the plant journal

SEB Society for Experimental Biology

The Plant Journal (2014) doi: 10.1111/tpj.12447

Cardamine hirsuta: a versatile genetic system for comparative studies

Angela S. Hay[1,*], Bjorn Pieper[1], Elizabeth Cooke[2], Terezie Mandáková[3], Maria Cartolano[1], Alexander D. Tattersall[2], Raffaele D. Ioio[1], Simon J. McGowan[4], Michalis Barkoulas[2,†], Carla Galinha[2], Madlen I. Rast[1], Hugo Hofhuis[1], Christiane Then[5], Jörg Plieske[6], Martin Ganal[6], Richard Mott[7], Jaime F. Martinez-Garcia[5,8], Mark A. Carine[9], Robert W. Scotland[2], Xiangchao Gan[1], Dmitry A. Filatov[2], Martin A. Lysak[3] and Miltos Tsiantis[1,*]

[1]*Max Planck Institute for Plant Breeding Research, Carl-von-Linné-Weg 10, 50829 Köln, Germany,*

[2]*Department of Plant Sciences, University of Oxford, South Parks Road, Oxford OX1 3RB, UK,*

[3]*Plant Cytogenomics Research Group, Central European Institute of Technology (CEITEC), Masaryk University, Brno CZ-625 00, Czech Republic,*

[4]*Computational Biology Research Group, The Weatherall Institute of Molecular Medicine, University of Oxford, The John Radcliffe Hospital, Oxford OX3 9DS, UK,*

[5]*Centre for Research in Agricultural Genomics CSIC-IRTA-UAB-UB, Campus UAB, Edifici CRAG, Bellaterra (Cerdanyola del Vallés), 08193 Barcelona, Spain,*

[6]*TraitGenetics GmbH, Am Schwabeplan 1b, D-06466 Gatersleben, Germany,*

[7]*Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK,*

[8]*Institució Catalana de Recerca i Estudis Avançats, 08010 Barcelona, Spain, and*

[9]*Natural History Museum, Cromwell Road, London SW7 5BD, UK*

Received 6 December 2013; revised 14 January 2014; accepted 16 January 2014.

*For correspondence (e-mails tsiantis@mpipz.mpg.de and hay@mpipz.mpg.de).

[†]Present address: Imperial College London, South Kensington Campus, London SW7 2AZ, UK.

SUMMARY

A major goal in biology is to identify the genetic basis for phenotypic diversity. This goal underpins research in areas as diverse as evolutionary biology, plant breeding and human genetics. A limitation for this research is no longer the availability of sequence information but the development of functional genetic tools to understand the link between changes in sequence and phenotype. Here we describe *Cardamine hirsuta*, a close relative of the reference plant *Arabidopsis thaliana*, as an experimental system in which genetic and transgenic approaches can be deployed effectively for comparative studies. We present high-resolution genetic and cytogenetic maps for *C. hirsuta* and show that the genome structure of *C. hirsuta* closely resembles the eight chromosomes of the ancestral crucifer karyotype and provides a good reference point for comparative genome studies across the Brassicaceae. We compared morphological and physiological traits between *C. hirsuta* and *A. thaliana* and analysed natural variation in stamen number in which lateral stamen loss is a species characteristic of *C. hirsuta*. We constructed a set of recombinant inbred lines and detected eight quantitative trait loci that can explain stamen number variation in this population. We found clear phylogeographic structure to the genetic variation in *C. hirsuta*, thus providing a context within which to address questions about evolutionary changes that link genotype with phenotype and the environment.

Keywords: *Cardamine hirsuta*, *Arabidopsis thaliana*, Brassicaceae, emerging model organism, comparative development, genetic map.

INTRODUCTION

The pervasive use of model organisms in biological research is driven in the large part by the powerful genetic tools available for these species. However, a key challenge in biology is to identify the genes that have generated biological diversity, as model organisms alone allow us to study only a tiny fraction of this diversity. Sequencing the genetic material of both model and non-model species is becoming increasingly simple with current technologies and empowers our ability to identify sequences that differentiate one organism from another. Unfortunately, only a

fraction of this sequence diversity contributes to phenotypic variation, making it difficult to establish causal relationships between changes in genotype and changes in phenotype. Genetic tools are, therefore, essential to test the evolved functions of DNA sequences but very few organisms have been developed as good genetic systems.

Parallel genetic studies in closely related species that span a relatively short evolutionary timescale have proven to be a powerful strategy to identify genetic changes that underlie phenotypic differences that are of evolutionary significance (Gompel *et al.*, 2005; Marcellini and Simpson, 2006; Jeong *et al.*, 2008; Wittkopp *et al.*, 2009; Frankel *et al.*, 2011). This approach has been successful because it allows causal genetic differences to be identified within an otherwise broadly comparable genotype to phenotype landscape. *Cardamine hirsuta* has emerged recently as one such genetic system for comparative studies of development with its close relative *Arabidopsis thaliana*. The dating of divergence times within the Brassicaceae has been hindered by limited reliable fossil data, however recent age estimates place the divergence of the *C. hirsuta* and *A. thaliana* lineages to some time between 13 million years ago (Mya), when *A. thaliana* and *Arabidopsis lyrata* diverged, and 34–43 Mya when the split between *Arabidopsis* and the *Brassica* complex has been estimated (Beilstein *et al.*, 2008, 2010; Couvreur *et al.*, 2010). Previous work in *C. hirsuta* has utilized a suite of experimental tools to genetically dissect leaf development pathways in these two species in order to understand the genetic causes of leaf shape evolution (Hay and Tsiantis, 2006; Barkoulas *et al.*, 2008; Blein *et al.*, 2008). *Cardamine hirsuta* shares many of the desirable traits found in the model species *A. thaliana*: it is a diploid and self-compatible annual plant with an abundant seed set, an 8-week seed-to-seed generation time and a small rosette growth habit that is amenable to large-scale cultivation. The *C. hirsuta* genome is estimated to be 1.5 times that of *A. thaliana* (Johnston *et al.*, 2005) and, importantly, *C. hirsuta* can be efficiently transformed by floral dip using *Agrobacterium tumefaciens* (e.g. transformation efficiency: *A. thaliana* Col-0, 0.3%; *C. hirsuta* Ox, 0.1%). This experimental tractability allows large-scale genetic screens and cross-species tests of gene function to be performed, in addition to sophisticated molecular genetic approaches that use the multitude of transgenic tools available in *A. thaliana* to manipulate and visualise gene expression and generate mosaics of gene function.

Cardamine hirsuta is a cosmopolitan weed, commonly called hairy bittercress or popping cress for its explosive seed dispersal. It is considered to be native to Europe and introduced worldwide; it is easily found on sandy or rocky soils and as a common weed of gardens, nurseries and disturbed ground (Rich, 1991; Lihova and Marhold, 2006). It is characteristically a winter annual but summer annuals are also common (Rich, 1991). Here, we describe *C. hirsuta* as an experimental system with the functional tools to ask questions about the genetic causes of phenotypic evolution in an in-depth and unbiased way. We present a genetic map, genome structure and recombinant inbred lines (RILs) for *C. hirsuta* and show that the genetic variation in *C. hirsuta* has clear phylogeographic structure. We demonstrate the utility of these genetic resources by mapping eight quantitative trait loci (QTL) that control stamen number, and use a combination of RNAseq data from the RIL founder accessions and synteny with the *A. thaliana* genome to predict candidate genes in these QTL regions. Finally, we describe several differences in morphology and physiology between *C. hirsuta* and *A. thaliana* that provide opportunities for comparative study and propose that the genetic tools presented here will allow *C. hirsuta* to be adopted by a broad range of plant biologists for comparative research.

RESULTS

Genetic map

To facilitate genetic studies in *C. hirsuta*, we constructed a linkage map using a RIL population derived from a cross between the reference accession from Oxford, UK (Ox) and a polymorphic accession from Washington State, USA (Wa) (Figure 1). In total, 288 molecular markers were designed based on 242 SNPs and 46 microsatellites that were likely to provide genome coverage based on synteny with the ancestral crucifer karyotype (Schranz *et al.*, 2006). The entire RIL population, which consists of 195 independent lines, was genotyped with these 288 markers and genotyping information from 183 F8 RILs was used to build an initial linkage map. Using a logarithm of odds score threshold of 6.5 or higher, all markers mapped to eight linkage groups (LG) that represented the eight chromosomes of *C. hirsuta* (Table S1). Single markers were selected in which multiple markers clustered together and a stringent selection regime excluded markers for which the nearest neighbour fit indicated mapping errors. Isolated markers under strong segregation distortion or with unexpectedly high levels of heterozygosity were also excluded, as these suggested technical errors. The resulting genetic map consisted of 212 markers with a total length of 886.6 cM (average spacing of 4.6 cM) with 26 markers per LG on average (range: 22 markers for LG2 – 33 markers for LG6) and the largest distance between two adjacent markers being 19.6 cM (Figure 1 and Table S1). With the genome size of *C. hirsuta* estimated at 225-Mbp by flow cytometry (Johnston *et al.*, 2005), the relationship between physical and genetic distance is approximately 250-kb per cM. Chi-squared tests (χ^2) for observed versus expected allele frequencies at all 212 marker loci on the genetic map revealed that 29% were significantly distorted

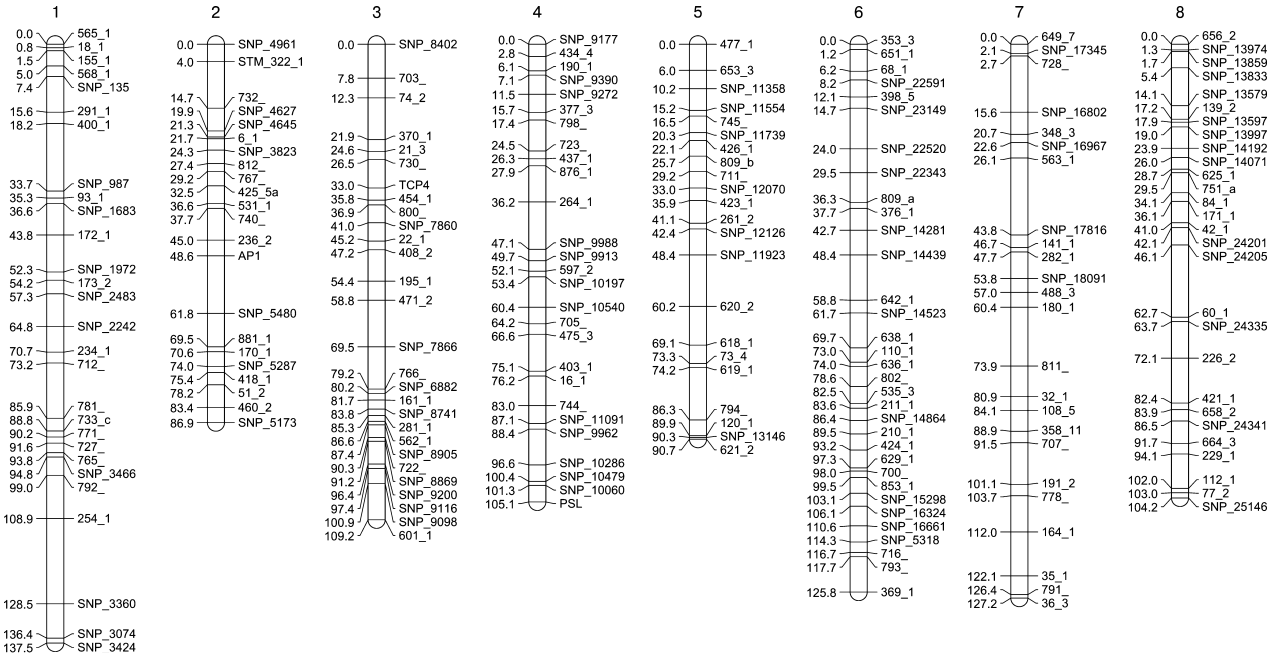

Figure 1. *Cardamine hirsuta* genetic map.
Eight linkage groups are numbered corresponding to the eight chromosomes in *C. hirsuta*. Horizontal bars along each chromosome indicate the positions of 212 molecular markers mapped in an Ox × Wa F8 recombinant inbred line (RIL) population. Marker positions in cM are shown to the left and the corresponding marker identifiers to the right of each linkage group.

