxyl methyltransferase; BPBT: benzoyl-CoA:benzylalcohol/2-phenylethanol benzoyltransferase; C4H: cinnamate 4-hydroxylase; CNL: cinnamate:CoA ligase; EGS: eugenol synthase; EOBII: EMISSION OF BENZENOIDS II; IGS: isoeugenol synthase; ODO1: ODORANT1; PAAS: phenylacetaldehyde synthase; PAL: phenylalanine ammonia lyase; BALDH: benzaldehyde dehydrogenase (proposed for petunia). Red arrows indicate CoA-dependant, non-β-oxidative pathway.

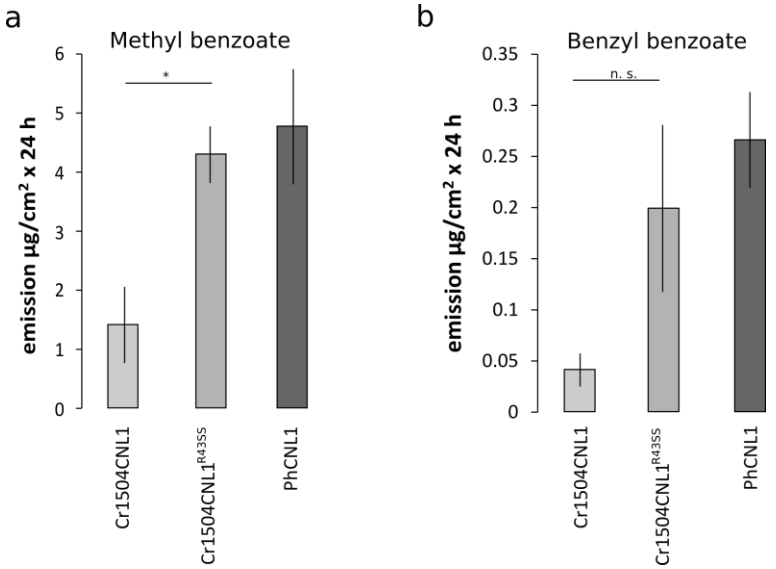

Figure 2.S2: Benzenoid emission after transient expression in *P. axillaris N.*

a) Methyl benzoate emission from flowers transiently overexpressing the indicated CNL enzymes. Values are mean ± SD from four measurements.

b) Benzyl benzoate emission from flowers transiently overexpressing the indicated CNL enzymes. Values are mean ± SD from four measurements.

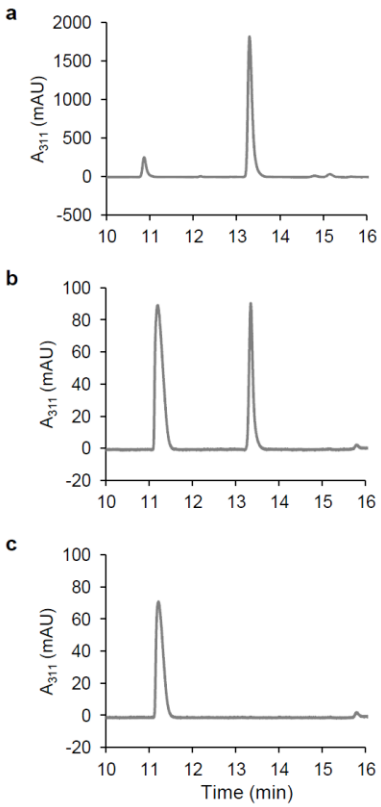

Figure 2.S3: HPLC assays of recombinant CNL enzymes.

a) Authentic cinnamoyl-CoA standard (0.1 mmol).

b) Assay of recombinant CgCNL1 enzyme.

c) Assay of boiled (denatured) recombinant CgCNL1 enzyme.

```
>NP_176764.1 acyl activating enzyme 12 [Arabidopsis thaliana]
>CDY22985.1 BnaC03g40840D [Brassica napus]
>XP_010474116.1 PREDICTED: butyrate--CoA ligase AAE11, peroxisomal-like [Camelina sativa]
>AOZ16462.1 truncated cinnamate-CoA ligase 1 [Petunia exserta]
>AEO52693.1 cinnamic acid:CoA ligase [Petunia x hybrida]
>XP_004232666.1 butyrate--CoA ligase AAE11, peroxisomal [Solanum lycopersicum]
>XP_006602034.1 butyrate--CoA ligase AAE11, peroxisomal [Glycine max]


CLUSTAL O(1.2.4) multiple sequence alignment


XP_010474116.1 [Cs]     MDKLELCKANNVPLTPIIFLKRASECYPNRTSIIYGNYRFTWPQTYDRCCRLAAFLLSLN        60
Col983 CNL1             MDKLALCEANNVPLTPITFLKRASECYPKQTSIIYGQTRFTWPQTYDRCCRLASSLLSFN        60
Cr1504 CNL1             MDKLVLCEANNVPLTPITFLKRASECYPKRTSIIDGQTRFTWPQTYDRCCRLASSLLSLN        60
CR80TR1-TS1 CNL         MDKLALCEANNVPLTLITFLKRASECYPKRTSIIYGQTRFTWPQTYDRCCRLASSLLSLN        60
Cg926 CNl1              MDKLVLCEANNVPLTPITFLKRASECYPKRTSIIYGQTRFTWPQTYDRCCRLASSLLSLN        60
NP_176764.1 [At]        MDNLALCEANNVPLTPITFLKRASECYPNRTSIIYGKTRFTWPQTYDRCCRLAASLISLN        60
CDY22985.1 [Bn]         MDNLALCDANNFPLTPITFLKRAAECYPNRTSIIYGQTRFTWPQTYDRCCRLTASLLSLN        60
XP_006602034.1 [Gm]     MDNLQKCQANYTALTPLTFLMRAAACYANRTSVIHEGTRFTWAQTYERCRRLAFSLRALN        60
AEO52693.1 [Ph]         MDELPKCGANYVPLTPLTFLTRAFKSYANRTSIIYAGARFTWEQTYKRCCRLASSLQSLN        60
XP_004232666.1 [Sl]     MDNLPKCGANYVPLTPLTFLTRASNSYANRTSIIYANVGFNWRETHERCCRLASSLRSLN        60
                        **:*  * **   ** : ** **  .* ::**:*     *.* :*:.** **:  * ::*

XP_010474116.1 [Cs]     ITRNDVVSILAPNVPVMYEVHFSVPMTGAVLNPINTRLDAKTIAVILRHAEPKIVFVDHE        120
Col983 CNL1             ITKNDVVSILAPNVPAMYEMHFSVPMTGAVLNPINTRLDAKTIAVIVRHAEPKILFVDHE        120
Cr1504 CNL1             ITKNDVVSILAPNVPAMYEVHFSVPMTGAVLNPINTRLDAKTIAVIIRHAEPKILFVDYE        120
CR80TR1-TS1 CNL1        ITKNDVVSILAPNVPAMYEVHFSVPMTGAVLNPINTRLDAKTIAVIIRHAEPKILFVDYE        120
Cg926 CNL1              ITKNDVVSILAPNVPAMYEVHFSVPMTGAVLNPINTRLDAKTIAVIIRHAEPKILFVDYE        120
NP_176764.1 [At]        IGKNDVVSVVAPNTPAMYEMHFAVPMAGAVLNPINTRLDATSIAAILRHAKPKILFIYRS        120
CDY22985.1 [Bn]         IRKNDVVSVLAPNIPAMYEMHFAVPMAGAVLNPLNTRLDAKSITTILRHAQPKIFCIDRS        120
XP_006602034.1 [Gm]     IARNDVVSVLAPNIPAMYEMHFAVPMAGAVLNTINTRLDAKNIATILRHSEAKVFFVDYE        120
AEO52693.1 [Ph]         IVKNDVVSVLAPNVPATYEMHFAVPMAGAVLNTINTRLDPMNIAIILKHSEAKLLFVDYE        120
XP_004232666.1 [Sl]     IVKNDVVSVLAPNVPAMLEMHFAVPMAGAVLNAINTRLDARNVALILKHSEAKIFFIDYE        120
                        * :*****::*** *.  *:**:***:***** :*****  .:: *::*:: *:. :  .

XP_010474116.1 [Cs]     FAPLIQEVFRLLSFGE-------------------SQAHPLIIFINEIGSTTNPSS-KEL        160
Col983 CNL1             FAPLIREVLRLLPSGE-------------------SQAYPRIIFINEIGSTAKPSS-KEL        160
Cr1504 CNL1             FASLIREVLRLLPSGE-------------------SQAYPRIIFINEISSTVKPYS-KEL        160
CR80TR1-TS1 CNL1        FASLIREVLRLLPSGE-------------------SQAYPRIIFINEISSTAKPSS-KEL        160
Cg926 CNL1              FASLIREVLRLLPSGE-------------------SQAYPRIIFINEIGSTAKPSS-KEL        160
NP_176764.1 [At]        FEPLAREILQLLSSED-------------------SNLNLPVIFIHEIDFPKRVSS-EES        160
CDY22985.1 [Bn]         FETLSREILHLLSFDD-------------------SKLNMLVIFIDETDLPKTISS-NEL        160
XP_006602034.1 [Gm]     YVSKAKEALRLLMDDNNNNNLKKGVPKPTNQPHSTTFSLPLVIVIDDINTPTRIRL-GEL        179
AEO52693.1 [Ph]         YLEKARKALELLMSTNFIT-----------AQNSKKISMPQVILIDDLYSPTRIQQQDQL        169
XP_004232666.1 [Sl]     YIDKAKKAIEILMSDF-------------------QMPMPLVVVIDDLDSPTGIRL-GEL        160
                        :    :: :.:*                          ::.*.:           :

XP_010474116.1 [Cs]     DYEGLIRETGTTPSLSATTFLVHDENDPISLNYTSGTTSEPKGVVVSHRGAYLSALSSII        220
Col983 CNL1             DYEGLIRKG--DPKVSSSMFHVHDEHDAISLNYTSGTTSEPKGVVISHRAAYLTVLSSII        218
Cr1504 CNL1             DYEDLIRKG--DPKVSSSMFCVHDEHDPISLNYTSGTTSEPKGVVISHRAAYLTVLSSII        218
CR80TR1-TS1 CNL1        DYEGLIRKG--DPKVSSSIFRVHDEHDPISLNYTSGTTSEPKGVVISHRAAYLTVLSSII        218
Cg926 CNL1              DYEGLIRKG--DPKVSSSMFCVHDEHDPISLNYTSGTTSEPKGVVISHRAAYLTVLSSII        218
NP_176764.1 [At]        DYECLIQRGEPTPLLLARMFCIQDEHDPISLNYTSGTTADPKGVVISHRGAYLSTLSAII        220
CDY22985.1 [Bn]         NYESLIRRGEPTSSMVAHMFRVQNEHDPISLNYTSGTTADPKGVVISHRGAYLSSLGAIM        220
XP_006602034.1 [Gm]     EYEQMVHHGNPNYFP---EGI-QDEWTPIALNYTSGTTSEPKGVVYSHRGAYLSTLSLIL        235
AEO52693.1 [Ph]         EYEQLVHQGNPEYAP---ENIVDDEWDPIVLNYTSGTTSEPKGVVYSHRGAFLSTLNTIM        226
XP_004232666.1 [Sl]     EYEQLVFQGNPEYVV---ENI-DDEWDPITLSYTSGTTSEPKGVVYSHRGAFLSTLSLIL        216
                        :** :: .             .:*   * *.******::***** ***.*:*: *. *:

XP_010474116.1 [Cs]     GWEMAICPVYLWTLPMFHGNGWTHTWSVAARGGTNICIRSVTALEIYKNIDLHEVTHMCC        280
Col983 CNL1             SMEMRLFPVYLWTLPMFHSNGWTHTWSVAARGGTNICMRNVTAQEIYKNIKIHDVTHMSC        278
Cr1504 CNL1             SMEMRLFPVYLWTLPMFHSNGWTHTWSVAARGGTNICMRNVTALDIYKNIELYNVTHMSC        278
CR80TR1-TS1 CNL1        SMEMRLFPVYLWTLPMFHSNGWTHTWSVAARGGTNICMRNVTALDIYKNIELYNVTHMSC        278
Cg926 CNL1              SMEMRLFPVYLWTLPIFHSNGWTHTWSVAARGGTNICMRNVTALDIYKNIELYNVTHMSC        278
NP_176764.1 [At]        GWEMGTCPVYLWTLPMFHCNGWTFTWGTAARGGTSVCMRHVTAPEIYKNIEMHNVTHMCC        280
CDY22985.1 [Bn]         GWEMGTCPVYLWTLPMFHCNGWTLTWSVAARGGTNVCIRHVTASEIFKNIQMHGVTHMCC        280
XP_006602034.1 [Gm]     GWEMGSEPVYLWTLPMFHCNGWTFTWGVAARGGTNVCLRTTAARDIYRNIVVHNVTHMCC        295
AEO52693.1 [Ph]         GWEMGTEPVYLWSLPMFHINGWTLTWGIAARGGTNVCIRNTTAQEIYSNITLHKVTHMCC        286
XP_004232666.1 [Sl]     GWEMGTEPVYLWSLPMFHCNGWTFTWGIAARGGTNVCIRNTTAQEIYSNIALHKVTHMCA        276
                        . **   *****:**:** **** **. ******.:*:* .:* :*: ** :: ****..
```

Figure 2.S4: Multiple sequence alignment of CNL1 protein sequences from different organisms.

Mutations in Cr80TR1-TS1 are highlighted in green, the causal mutation in Cr1504 is highlighted in yellow. Cs: *Camelina sativa*, At: *Arabidopsis thaliana*, Bn: *Brassica napus*, Gm: *Glycine max*, Ph: *Petunia x hybrida*, Sl: *Solanum lycopersicum*.

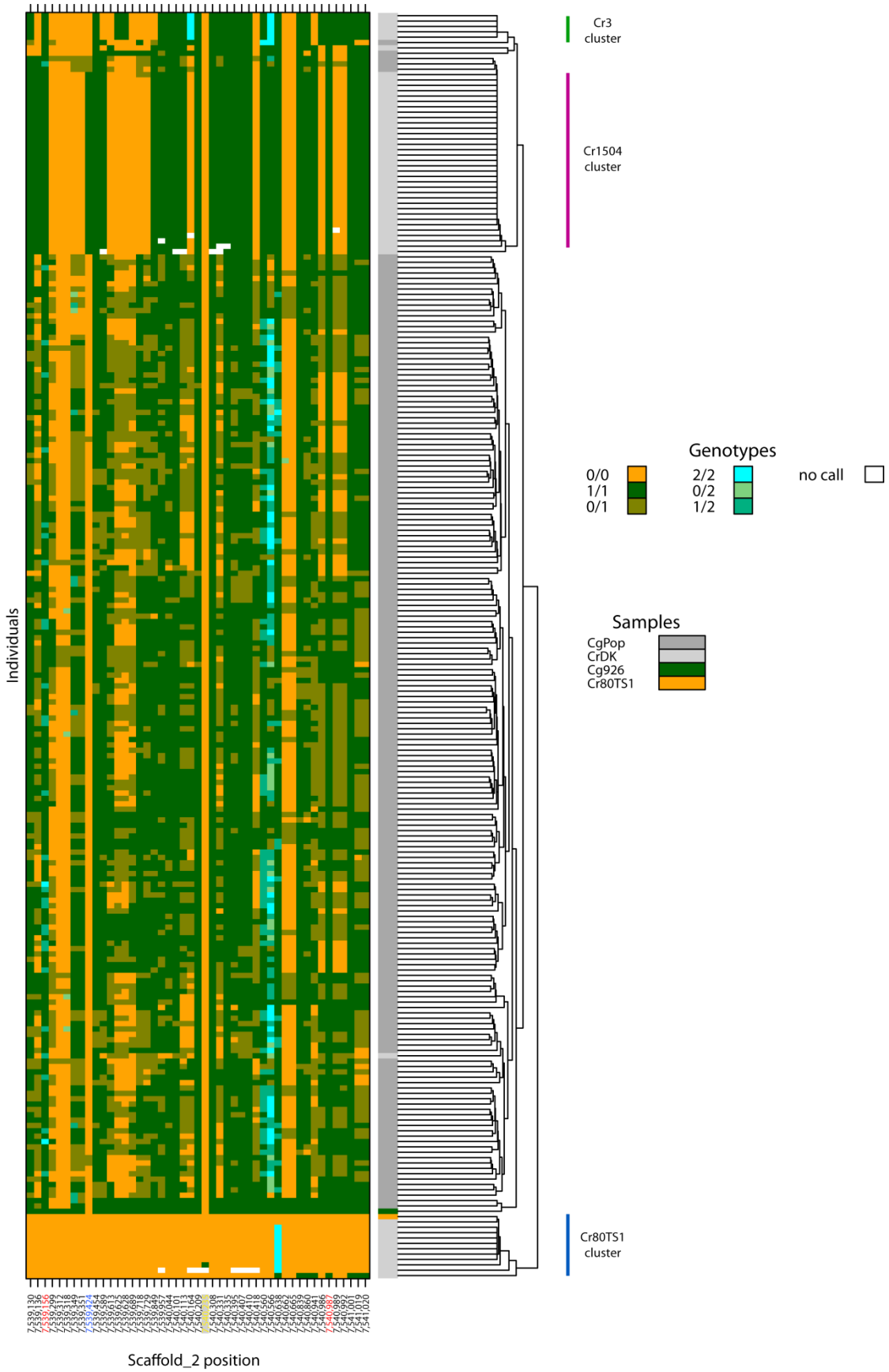
Cr3
cluster
Cr1504
cluster
Genotypes
0/0
1/1
0/1
2/2
0/2
1/2
no call
Samples
CgPop
CrDK
Cg926
Cr80TS1
Cr80TS1
cluster
Individuals
Scaffold_2 position

Figure 2.S5: Variant clustering of diploid CNL1 haplotypes in *C. grandiflora* and *C. rubella* accessions.

Clustering of graphical genotypes is shown for the resequenced *C. grandiflora* and *C. rubella* accessions. Different individuals are shown as rows. The identity of the samples is indicated by the coloured bars next to the cluster dendrogram on the right following the legend at the bottom. Base calls homozygous for the allele from Cr80TR1 are shown in orange (0/0), base calls homozygous for the allele from Cg926 in dark green (1/1) and heterozygous calls are olive green (0/1). Base calls at positions with more than two alleles are indicated as shown (2/2, 0/2, 1/2), '2' being the allele that is not found in either Cg926 nor Cr80TR1. Red font for the nucleotide positions indicates the mutations leading to the loss of enzyme activity in Cr80TR1, blue font indicates the causal mutation from Cr1504, yellow font the 4-bp deletion in CrTAAL. The third allele ('2') at position 7,539,156 in *C. grandiflora* accessions causes a conservative leucine-to-isoleucine exchange that segregates in *C. grandiflora*. The few *C. grandiflora* accessions clustering with the *C. rubella* accessions likely carry an introgressed *Cr1504 CNL1* haplotype (Sas et al., 2016). The three *C. rubella* clusters indicated on the right correspond to the ones shown in Figure 2. 2.

