Natural Enemies Drive Geographic Variation in Plant Defenses

Tobias Züst[1*†], Christian Heichinger[2], Ueli Grossniklaus[2], Richard Harrington[3], Daniel J. Kliebenstein[4,5] and Lindsay A. Turnbull[1]

Plants defend themselves against attack by natural enemies and these defenses vary widely across populations. However, whether communities of natural enemies are a sufficiently potent force to maintain polymorphisms in defensive traits is largely unknown. Here, we exploit the genetic resources of *Arabidopsis thaliana*, coupled with 39 years of field data on aphid abundance to: (i) demonstrate that geographic patterns in a polymorphic defense locus (*GS-ELONG*) are strongly correlated with changes in the relative abundance of two specialist aphids; and (ii) demonstrate differential selection by the two aphids on *GS-ELONG*, using a multi-generation selection experiment. We thereby show a causal link between variation in abundance of the two specialist aphids and the geographic pattern at *GS-ELONG*, which highlights the potency of natural enemies as selective forces.

Intraspecific genetic variation is essential in enabling species to respond rapidly to evolutionary challenges such as changing environmental conditions (*1*) or the emergence of novel pests and pathogens (*2*). This diversity often reflects the balance between the strength of local selection and the current and historical levels of population substructure and gene flow (*3, 4*). Geographic analyses of genetic variation in several plant species have revealed clear genetic signals of local adaptation (*5*), caused by differences in the selective regime among locations. These analyses are further supported by reciprocal transplant experiments, in which home genotypes generally outperform those transplanted from other populations (*6, 7*). While the drivers of local adaptation often remain unidentified, there is evidence that climate and soil can exert strong local selective pressures and play important roles in shaping large-scale genetic patterns (*8, 9*).

In contrast to the clear role of abiotic factors, there is little direct evidence that biotic forces, such as herbivory or competition, can lead to the maintenance of genetic variation across large geographic scales, despite the exceptional levels of polymorphism associated with genes involved in defense (*10, 11*). In theory, interactions between organisms and their natural enemies can lead to differences in the local selective regime because of geographic variation in the abundance or species composition of the enemy community (*3*). This spatial variation can affect defense if it is costly; e.g., if the average level of herbivory varies across populations, defended genotypes might dominate in heavily attacked populations while undefended genotypes would prosper when enemies are absent or rare (*12*). Another less studied effect is how defense might vary if plants are attacked by diverse collections of herbivore species that differ in feeding style and specialization. This could lead to higher levels of polymorphism in

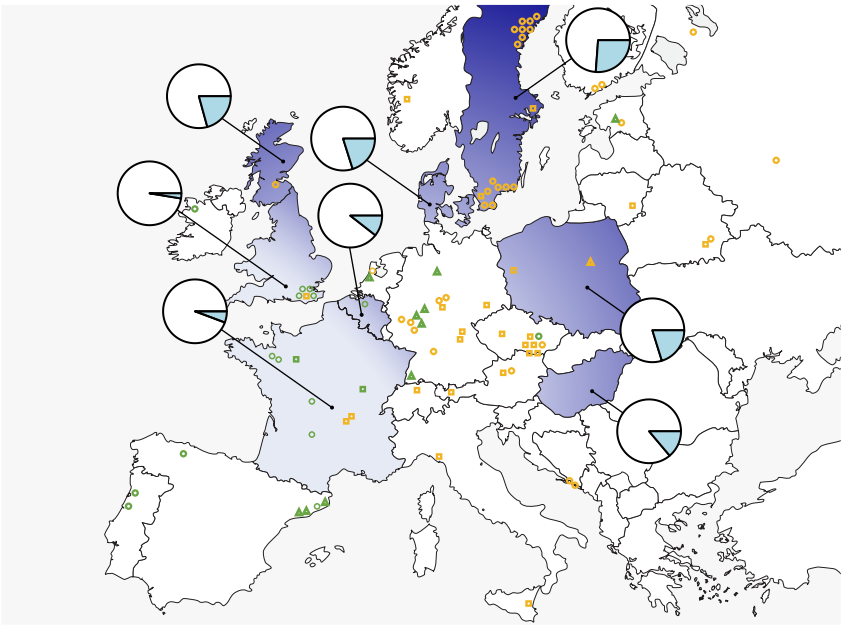

Fig. 1. Location of European *A. thaliana* accessions with known chemical profile. Symbol color indicates the *GS-ELONG* chemotype (orange: *3C*; green: *4C*) and symbol shape indicates the *GS-AOP* chemotype (square: *ALK*; circle: *OH*; triangle: *NULL*). For GS-ELONG the probability of finding *3C* populations increases strongly with longitude (binomial glm: $t = 5.11$, df = 85, $p < 0.001$) and weakly with latitude ($t = 1.75$, df = 85, $p = 0.084$). Countries with available aphid data are colored in blue. The shade of blue corresponds to the relative frequency of *L. erysimi* based on model predictions from a binomial GLM using data from 61 aphid suction traps. The relative frequency of *L. erysimi* increases strongly with longitude ($t = 5.03$, $p < 0.001$) and weakly with latitude ($t = 1.89$, $p = 0.060$). Piecharts indicate the observed average relative abundance of *B. brassicae* (white) and *L. erysimi* (blue) in each country.

[1] Institute of Evolutionary Biology and Environmental Studies and Zürich-Basel Plant Science Center, University of Zürich, Zürich, CH-8057, Switzerland. [2] Institute of Plant Biology and Zürich-Basel Plant Science Center, University of Zürich, Zürich, CH-8008, Switzerland. [3] Rothamsted Insect Survey, AgroEcology Department, Rothamsted Research, Harpenden AL5 2JQ, UK. [4] Department of Plant Sciences, University of California at Davis, Davis, California 95616, USA. [5] DynaMo Center of Excellence, Department of Plant Biology and Biotechnology, University of Copenhagen, DK-1871, Denmark.

*To whom correspondence should be addressed. E-mail: tobias.zuest@cornell.edu.

†Present address: Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY 14853, USA.

defense genes due to selection for specific defensive profiles matched to the predominant local herbivore or herbivore community (e.g., *13*). However, there is no direct evidence that variation in local herbivore communities represents a sufficiently strong selective pressure to favor specific defensive traits and maintain polymorphisms in defense-related genes.

The unparalleled genetic and molecular resources available for the model plant *Arabidopsis thaliana* make this species an ideal candidate to study the process of local adaptation to herbivores. The primary defensive trait in *A. thaliana* is a series of indolic and aliphatic glucosinolates, which are secondary plant metabolites with anti-herbivore properties (*14*). The accumulation and structure of aliphatic glucosinolates is mechanistically determined by alleles at the *GS-ELONG* locus that regulate the carbon side-chain elongation (*3C* or *4C*) (*15*) and by alleles at the *GS-AOP* locus that modify the functional group of the biologically active glucosinolate side-chain (*ALK*, *OH*, or *NULL*). The combination of these alleles yields six distinct chemotypes present in natural populations in varying proportions (*16*). Both individual glucosinolate compounds and full chemical profiles affect the susceptibility of a plant to specific herbivores (*17, 18*); hence the aliphatic chemotype is likely under differential, qualitative selection by herbivores. In contrast, accumulation of the main indolic glucosinolates in *A. thaliana* is highly plastic and modulated by a large number of small-effect genetic loci, which are therefore less likely to show clear signatures of selection (*19*).

We mapped geographic variation in the abundance of the six chemotypes within Europe from a set of 96 accessions (75 European) (*20*) with known chemical profiles (*16*) (Fig. 1). There was no apparent pattern in the distribution of the *GS-AOP* chemotypes, but for *GS-ELONG* the frequency of 3C to 4C chemotypes increases with both latitude and longitude (Figs 1, S1). If this pattern results from geographical variation in herbivore feeding pressure, we would expect it to be closely matched by variation in herbivore abundance patterns. While *A. thaliana* is attacked by a range of invertebrate herbivores, many of which preferentially feed on specific chemotypes (*17*), we hypothesized that the aphid species *Brevicoryne brassicae* and *Lipaphis erysimi* are likely drivers of these patterns as they are both abundant, mobile Brassicaceae specialists, yet differentially sensitive to environmental conditions (*21*). Fluctuations in aphid populations have been monitored since 1964 through the EXAMINE network (http://www.rothamsted.ac.uk/examine/) using suction traps which operate throughout the aphid flight season (*22*). We retrieved data on the two aphid species from 61 traps in eight European countries. These data revealed that the abundance of *L. erysimi* is usually lower than *B. brassicae*, but that the geographic pattern in the relative abundances of *L. erysimi* and *B. brassicae* closely mirrors the pattern at GS-ELONG (Figs 1 and S1). Variation in the relative abundance of these two specialist aphids could therefore underlie variation in the predominant *GS-ELONG* chemotype found in natural populations.

Since causal inferences are impossible from such correlative data, we tested the causality of aphid selection on *GS-ELONG*, carrying out a multi-generational selection experiment on populations of *A. thaliana* (*22*). We assembled 30 replicate populations from equal numbers of seeds from each of 27 natural accessions, including 6 of the 75 European accessions mapped above. Accessions were chosen to maximize variation in defense traits while including all six glucosinolate chemotypes in a range of genetic backgrounds (Table S1, Fig. S2). Over five generations, we consistently exposed populations to replicate (n = 6) treatments of a single specialist aphid species: either *B. brassicae* or *L. erysimi*; a single generalist aphid, *Myzus persicae*; a mixture of all three aphid species; and a no-aphid treatment. The generalist aphid was included as a negative control, since *M. persicae* is unresponsive to aliphatic glucosinolates (*23*) and we therefore would not expect it to exert directional selection on plant chemotype. The no-aphid treatment served as a control for other selective forces that were likely to affect the outcome of the experiment, such as intraspecific competition among accessions. Seeds were collected at the end of each generation with no mixing among populations and a subset was used to establish the next generation at a constant density. After five generations of repeated herbivore treatments we sampled 24 individuals from each population in generation 5 (144 individuals per treatment) and determined their genotype. To have a marker for changes in genotypic composition through time, we also measured leaf trichome density, a trait under strong genetic control (Fig. S3), on a representative sample of plants in all generations.