at the level $\alpha = 0.05$ (Table S1 and Figure S1). This transmission distortion did not affect the quality of the RILs for QTL mapping (see Figure 6) and was similar to that reported in the genetic map of another crucifer species *Boechera stricta* (Schranz *et al.*, 2007).

Genome evolution

To investigate the evolution of the *C. hirsuta* genome and to inform our genetic map construction, we analysed the karyotype of *C. hirsuta* by comparative chromosome painting (CCP). We used chromosome-specific BAC contigs from *A. thaliana*, arranged to represent 24 ancestral genomic blocks (A–X) and eight chromosomes (AK1–AK8) of the putative ancestral crucifer karyotype (ACK; Schranz *et al.*, 2006) to paint meiotic chromosomes of the *C. hirsuta* reference accession Ox (Figure 2). We found complete colinearity for all ancestral genomic blocks between the eight chromosomes of *C. hirsuta* and the ACK, with the exception of a translocation between chromosomes 6 and 8 that places block V at the top of chromosome 6, and blocks O and P at the top of chromosome 8 (Figures 2 and S2; Mandakova *et al.*, 2013a). We used this result to inform our linkage mapping and found that genetic markers confirmed the presence of a translocation between ancestral chromosomes 6 and 8. To further investigate this translocation, we used fluorescence *in situ* hybridization in *C. hirsuta* with 40 chromosome-specific *C. hirsuta* BAC clones that corresponded to seven genomic blocks (O, P, Q, R, V, W, and X) of *C. hirsuta* chromosomes 6 and 8 (Figure 3). These BAC probes were selected by directly mapping markers designed from BAC end sequences on our *C. hirsuta* linkage map or by prediction of BACs that belonged to each ancestral genomic block by BLAST analysis using the *A. thaliana* genome. We found that BACs that belonged to genomic block V hybridized to the same chromosome as BACs that belonged to blocks Q and R (Figure 3). Conversely, BACs that belonged to genomic blocks O and P hybridized to the same chromosome as BACs that belonged to blocks W and X (Figure 3). Thus, both our comparative approach using *A. thaliana* BACs and our direct approach using *C. hirsuta* BACs mapped the location of a translocation between the top arms of *C. hirsuta* chromosomes 6 and 8.

Comparative morphology of *Cardamine hirsuta* and *Arabidopsis thaliana*

To adopt *C. hirsuta* as a model system, it was important to gauge if comparative studies with *A. thaliana* were likely to provide general insights into phenotypic evolution. A comparative survey revealed an abundance of developmental traits that had diversified between the two species (Figure 4 and Table 1), and suggested that *C. hirsuta* would provide a tractable system in which to study the genetic basis for morphological diversity. In this survey, we compared the reference accessions *A. thaliana* Col-0 and *C. hirsuta* Ox and did not address trait variation within species. Leaf shape differs dramatically: *A. thaliana* leaves are simple with an entire margin while *C. hirsuta* leaves

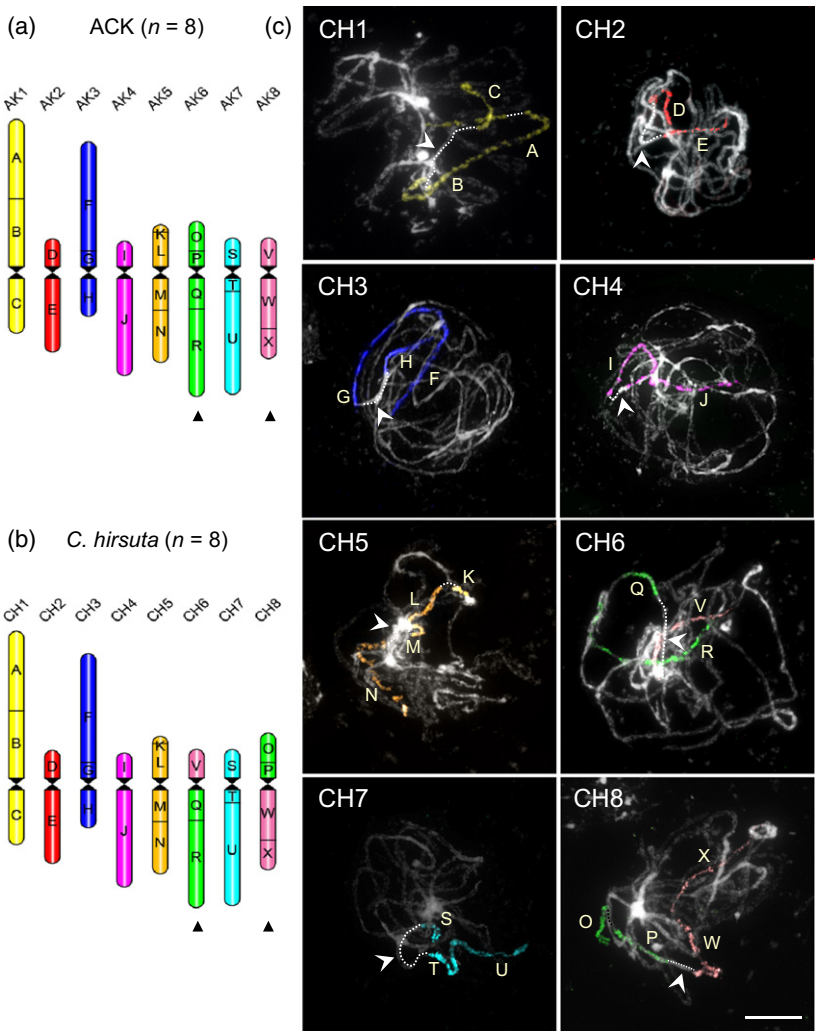

Figure 2. Comparative genome structure of *Cardamine hirsuta*.
(a) Ancestral crucifer karyotype (ACK) with eight chromosomes (AK1–AK8, colour coded) and 24 genomic blocks (A–X).
(b) Karyotype of *C. hirsuta* with eight chromosomes (CH1–CH8) and 24 genomic blocks (A–X). Note the origin of CH6 and CH8 via a whole-arm translocation between ancestral chromosomes AK6 and AK8 (chromosomes marked by arrowheads).
(c) The eight *C. hirsuta* chromosomes identified by comparative chromosome painting (CCP) using *Arabidopsis thaliana* BAC contigs arranged according to the 24 genomic blocks on eight chromosomes of ACK. Individual pachytene bivalents CH1 to CH8 were pseudo-coloured according to the *C. hirsuta* cytogenetic map (b). Arrowheads indicate centromeric regions. Scale bar: 10 μm.

are pinnately compound with leaflets arranged on either side of a central rachis (Figure 4a,b). Previous studies have started to reveal molecular mechanisms through which these leaf morphologies evolved and highlighted the importance of *cis*-regulatory evolution in this process (Hay and Tsiantis, 2006; Barkoulas *et al.*, 2008; Blein *et al.*, 2008). Unicellular trichomes are present on the leaves of both species and branched in *A. thaliana*, while *C. hirsuta* has unbranched trichomes (Figure 4c,d). In addition, trichomes are found on the stem of *A. thaliana* but not *C. hirsuta.* Shoot architecture differs such that *C. hirsuta* is more branched than *A. thaliana* and this architecture is likely to be governed by differential activity of axillary meristems (Figure 4e,f). In our standard growth conditions, more than twice the number of secondary branches were found on *C. hirsuta* branches when compared with *A. thaliana*; accessory branches were found in the axils of both rosette and cauline leaves in *C. hirsuta* but not *A. thaliana* (Figure 4e′,f′ and Table 1), this finding suggested that *C. hirsuta* has an increased capacity to initiate accessory shoot meristems. Stamen number is a reliable morphological character used to distinguish *C. hirsuta* from its close relatives as it has four medial stamens and lacks the two lateral stamens present in *A. thaliana* (Figure 4g,h). Petal number is also reduced in *C. hirsuta* compared with the four petals found in *A. thaliana* (Figure 4i–j′), and varies from 0 to 4 between individual flowers on a single plant. Petal shape differs between the two species, with *C. hirsuta* petals being more spoon-shaped with a shorter claw (Figure 4k,l and Table 1), a difference that possibly contributes to the more upright stance of *C. hirsuta* petals in the flower (Figure 4i,j).