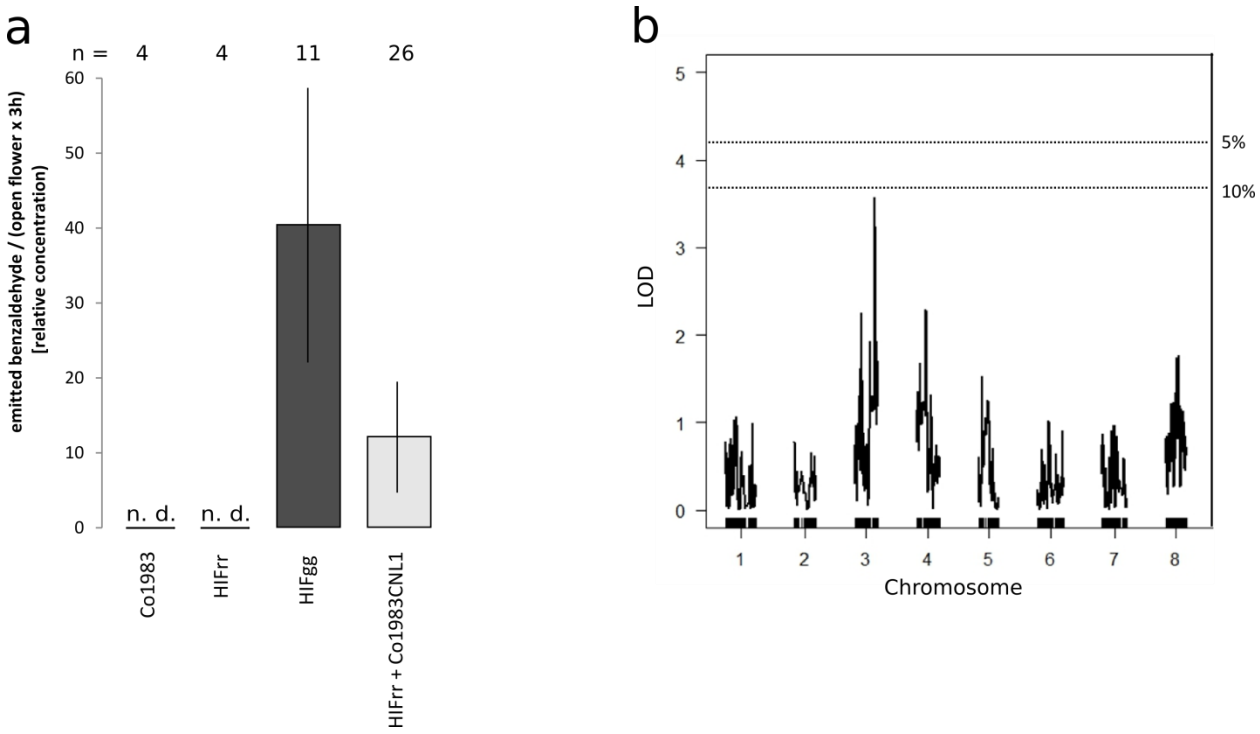

Figure 2.S6: Genetic basis of BAld loss in *C. orientalis*

a) Quantification of emitted BAld in the headspace of the indicated lines 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.

b) QTL mapping of BAld emission in the *C. grandiflora* x *C. orientalis* population. LOD score plot is shown; dashed, horizontal lines indicate the 5% and 10% significance threshold as determined by permutation testing. Ticks on the *x* axes show the positions of individual genetic markers.

2. 9. 2 Supporting Tables

Table 2.S1: Oligonucleotides used in this study

Name	Sequence	Description
OFJ1	GGTGAGAACATTAGTAGTATTGAGGTTG	Site-directed mutagenesis CrCNL1 primer with mutation F
OFJ2	CAACCTCAATACTACTAATGTTCTCACC	Site-directed mutagenesis CrCNL1 primer with mutation R
OFJ3	CATAGCCGTCATCATCCGCCATG	Site-directed mutagenesis CrCNL1 primer for whole fragment, BlpI F
OFJ4	GAACTTTCTTGGAACCGGTATCGTCC	Site-directed mutagenesis CrCNL1 primer for whole fragment, AgeI R
OFJ5	TAATACGACTCACTATAGGG	T7
OFJ6	AATTAACCCTCACTAAAGGG	T3
OFJ9	GTACCCGGGCTCGAGTTAATTCGAGGGAAGTTCATTTCAT	F SLiCE Ω+ML595
OFJ10	CATAAAACTAGTTTATCCATTGTAATTGTAAATAGTAATT	R Ω +CNL1
OFJ11	AATTACTATTTACAATTACAATGGATAAACTAGTTTTATG	F Ω +CNL1
OFJ12	CGACTCTAGAGGATCCTTAATTAAAGCCTTGAACTAACTT	R SLiCE CNL1+ML595
OFJ21	ATTGTTTGCTTCACATAACGCTAGTTTATCCAT	F Start CoCNL1
OFJ22	ATGGATAAACTAGCGTTATGTGAAGCAAACAATG	R Start CoCNL1
OFJ23	GGACCCATGGGCATGGATAAACTAGCGTTATG	NcoI site, start CoCNL1
OFJ24	TAAACTTTGCGGCCGCAAGCCTTGAATTAACATG	NotI Site, end CoCNL1 without stop codon
OFJ25	TAAACTTTGCGGCCGCAAGCCTTGAACTAACTTG	NotI Site, end SAS68CNL1 without stop codon
OFJ26	AATTACTATTTACAATTACAATGGATAAACTAGCGTTATG	F Ω end & SAS68+47 start for SLiCE
OFJ27	CATAACGCTAGTTTATCCATTGTAATTGTAAATAGTAATT	R Ω end & SAS68+47 start for SLiCE
OFJ30	TTAAAGCCTTGAATTAACATGATTC	R end CoCNL1
OFJ31	GAATCATGTTAATTCAAGGCTTTAATTAATTAATGGATCAGC TTT	F linearize pFJ5, overhang end CoCNL1
OFJ32	CATAACGCTAGTTTATCCATTGTAATTGTAAATAGTAATTG	R linearize pFJ5, end Ω, start CoCNL1
OFJ38	GCGCGCCATTAATTAATCCTACTAGTGGAGTGCTG	F pAS77, PacI, end CoCNl1 with terminator
OFJ42	GCTAAGCTTGAGCTCTAGAAGGTTAATTAATCCTACTAGTGG AG	F Slice pBAR CoCNL1
OFJ43	CGGGGGATCCATTAATTAATGGCGTCTCGACTTTGA	R Slice pBAR CoCNL1
OFJ46	GGTGAGAACATTAGTAGAATTGAGGTTG	rubella version point mutation CNL1 F
OFJ47	CAACCTCAATTCTACTAATGTTCTCACC	rubella version point mutation CNL1 R
JL1	CCTCTTACGCCTATAACGTTCTTGAAG	L16P mutation forward primer
JL2	CTTCAAGAACGTTATAGGCGTAAGAGG	L16P mutation reverse primer
JL3	CCTCTAACTCTCATAACGTTCTTGAAG	P16L mutation forward primer
JL4	CTTCAAGAACGTTATGAGAGTTAGAGG	P16L mutation reverse primer

Table 2.S2: Origin of *C. rubella* accessions and assignment to *CNL1* haplotype clusters

accession	latitude	longitude	region	Source	DK-SRA	Cluster
72.12	38.16	22.72		Koenig2018	DKCr172	Cr3
879	35.29	24.42		Koenig2018	DKCr6	Cr3
1408	35.29	24.42		Koenig2018	DKCr200	Cr3
1407-8	35.18	24.23		Koenig2018	DKCr36	Cr3
1411-3	35.29	24.42		Koenig2018	DKCr207	Cr3
6.26	44.56	0.35		Koenig2018	DKCr185	Cr1504
22.13	41.29	13.95		Koenig2018	DKCr189	Cr1504
23.9	40.72	15.2		Koenig2018	DKCr188	Cr1504
34.11	38.16	16.05		Koenig2018	DKCr173	Cr1504
77.16	38.18	20.57		Koenig2018	DKCr234	Cr3
84.15	38.48	21.43		Koenig2018	DKCr182	Cr1504
100.8	39.71	21.63	Coordinates of closest city (Kalabaka, Greece)	Weigel table	DKCr77	Cr1504
690	36.15	-5.58		Koenig2018	DKCr1	Cr1504
697	NA	NA	Italy	Guo2009	DKCr2	Cr1504
698	NA	NA	Italy	Guo2009	DKCr3	Cr1504
774	41.83	16	(Monte Gargano, Italy), MTE	Koenig2018	DKCr20	Cr1504
907	39.67	19.8		Koenig2018	DKCr7	Cr1504
925	39.67	20.85		Koenig2018	DKCr8	Cr1504
984	39.5	3		Koenig2018	DKCr9	Cr1504
1207	28.32	-16.57		Koenig2018	DKCr11	Cr1504
1208	28.32	-16.57		Koenig2018	DKCr12	Cr1504
1209	28.32	-16.57		Koenig2018	DKCr13	Cr1504
1311	42.88	-0.1		Koenig2018	DKCr15	Cr1504
1377	-34.67	-58.5		Koenig2018	DKCr16	Cr1504
1453	43.47	11.03		Koenig2018	DKCr18	Cr1504
1245-12	39.43	-6.33	Spain	Koenig2018	DKCr63	Cr1504
1249-11	38.78	-9.38		Koenig2018	DKCr52	Cr1504
1267-15	37.93	-6.88		Koenig2018	DKCr53	Cr1504
1314-10	47.32	5.02		Koenig2018	DKCr210	Cr1504
1316-5	47.1	4.93		Koenig2018	DKCr71	Cr1504
1319-3	47.37	3.98		Koenig2018	DKCr212	Cr1504
1321-1	46.95	4.3		Koenig2018	DKCr73	Cr1504
1354-12	45.88	10.83		Koenig2018	DKCr64	Cr1504
1504-11	28.67	-17.87		Koenig2018	DKCr38	Cr1504
1574-1	41.6	8.98		Koenig2018	DKCr30	Cr1504
1575-1	41.38	9.17		Koenig2018	DKCr40	Cr1504
1GR1	37.78	26.83	(Manolates, Greece)	Guo2009	NA	Cr1504
86IT1	40.62	14.37		Koenig2018	DKCr22	Cr1504
987-25	45.03	-0.58		Koenig2018	DKCr59	Cr1504
RIAH	NA	NA	Italy		DKCr202	Cr1504
1215	28.19	-16.19	Teneriffe, Spain	Guo2009	DKCr14	Cr1504
104.12	NA	NA		Weigel table	DKCr180	Cg
39.1	37.99	15.34		Koenig2018	DKCr83	Cr80TS1
75.2	38.09	22.17		Koenig2018	DKCr79	Cr80TS1
78.1	37.69	31.63		Koenig2018	DKCr168	Cr80TS1
79.1	37.73	21.69		Koenig2018	DKCr78	Cr80TS1
81.2	37.3	22.06		Koenig2018	DKCr228	Cr80TS1

83.1	38.44	21.42		Koenig2018	DKCr171	Cr80TS1
762	37.97	23.72		Koenig2018	DKCr4	Cr80TS1
844	35.2	24.23		Koenig2018	DKCr5	Cr80TS1
80TR1-TS1	41.02	28.96		Koenig2018	DKCr81	Cr80TS1
TAAL-1	36.61	4.11	Coordinates of closest city (Taguemont, Algeria)	Guo2009	DKCr82	Cr80TS1

2. 10 Transcriptome analysis of genes involved in the benzenoid pathway

The following analysis was not included in the manuscript. It is based on a data set by Natalia Wozniak and Adrien Sicard which was not ready for publication at the time of submission of my publication and therefore we decided to exclude it.

The selfing species *C. rubella* and *C. orientalis* do not emit benzaldehyde, the major compound of floral scent in *Capsella* (Sas et al., 2016). Our analyses lead to the suggestion that the loss of BAld emission in *C. orientalis* results from polygenic downregulation of the benzenoid pathway, as we could not detect a major QTL for scent loss in *C. orientalis,* the enzyme CoCNL1 is still functional and flowers of this species still contain considerable amounts of BAld. Other benzenoids emitted by *C. orientalis* flowers could not be robustly detected with the used techniques.

To identify candidate genes that could underlie the loss of BAld emission in *C. orientalis*, we analysed a transcriptomics data-set based on RNA-seq from three developmental stages (seedlings; young flower buds; older flower buds and open flowers) of the three *Capsella* species under study (Wozniak et al, in preparation). We focused on known genes from the benzenoid/phenylpropanoid pathway and transporters involved with active transport of volatiles (Figure 2.5).

Consistent with previous results, *CNL1* expression was not different between *C. grandiflora* and *C. rubella* (Sas *et al.*, 2016), and was only slightly reduced in *C. orientalis*. The phenylpropanoid-pathway genes with a significant downregulation in *C. orientalis* flowers were *Carub10018700* (the putative orthologue of the transcription factor gene *EOBII* that activates synthesis of phenylalanine; (Spitzer-Rimon *et al.*, 2010)*,* three genes encoding phenylacetaldehyde synthase (PAAS) isoforms (catalyses the formation of phenylacetaldehyde, a floral scent compound in *Petunia* (Kaminaga et al., 2006)), *Carub10028089* (the putative orthologue of *Petunia TE1*, encoding a peroxisomal thioesterase involved in the β-oxidative benzenoid pathway (Adebesin *et al.*, 2018)), and *BSMT1* (encoding a benzoic acid/salicylic acid carboxyl methyltransferase) (Figure 2. 5). However, the orthologue of the *Petunia ABCG1* gene encoding an ABC transporter that facilitates volatile emission from petunia flowers (Adebesin *et al.*, 2017) did not show any significant changes in expression in late-stage, volatile-emitting flowers between the three species. Thus, while it is tempting to speculate that the reduced expression of the putative *EOBII* orthologue in the selfing species may contribute to reduced availability of phenylalanine as the precursor to the phenylpropanoids, none of the other observed expression changes provides a compelling candidate for explaining the lack of BAld emission from *C. orientalis*

flowers. This is consistent with our conclusion from the above QTL mapping that no single major-effect mutation appears to be responsible for the lack of BAld emission in this species.

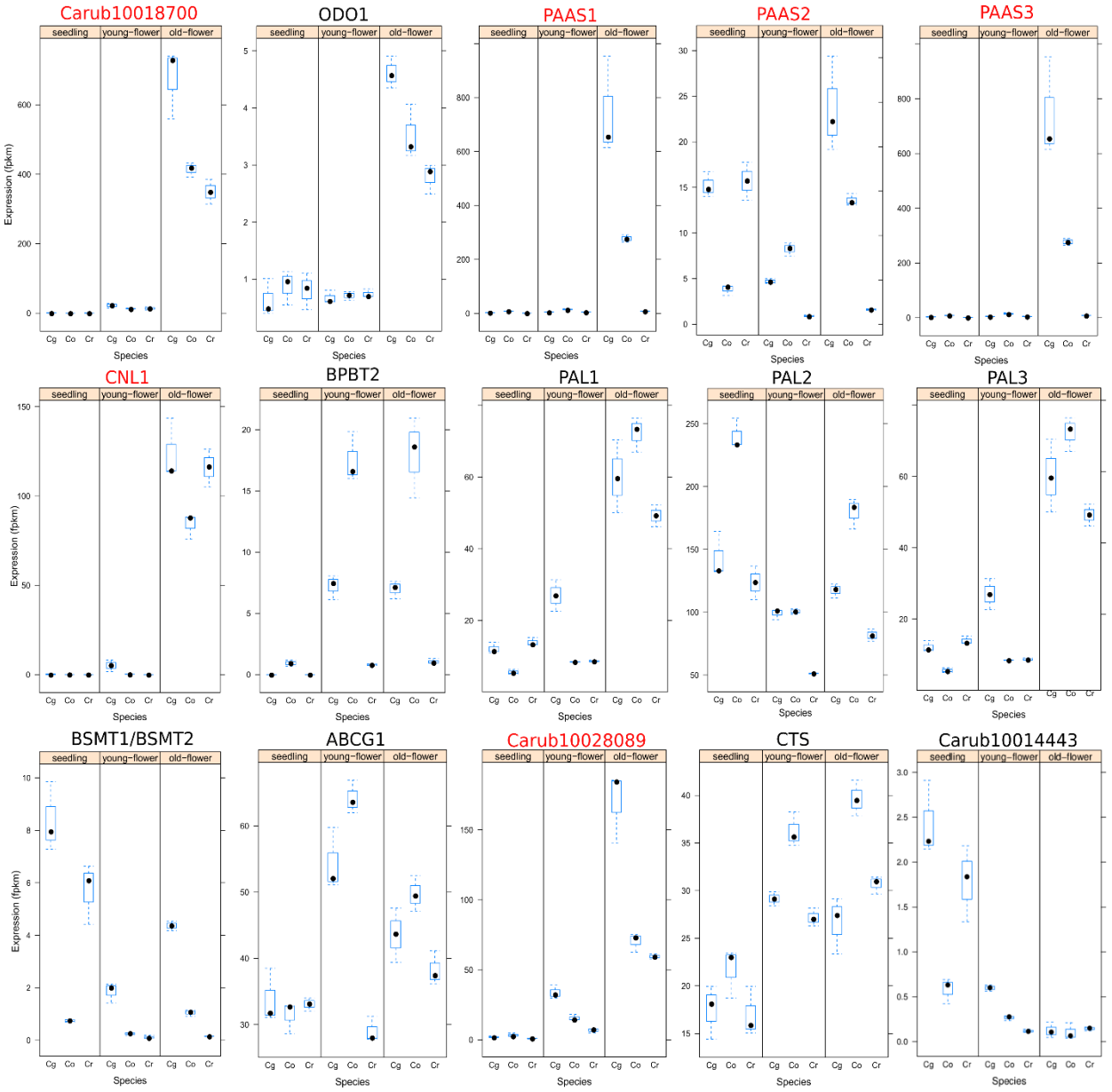

Figure 2.5: Expression levels of putative *Capsella* homologues for phenylpropanoid/benzenoid pathway genes and additional genes involved in BAld emission in seedlings, young flowers and old flowers of the three different standard accessions Cg926, Co1983 and Cr1504.

Genes with reduced expression in *C. orientalis* are highlighted in red. ODO1: ODORANT1; PAAS: phenylacetaldehyde synthase; CNL: cinnamate:CoA ligase; BPBT: benzoyl-CoA: benzylalcohol/ 2-phenylethanol benzoyltransferase; PAL: phenylalanine ammonia lyase; BSMT: benzoic acid/salicylic acid carboxyl methyltransferase; . ABCG1: orthologue of *Petunia* ATP-binding cassette transporter; CTS: orthologue of *Arabidopsis* peroxisomal ABC transporter COMATOSE (Bussell *et al.*, 2014). Carub10018700 is the putative orthologue of the transcription factor gene *EOBII* that activates synthesis of phenylalanine, Carub10028089 is the orthologue of *Petunia* thioesterase 1 and Carub10014443 is the putative orthologue of *PH4* (a MYB transcription factor known to regulate scent emission in *Petunia*).

3. Manuscript: A high-throughput amplicon-based method for estimating outcrossing rates

The methods part of this manuscript was written by me except the chapters about PCR primer design, sequence analysis and estimation of outcrossing rates. I contributed to the other parts of the manuscript. Experiments illustrated by figures 1, 2, 3, 4 and 5 were performed, analyzed and prepared for publication by me. I conducted the common-garden experiment including plant care and harvesting, DNA extractions from pooled seeds and the sequencing library preparation for two consecutive years. The bioinformatical analysis was done by Christian Kappel.

We are planning to submit this manuscript to Plant Methods in April 2019.

A high-throughput amplicon-based method for estimating outcrossing rates

Friederike Jantzen[1], Natalia Wozniak[1,2], Christian Kappel[1], Adrien Sicard[1,3], Michael Lenhard[1,4]

[1] University of Potsdam, Institute for Biochemistry and Biology, Karl-Liebknecht-Str. 24-25, house 26, D-14476 Potsdam-Golm, Germany

[2] present address: Ruhr University Bochum, Molecular Genetics and Physiology of Plants, Faculty of Biology and Biotechnology, Universitätsstraße 150, building ND North, D-44801 Bochum, Germany

[3] present address: Department of Plant Biology, Swedish University of Agricultural Sciences and Linnean Center for Plant Biology, Uppsala, Sweden.

[4] author for correspondence:

email: michael.lenhard@uni-potsdam.de

telephone: +49-331-9775580

email addresses:

jantzen1@uni-potsdam.de

adrien.sicard@slu.se

nwozniak@uni-potsdam.de

christian.kappel@uni-potsdam.de

3. 1 Abstract

- **Background:** The outcrossing rate is a key determinant of the population-genetic structure of species and their long-term evolutionary trajectories. However, determining the outcrossing rate using current methods based on PCR-genotyping individual offspring of focal plants for multiple polymorphic markers is laborious and time-consuming.
- **Results:** We have developed an amplicon-based, high-throughput enabled method for estimating the outcrossing rate and have applied this to an example of scented versus non-scented *Capsella* (Shepherd's Purse) genotypes. Our results show that the method is able to robustly capture differences in outcrossing rates. They also highlight potential biases in the estimates resulting from differential haplotype sharing of the focal plants with the pollen-donor population at individual amplicons.
- **Conclusions:** This novel method for estimating outcrossing rates will allow determining this key population-genetic parameter with high-throughput across many genotypes in a population, enabling studies into the genetic determinants of successful pollinator attraction and outcrossing.

Keywords

outcrossing, mixed mating, outcrossing rate, Capsella, amplicon sequencing

3. 2 Background

The rate at which individuals in a population outcross has a major impact on the genetic structure of the population and its responses to natural selection (1, 2). While outcrossing maximizes the heterozygosity in a population, selfing or inbreeding between relatives increases homozygosity. This, in turn, has a number of consequences, such as the phenotypic expression of recessive deleterious mutations, also known as inbreeding depression (3), and a reduced rate of effective recombination, as crossing over between homozygous chromosomes does not lead to the formation of genetically recombinant gametes (4). Over time, such a reduced effective rate of recombination leads to an increased length of haplotype blocks in linkage disequilibrium and of linked selection (4). In addition, inbreeding reduces the effective population size, and as a result, the relative importance of genetic drift increases compared to that of selection (5). The reduced efficacy of purifying selection also increases the risk of fixation of deleterious mutations and influences species extinction rates (6). Thus, the outcrossing rate is a key determinant of several population-genetic parameters with a major influence on long-term evolutionary trajectories of populations (2, 7).