Rapid adaptation occurred in the selection experiment, as evidenced by a progressive reduction in the effects of aphid feeding on final plant biomass in each generation (Fig. 2A). In line with the expected severity of aphid feeding based on previously reported population growth rates (*21*), *L. erysimi* caused the strongest reduction in plant biomass, while *M. persicae* was intermediate and *B. brassicae* had the least effect. The mixture treatment caused a similar reduction to *L. erysimi* alone, probably because aphid mixtures were dominated by this fast-growing aphid species. With each generation, trichome density decreased in the no-aphid treatment, while it remained at significantly higher levels in all aphid treatments (Fig. 2B). Adaptation to herbivore feeding was accompanied by considerable changes in the genotypic composition of populations, including the complete loss of nine genotypes (Fig. 3). There was a non-specific aphid effect on total indolic glucosinolates (lme: $F_{1,28} = 10.66$, $p = 0.003$), with plants in the no-aphid treatment producing on average 0.98 (± 0.03, SEM) µmol g^{-1}, and plants in aphid treatments producing 0.87 (± 0.03, SEM) µmol g^{-1}. In contrast, the different aphid treatments had a dramatic impact on the dominant aliphatic chemotypes within experimental populations. Significantly, the relative proportions of *3C* and *4C* chemotypes differed strongly among aphid treatments (Figs. 3, S4). After selection, populations of

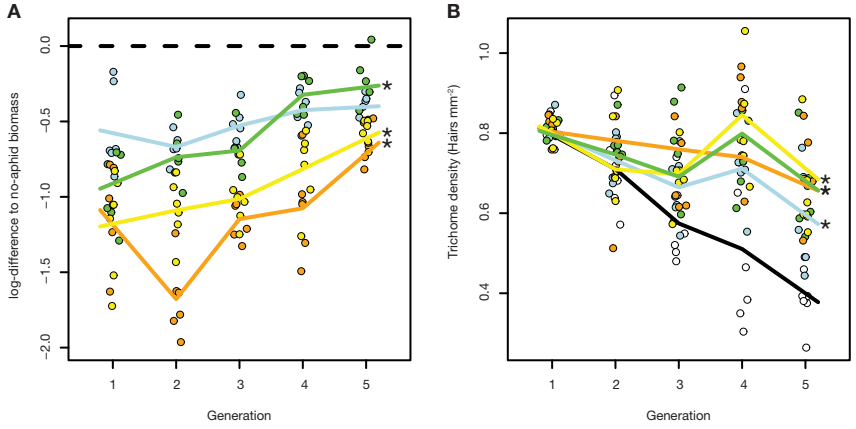

Fig. 2. A) Change in the negative impact of aphid treatments on final plant biomass over five generations, displayed as the log-difference to final plant biomass in the no-aphid treatment: *B. brassicae* (light blue); *M. persicae* (light green); *L. erysimi* (orange); and aphid mixture (yellow). Stars denote significantly less damage after five generations of selection (Table S2). B) Mean number of trichomes on the fourth leaf of 50 plants per population. Stars denote significant difference from the no-aphid treatment (black line) after five generations of selection.

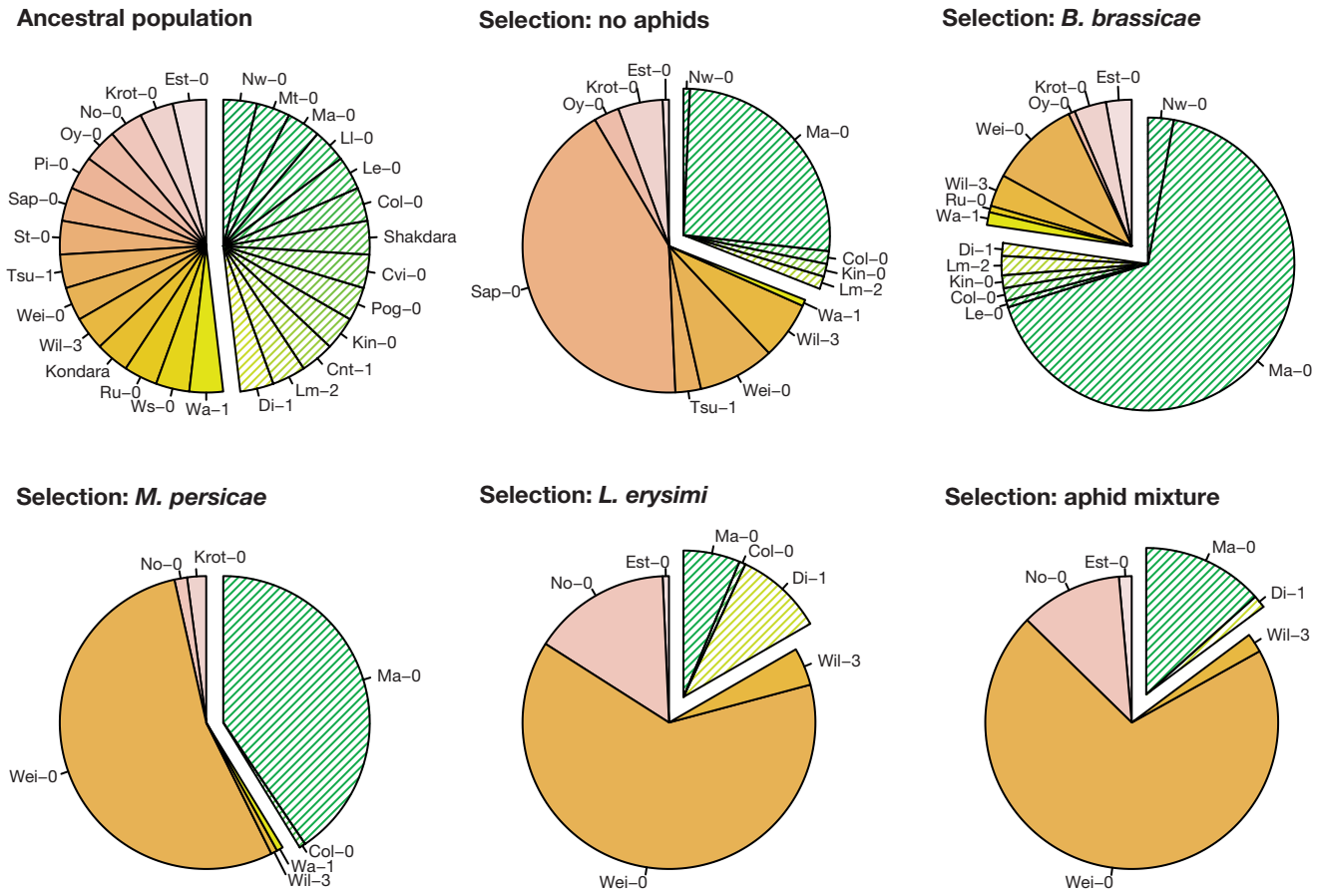

Fig. 3. Change in the composition of *A. thaliana* accessions, from equal proportions of 27 genotypes in the ancestral population to treatment-specific compositions after five generations of selection. Each chart gives mean genotype frequencies based on $n = 6$ replicate populations. *3C* chemotypes are indicated by solid, orange colors, while *4C* chemotypes are indicated by hatched, green colors.

the no-aphid treatment consisted of approximately two thirds *3C* and one third *4C* chemotypes. Specialist aphids selected for different chemotypes at *GS-ELONG*: the *4C* chemotypes strongly dominated in *B. brassicae* treatments (binomial glm, t = 3.08, df = 25, p = 0.002) and the *3C* chemotypes strongly dominated in both *L. erysimi* (t = 2.01, df = 25, p = 0.045) and the aphid mixture treatments (t = 2.21, df = 25, p = 0.027). The relative proportions of *3C* to *4C* chemotypes in populations exposed to the generalist aphid *M. persicae* did not differ from the no-aphid treatment (t = 0.18, df = 25, p = 0.858). Despite this similarity, the identity of the successful genotypes differed among treatments, with accession Sap-0 accounting for a large fraction of plants in the no-aphid treatment but being absent from all other treatments (Fig. 3). The genotypic composition of plant populations with *L. erysimi* and aphid mixtures was near-identical, confirming that *L. erysimi* dominated the mixture treatments, and suggesting that in co-founded populations, *L. erysimi* is the most important selective force. Most successful genotypes either had a *3C-OH* or a *4C-NULL* chemotype, and we found no individuals belonging to either alkenyl chemotype (*3C-ALK* or *4C-ALK*) in any treatment. Alkenyl chemotypes were common in generation 1 of the selection experiment (Fig. S2), and simulations of random sampling on the basis of observed population sizes reveal that their loss cannot be due to drift alone (Fig. S5) but rather was a consequence of selection.

To identify potential causes for the loss of particular genotypes, we measured size-standardized growth rate (SGR) as a measure of fitness, together with total aliphatic glucosinolate content and trichome density in a separate experiment on all 27 ancestral accessions. Alkenyl chemotypes expressed the highest levels of glucosinolates and were among the slowest growing genotypes overall (Fig. S6A). Alkenyl glucosinolates are an effective defense against leaf-chewing herbivores such as caterpillars (*24*), but their efficiency against specialist aphids remains largely unknown, while they have little effect on *M. persicae* (*23*). The loss of the alkenyl chemotypes therefore probably resulted from selection against a costly defense trait that provided insufficient benefits in our experiment. This cost-benefit balance is also the most likely reason for the difference in dominant genotypes between the no-aphid treatment and the aphid treatments (Fig. 3). The dominant genotype in no-aphid populations, Sap-0, was completely absent from all aphid treatments, indicating low fitness in the presence of herbivores. The Sap-0 genotype had the lowest trichome density of all non-glabrous accessions, and as trichome production had a growth cost (Fig S6B), its success can explain the observed decrease in trichome density in the no-aphid populations over time (Fig. 2B). Compared to other chemotypes, Sap-0 also produces low levels of glucosinolates, an additional indication that in the absence of herbivores, undefended, fast-growing genotypes will prosper.