Explosive pod shatter allows ballistic dispersal of *C. hirsuta* seeds over a distance of several metres while pod shatter in *A. thaliana* is non-explosive (Vaughn *et al.*, 2011). The dehiscent siliques of both species are very similar, but while *A. thaliana* valves remain flat at dehiscence, *C. hirsuta* valves curl rapidly, transferring stored mechanical energy to launch its seeds (Figure 4m,n). *Cardamine hirsuta* seeds are more than five times larger

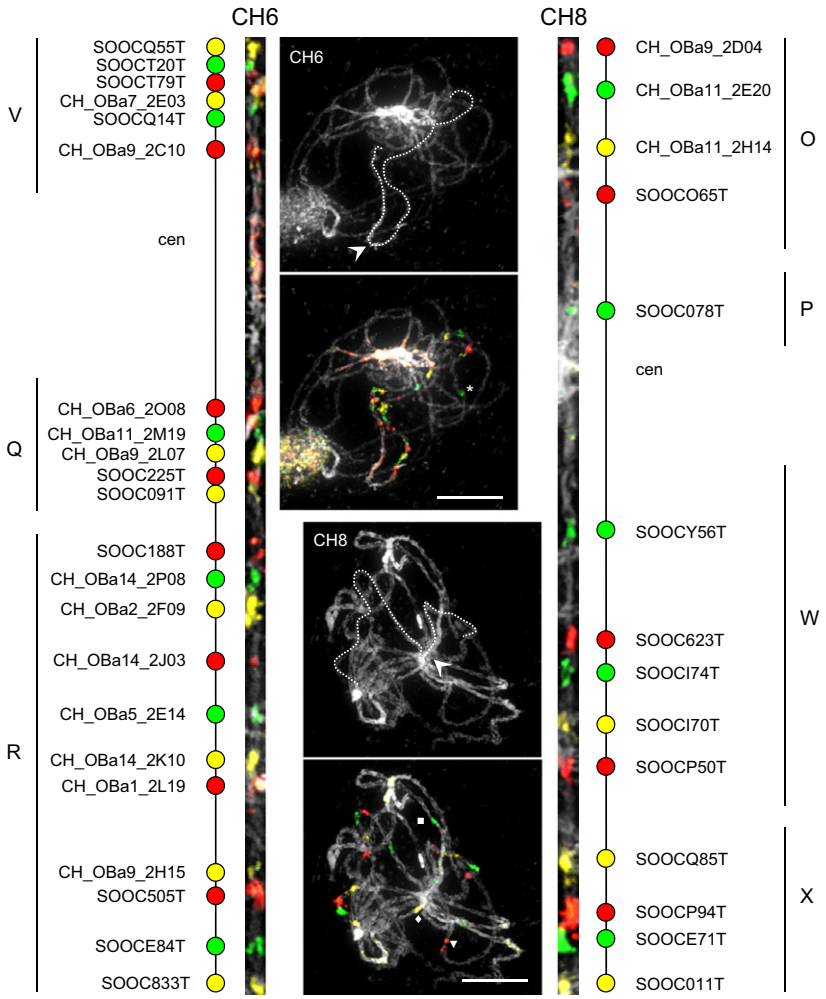

Figure 3. Whole-arm chromosome translocation in *Cardamine hirsuta*.
Differently labelled *C. hirsuta* BAC clones covering chromosomes CH6 and CH8 were hybridized to pachytene complements in *C. hirsuta*. BAC clones on CH6 have colinear position to genomic blocks V, Q and R, whereas BAC clones on CH8 correspond to blocks O, P, W and X. Centromeric regions (cen) are indicated by arrowheads and four BAC clones used in this experiment hybridized to chromosomes other than CH6 or CH8 (SOOCN07T *, SOOCK29T ■, SOOCP45T ♦, SOOC083T ▼). Scale bars: 10 μm.

than *A. thaliana* seeds and are flattened longitudinally to form a disk shape (Figure 4o,p and Table 1). The seed surface differs between species due to the central columella of the cells in the outer seed coat that forms raised bumps in *A. thaliana* and indented dimples in *C. hirsuta* (Figure 4o,p). This surface morphology may reduce drag on *C. hirsuta* seeds during ballistic dispersal in the same way that dimples reduce the aerodynamic drag on a golf ball (Davies, 1949). Embryo positioning in the seed also differs between species, with the embryo bent forwards in *C. hirsuta* seeds such that the flat surface of its two cotyledons sits parallel to the flat surfaces of the seed (Figure 4p), while in *A. thaliana* the embryo bends sideways. Embryo patterning is very similar between the two species and the larger size of *C. hirsuta* seeds compared with *A. thaliana* is reflected in larger embryo size throughout embryogenesis (Figure 4q,r).

The comparison of *C. hirsuta* and *A. thaliana* at the seedling stage highlights a number of morphological differences between species. *Cardamine hirsuta* seedlings are larger with a longer hypocotyl and adventitious roots that initiate at the junction of the hypocotyl and primary root and that are not found in *A. thaliana* (Figure 4s,t and Table 1). The primary root of *C. hirsuta* is thicker than *A. thaliana* due, firstly, to an extra cortical cell layer and, secondly, to additional cell files in each layer such that in *C. hirsuta* the cortex and endodermis comprise 12 cell files while in *A. thaliana* they comprise eight (Figure 4u,v and Table 1). The presence of starch granules is used as a marker for columella cell fate differentiation in the *A. thaliana* root because granules are reliably absent from stem cell initials and the quiescent centre, whereas all of these cell types accumulated starch granules in the *C. hirsuta* root (Figure 4w,x). This result emphasises the potential pitfalls of relying on characteristics that may be *A. thaliana*-specific for comparative development.

We also found differences in the physiological responses of *C. hirsuta* and *A. thaliana* seedlings to environmental and hormonal cues. *Arabidopsis thaliana* seedlings responded to simulated shade (i.e. white light enriched with far red) by hypocotyl elongation relative to growth under white light as reported previously (Figure 5a; Roig-

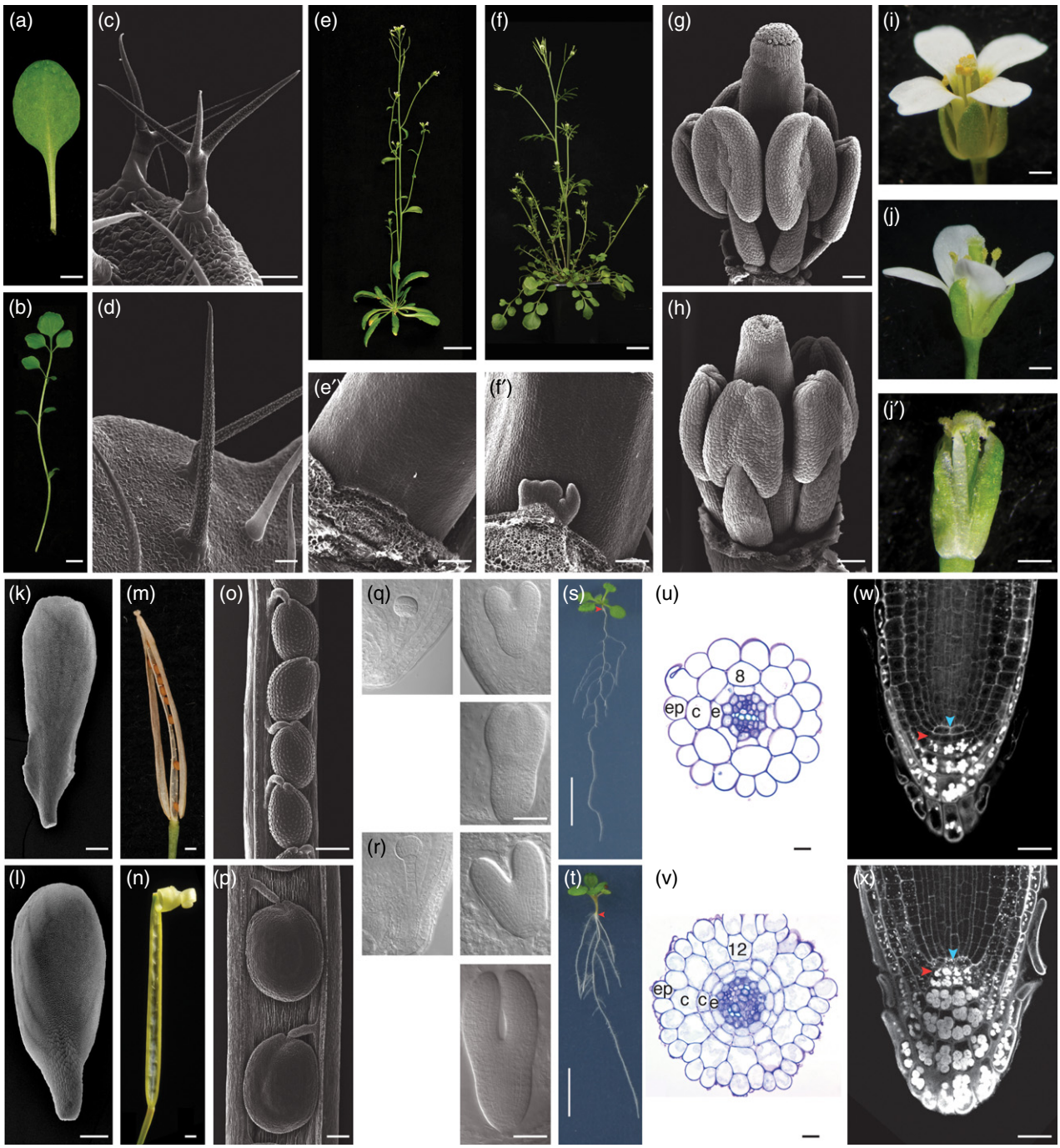

Figure 4. Comparative morphology.
Arabidopsis thaliana (a, c, e, e′, g, i, k, m, o, q, s, u, w) and *Cardamine hirsuta* (b, d, f, f′, h, j, j′, l, n, p, r, t, v, x) differ in leaf shape (a, b), leaf trichome branching (c, d), shoot branching (e, f), the presence of accessory meristems in the axil between branch and leaf (dissected off) (e′, f′), stamen number (g, h), petal number (i, j, j′), petal shape (k, l), fruit opening (m, n), seed size and shape (o, p), embryo size: globular, heart and torpedo stage embryos shown (q, r), hypocotyl length and adventitious root formation at the hypocotyl–root junction, indicated by the red arrow (s, t), number of cortex cell layers and cortical and endodermal cell files in the root (u, v) and starch granule accumulation in the quiescent centre (blue arrow) and columella stem cells (red arrow) (w, x). ep: epidermis, c: cortex, e: endodermis. Scale bars: 0.5 cm (a, b), 50 μm (c, d, e′, f′, u, v), 100 μm (g, h), 1 mm (i–j′, m, n), 2 cm (e, f), 250 μm (k, l, o, p), 1 cm (s, t), 25 μm (q, r, w, x).