In contrast to animals, where outbreeding enforced by dioecy is seen in the majority of species (8), most flowering plants are hermaphrodites (9). While many plant lineages have evolved genetic self-incompatibility and other mechanisms to enforce or promote outbreeding, mixed mating is very common in plants (10). In mixed-mating species, a fraction of the progeny of a plant is derived from selfing, while the rest is the result of outbreeding. Therefore, estimating the outcrossing rate of plants in a population is an important aspect of studies in plant reproductive systems.

Classically, the outcrossing rate is estimated by genotyping a large number of progeny individuals from a focal individual for several microsatellite or SNP markers and determining the fraction of genotypes that cannot have been produced by selfing at each marker (11, 12). From these data, rates of outcrossing and other parameters of the breeding system can then be estimated (13, 14). While this approach can provide a rich and nuanced picture of the breeding system in a population, it is laborious and thus not readily amenable to be used in a high-throughput manner. Examples for questions that require such a high-throughput approach would be the following. How does the breeding system of a species depend on different environmental conditions? Are rates of outcrossing stable within a population over different years? And how does variation in floral characteristics influence outbreeding rates? Answering this kind of question requires estimating outcrossing rates for a large number of focal individuals, which would be prohibitive when done by genotyping many progeny individuals per focal plant.

Concrete examples for the latter question are studies questioning the relevance of different traits presumed to help in pollinator attraction, such as large and showy petals, emission of floral scent, and nectar amount and composition (15). These traits often undergo large changes, when the breeding system changes from predominant outbreeding to selfing (16). This transition is generally accompanied by the evolution of the so-called selfing syndrome, comprising a reduction in flower size, especially that of petals, in scent and nectar production and in the ratio of pollen to ovules per flower. One example where the genetic basis of the evolution of the selfing syndrome is being studied is the genus *Capsella* (17, 18). This genus contains three diploid species, two of which (*C. rubella* and *C. orientalis*) represent independently derived selfers that have diverged from an outbreeding ancestor represented by present-day *C. grandiflora* between 100,000 and 200,000 years ago and between one and two million years ago, respectively (19-21). Several loci have by now been identified that have contributed to the reduction in petal size and in floral scent in *C. rubella* compared to *C. grandiflora* (22-24). While this is starting to shed light on the molecular basis and evolutionary history of selfing-syndrome traits, understanding the ecological consequences of changes in presumed pollinator-attraction traits remains a major challenge. That said, several key biological materials have become available as part of the process of gene identification - such as quasi-isogenic lines that only segregate for a very small chromosomal segment containing a causal gene, but are essentially isogenic otherwise - that will enable rigorously testing the effect of a given trait change on the interaction with pollinators and herbivores, including on the outcrossing rate. However, as outlined above, such studies would greatly benefit from a high-throughput method for estimating outcrossing rates from many individual plants differing in a gene and thus a trait of interest.

Against this background of work on *Capsella*, we set out to establish and evaluate a high-throughput compatible method for estimating and comparing outcrossing rates. This method is based on Illumina sequencing of PCR fragments amplified from pooled progeny individuals of a plant and estimating the outcrossing rate from the frequency of non-maternal haplotypes.

3. 3 Methods

3. 3. 1 Reagents

Gibberellic acid 4 + 7 (Duchefa Biochemie)
Ethanol (Carl Roth)
Paper bags for bagging plants (HERA)
Bird protection mesh (mesh size 25mm) (Zill GmbH Co. KG)
Insect protection mesh (mesh size 0.6mm) (Grow it)
2 ml and 1.5 ml tubes
96 well PCR plates (Sarstedt)
Foil/lids to seal plates
384 Well Lightcycler plates (Sarstedt)
Adhesive Optical film (Biozym Scientific)
Nuclease-free water
Liquid nitrogen
Quiagen DNeasy Plant Mini Kit (Quiagen)
AMPure XP beads (Beckmann Coulter)
Magnetic stand (Applied Biosystems)
KAPA HiFi Hotstart PCR Kit with dNTPs (Roche)
DMSO (Carl Roth)
ROX solution (ThermoFisher Scientific)
SYBR Green I Nucleic acid stain (Sigma-Aldrich)
TE buffer (10 mM Tris-HCl, pH 7.5)
QuantiFluor dsDNA System (Promega)
Tapestation reagents (Agilent Technologies)
Qubit reagents (ThermoFisher Scientific)
NextSeq Reagent Kit Mid Output 300 cycles (Illumina)

3. 3. 2 Equipment

Mortar and Pistil
Multichannel and single-channel pipettes
LightCycler 480 II (or similar) (Roche)
Mastercycler nexus (or similar) (Eppendorf)
Speedvac RVC 2-18 (Martin Christ Gefriertrocknungsanlagen)
NextSeq (Illumina)
Centrifuges (for plates and tubes)
2200 TapeStation (Agilent Technologies)
Qubit 2.0 (ThermoFisher Scientific)

3. 3. 3 Plant material and growth conditions

Two *Capsella* quasi-isogenic lines only differing in about 12 kb around the locus for loss of benzaldehyde emission were compared here (23). A self-compatible *Capsella grandiflora* line and qILs segregating for petal size were used as sources to provide pollen for outbreeding (17, 24). At the beginning of April 2017, seeds were sown on a soil-compost mix and watered with GA-supplemented water (gibberellic acid stock: 16,5 mg in 1 mL ethanol, 1:5000 dilution for watering) until germination. After sowing, seeds were stratified for four days in 4°C and then transported into an open greenhouse. Seedlings were pricked out into individual pots when they showed 4-6 leaves (approximately three weeks after germination) and were planted out in the plots just before starting to flower in mid-May. Bird or insect protection nets were installed. The set-up is shown in Figure 3.1. Plants flowered during June and July and were bagged in July to be harvested in August. When the oldest fruits started to ripen, individual plants were bagged and allowed to ripen for two more weeks before collection and preparation for DNA extractions. Due to weather conditions and infection with herbivores low numbers of seeds were harvested from each plant. Seeds from all plants with the same genotype within one plot were pooled and cleaned.

3. 3. 4 DNA extraction from pooled seeds

This method was optimized for extracting DNA from *Capsella* seeds. Depending on seed number/size it may need to be adapted. Other extraction methods such as CTAB (25) did not give any results.

1. Approximately 300 seeds were counted into a 2 ml Eppendorf tube to reduce contamination with sand and other dirt. Add sterilized water and incubate seeds for two days at 4°C.
2. Transfer seeds into mortar, remove water by pipetting. Wash 2-3 times with water to remove dirt if necessary.
3. Cool down mortar and pistil with liquid nitrogen and grind samples to a fine powder. Transfer powder into a new 2ml tube and add 800 µl Buffer AP1 of Quiagen DNeasy Plant Mini Kit. Store on ice until all samples are ground.
4. Proceed with following steps from Quiagen Kit manual, add double amount of buffer AP2 in step 3. For elution in step 12, 2x 50 µl buffer AE were used.
5. Test extracted DNA in a PCR. If there is no amplification and starting material contained a lot of soil or organic matter, try cleaning DNA samples with e.g. AMPure beads.
6. Transfer DNA samples into 96-well plate for further steps.

3. 3. 5 Design of PCR primers

Next-generation sequencing libraries relying on PCR-based approaches are known to be prone to numerous biases linked to amplicon length and GC-content, sequence heterogeneity at primer-annealing sites as well as copy number variation (26-29). To limit amplification biases, we focused our analysis on low polymorphic genes present in a single copy within genomes. To this end, we retrieved sequences from single-copy nuclear genes from *C. rubella* reference genome (https://phytozome.jgi.doe.gov/pz/portal.html) based on (30). Conserved sites within the *Capsella* genus were identified by comparing the sequence of 50 *C. rubella* and 193 *C. grandiflora* genomes. Primers were designed using Primer 3 plus in order to amplify amplicons of approximately 300 pb and to anneal to their templates at 55°C. We added 33 and 34 nt sequences complementary to the forward and reverse index primers to each of the gene-specific forward and reverse primers, respectively.

Sequencing quality will also depend on the heterogeneity of sequences as it is used during the first amplification cycles for cluster identification and phasing/pre-phasing calibration (31). Low-sequence heterogeneity may impair the distinction between different clusters and considerably limit the output of the sequencing run. A common solution to such issues is to co-sequence amplicon libraries with a heterogeneous control library (e.g. usually prepared from the bacteriophage PhiX genome and mixed at variable proportions, between 15 to 50%) with the drawback that a large number of reads will be lost as they will not correspond to the sequence of the target loci. Here, we introduce sequence heterogeneity by designing an additional primer pair for each target loci containing an additional nucleotide between the gene-specific sequence and those complementary to the indexing primers (Table 3.1). [Please note that we still added 15% PhiX genome because we were not sure if these primers would add enough heterogeneity]. Indexing primers from Illumina are shown in table 3.2.

Table 3.1: Primers for Primary PCR

Name	Sequence
1_Carubv10013869m_F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCACTCATCCATTCGGAAAT
1_Carubv10013869m_R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTTGGGGACAAGGTGCTAATC
2_Carubv10023806m_F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTACCGACCACATAGGCATCA
2_Carubv10023806m_R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAATGGCCGATTCTGCTTTTA
3_Carubv10018138m_F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCAAGCCAAAGTTTGATGCTT
3_Carubv10018138m_R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACTCGTCTGCAGTCATGGTG
4_Carubv10001640m_F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGAAGCGGATGGTTACAAAA
4_Carubv10001640m_R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGGCCAAGCTCACTCACATT
5_Carubv10001924m_F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGGGTTCAGATTGAGCGTAA

5_Carubv10001924m_R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAACTTGATCCTCTTTGGTACTGG
6_Carubv10023818m_F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTTCTTTTTCTGAGATTCCATTGCT
6_Carubv10023818m_R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGAAGCCTCTCCTGAGAAGTGA
7_Carubv10005658m_F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTCCAAGATCTGTGCTTGCTG
7_Carubv10005658m_R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCAGCTCCGGATGGTTAAAT
8_Carubv10006001m _F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTTTCAAAAGCTTTGCGTGAG
8_Carubv10006001m_R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGATGCTTCACGTTCACACCA
9_Carubv10006101m_F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTTCTATCCAAGGGCCATCA
9_Carubv10006101m_R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCCCATGGAAACTCCTTGTTG
10_Carubv10027375m_F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGATCCGTCGGCTCTTCTCTC
10_Carubv10027375m_R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAACCATGCCAATGCTTCATA
11_Carubv10011729m_F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGAGCAAGTCCCAAACAAAG
11_Carubv10011729m_R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCATTTCAAGCCGCTCTGG
12_Carubv10014733m_F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGCATTCGATCTCGATCTTG
12_Carubv10014733m_R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCGGTGGTGAAGACAACAATC
Primers targeting same sequences as above, but with added T or A before gene-specific sequence for heterogeneity	
1A_Carubv10013869m_F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGACCACTCATCCATTCGGAAAT
1A_Carubv10013869m_R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGATTGGGGACAAGGTGCTAATC
2T_Carubv10023806m_F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTTACCGACCACATAGGCATCA
2T_Carubv10023806m_R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTAATGGCCGATTCTGCTTTTA
3A_Carubv10018138m_F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGACAAGCCAAAGTTTGATGCTT
3A_Carubv10018138m_R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAACTCGTCTGCAGTCATGGTG
4T_Carubv10001640m_F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGGAAGCGGATGGTTACAAAA
4T_Carubv10001640m_R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTAGGCCAAGCTCACTCACATT
5A_Carubv10001924m_F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGATGGGTTCAGATTGAGCGTAA
5A_Carubv10001924m_R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAAACTTGATCCTCTTTGGTACTGG
6T_Carubv10023818m_F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTTTCTTTTTCTGAGATTCCATTGCT
6T_Carubv10023818m_R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTAGAAGCCTCTCCTGAGAAGTGA
7A_Carubv10005658m_F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGATCCAAGATCTGTGCTTGCTG
7A_Carubv10005658m_R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGATCAGCTCCGGATGGTTAAAT
8T_Carubv10006001m _F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTTTTCAAAAGCTTTGCGTGAG
8T_Carubv10006001m_R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTGATGCTTCACGTTCACACCA
9A_Carubv10006101m_F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGTTCTATCCAAGGGCCATCA
9A_Carubv10006101m_R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACCCATGGAAACTCCTTGTTG
11T_Carubv10011729m_F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGGAGCAAGTCCCAAACAAAG
11T_Carubv10011729m_R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCATTTCAAGCCGCTCTGG
12A_Carubv10014733m_F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGATGCATTCGATCTCGATCTTG
12A_Carubv10014733m_R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACGGTGGTGAAGACAACAATC

Table 3. 2: Primers for Indexing PCR (Indexed sequencing primers).

Name	Sequence
F1_MetaIndex	AATGATACGGCGACCACCGAGATCTACACTATAGCCTTCGTCGGCAGCGTC
F2_MetaIndex	AATGATACGGCGACCACCGAGATCTACACATAGAGGCTCGTCGGCAGCGTC
F3_MetaIndex	AATGATACGGCGACCACCGAGATCTACACCCTATCCTTCGTCGGCAGCGTC
F4_MetaIndex	AATGATACGGCGACCACCGAGATCTACACGGCTCTGATCGTCGGCAGCGTC
F5_MetaIndex	AATGATACGGCGACCACCGAGATCTACACAGGCGAAGTCGTCGGCAGCGTC
F6_MetaIndex	AATGATACGGCGACCACCGAGATCTACACTAATCTTATCGTCGGCAGCGTC
F7_MetaIndex	AATGATACGGCGACCACCGAGATCTACACCAGGACGTTCGTCGGCAGCGTC
F8_MetaIndex	AATGATACGGCGACCACCGAGATCTACACGTACTGACTCGTCGGCAGCGTC
R13_MetaIndex	CAAGCAGAAGACGGCATACGAGATGTCGTGATGTCTCGTGGGCTCGG
R14_MetaIndex	CAAGCAGAAGACGGCATACGAGATCGAGTAATGTCTCGTGGGCTCGG
R15_MetaIndex	CAAGCAGAAGACGGCATACGAGATTCTCCGGAGTCTCGTGGGCTCGG
R16_MetaIndex	CAAGCAGAAGACGGCATACGAGATAATGAGCGGTCTCGTGGGCTCGG
R17_MetaIndex	CAAGCAGAAGACGGCATACGAGATGGAATCTCGTCTCGTGGGCTCGG

3. 3. 6 PCR amplification and library generation

The method for PCR amplification and library preparation is based on a recent protocol for amplicon-based microbiome characterization (32). Further details and troubleshooting information can be found in (32).

This protocol is optimized for low sample DNA concentrations due to the availability and quality of starting material. For best results, minimize PCR cycles and select the samples with the highest concentrations in step 1c. Sample dilution might result in bottlenecking for low abundance alleles as described (32).

1. Primary PCR
    a) Create a sample dilution plate

        Vortex DNA samples and spin down in a centrifuge. Prepare a 384-well plate by pipetting 18 µl water in quadrant 2 (A02), 3 (B01) and 4 (B02). Dispense 10 µl of undiluted sample into quadrant 1 (A01). Generate a 10 fold dilution series by transferring 2 µl of each sample to quadrant 2. Pipet up and down ten times for mixing. Repeat for quadrants 3 and 4. In the end, you will have your undiluted sample in quadrant 1, 1:10 dilution in quadrant 2, 1:100 dilution in quadrant 3 and 1:1000 dilution in quadrant 4.

For primary PCR, pipet 3 µl of every sample and dilutions into a new 384-well plate which can be used in LightCycler 480 II. Start with the lowest concentration in quadrant 4 (B02) and then proceed with quadrant 3, 2 and 1 to use the same set of tips. Plates can be stored at -20°C or used for subsequent primary PCR.

b) Primary PCR

Prior to PCR, test primers individually and pooled for amplification of target genes. Figure 3.2A shows amplification products for 11 out of 12 primer pairs and primer pair 10 was excluded for further steps, in Figure 3.2B the pooling strategy was tested for two sets of primer pairs. For the primary PCR, a maximum of six different primer pairs were pooled (including primer pairs with added bases for heterogeneity, so 12 primer pairs overall).

Defrost the Kapa HiFi Hotstart kit reagents and the 384-sample plate if it was stored in the freezer. Vortex and centrifuge all reagents when thawed before using.

Prepare a 2x KAPA HiFi Hotstart qPCR master mix with the following ingredients: 1.2 µl 5x KAPA HiFi Fidelity buffer, 0.18 µl 10 mM dNTPs, 0.3 µl DMSO, 0.12 µl ROX (25 µM), 0.003µl 1000x SYBR Green, 0.12 µl KAPA HiFi Hotstart Polymerase, 0.3 µl forward primer pool (10µM), 0.3 µl reverse primer pool (10µM) and 0.48 µl nuclease-free water.

Dispense 3 µl of 2X KAPA HiFi Hotstart qPCR master mix in each reaction well on the 384-well plate containing DNA samples for a final volume of 6 µl.

Cover the plate, mix and spin-down. Start the following qPCR protocol on Roche LightCycler 480 II (or similar) after loading the plate: 95°C for 5 min, then 15 cycles of 98°C 20 sec, 55°C for 15 sec, 72°C for 1 min. Cycle number can be between 15 and 30 cycles but optimal results will be achieved by keeping the cycle number low. However, samples amplifying poorly could be amplified using more cycles. Plates can be stored at -20°C.

c) Analysis and choosing the best dilution for indexing PCR

The analysis was done manually. For more details on how to conduct it automatically, please refer to (32).

Compare the amplification curves for the different dilutions of the same DNA sample. Choose a sample concentration which is in the mid-to-late exponential phase at the last amplification cycle of the PCR for further steps. Samples should not have reached a plateau at the final cycle as this means overamplification. Figure 3.3A and B show two examples of amplification curves, due to low DNA concentrations samples were only in the early-exponential phase but were still used for further steps.

Prepare 96 well plate by distributing 18 µl water in each well. Spin down qPCR plate and transfer 2 µl of the appropriate dilution into the new plate to create a 1:10 dilution of the primary PCR. Mix and spin down. (If samples were in mid-to-late exponential phase after the final PCR cycle, transfer 5 µl of 1:10 dilution into a new plate containing 45 µl water to generate a 1:100 dilution of primary PCR. Use this instead of 1:10 dilution for further steps). Store at -20°C or progress to indexing PCR.

2. Indexing PCR

a) Picking an indexing scheme

Each sample needs to have an individual combination of i5 and i7 indices to make sure no overlap among pooled samples can occur. Prior to running the PCR an i5 and i7 dual-indexing scheme needs to be chosen, depending on how many samples will be pooled together for sequencing. The number of samples is equal to the number of index combinations needed.

See Illumina guide for more information on dual indexing: https://support.illumina.com/content/dam/illumina-support/documents/documentation/system_documentation/miseq/indexed-sequencing-overview-guide-15057455-04.pdf

For the design of Illumina adapter sequences see: https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/experiment-design/illumina-adapter-sequences-1000000002694-09.pdf

b) Indexing PCR

Prepare an oligo plate, adapted to your indexing scheme.

Make 10 µM dilutions of your 100 µM primer stocks in a 96 well plate by adding 30 µl 100 µM oligo stock to 270 µl water; forward primers were arranged in columns and reverse primers in rows.

For the 5 µM master plate containing 40 µl primer mix, add 10 µl of the 10 µM forward and reverse primer dilutions into a new plate and add 20 µl water. Mix well and use 2 µl for indexing PCR. See Supplemental Table 3.S1 for indexing scheme used in this study. [Please note that of the 38 different index combinations shown in Supplemental Table 3.S1, only 12 were used for samples analyzed in this study.]

Defrost the KAPA HiFi Hotstart PCR kit reagents and the primary PCR dilution plate (1:10 or 1:100) if necessary. Mix and spin down reagents.

Prepare a 3.33 KAPA HiFi Hotstart Indexing PCR master mix consisting of these ingredients: 2 µl 5x KAPA HiFi Fidelity buffer, 0.3 µl 10 mM dNTPs, 0.5 µl DMSO, 0.2 µl KAPA HiFi Hotstart Polymerase. Dispense 3 µl into wells of a 96-well PCR plate. Add 5 µl of diluted primary PCR product to the corresponding wells in the 96-well PCR plate that contains 3 µl of the indexing PCR master mix. Add 2 µl of 5 µM indexing primers from the prepared oligo plate for indexing. The final reaction volume is 10 µl.

Seal the plate, mix and spin down. Amplify in Mastercycler nexus (or similar) with the following conditions: 95°C for 5 min, 10 cycles of 98°C for 20 sec, 55°C for 15 sec, 72°C for 1 min. Centrifuge the plate to collect the samples after PCR program is complete. The plate can be stored at -20°C.