Despite known epistatic interactions between *GS-ELONG* and *GS-AOP* (*19*), our data suggest that aphid selection acts independently on the two loci. The magnitude and direction of selection exerted by the two specialist aphids on *GS-ELONG* in our experiment suggests a causal link between the observed cline in *GS-ELONG* across Europe and the changes in the relative abundance of the same aphids. Although *B. brassicae* is numerically

dominant across most of Europe, the faster-growing *L. erysimi* can inflict greater damage on plants and quickly dominates populations which are co-founded by both aphid species, thus even a modest change in relative abundance could cause the loss of *C3* populations. All plants in the selection experiment experienced strong intraspecific competition, and since growth rate is a good predictor of competitive ability (*25*) it is unsurprising that fast-growing plant genotypes were generally selected, while the slowest-growing alkenyl chemotypes were lost. Alkenyl chemotypes are, however, very common in natural populations, and could be maintained by other herbivores, for instance leaf-chewing caterpillars (*24*).

Ecological theory has consistently emphasized the role of natural enemies in maintaining diversity both within and among species, but convincing empirical evidence has been lacking. Here we demonstrate that even functionally similar herbivores such as different species of aphid have the potential to select for specific chemotypes and drive large-scale geographic patterns in plant defense profiles. It therefore seems likely that natural herbivore communities with their greater variety of feeding styles and specializations play a major role in shaping and refining the plant defenses observed in natural communities.

References and Notes:

1. G. Bell, A. Gonzalez, Adaptation and evolutionary rescue in metapopulations experiencing environmental deterioration. *Science* **332**, 1327-1330 (2011).
2. M. C. Fisher *et al.*, Emerging fungal threats to animal, plant and ecosystem health. *Nature* **484**, 186-194 (2012).
3. B. Charlesworth, M. Nordborg, D. Charlesworth, The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genet. Res.* **70**, 155-174 (1997).
4. M. Slatkin, Gene flow and the geographic structure of natural populations. *Science* **236**, 787-792 (1987).
5. Y. B. Linhart, M. C. Grant, Evolutionary significance of local genetic differentiation in plants. *Annu. Rev. Ecol. Syst.* **27**, 237-277 (1996).
6. A. Fournier-Level *et al.*, A map of local adaptation in *Arabidopsis thaliana*. *Science* **334**, 86-89 (2011).
7. R. Leimu, M. Fischer, A meta-analysis of local adaptation in plants. *Plos One* **3** (2008).
8. N. Alvarez *et al.*, History or ecology? Substrate type as a major driver of spatial genetic structure in Alpine plants. *Ecol. Lett.* **12**, 632-640 (2009).
9. A. M. Hancock *et al.*, Adaptation to climate across the *Arabidopsis thaliana* genome. *Science* **333**, 83-86 (2011).
10. J. Bergelson, M. Kreitman, E. A. Stahl, D. C. Tian, Evolutionary dynamics of plant R-genes. *Science* **292**, 2281-2285 (2001).
11. R. M. Clark *et al.*, Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana*. *Science* **317**, 338-342 (2007).
12. S. Y. Strauss, J. A. Rudgers, J. A. Lau, R. E. Irwin, Direct and ecological costs of resistance to herbivory. *Trends Ecol. Evol.* **17**, 278-285 (2002).
13. M. T. J. Johnson, A. A. Agrawal, Plant genotype and environment interact to shape a diverse arthropod community on evening primrose (*Oenothera biennis*). *Ecology* **86**, 874-885 (2005).
14. B. A. Halkier, J. Gershenzon, Biology and biochemistry of glucosinolates. *Annu. Rev. Plant Biol.* **57**, 303-333 (2006).
15. J. Kroymann, S. Donnerhacke, D. Schnabelrauch, T. Mitchell-Olds, Evolutionary dynamics of an *Arabidopsis* insect resistance quantitative trait locus. *Proc. Natl. Acad. Sci. USA* **100**, 14587-14592 (2003).
16. E. K. F. Chan, H. C. Rowe, J. A. Corwin, B. Joseph, D. J. Kliebenstein, Combining genome-wide association mapping and transcriptional networks to identify novel genes controlling glucosinolates in *Arabidopsis thaliana*. *PLoS Biol.* **9** (2011).
17. M. G. Bidart-Bouzat, D. J. Kliebenstein, Differential levels of insect herbivory in the field associated with genotypic variation in glucosinolates in *Arabidopsis thaliana*. *J. Chem. Ecol.* **34**, 1026-1037 (2008).
18. E. L. Newton, J. M. Bullock, D. J. Hodgson, Glucosinolate polymorphism in wild cabbage (*Brassica oleracea*) influences the structure of herbivore communities. *Oecologia* **160**, 63-76 (2009).
19. D. J. Kliebenstein, J. Gershenzon, T. Mitchell-Olds, Comparative quantitative trait loci mapping of aliphatic, indolic and benzylic glucosinolate production in *Arabidopsis thaliana* leaves and seeds. *Genetics* **159**, 359-370 (2001).
20. M. Nordborg *et al.*, The pattern of polymorphism in *Arabidopsis thaliana*. *PLoS Biol.* **3**, 1289-1299 (2005).
21. C. J. DeLoach, Rate of increase of populations of cabbage, green peach, and turnip aphids at constant temperatures. *Ann. Entomol. Soc. Am.* **67**, 332-340 (1974).
22. Materials and methods are available as supplementary material on Science Online.
23. J. H. Kim, G. Jander, *Myzus persicae* (green peach aphid) feeding on Arabidopsis induces the formation of a deterrent indole glucosinolate. *Plant J.* **49**, 1008-1019 (2007).
24. D. J. Kliebenstein, Secondary metabolites and plant/environment interactions: a view through *Arabidopsis thaliana* tinged glasses. *Plant Cell Environ.* **27**, 675-684 (2004).
25. S. Fakheran *et al.*, Adaptation and extinction in experimentally fragmented landscapes. *Proc. Natl. Acad. Sci. USA* **107**, 19120-19125 (2010).
26. R. Harrington *et al.*, in *Aphids in a new millennium* J. C. Simon, C. A. Dedryver, C. Rispe, M. Hullé, Eds. (Proceedings 6th International Aphid Symposium, Rennes, France, 2004) pp. 45-49.
27. L. R. Taylor, Synoptic dynamics, migration and the Rothamsted insect survey - presidential address to the British Ecological Society, December 1984. *J. Anim. Ecol.* **55**, 1-38 (1986).
28. E. D. M. Macaulay, G. M. Tatchell, L. R. Taylor, The Rothamsted insect survey 12-metre suction trap. *B. Entomol. Res.* **78**, 121-129 (1988).
29. R. Harrington *et al.*, Environmental change and the phenology of European aphids. *Glob. Change Biol.* **13**, 1550-1564 (2007).
30. N. Cocu, R. Harrington, M. Hullé, M. D. A. Rounsevell, Spatial autocorrelation as a tool for identifying the geographical patterns of aphid annual abundance. *Agr. Forest Entomol.* **7**, 31-43 (2005).
31. D. J. Kliebenstein *et al.*, Genetic control of natural variation in *Arabidopsis* glucosinolate accumulation. *Plant Physiol.* **126**, 811-825 (2001).
32. J. Hagenblad *et al.*, Haplotype structure and phenotypic associations in the chromosomal regions surrounding two *Arabidopsis thaliana* flowering time loci. *Genetics* **168**, 1627-1638 (2004).
33. S. Atwell *et al.*, Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* **465**, 627-631 (2010).
34. C. J. Bell, J. R. Ecker, Assignment of 30 microsatellite loci to the linkage map of *Arabidopsis*. *Genomics* **19**, 137-144 (1994).
35. K. Edwards, C. Johnstone, C. Thompson, A simple and rapid method for the preparation of plant genomic DNA for PCR analysis ANALYSIS. *Nucleic Acids Res.* **19**, 1349-1349 (1991).
36. R Development Core Team. R Foundation for Statistical Computing, Vienna, Austria (2010).
37. W. S. Rasband. U. S. National Institutes of Health, Bethesda, Maryland, USA (1997-2009).
38. T. Züst, B. Joseph, K. K. Shimizu, D. J. Kliebenstein, L. A. Turnbull, Using knockout mutants to reveal the costs of defensive traits. *Proc. R. Soc. B* **278**, 2598-2603 (2011).
39. J. C. Pinheiro, D. M. Bates, *Mixed-Effects Models in S and S-Plus*, Statistics and Computing (Springer, London, 2000).

Acknowledgments: This work was supported by the Forschungskredit of the University of Zürich to TZ and by Swiss National Science Foundation Grant 31-107531 to LAT, by US National Science Foundation grant DBI0820580 to DJK, an Advanced Grant of the European Research Council to UG, and a special grant of the University of Zürich in the commemoration of Prof. Christine B. Müller. The Rothamsted Insect Survey is supported by a BBSRC National Capability Grant and the Lawes Agricultural Trust. We thank members of the EU EXAMINE Project (EVK2-1999-00151) for use of data and A. A. Agrawal, S. West, J. Levine, and two anonymous reviewers for comments that improved the manuscript. All data presented in this paper and R code of all analyses are provided in the Supporting Online Material.

Supplementary Materials:

www.sciencemag.org
Materials and Methods
Figs. S1-S7
Tables S1, S2, S3
File with R code
Datafiles s3-s8

Supplementary Materials:

Additional Author notes: TZ and LAT proposed the original idea, performed statistical analyses, and wrote the paper; CH, UG, RH and DJK contributed data or provided molecular and chemical analyses; all authors contributed to revisions; TZ created the figures.

Materials and Methods:

Analysis of geographic patterns. We extracted coordinates data on the set of 96 natural accessions from Nordborg et al. (*20*) and combined it with glucosinolate data from Chan et al. (*16*). We analyzed the spatial pattern in *GS-ELONG* for Europe using a generalized linear model with binomial error structure and longitude and latitude as explanatory variables. There was no support for an interaction term; hence only main effects were included in the model. Data on aphid abundance were extracted from the database of the 'EXAMINE' network (*26*). This project emerged from the Rothamsted Insect Survey (*27*), which uses a network of 12.2 m high suction traps (*28*) throughout the UK to monitor aphids for research and extension work. Daily data are available for a large number of aphid species back to 1964 depending on the site. Other countries have adopted the same design of trap and the data have been assembled within a single database (*26*), facilitating pan-European analysis (*29*). There are currently 46 traps operating in ten countries, each representative of aphids flying within a surrounding area with a radius of more than 200 km (*30*). Many more traps have been operated in the past but have now been closed. This paper draws on samples from a total of 61 traps with varying extent of data coverage. Data from any particular trap in a given year were only used if at least one individual of both *Brevicoryne brassicae* and *Lipaphis erysimi* each were recorded in that location, since lack of data can be caused both by true absence of an aphid or by failure to monitor or record this species. We analyzed the aphid data with a binomial generalized mixed effects model (binomial glmer) for proportion data by treating the

total numbers of *L. erysimi* and *B. brassicae* as successes and failures, respectively. Trap identity was included as a random effect to account for both repeated measures and differences in the range of annual data. Longitude and latitude was treated as fixed effects, but as with *GS-ELONG*, there was no support for an interaction term.