Villanova *et al.*, 2007), while *C. hirsuta* hypocotyls were almost unresponsive to the same treatment (Figure 5a). In contrast, exogenous gibberellin (GA_3) stimulated hypocotyl elongation in both species (Figure 5b), this finding indicated that *C. hirsuta* responds to some but not all stimuli that trigger hypocotyl elongation in *A. thaliana*. We also

Table 1 Morphological diversity between *A. thaliana* and *C. hirsuta*

Morphology	*A. thaliana* (Col-0)	*C. hirsuta* (Ox)
Leaf shape	Simple	Compound
Leaf trichomes	Branched	Unbranched
Stem trichomes	Present	Absent
Shoot branches[a]		
Primary rosette	6.3 ± 0.3	10.0 ± 1.6
Secondary rosette	2.9 ± 0.3[f]	7.0 ± 0.3[f]
Accessory rosette	0	2
Primary cauline	8.0 ± 0.7	8.1 ± 0.4
Secondary cauline	1.6 ± 0.1 [f]	4.4 ± 0.3 [f]
Accessory cauline	0	5
Stamen number	6	4
Petal number	4	0–4
Petal width/petal length ratio[b]	0.29 ± 0.003[f]	0.40 ± 0.006[f]
Petal claw/petal length ratio[b]	0.60 ± 0.004[f]	0.33 ± 0.006[f]
Pod shatter	Non-explosive	Explosive
Seed shape	Oblong	Disk-shaped
Seed surface	Raised bumps	Indented dimples
Seed weight (mg)[c]	0.020 ± 0.000[f]	0.146 ± 0.005[f]
Seed area (mm^2)[d]	0.102 ± 0.001[f]	0.531 ± 0.009[f]
Mature embryo position in seed	Bent sideways	Bent forwards
Hypocotyl length (mm)[e]	1.35 ± 0.03[f]	3.82 ± 0.16[f]
Adventitious roots	Absent	Present
Primary root		
Cortex and endodermis cell files	8	12
Cortex cell layers	1	2
Starch granules	Columella cells	Columella cells, quiescent centre, columella stem cells

[a]Average number of primary and secondary branches per plant, total number of accessory branches observed (see Fig. S7 for description).
[b]N = 35 (*A. thaliana*), 32 (*C. hirsuta*).
[c]N = 16 407 (*A. thaliana*), 8630 (*C. hirsuta*).
[d]N = 56 (*A. thaliana*), 63 (*C. hirsuta*).
[e]N = 30 (*A. thaliana*), 33 (*C. hirsuta*); N = 6 (*A. thaliana*), 7 (*C. hirsuta*).
[f]Student's *t*-test *p*-value < 0.001.

compared the growth response of *C. hirsuta* and *A. thaliana* seedlings to the native auxin, indole-3-acetic acid (IAA), by assaying root meristem size. *Cardamine hirsuta* root growth showed the same bimodal dose response to increasing concentrations of auxin as reported previously for *A. thaliana*: 0.1 nM IAA stimulated growth and higher concentrations repressed growth (Figure 5c; Evans *et al.*, 1994). However, this auxin response was enhanced in *C. hirsuta*, as 1 nM IAA was sufficient to cause a significant reduction in *C. hirsuta*, but not *A. thaliana*, root meristem size (Student's *t*-test: *C. hirsuta* $P < 0.001$; *A. thaliana* $P = 0.3$; Figure 5c). To compare the dynamics of the auxin response between species, we quantified the stimulation of root growth in response to 0.1 nM IAA by measuring root meristem cell number at 9 and 24 h post-treatment. At 9 h post-treatment, cell number had increased 1.2-fold in *C. hirsuta* but was unchanged in *A. thaliana*, while by 24 h, cell number had increased in both species, this difference indicated a more rapid auxin response in *C. hirsuta* (Figure 5d). To compare the spatial dynamics of auxin response in the root tip between species, we analysed the expression of the auxin activity sensor *DR5::VENUS* in response to treatment with 10 nM IAA. We observed that *DR5::VENUS* accumulated in the stele and lateral root cap within 40 min of IAA treatment in *C. hirsuta* but not in *A. thaliana* (Figure 5e,f), this difference indicated a more rapid auxin response in these tissues in *C. hirsuta*.

QTL analysis of stamen number variation in *Cardamine hirsuta*

To investigate morphological variation at the intra-specific scale we analysed stamen number. The presence of four stamens is a derived trait in *C. hirsuta* and a reliable morphological characteristic for species identification, as most species within Brassicaceae bear six stamens (Rich, 1991). To understand whether this trait showed variation within *C. hirsuta*, we scored average stamen number in a species-wide survey of 40 *C. hirsuta* accessions grown under standard greenhouse conditions (Table S2). We found that average stamen number ranged from 4.00 to 5.05, and reflected variation in lateral stamen number (Figure 6a). The broad-sense heritability (H^2) of this trait was 0.85, and indicated that much of this phenotypic variation can be attributed to genetic variation.

To investigate the genetic basis of stamen number variation we used the RIL population, described in this paper, for QTL analysis. Average stamen numbers in the two founder accessions of the RIL population were significantly different (Ox: 4.24 ± 0.027 standard error of the mean (SEM) and Wa: 4.76 ± 0.1 SEM, Student's *t*-test $P < 0.001$). The RIL population had a mean average stamen number of 4.50, and showed transgressive variation with mean average stamen number per RIL ranging from 4.00 to 5.27 (Figure 6b). Therefore, these RILs captured the full extent of phenotypic variation found species-wide, despite the fact that the founder accessions did not represent phenotypic extremes. A considerable proportion of this variation had a genetic basis ($H^2 = 0.55$) and we detected eight QTL, numbered *SN1* through *SN8*, at which allelic variation in the RIL population had significant effects on average stamen number (Figure 6c and Table 2). A final additive QTL model with these eight QTL explained 47.8% of the total phenotypic variance.

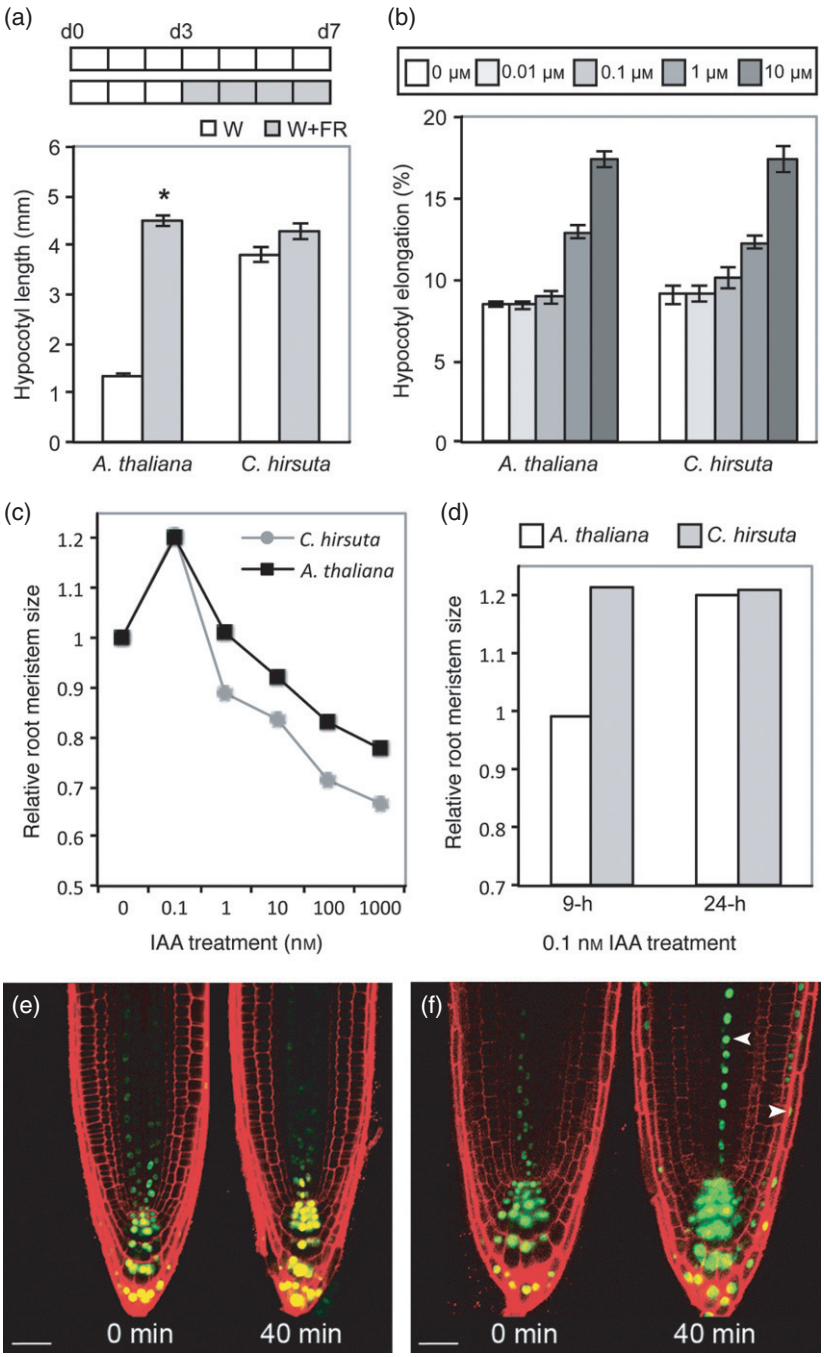

Figure 5. Comparative physiology.
(a, b) Hypocotyl elongation response of *Arabidopsis thaliana* and *C. hirsuta* seedlings to simulated shade (a) and gibberellin (GA_3) (b). (a) Hypocotyl length after 3 days (d3) grown in white light (W) and retained in W (white bars) or transferred to white/far red (W + FR, grey bars) until day 7 (d7); *Student's *t*-test $P < 0.01$. (b) Hypocotyl elongation, shown as percentage of etiolated hypocotyl length, in response to GA_3 dose (shaded bars refer to GA_3 concentrations in upper panel).
(c–f) Auxin sensitivity of *A. thaliana* and *C. hirsuta* roots. (c) Root meristem size, shown as a ratio between treated:untreated samples, in response to indole-3-acetic acid (IAA) dose. (d) Root meristem size, shown as a ratio between treated:untreated samples, 9 and 24 h after 0.1 nM IAA. *DR5::VENUS* expression (green) at 0 and 40 min after 10 nM IAA treatment in *A. thaliana* (e) and *C. hirsuta* (f), arrowheads indicate expression in the stele and lateral root cap. Cellular outlines are stained with propidium iodide (red) and all roots are imaged with the same confocal laser scanning microscope settings. Scale bars: 25 μm. Error bars indicate standard error of the mean (SEM).