3. Normalization and pooling

There are different options for normalization and pooling, depending on available systems and reagents. Please check protocol by Gohl et al. for further information (32).

a) QuantiFluor quantification of indexed samples

Calculate DNA concentrations of indexed samples following the manufacturer's protocol for QuantiFluor dsDNA system. Use two times 1 µl indexing PCR reaction per sample for quantification, so 8 µl are left for normalization.

b) Sample normalization

Choose a concentration for normalization. It depends on how concentrated/diluted your samples are. Calculate the amount of water for each sample to be added to a fixed amount of indexing PCR in order to get the desired concentration. For this experiment, 4 µl indexing PCR were used and 5 ng/µl were determined as desired concentration. The amount of water to add fluctuated from 0 to 90 µl, depending on sample concentration.

Seal plate, mix and spin down. The plate can be stored at -20°C.

c) Pooling and clean-up

Dispense identical volumes of all samples to be pooled in a through. Here, 3.5 µl of each sample normalized to 5 ng/µl were used. Mix and transfer to a 1.5 ml Eppendorf tube. Use Speedvac to collect pool in 20-100µl. Clean sample pool by using 1X AMPure XP beads (see Appendix 2 of (32)) and elute in 20 µl TE buffer.

d) Verify quality and size distribution of pooled library

The expected size distribution should be around 500bp, as primers for primary PCR were designed to keep variable regions between 300 - 400 bp long and approximately 170 bp were added for Illumina Sequencing (including sequences for read1 and read2, indices and i5/i7). Verify size range and distribution by running the library on 2200 TapeStation, following the manufacturer's protocol (Fig. 3.5).

Determine library concentration by using Qubit 2.0 Fluorometric Quantification and following the manufacturer's protocol.

Submit the library to your sequencing facility or follow instructions provided by (32).

3. 3. 7 Amplicon sequencing

Sequencing can be performed on a NextSeq sequencing platform using a NextSeq Reagent Kit MidOutput (300 cycles). The library is compatible with the standard sequencing primers, final amplicon structure is shown in figure 3.4.

3. 3. 8 Sequence analysis and estimation of outcrossing rates

Read pairs were associated to amplicons by parsing for the presence of forward and reverse primer at the beginning of reads 1 and 2 respectively using cutadapt version 2.1 (33). Primer sequences were removed, including T or A nucleotides that were added for heterogeneity (see Table 3.1). Reads resulting from primers without those T or A nucleotides were cut by one additional nucleotide at the end to get identical read lengths for both primer versions. Only reads corresponding to the expected length were kept: sequencing length - length of gene specific sequence part of the primer - one nucleotide for T or A. Obtained sequences for read 1 and 2 were combined into one fragment to be treated as haplotype further on. Haplotype occurrences were counted. Haplotypes with a frequency below 1% were excluded from further analyses as they are likely to be a consequence of random sequencing errors; we expect this not to shift haplotype proportions. Haplotypes with a frequency above 1%, but only present in one single sample were also excluded as they are likely to be a consequence of an early PCR amplification error. Roughly half of the fragments were filtered out that way (see Supplemental Table 3.2). The sum of remaining fragments per sample and amplicon were then used as a baseline to calculate (1) the proportions of parental (P1 or P2) haplotypes per sample and amplicon and (2) the proportions of P1 and P2 haplotypes for the polymorphic amplicon 6_Carubv10023818m. Data analyses were done using R (https://www.r-project.org). Illustrations were done using the R/lattice package (http://lmdvr.r-forge.r-project.org).

3. 4 Results

Plants of two *Capsella* quasi-isogenic lines (qILs) differing in about 12 kb around the *CNL1* locus that underlies the loss of benzaldehyde emission in *C. rubella* (23) were grown at the field site of the Botanical Garden of the University of Potsdam. The arrangement of the plants is shown in Figure 3.1. Each block contained two sets of six plants each of both genotypes arranged on opposite sides of a central area with inbred, self-compatible *C. grandiflora* plants and near-isogenic lines differing for the *SAP* locus that affects petal size (24). The plants in this central area served as pollen donors with different background genotypes. Three such blocks were protected from birds and rodents by a bird net, while three such blocks were covered by an insect-proof net.

We designed 12 primer pairs in exons of highly conserved genes that anneal to invariant nucleotide stretches across a large number of *C. grandiflora* and *C. rubella* genotypes. Primers were chosen in exons flanking an intron to maximize the sensitivity for detecting non-maternal genotypes, as intron sequences are generally more variable than exonic sequences. As shown in Figure 3.2, 11 out of the 12 primer pairs successfully amplified and two pools of six and five primer pairs were set up to minimize the number of required PCR reactions.

Genomic DNA was extracted from approximately 300 pooled seeds from the 12 scented and the 12 non-scented qIL plants per block. A previously described qPCR-based approach (32) was used to determine the optimal template concentration for the primary PCRs with the two pools of six and five primer pairs described above. Example results for this test are shown in Figure 3.3. In most cases, the undiluted sample was used for further steps.

Barcoding indices were introduced via a second, indexing PCR, resulting in final amplification products with the structure shown in Figure 3.4. Products from the two primary PCRs (with the six- and five-primer pair pools, respectively) were combined in equimolar ratios after this indexing PCR. An example of a final library pool as determined by Tape Station electrophoresis is shown in Figure 3.5.

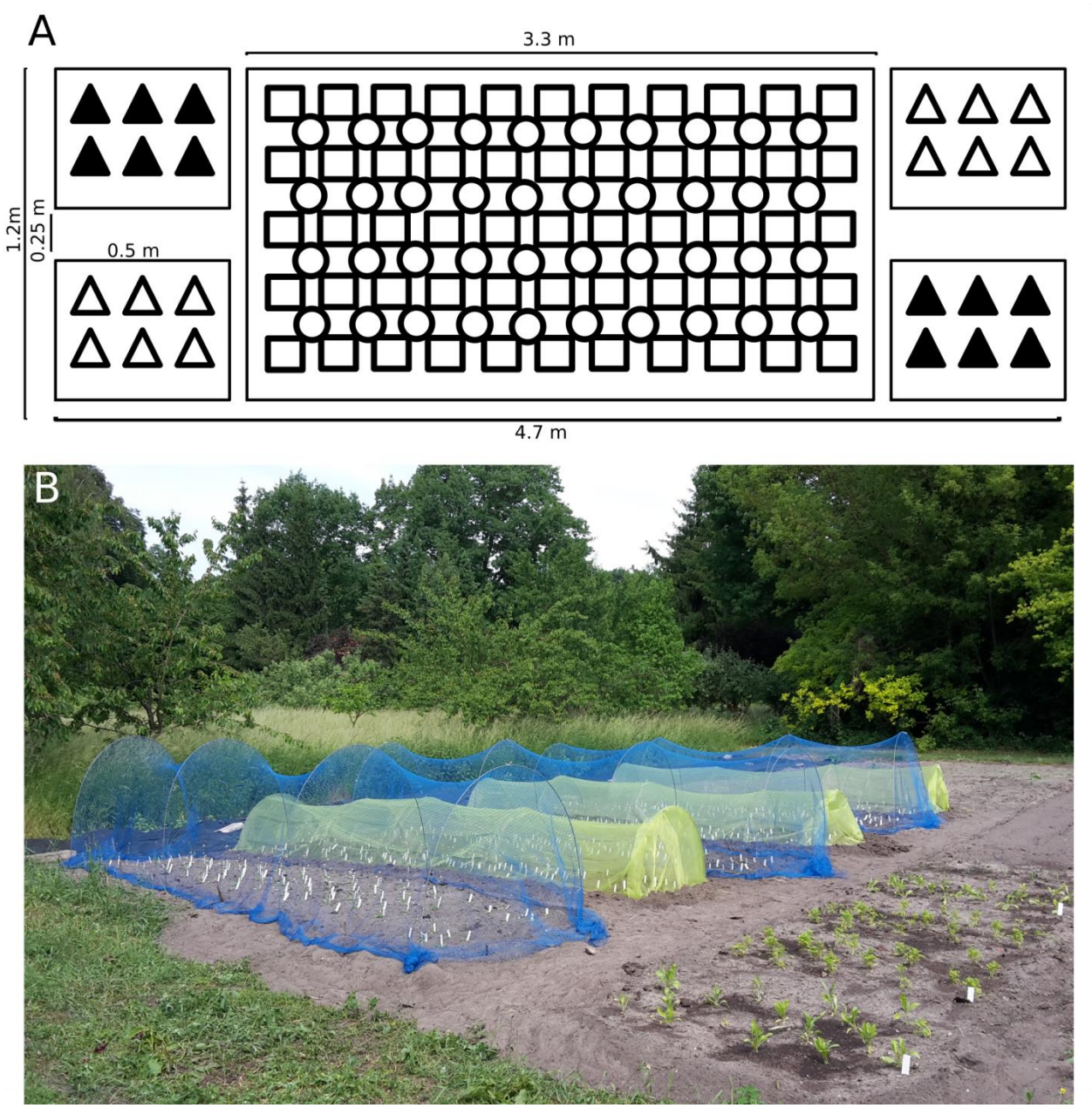

Figure 3.1: Plot set-up for common garden experiment.

(A) Detailed experimental design. Every plot consisted of five patches: two with six scented qIL plants (black triangles) and two with six unscented qIL plants (white triangles). The big area in the center contained other *Capsella* lines to provide pollen for outcrossing (circles - *Capsella grandiflora* self-compatible line, squares - NILs differing in petal size).

(B) Photo of common garden experiment from Spring 2017. Each plot was set up like described in 3.1A. In three plots plants were grown under insect-exclusion (green nets) and in the other three plots insects had access but plants were protected from other wildlife like birds, rodents and deer (blue nets).

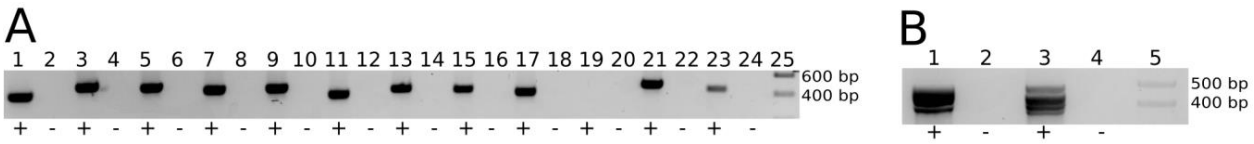

Figure 3.2: Primer-test for individual primers and pooling strategy.

(A) Primer sets 1-12 were tested with genomic DNA as template (odd-numbered lanes, '+') or in a no-template control reaction (even-numbered lanes, '-'). Primer set 10 (lanes 19 and 20) gave no PCR product and was therefore excluded in further steps.
(B) Pooled amplification of primer sets 1-5 (lane 1 and 2) and 6-12 (without 10, lane 3 and 4) with water controls.

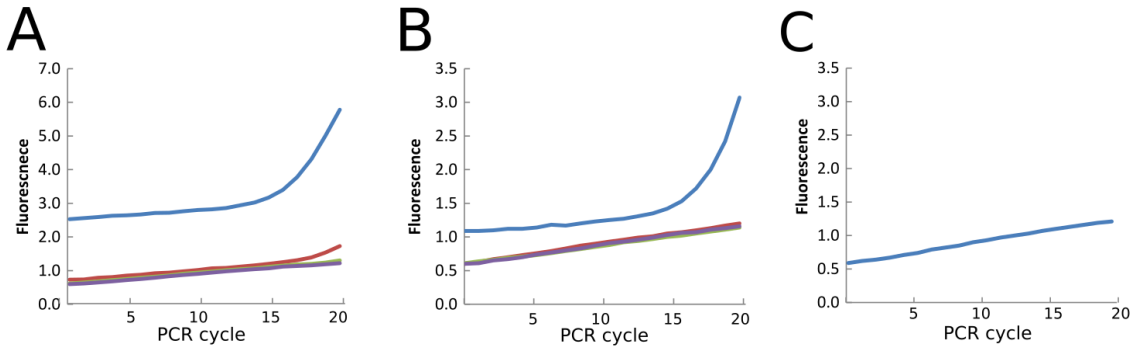

Figure 3.3: Examples for qPCR amplification traces.

For A and B, blue trace is the undiluted sample, red trace is 1:10 dilution, green is 1:100 and purple trace is 1:1000 dilution. DNA concentrations were low enough to not overamplify within 20 cycles. For more detailed information about cherry-picking using a robot please refer to the protocol provided by Gohl and colleagues (32).
(A) Undiluted sample (blue) is in mid-exponential phase and was used for further steps. 1:10 dilution is in early-exponential phase, the two lowest dilutions show no amplification.
(B) Undiluted sample (blue) is in early-to-mid-exponential phase and was used for further steps. Other dilutions did not amplify.
(C) Water blank control.

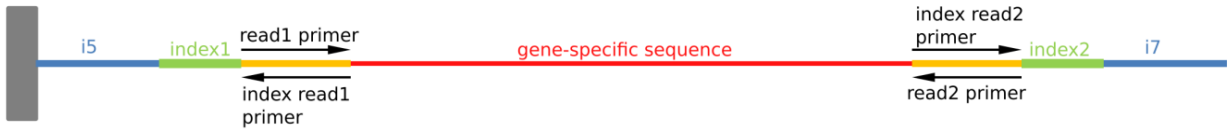

Figure 3.4: Structure of the final amplicon.

Gene-specific sequences are amplified during primary PCR (red), read1 and read 2 sequences (orange) are added to the gene-specific primers. Indices (green) and i5/i7 sequences (blue) are added during indexing PCR.

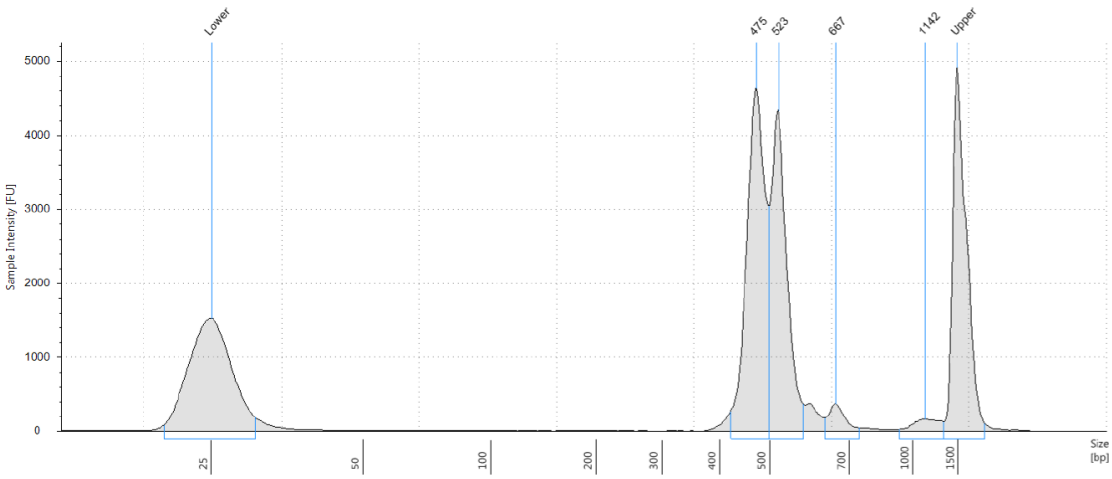

Figure 3.5: Example of a final library as analyzed with TapeStation.

The library fragments are between 400 and 600 bp long. Upper and lower standard are indicated.

Libraries were sequenced in 2x 150 bp paired-end mode, and the two reads were linked to create 300-bp fragments for analysis. Based on these, we defined major haplotypes as those present with at least 1% frequency in more than one sample. This excludes low-frequency sequencing errors, but also PCR-errors from early cycles, as these should be unique to individual samples. After such filtering, around 50% of fragments remained for analysis, except for sample 26, where only 35% fragments were retained (Supplemental Table 3.S2). We assigned these remaining fragments to either parental or non-parental haplotypes. Across all amplicons, the frequency of non-parental reads was consistently higher in the samples from plants grown under the bird-protection nets only, i.e. accessible to insects, than from those grown under insect-proof nets (Figure 3.6); in fact, in the latter no non-parental haplotypes could be detected for eight of the amplicons in five out of the six samples. In the samples from the bird nets, the frequency of non-parental haplotypes reached between 10 and 20% for eight of the amplicons, with very consistent estimates across the individual samples. For three of the amplicons, the frequency of non-parental haplotypes was below 10% in the bird-net samples, again with consistent estimates across the individual samples. We ascribe this difference between the two groups of amplicons to haplotype sharing with the other plants in the blocks that served as pollen donors for the outbred seeds, with at least some of these plants sharing the parental haplotypes with our lines at the three amplicons in question, thus rendering many outcrossing events undetectable. While the frequencies of non-parental haplotypes were higher for the samples from plants with the *C. grandiflora CNL1* haplotype than with the *C. rubella* haplotype for two of the replicates under the bird net, this was reversed in the third

replicate. Thus, overall there was no consistent difference in the estimated outcrossing rate between benzaldehyde-emitting and non-emitting plants in this one trial.

In summary, the strong and consistent difference in apparent outcrossing frequencies between samples under the two types of nets strongly supports the validity of our analysis method and indicates a surprisingly high rate of insect-mediated outcrossing in these self-compatible *Capsella* genotypes.

We also found that the two parental lines differed at one of the amplicons (number 6, locus Carubv10023818m, fig. 3.7), indicating that their genomic background was not fully isogenic. In principle, this difference could allow estimating outcrossing between the two parental lines. When considering only the two parental haplotypes, the plants with the *C. rubella* haplotype in the *CNL1* region appeared to have received more pollen with the alternative haplotype at amplicon 6 (i.e. from the plants with the *C. grandiflora* haplotype at *CNL1*) than vice versa under the bird nets. While this difference could suggest asymmetric pollen flow between the two lines, it could also be due to differential haplotype sharing with the other plants in the plots, in particular if the plants carrying the *C. grandiflora CNL1* haplotype shared their amplicon-6 haplotype with more of the other pollen-donor plants. To circumvent this issue of differential haplotype-sharing, only haplotypes found in neither of the two parental lines were counted for the analysis at amplicon 6 shown in Figure 3.6.

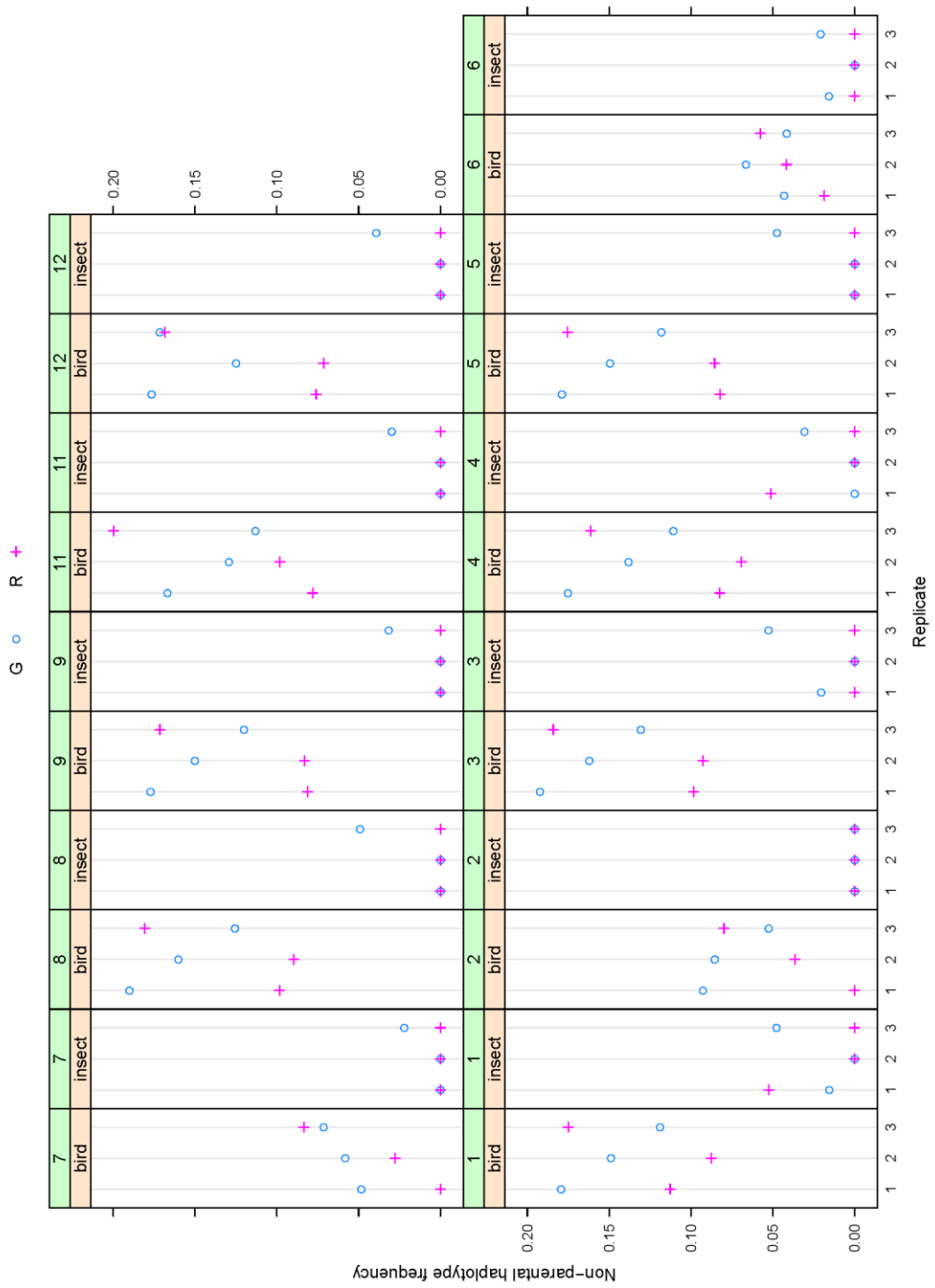

Figure 3.6: Non-parental haplotype frequencies across the amplicons.
Non-parental haplotype frequencies are plotted across the 11 amplicons. Amplicons are indicated by numbers '1' to '12', according to Table 3.1. Paired results for each of the three replicated blocks under the bird nets ('bird') and insect nets ('insect') are shown, with replicates numbers 1 to 3. 'G' and 'R' indicate samples homozygous for the *C. grandiflora* allele or the *C. rubella* allele in the *CNL1* region, respectively. For amplicon 6, only those haplotypes were counted as non-parental that were distinct from those found in either of the parental lines.