Study system. To create a genetically diverse ancestral population, we selected 30 accessions of *A. thaliana*, based on published information on glucosinolate profiles (*31*), flowering time (*32*), and trichome density (NASC, The European Stock Centre, www.arabidopsis.info). These lines are maintained in stock centers through selfing and are thus homozygous for most of their genome (*33*). We obtained seeds of all accessions through NASC and propagated them for one generation to amplify seed stocks and reduce potential confounding maternal effects. Three accessions completely failed to germinate; hence we used the 27 remaining accessions for the selection experiment (Table S1). We established laboratory stock cultures of the three aphid species *M. persicae*, *B. brassicae* and *L. erysimi.* Each culture was founded from a single adult aphid which we collected from naturally occurring Brassicaceae in the garden of the University of Zürich at the beginning of the experiment in spring 2009.

Selection experiment. Experimental populations were set up in individual cages made from Plexiglas® (Fig. S6) and were maintained at 18° C under a day/night cycle of 16h/8h. Plant populations in the first generation were established from 20 seeds per genotype (540 seeds in total), and in all subsequent generations 800 seeds were randomly selected to establish new populations in fresh soil. To establish plant populations at the beginning of a generation, seeds were evenly sprinkled into planting trays (18 x 35 cm) filled with 2.5 liters of standard germination soil (pre-mixed with vermiculite, GO M1, Tref Group, The Netherlands). Trays were thoroughly soaked with water, covered with transparent plastic film and cold stratified at 4° C for four days and placed into the cages inside the climate chamber afterwards.

Germination was measured in all populations and 15 days after sowing 50 random plants were sampled per population, removing the fourth leaf from 50 plants and counting the number of trichomes within a defined area on the adaxial surface of the leaf (Ø 4 mm). 17 days after sowing, aphid treatments were initiated by applying six aphids (two per species for aphid mixtures). After the introduction of aphids, cages were checked regularly for cross-contamination among treatments. Contaminations occurred rarely, and only toward the very end of the generational cycle and were therefore considered harmless due to already advanced plant senescence. Since plants in the no-aphid treatment senesced more slowly, contaminations in this treatment were treated with a systemic insecticide (ACTARA®, Syngenta, Switzerland) that killed aphids within a few days. All populations were harvested after 60 days when most plants had senesced. Every generation, the locations of cages in the growth chamber were re-randomized following a modified stratified random design.

Genotyping. We randomly selected 24 plants per population in generation 5 (144 per treatment), and genotyped them using a set of SSLP markers (*34,* see Table S3 for details). Plants were grown in a controlled climate chamber set to 18° C under a day/night cycle of 16h/8h and fifteen days after germination, we harvested two leaves from each plant, placed them into a test tube within a 96-tube rack format, and immediately froze the samples on dry ice for later DNA extraction. Several small glass beads (Ø 1 mm) were added to each tube and 96-tube racks were frozen in liquid nitrogen. Frozen samples were ground by shaking 2 x 30 sec (frequency = 30/sec), turning plates once by 180°, on a Mixer Mill Retsch MM300 (Retsch Technology GmbH, Germany) and DNA was extracted (*35*).

To distinguish the 27 ancestral accessions, we tested five polymorphic loci that produced different fragment size upon amplification (simple sequence length polymorphisms, SSLPs) using PCR (Table S3). We analyzed three markers (*nga6*, *nga172* and *ciw6*) on all plant samples, and a subset of plant samples with ambiguous results was analyzed with two

additional markers (either *nga111* or *ciw3*). PCR was first carried out on DNA of the 27 ancestral lines to establish the reference genotypes. All PCR reactions were carried out using 5 µl DNA and a final reaction volume of 25 µl. The PCR conditions were as follows: 94° C for 120 s (1x); followed by 35 cycles of 94° C for 30 s, 59° C for 20 s, 72° C for 30 s; and 72° C for 10 min (1x). The size of PCR products was analyzed on a QIAxcel® capillary electrophoresis system (Qiagen, Switzerland), using a high-resolution gel cartridge and standard Qiagen reagents. The 15bp/500bp QX Alignment Marker was used to align samples, as PCR product typically ranged between 120 and 250 bp. Following standard Qiagen protocols, we determined DNA fragment size using the BioCalculator Software, which is part of the QIAxcel® system. Plant samples were then assigned to ancestral genotypes using a discriminant analysis (function *lda* in R (*36*)) with fragment sizes of known genotypes as the training sample and unknown plant genotypes as the test sample.

Glucosinolate analysis. We analyzed the chemical profile of an additional set of 24 plants per population of generation 5 (144 plants per treatment). We grew plants in a controlled climate chamber set to 18° C under a day/night cycle of 16h/8h for 20 days, and then we harvested the six largest leaves of each plant. Leaves were placed onto a white plastic surface and photographed for size measurement. Immediately after photographing, all six leaves of one plant were put into a 1.4 ml test tube within a 96-tube rack format (Micronic, The Netherlands), containing 400 µl of 90% methanol, which inhibits the enzymatic breakdown of glucosinolates. Samples were then extracted and analyzed as described in Kliebenstein et al. (*31*). Leaf area of samples was measured from photographs using the open-source image processing software ImageJ (*37*). For a representative subset of genotypes, rosettes were harvested, dried and weighed to generate a calibration line. Using this line, all leaf areas were transformed into leaf masses. Glucosinolate profiles of plant samples were then assigned to

ancestral chemotypes using a discriminant analysis (function *lda* in R) with chemical profiles of known genotypes as training sample and profiles of generation 5 plants as test sample.

Phenotypic screening of *A. thaliana* genotypes. We measured several defense-related traits on the 27 ancestral accessions grown in a controlled climate chamber set to 18° C under a day/night cycle of 16h/8h. These were growth rate (as a descriptor of the competitiveness of a plant (see *25, 38*)), time of flowering, trichome density, and glucosinolate content. Growth rate was measured as biomass growth over the whole plant life using nine sequential harvests. For each harvest, we grew 2-3 individual plants for each of the 27 accessions and harvested plants on days 7, 9, 11, 14, 18, 22, 27, 30 and 38 after sowing. We fitted non-linear growth curves with the function *gnls* implemented in the *nlme* library for R (*36, 39*) to the total above-ground biomass data, using an asymptotic regression and the self-starting routine *SSasymp* (*38*). We calculated size-standardized growth rate (SGR) from the estimated model parameters at a common reference size (mean size 9 days after sowing).

Statistical analyses. All analyses were carried out in R 2.13 for Windows (*36*). For analyses of the selection experiment, aphid treatment was fitted as a five-level factor, with each of the four aphid levels being tested against the no-aphid treatment. All traits with multiple measures per population were analyzed with linear mixed-effects models (lme) implemented in the *nlme* library for R (*39*), using population as a random effect. The change in aphid impact on plant populations over time was analyzed in an lme-model of log-biomass as a function of treatment and generation. In this way, aphid treatment effects are expressed as differences on the log-scale, which is equivalent to log-ratios. Generation was treated as a factor to account for non-linearity in the relation with the response. Average aliphatic and indolic glucosinolate contents in each generation (based on composite leaf samples) were analyzed as log-concentrations in similar lme-models, while for trichome density, the absolute numbers were

analyzed. The overall proportion of the *GS-ELONG* chemotype in generation 5 was analyzed using a generalized linear model (glm) with a binomial error structure.

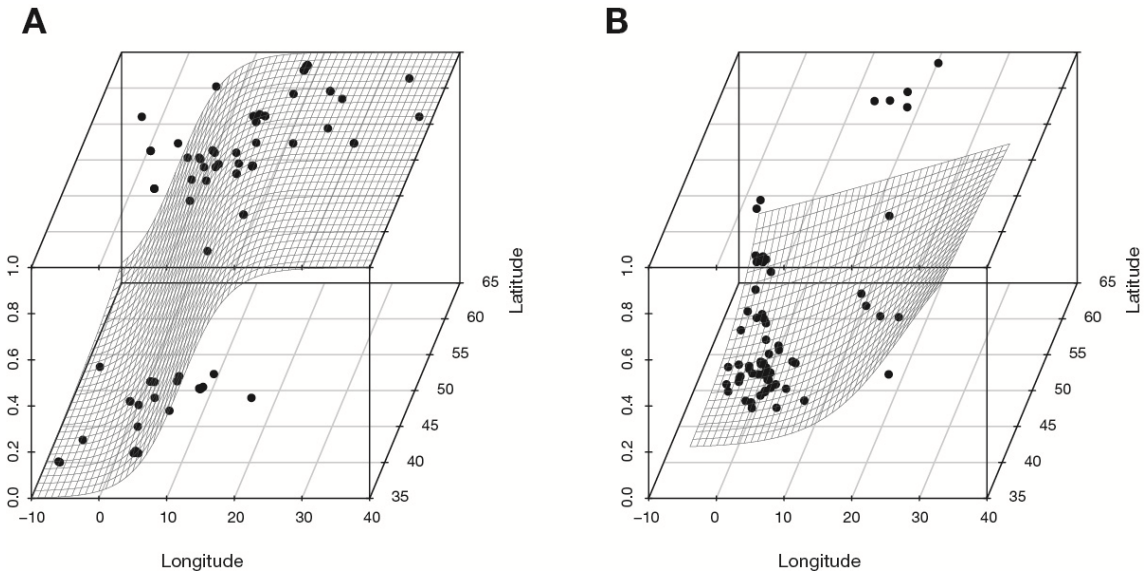

Figure S1. A) Three-dimensional plot of the distribution of *GS-ELONG* across Europe. *4C* chemotypes are represented as zero values and *3C* chemotypes as ones. Each dot represents one accession, and the hatched area is the model fit of a generalized linear model with binomial error structure (binomial glm), *GS-ELONG* as binary response, and longitude and latitude as explanatory variables. **B)** Three-dimensional plot of the relative proportion of the mean number of *Lipaphis erysimi* individuals per *Brevicoryne brassicae* individuals, captured in aerial suction traps distributed across Europe. Each dot represents one trap, and the hatched area is the model fit of a generalized linear mixed-effects model with binomial error structure (binomial glmer); the proportion of the two aphid species from several years of trapping as response, longitude and latitude as explanatory variables, and trap identity as a random effect to account for repeated measures at the same location over multiple years.