The two most significant QTL, *SN2* and *SN8*, with −log10(P) values of 9.65 and 9.54 respectively, also had the strongest allelic effects of +0.18 stamen each for homozygous substitutions of Ox alleles with Wa alleles. Together, these QTL accounted for nearly 40% of the explained variance. Only for *SN5* and *SN7* did the Wa allele reduce stamen number and their combined effects were much smaller than the sum of the remaining QTL effects that increased stamen number. Therefore, higher stamen number in the Wa accession compared with Ox could be explained in terms of QTL effects that increased stamen number. The least significant QTL was *SN1* with a −log10(P) value of 3.48 and an allelic effect of 0.11 stamen, which explained 3% of the total phenotypic variance (Table 2). However, *SN1* was found to be a component of a significant two-way interaction with *SN5* when all pairwise combinations of the main effect QTL were tested. A closer inspection of this interaction revealed that the allelic effects at *SN1* were stronger with Wa alleles present at *SN5* and that the allelic effects of *SN5* were fully condi-

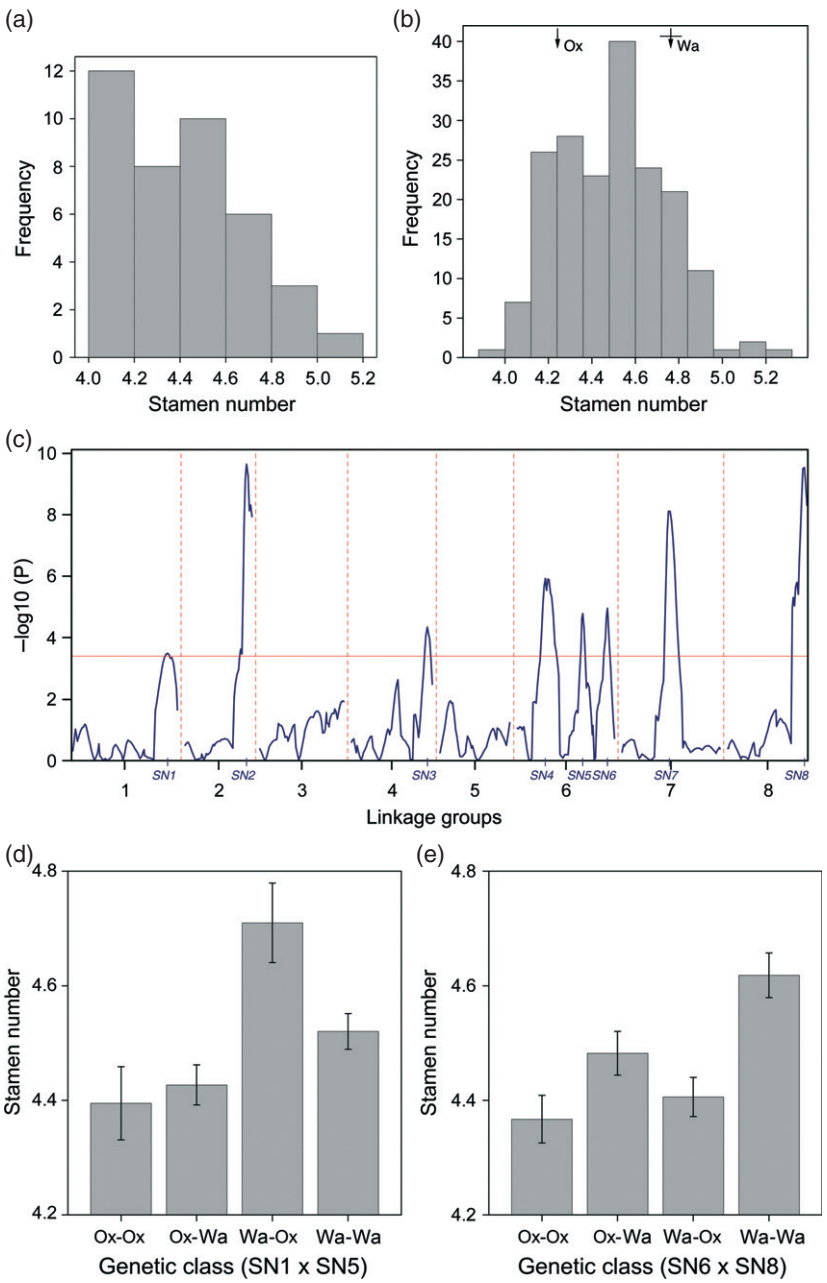

Figure 6. Quantitative trait loci (QTL) analysis of stamen number variation in *Cardamine hirsuta*.
(a, b) Phenotypic distribution of mean stamen number in (a) 40 different *C. hirsuta* accessions and (b) Ox × Wa RIL population, with accession and recombinant inbred line (RIL) frequencies on the y-axes respectively. Arrows and horizontal bars indicate average stamen number and standard error of the mean (SEM), respectively, for the founder accessions Ox and Wa.
(c) Genome-wide composite interval mapping scan for stamen number QTL (SN1–SN8) with the log transformed *P*-value (−log10(P)) plotted against genetic position of each linkage group. Horizontal line indicates genome-wide significance threshold allowing an error rate of 5%.
(d, e) Mean stamen number of the four possible homozygous genetic classes for the two significant two-way interactions that were detected: (d) SN1 × SN5, (e) SN6 × SN8. Bar height shows the mean stamen number of the allelic combination of the two components of each interaction respectively, shown on the *x*-axis, and error bars indicate SEM.

tional on Wa alleles being present at *SN1* (Figure 6d). One other significant two-way interaction was detected between *SN6* and *SN8*. At either of these loci the allelic effects were of a larger magnitude when Wa alleles were present at the other locus (Figure 6e). A final epistatic QTL model that included all additive QTL and the two interactions explained 49.5% of the total phenotypic variance. Therefore, the genetic basis of stamen number variation in this population could be attributed almost entirely to the detected effects of eight QTL.

Prediction of candidate genes in stamen QTL regions

To gain insight into the eight genetic intervals that contained stamen number QTL, we analysed RNAseq data from the two parental genotypes used in QTL mapping, Ox and Wa. We isolated RNA from shoot tissue following floral induction for two biological replicates of each parental genotype and used Illumina sequencing to generate 6.86-Gbp of sequence per sample. We made two predictions about the allelic differences that candidate genes may show between parental genotypes: first, genes may be differentially expressed; and, second, genes may contain non-synonymous nucleotide substitutions. To identify

Table 2 Characteristics of stamen number quantitative trait loci (QTL)

QTL	LG	Position (cM)	−log10 (P)	Effect	σ^2 Explained
SN1	1	124.6	3.5	0.11	3.0
SN2	2	79.9	9.6	0.18	10.3
SN3	4	98.5	4.3	0.11	4.5
SN4	6	36.3	5.9	0.13	6.4
SN5	6	85.0	4.8	−0.15	5.6
SN6	6	116.7	5.0	0.12	5.1
SN7	7	60.4	8.1	−0.16	9.3
SN8	8	100.0	9.5	0.18	11.2

−log10(P): log transformed *P*-value; Effect: allelic effects of the QTL for the homozygous substitution of Ox alleles with Wa alleles; σ^2 explained: variance explained by each QTL expressed as percentage of the total phenotypic variance.

Table 3 Prediction of candidate genes in stamen quantitative trait loci (QTL) regions

	Differentially expressed genes				Non-synonymous SNPs		
	Total		Stamen-expressed				
QTL	Up	Down	Up	Down	Total	STOP variants	Stamen-expressed genes
SN1	7	13	1	0	80	0	2
SN2	23	13	2	2	64	1	5
SN3	18	4	4	0	55	0	7
SN4	25	22	2	3	191	1	10
SN5	8	8	1	2	63	1	3
SN6	13	4	1	1	71	0	3
SN7	12	9	0	0	67	2	1
SN8	20	6	2	1	55	2	0
Total	126	79	13	9	646	7	31

candidate genes based on these predictions, we mapped *C. hirsuta* RNAseq reads to the *A. thaliana* reference genome (TAIR10, 27 416 protein coding genes), which gave 21 070 alignments. We found that 874 genes were differentially expressed between *C. hirsuta* Ox and Wa accessions with a false discovery rate (FDR) of 5% (Table S3). We also found 3138 non-synonymous single nucleotide polymorphisms (SNP) in expressed genes between these two *C. hirsuta* accessions, including 14 SNPs that introduced STOP codons and three SNPs that removed STOP codons (Table S4).

To analyse the subset of floral-expressed *C. hirsuta* genes located in stamen QTL regions, we identified syntenic regions of the *A. thaliana* genome that corresponded to 2-LOD intervals for the eight stamen QTL in *C. hirsuta*, and found 4339 genes. Of these genes, 126 were significantly up-regulated, 79 down-regulated, and 455 contained non-synonymous SNPs, including five genes with STOP codon variants (Table 3). These genes represented candidates for each QTL based on differential expression or amino acid differences between the parents. To further refine this gene list, we used previously published genome-wide datasets from *A. thaliana* (Wellmer *et al.*, 2004; Alves-Ferreira *et al.*, 2007) to identify 295 putative stamen-expressed genes in our eight stamen QTL regions. By these criteria we predicted stamen expression for 22 differentially expressed genes and 31 non-synonymous SNPs that were found in a stamen QTL region (Tables 3, S5 and S6). These candidates will inform future fine-mapping experiments aimed at identifying the genes that underlie these stamen number QTL in *C. hirsuta*.

Comparative transcriptomics

To investigate evolutionary processes that may have influenced the *C. hirsuta* transcriptome, we compared nucleotide divergence with related mustard species. We assembled *de novo* a floral transcriptome from 454 generated sequence reads and used the contigs as a BLAST query against *A. thaliana* coding sequences to produce 12 072 alignments with *A. thaliana* unigenes. Over 60% of these *C. hirsuta* contigs covered the full length of the corresponding *A. thaliana* coding sequences and over 90% were assigned at least one gene ontology (GO) term (Figures S3 and S4). As such, we assigned putative identification to a considerable proportion of the discovered *C. hirsuta* floral transcripts. Homologous genes were identified from *A. lyrata* and *Brassica rapa* with a reciprocal best BLAST hit approach; alignments of transcriptome sequences from *A. thaliana*, *C. hirsuta, A. lyrata* and *B. rapa* were created with MUSCLE software (Edgar, 2004) to give a dataset of 5603 homologous genes. *Cardamine hirsuta* and *A. thaliana* shared 96.3% amino acid sequence identity ($K_a = 0.037 \pm 0.024$) and were less diverged than either species was from *B. rapa* (see K_s values in Table 4).

To compare whether selective pressures differed between these four species we used a phylogeny-based analysis of K_a/K_s ratios (synonymous to non-synonymous substitution rates). The species-specific K_a/K_s ratios averaged across all 5603 genes were similar for different species ($K_a/K_s = 0.14$). However, we found that 121 genes showed significantly lower K_a/K_s ratios in *C. hirsuta* and *A. thaliana* (0.0630 ± 0.1274 and 0.086 ± 0.1414, respectively), when compared with the other two species (likelihood ratio test [LRT]; $P < 0.00001$), and reflected differences in the strength and/or type of selection that acted upon these genes in the two selfing species compared with the two out-crossing species.