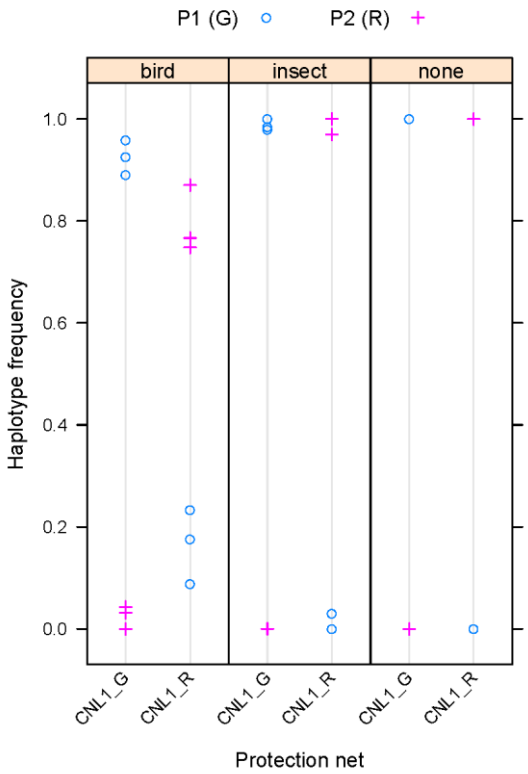

Figure 3.7: Frequency of the two alternative parental haplotypes in reads for amplicon 6 (Carubv10023818m).

Frequencies of the two alternative parental haplotypes at locus Carubv10023818m (termed 'P1 (G)' and 'P2 (R)') is plotted for the samples carrying the *C. grandiflora* allele (CNL1_G) or the *C. rubella* allele (CNL1_R) in the *CNL1* region, respectively. Samples are separated according to the protection net they were under. 'none' indicates samples from selfed parental plants grown in the absence of animal pollinators. Note that only the two parental haplotypes were considered for this analysis.

3. 5 Discussion

In this study, we have described a method for estimating outcrossing rates that lends itself to the analysis of many samples with high throughput and have validated the method using a common-garden experiment comparing the outcrossing rate between open-pollinated and insect-excluded plants. Our results clearly show that the method successfully detects insect-mediated outcrossing events and provides consistent estimates of outcrossing rates across replicated samples. While the analysis approach presented assumes that the maternal genotypes at the tested amplicons are known, the method can easily be adapted to the case when these are not, by preparing a parallel set of amplicon-sequencing libraries from genomic DNA of the mother plants to be analyzed.

The described method offers a major advantage regarding the time and effort required to estimate the outcrossing rate for many samples. For example, obtaining the estimates for the twelve samples in our study using the classical method would have required running more than 10,000 individual PCR reactions and analyzing the products by electrophoresis (assuming 100 progeny seeds were genotyped per sample). At the same time, the degree of pooling libraries derived from different samples for the sequencing run could easily be increased, enabling the analysis of more samples with very little extra effort. For example, when demanding on average a 10,000-fold coverage for each of ten amplicons per sample, hundreds of samples could be analyzed in parallel using a single NextSeq mid-output run. In principle, this would allow very fine-scale descriptions of how the outcrossing rate differs across a population to determine the effect of environmental influences or trait variation.

Compared with the single progeny-based approach these advantages concerning throughput come at a cost regarding the up-front investment in primer design and the precision of the estimates of outcrossing rates. As for primer design, one obvious technical source of error is differential amplification efficiency of different haplotypes due to mismatches in the primer-binding sites. For individual-based measurements, only very large differences in amplification efficiencies will cause an error, causing for example certain heterozygotes to be called as homozygotes for the more efficiently amplifying allele. By contrast, even slight amplification biases for different haplotypes can cause substantial error in the estimated outcrossing rates using the method described here. To circumvent this issue, some up-front sequencing of the haplotypes in the population in question may be necessary to identify highly conserved primer-binding sites. Following the strategy taken here, a cost-effective method for doing so would be to sequence PCR amplification products from highly conserved genes containing one or more introns from many pooled individuals in the population. Analysis of these sequences should allow

identifying invariant primer-binding sites flanking suitably polymorphic regions or adapting the primer design by incorporating polymorphic bases into the primers, if no fully invariant regions can be found.

As for the precision of the estimates, the present method will necessarily underestimate the true outcrossing rates, and it will do so more strongly than the individual-based method in most cases. Concerning each single amplicon, the true outcrossing rate will be underestimated by the combined frequency of the two maternal alleles in the pollen population, as any outcrossing event involving a pollen carrying one of the maternal alleles will be undetectable when considering a single amplicon. The individual-based method is better able to deal with this complication than the pool-based approach. This is because for the individual-based method a single marker with a non-maternal allele or haplotype is enough to classify an individual as resulting from outbreeding. By contrast, this information is necessarily lost in our pool-based approach; in a hypothetical example, if there were ten such outbred individuals, each with the diagnostic non-maternal haplotype in a different amplicon, these would all be detectable in the individual-based approach, but would only be counted as a single outbreeding event in the pool-based approach. Such a scenario of maternal haplotype-sharing at many of the amplicons will result for example from bi-parental inbreeding (13, 14).

The above bias means that the estimate closest to the true outcrossing rate will be obtained from the amplicon for which the combined frequency of the maternal haplotypes in the pollen population is lowest. In this regard, longer sequence reads appear preferable, as they will allow detecting a larger number of different haplotypes in the population, thus reducing the described effect. A further implication of the above is that outbreeding rates are strictly only comparable between individuals carrying the same maternal haplotypes at a given amplicon, as these will be affected by the above bias in the same manner. This is exemplified by amplicon 6 in our study, for which the two parental lines were polymorphic. Here, merely counting non-maternal haplotypes would have given very different estimates for the two parental lines for this amplicon. Thus, in light of these issues, if the aim is to characterize the reproductive system in a population with unknown genetic structure in great detail, considering also aspects like bi-parental inbreeding, the individual-based method remains the method of choice. By contrast, the pool-based method described here should be preferable, if the main aim is to obtain relative outcrossing rates from a large number of individuals and in situations where the above biases are likely to have a weak effect.

In summary, we have described a cost-effective method for the high-throughput estimation of outcrossing rates in plants. We see its major application in studies to correlate outcrossing rates with

environmental, morphological or physiological parameters across a large number of individuals, especially if the genetic structure of the population in question is known. This should enable connecting genetic differences that affect pollinator-attraction traits with effects on pollinator behavior in ecologically realistic settings.

3. 6 References

1. Hahn MW. Molecular Population Genetics. Oxford: Oxford University Press; 2018.
2. Hartfield M, Bataillon T, Glemin S. The Evolutionary Interplay between Adaptation and Self-Fertilization. Trends in Genetics. 2017 Jun;33(6):420-31.
3. Hedrick PW, Garcia-Dorado A. Understanding Inbreeding Depression, Purging, and Genetic Rescue. Trends Ecol Evol. 2016 Dec;31(12):940-52.
4. Nordborg M. Linkage disequilibrium, gene trees and selfing: an ancestral recombination graph with partial self-fertilization. Genetics. 2000 Feb;154(2):923-9.
5. Nordborg M, Donnelly P. The coalescent process with selfing. Genetics. 1997 Jul;146(3):1185-95.
6. Igic B, Busch JW. Is self-fertilization an evolutionary dead end? New Phytologist. 2013 Apr;198(2):386-97.
7. Wright SI, Kalisz S, Slotte T. Evolutionary consequences of self-fertilization in plants. Proc Biol Sci. 2013 Jun 7;280(1760):20130133.
8. Jarne P, Auld JR. Animals mix it up too: The distribution of self-fertilization among hermaphroditic animals. Evolution. 2006 Sep;60(9):1816-24.
9. Barrett SC. The evolution of plant sexual diversity. Nat Rev Genet. 2002 Apr;3(4):274-84.
10. Goodwillie C, Kalisz S, Eckert CG. The evolutionary enigma of mixed mating systems in plants: Occurrence, theoretical explanations, and empirical evidence. Annual Review of Ecology Evolution and Systematics. 2005;36:47-79.
11. Arista M, Berjano R, Viruel J, Ortiz MA, Talavera M, Ortiz PL. Uncertain pollination environment promotes the evolution of a stable mixed reproductive system in the self-incompatible Hypochaeris salzmanniana (Asteraceae). Annals of Botany. 2017 Sep;120(3):447-56.
12. Yin G, Barrett SCH, Luo YB, Bai WN. Seasonal variation in the mating system of a selfing annual with large floral displays. Annals of Botany. 2016 Mar;117(3):391-400.
13. Ritland K. Extensions of models for the estimation of mating systems using n independent loci. Heredity. 2002 Apr;88:221-8.
14. Koelling VA, Monnahan PJ, Kelly JK. A Bayesian method for the joint estimation of outcrossing rate and inbreeding depression. Heredity. 2012 Dec;109(6):393-400.
15. Raguso RA. Flowers as sensory billboards: progress towards an integrated understanding of floral advertisement. Current Opinion in Plant Biology. 2004 Aug;7(4):434-40.
16. Sicard A, Lenhard M. The selfing syndrome: a model for studying the genetic and evolutionary basis of morphological adaptation in plants. Ann Bot. 2011 Jun;107(9):1433-43.
17. Sicard A, Stacey N, Hermann K, Dessoly J, Neuffer B, Baurle I, et al. Genetics, evolution, and adaptive significance of the selfing syndrome in the genus Capsella. Plant Cell. 2011 Sep;23(9):3156-71.
18. Slotte T, Hazzouri KM, Stern D, Andolfatto P, Wright SI. Genetic Architecture and Adaptive Significance of the Selfing Syndrome in Capsella. Evolution. 2012 May;66(5):1360-74.
19. Bachmann JA, Tedder A, Laenen B, Fracassetti M, Désamoré A, Lafon-Placette C, et al. Genetic basis and timing of a major mating system shift in Capsella. biorxiv. 2018:425389.

20. Hurka H, Friesen N, German DA, Franzke A, Neuffer B. 'Missing link' species Capsella orientalis and Capsella thracica elucidate evolution of model plant genus Capsella (Brassicaceae). Mol Ecol. 2012 Mar;21(5):1223-38.
21. Koenig D, Hagmann J, Li R, Bemm F, Slotte T, Nueffer B, et al. Long-term balancing selection drives evolution of immunity genes in *Capsella.* eLife. 2019. 8.
22. Fujikura U, Jing R, Hanada A, Takebayashi Y, Sakakibara H, Yamaguchi S, et al. Variation in Splicing Efficiency Underlies Morphological Evolution in Capsella. Dev Cell. 2018 Jan 22;44(2):192-203 e5.
23. Sas C, Muller F, Kappel C, Kent TV, Wright SI, Hilker M, et al. Repeated Inactivation of the First Committed Enzyme Underlies the Loss of Benzaldehyde Emission after the Selfing Transition in Capsella. Curr Biol. 2016 Dec 19;26(24):3313-9.
24. Sicard A, Kappel C, Lee YW, Wozniak NJ, Marona C, Stinchcombe JR, et al. Standing genetic variation in a tissue-specific enhancer underlies selfing-syndrome evolution in Capsella. Proc Natl Acad Sci U S A. 2016 Nov 29;113(48):13911-6.
25. Doyle JJ. Isolation of Plant DNA from Fresh Tissue. Focus. 1990;12:13-5.
26. Dabney J, Meyer M. Length and GC-biases during sequencing library amplification: A comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. Biotechniques. 2012 Feb;52(2):87-+.
27. Krehenwinkel H, Wolf M, Lim JY, Rominger AJ, Simison WB, Gillespie RG. Estimating and mitigating amplification bias in qualitative and quantitative arthropod metabarcoding. Sci Rep-Uk. 2017 Dec 15;7.
28. Stadhouders R, Pas SD, Anber J, Voermans J, Mes THM, Schutten M. The Effect of Primer-Template Mismatches on the Detection and Quantification of Nucleic Acids Using the 5 ' Nuclease Assay. J Mol Diagn. 2010 Jan;12(1):109-17.
29. Polz MF, Cavanaugh CM. Bias in template-to-product ratios in multitemplate PCR. Appl Environ Microb. 1998 Oct;64(10):3724-30.
30. Duarte JM, Wall PK, Edger PP, Landherr LL, Ma H, Pires JC, et al. Identification of shared single copy nuclear genes in Arabidopsis, Populus, Vitis and Oryza and their phylogenetic utility across various taxonomic levels. Bmc Evolutionary Biology. 2010 Feb 24;10.
31. Fadrosh DW, Ma B, Gajer P, Sengamalay N, Ott S, Brotman RM, et al. An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. Microbiome. 2014 Feb 24;2.
32. Gohl DM, MacLean A, Hauge A, Becker A, Walek D, Beckman KB. An optimized protocol for high-throughput amplicon-based microbiome profiling. Protocol Exchange. 2016.
33. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnetjournal 2011;17:10.

3. 7 Declarations

Ethics approval and consent to participate

not applicable

Consent for publication

not applicable

Availability of data and material

The datasets generated and analyzed during the current study are available in the NCBI SRA repository under accession number PRJNA529581 (https://www.ncbi.nlm.nih.gov/sra/PRJNA529581).

Reviewer link:

https://dataview.ncbi.nlm.nih.gov/object/PRJNA529581?reviewer=eo3d1e198ofe1d0l1pk6r4tbpi

Competing interests

The authors declare that they have no competing interests.

Funding

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, and by grant SI1967/1-1 from the Deutsche Forschungsgemeinschaft to AS. The funders had no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Authors' contributions

AS, CK, ML conceived the study, FJ, NW, CK acquired and analyzed data, ML drafted the manuscript with input from all authors. All authors approved the final version of the manuscript.

Acknowledgements

We thank the staff of the Botanical Garden at the University of Potsdam for their help in preparing the field site and in plant care. We are grateful to Frauke Garbsch for optimizing the DNA extraction from seeds.

3. 8 Supplemental Information

Table 3.S1: Example indexing scheme.

	1	2	3	4	5	6	7	8	9	10	11	12
A	F1+R13	F1+R14	F1+R15	F1+R16	F1+R17							
B	F2+R13	F2+R14	F2+R15	F2+R16	F2+R17							
C	F3+R13	F3+R14	F3+R15	F3+R16	F3+R17							
D	F4+R13	F4+R14	F4+R15	F4+R16	F4+R17							
E	F5+R13	F5+R14	F5+R15	F5+R16	F5+R17							
F	F6+R13	F6+R14	F6+R15	F6+R16	F6+R17							
G	F7+R13	F7+R14	F7+R15	F7+R16								
H	F8+R13	F8+R14	F8+R15	F8+R16								

Table 3.S2: Read statistics for the amplicons (see attached Excel sheet on page 94).

The total number of fragments (i.e. paired reads) per sample is shown, as is the number and percentage of fragments mapped to the PCR amplicons, and the number and percentage of fragments used for haplotype calling. The latter excluded low-frequency fragments, as these most likely represent PCR or sequencing errors.

Number of fragments (i.e. paired reads) per sample

sample	total number of fragments
25	3108942
26	2071195
27	2566227
28	3558269
29	5067095
30	6786569
31	4790186
32	3191330
33	2943333
34	1850513
35	2980436
36	4459515
P1	3985730
P2	6022658

Number of fragments associated to amplicons

sample	total number of fragments	% of fragments per sample	1_Carubv10013869m	2_Carubv10023806m	3_Carubv10018138m	4_Carubv10001640m	5_Carubv10001924m	6_Carubv10023818m	7_Carubv10005658m	8_Carubv10006001m	9_Carubv10006101m	11_Carubv10011729m	12_Carubv10014733m
25	3048519	98.06%	538671	417977	342976	509375	445348	144121	24763	144247	325409	44538	111094
26	2025790	97.81%	309081	236326	225459	325747	288954	128744	32838	129331	239206	38996	71108
27	2510736	97.84%	310896	282931	197088	329383	264459	188598	30570	216029	470469	62711	157602
28	3487567	98.01%	598512	300905	319268	524366	482584	273668	52930	197409	515793	62566	159566
29	4970471	98.09%	811245	575982	470778	748509	639944	303824	55732	292230	728959	108392	234876
30	6652833	98.03%	1040239	684179	601282	929515	833852	503684	79790	420602	1057666	125100	376924
31	4678604	97.67%	554166	494002	387966	633755	570663	351887	65089	387604	822933	146509	264030
32	3122161	97.83%	313524	149859	166568	289514	262450	419945	137225	311163	690942	189048	191923
33	2881099	97.89%	346233	301732	248549	407923	369580	204750	35576	229095	502312	73873	161476
34	1812753	97.96%	238578	141906	153059	249392	243490	143821	44888	144627	295227	76483	81282
35	2918138	97.91%	414435	355612	286348	448940	400010	175442	38285	197839	392783	84531	123913
36	4362230	97.82%	686804	532271	485329	711036	631033	243885	52355	258008	507175	94713	159621
P1	3814975	95.72%	423960	326746	252204	438451	345528	380182	78318	397346	700752	191504	279984
P2	5884875	97.71%	772272	691941	589009	848424	684643	459683	146902	441954	683878	229592	336577

Number of fragments used for haplotyping (excluding low frequency fragments: technical and/or sequencing errors)

sample	total number of haplotyping	% of fragments used for ha	1_Carubv10013869m	2_Carubv10023806m	3_Carubv10018138m	4_Carubv10001640m	5_Carubv10001924m	6_Carubv10023818m	7_Carubv10005658m	8_Carubv10006001m	9_Carubv10006101m	11_Carubv10011729m	12_Carubv10014733m
25	1713005	56.19%	351953	173828	171617	282333	266823	94526	14040	70577	212284	20132	54892
26	707782	34.94%	132545	53645	75541	115919	101206	53185	12547	36392	93573	9122	24107
27	1445799	57.58%	204359	118710	99778	185081	163940	127881	17695	107456	314508	25748	80643
28	1943548	55.73%	381173	119114	151517	282639	283605	174450	29054	91092	331250	22560	77094
29	2684936	54.02%	500203	227668	219924	395314	368233	196639	31041	135157	455168	45957	109632
30	3844823	57.79%	682974	294736	303307	522848	513707	335324	46487	205875	701912	49375	188278
31	2244459	47.97%	296989	181357	158564	293279	288311	227571	35847	154228	449044	51031	108238
32	1431009	45.83%	168408	45545	66806	128330	119055	236527	66684	112669	352562	58578	75845
33	1429777	49.63%	201761	110929	105873	194379	193003	128070	19594	95043	283263	28089	69773
34	917939	50.64%	144490	49634	66592	123104	128660	87975	24161	60781	173029	24784	34729
35	1467388	50.29%	246821	132469	122474	220287	216846	112273	20596	84892	229478	30018	51234
36	1968378	45.12%	371753	171438	185724	311549	299829	147490	26646	96917	267112	26704	63216
P1	2139442	56.08%	282054	130153	123384	239366	206425	249764	44398	192200	456589	77879	137230
P2	3362499	57.14%	529463	294969	296662	479702	424959	306048	86425	223595	460848	88573	171255

4. Identifying a candidate gene underlying PAQTL1

4. 1 Introduction

As described before, Sicard et al. (2011) identified six QTLs underlying the reduction in petal area between the selfing species *C. rubella* and the outbreeder *C. grandiflora*, mirroring the polygenic basis for this trait. Causal genes underlying two of these QTL have been identified so far (Sicard et al, 2016; Fujikura et al., 2018), together explaining about 21.5% of variation between species. In a more recent comparative QTL study between *C. grandiflora*, *C. rubella* and *C. orientalis* it has been shown that eight QTLs underlie the size differences between *C. grandiflora* and *C. orientalis*, four of them overlapping with the QTLs identified in *C. grandiflora*/*C. rubella* (Wozniak, 2019). PAQTL1 explains 9.27% of variation between *C. grandiflora* and *C. rubella* and its confidence interval is between 12.9 -22.0 cM (Sicard et al., 2011). A QTL in that region was also detected in the *C. grandiflora x C. orientalis* F2 population, maybe providing another example for a 'hotspot' gene (Stern and Orgogozo, 2008). Once identified, the gene underlying this QTL could give more insights into the molecular basis of morphological evolution as this remains poorly understood, especially for polygenic traits as organ shape and size.