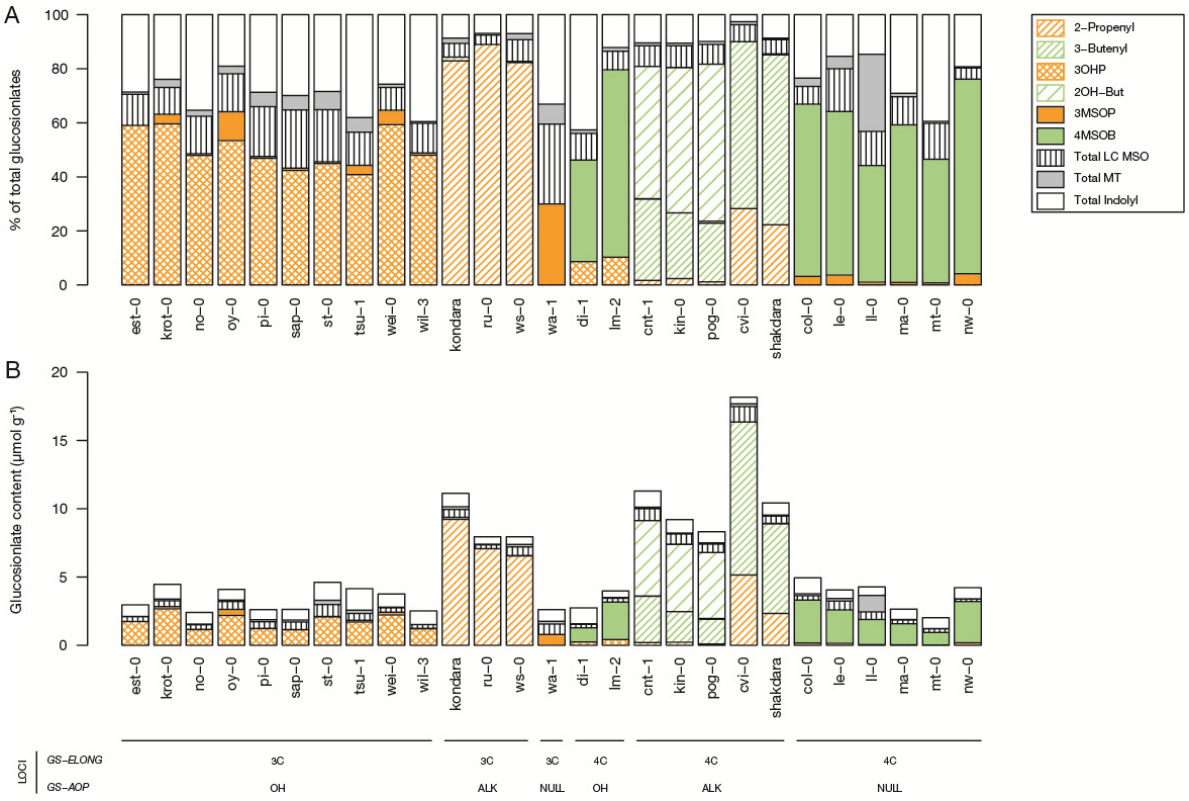

Figure S2. Natural variation in glucosinolate profiles in the 27 accessions (genotypes) of *A. thaliana* used for the selection experiment. Genotypes are ordered according to their chemotype on the basis of allelic variation at two loci, *GS-ELONG* (*3C* and *4C*) and *GS-AOP* (*NULL*, *ALK* and *OH*). Presented are **A)** relative and **B)** absolute concentrations of glucosinolates (µmol g^{-1} dry weight). Compounds with a *3C* side-chain are colored orange, while *4C* compounds are colored green. The biologically active functional group is indicated by the fill of each bar: solid (*NULL*), hatched (*ALK*), and cross-hatched (*OH*). Abbreviations of glucosinolate compounds: *3OHP,* 3-hydroxypropyl; *2OH-But,* 2-hydroxy-3-butenyl; *3MSOP,* 3-methylsulfinylpropyl; *4MSOB,* 4-methylsulfinylbutyl; *Total LC MSO,* long-chain methylsulfinyl; *Total MT,* methylthio.

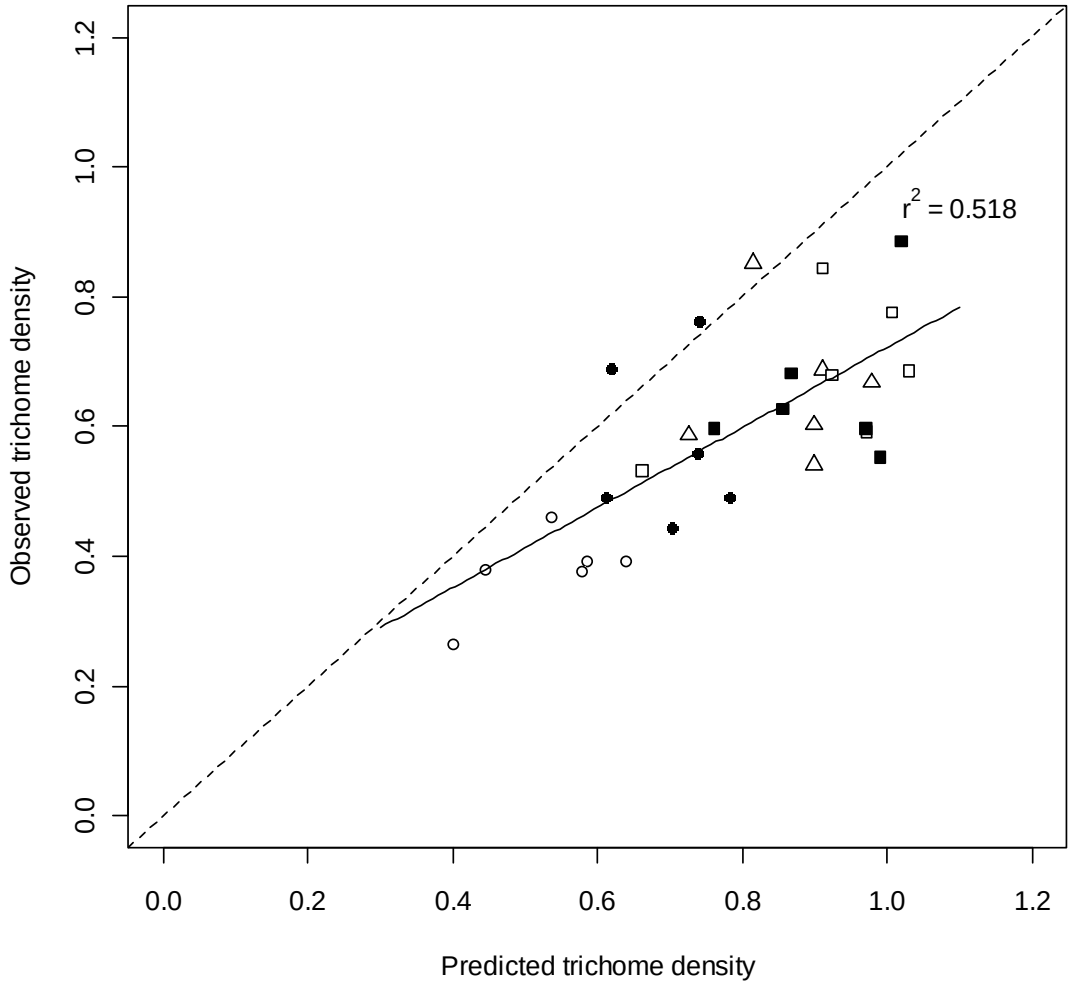

Figure S3. Plot of the predicted *versus* observed trichome density in generation 5 for all 30 populations ($r^2 = 0.52$, $p < 0.001$). Predicted trichome density is based on genotype frequencies in generation 5 and ancestral trait distributions. Symbols represent the five treatments: no aphids (open circles), *M. persicae* (open triangles), *B. brassicae* (filled circles), *L. erysimi* (open squares), and aphid mixture (filled squares). The dotted line indicates the 1:1 line, and the solid line is the actual model fit.