We further analysed this dataset of 5603 homologous genes for the presence of positive selection by comparing the fit to data of nested likelihood models allowing and not allowing for adaptive selection (Yang, 2007). A signal of positive selection was detected in 263 *C. hirsuta* genes

Table 4 Pair-wise synonymous (K_s) and non-synonymous (K_a) divergence between four crucifer species. K_s and K_a are below and above diagonal, respectively

	Arabidopsis thaliana		*Arabidopsis lyrata*		*Cardamine hirsuta*		*Brassica rapa*	
K_s/K_a	Mean	SD	Mean	SD	Mean	SD	Mean	SD
A. thaliana	–	–	0.015	0.013	0.037	0.024	0.064	0.034
A. lyrata	0.131	0.043	–	–	0.037	0.024	0.064	0.034
C. hirsuta	0.283	0.076	0.275	0.077	–	–	0.068	0.036
B. rapa	0.435	0.141	0.425	0.139	0.448	0.140	–	–

SD, standard deviation.

(LRT, $P < 0.05$), and after Sidak correction for multiple tests (Abdi, 2007) four genes remained significant: At2g27810.1, nucleobase-ascorbate transporter 12; At3g51150.1, ATP binding microtubule motor family protein; At4 g23640.1, potassium transporter; At4g32130.1, pollen-specific carbohydrate binding protein. Additional taxa would considerably enhance the sensitivity of this method to identify positively selected genes as candidates for functional analyses (Anisimova *et al.*, 2001; Yang and dos Reis, 2011).

Phylogeography and population structure

To understand if the genetic variation found in *C. hirsuta* is structured geographically we analysed the phylogeography of *C. hirsuta*. Overall, 163 accessions from across the putative native range of *C. hirsuta* and 15 accessions from its introduced range were sequenced for the chloroplast *ndhF-rpl32* spacer, to produce an aligned matrix of 662 bp, with length variable homopolymers excluded from the alignment. In total, 34 haplotypes were defined and the relationships between them investigated using statistical parsimony (Figure 7). Each non-singleton haplotype that occupied an internal node in the resulting network, together with its terminal derivatives, was defined as a haplotype group. Using this approach seven groups were defined (turquoise, blue, orange, pink, maroon, green and red; Figure 7a). Each group was restricted to a particular geographic area, that was broadly centred on an area of species endemism (Figure 7b).

A major split, supported by 10 mutations, is apparent between the turquoise haplotype group, which is endemic to the high mountains of East Africa, and all other haplotypes. The blue haplotype group has a distribution that is centred on the Atlantic fringe of Europe, and that extends into neighbouring areas and the islands of the North Atlantic; the orange haplotype group is restricted to the Atlas Mountains, Cyprus and the Zagros Mountains. The pink and maroon haplotype groups are primarily centred on the western Irano-Turanian floristic sub-region and the Mediterranean basin, with a few scattered individuals in East Africa, Romania, the northern edge of the Alps and the British Isles. The two most abundant haplotype groups (green and red) are also the most widespread; occurring all across Europe and in the case of the 'green group' also in south-west Asia. The 15 accessions from areas where *C. hirsuta* is considered introduced (USA, Japan, New

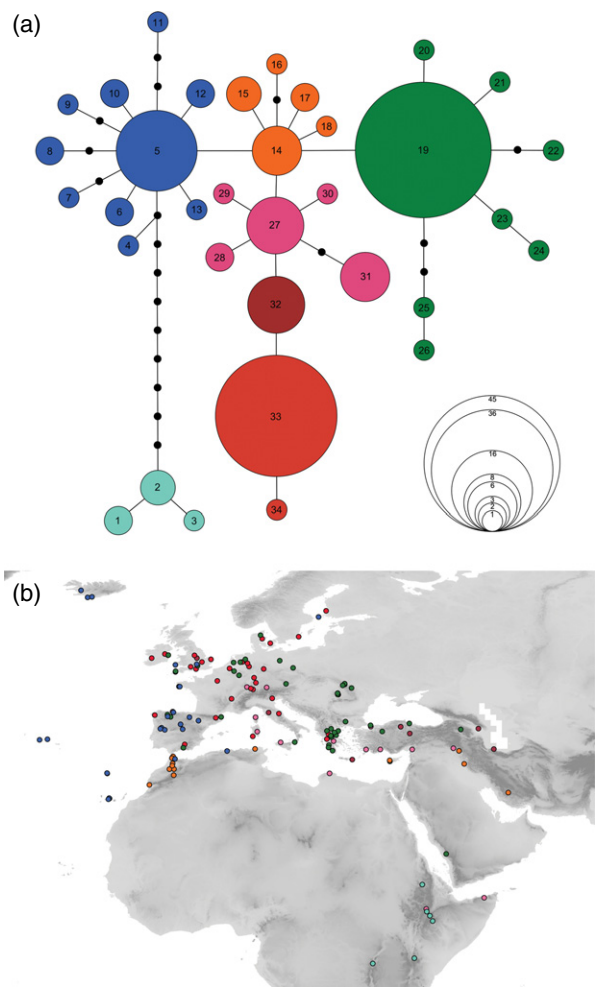

Figure 7. Phylogeography of *Cardamine hirsuta*.
(a) Statistical parsimony network of *C. hirsuta* ndhF-rpl32 haplotypes. Each circle represents a unique haplotype, with circle size proportional to the haplotype frequency (given by key in lower right corner). Connecting lines represent single mutations. Black dots indicate intermediate inferred haplotypes, either not sampled or extinct. Numbers within each circle indicate haplotype identity (1–34) and colour indicates haplogroup (seven in total).
(b) Map of accession collection localities in the native range of *C. hirsuta* in Eurasia with colours corresponding to those on the haplotype network.

Zealand and Australia) have no unique haplotypes and all have haplotypes that are common in Eurasia (haplotype numbers 5, 19, 27, 32, 33), a finding that is consistent with recent introduction.

To assess the distribution of genetic variation and population structure at the local scale we conducted nuclear SNP genotyping in four populations around Oxford (with eight to 11 individuals genotyped; Table S2). The overall level of polymorphism was comparable with that in *A. thaliana* populations (Schmid *et al.*, 2006; Kim *et al.*, 2007; Pico *et al.*, 2008) and inbreeding coefficient was very high ($F_{is} = 0.92$, $P < 0.0001$) as expected for a predominantly self-fertilising species. We genotyped 84 SNPs, of which 44 were polymorphic in the entire dataset; in each of the local populations the number of polymorphic SNPs varied from two to 28. Analysis of molecular variance (AMOVA; Excoffier *et al.*, 1992) revealed that 67.4% of variation was between the four local populations, 30% of variation was between individuals within the populations, and only 2.6% was within individuals (heterozygosity). This strong differentiation between closely neighbouring local populations was highly significant ($F_{st} = 0.67$, $P < 0.0001$). Extending the geographical scale to include two French populations (with three and 10 individuals genotyped) supported this pattern, as population subdivision at the local scale (between populations) was highly significant ($F_{st} = 0.581$, $P < 0.0001$). Furthermore, using the same set of 84 SNPs we found that average gene diversity over loci was only marginally higher for a global sample of 44 accessions (Table S2) (0.248 ± 0.1226), compared with that in the Oxford sample (0.197 ± 0.0981) (Tajima, 1983). These findings showed strong differentiation between populations, whether the scale was local or global; a larger scale experiment, in terms of markers and sampling, will give a more detailed view of *C. hirsuta* population structure.

DISCUSSION

Our understanding of the genetic causes of phenotypic evolution is hindered by the relative paucity of experimental platforms in which genetic tools can be utilized to study unambiguously the evolution of gene function in a genome-wide and unbiased fashion. Our data highlight the utility of *C. hirsuta* as a genetic system for such comparative studies. We showed that genetic diversity in natural *C. hirsuta* populations was accessible and structured geographically. We constructed a RIL population and genetic map and used these tools to identify eight QTL that explained the genetic basis of stamen number variation between RIL founder accessions. We found that there was considerable diversity in morphology and physiology between *C. hirsuta* and *A. thaliana* despite these species sharing over 96% amino acid identity. This diversity extended the potential reach of *C. hirsuta* beyond previous studies of leaf shape to allow general insights into phenotypic evolution (Hay and Tsiantis, 2006; Barkoulas *et al.*, 2008; Blein *et al.*, 2008).

The genome structure of *C. hirsuta*, with eight chromosomes and 24 conserved genomic blocks, is similar to the ACK and the extant genomes of *A. lyrata* and *Capsella rubella* (Hu *et al.*, 2011; Slotte *et al.*, 2013). As such, *C. hirsuta* provides a simple reference point for comparative genome studies across the Brassicaceae in which $n = 8$ is the most common and apparently ancestral base chromosome number in the family (Lysak *et al.*, 2006). Therefore, as a genetic system in the Brassicaceae, *C. hirsuta* offers a more ancestral counterpart to the five reshuffled chromosomes in *A. thaliana*. The tools presented here to analyse gene function in *C. hirsuta* are critical for the understanding of phenotypic consequences of changes observed in genes and genomes in the context of whole genome sequence information currently being generated for species in the Brassicaceae (e.g. Haudry *et al.*, 2013).

The whole-arm reciprocal translocation between ancestral chromosomes AK6 and AK8 that we describe in the reference Ox accession is characteristic for *C. hirsuta* as we found this translocation in six accessions for which we have constructed linkage maps and it was also reported recently using CCP (Mandakova *et al.*, 2013a,b). The same translocation has been found in other *Cardamine* species and in *Leavenworthia alabamica*, placing its origin ancestral to the divergence of these closely related genera in the tribe Cardamineae (Haudry *et al.*, 2013; Mandakova *et al.*, 2013a,b). The use of *C. hirsuta* BACs for fluorescence *in situ* hybridization in this study is an important validation of these previous findings from CCP with *A. thaliana* BAC probes, and suggests that CCP analysis in a greater sample of the Cardamineae tribe will clarify the origin of this translocation further. At the scale of a single *C. hirsuta* BAC, we found high microsynteny with *A. thaliana* and *A. lyrata*, this finding facilitated use of the well annotated *A. thaliana* genome as a comparative genomics tool, however colinearity was slightly disrupted in *C. hirsuta* compared with the two *Arabidopsis* species (Figure S5), and illustrated the different scales at which evolutionary remodelling of the *C. hirsuta* genomic landscape occurred: both sequence-scale and chromosome-scale rearrangements.

Stamen number is a morphological characteristic that is commonly used to identify *C. hirsuta* (Rich, 1991); we mapped eight QTL in a RIL population that determined variation in this trait. The RIL founder accessions did not represent extremes from the species range for this trait but transgressive segregation in the RILs captured the full extent of trait variation that we characterized species-wide. Thus, the eight QTL identified here considerably enhanced our understanding of the genetic basis of stamen number variation and indicated that the RILs and genetic map provide a good resource for natural variation studies. We used RIL founder transcriptomes to predict 53 candidate genes

for *C. hirsuta* stamen number QTL that included transcription factors, hormone and cell wall modifying enzymes (Tables S5 and S6). This RNAseq method to characterize QTL regions can be applied to any species in which a closely related reference sequence is available, and is a useful guide for subsequent fine-mapping experiments to identify the causal variants that underlie QTL.