After the transition to selfing, the reduction in *Capsella* petal area is highly specific to floral organs as demonstrated by Sicard et al., 2011. This is quite surprising as almost all known regulators of shoot-organ growth have a pleiotropic activity in both leaves and flowers. The arising question is how natural evolution brings organ-specific changes with only a universal tool-kit. Different hypotheses have been developed to provide answers to this question: either assuming mutations in a gene upstream, affecting the regulation of a pleiotropic gene in a given organ or mutations affecting the activity of this pleiotropic gene in an organ-specific manner (*cis*-regulatory changes) (Stern and Orgogozo, 2008; Gaunt and Paul, 2012). For morphological variation, the evolution of gene-regulatory sequences via *cis*-regulatory changes is considered to be the primary driver (Wray, 2007; Carroll, 2008), but only few examples exist where the downstream effects of such variation have been characterized in detail. One recent publication provides such an example for fruit shape evolution in *Capsella* (Dong et al., 2019). Here, the authors demonstrated that the variation in fruit morphology is directed by a regulatory diversification in one gene, leading to effects on hormone homeostasis and resulting in a morphological novelty.

During previous work it was tried to fine-map and identify the gene underlying PAQTL1, but this was not accomplished so far. Some results were contradictory, probably due to mis-genotyping or only small sample numbers. The region has been quite certainly confined to about 500 kb as shown in Fig. 4.1. To continue mapping this QTL to a gene, previous findings on mapping data should be confirmed by genotyping and phenotyping segregating lines generated by Runchun Jing. These lines were offspring from the RIL population generated by Sicard et al., 2011, for the original QTL study. Exploiting the availability of these plant lines it should be possible to quickly narrow down the position of the QTL to further characterize its underlying gene. After fine-mapping, functional studies will give more insights into the genetic basis of convergent evolution of morphology in the genus *Capsella*.

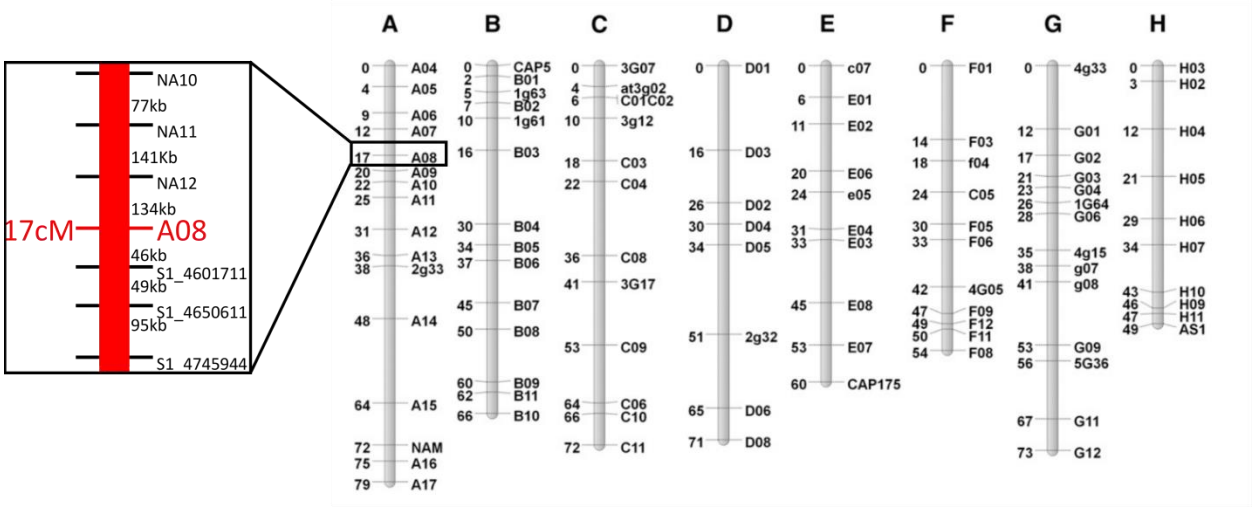

Figure 4.1: Genetic map based on the *C. grandiflora* x *C. rubella* RIL population. Modified from Sicard et al., 2011. Individual linkage groups are shown by grey bars, names of the used markers are on the right and the genetic distance in cM is shown on the left. Enlarged and in red is the region to which the QTL was mapped in previous work.

4. 2 Material and Methods

4. 2. 1 Plant Material and growth conditions

For this study lines generated by Runchun Jing were used. These were offspring from RIL lines used for QTL mapping as described by Sicard et al. (2011). Plants were grown under a long-day photoperiod (16h light/ 8h dark) wit temperature set to 21°C during day and 16°C during night. They were kept with 70% humidity and a light level of 150 µmol m^{-2} s^{-1}.

4. 2. 2 Molecular cloning

CrNUB and CgNUB were cloned from genomic DNA using OFJ35 and OFJ36 for the 3.7 kb promoter region and OFJ33 and OFJ37 for the 1212 bp gene and 573 bp terminator region. The PCR-amplified fragments were subcloned into a modified version of pBluescript II KS (StrataGen) using SLiCE fusion cloning (Zhang et al., 2012). The fragments were cloned into the AscI site of the plant transformation vector pBarMAP, a derivate of pGPTVBAR (Becker et al., 1992). The resulting vectors were named pFJ19 (containing CrNUB) and pFJ20 (containing CgNUB). The PCR-amplified fragments were also fused using SLiCE to generate promoter swap constructs in an analog way, resulting in pFJ21 (pCgNUB::CrNUB) and pFJ22 (pCrNUB::CgNUB).

4. 2. 3 Morphological measurements and statistical analysis

Petal area was measured using ImageJ from the digitalized images of dissected petals. Flowers for measurements were between the 15^{th} and 20^{th} flower of the main inflorescence and were taken when fully opened.

The statistical significance of differences in petal size was assessed by two-tailed Student´s *t* tests. The null hypothesis was rejected at $P < 0.05$. The significance level at which the null hypothesis could be rejected is indicated on each figure by the number of asterisks, with $P < 0.05$ *, $P < 0.01$ ** and $P < 0.001$ ***.

4. 2. 4 Oligonucleotides

Table 4.1: Oligonucleotides for mapping PAQTL1.

Name	Sequence	Description
ORJ19	AGTAGACTCTAAGTAAGCTAG	marker NA10 F
ORJ20	TAGGGACGGGACGTAGCACTT	marker NA10 R
ORJ21	GAAGATTCTGAGAGAAGACG	marker NA11 F
ORJ22	GGTCTAAATCTAATGAAGGTAC	marker NA11 R
ORJ27	AGACACGATATCAGGATTGC	marker NA12 F
ORJ28	CGAAGGTGCACAGATGCTGCA	marker NA12 R
A08 F	AGTTTTGTTTGCAGACGTTGA	marker A08 F
A08 R	ATTCACTATCATTGACCAACA	marker A08 R
ORJ57	TGAGCTAGATCCCAAATTCCAGA	marker scaffold_1.4601711 F
ORJ58	AGGCTGAGACTCTTTACAAGCC	marker scaffold_1.4601711 R
ORJ59	TGGCTTATTAAGTCGTGAGCCT	marker scaffold_1.4650611 F
ORJ60	TGTGGAAAAGGTTAACATCAGATG	marker scaffold_1.4650611 R
ORJ63	AATTGCAAGCGTGATCAAGTCC	marker scaffold_1.4745944 F
ORJ64	TGACTTTGCATGATTGGTCACA	marker scaffold_1.4745944 R
ORJ69	ACCTGCGTATACTTTCTCCGTC	marker scaffold_1.5199954 F
ORJ70	TTTCGGTTGGATTCGGGTTAGG	marker scaffold_1.5199954 R
A09 F	ATGTCTCCGGAAGCTTACGTTCTGT	marker A09 F
A09 R	TCAAGCGAAACCAACATTCCTTGG	marker A09 R
A10 F	TGACAAGCCACCGACTT	marker A10 F
A10 R	CAGACTGAGCCCTGGAGGA	marker A10 R
A11 F	AAAGCAAAAGGAAGAAGAGGAAA	marker A11 F
A11 R	GAAACGCACCGAAACACC	marker A11 R
A12 F	GACAACAACAGATCATTCCTAA	marker A12 F
A12 R	ATCTGAATCTTCTTCTTCTCG	marker A12 R

Table 4.2: Oligonucleotides used for cloning.

Name	Sequence	Description
OFJ33	ATGTTCTCTCTCTCTCTCCTTGC	F *NUB* gene
OFJ35	GACAACAAAGCTATGGAATCATG	F promoter *NUB*
OFJ36	GCAAGGAGAGAGAGAGAGAGAAC	R promoter *NUB*
OFJ37	GCGTGCAGACAAATAATTCAACTG	R *NUB* + terminator region
OFJ44	CAGTTGAATTATTTGTCTGCACGCGGCGCGCCATTAA TTAATGGATCAG	F to linearize AS77 for SLiCE
OFJ45	CCATAGCTTTGTTGTCTGCAGGTTTCGGCGCGCCTTC TAGCCAATTC	R to linearize AS77 for SLiCE

4. 3 Results

4. 3. 1 Confirming the position of PAQTL1

As previous results were not clear it was decided to take a conservative approach and first confirm the position of the QTL by genotyping and phenotyping different nearly-isogenic lines, segregating in a bigger region of interest. Especially the right side of the QTL was of interest as there were no clear results before.

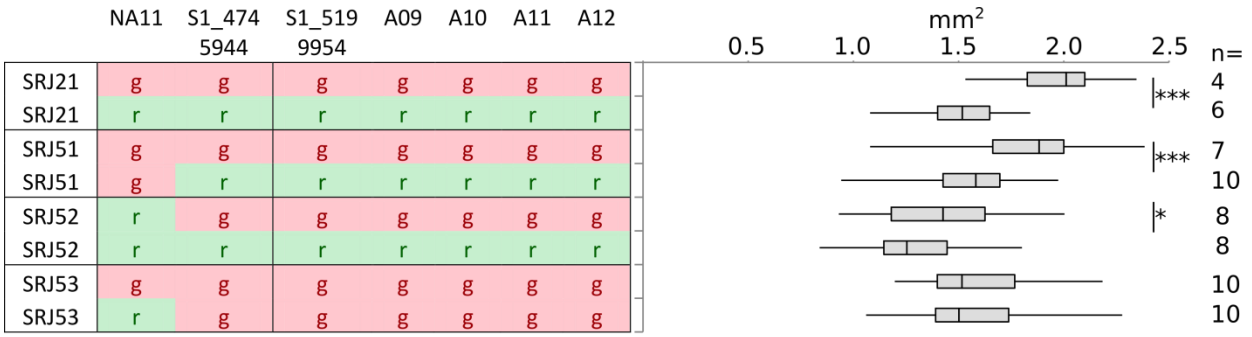

Figure 4.2: Confirming the region for PAQTL1. Four different segregating NILs were tested, for each plant 6-8 petals were measured. Homozygotes for both alleles were phenotyped, number of tested individuals per genotype is indicated. Box plots show the mean value (black line), the 25 and 75 percentiles (left and right bounds of the box); the whiskers show the location of maximum and minimum. Asterisks indicate statistically significant differences as determined by Student´s t test at $P < 0.05$ (*), $P < 0.01$ (**) and $P < 0.001$ (***).

Differences in petal area between the two genotypes showed that the QTL segregated in three out of four tested lines as shown in Figure 4.2. These data suggested that the QTL is to the right of marker NA11 and the petal area of plants homozygous for the *rubella*- allele in that region was significantly smaller than for plants carrying the *grandiflora*-allele.

Based on these findings, more lines were selected for genotyping and phenotyping to confirm previous mapping results for PAQTL1. In parallel, line SRJ51 was selected to find more recombinants in the segregating region for further fine-mapping, in case assumptions based on previous results could not be confirmed.

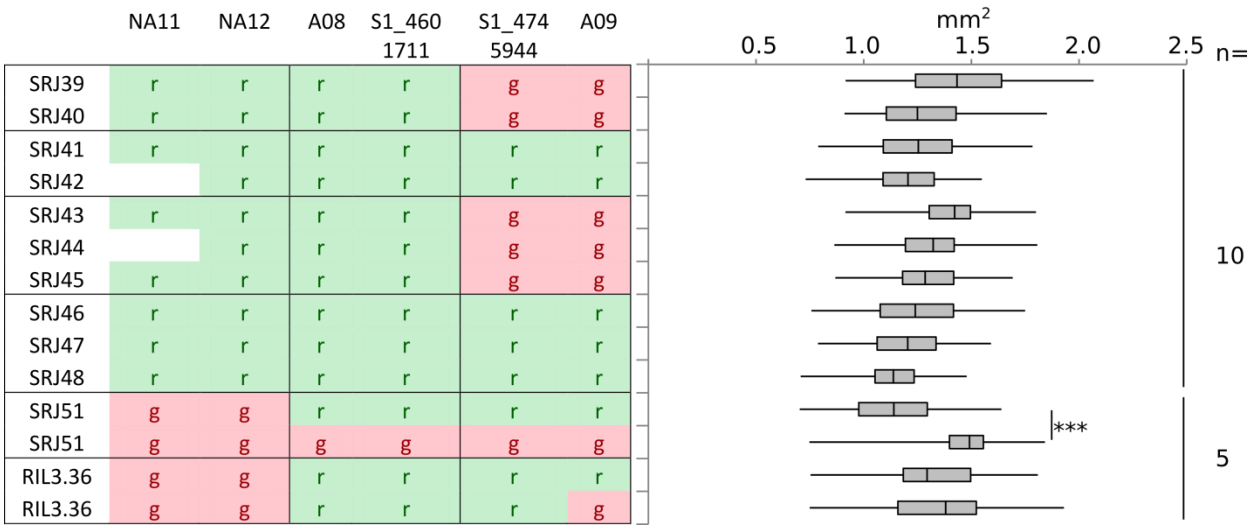

Figure 4.3: Narrowing down the position of PAQTL1. SRJ39-48 were selected based on previous results, 10 plants per line were phenotyped. SRJ51 shows clear segregation of the phenotype and RIL3.36 is a recombinant line found among SRJ51 plants which were genotyped to use for progeny-testing, 5 plants per line were phenotyped for each genotype.

To confirm that the QTL is upstream of marker S1_4745944, lines SRJ39-48 were selected, as no seeds for segregating plants from the original recombinant lines that this conclusion is based on were available. Lines RJ39-42 are offspring of the same recombinant line, either homozygous for the *rubella*- or *grandiflora*-allele from marker S1_4745944 on. Lines SRJ43-48 are offspring from a different plant with the same genotype in that region. There was no clear segregation of phenotypes among the two different groups and no significant difference between e.g. SRJ40 and 41 (Figure 4.3), so it was concluded that the QTL should be on the left of marker S1_4745944.

Additionally, line SRJ51 was genotyped with more markers and phenotyped in parallel with recombinant line 3.36, detected during screening for recombinants among seedlings of line SRJ51 (Figure 4.3). According to this data, the QTL should be to the right of marker NA12, leaving approximately 324 kb for further fine-mapping.

4. 3. 2 Identifying a candidate gene

As assumptions based on previous results have been proven to be right, a fast approach towards finding a candidate gene underlying PAQTL1 was taken. Following Ushio Fujikura´s data, the QTL was mapped to

a 100 kb interval between markers S1_4650611 and S1_4745944 and possible candidates in this region were evaluated. Table 4.3 contains all annotations for this region.

Table 4.3: Annotated genes or transcripts from *C. rubella* in the 100 kb interval.

Gene/transcript	Description
Carub10012013	Serine/threonine-protein kinase RIO // subfamily not named
Carub10011925	genomic DNA, Chromosome 3, TAC Clone: K13N2-related
Carub10012068	endosomal targeting BRO1-like domain-containing protein
Carub10008615	Serine/threonine-protein phosphatase 2A regulatory subunit A (PPP2R1)
Carub10011847	26S proteasome regulatory subunit, ATPase 3, interacting protein (PSMC3IP)
Carub10009293	Regulator of VPS4 activity in the MVB pathway protein
Carub10010335	no functional annotations
Carub10010302	expressed protein
Carub10010468	no functional annotations
Carub10010898	**zinc finger protein JAGGED-related**
Carub10008584	pentatricopeptide repeat-containing protein
Carub10009662	glyceraldehyde -3-phosphate dehydrogenase GAPC2, cytosolic
Carub10008945	serine/threonine-protein phosphatase 2A 55kDa regulatory subunit
Carub10012231	Protein of unknown function
Carub10012504	Protein of unknown function
Carub10011704	Ethanolaminephosphotransferase
Carub10012325	cytosolic sulfotranferase 1-related
Carub10009313	Tri helix transcription factor GT-1-related
Carub10012255	F-box domain (F-box) // F-box associated (FBA_1)
Carub10009445	Protein of unknown function
Carub10011549	Protein of unknown function
Carub10011246	Protein of unknown function
Carub10011634	Protein of unknown function
Carub10009278	F-box domain (F-box) // Leucine Rich Repeat (LRR_2) // FBD (FBD)
Carub10010947	ceramide synthetase (CERS)
Carub10011839	BZip protein (AtBZip48)-related
Carub10011177	zinc transporter 7
Carub10009802	3-methyladenine glycosylase 1-related
Carub10011693	Phytosulfokine precursor protein (PSK)
Carub10011982	no functional annotations
Carub10011972	no functional annotations
Carub10011195	alpha/beta-hydrolase-like protein
Carub10008310	PPR repeat (PPR) // PPR repeat (PPR_1) // PPR repeat family (PPR_2)
Carub10009802	3-methyladenine glycosylase 1-related
Carub10010856	root meristem growth factor 1-related
Carub10008594	phosphatidylinositol 3-kinase-related protein kinase
Carub10009590	18S pre-ribosomal assembly protein gar2-related protein

Carub10012096	F21F23.12 protein
Carub10011738	F21F23.12 protein
Carub10009460	F-box associated ubiquitination effector family protein
Carub10010036	6-phosphogluconolactonase 1-related
Carub10010185	no functional annotations
Carub10010435	peptidyl-prolyl isomerase E (cyclophilin E) (PPIE)
Carub10008872	CYTOCHROME P450 78A5-related
Carub10011148	nuclear transport factor 2 and RNA recognition motif domain-containing protein
Carub10009552	NINJA-family protein AFP1-related
Carub10008414	ribonuclease P subunit P38 // subfamily not named
Carub10008872	CYTOCHROME P450 78A5-related
Carub10009181	UPF0420 protein C16ORF58
Carub10008293	PPR repeat (PPR) // PPR repeat family (PPR_2)
Carub10009613	YqaJ-like viral recombinase domain (YqaJ)
Carub10012551	carbohydrate-binding X8 domain-containing protein
Carub10008535	inactive purple acid phosphatase 2-related
Carub10009998	alpha/beta-hydrolases superfamily protein
Carub10008622	methyltransferase PMT4-related
Carub10009863	protein KTI12 (KTI12)
Carub10009267	ELM2 domain-containing protein
Carub10012302	synaptosomal associated protein
Carub10010071	F7A19.1 protein-related
Carub10010793	no functional annotations
Carub10008191	F16A14.15-related
Carub10009699	leucine-rich repeat (LRR) family protein-related
Carub10010521	no functional annotations

The most promising candidate seemed to be Carub10010898, encoding a zinc-finger protein related to *JAGGED*. Zinc-finger proteins are among the most abundant proteins in eukaryotic genomes and they show diverse functions including DNA recognition, transcriptional activation and many more (Laity et al., 2011). *JAGGED* plays a role in shaping lateral organs in *Arabidopsis thaliana*, including petals (Dinneny et al. 2004) and it has been shown to play a role in the petal size reduction in *C. rubella* (Wozniak, 2019). It is a classical C_2H_2 zinc finger, expressed in the growing regions of lateral organs (Dinneny et al., 2004). Many of this type of zinc fingers are transcription factors that function by recognition of specific DNA sequences (Laity et al., 2011), making Carub10010898 an interesting candidate to cause gene expression changes, matching hypotheses for speciation gene candidates.