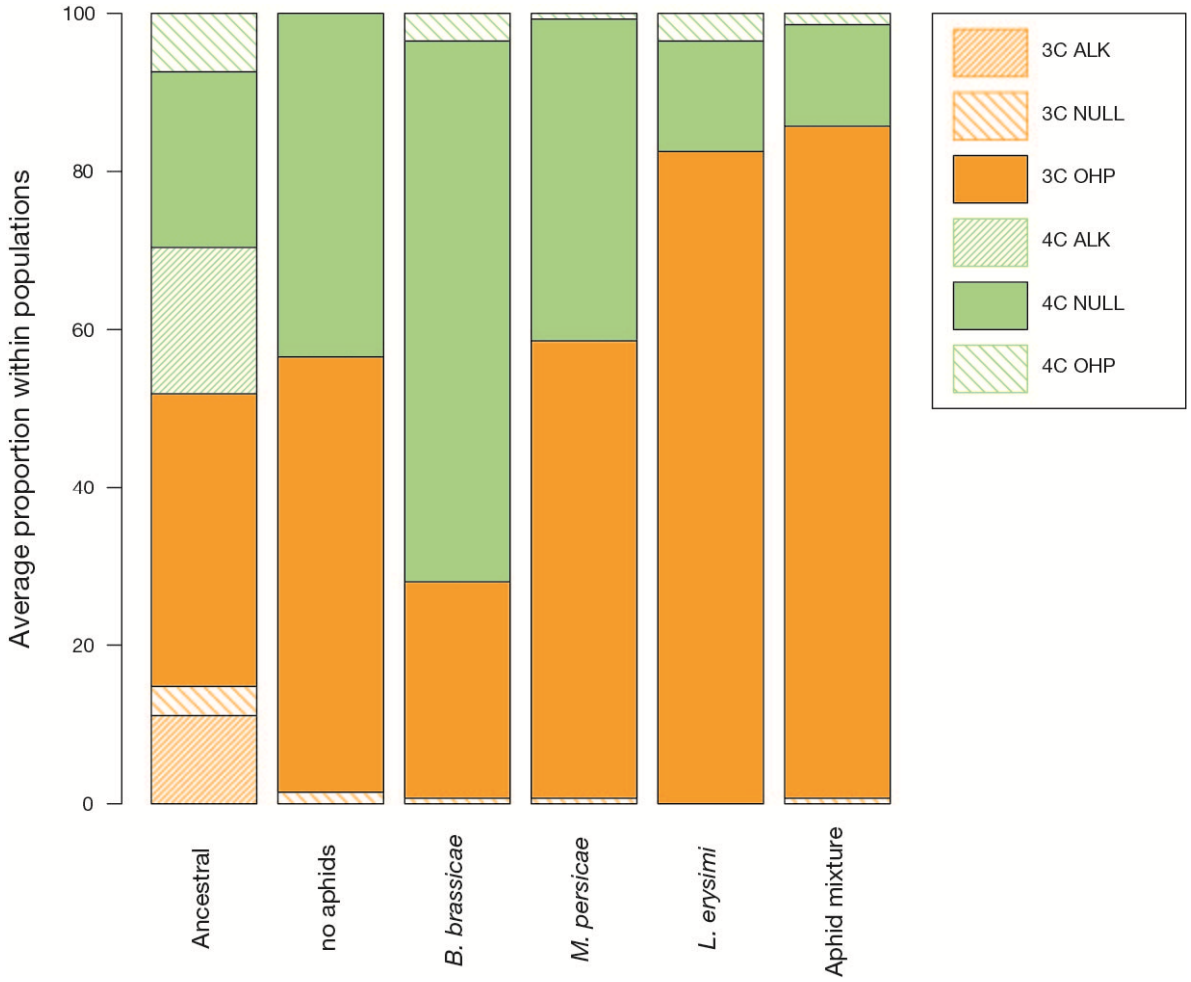

Figure S4. Mean frequencies of the six main chemotypes within plant populations before and after five generations of selection. Mean frequencies are based on HPLC analysis of 24 plants per population (144 plants per treatment). After selection, mainly the chemotypes *3C-OH* and *4C-NULL* remained in all treatments. Both alkenyl chemotypes were lost from all 30 populations, while *3C-NULL* (overall mean proportion: 0.71 %) and *4C-OH* (overall mean proportion: 1.82 %) remained at very low frequencies in all populations.

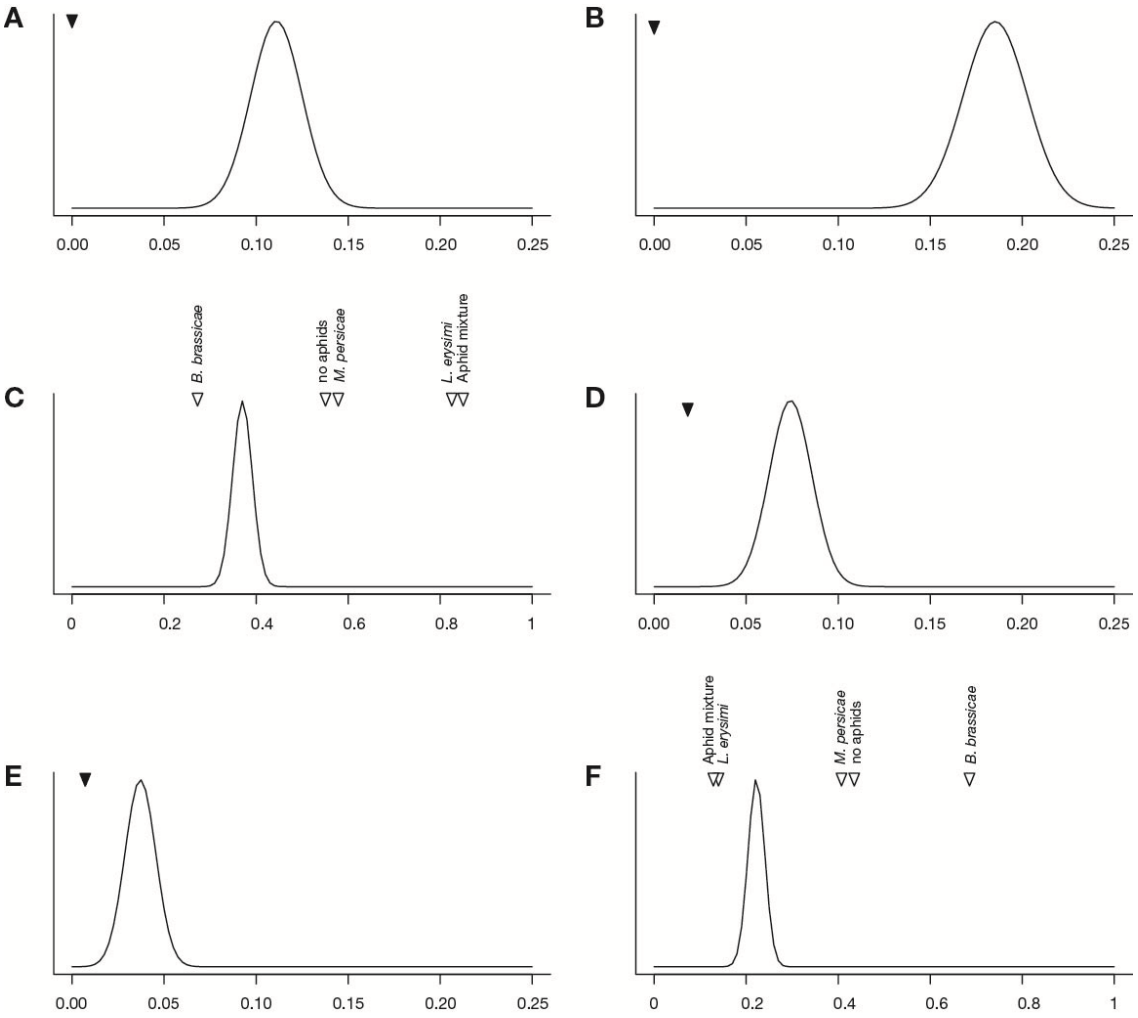

Figure S5. Distribution of expected chemotype frequencies. We generated distributions for the expected chemotype frequencies for **A)** *3C-ALK*, **B)** *4C-ALK*, **C)** *3C-OH*, **D)** *4C-OH*, **E)** *3C-NULL*, and **F)** *4C-NULL*, assuming no treatment differences and only random drift. This was achieved by randomly sampling genotypes based on their frequencies in previous generations only (in generation 1, each genotype had the same probability of being selected). The total population size was constrained to be the total number of seedlings observed in each population in each generation and we include the number of individuals actually genotyped as the final sample. Random sampling was repeated 10'000 times and chemotype frequencies were averaged across all 30 populations. For all observed chemotype frequencies, either the treatment-specific (empty triangles) or the overall average values (filled triangles) are shown, depending on significant treatment differences.

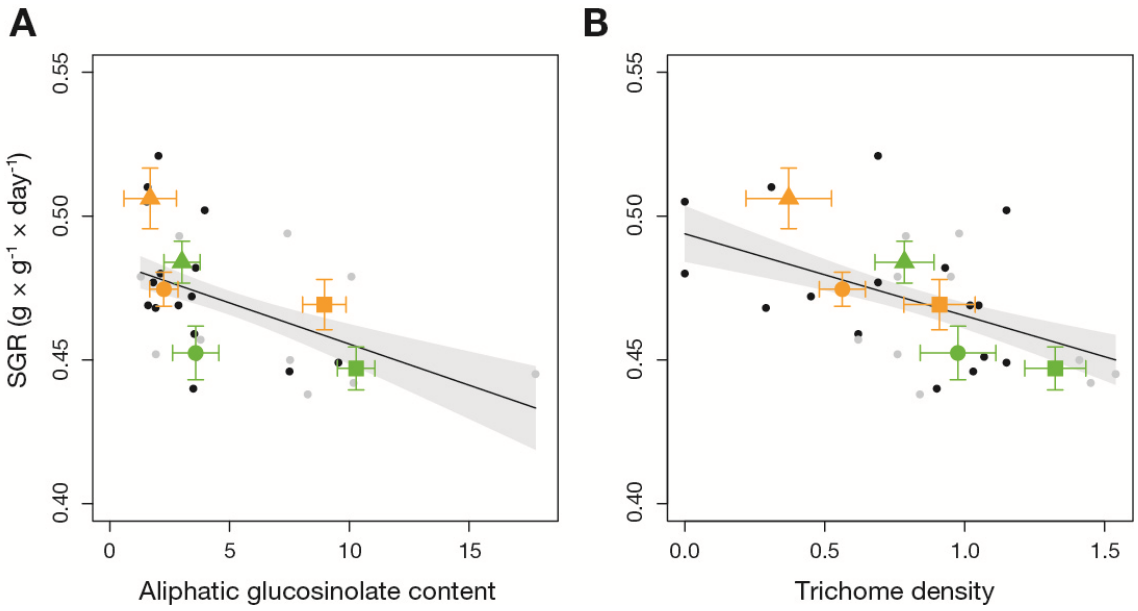

Figure S6. Trait values of the 27 ancestral genotypes. **A)** shows size-standardized growth rate (SGR) and aliphatic glucosinolate content, and **B)** SGR and trichome density. Solid lines are linear regressions with grey areas indicating ± 1 SEs. Genotypes extinct after five generations of selection are colored in grey. The chemotype means ± 1 SEs are overlaid, with symbol shape indicating the *GS-AOP* chemotype (square: *ALK*; circle: *OH*; triangle: *NULL*), and color indicating the *GS-ELONG* chemotype (orange: *3C*; green: *4C*).