The tools available in *C. hirsuta* can also be used to investigate the genetic basis of adaptation to environmental change. An example of environmental control of stamen number in *Cardamine* is found in the cleistogamous species *Cardamine kokaiensis*: chilling induces cleistogamy, such that flowers obligately self-pollinate, and these flowers show lateral stamen loss (Morinaga *et al.*, 2008). We compared the chilling-responsive genes that were previously reported in *C. kokaiensis* with candidate genes in our stamen QTL regions and found no overlap between these datasets. However, the use of *C. hirsuta* RILs to study the genetic basis of stamen number variation in response to the environment may help to understand genotype by environment interactions for this trait, and such findings may be relevant to the adaptive mating strategy of *C. kokaiensis*. Variation in lateral stamen number has also been reported in several accessions of *A. thaliana*, which typically has two lateral stamens (Muller, 1961; Smyth *et al.*, 1990). The comparison of the genetic architecture of stamen number variation between *A. thaliana* and *C. hirsuta* would help to understand the extent to which allelic variation at similar or different loci underlie intra-specific and inter-specific trait variation. Extension of this type of genetic analysis of phenotypic variation at multiple scales (inter-specific and intra-specific variation) throughout the Brassicaceae will reveal general versus particular principles of trait evolution at different evolutionary scales.

Cardamine hirsuta is a common, easily dispersed weed and yet clear geographic structuring of genetic variation is apparent in its native range. An understanding of this phylogeography provides an important framework for natural variation studies: it informs accession sampling, highlights examples of local adaptation and provides a context within which to understand phenotypic evolution in *C. hirsuta*. The phylogeographic patterns that we observed here are broadly concordant with studies in *A. thaliana* (Beck *et al.*, 2008), *Arabis alpina* (Koch *et al.*, 2006; Ansell *et al.*, 2011) and *Microthlaspi perfoliatum* (Koch and Bernhardt, 2004) that have been interpreted in the context of Pleistocene glacial dynamics with repeated glacial cycles restricting taxa to refugial areas, between which gene flow was low enough to enable populations to diverge; by drift and/or local adaptation (Petit *et al.*, 2003; Hewitt, 2004). The overlap of haplotype groups that we found in *C. hirsuta* is likely to be due to post-glacial range expansion from refugial areas, resulting in areas of secondary admixture (Hofreiter *et al.*, 2004), plus contemporary anthropogenic dispersal.

Our results have shown that some *C. hirsuta* haplotype groups are widespread whilst others are restricted to well characterized biogeographic regions such as the high mountains of East Africa and the Atlantic Fringe. This pattern of geographic structuring is not seen in *A. thaliana* (compare with Beck *et al.*, 2008), although direct comparisons must be cautionary given that the geographic sampling, the number of accessions sampled and the loci used are different. However, when we expanded the geographic sampling of *A. thaliana* by including *Atmyb2* and *PISTILLATA* sequences amplified from herbarium specimens from Ethiopia and Morocco with those from Beck *et al.* (2008), we found only two new haplotypes that differed from described haplotypes by one and two additional polymorphisms (Figure S6). Therefore, this additional geographic sampling did not alter considerably the previously described phylogeography of *A. thaliana* (Beck *et al.*, 2008). A more detailed understanding of the phylogeography of *C. hirsuta*, using additional accessions and genome-wide data, will shed light on these differences in the phylogeographies of *C. hirsuta* and *A. thaliana* despite their apparently similar geographic ranges, ecologies and mating systems.

The use of comparative studies in *C. hirsuta* and *A. thaliana* provided opportunities to understand both the evolutionary and ecological context of phenotypic diversity and may inform us about how the morphological and physiological differences that we identified between these species (Figures 4 and 5) relate to ecological differences. For example, explosive pod shatter, which is found in *C. hirsuta* but not *A. thaliana*, is a seed dispersal strategy that is likely to contribute to the success of *C. hirsuta* as a nursery weed and invasive plant (Yatsu *et al.*, 2003; Fain *et al.*, 2005). Shade avoidance is another trait that differs between *C. hirsuta* and *A. thaliana* seedlings and can play a major role in plant community dynamics. *Arabidopsis thaliana* seedlings perceive the shade cast by competing plants as a change in light quality and, like many shade-intolerant plants, respond by hypocotyl elongation (Martinez-Garcia *et al.*, 2010). *Cardamine hirsuta* seedlings, on the other hand, were unresponsive to shade. Future studies could investigate the full suite of shade avoidance responses in *C. hirsuta* and further address the difference in hypocotyl response between *C. hirsuta* and *A. thaliana* at the level of community responses, competition and co-existence in response to a specific environmental cue. In addition, previous studies have suggested that there is likely to be diversity in secondary metabolites such as glucosinolates between *C. hirsuta* and *A. thaliana*, which are important components of plant–microbe, plant–insect and plant–plant interactions in members of the Brassicaceae (Windsor *et al.*, 2005). Therefore, future work could address the

genetic basis of the adaptive morphologies that allow these species to exploit different ecological niches. Furthermore, *Cardamine* is a diverse and highly speciose genus, in which knowledge from the comparison of *C. hirsuta* and *A. thaliana*, coupled with NGS approaches, offer further opportunities to dissect the genetic basis of phenotypic diversity. In summary, we have provided a combination of genetic tools, workflows and reference information that will allow *C. hirsuta* to be used as an experimental system for comparative plant biology.

EXPERIMENTAL PROCEDURES

Short read sequence data from this article can be found in the NCBI Sequence Read Archive under accession number SRR882054. Full experimental procedures can be found in Method S1, which includes methods for plant cultivation, transformation and analysis of physiology and morphology (Figure S7), BAC library, RIL and genetic map construction, CCP and BAC fluorescence *in situ* hybridization (FISH) (Table S7), QTL, phylogeography and high throughput sequence analysis (Figures S8–S11).

ACKNOWLEDGEMENTS

We thank D. Bailey for advice; K. Marhold, S. West, M. Abraham, M. Thines and ABRC for materials; the Wellcome Trust Centre for the Human Genetics Oxford Genomics Centre and the Liverpool Centre for Genomic Research for sequencing; D. Lightfoot, D. Kudrna and R. Wing for BAC library construction; and J. Baker for photography. This work was supported by Biotechnology and Biological Sciences Research Council grants BB/D010977/1 and BB/F012934/1 to M.T., BB/H01313X/1 to A.H., Deutsche Forschungsgemeinschaft 'Adaptomics' grant TS 229/1-1 to M.T. and A.H., Max Planck Society core grant to M.T., Human Frontier Science Program grant to A.H., Natural Environment Research Council NE/E00489X/1 to D.F., Czech Science Foundation grant P501/10/1014 to M.L., Spanish Ministerio de Economía y Competitividad – FEDER grant BIO2011-23489 to J.M. and a Genetics Society Heredity Fieldwork grant to E.C. A.H. was supported by the Max Planck Society W2 Minerva programme and a Royal Society University Research Fellowship, and M.T. by a Royal Society Wolfson Merit award, Cluster of Excellence on Plant Sciences, and the Gatsby Charitable Foundation.

CONFLICT OF INTEREST

J.P. and M.G. are TraitGenetics GmbH employees; a company that offers molecular marker analysis services. TraitGenetics has no rights in any of the results presented in this paper.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Transmission distortion in the *Cardamine hirsuta* Ox × Wa RIL population.

Figure S2. Structure of *Cardamine hirsuta* chromosomes CH6 and CH8 revealed by CCP using *Arabidopsis thaliana* BAC contigs.

Figure S3. Proportion of coding region (CDS) length in *Arabidopsis thaliana* genes covered by cDNA sequence of *Cardamine hirsuta* homologues.

Figure S4. Distribution of the number of gene ontology (GO) terms assigned to *Cardamine hirsuta* cDNA contigs.

Figure S5. Microsynteny between *Cardamine hirsuta*, *Arabidopsis thaliana* and *Arabidopsis lyrata*.

Figure S6. Phylogeography of *Arabidopsis thaliana.* Modified from (Beck *et al.*, 2008).

Figure S7. Shoot branching scoring diagram.

Figure S8. Differential gene expression determined from RNAseq reads of RIL founder accessions using fold change and read count to compute P-values.

Figure S9. Distribution of *Cardamine hirsuta* cDNA contig lengths.

Figure S10. Proportion of annotated *Cardamine hirsuta* cDNAs is lower for shorter contigs.

Figure S11. The distribution of average pairwise differences between coding regions in *Cardamine hirsuta* and *Arabidopsis thaliana* shown separately for the 1st, 2nd and 3rd codon positions.

Table S1. Information for all molecular markers on the Ox × Wa F8 genetic map.

Table S2. Accession list of the plant material included in the phylogeography study with detailed source information and ndhF-rpl32 haplotype numbers.

Table S3. *Cardamine hirsuta* genes that are differentially expressed between RIL founder accessions.

Table S4. Non-synonymous SNP variants in *Cardamine hirsuta* RIL founder accessions, Ox and Wa.

Table S5. Differentially expressed genes between RIL founder accessions with predicted stamen expression in stamen number QTL regions.

Table S6. Non-synonymous SNP variants in *Cardamine hirsuta* RIL founder accessions, Ox and Wa, found in stamen number QTL regions in predicted stamen-expressed genes.

Table S7. *Cardamine hirsuta* BAC probes used for FISH.

Method S1. Full experimental procedures.

REFERENCES

Abdi, H. (2007) Bonferroni and Šidák corrections for multiple comparisons. In *Encyclopedia of Measurement and Statistics* (Salkind, N.J. ed.) Thousand Oaks, CA: SAGE Publications, pp. 1–9.

Alves-Ferreira, M., Wellmer, F., Banhara, A., Kumar, V., Riechmann, J.L. and Meyerowitz, E.M. (2007) Global expression profiling applied to the analysis of Arabidopsis stamen development. *Plant Physiol.* **145**, 747–762.

Anisimova, M., Bielawski, J.P. and Yang, Z. (2001) Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Mol. Biol. Evol.* **18**, 1585–1592.

Ansell, S.W., Stenoien, H.K., Grundmann, M., Russell, S.J., Koch, M.A., Schneider, H. and Vogel, J.C. (2011) The importance of Anatolian mountains as the cradle of global diversity in *Arabis alpina*, a key arctic-alpine species. *Ann. Bot.* **108**, 241–252.

Barkoulas, M., Hay, A., Kougioumoutzi, E. and Tsiantis, M. (2008) A developmental framework for dissected leaf formation in the *Arabidopsis* relative *Cardamine hirsuta*. *Nat. Genet.* **40**, 1136–1141.

Beck, J.B., Schmuths, H. and Schaal, B.A. (2008) Native range genetic variation in *Arabidopsis thaliana* is strongly geographically structured and reflects *Pleistocene glacial* dynamics. *Mol. Ecol.* **17**, 902–915.