BLASTing the sequence gave a significant alignment to *NUBBIN* in *Arabidopsis thaliana* and predicted zinc-finger proteins from other species. *NUBBIN* and *JAGGED* define stamen and carpel shape in

Arabidopsis and act redundantly in these tissues (Dinneny et al., 2006). *JAGGED* is detected in all floral organs and all layers of lateral organs but *jag* mutants show strong effects only in sepal and petal development, suggesting a lack of redundant factors there. *NUB* expression is restricted to the interior adaxial site of leaves, stamens and carpels and *nub* mutants show no significant phenotype (Dinneny et al., 2006). Double mutants of *jag* and *nub* show severely enhanced phenotypes with defects in leaf, stamen and carpel development compared to *jag* single mutants, underlining their redundant effects.

To test if the candidate gene *NUB* is underlying PAQTL1, *CrNUB* and *CgNUB* were cloned into a plant transformation vector. Both constructs are ready to be transformed to NILs containing the *rubella*-allele in that region for complementation assays. If *NUB* underlies PAQTL1, one would expect bigger petals in plants carrying the *CgNUB* allele compared to plants carrying *CrNUB*. Additionally, promoter swap constructs were prepared to test whether any functional difference between the two different *NUB* alleles is due to differences in the promoter region or in the coding sequence in transformed plants.

4. 4 Discussion and Outlook

During this work it was possible to partially confirm previous results on PAQTL1 and narrow down the underlying region to 324 kb. Based on previously made assumptions a promising candidate gene in that region was found, encoding for a homologue of *NUBBIN*, known from *A. thaliana* (Dinneny et al., 2006). The insights from *Arabidopsis* research seem to make Carub10010898 a strong candidate, but at the same time it does not give a perfect fit, as *A. thaliana NUB* is exclusively expressed in leaves, stamens, carpels and only briefly in petal primordia (Dinneny et al., 2006). Furthermore, *NUB* and *JAG* act redundantly in stamens and carpels and a study by Wozniak (2019) showed no interaction between PAQTL1 and PAQTL2, which encodes for *JAGGED* in *Capsella* (Wozniak, 2019). Also, *nub* mutants show no phenotype in *A. thaliana* (Dinneny et al., 2006) making it less likely to play an important role during petal development.

Characterizing and identifying genes underlying QTLs for selfing syndrome traits such as petal area leads to understanding basic molecular mechanisms for organ growth. This way, new regulators and genetic networks can be described and more light is shed onto the question how plants regulate organ specific growth with a universal toolbox. Changes in tissue-specific regulators or tissue-specific modificators of general regulators could be responsible. Through comparative studies among different species as carried out for *C. grandiflora*, *C. orientalis* and *C. rubella* it can be determined if convergent evolution of flower morphology followed the same or different genetic paths (Wozniak, 2019) and therefore allows testing to what extend genetic evolution is predictable. PAQTL1 was identified in both selfing species, possibly arguing for convergent morphology following at least partially the same genetic paths. It remains to be investigated what the main drivers of phenotypic evolution are.

By comparing phenotypic similarities and the evolutionary history of selfing-syndrome traits in selfers derived from two independent transitions to selfing it might be possible to determine if traits such as a reduction in petal size were by-products of the transition to selfing and the lack of selective force applied by pollinators or if they are results of resource re-allocation and positive selection. Considering that *C. rubella* is quite a young species compared to *C. orientalis* (Guo et al., 2009; Hurka et al., 2012) there might be positive selection for smaller petals as both selfers show a similar level of reduction, despite different divergence times.

During this study it was possible to generate tools to confirm or reject *NUBBIN* as a candidate underlying PAQTL1 by preparing constructs for plant transformation and by crossing recombinant plants to

generate a line which only segregates for a 100 kb region containing *NUB*. At the same time lines were selected for further recombinant screens and fine-mapping to either show clearly that *NUB* underlies PAQTL1 or to be able to narrow it down to a different candidate. Additionally lines containing different alleles of *JAG* and *NUB* were selected and crossed to see if there is any significant interaction resulting in size differences in floral organs in *Capsella*.

5. Concluding remarks

Self-fertilization arose multiple times independently among angiosperms and selfing species across different angiosperm genera show similar morphological and functional differences compared to their outbreeding sister taxa. The most prominent differences include a dramatic reduction of floral size and scent emission and are well described under the term selfing syndrome (Sicard and Lenhard, 2011). Because of its repeated occurrence, the transition to selfing is one of the main examples of convergent evolution in plants. To explain this wide-spread phenomenon of convergent evolution, it has been hypothesized that convergence can result from evolutionary re-targeting of the same genes and evidence for this has been found in many taxa (Stern and Orgogozo, 2013). One possible explanation is that these convergent changes result from mutations in the same genetic targets with minimal pleiotropic effect but also maximal phenotypic output and therefore allowing rapid adaptation to new conditions. If this is the case, molecular changes underlying adaption might be predictable, as there is only a limited number of possible targets fulfilling these requirements (Stern, 2013).

Often shifts to high selfing rates are accompanied by a strong reduction of genetic diversity, as it has been shown for *Capsella rubella* (Guo et al., 2009; Brandvain et al., 2013) and therefore it seems likely that morphological changes arising after a transition to selfing are based on *de novo* mutations rather than from standing genetic variation. The genus *Capsella* provides a great model to study the convergent evolution of the selfing syndrome as selfing arose at least twice independently, once in *Capsella orientalis* and once in *Capsella rubella*, allowing comparative studies between the ancestor-like *C. grandiflora* and the two derived selfers.

While the genetic basis of convergent evolution from morphological traits after the transitions to selfing in this genus was studied recently, less is known about the evolution of floral scent as a physiological trait. Therefore, taking advantage of previous studies about selfing syndrome evolution and scent loss in *Capsella rubella* (Sicard et al., 2011; Sas et al., 2016) it was possible to shed more light on the evolution of repeated scent loss in selfing *Capsella* species in this study.

First it was demonstrated that the loss of benzaldehyde emission in the selfing species *C. rubella* arose twice independently from missense mutations to *CNL1*.This conclusion was based on a combination of population genetic analysis, *in vivo* and *in vitro* experiments. This enzyme catalyzes the first committed step in the biosynthesis of benzenoids like BAld from cinnamic acid and is compartmentalized in the

peroxisome (Figure 2.S1). Its inactivation likely does not lead to an accumulation of toxic intermediates and induces no detectable pleiotropic collateral defects as the inactivation of other genes in the pathway might. This makes *CNL1* a good target to modify scent emission in plants and indeed it was inactivated twice independently in *C. rubella* and once in *P. exserta* (Amrad et al., 2016). In both species the inactivation resulted from loss-of-function mutations leading to a total lack of BAld emission. Such a severe phenotypic change could allow rapid adaptation. Traits increasing selfing efficiency are thought to evolve rapidly and therefore major mutations could be expected as it is the case for the loss of BAld emission. Attraction of insects can be disadvantageous for selfers as it could lead to pollen wastage, non-conspecific pollen presence and damage to flowers. Benzaldehyde is one of the most ubiquitous volatiles (Knudsen et al., 2006; Schiestl, 2010); and elicits different behavioral responses in pollinators and herbivores (Koschier et al., 2000; Schiestl, 2010); therefore, its elimination likely leads to a maximal phenotypic output regarding plant-insect interaction. All of this taken together leads to the conclusion that *CNL1* gives a great example of a hotspot gene for the convergent evolution of scent loss.

The population genetics analysis for *CNL1* also revealed two different genetic clusters within the *C. rubella* populations based on the two independently inactivated *CNL1* haplotypes identified here, arguing for two different geographic routes during the colonization of the Mediterranean. This partially confirms other studies where two different genetic clusters within the *C. rubella* populations have been identified (Guo et al., 2009; Koenig et al., 2019), giving more insight into the genetic structure of selfing populations and the spread of traits associated with the selfing syndrome. This study suggests that benzaldehyde emission has been lost twice independently due to *de novo* mutations and was not captured from standing genetic variation.

The question whether floral fragrance loss results from relaxed selective forces on traits for pollinator attraction or through other processes remains. Selective force towards scent loss might have been exerted by herbivory as some herbivores are attracted by benzaldehyde and other scent compounds (Koschier et al., 2000), letting plants without benzaldehyde emission benefit from less herbivorous attacks. With the deprivation of selection towards fragranced plants by pollinators, this might have advanced the evolution of scent loss. A different possibility is the loss of benzaldehyde emission through genetic drift instead. Further studies are necessary to exclude one of these options.

If benzaldehyde plays a role in pollinator attraction it should attract pollinators that bring non-self pollen and therefore have an impact on the genetic diversity of the offspring. To test for the role of an individual trait a new experimental procedure was established. Available tools such as quasi-isogenic

lines, only differing in the region underlying the loss of BAld emission, were combined with cutting edge sequencing technology accessible on-site to develop a method of estimating outcrossing rates and therefore testing the adaptive value of a trait under otherwise natural conditions. This newly developed method allows the analysis of many samples with a high throughput and detects successfully insect-mediated outcrossing events, allowing future experiments to determine how the outcrossing rate differs across a population under varying traits or to determine the effect of environmental influences. The experience gained during this first trial can be used to plan following experiments under natural conditions to better understand the effects of adaptive traits on population genetics and maybe allow untangling the interplay between demography and selection in self-fertilizing species in the future.

This study shed more light on the molecular basis of selfing syndrome evolution by using a comprehensive approach consisting of population genetics analysis, *in vivo* and *in vitro* experiments to determine the molecular basis of scent loss. The mutations leading to an abolished CNL1 activity in *C. rubella* have been identified, linking mutations on DNA level directly to a phenotypic outcome and showing that CNL1 has been re-targeted during the evolution of floral scent, even in different species (Sas et al., 2016; Amrad et al, 2016), making *CNL1* a good candidate hotspot gene. Wondering whether *CNL1* could be responsible for scent loss in the other derived selfer, *C. orientalis*, the CNL1 activity was investigated using different approaches, showing that CNL1 in *C. orientalis* is active and therefore arguing for a different mechanism being responsible for the loss or strong reduction of benzaldehyde emission in this species. This was unexpected, as both selfers probably derived from similar ancestral populations and were likely to evolve selfing in an analogous manner for reproductive assurance. However, CNL1 was clearly functional in *C. orientalis* and no major QTL underlying scent loss could be detected, arguing for a polygenic basis responsible for the lack or strong reduction of benzaldehyde emission in *C. orientalis*. These results demonstrate that parallel evolution can have strongly differing genetic bases and further studies are needed to answer the question why the genetic routes to similar phenotypic outcomes differed so much in the two selfers. To identify the mutations underlying scent loss in *C. orientalis*, further experiments are necessary. QTL mapping with recombinant-inbred lines (RILs) might be more successful to identify weaker mutations underlying scent loss than using pooled phenotypes of F3 families.

Internal pool measurements and a phenylalanine feeding experiment were conducted to narrow down the possible reasons for the loss of benzaldehyde emission in the standard *C. orientalis* accession. Plants are still able to produce and emit BAld, if the concentration of the precursor phenylalanine is high

enough, suggesting that the pathways leading to BAld emission remain functional. Possibly phenylalanine availability for scent emission is restricted in the selfer but further experiments are necessary to prove this, for example with labeled amino acids to compare how phenylalanine is metabolized in the different *Capsella* species. Phenylalanine is an important building block for proteins and a precursor for thousands of metabolites in plants (Vogt, 2010; Yoo et al., 2013). Maybe in *C. orientalis* this resource is reallocated to other processes than scent emission as selection for scent emission probably relaxed after the transition to selfing. It can be concluded that the convergent evolution of reduction in BAld emission is due to a different genetic basis than in *C. rubella,* suggesting that there are different ways for this trait to evolve independently and bringing similar outcome. It seems likely that genetic constraints are not responsible for the convergent evolution of scent loss/reduction in *C. orientalis*. If this was the case, one would assume similar underlying mechanisms in both selfing species. This suggests that the convergent evolution of scent loss is likely to be driven by ecological factors rather than by genetic constraints and evolutionary re-targeting in this species. Similar results have been obtained for the loss of (*E*)-β-ocimene in *C. rubella* and *C. orientalis* (Wozniak, 2019).

The mutations to *CNL1* in *C. rubella* give an example for coding changes being responsible for the evolution of scent loss, not *cis*-regulatory changes as it has been shown for petal size reduction (Sicard et al., 2016; Fujikura et al., 2018). Comparing these different molecular bases argues for the idea that physiological evolution might involve more coding changes than morphological evolution and that morphological evolution often results from *cis*-regulatory changes (Stern and Orgogozo, 2008).

To further elucidate the genetic basis of morphological evolution of selfing syndrome traits a comparative QTL study was conducted. The experiments by Sicard et al. (2011) and Wozniak (2019) showed that many loci affect the evolution of petal size and several were overlapping between the two selfing species. To find out whether the same genes underlie these QTL a shared QTL on chromosome 1 was chosen to be analyzed further. The position of PAQTL1, explaining about 9% of variation in petal area between *C. grandiflora* and *C. rubella,* was confirmed and tools for further fine-mapping of the underlying gene were prepared. Additionally a candidate gene, *NUBBIN*, was identified and cloned for further investigations. If PAQTL1 truly maps to *NUB* remains to be proven. As this QTL is shared between *C. rubella* and *C. orientalis* (Wozniak, 2019) it also needs to be shown if the same gene is responsible for petal size reduction in *C. orientalis*. If this is the case it would argue for shared molecular mechanisms influencing this trait in both selfers. Maybe there are genetic constraints for petal size evolution and

specific regulatory nodes in the network controlling flower size are more suitable for adaptation than others, making evolutionary re-targeting highly likely (Wozniak, 2019).

The genus *Capsella* proved to be an excellent model system to study the convergent evolution of the selfing syndrome. In this work an interdisciplinary approach was used, exploiting methods from molecular biology, biochemistry, genetics, bioinformatics and ecology, to better characterize scent loss as a selfing-syndrome trait. It was possible to identify its molecular basis in *C. rubella*, exclude different possibilities for the strong reduction/loss in *C. orientalis* and to establish an advanced method for further studies on characterizing the adaptive value of individual traits such as floral scent. Furthermore, a promising candidate gene was found, possibly explaining convergent evolution in petal area reduction in the two selfing species, *C. orientalis* and *C. rubella*, and the tools to further characterize it were prepared.

6. Additional Material and Methods

6.1 Chemicals

Chemicals were purchased from Carl Roth (Karlsruhe, Germany), Invitrogen (brand of Thermo Fisher Scientific), Promega (Mannheim, Germany), Serva (Heidelberg, Germany), Sigma-Aldrich (now Merck, Darmstadt, Germany) and Quiagen (Hilden, Germany).

Enzymes were purchased from Bioline (London, UK), Invitrogen (brand of Thermo Fisher Scientific), New England Biolabs (Frankfurt am Main, Germany), Roche Diagnostics (Grenzach-Wyhlen, Germany), Takara Bio USA Inc. (Mountain View, USA) and Thermo Fischer Scientific (Waltham, USA).

Materials for sequencing were ordered from Illumina (San Diego, USA).

Oligonucleotides were purchased from Sigma-Aldrich (now Merck, Darmstadt, Germany).

Antibiotics and herbicides for selective media were used in following concentrations:

Antibiotic/Herbicide	Dissolvent	Working concentration
Ampicillin	ddH_2O	100 µg/ml
Kanamycin	ddH_2O	50 µg/ml
Gentamycin	ddH_2O	25 µg/ml
Rifampicin	DMSO	80 µg/ml
Phosphinothricin (PPT)	ddH_2O	15 µg/ml

6.2 Disposable equipment

Materials for scent measurements were purchased from Chromatografie Zubehört Trott (Kriftel, Germany) and Brechbühler AG (Schlieren, Switzerland).

Lab consumables were purchased from Greiner Bio-One (Kremsmünster, Austria), Kisker Biotech (Steinfurt, Germany), Sarstedt (Nümbrecht, Germany), Starlab (Hamburg, Germany) and VWR (Radnor, USA). Materials for plant cultivation were ordered from Fitz Kausek (Mittenwalde, Germany).

6.3 Microorganisms

E. coli strains XL1- Blue and DH5α were used as vectors for cloning. For plant transformation, *A. tumefaciens* strain GV3101 was used.

6.4 Methods

6.4.1 Plant-related methods

Growing Capsella plants

Prior to sowing, seeds were sterilized using chlorine vapor. For this, seeds were distributed in 1.5 ml Eppendorf tubes and placed in a desiccator. A beaker containing a fresh solution of 50 ml sodium hypochlorite (12%) and 1.5 ml hydrochloric acid (37%) was added. The desiccator was closed for four hours. Once the sterilization process was finished, the lid was removed to allow evaporation of chlorine gas before proceeding to sow the seeds on agar plates with 0.5 x Murashige and Skoog medium, supplemented with giberellic acid (0.02 mM). After sowing, the plates were kept for four days in 4°C and then placed in a climate growth chamber with similar conditions as in the greenhouse until the first pair of real leaves grew (10-14 days). Afterwards plates with seedlings were transferred to 4°C for vernalization (depending on genotype for up to six weeks).

Phenylalanine-feeding experiment

L-Phenylalanine (Carl Roth) was dissolved in water for a 150 mMol solution. 20 ml solution (or water for controls) was filled into falcon tubes, three to five inflorescences were added and tubes were incubated for one and a half hours under the plant´s growing conditions. Subsequently tubes were transferred to scent collection apparatus and volatiles were collected for three hours.

6.4.2 DNA-related methods

DNA extraction for genotyping (High-throughput)

One leaflet or cotyledon per *Capsella* seedling was harvested into one well of a 96-well plate. 50 µl of genome extraction buffer were added (5 % SDS, 95 % TNE buffer) and the plate was covered with a silicone mat for PCR reactions. The plate was incubated at 95°C for 10 minutes, vortexed, incubated for 10 minutes at -80°C followed by a final incubation for 10 minutes at 95°C. 50 µl isopropanol were added, plate was sealed and gently vortexed. A centrifugation step at 4200 rpm for 30 minutes followed. The supernatant was discarded and the pellets washed with 100 µl 70% ethanol. After removing the ethanol the pellet was dried for 15 minutes. Following this, the pellet was resolved in 50 µl water and 2 µl were used for PCR reactions.

DNA extraction for high quality and purity

10-500 mg plant material was harvested into an Eppendorf tube and ground on liquid nitrogen to disrupt cells. 300 µl DNA extraction buffer (100 mM Tris pH 8.0, 50 mM EDTA pH 8.1, 1 M NaCl, 1.25 % SDS (w/v), added 0.1 mg/ml RNAse A before use) were added and samples were centrifuged at 4000 rpm for 1 minute. 150 µl 5 M potassium acetate were added and the samples were centrifuged for 15 minutes at 4000 rpm. During this step, fresh tubes containing 200 µl isopropanol were prepared and after centrifugation, 300 µl supernatant were carefully transferred and mixed with the isopropanol. A centrifugation step at 4000 rpm for 15 minutes followed. The supernatant was discarded and the DNA pellets were washed with 600 µl 70 % ethanol. After centrifugation for 15 minutes at 4000 rpm the ethanol was removed carefully and the pellets were dried for 10 minutes. The DNA pellet was dissolved in 100-150 µl water by mixing for 20 – 30 minutes at room temperature.

Cloning

Unless stated otherwise, DNA fragments were cloned into vectors using the SLiCE fusion cloning described by Zhang et al., 2012.

Sequencing

Sequencing was done by using a Ready 2 Run service of LGC Genomics (Berlin, Germany). The preparation of PCR-products and plasmids followed the manufacturer´s protocol.

6.4.3 RNA-related methods

Total RNA isolation with TRIzol reagent

For RNA isolation 4-5 top parts of inflorescences were collected into a 1.5 ml Eppendorf tube and ground with liquid nitrogen. TRIzol reagent (Invitrogen) was added and the manufacturer´s protocol was followed.

DNAse digest and cDNA synthesis

To degrade DNA in RNA samples they were digested with TurboDNAse from Invitrogen according to the manufacturer´s protocol. For cDNA synthesis by RT-PCR the Superscript III Reverse Transcriptase (Invitrogen) was used as described by the manufacturer. To only generate cDNA from mRNA, oligo$(dT)_{17}$ primer was used for each reaction.