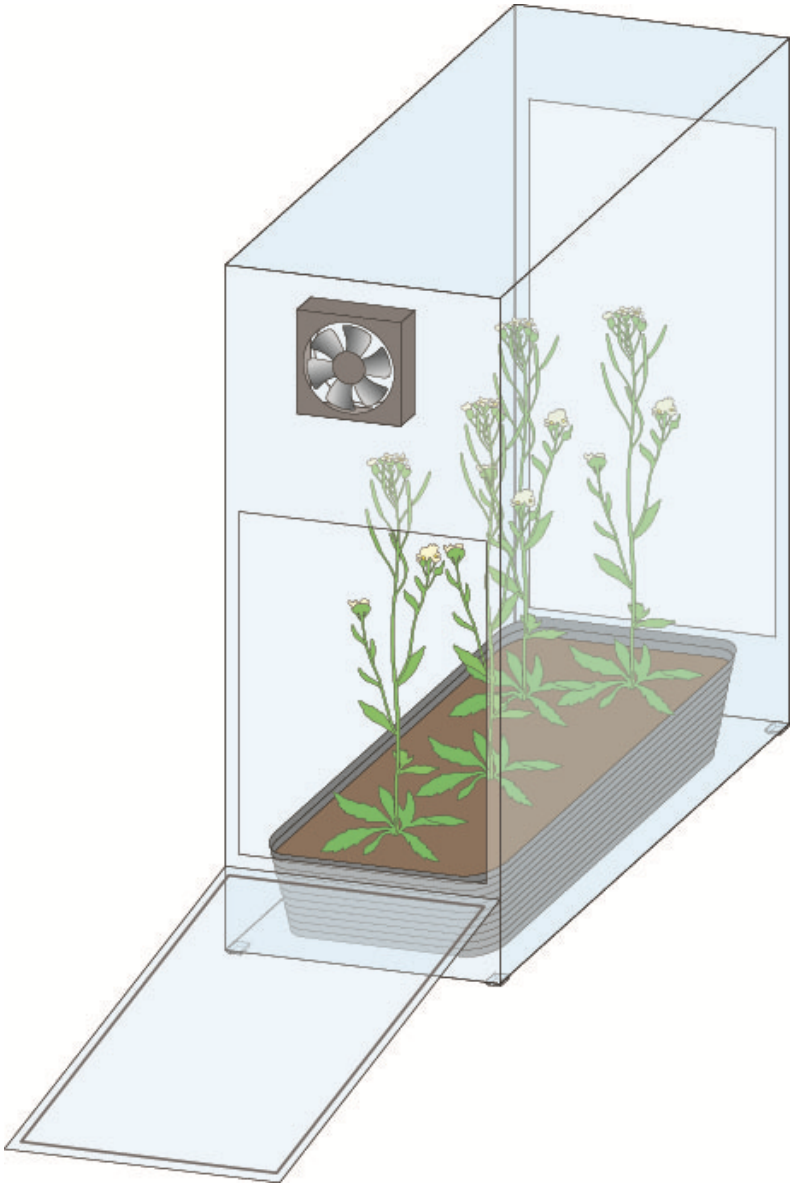

Figure S7. Schematic drawing of an experimental cage. Cages (20 x 36 x 50 cm w x l x h) were used to keep aphid treatments constrained to their respective plant populations in the selection experiment. Cages are made from 5 mm thick Plexiglass® with a square, netted hole in the back (18 x 30 cm) and a flap in the front of the cage for access. The flap is fitted with a rubber seal creating an insect-proof barrier when closed. Each cage is fitted with a ventilator, mounted on top of a circular netted hole (Ø 8 cm) on the front of the cage creating a linear airflow through the cage.

Table S1. List of the 27 genotypes used in the selection experiment. Bolting age and trichome density are mean values from approximately 20 plants. SGR is the average growth rate at a common reference size (mean size at 9 days after sowing). Glucosinolate content is based on the measurement of two plant samples per genotype, each sample consisting of six rosette leaves. All plants were grown in individual pots within a controlled climate chamber set to 18° C under a day/night cycle of 16h/8h.

Genotype	Country of Origin	Line ID	Bolting age [DAG]	Trichomes × $mm^{2\,-1}$	SGR [$g \times g^{-1} \times day^{-1}$]	Chemotype	Aliphatic glucosinolate content [$\mu mol \times g^{-1}$]	Indolic glucosinolate content [$\mu mol \times g^{-1}$]
cnt-1	United Kingdom	N1635	18.5 ± 1.1	1.45 ± 0.18	0.442	*4C-ALK*	10.150	1.184
col-0	United States	N1092	16.3 ± 0.7	1.15 ± 0.25	0.502	*4C-NULL*	3.903	1.155
cvi-0	Cape Verde Islands	N902	18.3 ± 1.0	1.54 ± 0.21	0.445	*4C-ALK*	17.684	0.476
di-1	France	N1108	14.1 ± 0.3	1.05 ± 0.18	0.469	*4C-OH*	1.588	1.169
est-0	Estonia	N1148	14.4 ± 0.7	0.00 ± 0.00	0.480	*3C-OH*	2.106	0.843
kin-0	United States	N1272	18.0 ± 1.7	0.84 ± 0.16	0.438	*4C-ALK*	8.259	0.973
kondara	Tadjikistan	N916	26.1 ± 4.0	0.95 ± 0.12	0.479	*3C-ALK*	10.151	0.966
krot-0	Germany	N3886	17.0 ± 0.8	0.90 ± 0.20	0.440	*3C-OH*	3.396	1.069
le-0	Netherlands	N1308	18.9 ± 1.0	0.62 ± 0.10	0.459	*4C-NULL*	3.490	0.623
ll-0	Spain	N1338	34.3 ± 0.6	0.62 ± 0.13	0.457	*4C-NULL*	3.764	0.628
lm-2	France	N1344	15.7 ± 0.7	1.07 ± 0.19	0.451	*4C-OH*	3.600	0.483
ma-0	Germany	N1356	14.8 ± 0.7	0.69 ± 0.14	0.521	*4C-NULL*	1.904	0.770
mt-0	Libya	N1380	16.1 ± 1.5	0.76 ± 0.12	0.479	*4C-NULL*	1.240	0.790
no-0	Germany	N1394	15.0 ± 1.0	0.69 ± 0.16	0.477	*3C-OH*	1.550	0.846
nw-0	Germany	N1408	16.4 ± 0.5	0.93 ± 0.12	0.482	*4C-NULL*	3.477	0.811
oy-0	Norway	N1643	17.7 ± 0.7	0.45 ± 0.10	0.472	*3C-OH*	3.318	0.776
pi-0	Austria	N1454	19.3 ± 1.0	0.76 ± 0.10	0.452	*3C-OH*	1.867	0.748
pog-0	Canada	N1476	17.9 ± 1.6	1.41 ± 0.17	0.450	*4C-ALK*	7.517	0.823
ru-0	Germany	N1496	16.8 ± 0.4	1.03 ± 0.11	0.446	*3C-ALK*	7.397	0.549
sap-0	Czech Republic	N1506	17.4 ± 1.1	0.29 ± 0.09	0.468	*3C-OH*	1.847	0.787
shakdara	Tadjikistan	N929	19.6 ± 2.2	1.15 ± 0.19	0.449	*4C-ALK*	9.540	0.903
st-0	Sweden	N1534	16.7 ± 1.5	0.79 ± 0.22	0.493	*3C-OH*	3.292	1.306
tsu-1	Japan	N1640	19.0 ± 3.5	0.56 ± 0.12	NA	*3C-OH*	2.568	1.572
wa-1	Poland	N1587	14.2 ± 0.4	0.31 ± 0.13	0.510	*3C-NULL*	1.746	0.866
wei-0	Switzerland	N3110	14.6 ± 0.7	1.02 ± 0.22	0.469	*3C-OH*	2.795	0.966
wil-3	Lithuania	N1598	15.3 ± 0.9	0.00 ± 0.00	0.505	*3C-OH*	1.519	0.994
ws-0	Russia	N1602	NA	0.98 ± 0.20	0.494	*3C-ALK*	7.395	0.557

Values are line means ± 1 SE

Table S2. Statistical tests for change in biomass and trichome densities. All tests are extracted from a linear model of log(Biomass) or a linear mixed-effects model of trichome density as response and treatment, generation, and the interaction as factorial explanatory variables. Values are *t*-tests on the treatment differences from the no-aphid treatment in generation 5.

	log(Biomass)		Trichome density	
M. persicae	$t = 4.16$	$p < 0.001$***	$t = 6.90$	$p < 0.001$***
B. brassicae	$t = 0.97$	$p = 0.333$	$t = 4.69$	$p < 0.001$***
L. erysimi	$t = 2.71$	$p = 0.008$**	$t = 7.36$	$p < 0.001$***
Aphid mixture	$t = 3.78$	$p < 0.001$***	$t = 6.96$	$p < 0.001$***

Table S3. List of primers used for genotyping of the set of 27 *A. thaliana* accessions.