Beilstein, M.A., Al-Shehbaz, I.A., Mathews, S. and Kellogg, E.A. (2008) Brassicaceae phylogeny inferred from *phytochrome A* and *ndhF* sequence data: tribes and trichomes revisited. *Am. J. Bot.* **95**, 1307–1327.

Beilstein, M.A., Nagalingum, N.S., Clements, M.D., Manchester, S.R. and Mathews, S. (2010) Dated molecular phylogenies indicate a Miocene origin for *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA*, **107**, 18724–18728.

Blein, T., Pulido, A., Vialette-Guiraud, A., Nikovics, K., Morin, H., Hay, A., Johansen, I.E., Tsiantis, M. and Laufs, P. (2008) A conserved molecular framework for compound leaf development. *Science*, **322**, 1835–1839.

Couvreur, T.L., Franzke, A., Al-Shehbaz, I.A., Bakker, F.T., Koch, M.A. and Mummenhoff, K. (2010) Molecular phylogenetics, temporal diversification, and principles of evolution in the mustard family (Brassicaceae). *Mol. Biol. Evol.* **27**, 55–71.

Davies, J.M. (1949) The aerodynamics of golf balls. *J. Appl. Phys.* **20**, 821–828.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797.

Evans, M.L., Ishikawa, H. and Estelle, M.A. (1994) Responses of Arabidopsis roots to auxin studied with high temporal resolution – comparison of wild-type and auxin-response mutants. *Planta*, **194**, 215–222.

Excoffier, L., Smouse, P.E. and Quattro, J.M. (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.

Fain, G., Atland, J. and Rinehart, T. (2005) Molecular and morphological characterization of Cardamine species. In Southern Nursery Association Research Conference, pp. 454–456.

Frankel, N., Erezyilmaz, D.F., McGregor, A.P., Wang, S., Payre, F. and Stern, D.L. (2011) Morphological evolution caused by many subtle-effect substitutions in regulatory DNA. *Nature*, **474**, 598–603.

Gompel, N., Prud'homme, B., Wittkopp, P.J., Kassner, V.A. and Carroll, S.B. (2005) Chance caught on the wing: *cis*-regulatory evolution and the origin of pigment patterns in *Drosophila*. *Nature*, **433**, 481–487.

Haudry, A., Platts, A.E., Vello, E. *et al.* (2013) An atlas of over 90,000 conserved noncoding sequences provides insight into crucifer regulatory regions. *Nat. Genet.* **45**, 891–898.

Hay, A. and Tsiantis, M. (2006) The genetic basis for differences in leaf form between *Arabidopsis thaliana* and its wild relative *Cardamine hirsuta*. *Nat. Genet.* **38**, 942–947.

Hewitt, G.M. (2004) The structure of biodiversity – insights from molecular phylogeography. *Front. Zool.* **1**, 4.

Hofreiter, M., Serre, D., Rohland, N., Rabeder, G., Nagel, D., Conard, N., Munzel, S. and Paabo, S. (2004) Lack of phylogeography in European mammals before the last glaciation. *Proc. Natl Acad. Sci. USA*, **101**, 12963–12968.

Hu, T.T., Pattyn, P., Bakker, E.G. *et al.* (2011) The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. *Nat. Genet.* **43**, 476–481.

Jeong, S., Rebeiz, M., Andolfatto, P., Werner, T., True, J. and Carroll, S.B. (2008) The evolution of gene regulation underlies a morphological difference between two *Drosophila* sister species. *Cell*, **132**, 783–793.

Johnston, J.S., Pepper, A.E., Hall, A.E., Chen, Z.J., Hodnett, G., Drabek, J., Lopez, R. and Price, H.J. (2005) Evolution of genome size in Brassicaceae. *Ann. Bot.* **95**, 229–235.

Kim, S., Plagnol, V., Hu, T.T., Toomajian, C., Clark, R.M., Ossowski, S., Ecker, J.R., Weigel, D. and Nordborg, M. (2007) Recombination and linkage disequilibrium in *Arabidopsis thaliana*. *Nat. Genet.* **39**, 1151–1155.

Koch, M. and Bernhardt, K.G. (2004) Comparative biogeography of the cytotypes of annual *Microthlaspi perfoliatum* (Brassicaceae) in Europe using isozymes and cpDNA data: refugia, diversity centers, and postglacial colonization. *Am. J. Bot.* **91**, 115–124.

Koch, M.A., Kiefer, C., Ehrich, D., Vogel, J., Brochmann, C. and Mummenhoff, K. (2006) Three times out of Asia Minor: the phylogeography of *Arabis alpina* L. (Brassicaceae). *Mol. Ecol.* **15**, 825–839.

Lihova, J. and Marhold, K. (2006) Phylogenetic and diversity patterns in *Cardamine* (Brassicaceae) – a genus with conspicuous polyploid and reticulate evolution. In *Plant Genome: Biodiversity and Evolution* (Sharma, A.K. and Sharma, A., eds.). Enfield: Science Publishers Inc., pp. 149–186.

Lysak, M.A., Berr, A., Pecinka, A., Schmidt, R., McBreen, K. and Schubert, I. (2006) Mechanisms of chromosome number reduction in *Arabidopsis thaliana* and related Brassicaceae species. *Proc. Natl Acad. Sci. USA*, **103**, 5224–5229.

Mandakova, T., Marhold, K. and Lysak, M.A. (2013a) The widespread crucifer species *Cardamine flexuosa* is an allotetraploid with a conserved subgenomic structure. *New Phytol.* **201**, 982–992.

Mandakova, T., Kovarik, A., Zozomova-Lihova, J., Shimizu-Inatsugi, R., Shimizu, K.K., Mummenhoff, K., Marhold, K. and Lysak, M.A. (2013b) The more the merrier: recent hybridization and polyploidy in *Cardamine*. *Plant Cell*, **25**, 3280–3295.

Marcellini, S. and Simpson, P. (2006) Two or four bristles: functional evolution of an enhancer of scute in Drosophilidae. *PLoS Biol.* **4**, e386.

Martinez-Garcia, J.F., Galstyan, A., Salla-Martret, M., Cifuentes-Esquivel, N., Gallemi, M. and Bou-Torrent, J. (2010) Regulatory components of shade avoidance syndrome. *Adv. Bot. Res.* **53**, 65–116.

Morinaga, S., Nagano, A.J., Miyazaki, S., Kubo, M., Fukuda, T.D.H., Sakai, S., Hasebe, M. and Fukuda, H. (2008) Ecogenomics of cleistogamous and chasmogamous flowering: genome-wide gene expression patterns from cross-species microarray analysis in *Cardamine kokaiensis* (Brassicaceae). *J. Ecol.* **96**, 1086–1097.

Muller, A. (1961) Zur Charakterisierung der Bluten und Infloresz- enzen von *Arabidopsis fhaliana* (L.) Heynh. *Kulturpflanze*, **9**, 364–393.

Petit, R., Aguinagalde, I., de Beaulieu, J.L. *et al.* (2003) Glacial refugia: hotspots but not melting pots of genetic diversity. *Science*, **300**, 1563–1565.

Pico, F.X., Mendez-Vigo, B., Martinez-Zapater, J.M. and Alonso-Blanco, C. (2008) Natural genetic variation of *Arabidopsis thaliana* is geographically structured in the Iberian peninsula. *Genetics*, **180**, 1009–1021.

Rich, T.C.G.(1991) *Crucifers of Great Britain and Ireland*. London: Botanical Society of the British Isles c/o The Natural History Museum.

Roig-Villanova, I., Bou-Torrent, J., Galstyan, A., Carretero-Paulet, L., Portoles, S., Rodriguez-Concepcion, M. and Martinez-Garcia, J.F. (2007) Interaction of shade avoidance and auxin responses: a role for two novel atypical bHLH proteins. *EMBO J.* **26**, 4756–4767.

Schmid, K.J., Torjek, O., Meyer, R., Schmuths, H., Hoffmann, M.H. and Altmann, T. (2006) Evidence for a large-scale population structure of *Arabidopsis thaliana* from genome-wide single nucleotide polymorphism markers. *Theor. Appl. Genet.* **112**, 1104–1114.

Schranz, M.E., Lysak, M.A. and Mitchell-Olds, T. (2006) The ABC's of comparative genomics in the Brassicaceae: building blocks of crucifer genomes. *Trends Plant Sci.* **11**, 535–542.

Schranz, M.E., Windsor, A.J., Song, B.H., Lawton-Rauh, A. and Mitchell-Olds, T. (2007) Comparative genetic mapping in *Boechera stricta*, a close relative of *Arabidopsis*. *Plant Physiol.* **144**, 286–298.

Slotte, T., Hazzouri, K.M., Agren, J.A. *et al.* (2013) The *Capsella rubella* genome and the genomic consequences of rapid mating system evolution. *Nat. Genet.* **45**, 831–835.

Smyth, D.R., Bowman, J.L. and Meyerowitz, E.M. (1990) Early flower development in Arabidopsis. *Plant Cell*, **2**, 755–767.

Tajima, F. (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics*, **105**, 437–460.

Vaughn, K.C., Bowling, A.J. and Ruel, K.J. (2011) The mechanism for explosive seed dispersal in *Cardamine hirsuta* (Brassicaceae). *Am. J. Bot.* **98**, 1276–1285.

Wellmer, F., Riechmann, J.L., Alves-Ferreira, M. and Meyerowitz, E.M. (2004) Genome-wide analysis of spatial gene expression in Arabidopsis flowers. *Plant Cell*, **16**, 1314–1326.

Windsor, A.J., Reichelt, M., Figuth, A., Svatos, A., Kroymann, J., Kliebenstein, D.J., Gershenzon, J. and Mitchell-Olds, T. (2005) Geographic and evolutionary diversification of glucosinolates among near relatives of *Arabidopsis thaliana* (Brassicaceae). *Phytochemistry*, **66**, 1321–1333.

Wittkopp, P.J., Stewart, E.E., Arnold, L.L., Neidert, A.H., Haerum, B.K., Thompson, E.M., Akhras, S., Smith-Winberry, G. and Shefner, L. (2009) Intraspecific polymorphism to interspecific divergence: genetics of pigmentation in *Drosophila*. *Science*, **326**, 540–544.

Yang, Z.H. (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591.

Yang, Z. and dos Reis, M. (2011) Statistical properties of the branch-site test of positive selection. *Mol. Biol. Evol.* **28**, 1217–1228.

Yatsu, Y., Kachi, N. and Kudoh, H. (2003) Ecological distribution and phenology of an invasive species, *Cardamine hirsuta* L., and its native counterpart, *Cardamine flexuosa* with, in central Japan. *Plant Spec. Biol.* **18**, 35–42.