Appendix: Oligonucleotide and vector lists

Table A1: Oligonucleotides used for qRT-PCR

Name	Sequence	Description
q75f	CGTTGCTGTCGTCTCGCTTC	Expression analysis for CNL1 F
q75r	ATCTTTGGCTCAGCATGGCG	Expression analysis for CNL1 R
TUBf	TGCACCAAGCAGCATGAAGA	Expression analysis for housekeeping gene F
TUBr	ATACTCGGCCTTGGAGATCCAC	Expression analysis for housekeeping gene R

Table A2: Vectors generated and used for this work.

Vector	Description	bacteria selection	plant resistance
pAS77	pBlueMLAPUCAP	Amp	
pBAR	pBARMAP	Amp	
pML595	pBARMAP with 35S-promoter	Kan	BASTA
pFJ1	pAS77 with Cr1504CNL1^{R453S} and endogenous promoter	Amp	
pFJ2	pBAR with Cr1504CNL1^{R453S} (for plant transformation)	Kan	BASTA
pFJ3	pML595, Ω, CgCNL1	Kan	BASTA
pFJ4	pML595, Ω, CrCNL1	Kan	BASTA
pFJ5	pAS77with Cr1504CNL1^{R453S}, Ω, cDNA	Amp	
pJET Co1983CNL1	for enzyme assay	Amp	
pJET Cr1504CNL1^{R453S}	for enzyme assay	Amp	
pJET CR80TR1	for enzyme assay	Amp	
pFJ6	ML595, Ω, Cr1504CNL1^{R453S}	Kan	BASTA
pFJ9	AS77 with CR80TR1and Ω	Amp	
pFJ10	AS77 with Co1983CNL1 and Ω	Amp	
pFJ11	ML595, Ω, CR80TR1	Kan	BASTA
pFJ12	ML595, Ω, Co1983CNL1	Kan	BASTA
pFJ13	pAS77 with Co1983CNL1 and endogenous promoter	Amp	
pFJ14	pBAR with Co1983CNL1 (for plant transformation)	Kan	BASTA
pJet CgCNL1^{S453R}	for enzyme assay	Amp	
pJet CR80TR1^{F543L}	for enzyme assay	Amp	
pJet Cg926CNL1^{L543F}	for enzyme assay	Amp	
pCS1	CgRIL205-20_CNL1 in pBAR	Kan	BASTA
pJET Cg926CNL1	For enzyme assay	Amp	
pJET Cr1504CNL1	For enzyme assay	Amp	
pFJ15	pAS77 with CrNUB (endogenous promoter)	Amp	
pFJ16	pAS77 with CgNUB (endogenous promoter)	Amp	
pFJ17	pAS77 pCg::CrNUB	Amp	
pFJ18	pAS77 pCr::CgNUB	Amp	

pFJ19	pBAR with CrNUB	Kan	BASTA
pFJ20	pBAR with CgNUB	Kan	BASTA
pFJ21	pBAR pCgNUB::CrNUB	Kan	BASTA
pFJ22	pBAR pCrNUB::CgNUB	Kan	BASTA

Acknowledgements

I would like to express my gratitude to my supervisor Michael Lenhard for his great support during my PhD study, ranging from giving excellent scientific advice to hands-on work in the lab.

Besides my supervisor, I would like to thank Alexander Vainstein and Florian Schiestl for taking the time to review my thesis.

Many thanks also to my collaborators, Alexander Vainstein and his group for great discussions and experimental support; Monika Hilker and Jona Höfflin for analyzing my floral scent samples and answering all my questions; Natalia Dudareva, Joe Lynch and Funmilayo Adebesin for their help with experiments, analysis and insightful discussions.

I am very thankful for all my fellow labmates and colleagues for creating a good working atmosphere, especially Melanie Teltow who kept the lab running. I would like to thank Natalia Wozniak, my co-conspirator working on scent emission, for the most helpful discussions. Thanks to Adrien Sicard and Christian Kappel for providing a lot of help and expertise in the *Capsella* field. I would like to thank Doreen Mäker for taking care of my plants in the greenhouse.

Many thanks to Frauke Garbsch, Luzie Bartsch and Florentine Ballhaus for their help with lab and harvesting work.

Last but certainly not least, I would like to thank my family, friends and my team for their unconditional support and understanding during the last three years and for finding ways to lift me up.

References

Abd El-Mawla AMA, Beerhues L. **2002**. Benzoic acid biosynthesis in cell cultures of Hypericum androsaemum. *Planta* **214**: 727–733.

Adebesin F, Pierman B, Wetzstein HY, Porter JA, Lynch JH, Boutry M, Ray S, Yanagisawa M, Alam I, Schuurink RC, *et al.* 2017. Emission of volatile organic compounds from petunia flowers is facilitated by an ABC transporter. *Science* **356**: 1386–1388.

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*.

Bachmann JA, Tedder A, Laenen B, Fracassetti M, Désamoré A, Lafon-Placette C, Steige KA, Callot C, Marande W, Neuffer B, *et al.* 2018. Genetic basis and timing of a major mating system shift in Capsella. *bioRxiv*: 425389.

Baker HG. **1967**. Support for Baker´s law - as a rule. *Evolution* **21**: 853–856.

Barrett SCH. **2002**. The evolution of plant sexual diversity. *Nature Reviews Genetics* **3**: 274–284.

Barrett SCH. **2003**. Mating strategies in flowering plants: The outcrossing-selfing paradigm and beyond. *Philosophical Transactions of the Royal Society B: Biological Sciences*. **358**: 991–1004.

Barrett SCH, Arunkumar R, Wright SI. **2014**. The demography and population genomics of evolutionary transitions to self-fertilization in plants. *Philosophical Transactions of the Royal Society B: Biological Sciences* **369**.

Becker D, Kemper E, Schell J, Masterson R. **1992**. New plant binary vectors with selectable markers located proximal to the left T-DNA border. *Plant Molecular Biology* **20**: 1195-1197.

Brandvain Y, Slotte T, Hazzouri KM, Wright SI, Coop G. **2013**. Genomic Identification of Founding Haplotypes Reveals the History of the Selfing Species Capsella rubella. *PLoS Genet* **9**: 1003754.

Brunet J. **1992**. Sex Allocation in hermaphroditic plants. *Trends in Ecology and Evolution* **7**: 79–84.

Bussell JD, Reichelt M, Wiszniewski AAG, Gershenzon J, Smith SM. **2014**. Peroxisomal ATP-Binding Cassette Transporter COMATOSE and the Multifunctional Protein ABNORMAL INFLORESCENCE MERISTEM Are Required for the Production of Benzoylated Metabolites in Arabidopsis Seeds. *Plant Physiology* **164**: 48-54.

Byers KJRP, Bradshaw HD, Riffell JA. **2014a**. Three floral volatiles contribute to differential pollinator attraction in monkeyflowers (*Mimulus*). *Journal of Experimental Biology* **217**: 614–623.

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

Carroll SB. **2008**. Evo-Devo and an Expanding Evolutionary Synthesis: A Genetic Theory of Morphological Evolution. *Cell* **134**: 25–36.

Charlesworth D, Morgan MT, Charlesworth B. **1990**. Inbreeding Depression, Genetic Load, and the Evolution of Outcrossing Rates in a Multilocus System with No Linkage. *Evolution* **44**: 1469.

Charlesworth D, Willis JH. **2009**. The genetics of inbreeding depression. *Nature Reviews Genetics* **10**: 783–796.

Cna'ani A, Muhlemann JK, Ravid J, Masci T, Klempien A, Nguyen TT, Dudareva N, Pichersky E, Vainstein A. 2015. Petunia x hybrida floral scent production is negatively affected by high-temperature growth conditions. *Plant Cell Environ* **38**(7): 1333-1346.

Darwin C. **1876**. The Effects of Cross and Self-Fertilisation in the Vegetable Kingdom. *John Murray, London.*

Davis CC, Endress PK, Baum DA. **2008**. The evolution of floral gigantism. *Current Opinion in Plant Biology* **11**: 49–57.

Dinneny JR, Yadegari R, Fischer RL, Yanofsky M, Weigel D. 2004. The role of JAGGED in shaping lateral organs. *Development* **131**: 1101-1110.

Dinneny JR, Weigel D, Yanofsky MF. **2006**. NUBBIN and JAGGED define stamen and carpel shape in Arabidopsis. *Development* **133**: 2285.

Dong Y, Jantzen F, Stacey N, Langowski L, Moubayidin L, Simura J, Ljung K, Ostergaard L. **2019**. Regulatory Diversification of *INDEHISCENT* in the *Capsella* Genus Directs Variation in Fruit Morphology. *Current Biology* **29**: 1038-1046.

Dötterl S, Vereecken NJ. **2010**. The chemical ecology and evolution of bee–flower interactions: a review and perspectivesThe present review is one in the special series of reviews on animal–plant interactions. *Canadian Journal of Zoology* **88**: 668–697.

Doubleday LAD, Raguso RA, Eckert CG. **2013**. Dramatic vestigialization of floral fragrance across a transition from outcrossing to selfing in Abronia umbellata (Nyctaginaceae). *American Journal of Botany* **100**: 2280–2292.

Douglas GM, Gos G, Steige KA, Salcedo A, Holm K, Josephs EB, Arunkumar R, Ågren 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* 112: 2806-2811.

Dudareva N, Pichersky E. **2000**. Biochemical and Molecular Genetic Aspects of Floral Scents. *Plant Physiology* **122**: 627-633.

Farré-Armengol G, Filella I, Llusià J, Peñuelas J. **2015**. Pollination mode determines floral scent. *Biochemical Systematics and Ecology* **61**: 44–53.

Fishman L, Kelly AJ, Willis JH. **2002**. Minor quantitative trait loci underlie floral traits associated with mating system divergence in Mimulus. *Evolution* **56**: 2138–2155.

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* **106**: 5241–5245.

Fujii S, Kubo KI, Takayama S. **2016**. Non-self- and self-recognition models in plant self-incompatibility. *Nature Plants* **2**.

Fujikura U, Jing R, Hanada A, Takebayashi Y, Sakakibara H, Yamaguchi S, Kappel C, Lenhard M. **2018**. Variation in Splicing Efficiency Underlies Morphological Evolution in *Capsella*. *Developmental Cell* **44**: 192–203.e5.

Gaunt S, Paul Y-L. **2012**. Changes in *Cis*-regulatory Elements during Morphological Evolution. *Biology* **1**: 557–574.

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

Glémin S, Ronfort J. **2013**. Adaptation and maladaptation in selfing and outcrossing species: New mutations versus standing variation. *Evolution* **67**: 225–240.

Gohl D, Gohl DM, MacLean A, Hauge A, Becker A, Walek D, Beckman KB. **2016**. An optimized protocol for high-throughput amplicon-based microbiome profiling. *Protocol Exchange*. doi:10.1038/protex.2016.030.

Goldberg EE, Kohn JR, Lande R, Robertson KA, Smith SA, Igić B. **2010**. Species Selection Maintains Self-Incompatibility. *Science* **330**: 493–495.

Goodwillie C, Ritland C, Ritland K. **2006**. The Genetic Basis of Floral Traits Associated With Mating System Evolution in *Leptosiphon* (Polemoniaceae): an Analysis of Quantitative Trait Loci. *Evolution* **60**: 491.

Goodwillie C, Sargent RD, Eckert CG, Elle E, Geber MA, Johnston MO, Kalisz S, Moeller DA, Ree RH, Vallejo-Marin M, *et al.* 2010. Correlated evolution of mating system and floral display traits in flowering plants and its implications for the distribution of mating system variation. *New Phytologist* **185**: 311–321.

Grossenbacher D, Briscoe Runquist R, Goldberg EE, Brandvain Y. **2015**. Geographic range size is predicted by plant mating system. *Ecology Letters* **18**: 706–713.

Guo Y-L, 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* **106**: 5246–5251.

Hartfield M, Bataillon T, Glémin S, Hartfield M. **2017**. The Evolutionary Interplay between Adaptation and Self-Fertilization. *Trends in Genetics* **33**.

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.

Kaminaga, Y., Schnepp, J., Peel, G., Kish, C. M., Ben-Nissan, G., Weiss, D., … Dudareva, N. (2006). Plant phenylacetaldehyde synthase is a bifunctional homotetrameric enzyme that catalyzes phenylalanine decarboxylation and oxidation. *The Journal of Biological Chemistry*, **281**(33): 23357–23366.

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.

Klahre U, Gurba A, Hermann K, Saxenhofer M, Bossolini E, Guerin PM, Kuhlemeier C. **2011**. Pollinator Choice in *Petunia* Depends on Two Major Genetic Loci for Floral Scent Production. *Current Biology* **21**: 730–739.

Klein AM, Vaissière BE, Cane JH, Steffan-Dewenter I, Cunningham SA, Kremen C, Tscharntke T. **2006**. Importance of pollinators in changing landscapes for world crops. *Proceedings of the Royal Society B: Biological Sciences* **274**: 303–313.

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. *The Plant Cell* **24**: 2015–2030.

Knudsen, J.; Tollsten, L.; Gunnar Bergstrom L. **1993**. Floral Scents - A Checklist Of Volatile Compounds Isolated By Head-Space Techniques. *Phytochemistry* **33**: 253–280.

Knudsen JT, Eriksson R, Gershenzon J, Ståhl B. **2006**. Diversity and distribution of floral scent. *The Botanical Review* **72**: 1–120.

Koenig D, Hagmann J, Li R, Bemm F, Slotte T, Nueffer B, Wright SI, Weigel D. **2019**. Long-term balancing selection drives evolution of immunity genes in *Capsella*. *eLife* **8**.

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.

Laity JH, Lee BM, Wright PE. **2001**. Zinc finger proteins: New insights into structural and functional diversity. *Current Opinion in Structural Biology* **11**: 39–46.

Lande R, Schemske DW. **1985**. The Evolution of Self-Fertilization and Inbreeding Depression in Plants. I. Genetic Models. *Evolution* **39**: 41.

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

Nunes CEP, Gerlach G, Bandeira KDO, Gobbo-Neto L, Pansarin ER, Sazima M. **2017**. Two orchids, one scent? Floral volatiles of Catasetum cernuum and Gongora bufonia suggest convergent evolution to a unique pollination niche. *Flora* **232**: 207–216.

Orndruff R. **1969**. Reproductive Biology in Relation to Systematics. *Taxon* **18**: 121–133.

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.

Raguso RA, Kelber A, Pfaff M, Levin RA, Lucinda A. 2007. Floral Biology of North American Oenothera Sect. Lavauxia (Onagraceae): Advertisements , Rewards , and Extreme Variation in Floral Depth. *Annals of the Missouri Botanical Garden* **94**: 236–257.

Raguso RA. **2008**. Wake Up and Smell the Roses: The Ecology and Evolution of Floral Scent. *Annual Review of Ecology, Evolution, and Systematics* **39**: 549–569.

Raguso RA. **2016**. Plant Evolution: Repeated Loss of Floral Scent — A Path of Least Resistance? *Current Biology* **26**: R1282–R1285.

Sas C, Müller F, Kappel C, Kent T V, 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.

Schoen DJ, Johnston MO, L'Heureux A-M, Marsolais J V. 1997. Evolutionary History of the Mating System in *Amsinckia* (Boraginaceae). *Evolution* **51**: 1090.

Sheehan H, Hermann K, Kuhlemeier C. Color and Scent: How Single Genes Influence Pollinator Attraction. *Cold Spring Harbor Symposia on Quantitative Biology*, Volume LXXVII.

Sicard A, Kappel C, Lee YW, Woźniak NJ, Marona C, Stinchcombe JR, Wright SI, Lenhard M. **2016**. Standing genetic variation in a tissue-specific enhancer underlies selfing-syndrome evolution in *Capsella*. *Proceedings of the National Academy of Sciences* **113**: 13911–13916.

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, Bäurle I, Lenhard M. **2011**. Genetics, Evolution, and Adaptive Significance of the Selfing Syndrome in the Genus Capsella. *The Plant Cell Online* **23**: 3156–3171.

Slotte T, Hazzouri KM, Stern D, Andolfatto P, Wright SI. **2012**. Genetic architecture and adaptive significance of the selfing syndrome in capsella. *Evolution* **66**: 1360–1374.

Slotte T, Hazzouri KM, Ågren JA, Koenig D, Maumus F, Guo Y-L, Steige K, Platts AE, Escobar JS, Newman LK, *et al.* 2013. The *Capsella rubella* genome and the genomic consequences of rapid mating system evolution. *Nature Genetics* **45**: 831–835.

Spitzer-Rimon B, Marhevka E, Barkai O, Marton I, Edelbaum O, Masci T, Prathapani N-K, Shklarman E, Ovadis M, Vainstein A. 2010. *EOBII*, a Gene Encoding a Flower-Specific Regulator of Phenylpropanoid Volatiles' Biosynthesis in Petunia. *The Plant Cell* **22**: 1961-1976.

Spitzer-Rimon, B., Farhi, M., Albo, B., Cna'ani, A., Ben Zvi, M. M., Masci, T., … Vainstein, A. 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**(12): 5089–5105.

Stebbins GL. **1970**. Adaptive Radiation Of Reproductive Characteristics In Angiosperms, I: Pollination Mechanisms. *Annu. Rev. Ecol. Syst.* **1**: 307-326.

Stern DL. **2013**. The genetic causes of convergent evolution. *Nature Publishing Group* **14**.

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

Takayama S, Isogai A. **2005**. Self-Incompatibility in Plants. *Annu. Rev. Plant Biol* **56**: 467–89.

Tsuchimatsu T, Shimizu KK. **2013**. Effects of pollen availability and the mutation bias on the fixation of mutations disabling the male specificity of self-incompatibility. *Journal of Evolutionary Biology* **26**: 2221–2232.

Tsuchimatsu T, Suwabe K, Shimizu-Inatsugi R, Isokawa S, Pavlidis P, Städler T, Suzuki G, Takayama S, Watanabe M, Shimizu KK. **2010**. Evolution of self-compatibility in *Arabidopsis* by a mutation in the male specificity gene. *Nature* **464**: 1342–1346.

Verdonk, J. C., Haring, M. A., Van Tunen, A. J., & Schuurink, R. C. 2005. Odorant1 regulates fragrance biosynthesis in Petunia flowers. *Plant Cell*, **17**(5): 1612–1624.

Vogt T. **2009**. Phenylpropanoid Biosynthesis. *Molecular Plant* **3:** 2-20.

Widhalm JR, Dudareva N. **2015**. A Familiar Ring to It: Biosynthesis of Plant Benzoic Acids. *Molecular Plant* **8**: 83–97.

Woźniak NJ, Sicard A. **2017**. Evolvability of flower geometry: Convergence in pollinator-driven morphological evolution of flowers. *Seminars in Cell and Developmental Biology* **79**: 3-15.

Wozniak N, Sicard A. **2018**. How to Become Selfish: Evolution and Adaptation to Self-fertilization in Plants. In: Origin and Evolution of Biodiversity. Cham: Springer International Publishing, 143–162.

Woźniak NJ. 2019. *Convergent Evolution of the selfing syndrome in the genus Capsella: Inferring the genetic basis and evolutionary history of selfing syndrome traits*. PhD thesis. University of Potsdam, Germany.

Wray GA. **2007**. The evolutionary significance of cis-regulatory mutations. *Nature Reviews Genetics* **8**: 206–216.

Wright SI, Slotte T, Kalisz S. 2013. Evolutionary consequences of self-fertilization in plants. *Proceedings of the Royal Society B* **280**: 20130133.

Yoo H, Widhalm JR, Qian Y, Maeda H, Cooper BR, Jannasch AS, Gonda I, Lewinsohn E, Rhodes D, Dudareva N. **2013**. An alternative pathway contributes to phenylalanine biosynthesis in plants via a cytosolic tyrosine:phenylpyruvate aminotransferase. *Nature Communications* **4**: 2833.

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

Affidavit

Hereby I declare that I wrote the PhD thesis “Molecular Basis and Adaptive Significance of Repeated Scent Loss in Selfing *Capsella* Species” independently and on my own. I clearly stated the contribution or substantial help from other people. Furthermore, I confirm that this thesis has not been submitted as part of another examination process, neither in identical nor in similar form.

Ich versichere hiermit, dass ich die Arbeit mit dem Titel “Molecular Basis and Adaptive Significance of Repeated Scent Loss in Selfing *Capsella* Species” selbstständig und ohne unzulässige fremde Hilfe erbracht habe. Ich habe den Beitrag oder die wesentliche Hilfe anderer deutlich angegeben und Zitate kenntlich gemacht. Des Weiteren bestätige ich, dass diese Arbeit weder in gleicher noch in ähnlicher Form für andere Prüfungsverfahren eingereicht wurde.

Potsdam, April 2019 Friederike Jantzen