SSLP Locus	SSLP Primer	Chromosome	PCR Primer Sequence	$MgCl_2$ concentration used in PCR reactions
nga111	NGA111F	1	TGTTTTTTAGGACAAATGGCG	1.5 mM MgCl2
	NGA111R		CTCCAGTTGGAAGCTAAAGGG	
ciw3	CIW3F	2	GAAACTCAATGAAATCCACTT	2.5 mM MgCl2
	CIW3R		TGAACTTGTTGTGAGCTTTGA	
nga72	NGA172_F	3	CATCCGAATGCCATTGTTC	2.5 mM MgCl2
	NGA172_R		AGCTGCTTCCTTATAGCGTCC	
nga6	NGA6_F	3	ATGGAGAAGCTTACACTGATC	1.0 mM MgCl2
	NGA6_R		TGGATTTCTTCCTCTCTTCAC	
ciw6	CIW6_F	4	CTCGTAGTGCACTTTCATCA	2.0 mM MgCl2
	CIW6_R		CACATGGTTAGGGAAACAATA	

```
# set the path for the working directory
setwd("c:/Documents/…")

# load general-purpose libraries
library(nlme)
library(lattice)
library(lme4)

##############################################################################
#######
## Spatial analysis: Chemotypes

## read in datafile
d1<-read.table("1226397s3.txt", header=T)

names(d1)
#Accession: Accession name
#Country: Country of collection
#Latitude: Coordinates of sampling location
#Longitude
#ID: ABRC accession number
#Elong: Allele at GS-ELONG
#AOP: Allele at GS-AOP

## transform ELONG data into a binary vector
d1$binom<-   as.numeric(d1$Elong=="3C")

## generalized linear model with a quasi-binomial error distribution
glm.1<-glm(binom ~   Longitude * Latitude  , family=quasibinomial, d1)
anova(glm.1, test="Chi")
# --> no support for an interaction term

## final model
glm.2<-glm(binom ~   Longitude + Latitude  , family=quasibinomial, d1)

## extract t-tests from summary table
summary(glm.2)


##############################################################################
#######
## Spatial analysis: Aphids

## load additional libraries
library(arm)

## read in datafile
d2<-read.table("1226397s4.txt", header=T)

names(d2)
#Country
#Trap
#Latitude
#Longitude
#Year: Year of record
#Brevicoryne: Number of B. brassicae captured
#Lipaphis: Number of L. erysimi captured
#Inoperative.days: Days per year trap was inoperative

## fit generalized linera mixed effects model using Trap as random effect
```

```
glme.1<-glmer(cbind(Lipaphis, Brevicoryne) ~  Longitude * Latitude + (1|
Trap), family=binomial,  data=d2)

## sampling the posterior distribution to summarize the variance
aphids<- sim(glme.1, 1000)

## calculating a p-value for the parameters)
apply(aphids@fixef, 2, function(x) 1-abs(sum(x>0)/1000-0.5)*2)
# --> no support for an interaction term

## final model
glme.2<-glmer(cbind(Lipaphis, Brevicoryne) ~  Longitude + Latitude + (1|
Trap), family=binomial,  data=d2)

## extract t-tests from summary table
summary(glme.2)


#############################################################################
#######
## Change in Biomass

## read in datafile
d3<-read.table("1226397s5.txt", header=T)

names(d3)
#Population:Experimental population
#Generation: Generation of the selection Experiment
#Treatment: Experimental treatment
#Seedlings: Number of germinated plants
#Sown: sown seed number
#Germination: Percent germination
#Biomass: Dried plant biomass at the end of a generation

## rearrange treatment so that aphids are tested against control
d3$Treatment<-relevel(d3$Treatment, ref="Control")

## linear model of log(biomass) as a function of treatment and Generation
lm.1<-lm(log(Biomass)~factor(Treatment)*factor(Generation), data=d3)

anova(lm.1)

## extract t-tests from summary table
summary(lm.1)

#############################################################################
#######
## Change in trichome density

## read in datafile
d4<-read.table("1226397s6.txt", header=T)

names(d4)
#Generation: Generation of the selection Experiment
#Population: Experimental population
#Plant: Identifier of plant measured
#Treatment: Experimental treatment
#Trichomes: density within a circle of 4mm diameter

## Exclude missing values
```

```
d4<-subset(d4, Trichomes!="NA")

## Treat generation as factor
d4$Generation<-as.factor(d4$Generation)

## Express trichomes per mm^2
d4$Hairs.std<-d4$Trichomes/(2^2*pi)

## Rearrange treatment so that aphids are tested against control
d4$Treatment<-relevel(d4$Treatment, ref="Control")

## Linear mixed effects model of trichome density as a function of
## treatment and generation. Population is treated as random effect, and
## a weights term is included to allow variance to differ in each
## generation
lme.1<-lme(Hairs.std~Generation*Treatment, data=d4, random= ~1
|Population, weights=varIdent(form=~1| Generation))

## extract t-tests from summary table
summary(lme.1)


##############################################################################
#######
## Simulations of random drift

## read in datafile
d3<-read.table("1226397s5.txt", header=T)

names(d3)
#Population: Experimental population
#Generation: Generation of the selection Experiment
#Treatment: Experimental treatment
#Seedlings: Number of germinated plants
#Sown: sown seed number
#Germination: Percent germination
#Biomass: Dried plant biomass at the end of a generation

## Treat generation as factor
d3$Generation<-as.factor(d3$Generation)

## Create a new data object, ordered by 1) Generation and 2) Treatment
sums.populations <- data.frame(aggregate(d3$Seedlings,
list(d3$Population, d3$Treatment, d3$Generation), sum, na.rm=T))


## Name columns
names(sums.populations) <- c("id", "treat", "gen", "seed")


#### Code for sampling
attach(sums.populations)

no.genotypes <- 27

## Vector containing numbers for the six chemotypes, repeated as many
## times as genotypes were present in the ancestral population
tr<-as.factor(c(rep(1,10),2,2,2,3,4,4,rep(5,5),rep(6,6)))

## Create empty dataframe to be filled with chemotype frequencies
```

```
chemotypes<-data.frame(Iteration=numeric(10000), "OH.3C"=numeric(10000),
"ALK.3C"=numeric(10000), "NUL.3C"=numeric(10000), "OH.4C"=numeric(10000),
"ALK.4C"=numeric(10000), "NUL.4C"=numeric(10000))

### Sampling loop
for (j in 1:10000){
## Print loop number
print(j)
if (.Platform$OS.type == "windows") flush.console()
Sys.sleep(1)


## Generate a matrix for the number of adults present in each generation

ob.seed <- matrix(c(seed[1:150], rep(24,30)), ncol= 30, nrow=6, byrow=T)
p.genos <- matrix(1/no.genotypes, ncol=27, nrow=30)

## Now pick gen.1 using equal probs per genotype
for (i in 1:6){
numbers <- rmultinomial(ob.seed[i,], p.genos)
p.genos <-  numbers/ob.seed[i,]
}

## add chemotype frequencies to empty dataframe in each loop
chemotypes[j,]<-c(j,as.numeric(tapply(colSums(numbers), tr, sum)))
}

detach(sums.populations)

#############################################################################
#######
## Genotype data

## read in datafile
d5<-read.table("1226397s7.txt", header=T)

names(d5)
#Sample: Plant sample identifier
#Population: Experimental population
#Treatment: Experimental treatment
#Accession: Genotype as determined by SSLP markers
#Chemotype: Ancestral chemotype of the identified genotype


#############################################################################
#######
## Chemotype data

## read in datafile
d6<-read.table("1226397s8.txt", header=T)

names(d6)

#Source: Either ancestral genotype or generation 5 population
#Treatment: Experimental treatment
#Chemotype: Chemotype based on chemical profile of the plant
#ELONG: Allele at GS-ELONG based on chemical profile of the plant
#AOP: Allele at GS-AOP based on chemical profile of the plant
#OH3: 3-hydroxypropyl
#MSO3: 3-methylsulfinylpropyl
```

```
#OH.S.Butenyl: 2(S)-hydroxy-3-butenyl
#MSO4: 4-methylsulfinylbutyl
#OH.R.Butenyl: 2(R)-hydroxy-3-butenyl
#Allyl
#MSO5: 5-methylsulfinylpentyl
#Butenyl
#MSO6: 6-methylsulfinylhexyl
#MT3: 3-methylthiopropyl
#MSO7: 7-methylsulfinylheptyl
#Pentenyl
#MT4: 4-methylthiobutyl
#MSO8: 8-methylsulfinyloctyl
#I3M: indolyl-3-methyl
#MOI3M4: 4-methoxy-indolyl-3-methyl
#MOI3M1: 1-methoxy-indolyl-3-methyl
#MT7: 7-methylthioheptyl
#MT8: 8-methylthiooctyl
#t.ali: total aliphatic glucosinolates
#t.ind: total indolic glucosinolates
#C3: total 3-carbon glucosinolates
#C4: total 4-carbon glucosinolates
#C7: total 7-carbon glucosinolates
#C8: total 8-carbon glucosinolates

## remove ancestral lines
d6<-subset(d6, Treatment!="Ancestral")
d6$Treatment<-factor(d6$Treatment)



### test for treatment effects on indole glucosinolates

## linear mixed effecs model of total indole glucosinolates as a function
## of the five treamtents
lme.2<-lme(t.ind~Treatment, random =~1| Source, data=d6)
anova(lme.2)
#--> no significant effect of treatment

## create variavle with '1' for aphid treatments and '0' for controls
d6$aphid<-as.numeric(d6$Treatment!="Control")

## linear mixed effecs model of total indole glucosinolates as a function
## of a two- level control/aphid variable
lme.3<-lme(t.ind~as.factor(aphid)   , random =~1| Source, data=d6)

anova(lme.3)
# aphid treatment is significant

summary(lme.3)


### Analysis of GS-ELONG proportions


## Complex loop to create a new dataframe with proportions of GS-ELONG
##for all 30 populations
d7<-data.frame(Population=numeric(30),
Treatment=numeric(30),Proportion.3C=numeric(30),
Proportion.4C=numeric(30))
```

```
for(i in 1:5){
for(j in 1:6){

n<-(i-1)*6+j

d7[n,1]<-
levels(factor(d6$Source[d6$Treatment==levels(d6$Treatment)[i]]))[j]
d7[n,2]<-  levels(d6$Treatment)[i]

tmp<-
tapply(d6$Chemotype[d6$Source==levels(factor(d6$Source[d6$Treatment==leve
ls(d6$Treatment)[i]]))[j]],
d6$ELONG[d6$Source==levels(factor(d6$Source[d6$Treatment==levels(d6$Treat
ment)[i]]))[j]], length)
tmp[is.na(tmp)] <- 0

d7[n,3]<-  as.numeric(tmp) [1]
d7[n,4]<-  as.numeric(tmp) [2]
}}

## Rearrange treatment so that aphids are tested against control
d7$Treatment<-as.factor(d7$Treatment)
d7$Treatment<-relevel(d7$Treatment, ref="Control")

## generalized linear model with binomial error structure
response<- cbind(d7$Proportion.3C, d7$Proportion.3C+d7$Proportion.4C)
glm.3<-glm(response ~ Treatment , family=binomial, data=d7)

## extract t-tests from summary table
summary(glm.3)
```